Proceedings of the 12th World Congress on Genetics Applied to Livestock Production (WCGALP)

Technical and species orientated innovations in animal breeding and contribution of genetics to solving societal challenges

edited by:
R.F. Veerkamp and Y. de Haas

Wageningen Academic Publishers

Proceedings of the
12th World Congress on Genetics
Applied to Livestock Production
(WCGALP)

Proceedings of the 12th World Congress on Genetics Applied to Livestock Production (WCGALP)

Technical and species orientated innovations in animal breeding, and contribution of genetics to solving societal challenges

edited by:
R.F. Veerkamp and Y. de Haas

Wageningen Academic
Publishers

This work is subject to copyright. All rights are reserved, whether the whole or part of the material is concerned. Nothing from this publication may be translated, reproduced, stored in a computerised system or published in any form or in any manner, including electronic, mechanical, reprographic or photographic, without prior written permission from the publisher,

Wageningen Academic Publishers,

P.O. Box 220,

NL-6700 AE Wageningen,

The Netherlands.

www.WageningenAcademic.com

copyright@WageningenAcademic.com

The individual contributions in this publication and any liabilities arising from them remain the responsibility of the authors.

The publisher is not responsible for possible damages, which could be a result of content derived from this publication.

e-EAN: 9789086869404

eISBN: 978-90-8686-940-4

DOI: 10.3920/978-90-8686-940-4

First published, 2022

© Wageningen Academic Publishers
The Netherlands, 2022

Section editors

Alison Van Eenennaam
Amanda Chamberlain
Andrea Wilson
Andres Legarra
Angela Canovas
Aniek Bouwman
Anna Sonesson
Antonio Reverter Gomez
Bart Ducro
Bas Rodenburg
Beatriz Villanueva
Ben Hayes
Biaty Raymond
Birgit Gredler-Grandl
Bjorg Heringstad
Bruno Santos
Christa Kühn
Christian Maltecca
Christian Reimer
Christina Rochus
Christine Baes
Christine Nicol
Claudia Kamphuis
Daniel Brown
Daniela Lourenco
Dieuwertje Lont
Diogo Hashimoto
Dirk Jan De Koning
Egbert Knol
Eileen Wall
Elisabetta Giuffra
Emily Clark
Emmeline Hill
Erling Strandberg
Esther Ellen
Fernando Cardoso
Filippo Miglior
Florence Phocas
Gabor Meszaros
Georg Thaller
Gerben de Jong
Gertje Petersen
Goutam Sahana
Gregoire Leroy
Gregor Gorjanc
Guilherme Rosa
Gustavo de los Campos
Gustavo Gandini
Gwendal Restoux
Han Mulder
Hanne Marie Nielsen
Hans Daetwyler
Hans Komen
Harmen Doekes
Helene Gilbert
Henk Bovenhuis
Hubert Pausch
Hugo Toledo Alvarado
Imke Tammen
Ines Adriaens
Irene van den Berg
Ismo Stranden
Jaana Peippo
Jack Dekkers
Jack Windig
Jan Lassen
Jan ten Napel
Jared Decker
Jeff Schoenebeck
Jens Vanselow
Jeremie Vandenplas
Jo Newton
Joanne Conington
Johan van Arendonk
John Bastiaansen
John Cole
John Henshall
John McEwan
Jörn Bennewitz
Julie Lamy
Julie Ojango
Julius van der Werf
Karel de Greef
Karen Marshall
Kathrin Stock
Kor Oldenbroek
Luca Fontanesi
Lucia Galvão de Albuquerque
Marcin Pszczola
Marco Bink
Marie-Pierre Sanchez
Mario Calus
Mark Camara
Mark Henryon
Marta Godia
Martien Groenen
Martijn Derks
Martin Lidauer
Martino Cassandro
Mathew Littlejohn
Mehdi Sargolzaei
Michael Aldridge
Miguel Perez
Mike Coffey
Mike McGrew
Mirte Bosse
Nick Serao
Nicolas Gengler
Ole Madsen
Oscar Gonzalez
Pascal Duenk
Paul Stothard
Pim Brascamp
Piter Bijma
Rachel Rupp
Raphael Mrode
Renzo Bonifazi
Richard Crooijmans
Ron Lewis
Ross Houston
Ruiz Felipe
Samuel Aggrey
Sarita Guy
Sipke Joost Hiemstra
Susanne Hermesch
Suzanne Rowe
Tadelle Dessie
Theo Meuwissen
Tom Druet
Torsten Pook
Trudee Fair
Victor Olori
Vinzent Boerner
Wendy Rauw
Zhao Guiping

Table of contents

Section editors 4
Table of contents 5

Part 1 – Plenaries **52**

1. Measure, predict and act; a paradigm of genetics applied to livestock production now and into the future 53
M.P.L. Calus, D.P. Berry and J.E. Pryce

2. Genetics of quantitative traits in natural populations: lessons from long term studies of wild animals 57
J.M. Pemberton

3. Discoveries and lessons from a genome-wide association study of human height in >5 million individuals 61
L. Yengo on behalf of GIANT Consortium (the height working group)

Part 2 – Young scientist career development **65**

4. History and author analysis of the World Congresses on Genetics Applied to Livestock Production 66
A.L. Van Eenennaam and S. Hermesch

5. What industry needs from an animal breeding graduate 70
T.J. Byrne, P.F. Fennessy and P.R. Amer

6. Young scientist career development: a global discussion on challenges, opportunities and solutions 72
J.E. Newton, M. Gòdia, B. Raymond and M.H.O. Toledo-Alvarado

Part 3 – Education **76**

7. Digital learning in postsecondary education: 30 transformational years 77
J.J. Lindburg

8. A panel discussion: broadening learning by incorporating industry professionals in education 80
J.B. Cole, R.A. Mrode, K.M. Olson and P. Knap

9. Conducting a national beef cattle genetics outreach program in the USA 82
K.D. Bullock, M.L. Spangler, R.L. Weaber, T.N. Rowan, M.M. Rolf, J.E. Decker, D.D. Loy, B.L. Golden, J.J. White and A.L. Van Eenennaam

10. Breeding Lab: a successful industry-academia collaboration to provide MSc consultancy training 86
J. Komen, J.A.M. van Arendonk and D.E. Lont

11. How a group of Merino breeders increased the rate of genetic gain by 134% in five years! 90
T. Granleese, J.H.J. van der Werf and S.J. Martin

12. YouTube applied to the education of genetics and genomics 94
G. Mészáros

13. Creating a good learning and sharing environment for bioinformatics 98
T. Klingström and J.I. Ohlsson

14. European master in animal breeding and genetics (EMABG): A training for research, practical, and transversal skills 102
G. Mészáros, P. Berg, T. Heams, H. Komen, D. Lont, H. Simianer, J. Sölkner, E. Strandberg and J. Tetens

Part 4 – Society – genetic solutions to achieve net zero carbon emission in livestock systems **106**

15. Strategies to mitigate greenhouse gas emissions from pasture-based sheep systems – an EU project consortium view 107
J. Conington, N. Lambe, F. Tortereau, F. McGovern, E. Navajas, I. De Barbieri, G. Ciappesoni, J. Jakobsen, E. Smith, J. Yates, Q. Le Graverand, K. McDermott, G. Steinheim, B. Aspeholen Aby, I. Dønnem, N. McHugh, L. Farrell, C. Marie-Etancelin, P. Johnson and S. Rowe

16. Comparing measurements of product output when defining emissions intensity 111
C.M. Richardson, F. Hely, M. Post, T.T.T. Nguyen, G.J. Nieuwhof and P.R. Amer

17. The genetic progress has effect on greenhouse gas emission intensities from Norwegian pork production 115
E. Gjerlaug Enger, H. Bonesmo, K. Martinsen, I. Andersen-Ranberg, D. Olsen and E. Grindflek

18. An approach to the genetic parameters of methane emission in Latxa dairy sheep 119
C. Pineda-Quiroga, I. Granado-Tajada, E. Ugarte, I. Goiri and A. García-Rodríguez

19. Ameliorating the contribution of livestock to global greenhouse gas production: dream or deliverable? 123
R. Hegarty, J.C. McEwan and S.J. Rowe

20. Genetically more efficient Australian dairy cows and sheep are higher emitters of methane per unit of food 127
B.J. Sepulveda, S.K. Muir, S. Bolormaa, I.M. MacLeod, M.I. Knight, R. Behrendt, L.C. Marett, M.H. Deighton, J.B. Garner, P.J. Moate, W.J. Wales, R.O. Williams, H.D. Daetwyler, B.G. Cocks and J.E. Pryce

21. The role of livestock genetics in addressing national GHG mitigation requirements 131
A.R. Cromie, D.P. Berry, L. Shalloo, T. Kirk and J. Crowley

22. Can genomics cope with a 30% reduction of methane emission from livestock in 10 years? 135
O. González-Recio, J. López-Paredes, A. Saborío-Montero, A. López-García, M. Gutiérrez-Rivas, I. Goiri, R. Atxaerandio, E. Ugarte, N. Charfeddine, J.A. Jiménez-Montero and A. García-Rodríguez

23. Low carbon emissions in future pig production – does a new breeding goal matter? 139
S. Zira, H.F. Olsen and L. Rydhmer

24. Individual animal carbon footprints in a national beef database: AniPrint 1.0 143
J.J. Crowley, L. Shalloo, R.D. Evans, J. Herron, P. Crosson, D.P. Berry, P. Owens and A.R. Cromie

Part 5 – Novel traits: environment and greenhouse gas 147

25. Can ruminal microbial information help improve selection for low-methane emitting dairy cows? 148
L. Zetouni, S. Roques, S.K. Kar, Y. de Haas, D. Schokker and M. Aldridge

26. Phenotypic differences and genetic parameters for methane concentration in BeefxDairy crossbred slaughter calves 152
K. Johansen, M. Kargo, M. Bjerring, P. Løvendahl and A.J. Buitenhuis

27. Genome-wide association study for mid-infrared methane predictions in Walloon dairy cows 156
H. Atashi, H. Wilmot, S. Vanderick, X. Hubin, H. Soyeurt and N. Gengler

28. Genetic parameters for feed efficiency, gas emissions, oxygen consumption and wool traits in Australian Merino 160
C.B. Marques, I. De Barbieri, J. Velazco, E.A. Navajas and G. Ciappesoni

29. Genetic selection of feed efficiency and methane emissions in sheep and cattle in Uruguay: progress and limitations 164
E.A. Navajas, O. Ravagnolo, I. De Barbieri, M.I. Pravia, I. Aguilar, M.O. Lema, B. Vera, P. Peraza, C.B. Marques, J.I. Velazco and G. Ciappesoni

30. Breeding for lowered methane emissions using rumen microbial profiles in sheep 168
T.P. Bilton, R. Jordan, M.K. Hess, S.M. Hickey, N. Amyes, P. Johnson, H. Henry, H. Baird, T. van Stijn, B. Veenvliet, W. Bain, G. Pile, T. Watson, E. Sandoval, J. Peers-Adams, B. Bryson, A.S. Hess, K.G. Dodds, J.C. McEwan and S.J. Rowe

31. Do we estimate dairy cows' methane production precisely enough? 172
M. Pszczola and T. Strabel

32. Heritability of methane emission in Norwegian Red cows based on measures from GreenFeed in commercial herds 176
K.B. Wethal, G.F. Difford, K. Winnberg, E. Norberg and B. Heringstad

33. Heritability and genetic correlations of enteric methane emissions of dairy cows measured by sniffers and GreenFeed 179
A.E. van Breukelen, M.N. Aldridge, R.F. Veerkamp, L. Koning, L.B. Sebek and Y. de Haas

34. Methane emission has low genetic correlations to lamb growth traits in Norwegian White sheep 183
J.H. Jakobsen, T. Blichfeldt, L.-B. Linneflaatten, M.O. Gløersen, L.E. Wallin and J.C. McEwan

35. Life cycle assessment to predict individual environmental impacts: towards selection for sustainable pig production 187
T. Soleimani, S. Hermesch and H. Gilbert

36. Genetic parameters for growth, feed intake and greenhouse gasses emissions in Italian Holstein-Friesian bulls 191
F. Tiezzi, R. Finocchiaro, L. Benzoni, J.B.C.H.M. van Kaam, M. Marusi, R. Bozzi and M. Cassandro

37. Association of genetic resistance to internal nematodes and production traits on feed efficiency and methane emissions in Corriedale lambs 195
E.A. Navajas, G. Ciappesoni, D. Gimeno, J.I. Velazco and I. De Barbieri

38. Rumen microbiota and host gene expression associate to predisposed milk urea concentration in Holsteins 199
H. Honerlagen, H. Reyer, D. Segelke, M. Oster, S. Ponsuksili, N. Trakooljul, B. Kuhla and K. Wimmers

39. Predicting nitrogen use efficiency of individual dairy cows by mid-infrared spectra 203
R. Shi, W. Lou, B. Ducro, A. van der Linden, H.A. Mulder, S.J. Oosting, S. Li and Y. Wang

40. Clearing the air between methane and commercial beef cattle in Ireland 207
C.V. Ryan, S.F. Kirwan, J.J. Crowley, D.C. Purfield, T. Pabiou and R.D. Evans

41. The impact of selection for divergence in methane emissions on alimentary tract measures of eight-month-old sheep 211
K.M. McRae, S.M. Hickey, G.J. Greer, B. Bryson, K.J. Knowler, W.E. Bain, S.N. Elmes, E.A. Young, J. Wing, E. Waller, J.C. McEwan and S.J. Rowe

42. Economic values for climate change responsive traits to develop a sustainable dairy cattle breeding goal in the tropics 215
C.K. Chepsiror, E.D. Ilatsia and T.O. Okeno

43. Phenotypic and genetic trends in American Angus associated with climate variability 219
G. Rovere, B.C.D. Cuyabano, B. Makanjuola, S. Kelly and C. Gondro

44. Genetic variation in resilience to climate effects on beef carcass traits 223
H. Bunning and E. Wall

45. Genetic parameters for rumination time and traits related to sustainable dairy production 227
L.S.F. Lopes, K. Houlahan, L. Alcantara, G.A. Oliveira Jr., D. Tulpan, F. Miglior, F.S. Schenkel and C.F. Baes

46. Genotype by temperature-humidity index interactions on milk production and udder health traits in the Montbeliarde cows 231
A. Vinet, S. Mattalia, R. Vallée, C. Bertrand, P. Bertuzzi, B.C.D. Cuyabano and D. Boichard

Part 6 – Challenges – resource allocation and genetics of feed intake and efficiency 235

47. Modelling multiplicativity in feed efficiency by regression on expected feed intake 236
M.H. Lidauer, E. Negussie, T. Mehtiö, P. Mäntysaari, T. Kokkonen and E.A. Mäntysaari

48. Alternative models to predict residual feed intake in Hereford breed and effects on their breeding values accuracy 240
M.I. Pravia, E.A. Navajas, I. Aguilar and O. Ravagnolo

49. Feed intake can be predicted as quantitative or qualitative traits 244
F. Tortereau, C. Marie-Etancelin, D. Marcon and J.L Weisbecker

50. Prediction of breeding values for feed intake in pigs using individual versus group records along with correlated traits 248
C. Zhang, R.A. Kemp, J.C.M. Dekkers, G.S. Plastow and H. Gao

51. Phenotypic and genetic associations between feed efficiency and FTIR milk-spectra 252
H.O. Toledo-Alvarado, R.J. Tempelman, M. Lopez-Cruz, M.J. VandeHaar, J.E.P. Santos, F. Peñagaricano, P. Khanal and G. de los Campos

52. Genetic relationships between weight loss in early lactation and daily milk production until 305 days in Holstein cows 256
T. Tribout, S. Minéry, R. Vallée, S. Saille, D. Saunier, V. Ducrocq, P. Faverdin and D. Boichard

53. Influence of body condition score genomic index on performance trajectories over the lactation period in Holstein cows 260
R. Lefebvre, P. Faverdin, S. Barbey, J. Jurquet, T. Tribout, D. Boichard and P. Martin

54. Feed efficiency and resource allocation trade-offs: theory, evidence and prospects 264
F. Douhard, R. Rupp and H. Gilbert

55. Mechanistic-based prediction of selection response on resilience and feed efficiency traits in dairy cattle 268
A. Bouquet, M. Slagboom, J. Thomasen, N.C. Friggens, M. Kargo and L. Puillet

56. Towards the characterisation of animal robustness by dynamic energy allocation indicators in fattening pigs 272
G. Lenoir, R. Muñoz-Tamayo, L. Flatres-Grall, I. David and N.C. Friggens

57. Milk metabolite profiles in goats selected for longevity support link between resource allocation and resilience 276
M. Ithurbide, H. Wang, C. Huau, I. Palhière, T. Fassier, J. Pires, T. Larsen, J. Cao, N.C. Friggens and R. Rupp

58. Lifetime feed efficiency and deep phenotypes from scarce feed intake records using the mechanistic LiGAPS-Dairy model 280
A. van der Linden, G.C.B. Schopen, B. Gredler-Grandl, G. de Jong, M. Sol, S. van der Beek and R.F. Veerkamp

59. Genetic analysis of feed efficiency in Atlantic salmon 284
A.K. Sonesson, B. Hatlen, B. Gjerde, G.F. Difford, P. Brunsvik, T.H.E. Meuwissen and A. Norris

60. Validation and genetic analysis of a feed efficiency criterion in French Lacaune ewes 288
C. Machefert, C. Robert-Granié, P. Hassoun, S. Parisot, C. Allain, D. Portes, G. Lagriffoul, J.M Astruc, and H. Larroque

61. Gene co-expression networks contributing to variation in residual feed intake in bovine hepatic tissue 292
K. Keogh, M. McGee and D.A. Kenny

62. Genetics parameters of feed efficiency in dairy goats, under commercial conditions 296
M. Chassier, F. Mosnier, R. Rupp, B. Bluet, A. Bailly-Salins and I. Palhière

63. Genetic analysis of predicted dry matter intake and gross feed efficiency in South African Holstein cows 300
M.A. Madilindi, O.T. Zishiri, B. Dube and C.B. Banga

64. Single-step genomic prediction for metabolic body weight of Nordic Red dairy cattle 304
R.M. Kempe, T. Mehtiö, A.-M. Leino, M. Koivula, T. Pitkänen, R.B. Stephansen, E.A. Mäntysaari, E. Negussie, J. Pösö, U.S. Nielsen, W.F. Fikse, G.P. Aamand and M.H. Lidauer

65. Application of imputed data and single-step GBLUP in the study of residual feed intake trait of Nellore cattle 308
J.A.II.V. Silva, M.H.V. Oliveira, A.M. Castilhos, A.M. Toro-Ospina, H. Srihi and L. Varona

66. Relationship between proxies of energy states and nitrogen use efficiency for Holstein cows in early lactation 312
Y. Chen, C. Grelet, S. Vanderick, H. Hu and N. Gengler

67. Genetic variation exists in nitrogen use efficiency for lactating cows in grass-grazed dairy systems 316
E. Tavernier, I.C. Gormley, L. Delaby, S. McParland, M. O'Donovan and D.P. Berry

68. Study of the causal relationship between the rabbit cecal microbiota and host's growth and feed efficiency 320
M. Piles, M. Mora, J.P. Sánchez, Y. Ramayo-Caldas, O. González-Rodríguez and M. Velasco-Galilea

69. Part of the functional rumen core microbiome is influenced by the bovine host genome and associated with feed efficiency 324
M. Martínez-Álvaro, J. Mattock, Z. Weng, R.J. Dewhurst, M.A. Cleveland, M. Watson and R. Roehe

70. Identifying diversity in cattle performance using Bayesian inference and a model 328
J.F. Ramirez, L. Puillet and N.C. Friggens

71. Feed efficiency of crossbreds of Montbéliarde, Viking Red, and Holstein with Holstein cows fed two alternative diets 332
B.J. Heins, G.M. Pereira and L.B. Hansen

72. Genetic and phenotypic associations among production, feed efficiency and immune traits in indigenous chicken of Kenya 336
S.A. Miyumo, C.B. Wasike, E.D. Ilatsia, J. Bennewitz and M.G.G. Chagunda

73. Novel genetic parameters to improve gRFI in dairy cattle using big data from multiple lactations and countries 340
R.B. Stephansen, P. Martin, B. Gredler-Grandl, C.I.V. Manzanilla Pech, G. Sahana, V. Ducrocq, C. Baes, F. Schenkel, R.F. Veerkamp and J. Lassen

74. Genetic parameters of feeding behaviour traits in parental lines of a mule duck 344
H. Chapuis, M. Lague and H. Gilbert

75. Estimates of genetic parameters for feeding behavior traits and its association with feed efficiency in Holstein cows 348
L. Cavani, W.E. Brown, K.L. Parker Gaddis, R.J. Tempelman, M.J. VandeHaar, H.M. White, F. Peñagaricano and K.A. Weigel

76. Protein efficiency is heritable but not genetically correlated with meat quality in pigs 352
E.O. Ewaoluwagbemiga, G. Bee and C. Kasper

77. Genetic analysis of digestibility traits in pigs measured by near-infrared spectroscopy 356
K.H. Martinsen, E.G. Enger, E. Grindflek, N.K. Afseth, L.T. Mydland, S. Wallen and T.H.E. Meuwissen

Part 7 – Society – contributions animal breeding to animal welfare 360

78. Genetic analysis of number of functional teats, piglet survival, and piglet weight in organic pigs 361
R.M. Zaalberg, T.T. Chu, J. Jensen and T.M. Villumsen

79. Genetic analysis of feeding behaviour and tail lesions 365
K.H. Martinsen, Ø. Nordbø, E.G. Enger, D. Olsen and E. Grindflek

80. Response to selection for social dominance including direct and indirect genetic effects 369
B. Tuliozi, E. Mancin, G. Gomez Proto, R. Mantovani and C. Sartori

81. Empowerment of the Italian small ruminant sector: new traits and tools toward a sustainable management 373
A. Negro, M. Cortellari, A. Bionda, S. Grande, A. Cesarani, A. Carta, N. Macciotta, S. Biffani and P. Crepaldi

82. How can breeding for improved animal welfare meet societal expectations and challenges? 376
A. Wallenbeck

83. Swine inflammation and necrosis syndrome (SINS) and its association with biting behavior after weaning 380
N.L. Leite, E.F. Knol, S. Nuphaus, S. Tsuruta, R. Vogelzang and D. Lourenco

84. Breeding for future social license for beef production 384
S.P. Miller and K.J. Retallick

85. Selection for piglet mortality and litter size in outdoor organic pig production systems 388
T.T. Chu, R.M. Zaalberg, J. Jensen and T.M. Villumsen

86. Heritability of daily activity over time in broilers 393
E.D. Ellen, M. van der Sluis, B. de Klerk, J. Henshall, Y. de Haas, T.B. Rodenburg and P. Bijma

Part 8 – Novel traits: health and well-being **397**

87. Resilience indicators based on individual feed intake from a high-health swine nucleus 398
C.D. Bierman, R.A. Kemp and J.C.M. Dekkers

88. Comparison of milk yield based resilience indicators across dairy cattle breeds 402
G. Bonekamp, M. Poppe, J. ten Napel, C. Kamphuis, Y. de Haas and I. Adriaens

89. Genotype-by-environment interactions for health traits across productivity level or health status of herds in Normande 406
S. Aguerre, A. Vinet and S. Mattalia

90. In search of a new tool for phenotyping tick resistance in cattle 410
O. Matika, S. Foster, N. Githaka, C. Mwendia, H. Brown, K. Watson, A. Djikeng and M. Birkett

91. Lameness in genomic evaluation for foot and claw disorders in Czech Holstein cows 414
L. Zavadilová, E. Kašná and Z. Krupová

92. The lasting effect of udder inflammation on milk lactose content in dairy cows 418
A. Costa, H. Bovenhuis, C. Egger-Danner, B. Fuerst-Waltl and M. Penasa

93. Estimation of genetic parameters for milk-related traits derived from automatic milking systems in Holstein cattle 422
L.S. Gloria, J.P. Boerman, V.B. Pedrosa, S.Y. Chen, M.E. Montes Gonzalez, J.S. Doucette and L.F. Brito

94. Identification of genome regions and promising candidate genes linked to innate immune capacity on young Holstein calves 426
P. Martin, S. Fresco, L. Jouneau, S. Taussat, A. Vinet, R. Lefebvre, D. Descamps, C. Ferret, C. Chottin, V. Pietralunga, T. Pezier, S. Lacroix-Lamandé, M. Boussaha, S. Barbey, V. Denizot, D. Boichard, S. Riffault and F. Laurent

95. Assessment of miRNAs in plasma and leucocytes as potential biomarkers for rumen health in cows fed forage and high-grain diets 430
O.E. Ojo, C. Pacífico, S. Johanns, S. Ricci, R. Rivera-Chacon, A. Sener-Aydemir, E. Castillo-Lopez, N. Reisinger, Q. Zebeli and S. Kreuzer-Redmer

96. Variance component estimation of longitudinal alpha diversity metrics of the ocular microbiome in preweaned beef cattle 434
A.D. Lakamp, A.C. Neujahr, M.M. Hille, J.D. Loy, S.C. Fernando and M.L. Spangler

97. Gene expression patterns observed from the skin of selected cattle breeds infested with *Amblyomma hebraeum* ticks 438
M. Letuka, D. Wragg, A. Bester-van der Merwe, K. Dzama, L.T. Nesengani, O. Matika, B. Dube and N.O. Mapholi

98. Combined effect of genetics and gut microbiota on variations in vaccine response in hens 442
A. Lecoeur, F. Blanc, D. Gourichon, N. Bruneau, T. Burlot, F. Calenge and M.H. Pinard-van der Laan

99. Resilience indicators based on daily milk yield data for genetic selection in dairy cattle 446
M. Poppe, H.A. Mulder, H. Hogeveen, C. Kamphuis, G. Bonekamp, M.L. van Pelt, E. Mullaart and R.F. Veerkamp

100. Genetic evaluation of heat tolerance in Italian Holstein breed 450
R. Finocchiaro, F. Galluzzo, J.B.C.H.M van Kaam, M. Marusi and M. Cassandro

101. Inferring genetic parameters for prenatal heat stress effects on calf diseases and cow productivity 454
T. Yin, K. Halli, S.F. Vanvanhossou, M. Bohlouli and S. König

102. A genetic evaluation for calf health: preliminary analysis 458
C. Lynch, F. S. Schenkel, K. Houlahan, N. van Staaveren, F. Miglior and C.F. Baes

103. Development of a single step genomic evaluation of health traits and relationships with dairy traits in French Holsteins 462
H. Leclerc, A. Barbat, R. Vallée, F. Guillaume, L. Manciaux, J.-B. Davière and M. Philippe

104. Genetic selection for reduced susceptibility to subclinical ketosis using mid-infrared predicted traits 466
A. Köck, L.M. Dale, A. Werner, M. Mayerhofer, F.J. Auer and C. Egger-Danner

105. Heritability of scurs in Wagyu and tropically adapted composite beef cattle 470
G. Moser and M.J. Kelly

106. Genetic parameters for social hierarchy of Nellore cattle using data from electronic feeding system 474
J.P.S. Valente, M. Deniz, K.T. de Sousa, R.C. Canesin, S.F.M. Bonilha, M.E.Z. Mercadante and L.T. Dias

107. Pigs' weighing order stability is favourably linked with tail biting and production: a new proxy for social structure? 478
W. Gorssen, R. Meyermans, S. Janssens and N. Buys

108. Application of machine learning algorithms to develop behavioural and stress phenotypes for sheep 482
H. Keshavarzi, D.L.M. Campbell and C. Lee

109. Heritability of social interactions in dairy cattle 486
L. Rönnegård, I. Hansson and W.F. Fikse

110. Uncertainty about the social network among dairy cows impairs estimation of indirect genetic effects 490
W.F. Fikse, I. Hansson and L. Rönnegård

111. Genetic and phenotypic variation and consistency of cow's preference and circadian use of robotic milking units 493
P. Løvendahl and A.J. Buitenhuis

112. Genetic and environmental influences on maternal calving difficulty in South African Holstein cattle 497
N.E. Mkhize, N.O. Mapholi, B. Dube, M.A. Madilindi, O. Matika and C.B. Banga

113. Genetic evaluation of gestation length in Italian Holstein breed 500
F. Galluzzo, J.B.C.H.M. van Kaam, R. Finocchiaro, V. Ferrari, M. Marusi and M. Cassandro

114. Genetic evaluation of cystic ovarian disease in Czech Holstein cattle 504
E. Kašná, L. Zavadilová and Z. Krupová

115. Entropy of broiler activity: individual variation and consistency 508
M. van der Sluis, L. Asher, T.B. Rodenburg, Y. de Haas, B. de Klerk and E.D. Ellen

116. Laying hens tweet: they also lead and follow – implications for genetic selection against gregarious nesting 512
L. Bécot, N. Bédère, T. Burlot and P. Le Roy

117. Analysis of social behaviors in large groups: simulation and genetic evaluation 516
Z. Wang, H.P. Doekes and P. Bijma

118. Divergent genetic selections for social attractiveness or tolerance toward humans in sheep 520
D. Hazard, E. Delval, S. Douls, C. Durand, G. Bonnafe, D. Foulquié, D. Marcon, C. Allain, S. Parisot and A. Boissy

119. Genetic parameters of novel behaviour traits derived from social network analysis in pigs 524
S. Agha, S. Foister, R. Roehe, S.P. Turner and A. Doeschl-Wilson

120. Genetic parameter estimation of various body temperature and respiration rate indicators in maternal-line pigs under heat stress conditions 528
H. Wen, J.S. Johnson, P.H.F. Freitas, J.M. Maskal, M.K. Byrd, F. Tiezzi, Y. Huang, A.P. Schinckel and L.F. Brito

121. Genetics of body weight and length of Large White piglets and relationship to maternal performance 532
L. Canario, Y. Billon, W. Hébrard and J. Demars

Part 9 – Novel phenotyping tools: machine learning **536**

122. Automated individual walking distance of group-housed broilers: a comparison between ground-truth, RFID, and video 537
J.E. Doornweerd, I. Fodor, G. Kootstra, R.F. Veerkamp, B. de Klerk, M. van der Sluis, A.C. Bouwman and E.D. Ellen

123. Automated processing and phenotype extraction of ovine medical images using deep learning and computer vision 541
J.F. Robson, S.J. Denholm and M. Coffey

124. VTag: automatic pipeline to annotate video data for pig phenomics studies 545
C.J. Chen, G. Morota and H. Cheng

125. Milk spectra selection for equations development to predict milk technological traits 549
M. Frizzarin, I.C. Gormley, A. Casa and S. McParland

126. Using supervised machine learning for honey harvest prediction 553
J. Ramirez-Diaz, A. Manunza, T.A. De Oliveira, T. Bobbo, V. Silva, P. Cozzi, S. Biffani, A. Stella and G. Minozzi

127. Validation of computer vision algorithms for classifying video segments applied to behavioural phenotyping of pigs 557
J.P. Steibel, J. Han, C. Chen, J. Siegford, T. Norton and D. Colbry

128. Automated estimation of pose features in broilers using computer vision 561
I. Fodor, J.E. Doornweerd, B. de Klerk and A.C. Bouwman

129. Predicting carcass cut yields in cattle from digital images using artificial intelligence 565
D. Matthews, T. Pabiou, R.D. Evans, C. Beder and A. Daly

130. Digital phenotyping to improve Atlantic salmon feed efficiency and welfare 569
G.F. Difford, S.K. Lindberg, K. Heia, B. Hatlen, E. Durland, B. Gjerde, R. Alvestad, V. Mota, C. Noble, A. Norris, S. Boison, N.K. Afseth, J.P. Wold and A.K. Sonesson

131. Prediction of roughage intake of dairy cows combining milk mid-infrared spectra and cow variables by deep learning 573
G.T. Gebregiwergis, N.K. Afseth, A. Kidane, E. Prestløkken and T.H.E Meuwissen

132. CFIT – Cattle Feed InTake – a 3D camera based system to measure individual feed intake and predict body weight in commercial farms 577
J. Lassen, J.R. Thomasen and S. Borchersen

133. Video-based analysis of dairy cow behaviour: detection of lying down and standing up 581
I. Adriaens, W. Ouweltjes, I. Hulsegge and C. Kamphuis

134. Image-based cattle conformation prediction using deep learning methods 586
D. Segelke, H. Alkhoder and J. Wabbersen

135. Prediction fat percentage and visceral weight from whole fish images with a multi-input neural network 590
Y. Xue, J.W.M. Bastiaansen and H. Komen

136. A case-control GWAS for the chicken plumage colour using a computer vision approach 594
S. Heo, S. Cho, J. Park, M. Kim, E. Cho, D. Jin, J. Cha, Y.J. Koh and J.H. Lee

137. Machine learning to predict somatic cell count at the subsequent test-day record in the Italian Mediterranean Buffaloes 598
T. Bobbo, R. Matera, G. Pedota, J. Ramirez-Diaz, A. Manunza, A. Stella, A. Cotticelli, G. Neglia and S. Biffani

138. Computer vision systems to advance high-throughput phenotyping in livestock 602
J.R.R. Dorea and G.J.M. Rosa

139. Tracking multiple cows simultaneously in barns using computer vision and deep learning 606
C. Kamphuis, I. Adriaens, W. Ouweltjes and I. Hulsegge

140. A computer vision system for individual tracking of group housed rabbits 610
J.P. Sánchez, I. Muñoz, O. González, M. Pascual, O. Perucho, P. Alsina and M. Piles

141. Clustering of Meishan and Large White sows for activity defined from image analysis, and influence on piglet mortality 614
O. Girardie, M. Bonneau, Y. Billon, J. Bailly, I. David, L. Canario

142. Using machine learning to predict feed intakes of meat sheep from animal traits and ruminal microbiota618
Q. Le Graverand, C. Marie-Etancelin, J.L. Weisbecker, A. Meynadier, D. Marcon and F. Tortereau

143. The magic of predicting shapes from DNA 622
M. Pérez-Enciso and L.M. Zingaretti

Part 10 – Novel phenotyping tools: sensor devices 626

144. Estimates of genetic parameters for milkability traits derived from automatic milking systems in North American Holstein cattle 627
V.B. Pedrosa, J.P. Boerman, L.S. Gloria, S.Y. Chen, M.E.M. Gonzalez, J.S. Doucette and L.F. Brito

145. Classifying the likelihood of conception in dairy cow with milk mid-infrared spectra before the first insemination 631
W. Lou, R. Shi, B. Ducro, A. van der Linden, H.A. Mulder, S.J. Oosting, L. Liu and Y. Wang

146. Use of tracking technology for phenotyping laying hens within cage-free housing 635
C.M. Montalcini, M.B. Petelle, S.G. Gebhardt-Henrich and M.J. Toscano

147. Learning behaviour of dairy cows in automated milking systems: genetic parameters and suggested candidate genes 639
R. Schafberg, F. Rosner, D. Oelschlägel and H.H. Swalve

148. Repeatability estimates of grazing and rumination activity of Merino sheep measured using wearable sensors 643
F. Almasi, M. Khansefid, H. Nguyen, A. Desai, J.E. Pryce and M. Stear

149. Deep data recording for developing resilience traits in Swiss cattle populations: a proof of concept and an application 647
B. Bapst, P. von Rohr and F.R. Seefried

150. Test-day genetic evaluations: a tool to measure herd resilience through monthly milk records 651
S. Mattalia, M. Rocland, S. Aguerre and M. Poppe

151. Genomic selection of postpartum anoestrus recorded with accelerometer collars to improve beef cattle fertility 655
E. Fontanillas, V. Prieur, A. Michenet, G. Auvray and J. Mante

152. Developing precision livestock farming in practice: using sensor time series data for breeding decision support systems 659
J.M.E. Statham, K.L. Burton, I. Adriaens, I. Lora, G. Cozzi, Y. de Haas, C. Kamphuis, L. Vedder, B. Loke and N.C. Friggens

Part 11 – Challenges – genetic control of infectious disease transmission and impact 663

153. Molecular mechanisms involved with influenza A nanovaccine immunogenicity in pigs 664
A.M.G. Ibelli, J.O. Peixoto, L.T. Fernandes, V. Haach, D. Gava, A.P. Bastos, F.N. Fonseca, M A.Z. Morés, M.E. Cantão, R. Schaefer and M.C. Ledur

154. Host variation in susceptibility to lumpy skin disease: gene expression analyses of experimentally infected cows 668
S. Van Borm, T. Klingström, A. Niazi, K. De Clercq, L. Mostin, A. Haegeman and D.J. de Koning

155. Relationship between resistance to a *Pasteurella multocida* experimental infection and production traits in rabbits 672
H. Garreau, F. Lantier, B. Bed'hom, E. Guitton, V. Helies, E. Helloin, C. Herbert, M. Maupin, R. Robert and M. Gunia

156. Quantification of environmental management systems for U.S. Katahdin sheep producers 676
B.C. Arisman, J.M. Burke, J.L.M. Morgan and R.M. Lewis

157. Bias and heritability of the autocorrelation based on longitudinal data used as resilience indicator 680
H.A. Mulder

158. The possibilities of estimating relationships between traits with and without micro-genetic environmental sensitivity 684
M.D. Madsen, J.H.J. van der Werf and S. Clark

159. Selective breeding for survival to *Vibrio harveyi* in European seabass (*Dicentrarchus labrax*) is robust to different feeding environments 688
M. Besson, R. Morvezen, A. Bestin, Y. François, A. Bajek, J. Brunier, T. Morin, M. Vandeputte, M. Dupont-Nivet, M. Izquierdo, D. Barkas and P. Haffray

160. Can we prevent pathogen adaptation when breeding disease resistant livestock? 692
A.D. Hulst, P. Bijma and M.C.M. De Jong

161. Indicators of disease resilience from complete blood count and *in vitro* immunoassays data from young-healthy pigs 696
X. Bai, J. Cheng, F. Fortin, J.C.S. Harding, M.K. Dyck, J.C.M. Dekkers, C.J. Field, C. Rogel-Gaillard, F. Blanc and G.S. Plastow

162. Evidence of naturally occurring genetic variation in host response to challenge with a highly pathogenic PRRSV strain 700
J.R. Dunkelberger, P.K. Mathur, E. Little, D. Hanson, S. Dee and E.F. Knol

163. A quantitative genetic theory for infectious diseases 704
P. Bijma, A.D. Hulst and M.C.M. de Jong

164. Breeding for general disease resistance: a selection experiment in rabbits 708
M. Gunia, J. Ruesche, P. Aymard, E. Gillet, C. Herbert, V. Helies, D. Savietto, R. Robert, L. Warin, H. Gilbert and H. Garreau

165. A single step genetic evaluation including causal candidate SNPs for resistance to paratuberculosis in Holstein cattle 712
M.P. Sanchez, T. Tribout, S. Fritz, R. Guatteo, C. Fourichon, L. Schibler, A. Delafosse and D. Boichard

166. Genetic parameters and QTL mapping for antibody response to paratuberculosis in a natural infected flock of sheep 716
M.G. Usai, S. Casu, T. Sechi, M.G. Cancedda, D. Pintus, S. Salaris, G. Mulas and A. Carta

167. The genetic basis of sex differences in immune responses in wild Soay sheep (*Ovis aries*) 720
E.A. Wimborne, A.M. Sparks, K. Watt, R. Sinclair, J.G. Pilkington, J.M. Pemberton, T.N. McNeilly, D.H. Nussey and S.E. Johnston

168. New tools and insights to enable breeding for reduced disease transmission 724
A. Doeschl-Wilson, G. Marion and C.M. Pooley

169. Transmission experiment in turbot shows high genetic variation in host infectivity affecting disease spread and survival 728
J.C. Prentice, C.M. Pooley, S. Tsairidou, R. Pong Wong, O. Anacleto, M. Saura, R.I. Bailey, G. Marion, B. Villanueva and A.B. Doeschl-Wilson

170. Impact of selection for disease resistance on the spread of infection in a simulated aquaculture population 732
S. García-Ballesteros, J. Fernández, A. Doeschl-Wilson and B. Villanueva

Part 12 – Challenges – across species and breed genomics genetics 736

171. Application of haplotype relationship matrices for genomic prediction in purebred and crossbred cows 737
M. Khansefid, M.H. Ferdosi, M.E. Goddard, M. Haile-Mariam, C. Schrooten, G. de Jong, E. O'Connor, H.D. Daetwyler, J.E. Pryce and I.M. MacLeod

172. *Bos taurus* and *Bos indicus* reduced representation bisulfite sequencing comparison showed high epigenetic variability in *Bos* subspecies 741
E. Capra, B. Lazzari, M. Milanesi, G.dP. Nogueira, J.F. Garcia, P. Ajmone Marsan and A. Stella

173. Evaluating the use of genotype information for the prediction of breed composition in purebred and crossbred cattle 745
C.A. Ryan, D.P. Berry, T. Pabiou and D.C. Purfield

174. Genomic prediction using breed origin of alleles model accounting for probabilities in the assignment of the alleles 749
A. Guillenea, G. Su, M.S. Lund and E. Karaman

175. Making the most of genetic relationships between Iberian dairy sheep breeds: exploring multi-breed genomic evaluation 753
I. Granado-Tajada, C. Pineda-Quiroga, E. Ugarte, M. Serrano, M.A. Jiménez, H. Marina, J.J. Arranz and M. Ramón

176. Online Mendelian Inheritance in Animals (OMIA) – future proofing of a globally used animal genetics knowledgebase 757
I. Tammen, M. Mather, Z. Li, J. Nothman, D.P. Vanichkina and F.W. Nicholas

177. Predicting the purebred-crossbred genetic correlation from phenotype and genotype data of parental lines in pigs 761
P. Duenk, Y.C.J. Wientjes, P. Bijma, M.S. Lopes and M.P.L. Calus

178. Multi-breed genomic predictions from commercial crossbred data 765
B.J Hayes, J. Copley, E. Dodd, E.M. Ross and G. Fordyce

179. Simultaneous mapping of quantitative trait loci in multiple subspecies using hybrid populations 769
C.L. Warburton, R.C. Costilla, B.N. Engle and B.J. Hayes

180. Strategies to find predictive variants that improve multi-breed genomic prediction in dairy cattle 773
I. van den Berg, R. Xiang, M. Khansefid, M.E. Goddard, J.E. Pryce, C.P. Prowse-Wilkins, A.J. Chamberlain and I.M. MacLeod

181. Impact of breed ancestry on the performance of crossbred dairy cattle 777
M.A. Jaafar,B.J. Heins, C. Dechow, C.P. Van Tassell and H.J. Huson

182. DNA pooling is a cost effective method of including commercial crossbred data in selection of purebreds 781
M.N. Aldridge, J. Marjanovic, J.M. Henshall, B. de Klerk, K. Peeters and Y. de Haas

Part 13 – Challenges – understanding and utilizing genetic diversity 785

183. Genetic variability in the individual inbreeding load: genomic prediction for artificial purging 786
L. Varona, D. López-Carbonell, H. Srihi, M. Ramírez, P. Martínez, M. Hermida, J. Altarriba and J. Casellas

184. Inbreeding depression in livestock: comparing trait groups and inbreeding measures 790
H.P. Doekes, P. Bijma and J.J. Windig

185. Global and local genome ancestry reveals signal of Merinization in Noire de Thibar sheep 794
I. Baazaoui, E. Ciani, S. Mastrangelo and S. Bedhiaf-Romdhani

186. Common heterozygosity-rich regions across the genomes of commercial and local goat breeds 798
F. Biscarini, A. Manunza, P. Cozzi and A. Stella

187. Markhor-derived introgression of *PAPSS2* confers high-altitude adaptability in Tibetan goats 802
C. Li, Y. Wu, B. Chen, Y. Cai, J. Guo, P. Kalds, Y. Chen, H. Pausch, J. Han, Y. Jiang and X. Wang

188. Patterns of inbreeding and selection using runs of homozygosity in North American dairy cattle 806
E.A. Lozada-Soto, F. Tiezzi, J.B. Cole, P.M. VanRaden and C. Maltecca

189. Exploring genetic diversity and selection signatures in indigenous African cattle breeds 810
J. Friedrich, R. Bailey, A. Talenti, K. Marshall, J. Prendergast and P. Wiener

190. Understanding and utilizing genetic diversity in Dual-Purpose Blue: genome-wide association for type traits 814
N. Gengler, H. Wilmot, S. Vanderick and H. Atashi

191. Genetic background of productive and adaptive features in the West African indigenous cattle breeds in Benin 818
S.F.U. Vanvanhossou, T. Yin, I.J. Giambra, K. Brügemann, L.H. Dossa and S. König

192. Genetic diversity and population structure of Slovenian local breed Drežnica goat 822
N. Pogorevc, M. Simčič, N. Khayatzadeh, J. Sölkner, B. Berger, D. Bojkovski, M. Zorc, P. Dovč, I. Medugorac and S. Horvat

193. Genetic architecture of a composite beef cattle population 826
E. Hay, S. Toghiani, A.J. Roberts, T. Paim, L.A. Kuehn and H.D. Blackburn

194. Variation in patterns of recombination result in genetic variation in intrachromosomal shuffling in the domestic pig 830
C. Brekke, S.E. Johnston, A.B. Gjuvsland and P. Berg

195. Identification of runs of homozygosity islands and functional analysis for Caqueteño Creole cattle 834
A.M. Toro-Ospina, J.A.II.V. Silva, G.P. Schettini, V.H. Vallejo and E.G. Ortiz

196. Candidate genes related to signatures of selection for body weight and harvest length in American mink 837
D.N. Do, K. Karimi, G. Hu and Y. Miar

197. Maintenance of genetic diversity in subdivided populations using different genomic coancestry matrices 841
E. Morales-González, B. Villanueva, M.Á. Toro and J. Fernández

198. Long-term changes in genetic mean and genic variance and underlying allele frequencies in breeding programmes 845
G.A. de Oliveira Junior, M. Spehar, F.S. Schenkel, C.F. Baes and G. Gorjanc

199. The long-term effects of genomic selection: allele frequency changes and fixation of loci 849
Y.C.J. Wientjes, P. Bijma, J. van den Heuvel, B.J. Zwaan, Z.G. Vitezica and M.P.L. Calus

200. The effect of recombination rate on genomic selection in simulation 853
M. Johnsson, A.M. Johansson, T. Klingström and E. Rius-Vilarrasa

201. Fine-scale study of meiotic recombination in cattle 857
C. Oget-Ebrad, G.C.M. Moreira, L. Karim, W. Coppieters, C. Charlier, M. Georges and T. Druet

202. Sex differences in recombination maps are associated with differential hotspot usage in sheep 861
H. Vassilief, M. Id Bella, D. Hazard, F. Tortereau, T. Faraut, S.E. Johnston and B. Servin

203. Is it possible to differentiate meat products of a local breed from those of its sister breed based on genotypes? 865
H. Wilmot, G. Glorieux, X. Hubin and N. Gengler

Part 14 – Genetic diversity and inbreeding 869

204. Assessing existing admixture in indigenous veld ecotypes and non-descript goat populations of Southern Africa 870
C. Visser and S.F. Lashmar

205. Estimation of sex chromosome inbreeding depression on milk production in cattle 874
I. Curik, H. Vostra-Vydrova, M. Shihabi, J. Sölkner and L. Vostry

206. Assessing genomic diversity and heterozygosity-rich regions of transboundary Nguni cattle of South Africa and eSwatini 878
E. van Marle-Köster, S.F. Lashmar, M. Okpeku, F.C. Muchadeyi, N.O. Mapholi and C. Visser

207. Infographics for management of inbreeding and kinship in small populations 882
M.A. Schoon, M. Neuteboom, S.J. Hiemstra, C. Danchin-Burge, J. Fernández and J.J. Windig

208. Y-chromosomal haplogroups from wild and domestic goats reveal ancient migrations and recent introgressions 886
J.A. Lenstra, VarGoats Consortium, I.J. Nijman, B.D. Rosen, P. Bardou, T. Faraut, T. Cumer, K.G. Daly, Z. Zheng, Y. Cai, H. Asadollahpour, B. Çınar Kul, W.-Y. Zhang, G. E, A. Ayin, M. Bakhtin, V.A. Bâlteanu, D. Barfield, H. Baird, B. Berger, T. Blichfeldt, G. Boink, S.R.A. Bugiwati, Z. Cai, S. Carolan, E. Clark, V. Cubric-Curik, M.I.A. Dagong, T. Dorji, L. Drew, J. Guo, J. Hallsson, S. Horvat, J. Kantanen, F. Kawaguchi, P. Kazymbet, N. Khayatzadeh, N. Kim, M. Kumar Shah, Y. Liao, A. Martínez, J.S. Masangkay, M. Masaoka, R. Mazza, J. McEwan, M. Milanesi, F. Md. Omar, Y. Nomura, N.-A. Ouchene-Khelifi, F. Pereira, G. Sahana, S. Sasazaki, A. Da Silva, M. Simčič, J. Sölkner, A. Sutherland, J. Tigchelaar, H. Zhang, Econogene Consortium, P. Ajmone-Marsan, D.G. Bradley, L. Colli, C. Drögemüller, C. Lei, H. Mannen, F. Pompanon, G. Tosser-Klopp and Y. Jiang

209. Characterization and effect of inbreeding on growth and reproductive traits in Nellore Cattle 890
L.F. Benfica, R.D. do Bem, E.R Carrara, L.F. Brito, D.A. Silva, R. Carvalheiro, J.N.S.G. Cyrillo and M.E.Z. Mercadante

210. Genomic diversity of Czech and Slovak native goat breeds 894
L. Vostry, H. Vostra-Vydrova, J. Rychtarova, B. Hofmanova, N. Moravcikova, R. Kasarda, M. Margetin, I. Drzaic, V. Cubric-Curik, J. Sölkner and I. Curik

211. Population structure of U.S. Holsteins allows for a snapshot of allele frequency changes and family specific SNPs 898
T.J. Lawlor, Y. Steyn, S. Tsuruta, Y. Masuda, D.A.L. Lourenco and I. Misztal

212. Genome-wide evaluation of runs of homozygosity in Indian Sahiwal cattle 902
S.K. Illa, S. Mumtaz, S. Mukherjee, S. Nath and A. Mukherjee

213. Following the hidden footsteps of past human migrations via the goat genome 906
A. Manunza, J. Ramirez Diaz, T. Bobbo, P. Cozzi, B. Lazzari, F. Biscarini, S. Biffani, A.M. Johansson, L. Grøva and A. Stella

214. Are autozygosity estimates from imputed genotypes reliable? 910
M. Ferenčaković, I. Curik, J.F. Garcia, J. Sölkner3 and R. Carvalheiro

215. Genetic diversity and population structure for resistance to mastitis in Braunvieh cattle 914
M.Z. Trujano-Chavez, R. Sánchez-Ramos, P. Pérez-Rodríguez and A. Ruíz-Flores

216. Evaluation of the Istrian cattle population structure by pedigree analysis 919
A. Ivanković, J. Ramljak, G. Šubara, M. Pećina, E. Šuran, Z. Ivkić and M. Konjačić

217. Population dynamics of a long-term selection experiment in chickens divergently selected for high or low body weight 923
S.J. Harrison, P.B. Siegel, C.F. Honaker and R.M. Lewis

218. Genetic management meets genomics 928
J.A. Woolliams and T.H.E. Meuwissen

219. Genetic characterization of the indigenous Sanga cattle of Namibia 932
D.A. Januarie, E.D. Cason and F.W.C. Neser

220. Hedging genetic diversity using a similarity matrix based on Mendelian sampling 936
A.A. Musa and N. Reinsch

221. Genetic basis of sexually divergent early life survival in red deer (*Cervus elaphus*) 940
L. Peters, J.M. Pemberton and S.E. Johnston

222. Changes in genomic inbreeding over half a century in Swedish Red and Swedish Holstein cattle 944
S. Eriksson, E. Strandberg and A.M. Johansson

223. Runs of homozygosity in Nguni and Bonsmara cattle populations 948
P. Soma, B.B. Kooverjee and M.M Scholtz

224. Estimation of old and new inbreeding and their effects on growth traits using pedigree and genomic information 952
P. Sumreddee, A. Ling, S. Toghiani, E.H. Hay, E. Hartono, S.E. Aggrey and R. Rekaya

225. Genome wide characterisation of runs of homozygosity on the main indigenous goat breed of Botswana 956
P.I. Monau, K. Raphaka and S.J. Nsoso

226. Candidate genes for disease, reproduction and meat quality traits in Portuguese native breeds 960
M.V.G.B. Silva, L.L. Verardo, M.A. Machado, J.C.C. Panetto, I. Carolino and N. Carolino

227. Runs of homozygosity harbour genes associated with muscle growth in Pietrain but not with fertility in Large White pigs 964
S.E. Palma-Vera, L. Derežanin, J. Fickel, H. Reyer, N. Reinsch, H. Henne and J. Schön

228. Analysis of runs of homozygosity of cattle living in different climate zones 968
L. Falchi, S. Mastrangelo, A. Cesarani, G. Senczuk, B. Portolano, F. Pilla and N.P.P. Macciotta

229. The genomic inbreeding trend over the last 25 years in three Italian pig breeds 972
G. Schiavo, S. Bovo, A. Ribani, S. Tinarelli, M. Cappelloni, M. Gallo and L. Fontanesi

230. Inbreeding and coancestry trends in Nordic Holstein 976
S. Tenhunen, J.R. Thomasen, L.P. Sørensen, G.P. Aamand, P. Berg and M. Kargo

231. Effect of genomic selection on genetic variability in chromosomal level in the Finnish Ayrshire 980
K. Sarviaho, P. Uimari and K. Martikainen

232. Genomic diversity of two indigenous Croatian goat breeds 984
I. Drzaic, V. Orehovački, L. Vostrý, I. Curik, N. Mikulec and V. Cubric-Curik

233. Consequences on the genome homozygosity of breeding program decisions to improve milk production in Manchega dairy sheep 988
A. Rubio-de Juan, C. Diaz, M.J. Carabaño, M.R. Fernandez-Santos and M. Ramón

234. Heterosis can be explained by thousands of partially recessive, deleterious mutations with very small effects at sites that are conserved across species 992
M.E. Goddard

235. Signatures of selection and runs of homozygosity in the genome of several fancy and meat rabbit breeds 996
M. Ballan, S. Bovo, G. Schiavo, M. Schiavitto, R. Negrini and L. Fontanesi

236. Investigating genetic redundancy as a source of genetic diversity and adaptability in the U.S. Holstein breed 1000
Y. Steyn, T.J. Lawlor, S. Tsuruta, Y. Masuda, A. Legarra, D.A.L. Lourenco and I. Misztal

237. Inbreeding depression and heterosis for live-weight and carcass traits in purebred and crossbred sheep 1004
P. Rafter, N. McHugh, K. McDermott and, D.P. Berry

238. Phylogenetic analysis and nucleotide diversity of 69 cattle breeds including German Black Pied cattle using WGS 1008
G.B. Neumann, P. Korkuć, D. Arends, M.J. Wolf, K. May, S. König and G.A. Brockmann

239. Partitioning of autozygosity in different age-based classes in cattle populations with different demographic histories 1012
T. Druet, L. Flori and M. Gautier

Part 15 – Genetic diversity and inbreeding: genebanking 1016

240. New FAO guidelines for the management of animal genetic resources 1017
P.J. Boettcher, P. Ajmone-Marsan, R. Baumung, J. Boes, L. Colli, C. Ginja, M. Honkatukia, J. Kantanen, G. Leroy and J.A. Lenstra

241. Genebanks for animal genetic resources in Europe: quality management system as a tool to improve their operation 1021
C. Danchin, D. Bojkovski, D.Duclos, M. Neuteboom, M. Schoon, F. Tejerina Ampudia, M. Tixier-Boichard, and S.J. Hiemstra

242. Genetic characterization of endangered indigenous pigs of Angola 1025
P. Sá, D. Santos, A. Leitão, J.M. Cordeiro, L.T. Gama and A.J. Amaral

243. Genetic diversity and inbreeding in US Jersey population and germplasm collection 1029
K. Srikanth, M. Neupane, M.A. Jaafar, C.W. Wolfe, H.J. Huson, C.P. Van Tassell and H.D. Blackburn

244. Development and utilization of the United States gene bank collection 1033
H.D. Blackburn, C.S. Wilson and C.D. Dechow

245. Mapping genetic diversity in European gene banks: preliminary results on chickens for the validation of IMAGE001 array 1037
M. Tixier-Boichard, N. Peynot , D. Duclos, S. Weigend, L. Monteagudo, A. Martinez Martinez, J.V. Delgado, R. Gonzalez Prendes, R. Crooijmans and G. Restoux

246. Conservation of an endangered pig breed using optimum contribution selection 1041
H.M. Nielsen and M. Kargo

247. IDI: an index to assess the usefulness of cryopreserved collections for the management of animal genetic diversity 1045
A. Jacques, D. Duclos, C. Danchin-Burge, M.J. Mercat, M. Tixier-Boichard and G. Restoux

248. Conservation through utilization in German Red and White dual-purpose cattle 1049
M. Slagboom, V. Milkevych, H. Liu, J.R. Thomasen, M. Kargo and C. Schmidtmann

249. Genetic rescue of small populations in the presence of deleterious variation 1053
C. Reimer, S. Weigend, J. Geibel and T. Pook

250. Genebank collections in relation to risk status in livestock species 1057
G. Leroy, H. Blackburn, C. Danchin-Burge, F. Tejerina Ampudia, S.J. Hiemstra, C. Ligda, P. Boettcher and R. Baumung

Part 16 – Challenges – use of whole genome sequence information 1061

251. Pangenomes of haplotype-resolved assemblies enable population-scale genotyping of cattle structural variation 1062
A.S. Leonard and H. Pausch

252. Comparison of methods for building pangenome graphs 1066
D. Crysnanto, A. Leonard and H. Pausch

253. Expanding the genomic toolkit: what does Oxford Nanopore sequencing have to offer? 1070
A. Hess, A. Caulton, R. Clarke, R. Brauning, K. McRae and S. Clarke

254. Successful trio-based reverse genetic screen in an endangered local cattle breed 1074
I.M. Häfliger, F.R. Seefried and C. Drögemüller

255. How does a harsh environment make wild and domestic animals evolve similarly? 1078
G. Romashov, A. Yurchenko, N. Yudin, and D. Larkin

256. Functional and association analyses of sequence data in a region of OAR 20 associated to paratuberculosis in sheep 1082
S. Casu, M.G. Usai, T. Sechi, C. Maestrale, S. Miari, S. Salaris, C. Ligios and A. Carta

257. Discovery of a genetic variant in serum amyloid A that might protect chickens from amyloid arthropathy 1086
C. Falker-Gieske, N.-F. Paul, M. Spourita, K. Gustmann, J. D. Gilthorpe and J. Tetens

258. Genetic background of the porcine immune system using SNP chip data and whole genome sequencing data 1090
C.M. Dauben, C. Große-Brinkhaus, H. Henne, A.K. Appel and E. Tholen

259. Deciphering the origins of Neolithic sheep from northern Iberian Peninsula 1094
M. Serrano, A. Fernández, C. González, J. Martínez-Blanch, M. Moreno-García, M. Rojo-Guerra and J.H. Calvo

260. Prioritisation of variants from whole-genome sequence data for validation of their association with traits of interest 1098
E. Molinero, R.N. Pena, J. Estany and R. Ros-Freixedes

261. Genome-wide association analyses and genomic prediction for pork meat quality traits using whole-genome sequence 1102
M. Heidaritabar, A. Huisman, M.C.A.M. Bink, P. Charagu and G. Plastow

262. A conditional multi-trait sequence GWAS of heifer fertility in tropically adapted beef cattle 1106
M. Forutan, B. Engle, M.E. Goddard and B.J. Hayes

263. Genomic prediction of purebred Merinos from an admixed terminal population using selected sequence variants 1110
N. Moghaddar A.A. Swan, P.M. Gurman, L. Li, and J.H.J. van Der Werf

264. The complete mitochondrial genome of a Peruvian creole cattle (*Bos taurus*) and its phylogenetic analysis 1114
C.I. Arbizu, R.D. Ferro-Mauricio, J.C. Chávez-Galarza, H.V. Vásquez, J.L. Maicelo, C. Poemape, J. Gonzales, C. Quilcate and F-A. Corredor

265. Unravelling the effect of structural variants from whole-genome sequence for depigmentation in Nellore cattle 1118
G. Vargas, H.H.R. Neves, N.A.M. Garzón, L.F.S. Fonseca, G.A. Fernandes Júnior, L.G. Albuquerque and R. Carvalheiro

266. On the advantage of identifying causal genetic variants for genomic prediction 1122
T.H.E. Meuwissen and M.E. Goddard

267. Fuzzy logic as a strategy for combining marker statistics to optimize preselection of high-density and sequence genotype data 1126
A. Ling, E.H. Hay, E. Hartono, S.E. Aggrey and R. Rekaya

268. Genetic evaluations and genomic characteristics for local cattle using genome sequences, 50K and a specific SNP chip 1131
K. May, M.J. Wolf, G.B. Neumann, P. Korkuć, G.A. Brockmann and S. König

269. Genomic prediction using whole-genome sequence data in intensely selected pig lines 1135
R. Ros-Freixedes, M. Johnsson, A. Whalen, C.Y. Chen, B.D. Valente, W.O. Herring, G. Gorjanc and J.M. Hickey

270. Single and multi-breed ssGBLUP using preselected variants from whole-genome sequence data in pigs 1139
S. Jang, R. Ros-Freixedes, J.M. Hickey, C.Y. Chen, W. Herring, I. Misztal and D. Lourenco

271. Application of low-pass sequencing to genomic prediction of egg quality in laying hens 1143
A. Wolc, J. Li, D. Lubritz, J. Arango, J. Fulton, P. Settar, K. Rowland and H. Cheng

Part 17 – Challenges – improving genomic prediction 1147

272. Accuracy of prediction for a genomic evaluation in rotational crossbreeding scheme (Montbéliarde × Holstein × Red Danish) 1148
P. Croiseau, R. Saintilan, A. Baur, I. Croue, V. Ducrocq, J.R. Thomasen, E. Karaman, D. Boichard, H. Leclerc and B.C.D. Cuyabano

273. Genomic evaluation for two-way crossbred performance in cattle 1152
Q. Mei, H. Liu, T. Xiang and O.F. Christensen

274. Improving accuracy of genomic prediction in distant populations by collecting sequence data over generations 1156
M.Y. Nawaz and C. Gondro

275. Comparison of genomic prediction methods for residual feed intake in broilers 1160
Z.X. He, S. Li, W. Li, J.Q. Ding, M.Q. Zheng, Q.H. Li, A.G. Fahey, J. Wen, R.R. Liu and G.P. Zhao

276. Impact of foreign phenotype data on single-step genomic evaluation of test-day protein yields for German Holsteins 1164
H. Alkhoder, Z. Liu, D. Segelke and R. Reents

277. Extend mixed models to multi-layer neural networks for genomic prediction including intermediate omics data 1169
T. Zhao, J. Zeng and H. Cheng

278. The holobiont 'predictome' of immuno competence in pigs 1173
J. Calle-García, Y. Ramayo-Caldas, L.M. Zingaretti, R. Quintanilla, M. Ballester and M. Pérez-Enciso

279. The long and short of Nanopore genomic prediction: the effect of read length on prediction accuracy 1177
H.J. Lamb, L.T. Nguyen, I.A.S. Randhawa, B.J. Hayes and E.M. Ross

280. Accuracy of genomic prediction of dry matter intake in Dutch Holsteins using sequence variants from meta-analyses 1181
B. Gredler-Grandl, B. Raymond, G.C.B. Schopen, P.K. Chitneedi, Z. Cai, C.I.V. Manzanilla-Pech, T. Iso-Touru, D. Fischer, S. Bolormaa, T.S. Chud, F.S. Schenkel, Y. Wang, C. Li, D. Hailemariam, B. Villanueva, A. Fernandez, C. Kuehn, G. Sahana, M.H. Lidauer, J.E. Pryce, O. González-Recio, G. Plastow, C.F. Baes, N. Charfeddine, Y. de Haas, R.F. Veerkamp and A.C. Bouwman

281. Genomic prediction of bull fertility in Italian Brown Swiss cattle 1185
H.A. Pacheco, M. Battagin, A. Rossoni, A. Cecchinato and F. Peñagaricano

282. On the use of SNPs of large effect to improve prediction accuracy in pigs 1189
M.S. Lopes, M. Derks, M. van Son, A.B. Gjuvsland, C.A. Sevillano, E. Grindflek and E.F. Knol

283. Including environmental variables in genomic models for carcass traits in Hanwoo beef cattle 1193
B.O. Makanjuola, G. Rovere, B.C.D. Cuyabano, S.H. Lee and C. Gondro

284. Considering chromosomal trait correlations improves accuracy of genomic prediction 1197
C. Schmidtmann, D. Segelke, J. Bennewitz, J. Tetens and G. Thaller

285. Effects of truncation and false positives in selection of markers for genomic prediction 1201
Z. Loh, J.H.J van der Werf and S. Clark

286. Genomic selection for breeding values under identity disequilibrium 1205
R.J.C. Cantet, B.K. Angarita Barajas, N.S. Forneris and S. Munilla

287. Effect of informative polymorphism on accuracy of genomic prediction and heritability 1209
M.S. Tahir, B.S. Olasege, K. Wockner, W.L.A. Tan, L.R. Porto-Neto and M.R.S. Fortes

288. Long-distance associations generate erosion of genomic breeding values of candidates for selection 1213
D. Boichard, S. Fritz, P. Croiseau, V. Ducrocq, B. Cuyabano and T. Tribout

289. Definition of reliabilities for models with metafounders 1217
M. Bermann, I. Misztal, D. Lourenco, I. Aguilar and A. Legarra

290. A new metric to assess reference populations for genomic selection in Australian beef breeds 1221
K.L. Moore, M.H. Ferdosi, C.G. Girard, S.F. Walkom and D.J. Johnston

291. Decreasing computing cost of categorical data analysis 1225
J. Hidalgo, I. Misztal, S. Tsuruta, M. Bermann, A. Garcia, K. Retallick and D. Lourenco

292. Within and across breed single step genomic prediction for somatic cell count including foreign information 1229
R. Mrode, M.W. Winters and M. Coffey

293. Using convolutional neural networks for image-based genomic prediction in mice 1233
B.C. Perez, A. Savchuk, P. Duenk, M.P.L. Calus and M.C.A.M. Bink

294. Selection for robustness: exploring the value of genomic prediction, reaction norm models and phenotyping strategies 1237
M. Ghaderi-Zefreh, A. Doeschl-Wilson, V. Riggio, O. Matika and R. Pong-Wong

Part 18 – Methods and tools: genomic prediction **1241**

295. Benchmark SNPs selection strategies to achieve optimum prediction in small population 1242
E. Mancin, C. Sartori, B. Tuliozi, G.G. Proto, R. Mantovani

296. Genetic correlations among longevity, fertility, udder health and type traits including or not genomic information 1246
R. Rostellato, L.-H. Maugan and V. Ducrocq

297. Alternative SNP weighting for genomic prediction methods in the presence of causative variants 1250
B.F. Santana, M. Riser and B. Fragomeni

298. Including selected sequence variants in genomic predictions for age at first calving in Nellore cattle 1254
G.S. Campos, D.A. Silva, H.H.R. Neves, D. Lourenco, G.A.F. Júnior, L.F.S. Fonseca, L.G. Albuquerque and R. Carvalheiro

299. Increasing accuracy of single-step evaluations of functional longevity including information from correlated traits 1258
L.-H. Maugan, T. Tribout, R. Rostellato, S. Mattalia and V. Ducrocq

300. Multi-trait genomic estimation of genetic parameters for growth and carcass traits of Duroc pigs 1262
E.C. Akanno, D.M. Thekkoot, C. Zhang, C. Bierman, G. Plastow and R.A. Kemp

301. Accounting for nuclear- and mito-genome in genetic evaluation and breeding of dairy cattle 1266
G.M. Fortuna, B.J. Zumbach, M. Johnsson, I. Pocrnic and G. Gorjanc

302. Using residual regressions to quantify and map signal leakage in genomic prediction 1270
B.D. Valente, G. de los Campos, A. Grueneberg, C.Y. Chen, R. Ros-Freixedes and W.O. Herring

303. Capitalizing on complex annotations in Bayesian genomic prediction for a backcross population of growing pigs 1274
F. Mollandin, H. Gilbert, P. Croiseau and A. Rau

304. Predictive ability of genomic prediction in layers when including CADD scores as genome function information 1278
M.C.A.M. Bink, M.P.L. Calus, M.F.L. Derks, J. Visscher and B.C. Perez

305. Across-sex genomic prediction using pre-selected SNP sets in cattle 1282
B.S. Olasege, M.S. Tahir, L.R. Porto-neto, B.J. Hayes and M.R.S Fortes

306. Genomic prediction with incomplete omics data 1286
E. Karaman, V. Milkeviych, Z. Cai, L. Janss, G. Sahana and M.S. Lund

307. A Bayesian hierarchical model to integrate a mechanistic growth model in genomic prediction 1290
H. Yu, J. van Milgen, E. Knol, R. Fernando and J. Dekkers

308. Mega-scale mixed models for genome-wide prediction with thousands of high-throughput phenotyping traits 1294
J. Qu, D.E. Runcie and H. Cheng

309. Nordic Holstein single-step test day model using metafounders and blending of foreign information 1298
M. Koivula, I. Strandén, G.P. Aamand and E.A. Mäntysaari

310. Effects of missing or wrong pedigree records on Single-Step predictions 1302
E.C.G. Pimentel, C. Edel, R. Emmerling and K.-U. Götz

311. An alternative method to consider a reference population in Single-Step SNP BLUP model without separating genomic terms 1306
T. Osawa, T. Baba, Y. Masuda, T. Kawahara and Y. Goto

312. Accuracy of pedigree and genomic predictions of meat quality in a multi-breed cattle using single and two step model 1310
S. Naderi, R. Evans and D.P. Berry

313. Single-step genomic prediction in small breeds: Finncattle case 1314
A.A. Kudinov, M. Koivula, I. Strandén, G.P. Aamand and E.A. Mäntysaari

314. Challenges to implement an interim genomic prediction for young animals in Wallonia, Belgium 1318
K. Wijnrocx, R.R. Mota, M. Dufrasne and N. Gengler

315. New insights into the computation of the inverse of the numerator relationship matrix for genotyped animals 1322
A. Maiwashe

316. GAPLS – a novel method to genomically rank bulls for daughter reproductive performance 1326
Y. Li, S. Lehnert, L.R. Porto-Neto, R. McCulloch, S. McWilliam, P. Alexandre, J. McDonald, C. Smith and A. Reverter

317. Performance of identity-by-state and identity-by-descent genomic models for meat traits in a cross-pig F2 population 1331
B.K. Angarita, J.P. Steibel, R.J.C. Cantet and N.S. Forneris

318. Modelling missing pedigree with metafounders and validating ssGBLUP evaluation for Uruguayan Holstein dairy cattle 1335
R.D. López-Correa, A. Legarra and I. Aguilar

319. Estimating metafounder relations for Finnish multibreed beef population 1339
M. Taskinen, A.-M. Leino, T.J. Pitkänen and E.A. Mäntysaari

320. A deregression method for a single-step SNP BLUP model using all genotype data 1343
Z. Liu and Y. Masuda

321. Variance component estimation for single-step genomic BLUP for Australian terminal sire sheep 1348
P.M. Gurman, L. Li, A.A. Swan, N. Moghaddar and J.H.J. van der Werf

322. Finetuning hyper-parameters increases the prediction accuracy in single-step genetic evaluation 1352
M. Neshat, Md.M. Momin, B. Truong, J.H.J. van der Werf , S. Lee and S.H. Lee

323. Comparison of single-step and multi-step evaluations for U.S. milk, fat, and protein 1356
R.R. Mota, A. Cesarani and P.M. VanRaden

324. Large-scale multiple-trait single-step marker model implementation 1360
B.L. Harris, R.G. Sherlock and M.A. Nilforooshan

325. An iterative method for deregressed proof in single-step genomic BLUP and its application to a dairy cattle population 1364
Y. Masuda, Z. Liu and P. Sullivan

326. Efficient approximation of reliability for genomic breeding values in large systems 1368
H. Eding

327. Evaluating the suitability of subjectively defined base populations 1372
J. ten Napel, J. Vandenplas, M. Calus, I. Strandén, M. Lidauer and R. Veerkamp

328. Maximum likelihood estimation of metafounder parameters for single and multiple breeds 1376
O.F. Christensen and A. Legarra

329. Exploring the statistical nature of independent chromosome segments 1380
M.K. Hollifield, M. Bermann, D. Lourenco and I. Misztal

330. Including local genomic breed proportions in genomic predictions for crossbred 1384
J.H. Eiríksson, G. Su and O.F. Christensen

331. Investigating the impact of APY core size and definition in single-step GBLUP evaluations 1388
R. Abdollahi-Arpanahi, D. Lourenco and I. Misztal

332. Strategies for choosing core animals in APY and their impact on the accuracy of single-step genomic predictions 1392
N.P.P. Macciotta, C. Dimauro, D. Lino Laurenco, A. Cesarani, L. Degano and D. Vicario

333. Genomic prediction accuracy: how low can we go? 1396
A. Reverter, P.A. Alexandre, Y. Li, B.C. Hine, C.J. Duff, A.B. Ingham and L.R. Porto-Neto

334. Genomic prediction in smallholder tropical dairy herds: modelling milk yield in first and later lactations 1400
E. Negussie, M.H. Lidauer, S. Meseret, A. Tera, E. Tessema, B. Jufar, T. Negera, A. Hassen, G. Gebreyohanes, O.A. Mwai and R. Mrode

335. Genomic prediction of wool shedding in a composite sheep flock using pooled allele frequencies and individual genotypes 1404
N. Vargas Jurado, J.W. Keele, L.A. Kuehn and R.M. Lewis

336. Optimal core definitions for the APY model with an example in large-scale pig dataset 1408
I. Pocrnic, F. Lindgren, D. Tolhurst, W.O. Herring and G. Gorjanc

337. Updating the core animals in the algorithm for proven and young in the American Angus Association national evaluations 1412
A.L.S. Garcia, S. Miller, S. Tsuruta, D. Lourenco, I. Misztal, D. Lu and K. Retallick

338. A critical aspect when using APY inversion with Single-Step GBLUP 1416
C. Edel, E.C.G. Pimentel, R. Emmerling and K.-U. Götz

339. Measures to quantify the accuracy and the erosion of genomic predicted breeding values 1420
B.C.D. Cuyabano, D. Boichard and C. Gondro

340. Model building pipeline to maximize the accuracy of breeding values prediction 1424
J. Jenko and Ø. Nordbø

341. Genomic prediction in multi-breed Irish beef cattle population using a breed origin of alleles model 1428
A. Guillenea, M.S. Lund, R. Evans, V. Boerner and E. Karaman

342. Breed-origin-of-alleles approach using summary statistics for multi-breed genomic prediction in dairy cattle 1432
J.B. Clasen, W.F. Fikse, G. Su and E. Karaman

Part 19 – Methods and tools: software and computing strategies 1436

343. Subsetted orthogonal data augmentation for fast parallel implementation of Bayesian models for whole-genome analyses 1437
L.H. Chen and G.S. Plastow

344. A simulator to evaluate gene editing assisted selection programs 1442
X-J. Yu, B.S. Dagnachew and T.H.E. Meuwissen

345. A database structural improvement for efficient trait variation curation in animal QTLdb and CorrDB 1446
Z-L. Hu, C.A. Park and J.M. Reecy

346. Efficient algorithms to identify duplicated genotypes in large datasets 1450
M.H. Ferdosi, S. Masoodi and M. Khansefid

347. Genotype error correction and the accurate identification of recombination sites 1454
M.M. Hindle and A. Kranis

348. A phasing algorithm targeted at dairy cattle populations having genotyped ancestors with many genotyped offspring 1459
M. Dahl and O.F. Christensen

349. Recent advances and future needs in genotype imputation 1463
P.M. VanRaden, D.J. Null and A.S. Al-Khudhair

350. Imputation of low coverage sequence data by utilization of pre-phased reads and HBimpute 1467
T. Pook and J. Geibel

351. Quantitative trait simulation using MeSCoT software 1471
V. Milkevych, E. Karaman, G. Sahana, L. Janss, Z. Cai and M.S. Lund

352. SIRE 2.0: a novel software tool to estimate host genetic variation in infectious disease transmission 1475
C.M. Pooley, G. Marion and A. Doeschl-Wilson

353. iVPSV: an intuitive visualisation platform for structural variation data 1479
H. Jung, B. Jeon, B. Hayes and E. Ross

354. A variational Bayes method for genomic prediction increases accuracy and computing speed 1482
J. Cheng, H. Cheng, R. Fernando, C. Maltecca, L. Ma, J. Dekkers and J. Jiang

355. Additions to the BGLR R-package: a new function for biobank size data and Bayesian multivariate models 1486
P. Pérez-Rodríguez and G. de los Campos

356. SFSI: a package to solve penalized selection indices and sparse genomic prediction 1490
M. Lopez-Cruz, G. de los Campos, B. Valente, C.Y. Chen, W. Herring and A. Grueneberg

357. INTERGEN: an efficient and flexible tool for large-scale genetic evaluation of complex animal and plant populations 1494
V.S. Junqueira, G.S. Campos and F.F. Cardoso

358. MiXBLUP 3.0 – software for large genomic evaluations in animal breeding programs 1498
J. Vandenplas, R.F. Veerkamp, M.P.L. Calus, M.H. Lidauer, I. Strandén, M. Taskinen, M. Schrauf and J. ten Napel

359. From raw sensor and automated data to genetic evaluation and validation in the cloud 1502
D. Schokker, I.N. Athanasiadis, M. Poppe, J. ten Napel, C. Kamphuis and R.F. Veerkamp

360. Indirect genomic prediction reduces computational costs in large-scale single-step evaluations 1506
I. Strandén, J. ten Napel, R.F. Veerkamp, R. Evans, S. Naderi, E.A. Mäntysaari and J. Vandenplas

361. Accounting for trait-specific genomic and residual polygenic covariances in multivariate single-step genomic evaluation 1510
K. Meyer

362. Avoiding double counting in genomic reliability calculations 1515
H. Ben Zaabza, M. Taskinen, E.A. Mäntysaari, T. Pitkänen, G.P. Aamand and I. Strandén

363. JWAS version 2: leveraging biological information and high-throughput phenotypes into genomic prediction and association 1519
H. Cheng, R. Fernando, D. Garrick, T. Zhao and J. Qu

364. ACCF90GS2: software for fast approximation of reliabilities of estimated breeding values in single-step GBLUP 1523
M. Bermann, D. Lourenco, A. Cesarani and I. Misztal

365. One for all: LMT – the linear models toolbox 1526
V. Boerner

366. Recent updates in the BLUPF90 software suite 1530
D. Lourenco, S. Tsuruta, I. Aguilar, Y. Masuda, M. Bermann, A. Legarra and I. Misztal

367. From data to genomic breeding values with the MiX99 software suite 1534
T.J. Pitkänen, H. Gao, A. Kudinov, M. Taskinen, E.A. Mäntysaari, M.H. Lidauer and I. Strandén

Part 20 – Statistical genetics: GWAS **1538**

368. Haplotypes affecting pulmonary arterial pressure in Angus cattle 1539
D. Lu, A. Garcia, S. Miller and K. Retallick

369. Detection of shared genomic regions between production and health traits in German Holstein cattle 1543
H. Schneider, D. Segelke, J. Tetens, G. Thaller and J. Bennewitz

370. Detecting the presence of genomic imprinting for carcass traits in cattle using imputed high-density genotypes 1547
D. Kenny, R.D. Sleator, C.P. Murphy, R.D. Evans and D.P. Berry

371. Evidence of genetic links between rumen bacteria and milk somatic cell score in Lacaune dairy sheep 1551
G. Martinez Boggio, A. Meynadier, R. Rupp, C. Allain, A.J. Buitenhuis and C. Marie-Etancelin

372. Leveraging GWAS models to map selection on complex traits in massive datasets 1555
T.N. Rowan, R.D. Schnabel and J.E. Decker

373. Fine-mapping young-stock survival QTL on chromosome 6 in Nordic Red Dairy Cattle 1559
Z. Cai, X. Wu, B. Thomsen and G. Sahana

374. Multibreed sequence level genome-wide association study of semen traits in tropical Australian cattle 1563
W.L.A Tan, L.R. Porto-Neto, A. Reverter and M.R.S. Fortes

375. Improved genome-wide associations using a breed-specific 200K SNP chip for German Black Pied (DSN) cattle 1567
P. Korkuć, G.B. Neumann, D. Arends, M.J. Wolf, K. May, S. König and G.A. Brockmann

376. Haplotype-based single-step GWAS for yearling temperament in American Angus 1571
A.C. Araujo, P.L.S. Carneiro, A.B. Alvarenga, H.R. Oliveira, S.P. Miller, K. Retallick and L.F. Brito

377. Genome-wide association analyses for various thermotolerance indicators measured during lactation in maternal-line pigs 1575
P.H.F. Freitas, J.S. Johnson, H. Wen, J.M. Maskal, M.K. Byrd, F. Tiezzi, Y. Huang, A.P. Schinckel and L.F. Brito

378. Investigation of GWAS variants associated with loin depth in commercial pigs 1579
S. Desire, M. Johnsson, R. Ros-Freixedes, C. Chen, W. Herring, G. Gorjanc, R.J. Mellanby, J. Hickey and M. Jungnickel

379. Multivariate GWAS revealed a critical region on chromosome 4 for developmental and feather growth traits in ducks 1583
E. Tarsani, S. Desire, O. Matika, K. McIntosh, A. Rae, A. Kranis and K.A. Watson

380. Investigation of genes associated with maternal weaning weight in South African Bonsmara beef cattle 1586
J.J. Reding, R.R. Van der Westhuizen, D.P. Berry and E. van Marle-Köster

381. Genome-wide genomic and functional association study for workability and calving traits in Holstein cattle 1590
T. Suchocki, B. Kosińska-Selbi, A. Żarnecki, M. Jakimowicz, W. Jagusiak, M. Morek-Kopeć and J. Szyda

382. Genomic region-specific association test identifies new loci for carcass traits in Hanwoo (Korean native cattle) 1594
D.J. Lee, D.H. Lee, J.M. Kang, Y.K. Kim, Y.J. Chung and S.H. Lee

383. Genetic mechanisms underlying feet and legs malformation in Nellore cattle: prioritization of GWAS results 1598
T.L. Silva, C. Gondro, P.A.S. Fonseca, D.A. da Silva, G. Vargas, H.H.R. Neves, I. Carvalho Filho, C.S. Teixeira, L.G. Albuquerque and R. Carvalheiro

384. Genome-wide association studies for milk yield in *Bos taurus* × *Bos indicus* crossbred cattle in India 1602
M. Al Kalaldeh, M. Swaminathan, V. Podtar, S. Jhadav, V. Dhanikachalam, A. Joshi and J.P. Gibson

385. Grouping of highly correlated variables improves performance of genomic evaluations 1606
D. Wittenburg and J. Klosa

386. A Monte-Carlo approximation of the null hypothesis for genome wide association sing the windows variance approach 1610
D. López-Carbonell, M. Ramírez, J. Altarriba, H. Srihi and L. Varona

387. Sequence-based association analyses on X chromosome in six dairy cattle breeds 1614
M.P. Sanchez, C. Escouflaire, A. Baur, C. Hozé, A. Capitan, S. Fritz and D. Boichard

388. Pleiotropic effects on carcass and meat quality traits in crossbred beef cattle 1618
F.M. Rezende, E. Rodriguez, J.D. Leal-Gutiérrez, M.E. Elzo, D.D. Johnson, C. Carr and R.G. Mateescu

389. Sequence-based multi-trait genome-wide association study for linear classification traits in Belgian Blue beef cattle 1622
J.L. Gualdron Duarte, A.S. Gori, G. Costa Monteiro Moreira, M. Georges, C. Charlier and T. Druet

390. Integrating imputed structural variants and tandem repeats in a GWAS for growth and carcass traits in F_2 pig crosses 1626
I. Blaj, J. Tetens, J. Bennewitz, G. Thaller and C. Falker-Gieske

391. Detecting QTL for two lowly correlated traits using multi-trait and meta-analysis approaches in Swiss Large White pigs 1630
A. Nosková, A. Mehrotra, N.K. Kadri, S. Neuenschwander, A. Hofer and H. Pausch

392. A Forward algorithm to identify credible sets in Bayesian variable selection models 1634
G. de los Campos, A. Samaddar, B. Valente, C.-Y. Chen, W. Herring and A. Grueneberg

Part 21 – Statistical genetics: machine learning, deep learning and artificial intelligence 1638

393. Multivariate application of artificial neural networks for genomic prediction 1639
S.O. Peters, K. Kızılkaya, M. Sinecen and M.G. Thomas

394. Comparison of GBLUP and machine learning breeding values for binary disease traits in working dogs 1643
J.A. Thorsrud, K.M. Evans, D.M. Holle, H.J. Huson

395. Predicting the polled genotype of ungenotyped animals using machine learning methods 1647
A.B. Gjuvsland

396. A machine learning approach to classify high and low feed efficient or grower mink 1652
A. Shirzadifar, G. Manafiazar, P. Davoudi, D.N. Do, G. Hu and Y. Miar

397. A machine learning approach for predicting the most and the least feed-efficient groups in beef cattle 1656
A. Shirzadifar, G. Plastow, J. Basarab, Y. Miar, C. Li, C. Fitzsimmons, M. Riazi and G. Manafiazar

398. Combination of SNP markers for discrimination of a target chicken population using machine learning algorithms 1660
E. Cho, S. Cho, M. Kim, T.K. Ediriweera, D. Seo and J.H. Lee

399. LASSO and SVM: an alternative approach to identify associated genome regions for simple and complex traits in cattle 1664
S. Widmer, F.R. Seefried, C. Drögemüller and P. von Rohr

400. Quantifying the functional conservation between human and pig using artificial neural networks 1668
J. Li, T. Zhao, Z. Pan, H. Zhou, L. Fang and H. Cheng

401. Using a U-net architecture to develop an automated ultrasound images analysis method in a rabbit breeding program 1672
C. Herbert, C. Ribas, K. Kashefifard, L. Flatrès-Grall and G. Lenoir

402. Machine learning and genetic improvement of animals and plants: where are we? 1676
D. Gianola, J. Crossa, O. Gonzalez-Recio and G.J.M. Rosa

403. Predicting pig digestibility coefficients with microbial and genomic data using machine learning prediction algorithms 1680
C. Carillier-Jacquin, V. Deru, L. Tusell, A. Bouquet, L. Jacquin and H. Gilbert

404. (Quasi) Multi-task support vector regression for whole-genome prediction of carcass traits in commercial broilers 1684
A.A.C. Alves, A.F.A. Fernandes, V. Breen, R. Hawken and G.J.M. Rosa

405. Genomic prediction of longitudinal body weights in pigs using a neural network 1688
Z. Ni, R.L Fernando, H. Yu, E.F. Knol and J.C.M. Dekkers

406. Compressing microbiota information using an autoencoder to predict growth traits in swine 1692
C. Maltecca, J. Jiang, J. Fix, C. Schwab C. Shull and F. Tiezzi

Part 22 – Society – genetics in the tropics **1696**

407. Providing genetics for the dairy industry in the tropics – the Brazilian Dairy Gir breeding program 1697
J.C.C. Panetto, M.V.G.B. Silva, M.A. Machado, R.S. Verneque, A.R. Fernandes, C.H.C. Machado, M.F. Martins, D.R.L.R. Faza, C.A.V. Borges, H.T. Ventura, M.A. Pereira, J.C. Oliveira, L.A. Glatzl Junior, A.O. Garcia and F.D. Leandro

408. Modelling phenotypic plasticity relative to a temperature humidity index as an indicator of adaptability in beef cattle 1701
W. Shaffer, J.A.H. Moreno, N.M. Bello, R. Weaber, J. Bormann and M. Rolf

409. Differentiated genomic regions in the Afrikaner and Brahman cattle of South Africa reveal parallel adaptive mechanisms 1705
S. Mdyogolo, M.D. MacNeil, F.W.C. Neser, M.M. Scholtz and M.L. Makgahlela

410. Genomic analysis of milk yield and heat tolerance in small holder dairy system of sub-Saharan Africa 1709
C.C. Ekine-Dzivenu, R. Mrode, R.D. Oloo, D. Komwihangilo, E. Lyatuu, G. Msuta, J.M. Ojango and A.M. Okeyo

411. Genetic analysis of traits needed to optimize improvement in beef cow-calf efficiency in Afrikaner cattle1713
M.M. Scholtz, M.D. MacNeil, M.C. Chadyiwa and F.W.C. Neser

412. Genetic diversity of the Cameroon Blackbelly sheep, an indigenous sheep from West Africa 1717
P. Wiener, M. Salavati, A. Djikeng, C.P. Van Tassell, B.D. Rosen, G.L. Spangler, G. Simo, V. Tanya, F. Meutchieye and E.L. Clark

413. Phenotypic characterisation of African chickens raised in semi-scavenging conditions 1721
K.M. Morris, K. Sutton, M. Girma, W. Esatu, B. Solomon, T. Dessie, L. Vervelde, A. Psifidi, O. Hanotte and G. Banos

414. Spatial modelling in genetic evaluation of South African Holstein cattle population 1725
I. Houaga, T.P. Oliveira, I. Pocrnic, E. Lavrenčič, C.B. Banga and G. Gorjanc

415. Testing phenotypes for degree of resilience using fluctuations in milk yield of dairy cows in sub-Saharan Africa 1729
R.D. Oloo, C.C. Ekine-Dzivenu, J.M.K. Ojango, R. Mrode, A.O. Mwai, and M.G.G. Chagunda

416. Egg production and genetic parameters for Tanzanian local chickens 1733
T. Magwaba, E. Mollel, J.R Mushi, M. Walugembe, E.N Amuzu-Aweh, G.H. Chiwanga, N. Chouicha, P.L. Msoffe, T. Kelly, R.A. Gallardo, H. Zhou, A. Wolc, S.J. Lamont, A.P. Muhairwa and J.C.M. Dekkers

417. Expanding the cattle reference graph genome 1737
A. Talenti, J. Powell, D. Wragg, E. Paxton, M. Chepkwony, A. Miyunga, R. Njeru, J.D. Hemmink, A. Fisch, B.R. Ferreira, J.A. Hammond, A.L. Archibald, P. Toye, T. Connelley, L. Morrison and J. Prendergast

418. Genetic structure and diversity of Jersey dairy cattle in Rwanda 1741
O. Opoola, F. Shumbusho, R. Mrode, D. Hambrook, M.G.G. Chagunda, D. Moran and A. Djikeng

419. Estimation of breed admixture levels in Sri Lanka cattle and its relation to performance traits 1745
M. Spoelstra, J. Sölkner, H. Bovenhuis, B. Babigumira, K. Periasamy and C.M.B. Dematawewa

420. Economic values for skin grade, days to market and number of hatchlings in the Australian saltwater crocodile industry 1749
S. Hermesch and S.R. Isberg

421. Exploiting phenotypic plasticity in animal breeding 1753
F.G. Kebede, H. Komen, T. Dessie, O. Hanotte, S. Kemp, C. Pita Barros, R. Crooijmans, M. Derks, S.W. Alemu and J.W.M. Bastiaansen

422. Whole genome sequencing reveals genetic diversity and heat-stress adaptation in Nigerian indigenous chickens 1757
A.A. Gheyas, M. Rachman, O. Bamidele, T. Dessie, J. Smith and O. Hanotte

Part 23 – Challenges – delivering genetic progress in systems around the world **1761**

423. Impact of industry effort to increase dairy heifer genomic testing in Australia 1762
M. Abdelsayed, M.M. Axfordand G.J. Nieuwhof

424. Weekly national dairy evaluations – a Vietnam-Australia collaboration 1766
G.J. Nieuwhof and L.T. Ong

425. The value proposition for using genomics to confirm parentage in Australian dairy systems 1770
J.E. Newton, M.M. Axford, J.E. Pryce, S.L. Bullen and M. Haile-Mariam

426. Changes in genetic trends for dairy cattle in the U.S. since the implementation of genomic selection 1774
F.L. Guinan, G.R. Wiggans, J.W. Dürr, H.D. Norman, J.B. Cole, C.P. Van Tassell, I. Misztal and D. Lourenco

427. Community based breeding programme for low input beef cattle production systems in the arid region of Argentina 1778
S. Munilla, J.M. Tessi and D.O. Maizon

428. A participatory approach to review and update the national breeding objectives for New Zealand dairy cattle 1782
B.F.S. Santos, G.M. Jenkins, J. Kok, K. Stachowicz and P.R. Amer

429. Genetic trends for carcass traits within breeds in a multi-breed evaluation 1787
S.A. Mulhall, R.D. Sleator, R.D. Evans and A.J. Twomey

430. Improving growth performance for dual purpose chickens for challenging conditions in Africa 1791
N. Duijvesteijn, L. Perrault, R. Atikpakpe, N. Garcia Martinez, J. Renckly, T. Arbeau and J.A.M. van Arendonk

431. Genomic selection in the US: where it has been and where it is going? 1795
K.J. Retallick, D. Lu, A. Garcia and S.P. Miller

432. Efficient genetic progress for quantitative traits through genomic selection 1799
S. Tsuruta, D.A.L. Lourenco and I. Misztal

Part 24 – Breeding programs **1803**

433. Strategies to support phenotyping for genomic selection R&D and implementation in beef and sheep industries 1804
R.G. Banks

434. The infrastructure for cattle data at the Swedish university of agricultural sciences, Gigacow 1808
T. Klingström, I. Ohlsson and D.J. de Koning

435. Optimizing sampling strategies for sib test schemes in a broiler breeder program 1812
C.A. de Hollander, V.P. Breen, J. Henshall, F.B. Lopes and M.P.L. Calus

436. Plasticity of breed direct and individual heterosis effects in beef cattle under extensive conditions 1816
G.M. Pyoos-Daniels, M.D. MacNeil, M.M. Scholtz, M.M. Seshoka, H.L. Weepener and F.W.C. Neser

437. Polled Accelerator – a unique application of genomic technologies to address a beef breeding challenge 1820
D.J. Johnston, M.H. Ferdosi, J. Cook and D.B. Savage

438. An approach for the design of breeding programs using genomics 1824
J.C.M. Dekkers, H. Su, L. Kramer and H. Yu

439. Should we breed cows like pigs? A new organisation of dairy cattle breeding 1828
M. Kargo, H.M. Nielsen, S. Østergård, R.D. Kring, M. Slagboom, J.B. Clasen and J.R. Thomasen

Part 25 – Operational breeding programs **1832**

440. Management of undesired alleles at multiple loci for a goat breeding program 1833
J. Raoul, J.M. Elsen, S.T. Rodríguez-Ramilo and I. Palhière

441. Prediction of gametic variance and its use in bovine breeding programs 1837
C. Hozé, A. Baur, S. Fritz and D. Boichard

442. Implementing genomic management of diversity and selection in French dairy goats 1841
I. Palhière, V. Gousseau and J.J. Colleau

443. Partition in long-term contributions of the genetic trend of Basco-Béarnaise dairy sheep 1845
S. Antonios, A. Legarra, J.M. Astruc, S.T. Rodríguez-Ramilo and Z.G. Vitezica

444. Introgression of prolificacy gene in Tunisian Barbarine sheep through marker-assisted selection scheme 1848
E. Arari, I. Baazaoui, N. Letayef, M. Lahmar, M. Hemdène, A. Haile, M. Rekik and S. Bedhiaf-Romdhani

445. Simulation of animal crossbreeding schemes using ADAM software 1852
H. Liu, H.M. Nielsen, M. Kargo, J.R. Thomasen, P. Berg, M. Slagboom, A.C. Sørensen and V. Milkevych

446. A tool for comprehensive implementation of simple and complex breeding programs 1856
B.P. Kinghorn, A.J. Kinghorn and J.L. Baller

447. Beyond scenarios – resource optimization for dairy cattle breeding programs (MoBPSopti) 1860
A. Hassanpour, H. Simianer and T. Pook

448. Predicting the outcome of a selection scheme in maternal traits 1864
A.V. Mäki-Tanila and J. Pietarinen

449. Inbreeding in circular and panmictic mating systems 1868
B. Guldbrandtsen

450. Genomic rotational crossbreeding including advanced optimum contribution selection methods in local dairy cattle breeds 1872
J. Stock, H. Esfandyari, D. Hinrichs, R. Wellmann and J. Bennewitz

451. Allele frequency differences at epistatic QTL explain different genetic trends in number of teats in two pig lines 1876
C.A. Sevillano, B. Harlizius, M.F.L. Derks, M.S. Lopes, M. van Son and E.F. Knol

Part 26 – Developing countries **1880**

452. Phylogenetic analysis of local cattle population in northern province of Sri Lanka 1881
K.G.C.B. Wijebandara, M.S. Kurukulasuriya, P.C.G. Bandaranayake and G.L.L.P. Silva

453. Accounting for heterogeneity of data in implementing genomic selection models applicable to crossbred cattle in India 1885
A. Guintard, L. Bourguignon, G. Restoux and V. Ducrocq

454. Connections among local nomenclature, phenotypes and genetic characteristics of native chickens in Cambodia: a synthesis 1889
P. Menghak and M.G.G. Chagunda

455. Experience in creating a female *Bos taurus* × *Bos indicus* reference population under smallholder system in India 1893
K.G. Bhave, S. Joshi, Y. Gaundare, A. Joshi, T. Shirsath, N. Punde, V. Potdar, V. Ratnakar and M. Swaminathan

456. Learnings and experiences from a community-based goat breeding program in India 1897
A. Schurink, S. Kumar, R. Chakraborty, T. Sawhney, C. Nimbkar and P. Amer

457. Response to index selection for temperate dairy cattle breeds in tropical Sri Lanka 1901
A.M. Samaraweera, J.H.J. van der Werf, V. Boerner and S. Hermesch

458. Challenges on construction and use of female reference populations for genomic selection under Indian smallholder systems 1905
V. Ducrocq, M. Swaminathan and J.P. Gibson

459. Bridging the gap in data from smallholder dairy systems; developing the Africa dairy genetic gains (ADGG) data platform 1909
J.M.K. Ojango, A.M. Okeyo, R. Mrode, E. Chinyere, G. Gebreyohanes, S. Meseret, D. Mogaka, E. Lyatuu, D. Komwihangilo, G. Msuta, S. Kahumbu, J.P. Gibson and S. Kemp

460. Attitudes of Peruvian llama and alpaca farmers towards genetic improvement 1913
M. Wurzinger and G. Gutiérrez

461. Development of a low-density SNP panel for local Ghanaian and Tanzanian chicken ecotypes 1917
M. Walugembe, E.N. Amuzu-Aweh, K.-S. Lim, Y. Wang, N. Chouicha, T. Kelly, A.P. Muhairwa, B.B. Kayang, P.L.M. Msoffe, A. Naazie, H.R. Otsyina, J.S. Mushi, R.A. Gallardo, H. Zhou, S.J. Lamont and J.C.M. Dekkers

462. Combining demographic, production and molecular investigations in genetic improvements: case study in indigenous chicken 1921
G.L.L.P. Silva, M.S. Kurukulasuriya, P.R.M.K. Fernando and C.M.B. Dematawewa

Part 27 – Breeding goals and selection strategies 1925

463. Relative importance of selection objective traits in participatory breeding programs – MultiAtr software 1926
R.N.B. Lobo

464. Validation of Angus HeiferSELECT using historical data 1930
P.A. Alexandre, L.R. Porto-Neto, B. Hine, A. Ingham, M. Samaraweera, C. Duff and A. Reverter

465. Full-scale simulation of the Norwegian red dairy cattle breeding programme to enable testing future breeding directions 1934
A. Ehsani, J. Jenko, H. Storlien, H. Melbo Tajet, I. Pocrnic, J. Obsteter and G. Gorjanc

466. Improving carcase value by incorporating primal weights into beef breeding objectives 1938
N.R. Sarker, B.J. Walmsley and S. Hermesch

467. A method for partitioning trends in genetic mean and variance to understand breeding practices 1942
T.P. Oliveira, J. Obšteter, I. Pocrnic and G. Gorjanc

468. A practical approach to predicting expected response to selection with a new index 1946
X. Zhang, P.R. Amer and C. Quinton

469. Theory and tools that are needed for the design and optimization of breeding programs 1950
P.R. Amer

470. General properties of the optimum selection index 1954
R. Wellmann

471. Advantages of modularisation for industry breeding objective development 1958
C.D. Quinton, P.R. Amer, T.J. Byrne and B.F.S. Santos

472. Enhancement of the terminal carcass production index to incorporate birth weight and lambing ease in Australian sheep 1962
E. Dehnavi and A.A. Swan

473. iGENDEC: a web-based decision support tool for economic index construction 1966
M.L. Spangler, B.L. Golden, S. Newman, L.A. Kuehn, W.M. Snelling, R.M. Thallman and R.L. Weaber

474. Analysis of causal structure of traits involved in sow lactation feed efficiency 1970
M. Mora, I. David, H. Gilbert, G.J.M. Rosa, J.P. Sánchez and M. Piles

Part 28 – Methods and tools: breeding goals 1974

475. Manipulating response to selection on many traits 1975
J.H.J. van der Werf and S.A. Clark

476. Selection for birth weight environmental variability in mice as a model to improve animal welfare in livestock species 1979
J.P. Gutiérrez, N. Formoso-Rafferty, C. Ojeda-Marín, L. El-Ouazizi El-Kahia, K.D. Arias and I. Cervantes

477. Sensitivity Analysis for a new Economic Selection Index for Chilean dairy cattle 1983
J.L. Chuma-Alvarez, H.H. Montaldo, D.A.L. Lourenco, F.J. Ruiz-López, M.E. Olivares and C. Lizana

478. Opportunities for animal genetics to foster the ecological transition of livestock farming systems 1987
D. Savietto, F. Douhard, H. Gilbert, R. Rupp, L. Fortun-Lamothe and A. Ducos

479. Defining animal breeding objectives and strategies within the framework of responsible research and innovation 1991
I. Olesen, A. Kettunen, G.K. Rosendal and H.M. Nielsen

480. On farm specific breeding goals for mating lead to higher genetic gains at herd level 1995
M. Bérodier Gomès, M. Brochard and V. Ducrocq

481. A bio-economic model to estimate the economic value of sow feed intake during lactation 1999
D.M. Thekkoot, R.A Kemp and J.C.M. Dekkers

482. Consequences of using different economic selection index methods on greenhouse gas emissions in beef cattle 2003
B.J. Walmsley

483. Impact of planning horizon length on the relative emphasis of traits in economic breeding goals 2007
H.F. Valasek, B.L. Golden and M.L. Spangler

Part 29 – Society – gene editing: ethics, economics and experiences 2011

484. Effect of alternative synchronisation treatments on the vaginal microbiota of ewes and their impact on pregnancy rate 2012
E.L. Reinoso-Peláez, O. González-Recio, C. González, M. Saura, A. Fernández, A. López-García, A. Saborío-Montero, J.H. Calvo, M. Ramón and M. Serrano

485. Modelling of direct-transfer vitrification of bovine embryos 2016
J. Lamy, A. de Wit, F. Fotinga-Meiring, E. Mullaart, L. Zak, E. Knol, E. Ellen and H. Woelders

486. Fertility analysis in stallions: a 35-year retrospective genetic study in the Argentinean Thoroughbred 2020
A. Karlau, A. Molina, A. Antonini and S. Demyda-Peyrás

487. Bull effects on *in vitro* embryo production quality traits are repeatable and heritable 2024
G. Gebreyesus, J.B. Secher, M.S. Lund, K. Kupisiewicz, M. Ivask and G. Su

488. Five years of optimizing the assisted reproduction protocol for European eel: what worked and what didn't? 2028
A.P. Palstra, P. Jéhannet, L.T.N. Heinsbroek and W. Swinkels

489. Random regressions for modelling semen production traits in HF purebred and crossbred bulls using Bayesian framework 2031
V.B. Ambike, R. Venkataramanan, S.M.K. Karthickeyan, K.G. Tirumurugaan, K.G. Bhave and M. Swaminathan

490. Removing the dopaminergic inhibition with CRISPR-Cas9 to study the progression of sexual maturation in zebrafish 2036
P. Jéhannet, A.A.C. De Wit, S. Brugman, Y. Zhang, J. Bogerd and A.P. Palstra

491. Societal aspects of animal breeding: a bibliometric analysis 2040
M. van der Sluis, K.H. de Greef and G. Bonekamp

492. The need for harmonized, risk-based regulation of gene editing in livestock 2044
A.L. Van Eenennaam and M.L. Mueller

493. Broadening the ethical debate on breeding innovations, public engagement and the role of the Democs game 2048
F.L.B. Meijboom, D.M. Bruce, K. Kramer and A. Bruce

494. Perceptions of genome editing in farm animals by livestock stakeholders 2052
R. Duclos, E. Delanoue, L. Journaux, D. Guéméné, M. Sourdioux and J.P. Bidanel

495. Production of a germline ablated bull using CRISPR/Cas9 to target *NANOS3* in bovine zygotes 2056
M.L. Mueller, B.R. McNabb, J.R. Owen, S.L. Hennig, P.J. Ross and A.L. Van Eenennaam

Part 30 – Challenges – use of multi-omics data **2060**

496. Using allele-specific expression to uncover *cis*-regulation in bovine muscle 2061
J.J. Bruscadin, T.F. Cardoso, W.J.S. Diniz, J. Afonso, M.M. de Souza, J. Petrini, V.H. da Silva, A. Zerlotini, L.L. Coutinho and L.C.A. Regitano

497. Enrichment of causative variants in tissue-specific and shared ATAC-Seq peaks in cattle 2065
C. Yuan, L. Tang, T. Lopdell, C. Oget-Ebrad, G. Costa Monteiro Moreira, J.L. Gualdron, Z. Cheng, M. Salavati, D.C. Wathes, M.A. Crowe, GplusE consortium, W. Coppieters, C. Charlier, T. Druet, M. Georges and H. Takeda

498. Preliminary investigation into the association of members of the rumen biome with production traits in Afrikaner cattle 2069
E.D. Cason, J. Alom, J. Barnard, M. Fair, P.D. Vermeulen and F.W.C. Neser

499. Maternal effect on the metagenomic composition determining the intramuscular fat content in rabbits 2073
A. Zubiri-Gaitán, M. Mora, C. Casto-Rebollo, M.A. Santacreu, A. Blasco, P. Hernández and N. Ibañez-Escriche

500. The gut microbiome profile varies among two maternal rabbit lines with different longevity 2077
I. Biada, N. Ibáñez-Escriche, A. Blasco and M.A. Santacreu

501. Comparing methods to summarize gut microbiota composition in estimating microbiability of host phenotypes in swine 2081
Y. He, C. Maltecca, J. Howard, Y. Huang, K. Gray and F. Tiezzi

502. Organoids as a research tool in animal breeding and nutrition 2085
L.M.G. Verschuren, D. Schokker, S. Verstringe, L. Kruijt, A.A.C. de Wit, E.G.T. van der Valk, S.K. Kar and E.D. Ellen

503. Simulation of host-microbiome evolution throughout a divergent selection experiment 2089
C. Casto-Rebollo, I. Pocrnic, G. Gorjanc and N. Ibáñez-Escriche

504. Rumen bacteria do not provide improved genetic evaluation of dairy traits in sheep 2093
G. Martinez Boggio, O.F. Christensen, A. Legarra, C. Allain, A. Meynadier and C. Marie-Etancelin

505. Phenotypic characterization of organoids derived from pig intestine segments 2097
F. Blanc, M. Mongellaz, F. Pepke, C. Bevilacqua, J. Riviere, M. Vilotte, E. Giuffra and G. Egidy

506. Can microbial data improve prediction of breeding values of efficiency traits in pigs fed conventional or fiber diets? 2101
V. Déru, F. Tiezzi, C. Carillier-Jacquin, B. Blanchet, L. Cauquil, O. Zemb, A. Bouquet, C. Maltecca and H. Gilbert

507. Investigating new technologies for on-site real-time sequencing for any animal scientist 2105
H. Ostrovski, R.P. Savegnago, W. Huang and C. Gondro

508. Impact of the bovine reference genome (ARS-UCD1.2) on functional annotation at genome, transcriptome and methylome level 2109
A. Cánovas, P.A.S. Fonseca, S. McKay and J.F. Medrano

Part 31 – Functional annotation of genomes: causal variants & gene discovery **2113**

509. Detection of genes influenced by environmentally responsive *cis*-regulatory variation in porcine immune cells 2114
E. Murani, F. Hadlich, S. Ponsuksili, and K. Wimmers

510. Insight into the most significant quantitative trait locus for calving traits on BTA18 in Holstein dairy cattle 2117
N. Dachs, M. Upadhyay, E. Hannemann, A. Hauser, S. Krebs, H. Blum, D. Seichter, I. Russ, L.J. Gehrke, G. Thaller and I. Medugorac

511. Finding potentially causal genetic factors for meat tenderness using eQTL study in Hanwoo cattle 2121
Y. Chung, K. Chung, P. Dinh, I. Choi and S. Lee

512. The application of mixed linear models for the estimation of functional effects on bovine stature based on SNP summary 2126
K. Kotlarz, B. Kosinska-Selbi, Z. Cai, G. Sahana and J. Szyda

513. Predicted isoforms of a Brahman cow revealed by full-length transcript sequencing 2130
L.T. Nguyen, Y. Cheng, R. Kuo, B.J. Hayes and E.M. Ross

514. Meta-analysis of genome-wide association studies for litter size in pigs indicates new causal variants 2134
E. Sell-Kubiak, M.F.L. Derks, M.S. Lopes, J. Dobrzański and T. Szwaczkowski

515. Selection signatures analyses revealed genes associated with adaptation in selected goat breeds in Kenya 2138
R.W. Waineina, T.O. Okeno, E.D. Ilatsia and K. Ngeno

516. Investigation of host genetic role in PCV2 infections 2142
H.R. Wijesena, K.M. Sutton, S.D. Kachman and D.C. Ciobanu

517. Olfaction: a critical force driving adaptive evolution in yaks 2147
A. Kour, S.K. Niranjan, S. Jayakumar, M. Pukhrambam and M. Sarkar

518. Functional analysis of selective sweeps to identify phenotypic traits in South African Nguni cattle 2151
A.A. Zwane, K.S. Nxumalo, M.L. Makgahlela, E. Van Marle-Koster and N. Maiwashe

519. Fine-mapping of candidate causal genes for tail length in the Merinolandschaf breed 2154
D.K. Lagler, E. Kunz, K. Eck, D. Seichter, I. Russ, C. Mendel, G. Lühken, M. Upadhyay and I. Međugorac

520. Massive detection of cryptic recessive genetic defects in livestock mining millions of life trajectories 2158
A. Guintard, F. Besnard, M. Cano, C. Hozé, H. Leclerc, M. Boussaha, C. Grohs, A. Barbat, S. Fritz, C. Escouflaire, J. Rivière, C. Péchoux, C. Danchin-Burge, G. Foucras, A. Relun, V. Plassard, M.-A. Arcangioli, S. Mattalia, D. Boichard and A. Capitan

521. A *de novo* missense mutation of COL1A1 causes osteogenesis imperfecta type 2 and premature delivery in Normande Cattle 2162
C. Grohs, J. Corbeau, J. Jourdain, M. Boussaha, V. Plassard, A. Barbat, C. Hamelin, J. Mortier, D. Boichard, R. Guatteo and A. Capitan

522. CRISPR-Cas9 gene editing enables functional evaluation of a host *Synaptogyrin-2* missense variant in PCV2b infection 2166
L.R. Walker, H. Vu and D.C. Ciobanu

523. Sequence-based GWAS meta-analyses for beef production traits 2170
M.P. Sanchez, T. Tribout, N. Kadri, P.K. Chitneedi, S. Maak, C. Hozé, M. Boussaha, P. Croiseau, R. Philippe, M. Spengeler, C. Kühn, Y. Wang, H. Pausch and D. Boichard

524. Deciphering genetic variants from whole genome affecting duodenum, liver and muscle transcriptomes in pigs 2174
D. Crespo-Piazuelo, O. González-Rodríguez, M. Mongellaz, H. Acloque, M.-J. Mercat, M.C.A.M. Bink, A.E. Huisman, Y. Ramayo-Caldas, J.P. Sánchez and M. Ballester

525. Detailed analysis of mortality rates in the female progeny of Holstein bulls allows the discovery of new dominant genetic defects 2178
F. Besnard, H. Leclerc, N. Cesbron, M. Boussaha, C. Grohs, N. Jewell, A. Pinton, H. Barasc, M. Femenia, L. Dorso, B. Strugnell, T. Floyd, C. Danchin, R. Guatteo, X. Hubin, S. Mattalia, D. Boichard and A. Capitan

526. Candidate genes for milk, growth and immune system traits in Brazilian Iberian-derived locally adapted cattle breeds 2182
L.L. Verardo, M.A. Machado, J.C.C. Panetto and M.V.G.B. Silva

527. Copy number variation analysis in Nguni and Bonsmara crossbred cattle 2187
B. Bhika Kooverjee, P. Soma, F.W.C Neser, M.A. van der Nest and M.M Scholtz

528. Resequencing of casein genes in Swedish Red Cattle with varying milk coagulation properties 2191
T.K. Vanhala, F. Lewerentz, L. Buhelt Johansen, C. Sibersen, M. Paulsson, M. Glantz and D.J. de Koning

529. Major histocompatibility complex class II DRB gene Exon-2 of Indian camel breeds show limited genetic diversity 2194
V. Prakash, B. Jyotsana and S. Suthar

530. Transcriptome profile of liver from pigs fed with two different levels of soybean oil revealed the NF-κB1 modulation 2199
S.L. Fanalli, B.P.M. da Silva, F.N. Ciconello, J. Reecy, J. Koltes, D. Koltes, L.C.A. Regitano, J.C.C. Balieiro, L.L. Coutinho, G.B. Mourão, G.C.M. Moreira, V.V. de Almeida, A. Luchiari Filho and A.S.M. Cesar

531. SPP1 gene polymorphisms within Intron-IV and Exon-IV region and their association with milk traits in Gaolao cattle 2203
R.R. Kate, D.S. Kale, J. Singh and D.V. Patil

532. Building *in vitro* tools for livestock genomics: chromosomal variation within the porcine PK15 cell line 2208
M. Johnsson, J.M. Hickey and M.K. Jungnickel

533. Conserved non-coding elements of the bovine genome 2212
N.K. Kadri and H. Pausch

534. A massively parallel reporter assay to screen bovine regulatory variants 2216
M.D. Littlejohn, T.J. Lopdell, A. Trevarton, J. Moody, K.M. Tiplady, K.A. Burborough, C.P. Prowse-Wilkins, A.J. Chamberlain, M. Goddard and R.G. Snell

535. Detection of causative genes for resistance against cardiomyopathy syndrome in Atlantic salmon using omics data 2220
M.L. Aslam, S. Gonen, S.A. Boison, E. Leder, M. Baranski, A. Norris, N. Robinson, S. Koga, T.K.K. Østbye and B. Dagnachew

536. Genetic regulation of ear size in Slovenian and Greek sheep breeds 2224
J. Klawatsch, D. Papachristou, P. Koutsouli, D. Seichter, I. Russ, M. Simčič, I. Bizelis and I. Medugorac

537. Large-scale 'omics fine-mapping identifies expression quantitative trait loci significantly affecting cattle phenotypes 2228
R. Xiang, M.E. Goddard, C.P. Prowse-Wilkins, C.M. Reich, B.A. Mason, J.B. Garner, L.C. Marett, I.M. MacLeod and A.J. Chamberlain

Part 32 – Functional annotation of genomes: structural variation and variants 2232

538. A functional copy number gain at indicine cattle footprint modulates actin-myosin and immune response genes in muscle 2233
V.H. da Silva, J. Afonso, N.S. Morosini, F.C. da Veiga, A.S.M. Cesar, B. Silva-Vignato, G.B. Mourão, A.Z. Neto, L.C.A. Regitano and L.L. Coutinho

539. Functional SNPs and INDELs within regulatory elements associated with mastitis in Holstein cow using -OMICs technologies 2237
V. Asselstine, J.F. Medrano, P. Stothard, F. Miglior, N.A. Karrow, C.F. Baes, F.S. Schenkel and A. Cánovas

540. Allele specific expression as an indication of ploidy in pig IPECJ2 and chicken SL-29 cell lines 2241
J. de Vos, R.P.M.A. Crooijmans, M.F.L. Derks, S.L. Kloet, M.A.M. Groenen and O. Madsen

541. Segregation of the structural variant responds for mastitis resistance in re-sequencing Nordic Holstein animals 2245
Z. Cai and G. Sahana

542. Rare CNVs in the bovine genome are not captured well by 50K density genotyping array SNPs 2249
Y.L. Lee, M. Bosse, W. Coppieters, R.F. Veerkamp, L. Karim, C. Oget-Ebrad, T. Druet, M.A.M. Groenen, M. Georges, A.C. Bouwman and C. Charlier

543. Genome-wide DNA methylation of Gonad in sex-biased zebrafish families 2253
S. Hosseini, N. Trakooljul, M. Hirschfeld, K. Wimmers, J. Tetens, A.R. Sharifi and B. Brenig

544. An approach to study the association between the blood cell methylome with feed efficiency traits 2257
A. López-Catalina, A. Bach, M. Gutiérrez-Rivas, R. Peiró-Pastor and Ó. González-Recio

545. Multi-dimensional functional annotation of bovine genome for the BovReg project 2261
G.C.M. Moreira, S. Dupont, D. Becker, M. Salavati, R. Clark, E.L. Clark, G. Plastow, C. Kühn and C. Charlier on behalf of the BovReg consortium

546. European project AQUA-FAANG: the epigenetic landscape of the Atlantic Salmon; focus on liver tissue2265
M.-O. Baudement, G.B. Gillard, T. Podgorniak, L. Grønvold, S. Lien and M.P. Kent

547. Integration of methylome and transcriptome profiles reveals candidate biomarkers for *S. uberis*-subclinical mastitis 2269
M.Q. Wang, M. Laterrière, P.L. Dudemaine, N. Bissonnette, D. Gagné, X. Zhao, J.P. Roy, M.A. Sirard and E.M. Ibeagha-Awemu

548. eQTL mapping in Brown Swiss bulls to identify variants associated with male fertility 2273
X.M. Mapel, M. Hiltpold, A. Lloret-Villas and H. Pausch

549. Bioinformatics workflow for the detection of eQTL in the cattle genome using Nextflow DSL2 2277
P.K. Chitneedi, F. Hadlich and C. Kuehn

550. Extensive functional genomics information from early developmental time points for pig and chicken 2281
H. Acloque, P.W. Harrison, W. Lakhal, F. Martin, A.L. Archibald, M. Beinat, M. Davey, S. Djebali, S. Foissac, S. Guizard, C. Guyomar, R. Kuo, C. Kurylo, O. Madsen, K. Miedzinska, M. Mongellaz, J. Smith, J. Smith, A. Sokolov, J. de Vos, E. Giuffra and M. Watson

551. Demonstrating the use of a mammalian methylation microarray for epigenome wide association studies in livestock 2285
A. Caulton, K.G. Dodds, K.M. McRae, C. Couldrey, S. Horvath and S.M. Clarke

552. Regulatory QTL and exon expression QTL in the mammary gland of dairy cows 2289
C.P. Prowse-Wilkins, T.J. Lopdell, R. Xiang, C.J. Vander Jagt, M.D. Littlejohn, A.J. Chamberlain and M.E. Goddard

553. H3K27me3 predicted regulation of gene expression linked to calcium quantification in Nelore muscle 2293
J. Afonso, W.J. Shim, M. Boden, N. Palpant, M.R.S. Fortes, W.J.S. Diniz, A.O. Lima, M.I.P. Rocha, T.F. Cardoso, J.J. Bruscadin, C.F. Gromboni, A.R.A. Nogueira, G.B. Mourão, A. Zerlotini, L.L. Coutinho and L.C.A. Regitano

554. Comparative analysis of CAGE-Seq across tissues reveals transcription start sites unique to cattle 2297
M. Salavati, R. Clark, D. Becker, C. Kühn, G. Plastow, G.C.M. Moreira, C. Charlier and E.L. Clark on behalf of the BovReg consortium

555. Transcriptomic characterisation of claw horn disruption lesions in the peripheral blood leucocytes of dairy cattle 2301
X. Dai, D. Xia, B.E. Griffiths, A. Anagnostopoulos, M. Barden, B. Li, M. Coffey, D. Werling, G. Oikonomou, G. Banos, and A. Psifidi

556. Evidences of INFt and ISG gene expression in leukocites of pregnant hair sheep ewes from a thermal stress experiment 2305
J.A. Gómez-Guzmán, J.F. Vázquez-Armijo, A.L. Lara-Rivera, A.M. Sifuentes-Rincón, R. Ramírez-Valverde, R. Nuñez-Domínguez, J.C. Martínez-González, W. Arellano-Vera, G.M. Parra-Bracamonte

557. Identification of differentially expressed transcription factors in Brangus skin infested with *Rhipicephalus australis* 2309
E. Mantilla Valdivieso, L. Nguyen, E. Ross, A. Raza, P. James, B. Hayes, N. Jonsson and A. Tabor

558. Unravelling regulatory variants affecting gene expression in four porcine tissues 2313
M. Gòdia, M.F.L. Derks, B. Harlizius, O. Madsen and M.A.M. Groenen

559. Unravelling the transcriptomic control of growth traits in sheep during prenatal and postnatal development 2317
S.A. Woolley, M. Salavati and E.L. Clark

560. Large-scale cis-eQTL analysis of gene expression in blood of young healthy pigs using PigGTEx 2321
L.M. Kramer, J. Teng, K.S. Lim, Y. Gao, H. Yin, L. Bai, G.E. Liu, Z. Zhang, L. Fang, G.S. Plastow, C.K. Tuggle and J.C.M. Dekkers

561. Unravelling the structural variability in chicken genomes by long-read sequencing 2325
J. Geibel, J. Schauer, A. Weigend, C. Reimer, D.-J. de Koning, H. Simianer and S. Weigend

562. Detection and characterisation of *de novo* structural variants in pigs 2329
M. Steensma, Y.L. Lee, A.C. Bouwman, C. Pita Barros, M.F.L. Derks, M.C.A.M. Bink, B. Harlizius, A.E. Huisman, R.P.M.A. Crooijmans, M.A.M. Groenen, H.A. Mulder and C.M. Rochus

563. Efficient detection of interchromosomal rearrangements in Holstein bulls using large SNP datasets 2333
J. Jourdain, A. Pinton, H. Barasc, A. Barbat, C. Hozé, M. Philippe, D. Boichard and A. Capitan

564. A pan-genome of commercial pig breeds 2337
M.F.L. Derks, A. Boshove, B. Harlizius, E. Sell-Kubiak, M.S. Lopes, E. Grindflek, E. Knol, M.A.M Groenen and A.B. Gjuvsland

Part 33 – Aquaculture 2342

565. Conserved QTL and chromosomal inversions affect resistance to columnaris disease in two rainbow trout populations 2343
F.C.F. Calboli, H. Koskinen, A. Nousianen, C. Fraslin, R.D. Houston and A. Kause

566. Multi-trait genome wide association study in correlated traits: fillet colour and body weight in Atlantic salmon 2347
P. Sae-Lim, S. Boison and M. Baranski

567. Implementation of genomic selection on production and quality traits and linkage disequilibrium in *Crassostrea gigas* 2351
A. Jourdan, P. Boudry, F. Enez, P. Haffray, F. Chenier, E. Vetois, F. Allal, P. Leroy, F. Phocas, J. Bugeon, L. Degremont and R. Morvezen

568. Genetic correlation between IP and cohabitation challenge test model for *Piscirickettsia salmonis* in Atlantic salmon 2355
C. Manosalva, M.L. Pavez, P. Sae-Lim, A. Norris, M. Baranski and S.A. Boison

569. Using 'population-specific haplotypes' to select for resistance to SRS in an admixed population of coho salmon (*Oncorhynchus kisutch*) 2359
T. Moen, J. Ødegård, F. Grammes, T. Knutsen, R. Deerenberg and D. Chichero

570. Potential of imputation for cost-efficient genomic selection for resistance to *Flavobacterium columnare* in Rainbow trout 2363
C. Fraslin, A. Nousiainen, H. Koskinen, A. Kause and R.D. Houston

571. Genomic prediction of complex traits in European sea bass: random cross-validation and leave-one-family-out validation 2367
S. Faggion, D. Bertotto, V. Bonfatti, M. Freguglia, L. Bargelloni and P. Carnier

572. Genetics of parasite attraction in Atlantic salmon: potential for group-level protection against sea lice 2371
J. Ødegård, S. Kjøglum, M. Medina, J.S. Torgersen, S.A. Korsvoll, R. Deerenberg, J.M. Yáñez, D. Cichero, P. Lopez and T. Moen

573. A major QTL affects resistance to viral nervous necrosis in farmed European seabass 2375
R. Mukiibi, D. Robledo, C. Peñaloza, S. Ferraresso, R. Franch, D. Bertotto, M. Freguglia, S. Laureau, F. Pascoli, A. Toffan, C. Tsigenopolous, L. Bargelloni and R. Houston

574. The potential indicator for resilience of Sukamandi strain tilapia grown in brackish water ponds 2379
M.H.F. Aththar, P. Setyawan, I. Imron, B. Gunadi, M. Camara, H. Komen and J.W.M. Bastiaansen

575. Exploiting genomics to infer population structure and genetic variability in wild and farmed seabream and seabass 2383
B. Villanueva, A. Fernández, R. Peiro-Pastor, C. Peñaloza, R.D. Houston, A.K. Sonesson, C.S. Tsigenopoulos, L. Bargelloni, K. Gamsiz, B. Karahan, E.Ö. Gökçek, J. Fernández and M. Saura

576. Comparison of tank and commercial sea-pen family evaluation of Chinook salmon (*Oncorhynchus tshawytscha*) in New Zealand 2387
M.R. Scholtens, K.G. Dodds, S.M. Clarke, S.P. Walker, M.L. Tate, R. Miller, T. Slattery and J.E. Symonds

577. Testing fat content or feeding behaviour as indirect selection criteria for feed efficiency using trout isogenic lines 2391
S. Pouil, T. Kernéis, E. Quillet, D. Lallias, L. Labbé, F. Phocas and M. Dupont-Nivet

578. The possibilities of using electronic sensors in aquaculture breeding 2395
W. Abbink, A. Palstra, W. Agbeti, G. Lembo and J. Komen

579. Genetic parameters and genome-wide association study of resistance to acute hyperthermia in rainbow trout 2399
H. Lagarde, D. Lallias, P. Patrice, M. Prchal, Y. François, J. D'Ambrosio, A. Acin-Perez, F. Cachelou, P. Haffray, M. Dupont-Nivet and F. Phocas

580. QTLs controlling swimming performance and their effect on growth in Nile tilapia (*Oreochromis niloticus*) 2403
X. Yu, S. Bekele Mengistu, H.A. Mulder, A.P. Palstra, J.A.H. Benzie, T. Quoc Trinh, M.A.M. Groenen, H. Komen and H.J. Megens

581. Optimizing genotyping effort in aquaculture breeding programs by pre-selection of candidates 2407
C. Roozeboom, B. Gulzari, H. Komen, M. Camara and J.W.M. Bastiaansen

582. Prediction of production traits by using body features of gilthead seabream (*Sparus aurata*) obtained from digital images 2412
B. Gulzari, A. Mencarelli, C. Roozeboom, H. Komen and J.W.M. Bastiaansen

583. Comparison of genomic selection barramundi breeding schemes for multiple environments with strong G×E 2416
M. Lillehammer, N.A. Robinson, D.B. Jones, C. Massault, S. Loughnan, H.S. Cate, P.J. Harrison, J.M. Strugnell, K.R. Zenger and D.R. Jerry

Part 34 – Avian **2420**

584. Genome-wide association studies of chicken caecal microbiota 2421
X. Dai, A.T. Hinsu, C. Dadousis, R.J. Pandit, M. Crotta, G. Limon, B. Fosso, J. Guitian, F.M. Tomley, P.G. Koringa, C.G. Joshi, D.P. Blake and A. Psifidi

585. Breeding for an optimised gut microbiota composition for the improvement of phosphorus utilisation in Japanese quail 2425
V. Haas, S. Vollmar, A. Camarinha-Silva and J. Bennewitz

586. High-resolution genotyping and analysis of *in silico* predicted proteins for classical MHC genes in Korean native chicken 2429
T.K. Ediriweera, P. Manjula, E. Cho, M. Kim, S. Cho and J.H. Lee

587. Genetic architecture of body weight and carcass traits in Ghanaian local chickens 2433
R.A. Kanlisi, E.N. Amuzu-Aweh, M. Walugembe, A. Naazie, H.R. Otsyina, N. Chouicha, T.R. Kelly, R. Gallardo, J. Dekkers, S. Lamont, H. Zhou and B.B. Kayang

588. Estimating environmental direct-maternal covariance in real data – illustration on broilers 2437
H. Romé, T.T. Chu, D. Marois, C.H. Huang, P. Madsen and J. Jensen

589. Host transcriptome response of meat-type chickens infected with *Eimeria maxima* and reared under heat stress environment 2441
S.E. Aggrey, A.F.A. Ghareeb, A.L. Fuller, M.C. Milfort, J.C. Foutz, G.H. Schneiders, W.S. Habashy, K. Adomako and R. Rekaya

590. Changes in genetic parameters for traits under genomic selection in poultry 2445
J. Richter, J. Hidalgo, V. Breen, R. Hawken, I. Misztal and D. Lourenco

591. Effects of age, line and diet on the occurrence of white striping lesions in broiler breast muscle 2448
L.T. Fernandes, J.O. Peixoto, M.S. Dal Pizzol, A.M.G. Ibelli, M.A.Z. Morés, F.C. Tavernari, A. Coldebella, M.E. Cantão and M.C. Ledur

592. Variation in male autosomal recombination in turkeys (*Meleagris gallopavo)* 2452
C.M. Rochus, B.J. Wood and C.F. Baes

593. Functional annotations of regulatory elements in the chicken genome 2456
Z. Pan, Y. Wang, L. An, Y. Bi, D. Guan, M. Delany, H.H. Cheng and H. Zhou

594. Time-series transcriptome analysis of broilers under *Eimeria tenella* infection reveals host-pathogen interactions 2460
M. Kim, Y. Chung, P. Manjula, D. Seo, S. Cho, E. Cho, T.K. Ediriweera, M. Yu, S. Nam, J.M. Heo and J.H. Lee

595. Genetic diversity of *Eimeria tenella* apical membrane antigen 1 in chickens from KwaZulu-Natal, South Africa 2464
A.J. Fatoba, O.T. Zishiri, D.P. Blake, S.O. Peters, J. Lebepe, S. Mukaratirwa and M.A. Adeleke

596. Chicken quantitative traits follow the omnigenic model 2467
P. Davoodi, A. Ehsani, R. Vaez Torshizi and A.A Masoudi

597. Association of polymorphism of *MAP3K8* with egg production traits of White Muscovy ducks (*Cairina moschata*) 2471
S.F. Bello, H.Xu, S. Zhang, K.D. Aloryi, R.O. Ahmed, A.A. Bahareldin, A.C. Adeola, M.A. Popoola, W. Zhu, D. Zhang, X. Zhang, C. Ji and Q. Nie

598. Genetic markers associated with bone quality in Rhode Island Red laying hens 2475
M. Sallam, H.A. McCormack, R.H. Fleming, P.W. Wilson, B. Andersson, M. Schmutz, C. Benavides, N. Dominguez-Gasca, E. Sanchez-Rodriguez, A.B. Rodriguez-Navarro, I.C. Dunn, D.J. De Koning and M. Johnsson

599. Determining the heritability of premaxillary bone shape and size within two populations of pure line laying hens 2479
S. Struthers, B. Andersson, M. Schmutz, H.A. McCormack, P.W. Wilson, I.C. Dunn, V. Sandilands and J.J. Schoenebeck

600. No G×E on egg qualities and body weight between cage and floor systems, implications for breeding programmes in laying hens 2483
N. Bédère, L. Bécot, T. Burlot and P. Le Roy

601. Growth and laying performance of local chicken breeds and their crosses fed with faba beans 2487
A.R. Sharifi, T. Nolte, S. Jansen, S. Weigend, D. Mörlein, I. Halle, W. Link and H. Simianer

602. Heterosis of egg-laying performance in chickens 2491
A.X. Ni, M.P.L. Calus, H. Bovenhuis, Y.Y. Sun and J.L. Chen

603. *In vitro* effects of probiotic *Bacillus* strains on adaptive immune cells of broiler chicken 2495
F. Larsberg, M. Sprechert, D. Hesse, G. Loh, G.A. Brockmann and S. Kreuzer-Redmer

604. Strain and age effects, as well as genetic parameters for adult traits in ostriches 2499
S.W.P. Cloete, A. Engelbrecht and Z. Brand

Part 35 – Bees and other insects **2503**

605. Economic value and production characteristics of table honey 2504
E.A. Frost, N.C. Chapman, R.G. Banks, S.F. Walkom and S. Hermesch

606. Evaluating the potential for mating control in honey bee breeding in three SE European countries (preliminary results) 2508
A. Uzunov, S. Andonov, B. Dahle, M. Kovačić, J. Prešern, G. Aleksovski, F. Jaman, M. Jovanovska, B. Pavlov, Z. Puškadija, J. Wegener and R. Büchler

607. A model for, and early implementation of, genomic selection in the Australian honey bee population 2512
N.C. Chapman, E.A. Frost and R.G. Banks

608. An over-time distribution map of the main *Apis mellifera* mitochondrial DNA lineage using honey as source of eDNA 2516
V. Taurisano, V.J. Utzeri, A. Ribani and L. Fontanesi

609. Using genomics to predict drone quality: why are there so many 'dud' male honey bees? 2520
G.P. Slater and B.A. Harpur

610. Development of genomic resources in black soldier fly (*Hermetia illucens* L.) via throughput DNA pool sequencing 2524
A.S.L. Donkpegan, A. Guigue, F.X. Boulanger, S. Brard-Fudulea, P. Haffray, M. Sourdioux and R. Rouger

611. The natural history of the black soldier fly, *Hermetia illucens*: insights from complete mitochondrial genome sequences 2528
J. Guilliet, G. Baudouin, N. Pollet and J. Filée

612. Impact of genetic selection in insect populations using different selection designs – a simulation study 2533
R.S.C. Rikkers, J.W.M. Bastiaansen, A.C. Bouwman and E.D. Ellen

613. Genetic parameters of black soldier flies estimated in full sib design 2537
A.C. Bouwman, J.E. Nugroho, D. Wongso, J. van Schelt, B.A. Pannebakker, B.J. Zwaan and E.D. Ellen

614. Towards selective breeding in insects – estimating genetic parameters with individual-level phenotypes and pedigree 2541
L.S. Hansen, S.F. Laursen, S. Bahrndorff, M. Kargo, T.N. Kristensen, G. Sahana, J.G. Sørensen and H.M. Nielsen

615. Genetic improvement program in black soldier fly for increased larval weight in industrial applications 2545
E. Facchini, K. Shrestha, E. van den Boer, P. Junes, G. Sader, K. Peeters and E. Schmitt

616. *Tenebrio molitor* genomic structure among available populations 2549
E. Eleftheriou, B. Vacherie, K. Labadie, C.G. Athanassiou, T. Rigaud, Y. Moret, T. Lefebvre and MA. Madoui

617. Inferring whole-genome tree sequences and population and demographic parameters of the Western honeybee 2552
J. Obšteter, A. Marinč, J. Prešern, D. Wragg and G. Gorjanc

618. Heritability of disease resistance to chronic bee paralysis, chalkbrood and nosemosis in the honeybee (*A.m.carnica*) 2556
A. Hoppe, R. Bernstein, M. Du and K. Bienefeld

619. Genetic analysis of royal jelly production and behaviour traits of honeybees 2560
B. Basso, T. Kistler, T. Gerez and F. Phocas

620. Comparison of heritabilities and estimated breeding values for three honeybee traits using dam and animal models 2564
G.E.L. Petersen, F.S. Hely, P.F. Fennessy and P.K. Dearden

621. Statistics for an accurate genome wide association study on *Varroa* resistance trait in a French honeybee 2568
S.E. Eynard, A. Vignal, Y. Agez, B. Basso, O. Bouchez, T. Bulach, Y. Le Conte, B. Dainat, A. Decourtye, L. Genestout, M. Guichard, F. Guillaume, E. Labarthe, R. Mahla, F. Mondet, M. Neudischko, F. Phocas, Y. Poquet, C. Sann, R.F. Serre, K. Tabet and B. Servin

622. New approach to identify Mendelian inconsistencies between SNP and pedigree information in the honey bee 2572
R. Bernstein, M. Du, A. Hoppe and K. Bienefeld

623. Investigation of parental meiotic contribution in honeybees (*Apis mellifera*) 2576
G.E.L. Petersen, A. Heywood, A.C.L. Walters, P.F. Fennessy and P.K. Dearden

624. SIMplyBee: stochastic simulator of honeybee populations and breeding programmes 2580
L. Strachan, J. Bubnič, J. Obšteter and G. Gorjanc

625. Assessment of uncertainty in direct and maternal genetic parameters estimates for honeybee colony phenotypes 2584
T. Kistler, E.W. Brascamp, B. Basso, P. Bijma and F. Phocas

626. Very local genotype by environment interaction in Austrian honey bees 2588
E.W. Brascamp, M. Rubinigg, R.F. Veerkamp and P. Bijma

Part 36 – Bovine beef 2592

627. Genetic prediction for growth traits in beef cattle using selected variants from imputed low-pass sequence data 2593
C.A. Russell, L.A. Kuehn, W.M. Snelling and M.L. Spangler

628. *Bos indicus* genome percentage effects gene expression profile and differentially expressed genes in Brahman cattle whole blood 2597
E.M. Ross, L.T. Nguyen, B.N. Engle, A.J. Chamberlain, E. Kho and B.J. Hayes

629. Effect of one copy of nt821 mutation in myostatin gene on ultrasound carcass traits in Senepol cattle 2601
G.R.O. Menezes, C.B. Souza, A.L.R. Lima, A. Gondo, F. Siqueira, A.A. Egito, L.N. Silva, J.A. Fernandes Junior and R.A.A. Torres Junior

630. Allelic diversity and effect of the intronic TBX20-191081 marker on temperament in beef cattle 2605
V.M. Del Real-García, A.M. Sifuentes-Rincón, F.A. Paredes-Sánchez, G. Ruíz-de la Cruz, E. Garza-Brenner, G.M. Parra-Bracamonte, H. Bernal-Barragán and F.A. Rodríguez-Almeida

631. Innovative phenotyping strategies exploiting immune response to heat stress conditions 2608
P. Lemal, M. Schroyen and N. Gengler

632. Epistatic interaction networks affecting feed efficiency and methane emission in Nelore cattle 2612
T.F. Cardoso, J.J. Bruscadin, W.J.S. Diniz, P. Banerjee, J. Afonso, J. Petrini, B.G.N. Andrade, J.M. Malheiros, T. Porto, A. Zerlotini, G.B. Mourão, L.L. Coutinho and L.C.A. Regitano

633. Host-genomically influenced ruminal microbial genes are temporally stable during the finishing phase in beef cattle 2616
J. Lima, M. Martínez-Álvaro, J. Mattock, M.D. Auffret, C.A. Duthie, M.A. Cleveland, R.J. Dewhurst, M. Watson and R. Roehe

634. A hypomorphic mutation in the *ATP2A1* gene increases muscle mass yet compromises meat quality of Belgian Blue cattle 2620
T. Rombouts, T. Druet, J.L. Gualdrón Duarte, N. Ahariz, K. Karim, W. Coppieters, S. De Smet, M. Georges and C. Charlier

635. Mitochondrial DNA copy number as a potential genetic indicator of growth and carcass traits in beef cattle 2624
L.P. Sanglard, L.A. Kuehn, W.M. Snelling, and M.L. Spangler

636. Genomic regions in Australian cattle associated with consumer satisfaction of beef 2628
A.M. Lynn, P. McGilchrist, H. Aliloo, J.H.J. van der Werf, R. Polkinghorne and S.A. Clark

637. A genome-wide scan for signatures of selection in Hanwoo and Angus cattle using whole-genome sequence data 2632
H.J. Kim, S. de las Heras-Saldana, S. Clark and J.H.J. van der Werf

638. Genetic analyses reveal the genetic architecture of the thermotolerance – production complex in beef cattle 2636
R.G. Mateescu, F.M. Rezende, K.M. Sarlo Davila, A.N. Nunez Andrade, A. Hernandez, M.A. Elzo and P.A. Oltenacu

639. The effect of allele ancestry on hip height and body weight of tropical beef cattle 2640
L.R. Porto-Neto, S. McWilliam, P.A. Alexandre, M.R.S. Fortes and A. Reverter

640. Mining of producer recorded data; using beef calf and cow live-weight data as a case study 2644
S. Walsh, N. McHugh, R.D. Evans, C. Murphy and D.P. Berry

641. Sexual precocity of purebred Charolais heifers 2648
S. Taussat, A. Vinet, D. Maupetit, D. Dozias, L. Chanteloup, G. Renand and P. Martin

642. Estimation of variance components for female longevity in Australian Angus cattle using random regression models 2652
H. Aliloo and S. Clark

643. Eye muscle area and lean muscle area are genetically different traits in high marbling Wagyu carcasses 2656
R.A. McEwin, M.L. Hebart, R. Tearle, J. Grose and W.S. Pitchford

644. Selection on yearling days to calving improves mature fertility traits in tropically adapted cattle 2660
M.L. Facy, M.L. Hebart, H. Oakey, R.A. McEwin,G.I. Popplewell and W.S. Pitchford

645. Estimation of maternal variance for slaughter traits 2664
A.J. Twomey and T. Pabiou

646. Biological types, direct and maternal heterosis and recombination effects on weaning weight of composite beef cattle 2668
J.B.S. Ferraz, R. Espigolan, F.S. Baldi, J.P. Eler, E.C.M. Oliveira, R. Nuñez-Dominguez and L.T. Gama

647. Breed differences, heterosis and recombination effects for calving interval in a composite beef cattle breed 2672
L.T. Gama, R. Espigolan, F.S. Baldi, J.P. Eler, E.C.M. Oliveira, R. Nuñez-Dominguez and J.B.S. Ferraz

648. Prediction ability for growth and maternal traits using SNP arrays based on different marker density in Nellore cattle 2676
F. Baldi, J.D. Rodríguez, R. Espigolan, R.B. Lôbo, R. López-Correa and I. Aguilar

649. Applying single step genomic evaluation and integrating local with foreign EBVs to increase calving performance 2680
S. Naderi, T. Pabiou and R.D. Evans

650. Understanding the genetics of fertility and temperament in Northern beef cattle using genomic technologies 2684
J.P. Copley, N.J. Corbet, J.M. Allen, A.R. Laing, G. Fordyce, M.R. McGowan, B.M. Burns, R.E. Lyons and B.J. Hayes

651. Genomic prediction with non-additive genetic effects for carcass weight and beef marbling in Japanese Black cattle 2688
K. Inoue, M. Nishio, Y. Inoue, M. Takeda and H. Hirooka

652. Reaction norm models contribute to the investigation of environmental sensitivity for weaning weight in Nellore cattle 2692
I. Carvalho Filho, D.A. da Silva, C.S. Teixeira, T.L. Silva, L.G. Albuquerque and R. Carvalheiro

653. Investigation on inheritance transmission modes using breeding values estimated from genomic or pedigree information 2696
H. Hirooka, M. Nishio, K. Inoue, M. Takeda and Y. Inoue

654. Genetic parameters for fatty acid traits of beef in Australian Angus 2700
C.J. Duff, A.M. Samaraweera, P. McGilchrist and S.A. Clark

655. Integration of beef cattle international estimated breeding values in the Italian evaluation 2704
R. Bonifazi, M.P.L. Calus, J. ten Napel, R.F. Veerkamp, S. Biffani, M. Cassandro, S. Savoia and J. Vandenplas

656. A genomic evaluation of genotype-by-environment interactions for foot scores in American and Australian Angus cattle 2708
A.B. Alvarenga, K. Retallick, A. Garcia, H.R. Oliveira, S.P. Miller, A. Byrne, L.F. Brito

657. Single-step GBLUP evaluation for categorical traits for beef cattle 2712
A. Nazari-Ghadikolaei, S. Eriksson and W.F. Fikse

658. Genetic parameters for mature cow body condition scores and fat and muscle traits in finishing beef cattle 2716
F. Weik, R.E. Hickson, S.T. Morris, D.J. Garrick and J.A. Archer

659. Decision support tools to support a more sustainable beef-on-dairy industry 2720
S.C. Ring, R.D. Evans, J. McCarthy, A. Cromie, and D.P. Berry

Part 37 – Bovine dairy – genetic evaluation methods 2724

660. Breed and heterosis effects for fatty acid composition of milk from dairy cows milked once- and twice-a-day 2725
I. Sanjayaranj, N. López-Villalobos, H.T. Blair, P.W.M. Janssen, S.E. Holroyd and A.K.H. MacGibbon

661. Long term effect of growth in rearing phase on reproduction and production trait in Murrah buffalo 2729
P. Joshi, A. Verma, G.R. Gowane, R. Alex, I.D. Gupta and D. Worku

662. Integrating parent-of-origin effects in random regression models for milk traits 2733
R.E. Jahnel, I. Blunk and N. Reinsch

663. Recombination rate and hotspots in Holstein cattle from Mexico 2737
E.I. Mejía-Melchor, A. García-Ruiz, C. Vásquez-Peláez, H.O. Toledo-Alvarado and F.J. Ruiz-Lopez

664. Longevity in genetic evaluation of Holstein cows in Uruguay 2741
N.F. Silveira, O. Ravagnolo and I. Aguilar

665. Comparison of two models for dairy herd improvement data in Taiwan 2745
Y.H. Lan, K.H. Lin, S.H. Ting, C.L. Su, P.A. Tu and E.C. Lin

666. Comparison of random regression test-day models for production traits of South African Jersey Cattle 2749
M.G. Kinghorn, E.D. Cason, V. Ducrocq and F.W.C. Neser

667. ROH describes the small holder dairy systems with Holstein cattle in Mexico 2753
A. García-Ruiz, J.G. Cortés-Hernández and F.J. Ruiz-López

668. How ssGBLUP became suitable for national dairy cattle evaluations 2757
I. Misztal D. Lourenco, S. Tsuruta, I. Aguilar, Y. Masuda, M. Bermann, A. Cesarani and A. Legarra

669. Options for evaluating multiple breeds in a single-step GBLUP for US dairy population 2761
A. Cesarani, D. Lourenco, P.M. VanRaden, E.L. Nicolazzi, A. Legarra, S. Tsuruta and I. Misztal

670. Integrating MACE breeding values into multiple-trait single step random regression test day evaluations of Aussie Reds 2765
V. Boerner, T. Nguyen and G. Nieuwhof

671. Optimum contribution selection in a dairy cattle population with different relationship matrices 2769
E. Gautason, G. Sahana, B. Guldbrandtsen and P. Berg

672. The rising popularity of embryo transfer in U.S. dairy cattle and implications for national fertility evaluations 2773
A.M. Miles, J.L. Hutchison and P.M. VanRaden

673. Simulating technical, economic and environmental performance of a dairy herd under selection in future climate conditions 2777
J.G. Burns, V. Eory, R. Mrode, G. Simm and E. Wall

Part 38 – Bovine dairy – genome informed breeding 2781

674. Understanding the genetic architecture of claw horn lesions in different lactation stages in Holstein cattle 2782
B. Li, M. Barden, V. Kapsona, E.S. Molano, A. Anagnostopoulos, B.E. Griffiths, C. Bedford, X. Dai, M. Coffey, A. Psifidi, G. Oikonomou and G. Banos

675. Shared response pattern in the mammary transcriptome after intramammary challenge with *E. coli* or *S. aureus* pathogen 2786
D. Becker, A. Heimes, R. Weikard, M. Hoedemaker, M.M. Meyerholz, H.J. Schuberth, W. Petzl, H. Zerbe, M. Schmicke, S. Engelmann and C. Kühn

676. Epistatic interaction between *KIT* and *MITF* causes splotchy face in Hereford × dairy crossbred calves 2790
S. Jivanji, A. Yeates, C. Harland, C. Gray, C. Couldrey, G. Worth, L. McNaughton, T. Fransen, T. Monehan, R. Spelman, D. Garrick and M. Littlejohn

677. Searching for genes involved in heat stress response in lactating cows through transcriptomics 2794
C. Diaz, M. Ramón, C. Gonzalez-Verdejo and M.J. Carabaño

678. Genomic regions exhibiting transmission distortions and copy number variants associated with stillbirth in cattle 2798
A.A.A. Martin, H.R. Oliveira, P.A.S. Fonseca, C.M. Rochus, F. Miglior, S. Id-Lahoucine, A. Cánovas, C.F. Baes and F.S. Schenkel

679. Mating allocations in Nordic Red Dairy Cattle using genomic information 2802
C. Bengtsson, H. Stålhammar, J.R. Thomasen, S. Eriksson, W.F. Fikse and E. Strandberg

680. Genome wide association study of blood cell count measures in Holstein cows 2806
C.J. Siberski-Cooper, J. Chinchilla-Vargas, L.M. Kramer, K.S. Lim, M.S. Mayes and J.E. Koltes

681. Investigation of intestinal atresia in a Jersey sire family 2810
T.R. Carthy, O.M. Keane, J.P. Hanrahan, D. Matthews, J. McEwan, S. Rowe and J. Mee

682. The genomic architecture of feed efficiency over the lactation of dairy cattle 2814
K. Houlahan, F.S. Schenkel, H.R. Oliveira, F. Miglior, G.A. Oliveira Jr., M.J. VandeHaar, K. Weigel, R.J. Tempelman and C.F. Baes

683. Population structure and signatures of selection in Italian Holstein through genome-wide analysis of imputed SNP data 2818
C. Persichilli, M. Marusi, J.T. van Kaam, R. Finocchiaro, G. Senczuk, M. Di Civita, F. Pilla and M. Cassandro

684. Genome-wide association studies for maturity rate in grazing dairy cows 2822
N. Lopez-Villalobos, H.B.P.C. Ariyarathne and R.E. Hickson

685. Selection, validation, and utilization of mitogenome SNP array information in cattle breeding 2826
V. Brajkovic, L. Bradic, K. Turkalj, D. Novosel, S. Ristov, P. Ajmone-Marsan, L. Colli, V. Cubric-Curik, J. Sölkner and I. Curik

686. Genomewide transcriptome profiling of milk derived primary bovine mammary epithelial cells after pathogen challenge 2830
T. Iso-Touru, D. Fischer, M. Kyläniemi, J. Tabell, A. Virta and J. Vilkki

687. Archetypal clustering reveals physiological mechanisms linking milk yield and fertility in dairy cattle 2834
E. Ooi, J.E. Pryce and M.E. Goddard

Part 39 – Bovine dairy – phenotype is king **2838**

688. Candidate genomic regions for forage dry matter intake in Swedish dairy cows using single-step genome-wide association 2839
G.M. Tarekegn, J. Karlsson, C. Kronqvist, K. Holtenius, S. Andonov, B. Berglund and E. Strandberg

689. Stillbirth, dystocia and weaning rates in Australian dairy calves 2843
M.M. Axford, J.E. Pryce, M. Khansefid and M Haile-Mariam

690. The relative importance of linear type traits in determining survival increases as Holstein-Friesian dairy cows age 2847
M. Williams, R.D. Sleator, C.P. Murphy, J. McCarthy and D.P. Berry

691. Genetic parameters of milk composition and fertility traits in cows milked once- or twice-daily in New Zealand 2851
J.M.D.R. Jayawardana, N. Lopez-Villalobos, R.E. Hickson and L.R. McNaughton

692. Identifying phenotypes to enhance stress resilience of dairy cows 2855
T.M. Sullivan, A. Sharma, K. Lamers, C. White, B.A. Mallard, A. Cánovas and N.A. Karrow

693. Estimation of genetic correlations between predicted energy balance and milk fatty acids of Holsteins in Japan 2859
A. Nishiura, O. Sasaki, M. Aihara, Y. Saito, R. Tatebayashi and H. Takeda

694. Associations between milk protein polymorphisms and milk production in Holstein-Friesian cows 2863
D.N. Kelly, F.L. Dunne, R.D. Evans and D.P. Berry

695. Genetic analysis of milking traits in Holstein cows based on automatic milking system data 2867
J. Kawakami, Y. Gotoh, T. Baba, S. Yamaguchi, S. Nakagawa, S. Yamaguchi, T. Oka and T. Kawahara

696. Fertility traits of Mexican Holstein cattle: genetic parameters, correlations and changes in genetic diversity over time 2871
C. Durán-Alvarez, A. García-Ruiz, R.A. Alonso Morales, L.E. Eguiarte , D. Piñero and F.J. Ruiz-López

697. Genetic parameters for residual feed efficiency in three dairy cattle breeds in commercial farms using 3D cameras 2875
C.I.V. Manzanilla-Pech, R.B. Stephansen, T. Andersen and J. Lassen

698. Insemination values to support mating decisions under dairy heifer calf herd size constraints 2879
A. De Vries, N. Bliznyuk, P. Sharma, Y. Han and P. Pinedo

699. Onset of heat stress and development of genomic predictions for heat tolerance in US Holsteins and Jerseys 2883
T.M. McWhorter, M. Sargolzaei, C.G. Sattler, M.D. Utt, S. Tsuruta, I. Misztal and D. Lourenco

700. Detection of genetic variability in cattle infectivity for bovine tuberculosis (bTB) 2887
E. Sánchez-Molano, D. Madenci, S. Tsairidou, M. Winters, A.P. Mitchell, G. Banos and A. Doeschl-Wilson

701. Dairy female fertility traits in seasonal and non-seasonal herds in Ireland 2891
K. Stachowicz, P.R. Amer, M.M. Kelleher, S.C. Ring and R.D. Evans

702. Genetic aspects of age at first calving in Italian Holstein dairy cows 2895
V. Ferrari, F. Galluzzo, R. Finocchiaro, J.B.C.H.M. van Kaam, M. Marusi, M. Penasa and M. Cassandro

Part 40 – Caprine/ovine **2900**

703. Genetic and genomic approaches to ewe's artificial insemination success in the Spanish Assaf and Manchega sheep breeds 2901
M. Serrano, M.A. Jiménez, M. Ramón and O. González-Recio

704. The impact of sire on variation in methane production and dry matter intake in sheep 2905
E. O'Connor, F.M. McGovern, E. Dunne, S.J. Morrison, T.M. Boland and N. McHugh

705. Milk transcriptome analysis to elucidate the impact of prepubertal nutrition in dairy ewes residual feed intake 2909
A. Suárez-Vega, B. Gutiérrez-Gil, G. Hervás, H. Marina, R. Pelayo, P.G. Toral, C. Esteban-Blanco, P. Frutos and J.J. Arranz

706. Inbreeding and effective population size of United States Katahdin sheep 2913
G.M. Becker, J.M. Burke, R.M. Lewis, J.E. Miller, J.L.M. Morgan, B.D. Rosen, C.P. Van Tassell, D.R. Notter and B.M. Murdoch

707. Genetic diversity and population structure of dairy goat populations in Kenya 2917
C.M. Kamidi, R.W. Waineina, C.B. Wasike, D.K. Ngeno and E.D. Ilatsia

708. Resolving the genetic structure of 18 Greek and neighbouring sheep populations 2920
A. Kominakis, E. Tarsani, A. Hager, G. Antonakos, I. Mastranestasis, D. Gkelia and I. Hadjigeorgiou

709. Genetic determinism of quality of lamb leather in Lacaune dairy sheep breed 2924
H. Larroque, E. Dugas, F. Woloszyn, F. Plisson-Petit, J. Gayraud, J.-C. Duchêne, J. Frayssignes, A. Gille, S. Brenot, D. Allain and L. Drouilhet

710. Heritability and genomic analysis of coagulation event in sheep milk 2928
G. Gaspa, A. Cesarani, F. Correddu, M. Congiu, C. Dimauro, A. Pauciullo and N.P.P. Macciotta

711. Genetic parameters across European and North American Alpine goats for two milk production and one udder type traits 2932
M. Teissier, L.F. Brito, F. Schenkel, B. Bapst, G. Bruni, P. Fresi, C. Robert-Granié and H. Larroque

712. Wool colour traits in Gotland sheep and association with a deletion in ASIP 2936
A.M. Johansson

713. Breeding initiatives around reduction of methane emissions in meat sheep – an international review 2940
N.R. Lambe

714. Measuring methane in sheep production systems – phenotypic factors affecting output 2944
F.M. McGovern, E. O'Connor, L. Farrell, E. Dunne, E. Wall and N. McHugh

715. Genetics of methane emissions in a multi-breed sheep population 2948
N. McHugh, E. O'Connor, D.P. Berry, E. Dunne, L. Farrell, P. Rafter, K. McDermott and F.M. McGovern

716. Modelling of productivity, profitability, and greenhouse gas emissions of Irish sheep flocks divergent in genetic merit 2952
L. Farrell, N. McHugh, T. Pabiou, A. Bohan, K. McDermott, L. Shalloo and J. Herron

717. Using milk fatty acids as biomarkers to improve feed conversion ratio in dairy sheep 2956
H. Marina, G. Hervás, R. Pelayo, P.G. Toral, A. Suárez-Vega, B. Gutiérrez-Gil, C. Esteban-Blanco, P. Frutos and J.J. Arranz

718. Six years of genomic selection have increased the genetic gain in French dairy sheep 2960
J.M. Astruc, G. Lagriffoul, A. Legarra and D. Buisson

719. Cross-validation of single-step genetic evaluation in U.S. Katahdin sheep 2964
A.J. McMillan, D.J. Brown, J.M. Burke, J.L. Morgan and R.M. Lewis

720. Partitioning of genetic trends by flock and gender in Istrian sheep breed 2968
M. Spehar, J. Ramljak, V. Poslon and A. Kasap

721. Comparison of GEBV accuracies predicted using the pedigree matrix of female and male potential reference populations 2972
A. Carta, M.G. Usai, S. Casu and S. Salaris

722. Quantification of the effect of incomplete reproduction data on the accuracy of selection 2976
W.J. Olivier

723. Using random regression models to model growth in sheep with irregular data 2980
E. Mutch, M. Coffey, G. Simm and R. Mrode

724. Impact of the order of polynomials in random regression model on the accuracy of genetic evaluation for weight traits 2984
U. Paneru, D.J. Brown, N. Moghaddar and J.H.J. van der Werf

725. Estimation of linkage disequilibrium in Pag and Istrian sheep breed: towards genomic optimum contribution selection 2988
A. Kasap, J. Ramljak and M. Špehar

726. Genome-wide association study for milk composition traits of Canadian Alpine and Saanen goats 2992
E. Massender, H.R. Oliveira, L.F. Brito, L. Maignel, M. Jafarikia, C.F. Baes, B. Sullivan and F.S. Schenkel

727. Evaluating performance of Suffolk, Columbia, and crossbred lambs at birth and weaning 2996
R.M. Lewis, N. Vargas Jurado, D.J. Brown, D.R. Notter and J.B. Taylor

728. Genetic selection for sensory eating quality of lamb using consumer assessments 3000
S.Z.Y. Guy, S.I. Mortimer, L. Pannier, P. McGilchrist, D.J. Brown, D. Pethick and A.A. Swan

729. Genetic characterization and connectedness of dairy goats in Canada, France, Italy and Switzerland 3004
M. Teissier, L.F. Brito, F. Schenkel, G. Bruni, P. Fresi, B. Bapst, C. Robert-Granié and H. Larroque

730. Value of data from ram breeding flocks as an industry reference population for Australian sheep 3008
P. Alexandri, S.F. Walkom, A.A. Swan, J.H.J. van der Werf and D.J. Brown

731. Pairwise principal component analysis of African goat breeds identifies loci involved in meat and/or milk production 3012
S.A. Rahmatalla, D. Arends, G.M. Tarekegn, R. Onzima and G.A. Brockmann

732. Linking genetic merit to phenotypes: a validation of the terminal and replacement Irish sheep breeding objectives 3016
T. Pabiou, K. McDermott, A. Bohan, L.J. Farrell and N. McHugh

733. Influence of a nutritional restriction in dairy ewe lambs on the response to a later inflammatory intramammary challenge 3020
R. Pelayo, H. Marina, A. Suarez-Vega, C. Esteban-Blanco, G. Foucras, J.J. Arranz and B. Gutierrez-Gil

734. Effects of breeding for scrapie resistance in Bavarian sheep breeding populations 3024
J. Buitkamp and J. Semmer

735. Options for the selective breeding for resistance to gastro-intestinal nematodes of Sarda breed sheep 3028
S. Salaris, S. Casu, M.G. Usai, A. Scala and A. Carta

736. Investigation of the Scrapie haplotypes in the South African Merino population 3032
J.J. Reding and B.E. Mostert

737. Toward a genetic evaluation for functional longevity in French dairy sheep breeds 3037
D. Buisson, J.M. Astruc, L. Doutre and I. Palhière

738. Maturity, a new indicator to improve longevity of Saanen dairy goats in France 3042
M. Arnal, I. Palhière and V. Clement

739. The relevance of spermatozoal RNA and quality in livestock production 3046
M.J. Hodge, S.J. Rindfleish, C.P. Stephen, S. de las Heras-Saldana and S.D. Pant

740. Exploring the vomeronasal organ transcriptome in Rasa Aragonesa rams with different sexual behaviour 3050
K. Lakhssassi, I. Ureña, B. Marín, M.P. Sarto, B. Lahoz, J.L. Alabart, J. Folch, M. Serrano and J.H. Calvo

741. Elucidating transcriptomic differences in perirenal fat between male and female suckling lambs 3054
M. Alonso-García, J.J. Arranz, H. Marina, R. Pelayo, C. Esteban Blanco, B. Gutiérrez-Gil and A. Suárez-Vega

742. Measures of diversity based on autosomal and sex chromosomes of Valachian sheep 3058
M. Mészárosová, N. Moravčíková, G. Mészáros, I. Pavlík, M. Margetín and R. Kasarda

743. The data quality score: objective assessment of data quality for Australian sheep breeders 3062
D.J. Brown, E.J. McCrabb, P.E. Bradley, I.J. Rose, R.G. Banks and S.Z.Y. Guy

744. Impact of genotypic information on genetic evaluation accuracy for mastitis and footrot in the UK Texel Sheep 3066
K. Kaseja, S. Mucha, E. Smith, J. Yates, G. Banos and J. Conington

745. Characterization of homozygous haplotype deficiency compromising fertility traits in Manech Tête Rousse dairy sheep 3070
M. Ben Braiek, C. Moreno-Romieux, C. Hozé, J-M. Astruc and S. Fabre

746. Genetic analysis of colostrum and passive immune transfer in divergent goat lines selected for functional longevity 3074
M. Wicki, C. Huau, T. Fassier and R. Rupp

747. Selection signatures in South African indigenous sheep using the HAPFLK approach 3078
I.N. Malan, S.W.P Cloete and A.H. Molotsi

748. An across-flock genetic analysis of components of reproduction in four South African Merino resource flocks 3081
C.L. Nel, A.A. Swan, K. Dzama, W.J. Olivier, M.A. Snyman and S.W.P. Cloete

Part 41 – Companion animals **3085**

749. Diversity status and future breeding strategy for Tatra hound in Slovakia 3086
R. Kasarda and N. Moravčíková

750. Genome-wide association study for ectopic ureter in three dog breeds 3090
A. Stella, V. Cassani, G. Majolino, P. Cozzi, S. Biffani, F. Biscarini and G. Minozzi

751. Linkage disequilibrium variability in the Czechoslovakian wolfdog genome as a result of long-term artificial selection 3093
N. Moravčíková, R. Kasarda, L. Vostrý, H. Vostrá-Vydrová, J. Vašek and D. Čílová

752. Applications of single-molecule real-time isoform sequencing (Iso-Seq) for unravelling complexity of dog transcriptomes 3097
W. Zhang, L. Eory, M. Salavati, E. Clark, J. Smith, A.L. Archibald and J.J. Schoenebeck

753. Genome-wide signals of positive selection identified in Livestock Guardian Dogs 3101
M. Shihabi, V. Glavaš, V. Cubric-Curik, M. Zorc, P. Dovč and I. Curik

754. Implementation of single-step GBLUP for complex behaviour traits in Labrador Retrievers used as service dogs 3105
M.M. Riser, E. Leighton, J. Russenberger, C. Moser and B. Fragomeni

755. Longevity and major causes of death are heritable in Irish Wolfhound 3109
C.J.A. van Gemert, J.J. Windig and P. Berg

756. The Vaika project: a multidisciplinary study of canine aging and the genetic mechanisms influencing the aging process 3113
H.J. Huson, E.L. Andrianova, D.I. Fleyshman, L. Brodsky, J.P. Loftus, J.J. Wakshlag and A.V. Gudkov

757. Genomic diversity in selected Austrian dog populations 3117
G. Mészáros, M. Mészárosová and A. Geretschläger

758. Breed differences in the use of tools for dog breeding 3121
M. Janes, T.W. Lewis J.J. Ilska, G. Gorjanc and J.A. Woolliams

759. Inbreeding and litter size in Dutch pedigreed dogs 3125
J.J. Windig, M.L. Margarita and H.P. Doekes

Part 42 – Equine **3129**

760. Genetic analysis of three horse breeds of Greece 3130
A. Kominakis, N. Kritikos, A. Simpson, T. Sdourlias, E. Iliopoulou, V. Karagianni, D. Pappas and A.L. Hager-Theodorides

761. Genetics of new conformation traits in Italian Heavy Draught Horse foals 3135
R. Mantovani, E. Mancin, B. Tuliozi, G. Gomez Proto, G. Pigozzi and C. Sartori

762. Effects of selection on breed composition of the Spanish Sport Horse population 3139
E. Bartolomé, S.T. Rodríguez-Ramilo, J. Fernández and M. Valera

763. Early selection criteria for improved longevity of jumping horses 3143
A. Ricard, S. Deretz and B. Dumont Saint Priest

764. Recent advances in genomics of equine health, welfare and performance 3147
S.C.M. Mikko

765. Microsatellite allele imputation from SNP genotypes for parentage verification in sport horses 3151
H. Crichan, C. Engler, E. Goulas, S. Dhorne-Pollet, M. Addes and A. Ricard

766. Equine parentage control: bridging the gap from a microsatellite to a SNP based approach 3155
W. Nolte, H. Alkhoder, M. Wobbe, K.F.S. Stock, E. Kalm, S. Vosgerau, N. Krattenmacher, G. Thaller, J. Tetens and C. Kühn

767. Haplotype blocks and heterozygosity rich regions on ECA2 in Swedish Warmblood horses 3159
M. Ablondi, S. Eriksson and S. Mikko

768. Single step genomic evaluation for horses based on a multi-breed reference population 3163
M. Wobbe, H. Alkhoder, Z. Liu, S. Vosgerau, N. Krattenmacher, M. von Depka-Prondzinski, E. Kalm, R. Reents, W. Nolte, C. Kühn, J. Tetens, G. Thaller and K.F. Stock

769. Methodology to integrate pedigrees of two Belgian Warmblood studbooks and its importance for genetic evaluation 3167
L. Chapard, N. Buys and S. Janssens

Part 43 – Porcine **3171**

770. Estimation of genetic parameters and trends of reproduction traits of Korean Landrace and Yorkshire pigs 3172
S.S. Lee, M. Alam, H.K. Chang and T.J. Choi

771. Genetic parameter estimates of heat tolerance in number of born alive in Landrace, Large White, and Duroc pigs 3176
T. Okamura, Y. Fukuzawa, M. Nishio, A. Arakawa, K. Ishii, H. Takahashi, S. Ogawa and M. Satoh

772. Bayesian analysis of paternal and maternal gametic effects in a reciprocal cross between two Iberian varieties 3180
H. Srihi, N. Ibáñez-Escriche, J. Casellas, J.L. Noguera, P. Hernández, M. Martín de Hijas, M. Vázquez-Gómez, S. Negro, J.P. Rosas and L. Varona

773. Heritabilities and genetic correlations of litter traits at farrowing and semen production traits in Duroc pigs 3184
S. Ogawa, M. Kimata, M. Tomiyama and M. Satoh

774. Genomic analysis of quality and morphological sperm parameters in Piétrain breeding boars 3188
C. Großе-Brinkhaus, Z.M. Kröhl, M. Friedrichs and E. Tholen

775. Differential expression in the porcine thymus upon infection by PRRSV, influenza B or their coinfection 3192
K.M. Sarlo Davila, Y. Sang, W. Ma and L.C. Miller

776. Genetic associations between survival at different parities in commercial crossbred sows 3196
B.G. Poulsen, N.G. Leite and D. Lourenco

777. Piglet Survival is lowly heritable in DanBred Landrace pigs 3200
X. Guo, A.C. Sørensen, M. Henryon and T. Ostersen

778. Feasibility of reducing mortality of pigs from birth to slaughter by genetic selection 3204
G. Su, T. Liu, O.F. Christensen, M.S. Lund and B. Nielsen

779. Genotyping dead animals improves post-weaning survival of pigs in breeding programs 3208
M. Sharif-Islam, J.H.J van der Werf, M. Henryon, T.T. Chu, B.J. Wood and S. Hermesch

780. GBP5 PRRSV resistance gene had no effect on pigs' infectivity or susceptibility in a trial simulating natural infections 3212
M.E. Chase-Topping, G. Plastow, J. Dekkers, Y. Fang, V. Gerdts, J. van Kessel, J. Harding, T. Opriessnig and A. Doeschl-Wilson

781. Unravelling the genetic basis governing the porcine metabolism 3217
S. Bovo, G. Schiavo, F. Fanelli, A. Ribani, F. Bertolini, V.J. Utzeri, M. Gallo, G. Galimberti, S. Dall'Olio, P.L. Martelli, R. Casadio, U. Pagotto and L. Fontanesi

782. The value of innate sow appetite as a model trait for maternal breeding objectives 3221
L. Vargovic, K.L. Bunter and S. Hermesch

783. Genomic regions associated with backfat thickness show pleiotropic effect on osteochondrosis in pig 3225
M. van Son, M.F.L. Derks, M.S. Lopes, C.A. Sevillano, B. Harlizius and E. Grindflek

784. Non-genetic correlations between production traits and litter traits in pigs are important to account for 3229
T. Ostersen, M. Henryon, X. Guo, O.F. Christensen, G. Su and A.C. Sørensen

785. Breeding for component traits of litter size at day 5 increases piglet survival while maintaining litter size at day 5 3233
M. Henryon, T. Ostersen, X. Guo, G. Su and A.C. Sørensen

786. Phenotyping strategies for an efficient and holistic approach to reduced boar taint through genomic selection 3237
A. Markey, C. Burgeon and N. Gengler

787. Genome-wide association study for oxidative stress indicators in replacement gilts 3241
S.A. Breuer, J. Chinchilla-Vargas, B.J. Kerr and K.J. Stalder

788. Genetic determinism of boar taint in the French Landrace pig breed 3245
C. Larzul, M.-J. Mercat, C. Hassenfratz, C. Carillier-Jacquin, R. Comte, B. Blanchet, I. Louveau, S. Boulot and A. Prunier

789. A comprehensive overview of mitochondrial DNA insertions in the nuclear genome of the pig 3249
M. Bolner, M. Ballan, S. Bovo, G. Schiavo and L. Fontanesi

790. To Box-Cox or not to Box-Cox: phenotypic and genomic evaluation of litter size variability in Landrace pigs 3253
G. Cieleń, E.F. Knol, M.S. Lopes E. Sell-Kubiak

791. Genetic parameters of feeding behaviour and feed efficiency traits in Pietrain 3258
P. Núñez, S. Gol, J. Reixach and N. Ibáñez-Escriche

792. Genetic analysis of hoof colour in Duroc pigs 3262
J. Vila, O.O.M. Iheshiulor, M.C.A.M. Bink and A.E. Huisman

793. Filling information gaps in swine crossbreeding schemes by imputing non-genotyped F_1 animals to improve genetic gain 3266
G.M. See, J.S. Fix, C.R. Schwab and M.L. Spangler

794. Distinctive gene-expression profiles characterise tail-biting precursors in pigs under dietary protein restriction 3270
L. Couteller, L. Roch and C. Kasper

795. Analyses of heat stress effects on the farrowing rate of Iberian sows 3274
M. Muñoz, G. Matos, G. Gómez and J.M. García-Casco

796. Whole-genome screening for resilience against PRRSV outbreaks in breeding sows 3278
M. Laplana, R. Ros-Freixedes, J. Estany, L.J. Fraile and R.N. Pena

797. Genetic relationships among immune response traits of young healthy pigs evaluated by immunoassays 3282
V. Bhatia, J. Schmied, J. Cheng, X. Bai, B. Mallard, F. Fortin, J.C.S. Harding, M.K. Dyck, G.S. Plastow, C.J. Field, C. Rogel-Gaillard, F. Blanc, PigGen Canada and J.C.M. Dekkers

798. Sequencing-based genome-wide association study for resilience indicators in growing pigs 3286
H. Laghouaouta, R.N. Pena, R. Ros-Freixedes, L. Fraile and M. Laplana

799. First parity indicator traits for longevity in crossbred sows 3290
M.W. Iversen and E.G. Enger

800. protiPig: Genomic analyses of traits related to nitrogen utilisation efficiency in pigs 3294
M. Schmid, R. Weishaar, D. Berghaus, E. Haese, N. Sarpong, A. Kurz, J. Seifert, A. Camarinha-Silva, M. Rodehutscord and J. Bennewitz

801. Genomic prediction of ham weight loss during dry-curing in heavy pigs 3298
V. Bonfatti, S. Faggion and P. Carnier

802. Meta-analyses for boar taint compounds in two purebred maternal lines and Piétrain-sired crosses 3302
I. Brinke, C. Große-Brinkhaus, K. Roth, M.J. Pröll-Cornelissen, I. Schiefler and E. Tholen

Part 44 – Free communications **3306**

803. A data-architecture to monitor and collect cow-individual methane emissions real-time from commercial dairy farms 3307
I. Hulsegge, D. Schokker, G.G.J. Klandermans and C. Kamphuis

804. An ontology approach for the collection and management of genetic traits and milk protein data 3311
C. Franzoni Migliorati, J.B.C.H.M. van Kaam, R. Finocchiaro, C.A. Lombardi, M. Marusi and M. Cassandro

805. Genotyping errors in SNP markers: types, consequences, and a novel detection method 3315
W.M. Rauw, M.G. Gil, O. Torres and L. Gomez-Raya

806. LncRNAs variability in skeletal muscle of Polish Landrace boars 3319
B. Hofman, J. Szyda, M. Frąszczak and M. Mielczarek

807. Inclusion of average rainfall in genetic evaluation of SA Holsteins to mitigate genotype by environment interaction 3323
M. van Niekerk, F.W.C. Neser, J.B. van Wyk and V. Ducrocq

808. Development of an automatic method to estimate IMF for fattening pigs in breeding program 3327
L. Flatrès-Grall, C. Ribas, K. Kashefifard, C. Herbert and G. Lenoir

809. Assessing genomic prediction of economic trait in alpacas: a simulation study 3331
G.C. Mamani, B.F. Santana and D. Jarquin

810. Somatic cell count and its relationship with udder type traits in Italian Mediterranean Buffaloes (*Bubalus bubalis*) 3335
M.M. Gómez, R. Cimmino, Y. Gombia, G. Zullo and S. Biffani

811. Estimating vertical path coefficients of transmission of non-genetic inherited effects from a similarity matrix 3339
I. David, C. Larzul and A. Ricard

812. Building on Falconer's work on environmental sensitivity and scale 3343
D.L. Waters, J.H.J. van der Werf and S.A. Clark

813. Impact of a non-Gaussian dependence structure on REML estimation of the bivariate genetic animal model 3347
T. Rohmer, A. Ricard and I. David

814. Impact of genotyping strategy on the accuracy of genomic prediction for survival in pigs 3351
T. Liu, B. Nielsen, O.F. Christensen, M.S. Lund and G. Su

815. Effect of a *TLR1A* polymorphism on natural antibodies in poultry 3355
H. Bovenhuis, T.V.L. Berghof M.H.P.W. Visker, J.A.J. Arts, J. Visscher, J.J. van der Poel and H.K. Parmentier

816. Geometry meets genomics 3360
L. Gomez-Raya, W.M. Rauw and A. Hess

Author index 3365

Part 1

Plenaries

1. Measure, predict and act; a paradigm of genetics applied to livestock production now and into the future

M.P.L. Calus[1*], D.P. Berry[2] and J.E. Pryce[3,4]

[1]Wageningen University & Research, P.O. Box 338, 6700 AH Wageningen, the Netherlands; [2]Teagasc, Moorepark Research Center, Fermoy, Ireland; [3]Agriculture Victoria Research, Agribio, 5 Ring Road, Bundoora, VIC 3083, Australia; [4]School of Applied Systems Biology, La Trobe University, Bundoora, VIC 3083, Australia; mario.calus@wur.nl

Abstract

Rapid developments in phenotyping and genomic technologies are contributing to lakes of data on animals. Concurrent developments in prediction algorithms are leading to a proliferation of (genomic and phenotypic) predictions on individual animals. Using genetics applied to livestock to reduce the impact of livestock on climate change, while addressing livestock welfare concerns and meeting more sophisticated and diverse nutritional demands of the consumer, are likely to continue to be emerging trends in this and future WCGALP. Being able to measure, predict and act will continue to be the fundamental principles to achieving these objectives. Failure to do so will jeopardise the success of animal breeders into the future, especially if growing numbers of consumers favour entirely plant-based sources of nutrition.

Introduction

Livestock breeding programs can be broadly summarised by the principles of *Measure, Predict* and *Act*. This paper takes a journey through these three components highlighting the advances made, but also the gaps in knowledge. Being able to collapse the deluge of predictions constantly streaming towards producers into actionable information will be one of the greatest challenges to achieving meaningful change. Animal breeders are well versed in not only collapsing vast quantities of data on a whole gamut of traits into a single index value per animal, but also providing the necessary communications/extension and validation of such tools to build trust among stakeholders.

Measure

The breadth of livestock breeding objectives has generally been limited to those traits with data available, either directly or indirectly; health, fertility and growth traits comprised between 12% (1974) and 30% (2006) of the WCGALP papers to-date (Figure 1a). Focussing on those traits may disadvantage traits that are less well recorded (e.g. feed intake, environment), especially in breeds/strains that are less numerous and subsequently have smaller reference populations, which may be exacerbated in a paradigm of market-driven investment in genetic improvement.

The biggest challenge for the current generation of animal breeders is most likely to rapidly address the contribution of livestock to the environment; with livestock being the most significant anthropogenic methane source. Ruminants emit methane as a by-product of rumen fermentation, while poultry are the second largest producer of ammonia, phosphorus, nitrogen, carbon dioxide, and methane, meaning that the challenge spans the main-stream species considered at WCGALP conferences (e.g. de Haas *et al.* 2018, Sell-Kubiak *et al.* 2018). Focusing on the efficient use of feed (especially from sources that cannot be used for human consumption), will likely continue to feature in all animal breeding programs. Feed efficiency was, in fact, a hot topic across species at the 2014 and 2018 WCGALP and continues to be in 2022, with over 5% of papers presented being on the topic (Figure 1b). This rise is likely due to the increased interest in these traits in cattle breeding, while measuring traits associated with feed efficiency has been common in pigs and poultry for a long time. Figure 1b also shows how papers on methane emissions have been

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_816, © M.P.L. Calus *et al.* 2022

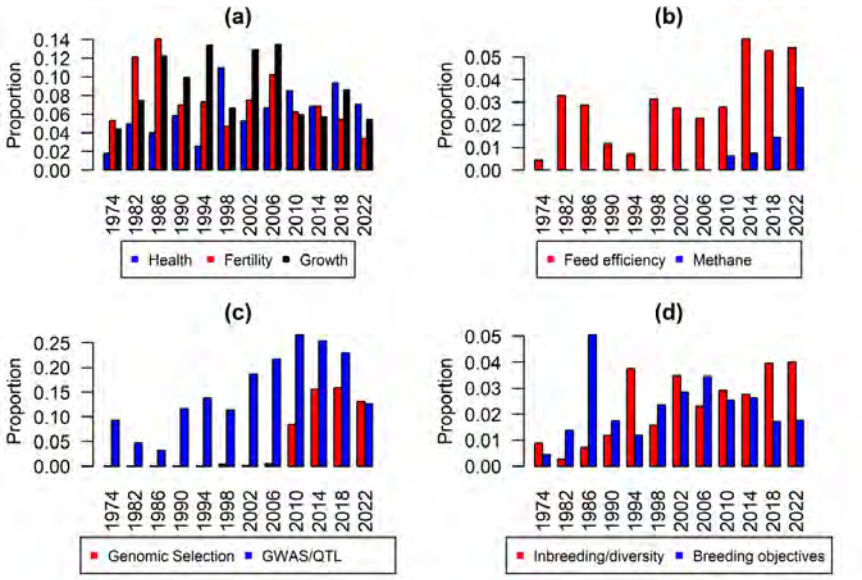


Figure 1. Representation of traits (a & b) and topics (c & d) for each WCGALP year.

increasing exponentially from the 2014 to 2022 WCGALP. The WCGALP2022 sees an increased focus on selecting for a reduction in all pollutants, with ingenious phenotyping strategies and the use of indirect predictors being used as ways to increase the accuracy of genomic prediction anticipated.

Consumers are increasingly demanding high-quality nutritious products. Genetics can be used to select for the best nutritional profile using data from technologies to measure these characteristics. We may see food being tailored to match a person's own requirements. We envisage that animal welfare and sustainability will continue to be important themes. Notable examples are previously observed deterioration in fertility (in dairy cows) and rise of leg weakness (in broiler chickens). Both are, of course, examples of unintended consequences of selection and a warning to animal breeders. To counteract these trends, health and fitness traits have been included in breeding objectives for 20+ years in many cases.

Across many traits discussed, and even ones not thought of yet, it seems very likely that sensors and automated phenotyping will become the primary strategy to collect data for animal breeding purposes and there will be many more high-tech approaches presented at WCGALP2022. For example, CT scanning has been used for non-invasive measurements of (small) carcases for quite a long time, but such precision phenotyping technologies are coming into their own through its combination with genomic technologies (Byrne and Amer 2018). The challenge from this will be processing high volumes of data, data quality control steps and combining it in an informative way with other omics data.

Predict

Genomic selection (GS) has undeniably accelerated the wide-spread use of individual genome-wide information in animal breeding programs, facilitated by the availability of affordable 50k genotyping arrays, all within just one decade. This is supported by the rapidly increasing number of GS papers since WCGALP2006, for the first time being more numerous than GWAS papers at WCGALP in 2022, the latter declining after peaking at WCGALP2010 (Figure 1c). Methodological and technological developments have enabled computational requirements to stay within reasonable limits, despite reference populations

increasing in size exponentially. We expect several reported applications using >50M genotyped animals at WCGALP2026.

Sequencing strategies in the past decade took either the 1000-bulls' genome approach by genotyping key ancestors in the population at high coverage and imputing to other genotyped animals, or a low-pass sequencing approach, genotyping very many animals at (ultra)low coverage followed by imputation taking advantage of high coverage at the haplotype level (Hickey *et al.* 2014). Large scale individual level actual sequence data, including long reads, will be available soon. Unlocking its full potential include routine detection of chromosome anomalies, optimizing selection and mating considering deleterious alleles, and frequent GWAS for all breeding goal traits. We envisage that GWAS results at base pair resolution will become standard output of genomic predictions, which will develop from being a pure *prediction machine* to also be a *discovery machine*. Combining this phenotype-driven discovery with biological information based on increasingly more accurate functional annotation maps, will enable continuous refinement of SNP sets used in genomic prediction (Goddard *et al.* 2018). Finally, other omics data including microbiome, gene expression, and epigenetic profiles, will help by enabling more accurate phenotyping, refining functional annotation maps, and more accurately dissecting phenotypic variances into its underlying components.

Thirty years ago, the idea of 'velo genetics' or germ-line selection was described, with the promise to reduce generations intervals drastically (Georges and Massey 1991). GS indeed has enabled embryo selection in practice. The application of embryo selection based on GS, however, mostly enables to further increase selection intensity, but not to further shorten generation intervals. Genome editing has been a futuristic topic for a long time, but is now part of the available toolbox thanks to CRISPR-CAS. The eventual contribution of gene editing will rely heavily on the ability to identify sufficient numbers of loci that are worthwhile targeting (Simianer *et al.* 2018) coupled with societal acceptance of the technique. With the foreseen *discovery machine* as a 'by-product' of GS, and the increasingly more accurate annotation maps, eventually a list of targets worthwhile pursuing is expected to be available.

Act

The breeding goals of the future will need to explicitly consider three consumer-interest areas of environmental responsibility, animal welfare and nutritive characteristics and product quality (Byrne and Amer 2018). While productivity and feeding the world are still of great importance, ignoring consumer-interests is incredibly risky. Approaches to appropriately weight these features in breeding goals will be important, prompting greater research into the optimal construction of breeding objectives and more so than has been presented to-date (Figure 1d).

Predictions in themselves are merely outputs and must therefore be acted upon to create impact. Whereas artificial intelligence (AI) is being heralded as a disruptive technology to the prediction processes, the field of eXplainable Artificial Intelligence (XAI) could have a role helping comprehend and trust the predictions. Moreover, prescription (i.e. how to deal with predicted outcomes) and, more importantly, the ability to diagnose the root of sub-optimal performance will be crucial in the future as decision support tools. Breeders will have to engage more with the developers of decision support systems to: (1) ensure genetic merit is included as a feature in prediction algorithms; (2) ensure that such support systems diagnoses the root of the sub-optimal performance – such support systems must also be able to detect risk into the future given the current trajectory; and (3) note any remedial measures undertaken for consideration either in data editing or statistical modelling of the resulting genetic evaluation.

While strategies to control inbreeding were popular in the WCGALPs of the 1990s (e.g. optimal contribution selection (Meuwissen 1997); Figure 1d), there is likely to be a resurgence in the future, especially in ways that are easy to implement and utilise genomic information.

Outlook

Breeding is a well proven technology to deliver genetic change; of course, this genetic change has not always been favourable. Whereas the breeding programs of some species has consolidated considerably (e.g. pigs, poultry, aqua), the breeding programs of other species, most notably ruminants, are still largely underpinned by national genetic evaluations. Of strategic importance is data ownership, which is becoming a contentious issue and as yet, has not been unequivocally addressed in ruminant breeding programs. The time is approaching for mega-corporations to realise the power of genetic gain in delivering food security (i.e. power) while also remaining stewards of the environment. Many of these corporations (e.g. Amazon, Google, Facebook) already have the analytical expertise and potential access to data. More importantly they are now part of everybody's (i.e. consumers) home and have strong influential powers through algorithms like recommender systems. While this will jeopardise the *modus operandi* of current ruminant breeding programs, consolidation may lead to new opportunities and exciting challenges. Nevertheless, issues on data ownership and propriety evaluations could impact what is published at future WCGALP.

The biggest challenge faced by livestock production in the near future, is retaining societal acceptance, at least partly by contributing to healthy tailor-made diets and decreasing its environmental footprint. The world is awash with commentary of silver bullets to grand challenges; breeding programs have proven time and time again to deliver on grand challenges but will need to be more vocal. Now is the time to capitalise on the full potential of highly technological and data intensive breeding programs to deliver on carbon neutrality, in close dialogue with its stakeholders and in particular the consumers.

References

Byrne T., and Amer P.R. (2018) Proc. of the 11th WCGALP, Auckland, New Zealand.

de Haas Y., Wall E., Garnsworthy P., Kuhla B., Negussie E., *et al.* (2018) Proc. of the 11th WCGALP, Auckland, New Zealand.

Georges M., and Massey J.M. (1991) Theriogenology 35(1):151-159. https://doi.org/10.1016/0093-691X(91)90154-6

Goddard M.E., MacLeod I.M., Kemper K.E., Xiang R., Van den Berg I., *et al.* (2018) Proc. of the 11th WCGALP, Auckland, New Zealand.

Hickey J.M., Gorjanc G., Cleveland M.A., Kranis A., Jenko J., *et al.* (2014) Proc. of the 10th WCGALP, Vancouver, Canada.

Meuwissen T. (1997) J Anim Sci 75(4):934-940. https://doi.org/10.2527/1997.754934x

Sell-Kubiak E., Wimmers K., Reyer H., and Szwaczkowski T. (2018) Proc. of the 11th WCGALP, Auckland, New Zealand.

Simianer H., Pook T., and Schlather M. (2018) Proc. of the 11th WCGALP, Auckland, New Zealand.

2. Genetics of quantitative traits in natural populations: lessons from long term studies of wild animals

J.M. Pemberton

Institute of Evolutionary Biology, University of Edinburgh, EH9 2FL, United Kingdom; j.pemberton@ed.ac.uk

Abstract

From around the 1950s, a number of long term studies of wild animal populations were initiated by researchers interested in the behaviour and ecology of natural populations. These studies focus almost exclusively on vertebrates, and in many cases involve capturing individuals for marking and hence opportunities for tissue sampling. As genetic and genomic techniques have developed, enabling the recovery of pedigrees and now genomic estimates of relationship and inbreeding, these projects have pivoted to more evolutionary topics, especially the genetics of quantitative traits, where statistical techniques developed by livestock breeders have been very influential. I will illustrate some of the advances in the field with examples from long term studies of red deer on the Isle of Rum, and Soay sheep on St Kilda, both in Scotland.

Introduction

Evolutionary biologists and ecologists have long been interested in the relative roles of genetic and environmental factors in determining trait variation and their influence on population and evolutionary dynamics, including the maintenance of genetic variation. Long term studies of individuals commonly involve the tracking of individuals from birth through all breeding attempts, to death, yielding data on individual breeding success, survival and often several fitness-related morphological or physiological traits, and offer a window on these issues. Until the 1990s, the genetics of quantitative traits in *in situ* natural populations was a modest research field, focused mainly on passerine birds, where pairs of individuals feeding chicks at a nest (the 'social parents') were assumed to be the genetic parents and heritabilities were estimated using parent-offspring regression (e.g. Boag and Grant (1978); Van Noordwijk *et al.* (1988)). A key issue was actually determining genetic relationships between individuals, especially paternity: molecular techniques soon showed that extra-pair paternity was common in birds, and behavioural data were likely to be even poorer in predicting paternity in, for example, polygynous species. With the advent of DNA fingerprinting, microsatellites and now SNPs and appropriate inference methods (e.g. Marshall *et al.* (1998); Wang (2004); Huisman (2017)), the recovery of pedigrees is now robust, if still labour-intensive.

Once pedigrees became available, a second breakthrough came from animal breeding: the animal model (Meyer 1991) enabled efficient use of all relationship data to estimate heritabilities and genetic correlations in natural populations (Kruuk 2004). Another major benefit of the animal model approach is that it allows simultaneous estimation of other variance components such as year, spatial and maternal effects (Kruuk and Hadfield 2007). These sources of variation, initially regarded to some degree as nuisances, are important subjects of study in their own right.

Here, I will address four of the several areas in which the field has made progress in recent years – for wider reading see (Charmantier *et al.* 2014). These are: (1) quantifying inbreeding depression; (2) quantifying additive genetic variation; (3) genetic architecture; and (4) evolutionary dynamics of quantitative traits. I will illustrate these topics with examples drawn from two long term studies, of red deer on the Isle of Rum, and of wild Soay sheep on St Kilda. In these projects as many study area individuals as possible are captured at birth for permanent marking and sampling and then followed for the rest of their lives. In the case of the deer, we do not usually handle animals again, so besides reproduction and survival, birth, *post mortem*

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_2, © J.M. Pemberton 2022

skeleton and cast antler measures are the main phenotypic traits available. In the case of the sheep, we attempt to recapture and phenotype as many individuals as possible each summer and so we have a larger array of study traits. In both study populations, all sampled individuals are genotyped on 50K SNP arrays yielding around 38K polymorphic sites in each case.

1. Quantifying inbreeding depression. Close pedigree inbreeding is rare in most wild vertebrate populations, but in both study populations there is extensive low-level inbreeding. Past analyses using (incomplete) pedigrees and microsatellite heterozygosity indicated that there was inbreeding depression in both populations, but its magnitude remained unclear until genomic approaches to estimating individual inbreeding became available. Using a genomic inbreeding coefficient (F_{grm}; Yang *et al.* (2011)) in the deer, we found that inbreeding depresses birth weight, juvenile survival (independently of birth weight), and annual breeding success in both sexes and offspring survival in females (Huisman *et al.* 2016). As a consequence of these effects, inbreeding depression across the lifespan is strong: a female with F_{grm}=0.125 has an average 75% reduction in lifetime breeding success (LBS) compared with an average female with F_{grm}=0. The equivalent figure for males is a 95% reduction in LBS. The same approach found inbreeding depression in several traits in Soay sheep (Bérénos *et al.* 2016). More recently, following imputation of the Soay sheep data to >400K SNPs, we have used a 'runs of homozygosity' approach to quantify inbreeding depression in survival in the Soays and demonstrate that while negative effects are scattered throughout the genome, there are five genome regions with genome-wide significant effects on survival (Stoffel *et al.* 2021a). In further analysis we found that most of the inbreeding depression lies within longer runs, consistent with the idea that the haplotypes forming short runs have been purged of deleterious mutations (Stoffel *et al.* 2021b).

2. Quantifying additive genetic variation for traits. Estimating heritabilities and genetic correlations with precision is a necessary requirement if we are to predict how populations will respond to natural selection. In recent decades many estimates of trait heritabilities have been made in natural populations (Postma 2014). Comparisons between traits enabled us to test the prediction that traits under stronger directional selection would show lower heritabilities (due to attrition of genetic variance by selection). In the deer we did indeed find this pattern: heritabilities are inversely correlated with a trait's association with LBS, our measure of fitness (Kruuk *et al.* 2000). However, this was not for the expected reason: other components of variance (presumably, mainly environmental) increase as one considers ever more integrative traits like components of fitness, and hence the decline in heritability was driven by increasing phenotypic variance, with additive genetic variance showing no consistent pattern.

The genomic era offers the possibility of pedigree-free estimation of heritabilities and genetic correlations using genomic relatedness. In human cohorts selected to be unrelated for GWAS, fitting information from all SNPs underestimates trait heritabilities, for example for height (Yang *et al.* 2010), so it was of interest to know how it performed in natural populations including a wide range of pairwise relatedness and consequently higher linkage disequilibrium. We conducted one of the first comparisons of heritabilities using pedigree versus genomic relatedness in Soay sheep and found that estimates coincided across most traits and age classes (Bérénos *et al.* 2014). The approach offers the potential for wild quantitative genetics to extend into taxa such as invertebrates that cannot be followed so easily in the wild, although note that parentage information will always be required to quantify maternal effects.

3. Genetic architecture. As with livestock and humans, there has been great interest in whether it is possible to locate genes of major effect underlying quantitative traits in the wild, since this would allow detailed studies of selection and response at specific loci. We have carried out a number of GWAS analyses (controlling for relationship) and regional heritability studies across the deer and sheep data, with mixed

results. In Soay sheep, most morphological traits seem to be under highly polygenic control, though two regions contributing to leg length, on chromosomes 16 and 19, are robust and seem likely to have come from a documented introgression event (Bérénos *et al.* 2015). In deer, no regions were found underpinning either neonatal traits (J. Gauzere pers. comm.) or several measures of antlers (Peters *et al.* 2022). On the brighter side, a major locus underpins both the horn polymorphism and horn size in the Soay sheep, and selection analysis using inferred genotypes showed that the polymorphism is maintained by heterozygote advantage (Johnston *et al.* 2013). Using SNP and pedigree data we have also derived individual autosomal recombination rate (ARR) as a trait and shown that it maps to genomic regions containing the genes REC8 and RNF212B, adding to the evidence that these genes control variation in recombination rate in mammals (Johnston *et al.* 2016; Johnston *et al.* 2018).

4. Evolutionary dynamics of quantitative traits. The ultimate aim of most studies of quantitative traits in natural populations is to understand the evolution of traits and the maintenance of genetic variation. Some traits change in the predicted direction: birth date is selected to be earlier in deer and the pedigree breeding values have got earlier over recent years (Bonnet *et al.* 2019). However, and in sharp contrast to animal breeding, a common observation in natural populations is that there is strong directional selection on a heritable trait, but no evidence of phenotypic change in response to selection. This is known as the 'paradox of stasis' and many possible explanations for it have been investigated, including in the deer and sheep. One possibility is that a genetic trend may be masked by an environmental trend affecting a trait by plasticity and in the opposite direction. Such 'cryptic evolution' was suggested to be the case for Soay sheep body size (Wilson *et al.* 2007) which is under positive selection but in the field the mean body sizes are getting smaller. A recent analysis of Soay sheep body weight using genomic prediction finds that the breeding values are indeed increasing (Hunter *et al.* 2022). Another explanation lies in environmental variation causing a phenotypic but not a genetic correlation between a trait and fitness. This appears to be the case for antler size and LBS in deer (Kruuk *et al.* 2002). A third possibility is the potential for constraints arising from antagonistic selection or negative genetic correlations, for example between the sexes or between traits, but in deer at least these do not seem to be very constraining (Morrissey *et al.* 2012; Walling *et al.* 2014). Finally, constraints could also arise via maternal effects, but a recent analysis of deer birth weight, which has a large maternal genetic component, found this too to be unlikely (Gauzere *et al.* submitted).

Conclusions

Advances in genomic and statistical analysis techniques have led to great progress in the study of quantitative traits in the wild (Charmantier *et al.* 2014). A major difference from livestock production is that appreciable plastic effects due to environmental heterogeneity in time and space are to be expected at every turn and must be handled appropriately. Due to relatively small sample sizes and the scarcity of large effect loci, GWAS approaches in natural populations seem likely to yield many flat 'Uppsala' plots for some time to come. Genomic prediction has yet to be widely exploited but holds considerable promise (Ashraf *et al.* 2021).

References

Ashraf B., Hunter D.C., Bérénos C., Ellis P.A., Johnston S.E. *et al.* (2021) Mol. Ecol. n/a(n/a). https://doi.org/10.1111/mec.16262

Bérénos C., Ellis P.A., Pilkington J.G., Lee S.H., Gratten J. *et al.* (2015) Mol. Ecol. 24(8)**:** 1810-1830. https://doi.org/10.1111/mec.13146

Bérénos C., Ellis P.A., Pilkington J.G. and Pemberton J.M. (2014) Mol. Ecol. 23(14)**:** 3434-3451. https://doi.org/10.1111/mec.12827

Bérénos C., Ellis P.A., Pilkington J.G. and Pemberton J.M. (2016) Mol. Ecol. 25(13)**:** 3152-3168. https://doi.org/10.1111/mec.13681

Boag P.T., and Grant P.R. (1978) Nature 274(5673)**:** 793-794.

Bonnet T., Morrissey M.B., Morris A., Morris S., Clutton-Brock T.H. *et al.* (2019) PLoS Biol. 17(11)**:** 23. https://doi.org/10.1371/journal.pbio.3000493

Charmantier A., Garant D. and Kruuk L.E.B. (Editors) (2014) *Quantitative Genetics in the Wild*. Oxford University Press, Oxford.

Gauzere J., Pemberton J.M., Morris A., Morris S. and Walling C.A. (submitted) Evol. Lett.

Huisman J. (2017) Molecular Ecology Resources 17)**:** 1009-1024. https://doi.org/10.1111/1755-0998.12665

Huisman J., Kruuk L.E.B., Ellis P.A., Clutton-Brock T. and Pemberton J.M. (2016) Proc Natl Acad Sci USA 113(13)**:** 3585-3590. https://doi.org/10.1073/pnas.1518046113

Hunter D., Ashraf B., Bérénos C., Ellis P., Johnston S. *et al.* (2022) Mol. Ecol.**:** 2021.2001.2006.425564. https://doi.org/10.1101/2021.01.06.425564

Johnston S.E., Berenos C., Slate J. and Pemberton J.M. (2016) Genetics 203(1)**:** 583-+. https://doi.org/10.1534/genetics.115.185553

Johnston S.E., Gratten J., Berenos C., Pilkington J.G., Clutton-Brock T.H. *et al.* (2013) Nature 502(7469)**:** 93-+. https://doi.org/10.1038/nature12489

Johnston S.E., Huisman J. and Pemberton J.M. (2018) G3-Genes Genomes Genetics 8(7)**:** 2265-2276. https://doi.org/10.1534/g3.118.200063

Kruuk L.E.B. (2004) Philosophical Transactions of he Royal Society of London Series B-Biological Sciences 359(1446)**:** 873-890.

Kruuk L.E.B., Clutton-Brock T.H., Slate J., Pemberton J.M., Brotherstone S. *et al.* (2000) Proceedings of the National Academy of Sciences of the USA 97(698-703.

Kruuk L.E.B., and Hadfield J.D. (2007) J. Evol. Biol. 20(1890-1903.

Kruuk L.E.B., Slate J., Pemberton J.M., Brotherstone S., Guinness F. *et al.* (2002) Evolution 56(8)**:** 1683-1695.

Marshall T.C., Slate J., Kruuk L.E.B. and Pemberton J.M. (1998) Mol. Ecol. 7(5)**:** 639-655.

Meyer K. (1991) Genetics Selection and Evolution 23(67-83.

Morrissey M.B., Walling C.A., Wilson A.J., Pemberton J.M., Clutton-Brock T.H. *et al.* (2012) Am. Nat. 179(4)**:** E97-E114. https://doi.org/10.1086/664686

Peters L., Huisman J., Kruuk L.E.B., Pemberton J.M. and Johnston S.E. (2022) Mol. Ecol. 31(4)**:** 1281-1298. https://doi.org/10.1111/mec.16314

Postma E. (2014) Four decades of estimating heritabilities in wild populations: improved methods, more data, better estimates?, pp. 16-33 in *Quantitative Genetics in the Wild*, edited by Charmantier A., Garant D. and Kruuk L.E.B. Oxford University Press, Oxford.

Stoffel M.A., Johnston S.E., Pilkington J.G. and Pemberton J.M. (2021a) Nature Communications 12(1)**:** 10. https://doi.org/10.1038/s41467-021-23222-9

Stoffel M.A., Johnston S.E., Pilkington J.G. and Pemberton J.M. (2021b) Evol. Lett. 5(3)**:** 187-195. https://doi.org/10.1002/evl3.229

Van Noordwijk A.J., Van Balen J.H. and Scharloo W. (1988) Genet. Res. 51(149-162.

Walling C.A., Morrissey M.B., Foerster K., Clutton-Brock T.H., Pemberton J.M. *et al.* (2014) Genetics 198(4)**:** 1735-+. https://doi.org/10.1534/genetics.114.164319

Wang J.L. (2004) Genetics 166(4)**:** 1963-1979.

Wilson A.J., Pemberton J.M., Pilkington J.G., Clutton-Brock T.H., Coltman D.W. *et al.* (2007) Evol. Ecol. 21(3)**:** 337-356.

Yang J., Benyamin B., Mcevoy B.P., Gordon S., Henders A.K. *et al.* (2010) Nat. Genet. 42(7)**:** 565-U131. https://doi.org/10.1038/ng.608

Yang J., Lee S.H., Goddard M.E. and Visscher P.M. (2011) Am. J. Hum. Genet. 88(1)**:** 76-82. https://doi.org/10.1016/j.ajhg.2010.11.011

3. Discoveries and lessons from a genome-wide association study of human height in >5 million individuals

L. Yengo on behalf of GIANT Consortium (the height working group)

Institute for Molecular Bioscience, The University of Queensland, 306 Carmody Road, St Lucia, Brisbane, QLD 4072; Australia; l.yengo@imb.uq.edu.au

Abstract

In this lecture, I will present findings from the largest, to date, genome-wide association study of human height conducted in >5 million individuals of diverse ancestries. This gargantuan effort has led to identify >12,000 independent associations with common SNPs with a frequency >1%. These height-associated variants are clustered within ~7,000 genomic loci enriched for genes which mutations cause abnormal skeletal growth. Across five ancestry and ethnic groups, we found that >90% of the common variants SNP-based heritability is explained by variants within these 7,000 loci, which cumulatively cover ~21% of the human genome. A genetic predictor combining these 12,000 height-associated variants yields a prediction accuracy of R^2~40%, while combining that predictor with the average height of parents reaches an unprecedented accuracy of R^2~54%. Overall, our study proposes a saturated map of genomic loci likely containing most of the genetic variation responsible for the missing heritability of human height.

Introduction

Over the past 15 years, genome-wide association studies (GWAS) have revolutionised our understanding of the genetic architecture of complex traits by allowing the discovery of tens of thousands of genetic variants associated with thousands of traits. GWAS of human height (a classic polygenic trait) had hitherto identified over 3,000 independent SNPs associated with inter-individual differences,[1] explaining about half of the SNP-based heritability.[2] Sample size is the main factor driving GWAS discoveries. However, how large experimental samples need to be before we can account for the totality of the SNP-based heritability has long remained unknown. Here, we perform a GWAS of human height in ~5.4 million individuals with diverse ancestries (Yengo 2022[3]) and show empirically that our sample size is sufficient to map autosomal genomic loci, where genetic variation accounts for 90 to 100% of the SNP-based heritability.

Materials & methods

We analysed data from 5,380,080 individuals whose ancestries were predominantly European (EUR; 76% of total sample), East-Asian (EAS; 8.8%), Admixed Hispanic ethnicity (HIS; 8.5%), African (AFR; 5.5%) and South-Asian (SAS; 1.4%). We tested association at 1.3 million SNPs catalogued in the HapMap 3 project, which includes over 1 million with a minor allele frequency (MAF) larger than 1% in each of the five ancestral groups. We first meta-analysed cohorts within each ancestral group using a fixed effect meta-analysis then strategy, then meta-analysed summary statistics from all 5 groups of cohorts using fixed effect approach. In total we performed 6 meta-analyses: 5 within-ancestral groups and 1 across-ancestral groups. For each meta-analysis, we identified sets of genome-wide significant (GWS) conditionally independent associations using the approximate conditional and joint multiple-SNP analyses,[4] as implemented in GCTA[5].

We defined GWS loci as genomic regions centred around each GWS SNP and including all SNPs within 35 kb on each side of the lead GWS SNP. Overlapping GWS loci were merged so that the number and cumulative length of GWS loci are calculated on non-overlapping GWS loci. To estimate the total variance explained by SNPs within GWS loci, we: (1) partitioned all SNPs from the HapMap 3 panels into two groups (SNPs within those 7,209 loci versus SNPs outside those 7,209 loci); (2) calculated genetic relationship matrices based on these two sets of SNPs; and (3) used these matrices in a GREML analysis to estimate

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_816, © L. Yengo 2022

heritability. The cumulative length of non-overlapping GWS loci is ~647 Mb, i.e. ~21% of the genome assuming a genome length of ~3,000 Mb.

Finally, we used joint SNP effects from our GWAS meta-analysis (re-estimated using GCTA) to calculate a genetic predictor of height from GWS SNPs only.

Results

We identified 9,863, 1,888, 918, 493 and 69 independent genome-wide significant (GWS; $P<5\times10^{-8}$) associations in the EUR, HIS, EAS, AFR and SAS groups, respectively (Table 1). We found a strong genomic co-localisation of associations identified in each population and a correlation of estimated SNP effects >0.64 across all pairs of ancestral groups. Next, we detected 12,111 conditionally independent associations in a multi-ancestry meta-analysis including all study participants. We identified various genomic regions with elevated density of associations. For example, we found on chromosome 15 near the *ACAN* gene a 200 kb-long genomic segments containing 25 independent associations. More generally, high density of associations within 100 kb-long genomic window is associated with the presence of an autosomal gene curated from the Online Mendelian Inheritance in Man (OMIM) database[6], which pathogenic mutations are known to cause syndromes of abnormal skeletal growth (>2.5× enrichment; $P<0.001$).

The 12,111 GWS associations were clustered within 7,209 non-overlapping GWS loci) covering ~21% of the human genome. Using a partitioned GREML analysis we found that >90% of the common variants SNP-based heritability in 5 ancestry groups can be explained by SNPs located within GWS loci. For European ancestries, which constitutes the majority of our discovery sample, that proportion reaches 100% of the SNP-based heritability. Importantly, these observations suggest that although the majority of HapMap3 SNPs within GWS loci are not GWS themselves (only 12,111 out of ~300,000 SNPs; Figure 1), yet-to-be-discovered height-associated variants are most likely located within these GWS loci.

Finally, we used estimated SNP effects from our GWAS meta-analysis at these 12,111 height-associated SNPs to calculate a polygenic score (PGS) of height (Methods). This PGS explains 40% of height variance in European ancestries individuals (76% of our discovery sample; Methods) but shows a reduced accuracy in individuals whose ancestries are more genetically distant to EUR. In particular, the prediction accuracy of the height PGS is only ~10% in individuals with predominantly African ancestries. Note that a 4-fold reduction in prediction accuracy is consistent with previous studies[7,8] and largely explained by linkage disequilibrium and allele frequency differences between European and African ancestries.[8]

Table 1. [Extracted and simplified from Yengo (2022)] Summary of results from within-ancestry and multi-ancestry GWAS meta-analyses. GWS: Genome-wide significant ($P<5\times10^{-8}$). COJO SNPs: near independent GWS SNPs identified using an approximate conditional and Joint analysis implemented in the GCTA software.

Cohort main ancestry or ethnicity	Number of studies	Sample size	Number of GWS COJO SNPs	Number of GWS loci
European	173	4,080,687	9,863	6,386
East-Asian	56	472,730	918	821
Hispanic	11	455,180	1,888	1,599
African	29	293,593	493	436
South Asian	12	77,890	69	66
Multi-ancestry meta-analysis	281	5,314,291*	12,111	7,209

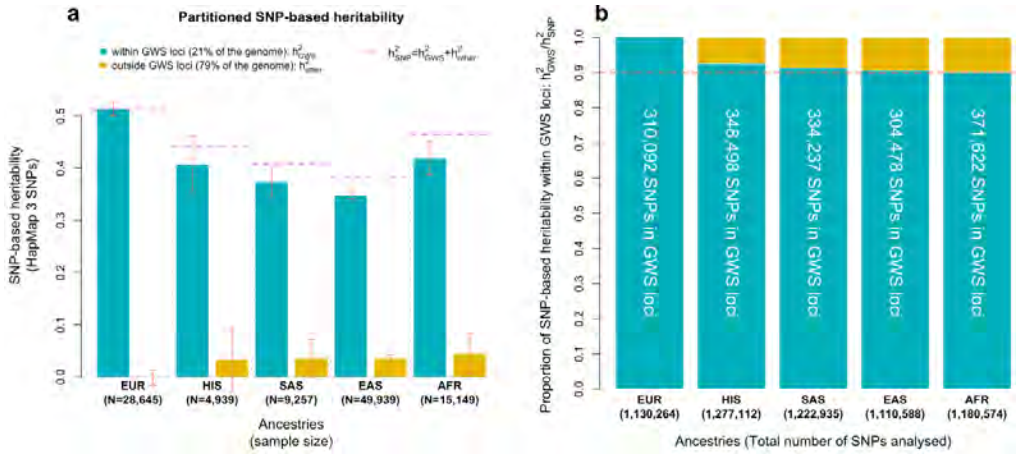


Figure 1. Variance of height explained by HapMap 3 SNP within genome-wide significant (GWS) loci. [Extracted and simplified from Ref.[3]]. Panel a shows stratified SNP-based heritability (h^2_{SNP}) estimates obtained after partitioning the genome into SNPs within 35 kb of a GWS SNP ('GWS loci' label) vs SNPs >35 kb away from any GWS SNP. Analyses were performed in samples of five different ancestry/ethnic groups: European (EUR: meta-analysis of UK Biobank (UKB) + Lifelines study), African (AFR: meta-analysis of UKB + PAGE study), East-Asian (EAS: meta-analysis of UKB + China Kadoorie Biobank), South-Asian (SAS: UKB) and Hispanic group (HIS: PAGE). Panel b shows that >90% of h^2_{SNP} in all ancestries is explained by SNPs within GWS loci identified in this study.

Discussion

In this study, we performed the largest, to date, GWAS of any human trait and thereby identified >12,000 associations explaining >90% of the common variants SNP-based heritability of height. A genetic predictor of height derived from these findings reaches an accuracy of 40% in European ancestry individuals, which rivals that of the average height of parents (i.e. 43%). Interestingly, we showed that these two predictors can be linearly combined to reach an accuracy of 54% and that the optimal weights only depend on the accuracy of the SNP-based predictor, the narrow sense heritability and the spousal correlation for traits, like height, subjected to assortative mating in the population.

By quantifying the relationship between sample sizes and GWAS discoveries, we found that prioritisation of the relevant biological pathways and gene sets is possible at relatively small sample sizes (n<250,000), while identification of variants almost linearly increases with sample size, thereby contributing to increase the density of associations within previously identified loci and reflecting the large allelic heterogeneity for height-associated haplotypes.

Our study has a few limitations, many of which are being currently addressed. The first one is the over-representation of European ancestries in our discovery samples, which limits the transferability of some of our findings to other human populations. Another limitation is that our study focused on common genetic variation (>1%), while our large sample size could, in principle, yield enough power to detect associations involving lower frequency SNPs. However, our large sample is in fact an aggregation of many small studies, which creates a challenge to distinguish true signals from study-specific artefacts.

Altogether, our study reveals that extremely large sample sizes can resolve the missing heritability problem by mapping specific genomic regions likely containing most of the causative genetic variation for complex traits like human height.

References

Lui J.C., Nilsson O., Chan Y., Palmer C.D., Andrade A.C. *et al.* (2012) Human Molecular Genetics 21(23): 5193-5201. https://doi.org/10.1093/hmg/dds347

Martin A.R., Kanai M., Kamatani Y., Okada Y., Neale B.M. *et al.* (2019) Nature Genetics 51(4): 584-591. https://doi.org/10.1038/s41588-019-0379-x

Wang Y., Guo J., Ni G., Yang J., Visscher P.M. *et al.* (2020) Nature Communications 11(1): 3865. https://doi.org/10.1038/s41467-020-17719-y

Yang J., Benyamin B., Mcevoy B.P., Gordon S., Henders A.K. *et al.* (2010) Nature Genetics 42(7): 565-569. https://doi.org/10.1038/ng.608

Yang J., Ferreira T., Morris A.P., Medland S.E., Madden P.a.F. *et al.* (2012) Nature Genetics 44(4): 369-375. https://doi.org/10.1038/ng.2213

Yang J., Lee S.H., Goddard M.E. and Visscher P.M. (2011) Am J Hum Genet 88(1): 76-82. https://doi.org/10.1016/j.ajhg.2010.11.011

Yengo L., Sidorenko J., Kemper K.E., Zheng Z., Wood A.R. *et al.* (2018) Hum Mol Genet 27(20): 3641-3649. https://doi.org/10.1093/hmg/ddy271

Yengo L., Vedantam S., Marouli E., Sidorenko J., Bartell E. *et al.* (2022) bioRxiv: 2022.2001.2007.475305. https://doi.org/10.1101/2022.01.07.475305

Part 2

Young scientist career development

4. History and author analysis of the World Congresses on Genetics Applied to Livestock Production

A.L. Van Eenennaam[1*] and S. Hermesch[2]

[1]Department of Animal Science, University of California, 1 Shields Ave, Davis, CA 95616, USA; [2]AGBU, a joint venture of NSW Department of Primary Industries and University of New England, 2351, Armidale, Australia; alvaneenennaam@ucdavis.edu

Abstract

It is almost fifty years since the first World Congress on Genetics Applied to Livestock Production (WCGALP) was convened in Spain in 1974. This paper, targeted for the early-career scientist session, details the history and evolution of WCGALP, and outlines gender composition of the Permanent International Committee, session chairs, plenary and invited speakers, and authors. Data containing 24,351 author records from all WCGALP proceedings, and old programs were used for the analyses. In the early years, there were few women involved in any aspect of WCGALP, however, more recent congresses have featured an increasing proportion of women on WCGALP committees and serving as session chairs. Based on recent data, gender is not playing a determining role in selection of submitted papers for oral presentations. We recommend greater gender diversity on WCGALP committees and encourage early-career scientists to actively participate in WCGALP conferences to further increase the diversity of future congresses.

Introduction

The idea of World Congresses for specialists working in livestock production was hatched in Spain in the 1960s. It started with a World Congress on Animal Feed held in Madrid in 1966 and 1972. Then in 1974 the first World Congress on Genetics Applied to Livestock Production (WCGALP) was held in Madrid. Starting with the second Congress in 1982, also held in Madrid, the event has been held every four years at locations throughout the world, and this Congress in Rotterdam in 2022 is the twelfth.

The governance of the first WCGALP was under the auspices of the local organizing committee, chaired by Carlos Luis de Cuenca, from the University of Madrid. Alan Robertson from the University of Edinburgh joined Carlos Luis de Cuenca in creating a Permanent International Committee (PIC) for WCGALP in 1974, and he was the first President of this committee which has traditionally been primarily responsible for evaluating bids for the location of the next WCGALP. William (Bill) Hill also from the University of Edinburgh, later served as President of this committee for many years. Honorary members of the PIC were Gordon Dickerson and Bill Hill.

The format of WCGALP has varied throughout the years. At the first WCGALP there were six main topics: (1) factorial genetics; (2) cytogenetics; (3) biochemical genetics; (4) quantitative genetics; (5) genetic modelling with laboratory animals; and (6) applied genetics to breeding programs with separate sessions for dairy cattle, beef cattle, sheep, swine/pig and poultry breeding programs. The format at that event included an invited speaker leading off each of these sessions, and invitations to speak were typically made on the recommendation of the local session organizers. Subsequent conferences followed a similar model, some featuring invited plenary speakers, and over time new fields such as transgenic animal research, molecular genetics, and genomics were included as they became established. At WCGALP 2022, the focus of sessions has been shifted to also include addressing a diverse range of societal challenges, relevant for animal breeding, rather than focusing solely on development of new methodologies. This change in topics at Congresses reflects the tremendous developments in the field of genetics and animal breeding in the almost 50 years since the first WCGALP, and moreover profound changes in both society and academia during this time.

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_4, © A.L. Van Eenennaam and S. Hermesch 2022

Materials & methods

Information about authors of papers presented at WCGALP was obtained from the WCGALP digital archive (http://www.wcgalp.org/). Data contained 24,351 author records detailing the author's name, the year the paper was presented, the title of the paper, and the volume in which the proceedings paper was published. The first name of authors was run through genderize.io (https://genderize.io/), to predict the gender of a person given their name. The gender of plenary speakers prior to 2014, and the members comprising the PIC were manually curated. Further, individuals who were authors on 15 or more WCGALP proceedings papers were examined as a group. If gender was unassigned for authors it was curated manually based on first-hand knowledge, or using the internet-based ISI Web of Science to search for other same-authored papers that provided first names, and/or Google to locate profiles or images of them. Prior to 2014, only initials were available for WCGALP authors, which made it difficult to designate the gender of these authors and gender was assigned unknown. Hence there is less confidence associated with data prior to these two most recent conferences.

Results

For many years both the PIC and the local organising committees (LOC) were comprised predominantly of men. To our knowledge, Helen Newton Turner was the first female invited session speaker in 1974 and member of the PIC, Susanne Hermesch was the first woman to be voted President of this Committee in 2018, and Yvette de Haas was the first female co-leading a LOC in 2022. There has been a continuous increase in the proportion of women on the PIC of WCGALP from 5% in 2010, 15% in 2014, 26% in 2018 and 33% in 2022.

The local organizing committee decides on plenary speakers and session chairs which have both included more women in 2014 and 2018 (Table 1), reflecting an increased awareness of gender balance in key-note presentations or positions. The first invited female plenary speaker was Jenny Marshall Graves in Australia in 1998. The number of proceedings papers increased over time (Table 2), with the exception of conferences in Australasia where there was a decrease, likely due to the travel involved for the majority of European and North American participants to attend these conferences. There has been an increase in the number of authors per paper, mirroring a trend whereby the average number of co-authors per paper published has increased steadily over the past century in all scientific disciplines.

A total of 201 authors had 15 or more papers presented at multiple WCGALP. Jack Dekkers who first presented in 1994 was the most prolific author with 64 papers, closely followed by John Woolliams, who first presented in 1990, with 62. The most prolific female author was Lucia Albuquerque with 35 papers who first presented in 1994. Two authors, Mike Goddard and Moshe Soller, had papers at each of the first 11 WCGALP spanning 1974 to 2018, while Karin Meyer was the female author with the highest WCGALP attendance, participating in conferences from 1982 to 2018. Figure 1 shows the breakdown of the most prolific authors by gender. There was a gender difference ($P<0.005$) in the average number of papers per author in this highly prolific group. For example, of authors in this group that were first-time authors in 2002 and are therefore likely of a similar age, male authors ($\bar{x}=24.2$ (± 1.4), n=35) had on average 5-6 more papers than female authors ($\bar{x}=18.7$ (± 0.8), n=16). Of the 2006 first-time authors, 10 male and one female author had 15 or more papers by 2018, and of the 2010 first-time authors only two male authors had 15 or more papers by 2018.

Encouragingly, of the 2014 first-time authors, a single female author, Daniela Lourenco, had 15 or more papers by the 2018 congress. Approximately one quarter of all the 2014 first-time authors (673 men, 467 women), returned as an author on a paper presented at the 2018 conference, 151 and 136, respectively. Of the 801 papers that were originally submitted to the 2018 New Zealand conference, 59% from male authors

Table 1. The number (N) of plenary speakers, session chairs and invited speakers along with percentage (%) of females for each group where gender of first author was known.

Year	Location	Plenary		Chairs		Invited	
		n	%	n	%	n	%
1974	Madrid, Spain	no plenaries		10	0	67	2
1982	Madrid, Spain	no plenaries		10	0	48	0
1986	Nebraska, USA	no plenaries		22	0	74	0
1990	Edinburgh, Scotland	4	0	34	3	74	4
1994	Guelph, Canada	no plenaries		15	0	73	3
1998	Armidale, Australia	5	20	29	7	58	12
2002	Montpellier, France	3	0	38	10	50	8
2006	Belo Horizonte, Brazil	3	0	22	0	78	13
2010	Leipzig, Germany	8	13	34	18	44	22
2014	Vancouver, Canada	5	0	36	6	108	24
2018	Auckland, New Zealand	4	50	53	23	34	24
2022	Rotterdam, the Netherlands	6	50	156	32	15	27

Table 2. Number of papers (P) and authors (A), authors per paper (A / P) and number of authors with gender known or unknown (N/A) for the WCGALP. Cells shaded dark grey are associated with low confidence in the correct gender identification of the author.

Year	Location	P	A	A / P	Males	Females	N/A	%females[1]
1974	Madrid, Spain	226	399	1.8	209	3	187	1%
1982	Madrid, Spain	364	755	2.1	403	6	346	1%
1986	Nebraska, USA	277	627	2.3	343	11	273	3%
1990	Edinburgh, Scotland	344	784	2.3	435	25	324	5%
1994	Guelph, Canada	946	1,229	1.3	649	27	553	4%
1998	Armidale, Australia	257	858	3.3	419	25	414	6%
2002	Montpellier, France	949	3,812	4.0	1,984	86	1,742	4%
2006	Belo Horizonte, Brazil	868	3,899	4.5	1,978	161	1,760	8%
2010	Leipzig, Germany	789	3,477	4.4	1,930	186	1,361	9%
2014	Vancouver, Canada	946	4,456	4.7	2,893	1,239	324	30%
2018	Auckland, New Zealand	753	4,055	5.4	2,580	1,187	288	32%

[1] Percentage based on authors with gender known.

and 41% from female authors, around 80 were withdrawn, with an approximately equal number from both genders. Of those papers accepted, 38% were presented in an oral session, irrespective of the gender of the author.

Discussion

The rationale for preparing this paper for the early-career scientist session was our experiences and observations as female scientists both with a PhD awarded in 1997 in a traditionally-male scientific discipline. The dearth of women in the WCGALP congresses initially can possibly be explained by the number of women actively participating in the field over time. Nevertheless, it has generally been observed that the proportion of female speakers differs dramatically by the gender of the organizer, with male-

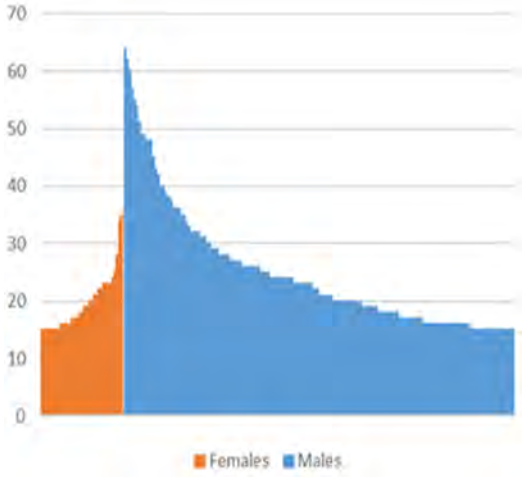


Figure 1. Number of WCGALP proceedings papers sorted by author gender for authors with at least 15 papers.

organized symposia having fewer female speakers than those organized by women, or a mix of genders (Isbell *et al.*, 2012). Further, blinded submissions of abstracts (as to author and institution) was also found to result in more women being chosen as symposium speakers (Newkirk *et al.*, 2005). We recommend improving the involvement of women, early-career scientists, scientists from developing countries and other minorities in key positions involved in the planning and organisation of future WCGALP events to ensure greater diversity of invited speakers. It would also be helpful to ask authors include their ORCID digital identifier when submitting proceedings papers to facilitate analyses such as those presented in this paper in the future.

There are some encouraging trends in the data we analysed. The days of all-male plenary sessions seem to be a thing of the past, and the proportion of female session chairs and invited speakers gradually increased over time. And based on recent data, gender is not playing a role in selection of submitted papers for oral presentations, nor in the observed willingness of conference attendees to attend subsequent conferences. This offers tremendous opportunities for early-career scientists to raise their profile at WCGALP events independent of gender. Inviting speakers is risky for session chairs and organisers, if the speaker is not well known. This naturally decreases the likelihood of unfamiliar scientists to be invited. We encourage all early-career scientists to speak up at WCGALP, to ask questions and to get noticed. We believe that both of these aspects, greater gender diversity of organising committees or chairs and increased courage of early-career scientists to get noticed will lead to greater diversity of session chairs, invited speakers, authors and attendees at future WCGALP conferences.

Acknowledgements

We thank William Hickson for his help as a 'friendly computer geek' and Rea Alexandri, Jean-Michel Elsen, Dorian Garrick, Steve Miller, Pieter Knap, Matt Spangler and Brian Wickham for their support and help.

References

Isbell L.A., Young T.P., and A. H. Harcourt A.H. (2012). PLoS One 7(11):e49682. https://doi.org/10.1371/journal.pone.0049682

Newkirk M.M., Richie E., and Lunney J.K. (2005). Nature Immunology 6(9):855-855. https://doi.org/10.1038/ni0905-855

5. What industry needs from an animal breeding graduate

T.J. Byrne[1*], P.F. Fennessy[2] and P.R. Amer[2]

[1]AbacusBio International Limited, Roslin Innovation Centre, Easter Bush, Midlothian, EH259RG Edinburgh, United Kingdom; [2]AbacusBio Limited P.O. Box 5585, Dunedin, New Zealand; tbyrne@abacusbio.co.uk

Abstract

The historical focus of animal breeding post-graduate qualifications on breeding value estimation methods, and genomic approaches more recently, has resulted in a shortage of people with the necessary skills to design and lead industry breeding programs. While numeracy is important for a complete understanding of animal breeding, prioritisation of mathematical and statistical skills for post-graduate animal breeding training often discourages suitable young people from entering the field. The different incentives and structures of industry and academia also means that career paths in academia and industry are very different and transitioning from one to the other can be challenging. This paper outlines the skills we believe are required for the future, both within an agricultural genetic improvement-based consultancy business, and for our clients.

Introduction

AbacusBio is a consultancy business with approximately 50 people based in New Zealand, the United Kingdom and North America. We have clients across the globe, and a multinational team representing 16 nationalities and many different cultural perspectives. While our consultancy work covers many topics, we have a long history working to support the design and practical function of animal breeding programs. We focus on building long term relationships with our clients and have worked in many different structures and cultures within our client organisations.

In this opinion piece, we outline the key skills we believe are required to build well rounded animal breeders that are capable of leading livestock breeding programs through the next 50 years and beyond.

Key future themes

The divide between academia, government and industry continues to grow. However, all three will be key contributors to the genetic gain achieved over the next 50 years. Academia makes available new ideas and methodologies, and in some cases can be directly responsible for their implementation. For many programs, government is a key source of research funds, but it is also still common for both government regulations and Research and Development Investment to have a significant impact on breeding scheme designs. Industry in this context could be a team working on national infrastructure that supports animal evaluation and genetic improvement, or alternatively a private company competing to survive in a competitive market for gene stocks.

While the potential for introduction of new and disruptive technologies (e.g. historically we have seen BLUP, Genomic Selection, Embryo Transfer, Semen Sexing and Gene Editing) remains, navigation of a number of key tensions will become increasingly important for livestock breeding programs in the future. These tensions relate to the trade-off between short term genetic gain and long-term genetic diversity, the trade-off between genetic change in traits important to society (e.g. environmental and animal welfare influencing traits) and those traits that directly drive farming convenience and competitiveness), and the opportunity to import superior genetic merit from other countries and populations, versus the effort required to directly improve the local population. Moreover, there are financial (e.g. cost-benefit analysis) and operational considerations that span technical and business/organisational spheres.

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_5, © T.J. Byrne *et al.* 2022

Successful navigation of these issues will require a new generation of animal breeders, who have a broad range of life experiences, skills, and perspectives.

Training needs

Young professionals need a basic background in quantitative genetics and animal breeding principles, as offered by many undergraduate courses in universities. From here a working knowledge of the principles of mixed model statistical analysis is important, but there needs to be recognition that young professionals are more likely to follow a well-rounded animal breeding career than just crunching large data sets through novel statistical analysis procedures. Principles of trait prioritisation, inbreeding management, breeding scheme design and cost benefit analysis should also be taught in sufficient detail that young professionals can go on to apply these critical tools in their future careers.

Desired attributes

The following attributes are highly desirable when we recruit early career professionals into our business:

- team-work;
- numeracy and literacy;
- enquiring mind;
- willingness to challenge existing ideas and processes, but to do so in a safe and professional way;
- flexibility in approach to team and client needs;
- a desire for constant learning and professional growth;
- recognition of the importance of business principles in the practical application of science.

In addition to these attributes, we prefer applicants who can demonstrate through previous experience:

- some degree of existing professional network, gained through work experience and or industry internship;
- a background knowledge of one or other farming/ food production system, and the culture of the people working within that system;
- a diversity of experiences gained either through travel, or from having overcome challenging life circumstances;
- a PhD or Masters thesis that demonstrates that a diversity of topics has been addressed, therefore requiring a range of skills to have been mastered to achieve it;
- a demonstrated ability of being able to switch from an academic career to an industry role.

Discussion

While universities provide valuable skills and knowledge, the cultural drivers and performance incentives in academia are very different to those in industry. It is therefore essential that early career professionals seek industry placement opportunities as part of their education. Where possible, combining these experiences with international travel makes for very valuable life experiences that serve people well for the duration of their career.

There is a growing push towards collaboration between universities and industry, as the talent shortage in animal breeding intensifies. With the convergence of plant and animal breeding approaches that is occurring due to genomics, more interchangeability between plant and animal breeding careers can be expected in the future.

Our experience suggests that several structured and unstructured approaches to early career development can build more diverse and valuable talent for the future of animal breeding. The value of mentorship, on-the-job training/internships, and exposure to different career paths cannot be understated.

6. Young scientist career development: a global discussion on challenges, opportunities and solutions

J.E. Newton[1,2*], M. Gòdia[3], B. Raymond[4] and M.H.O. Toledo-Alvarado[5]

[1]Agriculture Victoria Research, AgriBio, 5 Ring Road, Bundoora 3083 Australia; [2]Action for Agriculture, P.O. Box 241, Jamberoo 2533, Australia; [3]Wageningen University & Research, Droevendaalsesteeg 1, 6708 PB, Wageningen, the Netherlands; [4]Chappes Research Centre, Limagrain Europe, Bâtiment 1, Route d'Ennezat, P.O. Box 63720, Chappes, France; [5]Department of Genetics and Biostatistics, National Autonomous University of Mexico, Ciudad Universitaria, P.O. Box 04510, Mexico City, Mexico;
jo.newton@agriculture.vic.gov.au

Abstract

Attracting, supporting, and retaining the next generation of young scientists into the field of animal breeding and genetics is important for the future of global livestock systems. The introduction of a Young Scientist Career Development session at the 12th WCGALP provides an invaluable platform to discuss the challenges, but more importantly, opportunities and solutions to foster the development of our next generation of animal breeders. A highlight of this session will be a facilitated panel discussion with Assoc. Prof. Daniela Lourenco (University of Georgia), Prof. Jennie Pryce (Agriculture Victoria Research and LaTrobe University), Prof. John Hickey (Bayer Crop Science) and Prof. Johan van Arendonk (Hendrix Genetics). These challenges will not be solved through a single panel discussion, or conference session. However, providing a platform for the WCGALP community to come together and work collaboratively on solutions to support young scientist career development is an important first step.

Introduction

Livestock industries are of utmost importance in modern society. This positioning has in part been achieved due to advances in the biological sciences – including nutrition, genetics, physiology and disease control. Today, livestock systems globally are facing numerous challenges and the field of animal breeding can play a role in meeting these challenges. Collectively, Animal Scientists, animal breeding companies, universities and governmental institutions need to ensure that the necessary young professionals are attracted and retained to the sector given the type of future leadership and research that the livestock industry requires.

Achieving this requires knowledge of challenges and opportunities that young animal breeders face as well as awareness of what constitutes a desirable career pathway for today's university graduates. The major factors determining the appeal for a particular career that were proposed some years ago (Harwin, 1979) included:

1. vocation/job opportunities;
    a. public sector;
    b. private sector;
    c. self-employed persons;
2. remuneration or earning power;
3. job satisfaction.

A number of other considerations are important to today's early career professionals. This includes job stability, career development, University/company recognition and reputation, work location, work-load and flexibility, parental leave arrangements and alignment between personal and organisational values. The benefits of mentoring, and access to career development for job satisfaction are well documented (Sinha,

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_6, © J.E. Newton *et al.* 2022

2020) and are also applicable to the field of animal breeding. Conversely, lack of career stability and long-term or permanent job contracts as well poor work-life balance are key reasons given for early and mid-career animal scientists to leave research positions (Mayberry *et al.* 2021).

Historically, marginalised groups have been underrepresented across STEM disciplines (Wanelik *et al.* 2020) so any conversation about career development should also consider the work still needed in diversity equity and inclusion, including gender equity. While some progress has been made, in part due to initiatives like the Athena Swan Charter (United Kingdom & Ireland) and similar schemes in North America and Australia, there is a long way to go in achieving STEM and animal breeding communities that embody best practice Diversity, Equity and Inclusion. Just 40% of respondents in *Nature's* 2021 salary and job satisfaction survey felt their employers were doing enough to promote diversity (Woolston, 2021). In addition, survey respondents who did not identify as white were more likely to report experiences of discrimination, bullying and harassment in the workplace (Woolston, 2021). Ethnicity and socioeconomic background have also been shown to impact early career progression in STEM (Wanelik *et al.* 2020). In Australia, in 2020 women who graduated with postgraduate STEM qualifications in agricultural and environmental studies earnt $AU 23,000 less than male graduates (Australian Government, 2021). In addition, only ~25% of all agricultural scientists in Australia are women with women also underrepresented in senior leadership roles.

The aim of introducing a Young Scientist Career Development session at the 12th WCGALP in Rotterdam, the Netherlands, 2022 is to provide a platform for a global discussion on the challenges and opportunities for young scientist career development.

Panel discussion

Given the diversity of career pathways in animal breeding available to young scientists, the inaugural Young Scientist Career Development session at the WCGALP 2022 will feature a panel discussion. Panellists will be asked to reflect on both the challenges and opportunities for young scientist career development. During the discussion, panellists will be asked to consider defining moments in their own careers, advice for young scientists, solutions to the challenges faced by young scientists and what they see as the key opportunities for the next generation of animal scientists.

A shortlist of potential panellists was identified by session co-chairs based on a number of criteria including; contribution and profile in the field of animal breeding and genetics, experience across academia, industry and government, diversity in geographical location, gender, experience and research/work focus areas.

Panellists

In this workshop, we will include 4 panellists to discuss their experience in the animal breeding sector, Assoc. Prof. Daniela Lourenco, Prof. Jennie Pryce, Prof. John Hickey and Prof. Johan van Arendonk. Brief biographies are given below.

Assoc. Prof. Daniela Lourenco, Associate Professor, University of Georgia, United States of America.
Assoc. Prof. Daniela Lourenco, is an Associate Professor in Animal Breeding, Genetics, and Genomics at the University of Georgia in the United States. She was born and raised in Brazil, where she earned her MS and PhD degrees in animal breeding and genetics from Maringa State University. She has been working in this field since 2004, and her current interests include the use of genomic information to increase rates of genetic progress, the development of methods for genomic evaluations, and the use of computational algorithms to analyze large data. Her research is focused on genomic selection in beef and dairy cattle, swine, poultry, and fish. Daniela's group has developed and implemented single-step genomic evaluations for several breeding companies and breed associations. She has published over 360 scientific papers and

proceedings, including over 125 refereed journal publications. Daniela has served as advisor or co-advisor for 12 Masters and PhD students, committee member for 18 graduate students at The University of Georgia and 14 at international universities, has received 27 visitors from several countries, and has supervised 9 postdocs. Highly active in all projects developed by the animal breeding group at University of Georgia since 2012, and has been a principal investigator in research projects with budgets up to $500,000.

Prof. Jennie Pryce, Principal Research Scientist, Agriculture Victoria Research and Professor, School of Applied Systems Biology, La Trobe University, Australia. Professor Jennie Pryce is a geneticist who is well known for her work in dairy genetics and genomics. Jennie is principal research scientist for Agriculture Victoria and Professor of animal genetics at La Trobe University. Jennie has dedicated her life to dairy herd improvement and is very passionate about research. She grew up on a dairy farm in Shrewsbury (UK) where at a young age she bred pedigree Holstein dairy cattle under Severnvale Holsteins prefix. Jennie received her BSc (Hons) first class and PhD from The University of Edinburgh (UK). Prior to her work in Australia, she was employed as a dairy geneticist by the Scottish Agricultural College (UK) and as a scientist by the dairy breeding company Livestock Improvement Corporation (New Zealand). In 2008 Jennie made Melbourne her home and started a research position with the Department of Primary Industries (now Agriculture Victoria Research). Jennie now leads the DairyBio animals research program and is responsible for the animals' program. This innovation is critical to the dairy industry – regional Victoria's largest employer and a $1 billion export industry for the state. Jennie is instrumental in development and delivering of successful research, working to build industry groups' understanding and adoption, and led development of tools for direct benefit on farm by most farmers. Her position closely aligns with her main research interests of genetic improvement of traits associated with health, welfare and environmental impact in dairy cows.

In 2016 Jennie was the first non-North American recipient of the prestigious American Dairy Science Association J.L. Lush Award for Animal Breeding and Genetics and in 2019, 2020 and 2021 was named as Australia's top researcher in the field of animal husbandry by research analytics firm League of Scholars; this award is based on number of citations for research papers published in the top 20 journals in each field over the past 5 years.

Prof. John Hickey, Bayer Crop Science, Spain. Prof. John Hickey did his PhD studies in Wageningen University and Research, the Netherlands, on genetic evaluations of multiple breed cattle populations followed by post-doctoral positions in Australia (University of New England) and Mexico (International Maize and Wheat Improvement Centre – CIMMYT). In Australia he developed methods and algorithms, with accompanying genotyping strategies to impute genotype data in livestock populations. This work has subsequently been implemented in some of the largest animal breeding programs globally. In Mexico he researched aspects of genomic selection in plant breeding programs. Subsequently, John held the Chair of Animal Breeding at The Roslin Institute which is part of The Royal Dick School of Veterinary Studies at the University of Edinburgh in the United Kingdom. While at the Roslin Institute, his research group worked on breeding projects in pigs, chickens, cattle, groundnut, forest trees, strawberry, maize, wheat and other species. Software and algorithms developed by John Hickey and colleagues underpin aspects of several of the largest breeding programs globally. Since 2020, John joined Bayer Crop Science as Head of Corn Product Design, based in Barcelona, Spain.

Prof. Johan van Arendonk, Hendrix Genetics, the Netherlands. Prof. Johan van Arendonk received his PhD-degree at Wageningen University in 1985. From 1985 onwards, he worked at the Department of Animal Breeding of the Wageningen University. From 1985 until 2015, he worked in different roles at Wageningen University in the domain of animal breeding and genetics. In 1998, he was promoted to

personal chair and in 2002 he became full professor in Animal Breeding and Genetics In that period, he (co-) supervised around 90 PhD students and (co-)authored 350 papers in scientific journals. The emphasis in his career at Wageningen University changed from doing research himself, to training and coaching the new generation, to building and managing a team of professionals. During the last 5 years at Wageningen, he combined his role as chair of the animal breeding and genetics group with that of Dean of Science, in which he had a university-wide responsibility. In January 2016, he started as Chief Innovation & Technology Officer at Hendrix Genetics a global multi-species animal genetics and technology company. In that role, he contributes to strengthening the research and product development activities for the different species. In April 2020, he joined the executive committee of the company where he is responsible for Research & Development, the company's Sustainability Program and Innovation portfolio.

Conclusions

Given the many challenges young scientists face globally, not just in the field of animal breeding and genetics, the young scientist career development session represents a valuable addition to the WCGALP 12th Conference Program. We are indebted to our panellists for donating their time to participate in this session. The challenges in attracting, retaining and supporting the next generation of animal breeders will not be solved through a single panel discussion, or conference session. However, by creating space for these important discussions we are providing a platform for the whole of the WCGALP community to come together and work collaboratively on solutions to addressing these challenges.

References

Australian Government (2021). STEM Equity Monitor. Data Highlights 2021. Australian Government Department of Industry, Science, Energy and Resources, Australia

Harwin G.O. (1979). S. Afr. J. Anim. Sci 9:157-162

Mayberry D., Hatcher S., and Cowley F. (2021) Anim. Prod. Sci. 61(3):201-207. https://doi.org/10.1071/AN20115

Sinha R. (2020) In: Career Development and Job Satisfaction. IntechOpen, Chapter 2.

Wanelik K.M., *et al.* (2020) Ecol. Evol. 10(14): 6870–6880. https://doi.org/10.1002/ece3.6423

Woolston C. (2021) Nature. 600:177-179. https://doi.org/10.1038/d41586-021-03043-y

Part 3

Education

7. Digital learning in postsecondary education: 30 transformational years

J.J. Lindburg

University of Nebraska System, 3835 Holdrege Street, Lincoln, NE 68583, USA; jlindburg@nebraska.edu

Abstract

Digital education has brought sweeping change to postsecondary education since its foundation in the early 1990s. Understanding its origins and complexities; perceptions of faculty and students; and best practices for shaping and supporting the next generation digital learning environment are critical elements for faculty and administrators to consider in positioning our modern universities for post-pandemic success.

Introduction

Digital Education – sometimes also referred to as online learning, distance education, blended learning, hybrid courses, flipped classroom, remote learning, asynchronous delivery, or technology-infused teaching – has its roots in correspondence education, where a century ago institutions would mail packets of learning materials to an individual who would return completed assignments for grading at the end of a term without faculty-to-student or student-to-student engagement. In the 1980s, this model shifted to satellite education, where a faculty member would teach a classroom of students on a physical campus, and their lecture would be broadcast live to room(s) of students in another physical location.

The invention of the internet in 1992 made digital education as we know it today a possibility (Harasim, 2000), and experimental faculty began shaping some of the earliest forms of online course as we know them today on homegrown webpages in the late 1990s and early 2000s. The 21st century began with a 'paradigm shift in attitudes toward online education' (Harsim, 2000), which brought a rapid expansion of courses, programs, and even new postsecondary institutions. In the past two decades, a more modern emergence of digital learning has become a staple of postsecondary education nationally and internationally, in which faculty use a learning management system to deliver course materials, communicate with students, receive assignments, provide feedback, issue grades, and manage general activities of a course. The learning management system has become critical to teaching and learning (Pomerantz *et al.*, 2018), with most institutions running at least one learning management system (Lang and Pirani, 2016), and the rate of overall adoption of learning management systems rivaling only that of cars and cellphones in the United States. Prior to the COVID-19 pandemic, digital education was reaching 6.9 million students in the United States, or 35% of students in the nation (IPEDS, 2018).

Discussion

Broad debates about the efficacy of digital education have long plagued the modality from being fully embraced by all sectors of universities. Faculty argue that engagement is sub-standard, that learning objectives are not effectively met when students study exclusively online, and that the personal and professional development growth capabilities of students are not fully realized when students are learning at a distance. These concerns were further exacerbated by the fact that for-profit institutions dominated the market in the early 2000s, and largely focused on enrollment and marketing tactics rather than rigor, an institution's reputation, faculty credentials and preparedness, or quality of course design. Yet, despite these ongoing debates, students continued to opt into these digital courses and programs that increased their flexibility and access in order to attain a degree. Students' perception of the quality of online learning is also important to consider as they continue to opt into this modality. According to Van Wart *et al.* (2020), factors shaping a student's perception of the quality of online learning include 'their own sense of learning

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_7, © J.J. Lindburg 2022

achievement, satisfaction with the support they receive, technical proficiency of the process, intellectual and emotional stimulation, comfort with the process, and sense of learning community.'

In the past decade, there has been a surge in technology development as related to digital education. Rudimentary webpages and cumbersome, clunky learning management systems gave way to the 'next generation digital learning environment' we see curated and managed carefully at leading institutions. The next generational digital learning environment addresses five dimensions – (1) interoperability and integration; (2) personalization; (3) analytics, advising, and learning assessment; (4) collaboration; and (5) accessibility and universal design (Pomarantz *et al.*, 2018). To support a robust, learner-centric next generation learning environment, three broad considerations must be made: a strong infrastructure to allow for meaningful engagement, humanizing online experiences, and emphasizing access and affordability.

On top of this strategic infrastructure exists tremendous capabilities for innovation and personalized, meaningful engagement at both the peer and faculty-to-student interaction. This engagement in a virtual environment proved to be critical during the COVID-19 pandemic. As the instructional technologies and pedagogical practices that faculty relied on to complete the pivot to remote learning in Spring 2020 not only helped to engage students, but also aided in mitigating student stress. These practices and strategies were not just acts of good teaching, but were acts of compassion during a crisis (Anderson, 2021). As Martin and Bollinger (2018) point out, 'student engagement increases student satisfaction, enhances student motivation to learn, reduces the sense of isolation, and improves student performance in online courses.'

The next generation digital learning environment will be a more humanized space. Humanizing is a teaching strategy that 'removes affective and cognitive barriers from online learning, and provides a pathway towards a more equitable future for higher education that supports the success of all students' (Pacansky-Brock *et al.*, 2020). Some strategies that are at the core of humanizing the online environment include instructor-student relationships, welcoming visuals, warm pre-recorded communications, trust, a 'liquid syllabus,' and a culture of care (Pacansky-Brock *et al.*, 2020). Stock photos of students learning in isolation that are sometimes associated with the digital learner today will give way to a connected, holistic, student-centric experience that engages learners, fosters a sense of community, and accomplishes course objectives.

Finally, the next generation digital learning environment paves the way for true mission-centric commitments to student access and affordability. Efforts like the use of free or reduced-cost digital course materials integrated into the learning management system in lieu of traditional textbooks – commonly referred to as Open Educational Resources – have been accelerated at large institutions like Ohio State University, who has saved students over $20M through this practice. These types of strong commitments to student success and degree attainment by way of innovation within a strategic, well-positioned, interoperable learning management environment will be critical efforts for postsecondary institutions in the future.

As universities continue to shape their digital learning footprint, they must remember that the COVID-19 pandemic exposed the concept of the 'digital divide,' or equal access to technology. The World Economic Forum reports that half of the world's population has no access to the internet, fewer than one in five people in the least developed countries are connected, and that the digital divide impacts women more than men (Broom, 2020). The second element of the digital divide is digital readiness skills. The Pew Research Center reports that just 29% of the population in the United States have high levels of digital skills, with 70 million individuals in the country not prepared for robust online use like we would find in a university digital learning environment (Horrigan, 2016). Even for traditional college-age students, the second layer of the digital divide is staggering in that age is not an accurate predictor of digital readiness but rather education

perpetuates digital literacy skills, meaning that our traditional college students may in fact be digital natives and accustomed to using technology, but not in the capacity needed to be effective in navigating postsecondary digital education (Kirschner and De Bruyckere, 2017).

In conclusion, digital education has brought sweeping change to postsecondary education since its foundation in the early 1990s. Understanding its origins and complexities; perceptions of faculty and students; and best practices for shaping and supporting the next generation digital learning environment are critical elements for faculty and administrators to consider in positioning our modern universities for post-pandemic success.

References

Anderson, J. (2021) Out of Crisis, Compassion: Using Instructional Technologies to Alleviate Student Stress. EDUCAUSE Review. Available at: https://er.educause.edu/articles/2021/8/out-of-crisis-compassion-using-instructional-technologies-to-alleviate-student-stress

Broom, D. (2020) Coronavirus has exposed the digital divide like never before. Available at: https://www.weforum.org/agenda/2020/04/coronavirus-covid-19-pandemic-digital-divide-internet-data-broadband-mobbile/

Harasim, L. (2000) Internet and Higher Education 3(1-2):41-61. https://doi.org/10.1016/S1096-7516(00)00032-4

Horrigan, J., (2016) The meaning of digital readiness. Pew Research Center. Available at: https://www.pewresearch.org/internet/2016/09/20/the-meaning-of-digital-readiness/

Kirschner, P.A., and De Bruyckere, P.D. (2017) Teaching and Teacher Education 67:135-142. https://doi.org/10.1016/j.tate.2017.06.001

Lang, L., and Pirani, J.A. (2016) The 2015 Enterprise Application Market in Higher Education: Learning Management Systems. Research report, Louisville, CO, USA: ECAR. Available at https://library.educause.edu/-/media/files/library/2016/12/cdsm816d.pdf

Martin, F., and Bollinger, D.U. (2018) Online Learning 22(1):205-222. https://doi.org/10.24059/olj.v22i1.1092

National Center for Education Statistics (NCES) (2018) Integrated postsecondary education data system. Available at: https://nces.ed.gov/ipeds/

Pacansky-Brock, M., Smedshammer, M., and Vincent-Layton, K. (2020) Humanizing online teaching to equitize higher education. Current Issues in Education, 21(2). Available at: http://cie.asu.edu/ojs.index.php/cieatasu/article/view/1905.

Pomerantz, J., Brown, M., and Brooks, D.C. (2018) Foundations for a Next Generation Digital Learning Environment: Faculty, Students, and the LMS, Research report, ECAR, Louisville, CO, USA. Available at https://www.educause.edu/ecar/research-publications/foundations-for-a-next-generation-digital-learning-environment-faculty-students-and-the-lms/introduction

Van Wart, M., Ni, A., Medina, P., Canelon, J., Kordrostami, M. *et al.* (2020) International Journal of Educational Technology in Higher Education 17(53). Available at: https://doi.org/10.1186/s41239-020-00229-8

8. A panel discussion: broadening learning by incorporating industry professionals in education

J.B. Cole[1*], R.A. Mrode[2,3], K.M. Olson[4] and P. Knap[5]

[1]URUS Group LP, 2418 Crossroads Drive, Suite 3600, Madison, WI 53718, USA; [2]International Livestock Research Institute (ILRI), BOX 30709-01001, Nairobi, Kenya; [3]Scotland Rural College, W Mains Rd, Edinburgh EH9 3JG, United Kingdom; [4]ABS Global, 1525 River Road, Deforest, WI 53532, USA; [5]Genus-PIC, Ratsteich 31, 24837 Schleswig, Germany; john.cole@urus.org

Abstract

Educating future animal breeders and geneticists is fundamentally important. During this training, young scholars, both those completing graduate and post-doctoral studies and those establishing their careers, benefit from understanding the range of professional opportunities available, and the skills they require. This includes exposure to the needs of various livestock sectors. Through a panel discussion in the Education session, industry-based experts engaged in developing nation, extensive, and intensive production systems will share their experiences and discuss opportunities to serve a range of livestock industries. In an interactive session, the moderator will present questions to the panel. This will be followed by an open discussion with the audience delving further into the points raised. The goal of the session is to broaden the attendees' understanding, particularly younger scholars, of the skills needed and professional opportunities available to meet the needs of livestock industries globally in animal breeding and genetics.

Introduction

Study of the theory and application of animal breeding and genetics is a centrepiece of graduate and post-doctoral training programmes in animal breeding and genetics. Yet, to fully meet the demands of modern agriculture, exposure to the diverse opportunities available across livestock sectors needs to part of that learning. Including industry-based professionals in the educational process is one way to fill that gap.

As part of the Education session at the WCGALP 2022, through a moderated panel discussion, industry-based experts engaged in developing nation, extensive, and intensive production systems will share their experiences and discuss opportunities to serve a range of livestock industries. In an interactive session, the moderator will present questions to the panel. This will be followed by an open discussion with the audience delving further into the points raised. The goal of the session is to broaden the attendees understanding, particularly younger scholars, of the skills needed and professional opportunities available to meet the needs of livestock industries globally in animal breeding and genetics.

Materials & methods

The session will be led by a moderator, who will present a series of questions to the panel members to consider. This will be followed by an open discussion with attendees, allow them opportunity to seek further insights from the panel members on the points raised, and to share their own experiences.

Panel members and moderator. The panel members were invited due to their recognized expertise in engaging with developing nation, extensive, and intensive production systems.

Dr Raphael Mrode holds a joint position at the International Livestock Research Institute, Nairobi, Kenya and Scotland Rural College, Edinburgh, United Kingdom. He has been lecturing on the Edinburgh University MSc course on Quantitative Genetics and Genome (QGG) since 2005, is a co-teacher at the international course on QGG at ILRI-BeCA, Kenya and has been principal instructor at similar courses in

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_8, © J.B. Cole *et al.* 2022

South Korea, Romania, Turkey, Nigeria, and Kenya in the last 12 years. He is the author of the book, 'Linear models for the Prediction of Breeding Values', which used in many animal breeding courses.

Dr Katie Olson is the Global Director of Dairy Product Development at ABS Global. She is responsible for creation of dairy genetics and germplasm, developmental dairy genetics research, product validation and elements of genetic services. She is on the board of directors for both the Council of Dairy Cattle Breeding, as well as the National Association of Animal Breeders (NAAB), and is a former chair of the board for NAAB.

Dr Pieter Knap studied animal breeding and aquaculture in Wageningen (NL) and has worked in commercial pig breeding since 1981: currently as Genetic Strategy Manager with Genus-PIC (based first in Scotland and now in Germany) and previously with the predecessor of Topigs (NL), and with Norsvin International (NO). From 2002 to 2013 he was also (vice-)president of the commission for pig production of EAAP.

The sessions moderator is Dr John Cole. Dr Cole is the PEAK Sr. VP for Research & Development at URUS Group LP, where he oversees the genetics and reproductive biology research for the largest producer of bovine genetics in the world. He was previously a Research Geneticist (Animals) for the U.S. Department of Agriculture's in-house research arm, the Agricultural Research Service, for 17 years. During his tenure at USDA, he was part of the team that developed genomic selection for U.S. dairy cattle, and his research interests include calving and functional traits and genetic diseases.

Aim and target audience. The aim of the panel discussion is to consider animal breeding education today and in the future, and to identify strengths and challenges in the training process to meet the needs of various industry sectors. It is envisioned that the session will especially attract graduate, post-doctoral and early career scientists. In addition, it is hoped that more experienced scientists will join the discussion to offer their own expertise and experience to the dialog.

Discussion

Although it is imagined that the conversation between the moderator, panel members and attendees will be free flowing, several potential topics for discussion are:

- Incorporating experiential learning, including industry internships, in instruction.
- Including greater focus on globalisation in the content of the animal genetics and genomics curriculum.
- Recruiting diverse and dispersed students (different cultural backgrounds and geographies) into the discipline.
- Integrating industry experts into instruction in animal breeding and genetics.
- Incorporating innovations in pedagogy into the animal breeding and genetics classroom.
- Overcoming the fear of the quantitative aspects of animal breeding and genetics among students.
- Finding the right balance in education between the foundations and current technologies affecting animal genetics and genomics.

The success of the session depends on active engagement among the panel members and attendees. The participation of graduate, post-doctoral, and early career scientists in the discussion will be particularly encouraged.

9. Conducting a national beef cattle genetics outreach program in the USA

K.D. Bullock[1*], M.L. Spangler[2], R.L. Weaber[3], T.N. Rowan[4], M.M. Rolf[3], J.E. Decker[5], D.D. Loy[6], B.L. Golden[7], J.J. White[8] and A.L. Van Eenennaam[9]

[1]University of Kentucky, 807 WP Garrigus, Lexington, KY 40546, USA; [2]University of Nebraska – Lincoln, P.O. Box 830908, A218h ANSC, Lincoln, NE 68583, USA; [3]Kansas State University, 232 Weber Hall, Manhattan, KS 66506, USA; [4]University of Tennessee, 256 Brehm Animal Science Building, 2506 River Dr, Knoxville, TN 37996, USA; [5]University of Missouri, 920 East Campus Dr, Columbia, MO 65211, USA; [6]Iowa State University, 313E Kildee Hall, Ames, Iowa 50011, USA; [7]Theta Solutions, LLC, 4745 Plover St NE, Lacey, WA 98516, USA; [8]National Cattlemen's Beef Association, 9110 East Nichols Avenue Suite 300 Centennial, CO 80112, USA; [9]University of California – Davis, 2113 Meyer Hall, One Shields Ave, Davis, CA 95616, USA; dbullock@uky.edu

Abstract

Reductions in beef genetics Extension specialists and outreach funding in the US has led a group of beef genetics Extension faculty to coordinate their efforts and develop a national Extension program to meet beef producers and Extension educators' needs in beef genetics programming. This effort, collectively called eBEEF, has utilized four platforms to provide these outreach efforts: Beef Improvement Federation, National Beef Cattle Evaluation Consortium, eBEEF.org website and the National Cattlemen's Beef Association education program. Materials provided through eBEEF include the publication of a Sire Selection Manual, major contributions to the Beef Improvement Federation Guidelines for Uniform Beef Improvement Programs, 32 factsheets, 165 videos, over 18 train-the-trainer webinar series, 5 direct producer education webinars, and numerous in person educational programs. This collaboration has proven to be an effective model to provide beef genetics outreach programming to a national audience.

Introduction

The Cooperative Extension Service in the United States was established in 1914 by the Smith–Lever Act and one of its primary functions was to inform farmers and ranchers about current developments in agriculture (https://nifa.usda.gov/cooperative-extension-system). Extension is funded by both federal and state dollars at an amount determined by the population of the state and the level of agriculture in that state. Funding is provided to Land Grant universities in each state and that institution determines what specific areas of agriculture will have an Extension presence in their state. The structure of state Extension systems across the country varies, but in general each county or area within the state has an Extension agent who consults directly with farmers and ranchers with support from state-wide specialists with expertise in specific species/disciplines. State specialists in the area of animal production generally have a PhD, but occasionally will have a Master's degree. The decision of what specialization is needed regarding animal production in each state is typically determined by the Animal Sciences Department and College of Agriculture administration.

Over the past 30 years there has been a decline in the number of specialists in Extension and a more drastic decline in the number of geneticists hired into those roles. Currently in the US there are approximately 10 faculty members with Extension appointments in beef cattle genetics and all have other obligations in research, instruction, and/or administration. The root cause of the decrease in beef Extension geneticists is likely multifaceted including a shortage of trained quantitative geneticists (Green, 2009; Lewis *et al.*, 2014), an increased academic emphasis on nutrition and, within genetics, from quantitative to molecular genetics, a lack of interest from genetics graduates to pursue a career in Extension, and increased competition by

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_9, © K.D. Bullock *et al.* 2022

industry to hire PhD trained geneticists. In addition, funding resources to conduct Extension programming have also decreased, and increasingly those funds have been tied to federally funded integrated projects that include research, teaching and extension components, with research typically receiving the majority (i.e. 2/3) of the funding apportionment.

With the absence of a beef genetics Extension specialist in many states, and no national program to provide educational resources, a group of Extension geneticists teamed together to coordinate efforts to meet the genetics needs of beef producers nationally in an endeavour called eBEEF. This paper will outline how this group coordinated their efforts to provide educational materials and discuss the effectiveness of this effort.

Materials & methods

Currently, there are four primary outlets utilized by eBEEF to provide outreach to a national audience of beef farmers and ranchers: Beef Improvement Federation (BIF), National Beef Cattle Evaluation Consortium (NBCEC), eBEEF.org website, and the National Cattlemen's Beef Association (NCBA) education program.

Beef Improvement Federation. The Beef Improvement Federation was founded in 1968 'to connect science and industry to improve beef cattle genetics' (https://beefimprovement.org). The federation consists of state/provincial beef improvement organizations and beef cattle associations from the US and Canada. This is one of the few organizations that brings together research scientists, Extension, allied industry, and beef cattle producers to collectively pursue improvements in beef cattle selection. The outreach mechanisms of the organization include an annual research symposium and Guidelines for Uniform Beef Improvement Programs (http://guidelines.beefimprovement.org/) available in a wiki format. eBEEF is heavily involved in BIF with four members currently serving on the BIF board of directors, three serving as workgroup chairs for the annual symposium (http://www.bifconference.com), and three serving on the Guidelines for Uniform Beef Improvement Programs Drafting Committee. Consequently, eBEEF members are involved with the program development, and often serve as speakers, for the annual research symposium and have been intimately involved in the development of the guidelines and transitioning them to a wiki format.

National Beef Cattle Evaluation Consortium. Coordination of national beef genetics centric Extension programming originated through the NBCEC Educational Programs (https://www.nbcec.org) (Bullock *et al.*, 2010). Through this federally funded organization, outreach was achieved through several programs including publishing the first two editions of the NBCEC Sire Selection Manual (Bullock *et al.*, 2021), and the Brown Bagger series (2004-2021), which is an annual train-the-trainer online educational program targeted at Extension professionals, breed association representatives and industry educators (Bullock *et al.*, 2010). Additionally, industry thinktanks are hosted comprised of researchers, Extension professionals and beef producers focused on targeted traits for potential inclusion in the national cattle evaluation, and in person educational programs are held at national and regional venues. The federal funding for this organization was terminated in 2012. However, the Brown Bagger series has continued on a no cost basis and other funds were accessed to publish a third edition of the NBCEC Sire Selection Manual (Bullock *et al.* 2021). This 52-page, peer-reviewed manual was written by Extension specialists and outlines basic genetic principles and applications with a target audience of beef cattle seedstock and commercial producers.

eBEEF.org Website. The eBEEF.org website (http://ebeef.org) was initiated in 2015 as a partnership among faculty at 6 U.S. Land Grant universities with the goal of providing beef producers and beef Extension educators/agents with resources related to beef cattle genetics. This also corresponded with a time in which several large U.S. Department of Agriculture integrated grants (e.g. bovine respiratory disease, female fertility, feed efficiency) were nearing an end and a common platform for offering information from these grants to producer clientele was needed. The website contains factsheets (32), videos (165), FAQs (49),

and an e-newsletter (7). The website was initially housed within a website known as eXtension, a national extension effort predicated on the voluntary labor of Extension specialists and educators to create and deliver content through the website. The eBEEF specialists provided content through the Beef Genetics and Genomics Community of Practice. However, eXtension was disbanded in 2020, and the eBEEF website was moved to the University of California – Davis.

National Cattlemen's Beef Association. The NCBA *(*https://www.ncba.org*)* is the national trade organization for beef cattle producers in the U.S. Their contributions to the beef industry include governmental policy, improved marketing, and producer education. eBEEF is heavily involved in NCBA's producer education that pertains to beef cattle genetics. Collaborations include invited presentations at the annual Cattle Industry Convention and NCBA Trade Show's Cattlemen's College, appearances on NCBA's Cattlemen to Cattlemen television program, and presentations on national webinars hosted by NCBA.

Results

Beef Improvement Federation (BIF). The BIF Guidelines for Uniform Beef Improvement Programs officially moved to a wiki format in May 2020. During 2021, the site was viewed by 8,104 users from a worldwide audience. The US was the largest user with 5,315 views, followed by Canada, Australia, and China with 453, 198 and 188, respectively. Excluding the homepage, the pages with the greatest views were the topics of international animal identification letters (1,506 views) and body condition scores (1,445 views).

eBEEF. The numbers of visits to the eBEEF.org site was 5,953, 21,072, 29,767, and 26,612 in 2015, 2016, 2017, and 2018, respectively. Since the migration from eXtension to the site hosted by UC-Davis in 2020 the website has had 9,041 views through December 2021.

There are currently 32 factsheets posted on the website under four categories: Beef Cattle Selection, Use of DNA Information, Simple (single gene) Traits, and New Technologies. Three new factsheets were posted in 2021. As an interesting snapshot of what factsheets have been viewed, in 2017 the factsheet with the most views was 'Color patterns in crossbred beef cattle' with 13,675 views; the second was 'The genetics of horned, polled and scurred cattle' with 8,147; while the third was 'How to get started with DNA testing' with 2,371. These data indicate an appetite for information on simple qualitative traits, and a need for educational material outlying fundamental basics of genetics for those with little familiarity with beef cattle genetics and breeding. The number of views by video was highly variable ranging from over 1000 to as few as 4. The videos range from technical content and long (over 1 hour), to short, concise pieces of 'how to' information, which likely impacts viewership.

Much of the information found on the website of partner organizations (NCBEC, BIF and NCBA) is crosslinked on eBEEF.org and vice versa. For example, the NBCEC Sire Selection Manual is available on both the NBCEC website and eBEEF.org. In addition to distributing the 3rd edition of this manual as a bound paperback book at the 2021 BIF meeting, the manual was made available as a pdf on the websites in April 2021, and as of January 2022 it had been downloaded 774 times between the eBEEF.org and NBCEC.org sites and approximately 1000 hardcopies have been distributed.

National Cattlemen's Beef Association (NCBA). In 2018, the eBEEF team presented a webinar series hosted by NCBA Producer Education that consisted of four sessions: (1) Fake News: EPDs Don't Work; (2) Show me the Money! Are there EPDs for Profit?; (3) The Four S's of Crossbreeding: Simple, Structured, Successful, and Sustainable; and (4) Putting the Tools to Use: Buying Your Next Bull. Over 650 producers participated live and YouTube recordings received over 3,300 views. Survey results rated relevance to beef

producers between 4.3 and 4.5 on a 5-point scale. Participation in eBEEF live presentations at NCBA Cattlemen's Colleges are approximately 800.

Discussion

During a time of reduced funding and a dwindling faculty base of Extension personnel with training in genetics, a team of Extension beef cattle geneticists organized under the name eBEEF provide educational resources to beef producers and Extension educators across the US. The mission of eBEEF is 'to foster a research and outreach community, engage beef cattle producers and ag professionals through trainings and publications, and support research and outreach projects' (https://ebeef.ucdavis.edu/about).

Collaborative efforts of eBEEF resulted in the development of educational materials that were made available across a variety of platforms to a national audience. Moreover, the eBEEF model reduced duplication of effort by the limited number of Extension personnel with genetics training, increased consistency of messaging, and consolidated producer educational material into a single location. Through electronic and limited print publications, train-the-trainer webinars, and direct producer education both virtually and at national venues, eBEEF has established a highly visible, impactful, and cost effective national educational program on beef cattle genetics. Although there have been successes demonstrated through this low-cost model of outreach, avenues for sustained support are needed. As important are mechanisms to ensure that junior faculty receive academic credit for such collaborations, and that well-trained genetics faculty can be recruited into Extension roles. Extension funding sources and personnel to deliver outreach programs in the future remain two primary areas of concern.

References

Bullock K.D., Lewis R.M., Weaber R.L, Lockee B.B, Strohbehn D.R. *et al.* (2010) Proc. of the 9th WCGALP, Leipzig, Germany.

Bullock K.D., Spangler M.L., Van Eenenaam A.L., Weaber R.L., Rolf M.M., *et al.* (2021) NBCEC Sire Selection Manual Available at: https://ebeef.ucdavis.edu/

Green RD (2009) J. Anim. Sci. 87(2):793-800.

Lewis R.M., Spangler M.L., Lockee B.B., Enns R.M., Enns K.J., *et al.* (2014) Proc. of the 10th WCGALP, Vancouver, Canada.

10. Breeding Lab: a successful industry-academia collaboration to provide MSc consultancy training

J. Komen[1*], J.A.M. van Arendonk[2] and D.E. Lont[1]

[1]Wageningen University & Research, P.O. Box 338, 6700 AH Wageningen, the Netherlands; [2]Hendrix Genetics, Spoorstraat 69, 5831 CK Boxmeer, the Netherlands; hans.komen@wur.nl

Abstract

A new course Breeding Lab was developed and introduced, to better prepare students for the diverse job market. The course was set-up by Wageningen University in collaboration with industrial partner Hendrix Genetics. Both organisations played a key role in the course. The academic partner focused on the development and application of skills and competences necessary for solving real life issues in animal breeding and genetics. The industrial partner was responsible for formulating the real-life issue to be addressed, technical supervision of the students and evaluation. The two partners involved in the course and the students found the course as a valuable addition to the curriculum.

Introduction

An important task and challenge of academia is to prepare students for the future job market. Results of the yearly Dutch National Student Questionnaire consistently show students' wish for better preparation for the diverse job market. To respond to this, we recently developed and introduced the course 'Breeding Lab' in close collaboration between industry and academia. This course was offered for first time in 2019-2020 to students of the European Master in Animal Breeding and Genetics (EMABG). The enrolled MSc students worked on a group assignment in an industry setting, allowing them to learn and apply competences relevant for solving real life issues that were difficult to develop on campus.

In this paper, we will explain the set-up and learning outcomes of the breeding lab, and the different educational formats used to achieve the learning outcomes. Subsequently, we will focus on the experiences at both sides: the industry partner, Research & Development team of Hendrix Genetics, and the academic partner, the Animal Breeding and Genomics group of Wageningen University. In the discussion we focus on the experiences of students and give an outlook on expanding the course to accommodate larger numbers of students.

Materials & methods

Breeding Lab: set-up and learning outcomes. The idea of the Breeding Lab was developed over time, following consultation of several R&D directors who expressed the need to have students better prepared for their professional careers in industry, in terms of professional skills and competences. At the same time, results of the Dutch National Student Questionnaire (NSE) showed that students also felt the need to be better prepared for the diversity in the future job market. Combining the two, led to the development of the pilot project 'Breeding Lab' by the Animal Breeding and Genomics group of Wageningen University in 2018. This course was included as an academic consultancy course in the curriculum for the EMABG. Wageningen University is one of the six partners in the EMABG consortium, and one of the two to provide education to first-year students within that master program.

The aim of Breeding Lab is to prepare students for a professional career in industry or (non)governmental organizations. This is achieved by letting students work on an assignment, provided by the industrial partner. The emphasis in the assignment lies on solving a problem in teams and experiencing, learning, and acquiring the professional competences that are needed to successfully operate in an (inter)national,

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_10, © J. Komen *et al.* 2022

multidisciplinary team. An important part of the course is devoted to understanding, describing, and reflecting on the competences that are needed to successfully operate in a professional environment:

- sensitive to management structures, politics and governance;
- personal communication skills;
- collaboration and personal effectiveness when working in groups;
- dealing with setbacks, disappointments, and feedback.

The course is scheduled as a combination of a 'group internship' at Hendrix Genetics, and training and exercises related to professional skills. In a period of 4 weeks, groups of students work full-time on an animal genetics-related assignment. This assignment fulfills the following criteria: it should be a commissioned multidisciplinary assignment, concern an authentic case, offer a certain level of complexity, reflect various interests and stakeholders, and it should require an academic (master) level to address the issue.

The learning outcomes: after successful completion of the course students will be able to:

1. Contribute at an academic level to the execution of a multidisciplinary project, both in terms of process and content, related to the discipline of Animal breeding, by gathering, selecting, and analysing information and integrating this into a proposed solution to the assignment given by the commissioner.
2. Report, discuss and defend viewpoints and conclusions in a professional, academic and a collaborative manner before representatives of the commissioner and other stakeholders involved (demonstrate the ability to engage in perspective-taking).
3. Reflect on their personal functioning in and contribution to a professional team.
4. Demonstrate academic competences to execute the team assignment, within the complex (inter) national environment of a professional organisation.
5. Reflect on the work culture of the organisation and its decision-making process.
6. Further analyse the competences that are needed to be successful in a professional environment, as encountered during the assignment, and reflect on these in a self-evaluation.

From an educational point of view, the Breeding Lab can be seen as a 'pressure cooker', due to its short time frame of 4 weeks. It requires students to apply the 'learning by doing' mentality and to have their hands on the assignment from the very first moment. It also requires collaboration and assignment of specific tasks within the teams, and knowing about and respecting each other's strengths and weaknesses.

Professional skills training modules are part of the course, and are given by professionals from the 'Education and Learning Studies' chair group of Wageningen University. The first of the three skills training modules, mostly focusing on intercultural awareness, is planned before the students start the 4-week period. The other two skills training modules are given along the way, and focus on team roles (training 2) and conflict resolution skills (training 3).

Intercultural and collaboration skills are an important part of the course. The target group for the Breeding Lab in its current form, are the first-year MSc students of the European Master in Animal Breeding and Genetics. Given the nature of that programme, this is a highly diverse group of students, representing many different nationalities (Table 1). Hence, intercultural skills, as well as communication and collaboration skills, are an important aspect of the competences to be developed and applied in the course.

Each student team of 4-5 students is assigned to an experienced R&D member from Hendrix Genetics. These experts are indispensable in supervising the content of the assignment. This entails clarifying the content of the assignment, getting the right (company) information at hand, and making sure the students keep heading in the right direction leading to a recommendation or outcome that is useful for the company.

At the same time, the teams get coaching from experienced teachers from Wageningen University, focusing on the group process, personal development of team members and reflection on roles and tasks within the teams.

Results

The course started in academic year 2019-2020 and has run two times, in two consecutive years. The third offering is taking place at the time of writing (January 2022). The composition of the groups in terms of nationalities is given in Table 1.

Experiences from the viewpoint of industry-partner Hendrix Genetics. As Hendrix Genetics, we recognize the importance of introducing students during their training to practices in the industry. That is why we, like colleagues from other breeding organisations, provide guest lectures in course at the university and do offer internships for students. We are very pleased with this new course which gives us the opportunity to introduce a larger group of students to the way of working in a company. The interaction with the students and the results of the assignment help us in finding new or better solutions. Not only the students but also the geneticists involved in the supervision benefit from the interactions during the Breeding Lab.

Experiences from the perspective of Wageningen University. We started the course in 2019-2020 from scratch, with two groups of motivated and very diverse students and in good collaboration with Hendrix Genetics. The course has taught us once again how important students' personal skills are, for a student team to be successful. One of our tasks as coaches is to make students aware of the difference between knowledge-based skills and the personal skills required to be successful in a team. Coaching students on this type of behaviour is sometimes challenging but is also very rewarding. Our goal is to establish an open atmosphere in which students can share their successes as well as their doubts. An important aspect of the Breeding Lab is its international composition. Students are generally unaware of the cultural differences that can lead to different attitudes and types of behaviour (indirect vs direct communication, time management, etc.) during teamwork. At worst, they can lead to counterproductive conflicts and group failure (Meyer, 2010). The first skills training course is therefore fully dedicated to cultural differences and creating understanding. The students, (especially Dutch students), experience this training in many cases as an 'eye opener'.

Table 1. International classroom in Breeding Lab, academic years 2020-2021, 2021-2022, and 2022-2023. Table shows an overview of students' nationalities per edition.

	Year		
Continent	**2019-2020**	**2020-2021**	**2021-2022**
Europe	Dutch (3)	Dutch (1)	Dutch (4)
	Austrian (1)	Irish (1)	German (1)
		Hungarian (1)	
Africa	Nigerian (1)	Ethiopian (1)	Ethiopian (1)
		Ghanaian (1)	
Asia	Bhutanese (1)	Chinese (1)	Mongolian (1)
	Bangladeshi (1)		Indian (1)
			Bhutanese (1)
Latin America	Bolivian (1)	Paraguayan (1)	
		Brazilian (1)	
Total number of students	8	8	9

Due to the pressure cooker nature of the course, we observe quite steep learning curves with our students. We therefore believe that the course is a valuable addition to the MSc programme.

Discussion

We asked for extensive feedback from students of both completed editions 2019-2020 and 2020-2021. Students indicated that they enjoyed a lot working on an assignment with a real company and taking a look into the corporate life of animal breeding. They particularly liked the skills training sessions. The intercultural fingerprint as well as conflict management skills training were equally preferred to each other by the students. This was due to simply interest and experiencing conflicts within the team, respectively. The students generally did experience good group feeling and considered the group size of 4-5 students as adequate in relation to the project. However, some students felt tension within the group and were convinced the skills training, especially regarding conflict management, could have helped them more with such issues.

Quotes from students: 'Great to have complete freedom to determine your own schedule and work on the assignment as a team. A lot of responsibility, a steep learning curve.' and 'I liked the set-up of the course where we had to solve a problem for a company. This was unlike my previous courses, which consisted of lectures and exams.'

One of the objectives of the Breeding Lab was to prepare students better for the job market, e.g. working in a breeding company. The assignments of the Breeding Lab appear to do just that: students work on real life problems, and meet and talk with R&D employees on how they perceive working in a breeding company. Students are generally surprised to find that working in an R&D environment can be equally challenging intellectually, as working in academia (e.g. as PhD student). Many EMABG students continue by doing an internship with Hendrix-Genetics.

As an outlook on future, the Breeding Lab is a clear success as non-EMABG students are also increasingly registering for this course. Students are attracted by the learning-by-doing approach and the high level of independence. The fact that the problems are real life problems that need a solution, rather than pre-cooked exercises makes all the difference for them. The set-up of the course can easily accommodate larger numbers of students. A potential bottleneck is the time that industrial partners can devote to assist/support the students. The current period of 4 weeks might be too short for the partner to get value out of the project.

References

Meyer E. (2016) The culture map: Decoding how people think, lead, and get things done across cultures. PublicAffairs, New York, US.

11. How a group of Merino breeders increased the rate of genetic gain by 134% in five years!

T. Granleese[1*], J.H.J. van der Werf[2] and S.J. Martin[3]

[1]NSW Department of Primary Industries, Grafton NSW 2460, Australia; [2]University of New England, Armidale NSW, 2351, Australia; [3]SheepMetriX Pty Ltd; P.O. Box 1533, Young NSW, 2594, Australia; tom.granleese@dpi.nsw.gov.au

Abstract

The DNA Stimulation Project was a major genetics adoption program in Merino sheep breeding enterprises running from January 2018 to June 2022. Historically, the Australian Merino population have had low rates of genetic gain. This project used a collaborative approach between Meat and Livestock Australia's Donor Company, University of New England, MerinoLink Limited, Merino breeders, scientists, and genetic service providers. A bold project target was set to double the rate of genetic gain by 2022 for the group of 30 breeders who collectively invested over AUD$1.7 million in total and was co-matched by Meat and Livestock Donor Company. The project focussed on working with the project participants to strategically use the genetic and genomic tools currently available. With the ambitious goal being reached a year early (increase of 134% in lifetime of project), the key component to this success was the willingness of project participants to engage and the coaching approach by the project team.

Introduction

On average genetic gain in Australian Merinos has been historically low with very large variations across the industry (Granleese *et al.* 2018). Underlying influences have been caused by a multitude of factors including inaccurate breeding values (Stephen *et al.* 2018) and/or a lack of understanding of selection theory.

The DNA Stimulation Project aimed to build understanding and implementation of genetic selection tools with the project participants. The project focussed on capacity building and working collaboratively at all levels (scientists, genetic service providers, and breeders) across the industry to communicate how to use software tools more effectively for assisting in the design of breeding programs to increase the rate of genetic gain for participants.

During the DNA Stimulation Project, project participants have been faced with significant challenges including unprecedented drought in many parts of Australia and restrictions due to COVID-19. The project team were able to adapt to the challenges and continued to deliver the project requirements.

The aim of the DNA Stimulation Project was to double the annual rate of genetic gain by 2022 (starting 2018) in line with Meat and Livestock Australia's National Livestock Genetic Consortiums aims. This paper outlines methods the project used to almost reach the target of doubling the rate of genetic gain.

Materials & methods

Merino breeders were notified of a project call that would subsidise genomic testing for their seedstock flocks. Given Merino breeders have historically low pedigree recording, this was a vehicle to fast track pedigree collection. The project expectations agreed that in exchange for subsidised genotyping (funding is 58% from the project participants and 42% from the Meat and Livestock Donor Company) that they would form a network to upskill via a series of intense coaching and workshops. Scientists were funded to link to private consulting genetic service providers, the national sheep genetic evaluation group – Sheep Genetics

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_11, © T. Granleese *et al.* 2022

and the Merino breeders. The aim was to create links between all levels of data collection, Australian Sheep Breeding Values (ASBVs) delivery and research and development.

The Project had 30 ram breeders (accounting for 40% of the annual Merino Sheep Genetics database intake of animals) and 15 participating commercial breeders who breed their own rams, 43 commercial breeders who purchased rams and seven private consulting genetic service providers. The project participants were located across New South Wales, Western Australia and Victoria. In total this group represented over 10,000 rams sold annually (contributing to >2.5 million sheep born annually), 15,000 doses of semen sold annually to seedstock and commercial breeders, as well as over one million sheep bred directly by the project participants.

The extension and adoption process include a combination of face-to-face workshops, intensive one-on-one meetings, webinars, phone calls, personal emails, e-newsletters and group email updates. Input into these processes include key personnel within the DNA Project Team, University of New England, Sheep Genetics, Meat and Livestock Australia and participants in the project, including breeders and genetics service providers.

Key tools used with project participants include DNA parentage to capture pedigree, low density genomic tests for pedigree collection and/or genomic selection, ASBVs for selection, MateSel (Kinghorn, 2011) to optimise mating decisions in large breeding programs, Sheep Genetics Ramping Up Genetic Gain report (RUGG) to identify data collection problems, Flock Profile – a genomic only estimation of flock average (Swan *et al.* 2018) for commercial breeders, Ram Team Manager (RamSelect.com.au) for commercial ram buyers to improve their ram teams each year for their breeding objective, Rampower within flock indexes for commercial breeder collecting phenotypes but not pedigree and percentile band tables for context of breeding values.

Each project participant worked through the process of setting and documenting an achievable breeding objective which is revisited each workshop or meeting with the DNA project team. Key questions covered were: – are they on track, what is working well (or not), what can be improved, and how are project participants measuring progress?

To benchmark progress the DNA Stimulation Project used rates of genetic gain were ASBVs generated by Sheep Genetics. The project also used direct workshop feedback to track the effectiveness of the project.

Genetic improvement was benchmarked using a standard breeding objective using Sheep Genetic's Merino Production Plus (MP+) economic index. 'Doubling the rate of genetic gain' was measured by comparing the five-year genetic gain average (2012-2017) compared to the project four-year genetic gain average (2017-2021).

Results

Genetic gain increased by 134% in ram breeders comparing the four-year period of the project to the previous five years prior to the project. Another of the key outcomes of the project with the strategic use of DNA testing technologies was the increased number of ram breeder participants submitting data to Sheep Genetics with full pedigree information. There was an increase in the number of project participants submitting full pedigree from 25 to 70% between January 2018 and January 2022.

At the beginning of the project there were large differences in average genetic merit between the ram breeder participants in the DNA Project, with the lowest average MP+ index 110 (90th percentile) and the highest 167 MP+ index points (1st percentile) at the commencement of the project in 2017. The current MP+ index range between the ram breeders is 138 (70th percentile) to 188 (1st percentile) index points. The diversity in the group facilitated sharing and swapping experiences to allow informal mentoring within the group.

Discussion

The key component to the success of this educational project was providing links between breeders, private genetic service providers and scientists. The interaction between all parties allowed flow of information benefiting all parties and breaking down communication and trust barriers. Furthermore, engaging private genetic service providers who had clients within the project gave extra incentive for them to up-skill. The up skilling of these professionals' results in the multiplier effect with more educated messages and advice being passed on to many of their clients.

Trust within the project was initially built on 'across-the-kitchen table' meetings with ram breeders where they expressed their frustrations with genetic evaluation and barriers that stopped them collecting crucial data for higher quality ASBVs. With this knowledge the team scientist and service provider were able to provide direct feedback as to why not collecting data points was affecting their ASBVs and practical solutions to their problems. Once participants were listened to in a 'safe' and private space, their willingness to change and adopt occurred quickly.

Access to the same genetic service provider as well as scientist at any time in any medium (i.e. email, phone call, text message, Zoom meetings) was another key step in helping breeders upskill and adopt better practice. Further group educational activities held were annual ram breeder workshops where a range of scientific developments, direct feedback, setting S.M.A.R.T. breeding objectives and peer-to-peer learning through social networking and project case studies. Before and after each workshop the project team would get breeders to fill in questionnaires rating their understanding of how ASBVs are calculated and how much they use ASBVs in selection decisions. Prior to the project starting the group average responses and post project average responses are summarised in Table 1. Further into the project as breeders built a depth a pedigree in their flocks they were in a position to use Matesel (Kinghorn 2011) where matings are optimised to maximise genetic and limit future co-ancestry. A series of Matesel workshops were run in the last three years of the project. Either breeders ran their own Matesel at joining or employed their genetics service provider to do this for them which helped accelerate rates of genetic gain given Matesel uses index selection.

The project observed a direct relationship between attending educational activities and rates of genetic gain. Figure 1 demonstrates that the more engaged ram breeder project participants were with the project the higher the rates of genetic gain. This trend validates the intense relationship the breeders have with the

Table 1. Ongoing mean scores at set times throughout project of breeders' knowledge and use of ASBVs throughout the project.

	Understanding of ASBVs	Use of ASBVs in selection decisions
1-Jan-2018	3.6	6.4
1-Jan-2019	6.8	7.5
1-Jan-2020	7.4	8.4
1-Jan-2021	8.3	8.7
1-Jan-2022	8.5	8.9

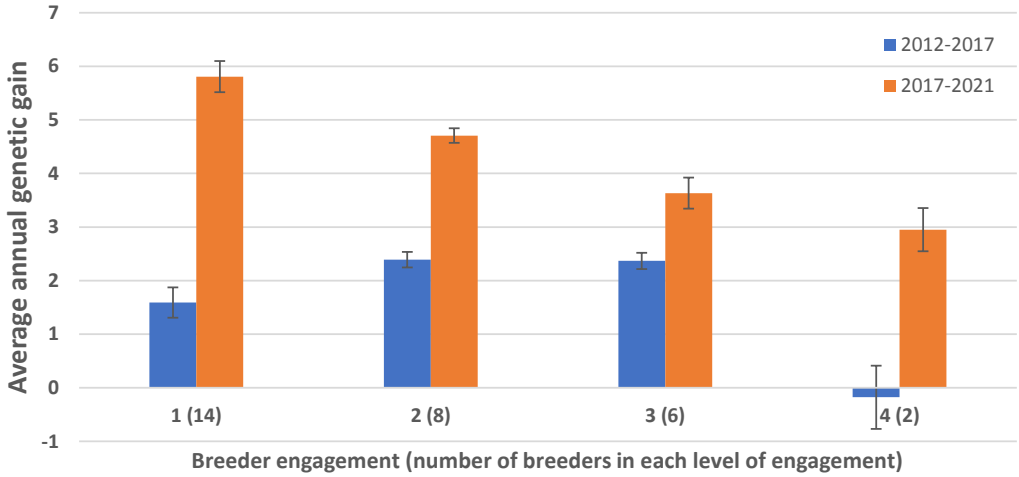


Figure 1. Blue bars represent five-year average annual genetic gain (2012-2017) pre-project and orange bars represent three-year average annual genetic gain (2017-2021) during project for studs grouped by project engagement. 1 – Fully engaged: attends all workshops, calls project leaders, asks for one-on-one meetings; 2 – Engaged: Attends most workshops and/or some extra meetings; 3 – Sometimes engaged: Sporadic attendance to workshops, rarely uses extra opportunities; 4 – Never engaged: Do not turn up to any workshops or extra activities.

project team. A key to keeping breeders engaged was using a co-investment model. Initially the project used the motivation of subsidised genomic testing to 'lure' breeders into the project. With Meat and Livestock Australia offering dollar-for-dollar investment with project partners, the project lead was able to 'skim' small amounts of money off each subsidised test to help fund a genetics service provider's and scientist's time for the life of the project. Furthermore, breeders agreed to pay an annual fee of $2,000 AUD/yr to: (1) help fund workshops and time of staff on the project; and (2) maintain interest and investment in the project after the initial benefit of subsidised genotyping.

Summing up, the DNA Stimulation project was a success in increasing the rate of genetic gain in Merinos. Furthermore, the project had long-lasting implications with the up skilling of genetic service providers and breeders to continue the pursuit of improving breeding objectives and better facilitating clients.

References

Granleese T., Clark S.A., Duijvesteijn N., Bradley P.E., and Van der Werf J.H.J. (2018) An. Prod. Sc. https://doi.org/10.1071/AN17720.

Kinghorn B.P. (2011) GSE. https://doi.org/10.1186/1297-9686-43-4

Stephen, L., Brown D., Jones C., and Collinson C. (2018). Proc. of the 10th WCGALP, Auckland, New Zealand.

Swan A.A., Gurman P.M., Boerner V., Brown D.J., Clark S.A., *et al.* (2018) Proc. of the 10th WCGALP, Auckland, New Zealand.

12. YouTube applied to the education of genetics and genomics

G. Mészáros

University of Natural Resources and Life Sciences, Vienna, Gregor-Mendel-Straße 33, 1180 Vienna, Austria;
gabor.meszaros@boku.ac.at

Abstract
Educational videos are a form of asynchronous learning, with their own set of advantages and disadvantages. They are available at any time for students to watch and learn. On the other hand, practical experiences show that the entire duration of videos are not watched. This paper gives practical insight on viewer habits for educational videos, using the 'Genomics Boot Camp' YouTube channel, as an example. Moreover, the aspects of educational video placements on commercial platforms, and their potential innovative use for standard research project dissemination activities are discussed.

Introduction
It is not easy to write this paper. You, as a reader, might say that I write it only to promote my personal brand on a world-wide stage, in front of the most relevant audience there is, for an educational resource focused on genetics and genomics. Certainly, any promotion activities are very useful in this line of work, as you will see in this paper. What you will also see, however, is that in the business of making videos an even more important part is providing a value to people who decide to invest their time into your work.

This paper is to provide my experiences from managing the 'Genomics Boot Camp' YouTube channel, that focuses on teaching theoretical and practical aspects of genomics. I will compare my experiences to the typical classroom setting in face-to-face and online forms, discuss various formats and opportunities such a platform provides, including reach to world-wide audiences across disciplines and the general public. I will also outline how educational videos could be used for classical project dissemination and communication efforts. At the very end of the paper you will find another experimental approach that might be utilized in science communication.

Materials & methods
The paper was based on personal experiences from almost one year of publishing educational video on the YouTube platform. The first public video on the Genomics Boot Camp channel was released on 24 January 2021.

The statistics on viewer counts and engagement were taken from the YouTube analytics application programming interface (API), which retrieves YouTube Analytics data for the channel. The view statistics and visualizations of audience retention were taken from the YouTube Studio. The retention rate was defined as the percentage of viewers still watching at any given time point. The shaded areas in the figures were the average retention rates computed from the most recent videos at the time. At the time of writing of this paper, the channel had 1,700 subscribers, 81 public videos, and 64,600 lifetime views. In the last 90 days there were 30,000 views from 11,800 unique viewers. Overall, numbers related to viewership and audience participation indicate increasing engagement, but the unique viewers are available only for the last 90 days, due to limitations in the YouTube Studio.

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_12, © G. Mészáros 2022

Results & discussion

My initial motivation to start this channel was my experience that people need a simple, down to earth explanation on how to deal with single nucleotide polymorphism (SNP) data. This apparent need, at the time, was coming from people involved in livestock research. I kept getting very similar questions and requests from students eager to do their MSc thesis, but also more senior colleagues, who were not closely working in practical genomics before. Soon after launching the channel the audience has broadened, with people from human and plant genetics, genomics, and other fields of biology joining in. Also, the audience broadened from my students and colleagues I knew to the worldwide level.

Videos in teaching and the viewing behaviour. The video lectures and teaching materials are a form of asynchronous learning, which allow learning in different times and locations. Viewers could ask questions, which are regularly answered via the comment section of each video, which are also visible to all future viewers. Such resources were not meant as replacement of traditional learning, but to provide additional resources that students might seek, in form of clarifications or practical tutorials. Similarly, Noetel *et al.* (2021) found that adding video to existing teaching led to strong learning benefits.

The student viewing behaviour of videos is very different from a synchronous, and especially face-to-face classroom setting, where the students sit through entire lectures, typically in length of 1-1.5 hours or more. In asynchronous setting the view durations are shorter. Black (2020) found average view durations between 14 and 25 minutes, for lectures lasting from 46 minutes to 1.5 hour, with relatively longer views for longer videos. There was no optimal length of videos suggested, but generally shorter videos are suggested over longer ones. Still, the creation of, for example, 10-minute videos does not guarantee that students watch it all. In my experience, the average view duration is similar in proportions, regardless of the video length. Using the two examples shown below, the view duration was 2:42 for the video with a below average retention rate (Figure 1, total length: 10:30), and 5:36 for the video with above average retention rate (Figure 2, total length: 10:38). Similar patterns are visible throughout all videos on the channel. Here I must note that average view times were likely influenced by two additional factors. The first one is the large number of users that click away after few seconds in the video, thus greatly decreasing the overall average view duration. The likely reason was that they quickly found that the content of the video was not what they expected and moved on. The second factor could be that the videos are not part of any formal course, thus it is not mandatory for the students to watch them in full. If they had a specific question that was answered within the video, they will likely click away.

Another aspect of learning was the attention span of students during lectures. It is very challenging to measure this behaviour in a classroom setting. Early sources (Hartley *et al.*, 1978) mention an attention span of 10-15 minutes, but Bradbury (2017) contradicts this measure, and suggests waxing and waning attention throughout the lecture, rather than a fixed time period.

Similar behaviour was noticeable, and measurable on the videos related to the retention rate (Figure 1 and 2). In both videos there is a noticeable sharp decline right at the beginning, suggesting viewers clicking away. The exact reason is unknown, possibly due to unintended clicks on the video, or the viewers quickly finding out the content is not what they expect. The retention could be improved by certain techniques, some of which were applied in the video with better retention. Examples of these techniques include a setup of video chapters, which see slight increases in viewership (dashed lines on Figure 2). Regardless of the perceived quality of the video, however, the retention rate seems to decline over time, so viewers do not watch the entire video, from beginning to the end, as one might expect. Similar decline was also shown by Lau *et al.* (2018), which suggest this was a universally applicable viewer behaviour.

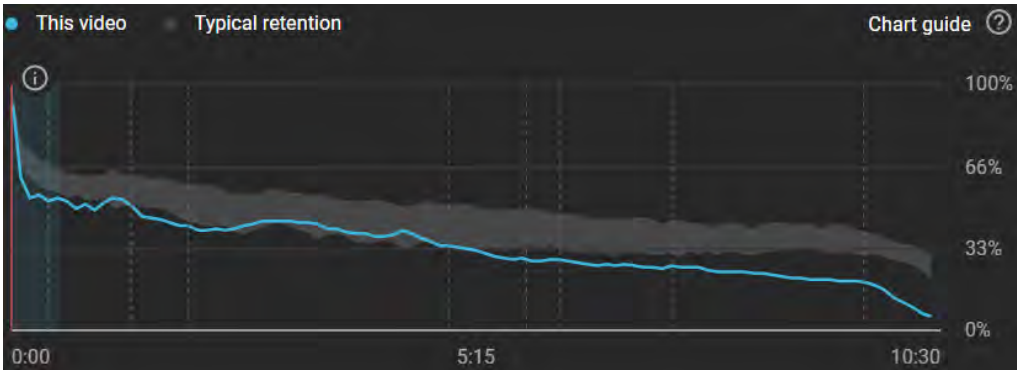


Figure 1. Video with a below average retention rate. Vertical lines indicate video chapters. Views: 6265, Title: 'Genomics in practice – How to start PLINK'.

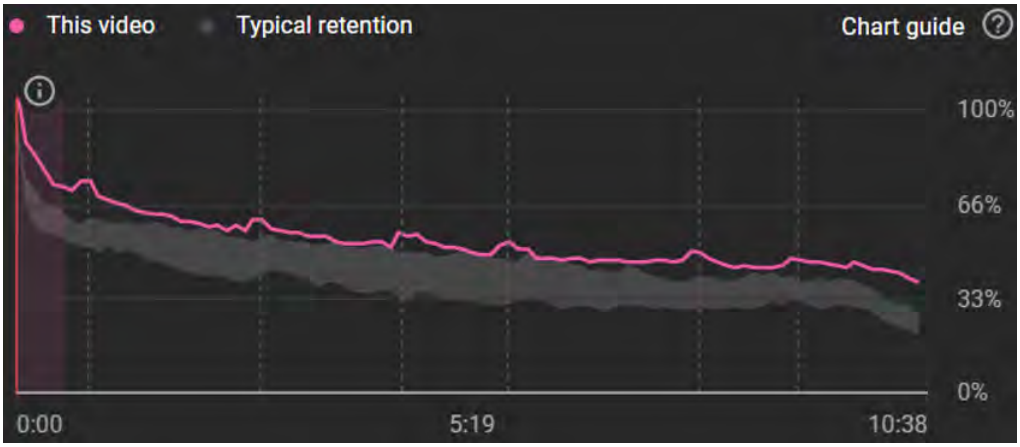


Figure 2. Video with an above average retention rate. Vertical lines indicate video chapters. Views: 274, Title: 'Get organized with your data on your computer'.

Practical considerations. A distinct advantage of video materials on commercial platforms like YouTube is that the setup and maintenance of such platform do not require investment from lecturer's side. YouTube also actively promotes the videos to potentially interested new viewers via so-called 'impressions', i.e. the number of times the video thumbnail was shown to users. From these impressions, only a small percentage actually click (so called 'click through rate'). At the time of writing the monthly impressions are over 120K (ca. 4K daily), which at overall 5% click through rate meant 5.5K views, solely from the results of the algorithm. This is not a lot by any means: large channels have their daily impressions in millions! Still, even these lower numbers are only accumulating, and consistently growing over time in an accelerating manner. According to my experiences related to reaching audiences, this is a distinct advantage of YouTube, compared to other websites. Static websites have stable content, but they need to be discovered, which requires constant and consistent efforts. Other social websites, like Twitter could drive considerable audiences using a similar algorithm, but the promotion of any single message (tweet) is only for a short duration (a few days maximum). Opposing to this, on YouTube a well-done video is continuously suggested for years.

Unconventional use of video lectures. Another way such educational videos could be used is for dissemination activities of standard research project. In my opinion this is currently a largely under-utilized opportunity, with distinct advantages for both the research projects (to reach more people) and the viewers (to learn from experts in the field). Conventionally, dissemination activities rely on sharing results and outcomes of research projects to different audiences, ranging from general public to the wider scientific community. It is perhaps only my impression, but most of the time such reporting goes something like 'We from Project X did a lot of work in Y, so we advanced the science.' This is great. It is shared and reposted mostly by people who were involved, perhaps their colleagues, institutions, and maybe a several other interested people. But I think it could be improved, incorporating the message in an educational video with: 'This is how you do Y. Here are some open access data to do it. And by the way, we are from Project X.' While the first approach provides the information to the community, the second one provides actual value, and a good reason why viewers should engage and share the material, even if they don't really care about Project X. If the topic is of interest, novel or even a very often used standard procedure, wide scale feedback is very likely. Case in point: One of the most viewed videos on the Genomics Boot Camp channel is a tutorial how to do a PCA plot. In all projects, there is a great amount of expertise that could be easily utilized for similar purpose, creating lasting footprints of projects and brands, in the form of videos that people actively search for.

Finally, the promised experimental approach, that might be utilized in science communication, or for other similar purposes. Parallel with the submission of this paper I made an accompanying video with additional commentary. The inspiration for this approach came from the annotated version of Watson and Crick (1953) on the Exploratorium website. Here the annotation gives an entirely new dimension to the paper, providing historical context and supporting information that greatly enhance the pleasure of reading. Similarly, in the case of this paper, you could gain some accompanying information by searching for the 'YouTube applied to the education of genetics and genomics' video. The premiere of this video was set to the beginning of the WCGALP 2022 conference. In general, such annotations or commentaries in audio-visual or written forms could provide context for various audiences. In education, setting the scene, and briefly explaining the relevance of published materials, might similarly improve their understanding by the readers, as opposed to a 'Read this' directive.

For the future a video series on genomics theory basics will be completed, complementing the existing practical tutorial. The overall plan is a continued weekly release of videos that are attractive to learners.

References

Black M. (2020) Appl. Econ. Teach. Res. 2(5):18-25. https://doi.org/10.22004/ag.econ.308057

Bradbury N. (2017) Acad. Med. 92(4):428. https://doi.org//10.1097/ACM.0000000000001584

Exploratorium, Annotated version of Watson and Crick paper. Available at: https://www.exploratorium.edu/origins/coldspring/printit.html

Hartley J., Davies I.K. (1978) Program. Learn. Educ. T. 15:207–224. https://doi.org/10.1080/0033039780150305

Lau, K.H.V., P. Farooque, G. Leydon, M.L. Schwartz, R.M. Sadler, *et al.* (2018) Med. Teach. 40:91–98. https://doi.org/10.1080/0142159X.2017.1395001

Noetel, M., Griffith, S., Delaney, O., Sanders, T., Parker, P. *et al.* (2021). Rev. Educ. Res. 91(2):204–236. https://doi.org/10.3102/0034654321990713

13. Creating a good learning and sharing environment for bioinformatics

T. Klingström* and J.I. Ohlsson

Department of Animal Breeding and Genetics, Swedish University of Agricultural Sciences, Box 7023, 750 07 Uppsala, Sweden; tomas.klingstrom@slu.se

Abstract

Most bioinformatics researchers create their own working environments depending on their training, access to computing resources and what advice they receive from other researchers. This makes the development of training programs difficult as much knowledge and work is locked into each researcher's unique computing environment for data analysis.

Using software container solutions such as Biocontainers and a 'one image per task' coding pattern when developing a workflow, it is easy for an application expert to rewrite workflows for research, training and dissemination of new workflows for analysing data. This paper describes the reasoning behind this approach which is now being tested in a collaboration between livestock researchers in Sweden and South Africa.

Introduction

Advancements in high-throughput technology for genotyping and phenotyping mean that data acquisition, storage, analysis and dissemination is an increasingly large part of the work in agricultural research. In a survey at the Swedish University of Agricultural Sciences, PhD students spend 25% of their time dealing with challenges caused by a lack of training or tools necessary to perform data analysis (Klingström, 2017). Developing and providing the tools and training necessary to work efficiently with data is therefore an important task for the university to improve research efficiency. In the Life Sciences, specialisation in this area falls under the designation of bioinformatics which is a research discipline joining together biology and computer science. Depending on the context the subject of bioinformatics can be viewed from multiple perspectives:

- As a tool supporting the development of knowledge in other fields of the Life Sciences.
- As an independent research subject pushing the boundaries of human knowledge of its own accord.
- As an interface between biology and computer science providing biological inquiries access to high-powered computing resources.

The different perspectives make it difficult to disseminate the wide array of tools available for bioinformatics. Tools for researchers specialising in bioinformatics often prioritise flexibility and control while researchers who only occasionally use bioinformatics require tools prioritising ease of use and high quality training materials. This creates difficulties in sharing bioinformatics workflows between research groups. In the B3Africa project (Klingström *et al.*, 2017), we worked on how to enable interdisciplinary collaborations incorporating all three perspectives to enable researchers to better disseminate bioinformatics tools and apply them to available datasets. One of the key conclusions of the project was that Galaxy (Jalili *et al.*, 2020) is a versatile solution which helps us break down the matter of bioinformatics into three distinct components:

- A single user interface which can be used by researchers with little training to perform bioinformatics research using virtually any open-source bioinformatics tool available.
- A data management platform handling data access, metadata and how data can be accessed from local storage, public repositories or cloud storage.
- Providing users with access to a suitable computing environment with the bioinformatics software necessary to perform the desired data analysis.

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_13, © T. Klingström and J.I. Ohlsson 2022

A weakness with Galaxy is that the graphical user interface does not provide the flexibility and control desired by most bioinformaticians. This means that any Galaxy server requires dedicated staff scientists who provide training and manage the server to install new tools on behalf of the user. Thus, the efficiency gains from using networked Galaxy resources are reduced or lost as tools are configured for every specialized use, creating an environment with barriers to communication and productivity. This can be avoided by using containerized software in a computing environment shared between the Galaxy server and bioinformaticians using their own preferred workflow management software.

Software containers derive their name from the intermodal shipping container, which is nowadays a ubiquitous component of the international shipping system (Figure 1). Instead of requiring a specifically designed vehicle for each kind of goods transported, vehicles of today are often designed to carry a standardized shipping container. Likewise, software containers ensure that any containerized software has the environment and configuration necessary to function the same on different servers. With modern computers and software, pretty much any laboratory with bioinformaticians can host a data management platform and the software containers to analyse small datasets while access to computing power, data transfer capacity and working memory (RAM) limit the size of datasets which can be processed within a reasonable timeframe. Using software containers, it is thereby possible to fully test a workflow on a subset of data before applying for funding or access to computing power in a shared e-infrastructure for computing, and using the same workflow to analyze larger datasets.

There are a limited number of researchers available for training and the large number of different specialisations within bioinformatics makes it hard for a single university to cover the width of applications for bioinformatics in current-day biology research. The containerized software model makes it easier to document and update your computing environment, and to accommodate workflows and training material developed by other labs. The Galaxy Training Network is a network of researchers who all provide training material which can be used on many different Galaxy servers (Batut *et al.*, 2018). By using software containers on the local server it is possible to keep local computing resources up to date and ensure that researchers can execute the workflows they've learned from the Galaxy Training Network.

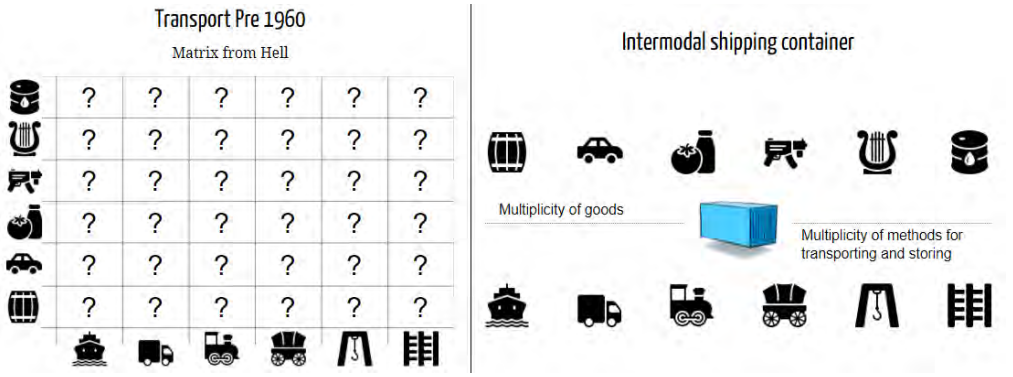


Figure 1. How the intermodal shipping container solved the matching matrix from hell in international shipping. An adaption from 'Docker and Galaxy' by Bérénice Batut and Björn Grüning, licensed under CC BY-ND 4.0.

The Galaxy platform is an excellent learning tool as it enables researchers to quickly start working with different tools and understand the different steps of data processing in a bioinformatics workflow. It is however not popular as an everyday tool for bioinformaticians developing workflows as it is inherently less flexible than scripting and command line tools. There is a plethora of options for bioinformaticians setting up their working environment but workflow managers such as Nextflow (Di Tommaso *et al.*, 2017) are gaining popularity as they provide the flexibility of scripting while improving transparency and reusability of code.

Results

To accommodate these different requirements, we are testing a simple coding pattern to ensure that workflows developed in Nextflow can easily be transformed to run as workflows on Galaxy systems. By using the large repository of Biocontainers (da Veiga Leprevost *et al.*, 2017) which contain container images for a vast number of commonly used bioinformatics tools and a 'one image per task' pattern, it is possible to ensure that workflows developed in Nextflow can also be run on a Galaxy server as both workflow managers can be configured to run Biocontainers on any computing environment that supports Docker or Apptainer (Singularity) containers. The Nextflow scripts can then easily be converted into Galaxy workflows and configured to submit jobs to the same scheduler on a computing cluster, meaning that actual computations will be run using the same software and compute nodes but jobs are defined and initiated by two different workflow management systems with different user requirements. This approach means that developed bioinformatics workflows can be quickly disseminated to other researchers as Galaxy workflows and training material be written as necessary to provide researchers with the knowledge and tools to perform their research.

Developing workflows with Nextflow we use four simple rules for coding (Figure 2).

1. Data is stored on a file storage system. Nextflow is configured so that containers are mounted on the file storage area and can access, modify and store the files used and created by the workflow.
2. Containers are loaded from a registry (usually Biocontainers). Each container contains the software necessary to execute *one task* within the workflow (e.g. quality control, trimming, alignment, etc.).
3. Configuration scripts are set to let the workflow engine take the workflow code, pass the job to the right system and execute the workflow with output being stored on the storage area.
4. All written code is to be commented and committed to the SLU Global Bioinformatics Centre Github. This is where everything is documented so that workflows upon completion of a project can be implemented in Galaxy.

By adhering to these principles bioinformatics research can quickly be turned into reusable workflows which can be adapted and used within the many different projects carried out on different species and breeds of relevance to agricultural researchers.

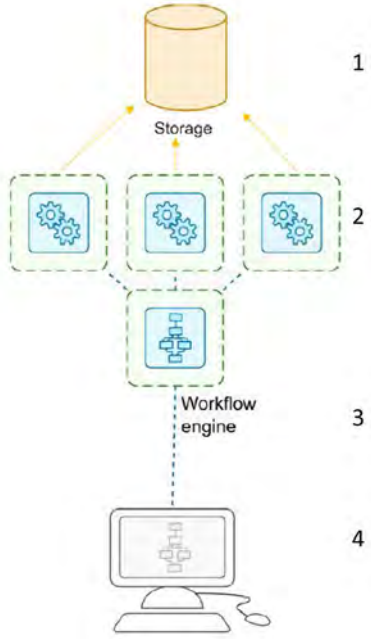


Figure 2. The Nextflow working environment. A workflow consists of an orchestrated and repeatable pattern of operations through which a piece of work is processed from initiation to completion. With the 'one image per task' pattern, a workflow engine like Galaxy or Nextflow mount a succession of software containers on the file storage which each execute the defined set of steps necessary to process the data (Spjuth *et al.*, 2021).

Acknowledgements

The authors acknowledge the support of the Galaxy community, the Freiburg Galaxy Team and Björn Grüning, Bioinformatics, University of Freiburg (Germany) funded by the Collaborative Research Centre 992 Medical Epigenetics (DFG grant SFB 992/1 2012) and the German Federal Ministry of Education and Research BMBF grant 031 A538A de.NBI-RBC. The exchange project with the Agricultural Research Council in South Africa has been funded by a STINT Initiation grant for Joint South Africa-Sweden Research Collaboration (SA2018-7728)

References

Batut B., Hiltemann S., Bagnacani A., Baker D., Bhardwaj V., *et al.* (2018). Cell Systems 6, 752-758.e1. https://doi.org/10.1016/j.cels.2018.05.012

da Veiga Leprevost F., Grüning B.A., Alves Aflitos S., Röst H.L. (2017). Bioinformatics 33, 2580–2582. https://doi.org/10.1093/bioinformatics/btx192

Di Tommaso P., Chatzou M., Floden E.W., Barja P.P., Palumbo E., Notredame C. (2017). Nat Biotechnol 35, 316–319. https://doi.org/10.1038/nbt.3820

Jalili V., Afgan E., Gu Q., Clements D., Blankenberg D., Goecks J., et. al (2020). Nucleic Acids Res 48, W395–W402. https://doi.org/10.1093/nar/gkaa434

Klingström, T., (2017). Acta Univ. agric. Suec., Silv. 1652-6880; 2017:99 Sveriges lantbruksuniversitet., Uppsala.

Spjuth O., Capuccini M., Carone M., Larsson A., Schaal W., *et al.* F1000Research 2021, 10:513 https://doi.org/10.12688/f1000research.53698.1

14. European master in animal breeding and genetics (EMABG): A training for research, practical, and transversal skills

G. Mészáros[1], P. Berg[2], T. Heams[3], H. Komen[4], D. Lont[4], H. Simianer[5], J. Sölkner[1], E. Strandberg[6] and J. Tetens[5*]

[1]University of Natural Resources and Life Sciences Vienna, Gregor-Mendel-Straße 33, 1180 Vienna, Austria; [2]Norwegian University of Life Sciences, Universitetstunet 3, 1430 Ås, Norway; [3]AgroParisTech, 16 rue Claude Bernard, 75231 Paris, France; [4]Wageningen University & Research, P.O. Box 338, 6700 AH Wageningen, the Netherlands; [5]Georg-August-Universität Göttingen, Albrecht-Thaer-Weg 3, 37075 Göttingen, Germany; [6]Swedish University of Agricultural Sciences, Dept of Animal Breeding and Genetics, Box 7023, 750 07 Uppsala, Sweden; jens.tetens@uni-goettingen.de

Abstract

The European Master in Animal Breeding and Genetics (EMABG) is an international master program operating on a global scale, with the aim to train experts and future leaders for research and industry. The training and education takes place at six European universities, within four study tracks with carefully designed curricula. One of the central parts of the MSc programme is the integration of teamwork activities and training of transversal skills to increase the employability of the graduates. To achieve this goal several major events are incorporated into the study programme during the entire duration of the two years. Also, strong emphasis is placed on integration of experts from the industry in the form of talks during joint events, internships, and direct work during special courses, as well as the final thesis.

Introduction

The general aim of the livestock sector is to meet the aspirations of the world's population for increased availability of animal products in a sustainable manner, while ensuring food safety, animal welfare, and maintenance of genetic diversity. To meet the challenge of the growing demands for food, livestock production needs to double in the coming thirty years, while halving its environmental impact (FAO, 2017). This implies that livestock production needs to improve efficiency of production, robustness and welfare of animals, and quality of animal products. Genetic improvement programmes help to meet these challenges by selecting animals for increased production with less input, reduced environmental impact per unit of product, and improved animal health and welfare. In the past, typically more than 50% of improvements of livestock productivity were caused by genetic progress through breeding. This illustrates both the need and opportunity for innovations in farm animal genetic improvement. Accordingly, training of young professionals addressing current and future challenges in animal breeding is needed.

European Master in Animal Breeding and Genetics

The European Master in Animal Breeding and Genetics (EMABG) presents a programme that results from intense discussion among six European partner universities. The programme is innovative in that it combines e-learning tools with training that follows the most recent scientific developments, effectively combining the expertise of each partner. It provides a rich learning environment through intense interaction with industry partners and NGOs, and offers flexible and personal learning paths (Figure 1). Summarizing all intakes and editions, a total of 161 students from 54 countries were trained in the program.

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_14, © G. Mészáros *et al.* 2022

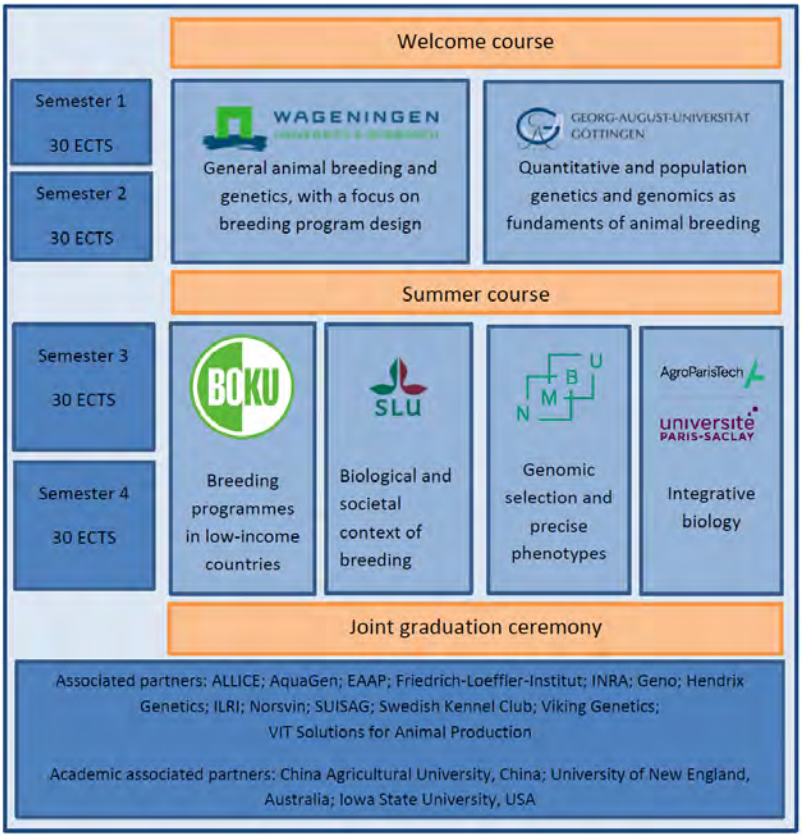


Figure 1. The structure of the EMABG MSc course.

The common vision of the programme is to deliver graduates that are experts in animal breeding and genetics, but with different foci, depending on their chosen study track. This makes them suitable for, e.g. managing breeding programmes in low-income countries, conducting breeding value estimations as required by industry, or responding to societal needs to combat climate change or improving animal welfare through breeding. The ultimate goal is to train a new generation of graduates that are fit to respond to current and future needs of the animal breeding sector in a global context.

The EMABG is a joint master programme that fills a real need for the industry sector and the whole society. It is designed to answer the scientific, practical, and societal challenges of animal breeding and genetics. These challenges can be summarised as follows:

- meet the increased demand for animal and fish products, while preserving or improving product quality and animal welfare, and minimizing the environmental impact;
- preserve natural resources, with special attention to biodiversity;
- make efficient use of novel tools and technologies to reach these goals while maintaining or improving societal acceptance of animal breeding;
- develop sustainable animal breeding programmes that contribute to improved livelihood of farmers and efficient food chains;
- develop sustainable breeding programmes that contribute to improved health and welfare of companion animals, including populations in zoos and nature reserves.

It is a common vision of EMABG partners to meet these challenges by intensive cooperation in the education and training of students, further enhanced by the complementarity among participating universities. Thus, the EMABG was set up to:

- offer quality higher education with a European perspective in the field of animal breeding and genetics, which is attractive to both programme (EU) and partner country nationals;
- enhance the quality of training in animal breeding and genetics and to strengthen attractiveness of the European Higher Education Area;
- contribute to solving societal needs by developing the qualifications of the students so they possess the appropriate skills, are open-minded and internationally experienced.

Joint events for the entire master course. There are three jointly organized courses, and a jointly supervised thesis between first-year and second-year universities (Figure 1). The Welcome course and the Summer course allow for personal interactions between all EMABG students and teaching staff from all partner universities. The teaching activities during the Welcome and Summer courses are covered by lecturers from all partners, associated partners and invited lecturers from academia and industry. The Breeding lab is organized at both WU and UGOE, and involves close collaboration with industry partners in specific group assignments. In addition to these courses, a transversal educational project was set up, where the students expand their knowledge and review various aspects of animal breeding and genetics in the context of the UN Sustainable Development Goals (SDGs).

The Welcome course is the highly successful cornerstone of the EMABG programme (according to regular feedback from students), and the defining element of academic and cultural induction of the students to the European higher education environment. This introductory event was developed to ensure a proper academic orientation of international students. EMABG students are introduced into the academic system in Europe, get the opportunity to establish personal contacts and friendships among each other, and get settled in before the start of the academic year. The students are also introduced to the European and global dimension of farm-animal breeding. During teaching and group work, close attention is paid to cultural differences and to rights and responsibilities of students and teachers in the EMABG. The course includes an intensive skills training part, which introduces students to topics ranging from poster and presentation skills to information literacy and scientific writing. Staff from all six partners participates in this event, which makes it possible to introduce each partner university and discuss the individual learning paths and study track choices of students.

The Breeding lab, is a course that is scheduled in the first and second semester of studies, simultaneously organised by WU and UGOE. Groups of students work on an assignment in close collaboration with associated industry partners. The aim of this course is to familiarize students with the competences and skills that are needed to work in a professional environment. The assignment is solved in interactive group work under the supervision of industry and academic staff from the partner universities. Examples are the design of a breeding programme for a specific environment or market, the definition of a new breeding goal, or the validation of a novel technology for breeding purposes. The aim is to put the knowledge of different elements of animal breeding acquired in the first year into a practical context. This helps the students to identify preferences for the choice of master projects.

The Summer course takes place between the second and third semester, overlapping with the Welcome course. The Summer course, titled 'Contribution of Animal Breeding to Global Food Security', is also offered as part of a summer school within the framework of the Euroleague for Life Sciences (ELLS) network, thus open to a worldwide audience. The content of the summer course is focused on widening the knowledge of students in terms of societal context of animal breeding and its practical relevance, with

participation of industry representatives and other associated partners. This increased networking outside of the academic environment makes it possible for the students to decide on their final preferences on the master thesis topic, including the possibility to collaborate with industry to solve currently relevant problems. The personal interactions also enable the screening of students for potential new hires by the industry, thus increasing the employability of EMABG graduates.

The overlap of the summer course with the Welcome course enables the two student groups to meet in a crucial time point in their studies. In this joint event the summer course students share their experience and give valuable advice to their younger colleagues regarding the habits, expectations, and workplace culture at the first-year universities. Intensifying the connectedness and interpersonal links between students in their mid-studies and those just starting further increases the cohesion of the EMABG programme.

The EMABG student graduation is a one-day event when the students from all second-year universities reunite to present their Master theses, celebrate their success, and receive their EMABG Diploma certificate. Graduates present their MSc theses in a conference-like setting, potentially also including associated partners and industry representatives, family members (online) and other guests. This small conference gives an additional possibility for the students to highlight their professional qualities and connect to potential employers. According to our EMABG alumni survey in 2017, about 40% of the graduates were already employed at the time of their graduation; this increased to over 90% six months after their graduation. The graduation itself is held as a formal ceremony, with the attendance of representatives from all EMABG partner universities, as well as senior officials from the organizing university.

A new element compared to previous iterations of the EMABG programme is the setup of the transversal projects on UN SDGs. This joint student project increases the interactions between the students within and between universities. One of the core elements of the transversal project is the preparatory group work within each university, supervised by each EMABG partner. Within this project the students work on competences that are needed in a professional working (industry) environment. They experience workgroup dynamics and aim-oriented working plan activities. The necessity of these transversal skills was highlighted within our needs analysis survey among potential employers. After course activities at each partner university the students engage in a joint videoconference to present their work and views on components of SDGs. This meeting gives an additional work-related meeting opportunity for the students in mid-semester.

Conclusions

The EMABG is an international master programme aimed at training experts and future leaders for the animal breeding research and industry. This training is being done at six partner universities, along four study tracks that cover various aspects and nuances for the breeding of livestock, fish, and companion animals on the global scale. A strong emphasis is placed on integration of industry experts at various stages of the two years of the MSc, so the students obtain the necessary teamwork, collaboration, and transversal skills. These skills, along with the strong theoretical background and practical computation skills are the foundation of their future success and increased employability.

References

FAO. 2017. The future of food and agriculture – Trends and challenges. Rome.

Part 4

Society – genetic solutions to achieve net zero carbon emission in livestock systems

15. Strategies to mitigate greenhouse gas emissions from pasture-based sheep systems – an EU project consortium view

J. Conington[1*], N. Lambe[2], F. Tortereau[3], F. McGovern[4], E. Navajas[5], I. De Barbieri[5], G. Ciappesoni[5], J. Jakobsen[6], E. Smith[7], J. Yates[7], Q. Le Graverand[3], K. McDermott[8], G. Steinheim[9], B. Aspeholen Aby[9], I. Dønnem[9], N. McHugh[4], L. Farrell[4], C. Marie-Etancelin[3], P. Johnson[10] and S. Rowe[10]

[1]SRUC Easter Bush, Roslin Institute Building, Edinburgh EH25 9RG, United Kingdom; [2]SRUC, Hill and Mountain Research Centre, Crianlarich, West Perthshire, FK20 8RU, United Kingdom; [3]GenPhySE, Université de Toulouse, INRAE, Castanet Tolosan, France; [4]Teagasc, Animal and Grassland Research and Innovation Centre, Athenry, H65 R718, Co. Galway, Ireland; [5]Instituto Nacional de Investigación Agropecuaria, INIA Las Brujas, 90100, Rincón del Colorado, Canelones, Uruguay; [6]Norwegian Association of Sheep and Goat Breeders, Box 104, 1431, Ås, Norway; [7]The British Texel Sheep Society, Stoneleigh Park, Warwickshire, CV8 2LG, United Kingdom; [8]SheepIreland Highfield House, Shinagh, Bandon, Co. Cork, P72 X050 Ireland; [9]Norwegian University of Life Sciences, Universitetstunet 3, 1433 Ås, Norway; [10]AgResearch, Invermay Agricultural Centre, Private Bag 50034, Mosgiel 9053 New Zealand; joanne.conington@sruc.ac.uk

Abstract

An international project ('Grass To Gas', 2019-2024) to combine expertise and generate new knowledge for the reduction of the potent methane (CH_4) greenhouse gas (GHG) from sheep is underway. Aims are to validate predictors of feed intake, methane emissions and feed efficiency, to investigate the association between feed efficiency and methane emissions measured indoors and outdoors, to explore the opportunity of using genetics and genomics (animal and microbiome) to reduce methane emissions in pasture-based sheep systems and quantify the economic and environmental benefits of more feed-efficient and lower GHG-emitting sheep linked to their microbiome. The potential impact is to deliver applied, sustainable solutions to reduce methane emissions for the international sheep breeding community, by bringing together the latest precision livestock monitoring and molecular technology to identify novel selection targets and potentially candidate genes.

Introduction

International commitments to reduce methane emissions to mitigate the impacts of climate change have been made by over 50 countries through the Paris agreement to limit global temperature rises to less than 1.5 degrees C; the recent Global Methane Pledge is to lower 2020 methane emissions level by 30% by 2030. Effective methane mitigation solutions are urgently needed to reduce CH_4 emissions for sustainable sheep production systems worldwide. In New Zealand, it has been shown that selective breeding can be used to achieve cumulative reductions in CH_4 per animal (Rowe *et al.* 2021), but there is no consensus about how best to implement direct and indirect measurements of CH_4 into breeding programmes internationally.

Materials & methods

Research objectives are designed to explore and validate results emanating from different countries undertaking replicated studies using native or economically important breeds. In the UK where new studies on individual animal feed intake indoors and outdoors are taking place, 250 Texel cross lambs sired by performance-recorded rams that had also been CT scanned as a lamb were monitored post-weaning for 6 weeks using CIEL-funded Biocontrol individual feed intake monitoring equipment in 2021 (and replicated in 2022). Their growth, body composition and rumen volumes (as determined by CT) at the start and finish of the trial (14 and 20 weeks of age on average) were also recorded. Paired lambs (related half- or full-sibs of those monitored indoors) were grazed out of doors on mixed perennial ryegrass swards and growth, body

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_15, © J. Conington *et al.* 2022

composition and grazing offtake measured, which was also replicated in Turkey using 250 Akkaraman sheep. In Norway, the effects of roughage quality (grass silage and freshly cut grass) and sheep breed on enteric CH_4 were investigated. Twenty Norwegian White sheep (NWS) and 20 Old Norwegian spael (ONS) were housed in individual pens and fed *ad libitum* roughage. Feed intake was monitored using the BioControl system. In both experiments, repeated 50-minute CH_4 concentrations were captured in portable accumulation chambers (PAC). Each ewe in the grass silage experiment had 30 PAC measurements, while each ewe in the fresh grass experiment had 21 measurements. Individual animal CH_4 measurements were collected on 454 NWS adult females from 44 breeding flocks from 12 ram circles. Detailed results of these measurements are reported in a separate, related paper (Jakobsen *et al.* WCGALP, 2022).

A validation of the n-alkane technique in Ireland estimated dry matter intake (DMI) of mixed breed sheep grazing perennial ryegrass swards. A total of 242 *in vivo* DMI measurements were obtained. Lin's concordance correlations were used to examine the relationship between the ratio of C_{33} / C_{32} n-alkane when predicting DMI. A series of experiments were carried out using 60 nulliparous ewe lambs whereby to measure CH_4, O_2 and CO_2 levels were measured over 50 minutes using PAC. Data were recorded on multiple days pre- and post-respiration chamber measurements. PAC measurements were subsequently collected on 2,404 animals varying in life-stage, breed, sire line, production type and diet type to quantify genetic variation in methane output within the commercial Irish sheep population.

Feed efficiency and CH_4 emissions have also been measured in Uruguay in 940 Australian Merino, 223 Corriedale, 360 Dohne Merino and 77 Texel lambs, which are part of to the national genetic evaluations, with the objective of investigating phenotypic and genetic associations among these traits, as well as with others relevant for sustainable sheep farming. Individual feed intake is recorded by Intergado® automated feeders during 56-day tests in which lambs are fed *ad libitum* with lucerne haylage. At the end of each test, CH_4 and CO_2 emissions are measured twice using PACs.

In France, male lambs of the Romane breed were measured at 3 months of age for performance, feed intake and feeding behaviour. These lambs emanated from two divergent lines for Residual Feed Intake (RFI) under a concentrate diet. The same measurements were also taken using a mixed diet. A protocol for GHG emissions phenotyping with GreenFeed (C-Lock) was defined. After each feeding period, blood, rumen fluid and faeces were sampled. Metabolomics analyses were undertaken on plasma and rumen fluid, and ruminal microbiota was sequenced for the 16S and 18s rRNA genes with the objective of identifying proxies of feed intake and GHG emissions, by using multivariate analyses (PLS, PLS-DA, Random Forest, SVR). In 2020, rumen volumes of 24 extreme Romane male lambs were CT-scanned to determine body composition and rumen size.

In New Zealand, animals selected divergently for methane emissions have also been measured for rumen microbial composition, milk, meat and wool composition, growth and slaughter traits, faecal egg counts and physiology of the alimentary tract. Part of this program has involved over 1000 female lambs measured for feed efficiency and feed intake traits to estimate relationships between methane emissions, microbial composition and gut related physiology. During this project, the role of New Zealand is to share protocols and data and to supply newly developed rapid sequencing technology of rumen microbes (Hess *et al.*,2020). Details of the feed efficiency programme are described in Johnson *et al.* (2018) and associated PAC measurements in Jonker *et al.* (2018).

Results

For feed intake in the UK after only 1 year of trial work, the Texel-X lambs showed significant between-sire variation in both average daily gain (ADG) and dry matter intake (DMI). At the sire level, initial results did not show a strong relationship between ADG and DMI. Live weight explained 41% of the variation in rumen volume (a potential methane predictor), with considerable variation in live weight adjusted rumen volume (CV 22 and 29% at each scan age), which differed significantly between sires, but was not significantly related to DMI in initial analyses. In the Norwegian grass silage experiment, DMI of early and average cut silage was 2.479 kg and 1.835 kg for NWS, and 1.337 kg and 1.143 kg for ONS, respectively. The difference in DMI was also reflected in enteric CH_4, being 1.65 vs 0.95 grams CH_4/hour for NWS and ONS respectively. When corrected to a mean DMI, ONS still had lower enteric CH_4 compared to NWS, (1.09 vs 1.52 grams CH_4/hr). Similarly, the early cut grass silage resulted in lower CH_4 emissions compared to the average, being 1.28 vs 1.32 grams CH_4/hour. Results from the 4,534 gas measurements of breeding females from 44 NWS flocks revealed mean CH_4 g/hr of 1.37 (s.d 0.28) and heritability CH_4/hr of 0.17 (se 0.04).

When estimating feed intake from grass, using the n-alkane technique in Ireland, the Lin's concordance correlation coefficient and the r-squared calculated between n-alkane estimated DMI using the ratio of C_{33} / C_{32} and measured DMI was 0.59 and 0.79, respectively. Offering animals low and medium herbage mass provided a more accurate estimate of DMI using the n-alkane technique. When investigating the measurement of CH_4 output, the time of day that animals were measured had no effect. For all three gases (CH_4, oxygen and carbon dioxide), the between-day variance (39 to 40%) accounted for a larger proportion of total variance compared to between-animal variance. The correlation between CH_4 production measured in PAC and the respiration chambers was 0.55, similar results were reported in New Zealand (Jonker *et al.*, 2018). When the data edits were applied to the dataset based on the rate of gaseous emissions 44 animals were retained, a higher correlation of 0.66 was estimated.

In determining the most appropriate model for the estimation of residual feed intake (RFI) using Uruguayan data, important fixed effects include pen-trial, average daily gain and metabolic body weight, whilst body composition and fleece growth traits were not significant (P>0.05) (Marques *et al.*, 2021). Amarillo *et al.* (unpublished data) also found the potential of reducing by 7 days the length of the feed efficiency test, without compromising the accuracy of RFI estimates, with R^2 being around 70%. The comparison of high, medium, and low RFI groups of lambs confirmed that more efficient animals are those with lower feed intake at similar performance levels in all breeds. Phenotypic Correlation coefficients in Corriedale, Dohne and Merino were moderate to high between CH_4 emissions and feed intake (0.8 to 64) but weaker with RFI (0.01 to 0.20). No differences in RFI were found between resistant (R) and susceptible (S) Corriedale sheep to internal parasites selected from divergent selection lines (Ferreira *et al.*, 2021). A Principal Component Analysis including body weight, average daily gain and gas emissions measured by PACs indicated the possibility of estimating feed intake, with a precision above 72% in allocating animals to high, medium, and low observed feed intake groups. More detailed results are in Le Graverand *et al.* WCGALP 2022.

In New Zealand, Johnson *et al.* (2018) reported that the most appropriate models to estimate RFI included fitting cohort, average daily gain and metabolic mid-weight (P<0.05) reaching an accuracy of 0.88. It also showed that whilst reduced periods of feed intake data could be used to estimate RFI, to obtain the most accurate average daily gain estimates the full 42d test period was most appropriate. The heritability estimates for CH_4 measured as absolute emissions and CH_4 yield (CH_4/CH_4+ CO_2) were 0.26±0.04 and 0.24±0.03, with the heritability estimate for RFI 0.46±0.13. The phenotypic (r_p) and genetic (r_g) correlation between CH_4 and the weight of the lambs were 0.62±0.03 and 0.84±0.05 respectively for absolute CH_4 emissions and 0.42±0.10 and 0.07±0.05 for methane yield. The phenotypic and genetic correlation between CH_4 and RFI were -0.18±0.12 and 0.08±0.05 respectively for absolute CH_4 emissions and -0.45±0.13 and -0.25±0.04 for

methane yield. There was a negative r_p between RFI and ultrasound fatness at the start of the measurement period (-0.23±0.04), and a positive r_p for RFI and change in fatness across the trial (0.21±0.04).

After three generations of divergent selection (2 s.d.) for RFI in France, more efficient animals (low RFI) ate 10% less concentrate, had shorter daily feeding duration with longer time intervals between feeder visits than less efficient (high RFI) animals. Ruminal microbiota and metabolomics do not yet show any significant difference between efficient and less efficient animals, regardless of the diet, however, more efficient animals have higher relative concentrations of citrate and malate and lower relative concentrations of amino acids in their blood than less efficient animals. Efficient animals also tended to have larger rumens, when expressed in relation to body weight (*P*-value=0.09). The correlation between indoor concentrate intake and indoor forage intake is positive and that with RFI are positive, being 0.22 and 0.25 respectively. In New Zealand, it has been shown that divergent selection for methane is associated with rumen size with high emitters of methane having larger rumens.

Discussion

One of the most successful outcomes for the project is truly independent validation of methods and protocols to monitor complex phenotypes associated with environmental impact. Although a full meta-analysis of data is yet to take place, sharing of methods, protocols and equipment has ensured that studies are comparable. Very similar heritabilities and repeatabilities from the analysis of methane emissions from PACs across different countries and sheep breeds indicate that the same trait is being evaluated and that global comparisons are possible using co-developed or co-validated proxy methods. The roles of feed intake, feed quality, liveweight, body composition and health status in accounting for variations between animals' performance and CH_4 emissions has been central to the investigations so far in this project. The use of PACs in experimental as well as commercial settings has generated data to optimise breeding programmes of the future. Participating countries should consolidate their reference populations locally and internationally to augment the role that genomic selection will have in tackling the problem of GHG emissions from livestock.

Acknowledgements

Each country receives funding from their national sponsoring organisation, under the banner of the Joint 2018 call of the three ERA-NETs (SusAn, FACCE ERA-GAS and ICTAGRI 2).

References

Ferreira, G. F., Ciappesoni, G., Castells, D., Amarillo-Silveira F., Navajas, E. A., *et al.* (2021). Anim. Prod. Sci. 61(8): 754–760. https://www.publish.csiro.au/AN/AN20121

Hess M., Rowe S., Van Stijn T.C., Henry, H.M., Hickey S.M. *et al* 2020, PLoS ONE 15(4): e0219882. https://doi.org/10.1371/journal.pone.0219882

Johnson, P.L., Knowler, K., Wing, J., Hickey, S., Johnstone, P. (2018). Proc. of WCGALP 11: 608, Auckland, NZ, 11-16 Feb.

Jonker A., Hickey S., Rowe S., Janssen P., Shackell G., Elmes, S. McEwan, J. (2018). J. Anim Sci, 96: 3031-3042. https://academic.oup.com/jas/article/96/8/3031/4993361?login=true

Rowe, S.J., Hickey, S.M.; Johnson, P.L.; Bilton, McEwan, J.C.; *et al.* Proc. of AAABG (2021) Adelaide, Australia, 2-4th Nov. http://www.aaabg.org/aaabghome/proceedings23.php

16. Comparing measurements of product output when defining emissions intensity

C.M. Richardson[1*], F. Hely[1], M. Post[1], T.T.T. Nguyen[2], G.J. Nieuwhof[2] and P.R. Amer[1]

[1]AbacusBio International Limited, Roslin Innovation Centre, Easter Bush, Midlothian, EH25 9RG, Edinburgh, Scotland, United Kingdom; [2]DataGene Ltd, AgriBio, Bundoora, 3083, Victoria, Australia; crichardson@abacusbio.com

Abstract

Various definitions of methane emission traits have been suggested for use within animal breeding programs, including methane production, methane yield, and methane intensity. However, when calculating emissions intensity multiple product output definitions could be used. We compared different definitions of product outputs used to define methane intensity for the purpose of evaluated animal breeding strategies. Three definitions of product outputs were explored: energy corrected milk (ECM), fat and protein corrected milk (FPCM), and protein equivalents (PROT-eq). Emission intensity estimated for each product output definition was 0.57 kg CO_2e/ kg ECM, 0.57 kg CO_2e/kg FPCM and 12.67 kg CO_2e/kg PROT-eq. While ECM and FPCM are primarily used to describe production levels in terms of energy content, PROT-eq consider products in terms of the value to the farmer in economic terms. We therefore recommend that PROT-eq should be used as the primary metric for defining emissions intensity in dairy industries.

Introduction

Multiple definitions of methane emission traits have been suggested for use within animal breeding programs. These definitions include methane production, methane yield, methane intensity and residual methane. In 2016, Australia committed to reduce Greenhouse Gas (GHG) emissions by 26 to 28% of 2005 levels by 2030 (Paris Agreement, 2016). Dairy Australia has set an industry goal of reducing its emission intensity (GHG per unit of product) by 30% by 2030 (Dairy Australia, 2021). Methane emissions intensity is therefore an alternative breeding objective that could be used in the national selection index to balance selection of traits which influence both the emissions and product output associated with traits available for selection simultaneously. This methane definition favours a more net neutral state and is especially efficient for systems where either animal inventories or production levels are capped, as both a reduction in emissions intensity and gross emissions may be achieved (Lahart *et al.*, 2021). While it has been demonstrated that selection for any methane trait may result in progress in all methane traits (Richardson *et al.*, 2021a), one challenge with reaching these emission reductions goals are the various definitions for emissions traits. Previous studies have explored the different methane emissions trait definitions, in particular residual methane production (Richardson *et al.*, 2021b), however, variations of the other methane traits definitions have not been explored. This study builds on our previous research by exploring the impact of alternative ways of quantifying milk output when defining emissions intensity. Methane intensity is defined as the amount of methane per kilogram of product, where the amount of product has been defined under multiple measures including the energy corrected milk (ECM), fat and protein corrected milk (FPCM), or protein equivalents (protein-eq). The aim of this research was to compare the different definitions of product outputs used to define methane intensity for the purpose of evaluated animal breeding strategies with modified weightings on existing traits to reduce GHG emissions.

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_16, © C.M. Richardson *et al.* 2022

Materials & methods

Australian dairy cow emissions and product output. The yearly gross methane emissions produced by an average Australian dairy breeding cow was previously estimated by Richardson *et al.* (2021b) using the following equation:

$$E = \sum_{i=1}^{c} n_i F_i \alpha Y_i \quad (1)$$

where E is the total gross methane emitted per breeding cow across c classes (replacement and cow), n_i is the number of animals in a given stock class i per breeding female in the average Australian dairy herd, α is the methane carbon dioxide equivalent coefficient per kg of DMI (0.532 kg CO_2/kg DM), F_i is the amount of feed (in kg DM) per animal of the stock class i, and Y_i is a binary indicator variable that takes a values of 1 or 0 depending on whether the class i is considered in the total definition of methane emissions, respectively. In this case, the animal classes considered were milking cows and replacement heifers as male calves have marginal value in the Australian dairy industry.

Definitions of product outputs. A total of three definitions of product outputs were explored: energy corrected milk (ECM), fat and protein corrected milk (FPCM), and protein equivalents (PROT-eq). These were estimated using the following equations:

$$\text{FPCM} = \text{Milk kg} \times ((0.1226 \times \text{Fat \%}) + (0.0776 \times \text{Protein \%}) + 0.2534) \quad (2)$$

$$\text{ECM} = (\text{Milk kg} \times ((376 \times \text{Fat \%}) + (209 \times \text{Protein \%}) + 948))/3{,}138. \quad (3)$$

$$\text{PROT-eq} = (k_{fat} \times \text{kg Fat}) + (k_{protein} \times \text{kg protein}) + (k_{milk} \times \text{kg milk}) \quad (4)$$

where k_x is defined as the value of each product x expressed as a proportion of the market value of protein. Production parameters in this study were assumed to be representative of the average Australian dairy cow, with an average 305-d production of 6,861.00 kg milk at 4.00% fat, 3.30% protein and 4.80% lactose (DataGene Ltd., 2017).

Calculating k-values for milk and component traits. In-line with previously published studies (Richardson *et al.*, 2021; Zhang *et al.*, 2019), the product classes considered in this study are milk and its components (fat and protein). Similar to the concept that GWP is used to convert methane into CO_2 equivalents (CO_2e), product conversion factors, k_x, convert milk and its component traits into a single output. This allows for multiple product classes to be considered and evaluated, while maintaining a single product class for comparison. The product classes used for comparison in this study are the three definitions of product outputs. The product conversion factors for ECM and FPCM were estimated similarly, where the relative value of each product class was determined by calculating the difference in ECM or FPCM calculated when each product was independently increased by 1 kg. In comparison, PROT-eq conversion factors were based on component values of $6.76, $2.08 and -$0.11, for protein, fat and milk, respectively (Byrne *et al.* 2016). The PROT-eq value for milk volume is constituted by the milk volume percentage of lactose (4.8%) to avoid double counting the value of milk fat and milk protein, when milk volume is increased.

Calculating emissions intensity. Emissions intensity was calculated using the gross emissions per animals as estimated by Richardson *et al.* (2021b), divided by each of the suggested product outputs, and calculated as:

$$EI_i = \frac{E\ (kg\ CO2e)}{P_i\ (kg)} \quad (5)$$

where EI is the emissions intensity value for each of the product output *i* classes (ECM, FPCM, PROT-eq), E is the gross emissions measured in kg CO_2e, and P_i is the total product output for *i* classes.

Results

Product output and the respective emissions intensity for each product output definition are presented in Table 1. Similar product outputs were estimated for ECM and FPCM of 6,869.09 kg and 6,860.18 kg, respectively, whereas the product output expressed in PROT-eq was 306.15 kg. Table 2 presents the implied values of fat and milk relative to protein under alternative milk product output definitions. The change in kg of milk and its components due to a 1 kg increase in ECM or FPCM was 6.66 and 7.76 kg protein, 1.98 and 12.26 kg fat, and 1.00 and 0.99 kg milk. The PROT-eq product conversion factors for milk, fat and protein were -0.0008, 0.31 and 1, respectively.

Discussion

The estimated total product output is highly dependent on the metric of product output chosen, and the resulting emission intensity values are variable and driven by different factors. Thus, choice of different definitions of milk output (i.e. ECM, versus FPCM versus PROT-eq) is likely to lead to quite radically different responses to selection according to the different definition of emissions intensity. The fat and protein corrected milk definition of emissions intensity places a higher weighting on milk fat than on milk protein as a productive output. Use of energy corrected milk for the emissions intensity value is almost identical. These are both ways of correcting milk output based on energy composition of the components, and ignore the fact that protein is typically valued more highly as a milk output than fat, even though it has a lower energy content. There is a wide range in weightings on fat relative to protein across national

Table 1. Product output and methane emissions intensity using ECM, FPCM, PROT-eq.

	Units	Total product	EI[1] (kg CO_2e/kg product)
ECM[2]	kg	6,869.09	0.57
FPCM[2]	kg	6,860.18	0.57
PROT-eq[2]	kg	306.15	12.69

[1] EI is the emission intensity.

[2] ECM is energy corrected milk, FPCM is fat and protein corrected milk, and PROT-eq is protein equivalents.

Table 2. Implied values of fat and milk relative to protein under alternative milk product output definitions.

	Protein (kg)	Fat (kg)	Milk (kg)
ECM	1.00	1.98/6.66 =1.8	1/6.66 = 0.15
FPCM	1.00	12.26/7.76 = 1.58	0.99/7.76 = 0.13
PROT-eq	1.00	0.31	-0.0008

[1] ECM is energy corrected milk, FPCM is fat and protein corrected milk, and PROT-eq is protein equivalents.

selection indexes of different countries. For example, the USA and New Zealand national indexes assume a higher market value for milk fat than for milk protein, while many other national indexes (including Australia) assume a market value for fat which is less than half of the market value of protein.

Defining an emissions intensity breeding objective in terms of ECM or FPCM places a significantly larger and positive value and resulting emphasis on milk compared to PROT-eq. This is in direct conflict with the majority of milk payment systems in Australia which penalise milk volume (Axford *et al.*, 2021).

Conclusions

ECM and FPCM are definitions of product output primarily used to describe production levels in terms of energy content. While PROT-eq on the other hand consider products in terms of the value to the farmer and economics. Therefore, while ECM and FPCM are currently widely used internationally, we recommend that ECM and FPCM should be abandoned as metrics for defining emissions intensity in dairy industries.

References

Axford M.M., Williams P.W., Abernethy D.P., and Nieuwhof G.J. (2021). Proc. of the 23rd AAABG, Armidale, Australia

Byrne T.J., Santos B.F.S., Amer P.R., Martin-Collado D., Pryce J.E., *et al.* (2016). J. Dairy Sci. 99:8146–8167. https://doi.org/10.3168/jds.2015-10747

DataGene Ltd. 2017. Australian dairy herd improvement report. Accessed May 21, 2020. https://datagene.com.au/ct-menu-item-13/industry-statistics?fid=18.

Lahart, B., Shalloo L., Herron J., O'Brien J., Fitzgerald R. *et al.* (2021). J Dairy Sci. 104: 8039–8049. http://dx.doi.org/10.3168/jds.2020-19618.

Paris Agreement. 2016. Accessed Dec 11, 2020. https://unfccc.int/process-and-meetings/the-paris-agreement/the-paris-agreement.

Richardson C.M., Nguyen T.T.T., Abdelsayed M., Moate P.J., Williams S.R.O., *et al.* (2021a). J. Dairy Sci. 104:539-549. http://dx.doi.org/10.3168/jds.2020-18565

Richardson C.M., Amer P.R., Hely F., van den Berg I., Pryce J.E. (2021b). J. Dairy Sci. 104:10979-10990 https://doi.org/10.3168/jds.2021-20348

17. The genetic progress has effect on greenhouse gas emission intensities from Norwegian pork production

E. Gjerlaug Enger[1*], H. Bonesmo[2], K. Martinsen[1], I. Andersen-Ranberg[1], D. Olsen[1] and E. Grindflek[1]

[1]Norsvin, Storhamargata 44, 2317 Hamar, Norway. [2]Norwegian Institute of Bioeconomy Research (NIBIO), Division of Survey and Statistics, P.O. Box 115, 1431 Ås, Norway; eli.gjerlaug@norsvin.no

Abstract

The environmental sustainability of food production systems, including net greenhouse gas (GHG) emissions, is of increasing importance. The objective for this study was to investigate whether an IPCC methodology-based model could reflect the effects in reduction of GHG emissions from the genetic progress in pork production. Pig production data from 654 Norwegian farms were used in the model and results revealed that the most important traits for GHG emissions were feed conversion rate, weaned piglets per litter and mortality. Estimated genetic gains for 2021 for these traits resulted in an estimated yearly reduction in GHG emissions of 1.4 and 1.9% for Landrace and Duroc, respectively. Further a study of the same farms, based on data from 2014 to 2019 showed that estimated net GHG emissions decreased from 2.49 to 2.34 kg CO_2 equivalents/kg carcass weight, with an average of 1.2% annually over the investigated period.

Introduction

The environmental sustainability of food production systems, including net greenhouse gas (GHG) emissions, is of increasing importance for diet recommendations. Thus, to preserve a proportion of animal-based food in such recommendations, it is necessary to further reduce the net GHG emission intensities from animal husbandry systems. As monogastric animals, pigs have low enteric methane (CH_4) production compared to ruminants. The largest sources of net GHG emissions from pig production systems are manure storage, including nitrous oxide (N_2O) and CH_4, and from the feed production, both N_2O and carbon dioxide (CO_2) are included.

In Norwegian pork production, animal performance is high both in terms of reproduction and growth in addition to low mortality. Following the introduction of genomic selection in the pig breeding program in 2014 (Nordbø *et al.*, 2014), the progress in animal performance has been further increased (Ingris, 2019). Animal health status is high due to the closed national population of animals, which has contributed to the eradication of several swine diseases common in the rest of the world (Grøntvedt *et al.*, 2016). Norwegian pig farms are small, based on family farm units, and the pig production has higher requirements for animal welfare than in most other European countries (Falk and Hofshagen, 2019).

A hybrid sow is often a cross between two maternal breeds; in Norway, this is normally Norsvin Landrace and Topigs Norsvin Z-line (Yorkshire breed). The commonly used sire line is Norsvin Duroc. The aim of this study was to quantify how much the genetic progress contributed to reduction in GHGs, as well as to investigate whether this could be measured in commercial herds.

Materials & methods

The estimates of GHGs include emissions of enteric and manure storage methane (CH_4), direct and indirect nitrous oxide (N_2O) from the pig-barn and manure storage, and emissions from feed production. The emissions from feeds, both domestic and imported, include estimates for direct and indirect N_2O from both manure and synthetic fertilizer, soil carbon change (CO_2), and direct and indirect energy use (CO_2

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_17, © E. Gjerlaug Enger *et al.* 2022

equivalents (eq.)). In addition, the estimates of GHG emissions intensity encompasses the GHG emissions from the production process and transportation of purchased feeds, and GHG emissions caused by direct energy use in pig-barns. All age classes of the pig's life are included, as well as sow and replacement of gilts. All GHG emissions are expressed as CO_2 eq., to account for the global warming potential of the respective gases given a time horizon of 100 years: CH_4 kg × 25 + N_2O kg × 298 + CO_2 kg × 1 (IPCC, 2007).

The points explain the course of the trial:

1. The IPCC methodology-based model was developed and implemented in a GHG calculator for pig farmers (Bonesmo and Enger, 2021; https://klimasmartlandbruk.no).
2. The GHG calculator was designed to use readily available production parameters from pig farms. The data from commercial farms in this study were obtained from the Norwegian litter recording system (Ingris, Animalia AS).
3. To estimate the effect of the different parameters in the model, a sensitivity test was performed with data from 2019 at 654 Norwegian pig farms. One production parameter at the time were improved with one standard deviation and change in estimated GHG emissions was estimated.
4. The production parameters were ranked based on the effect they had on the estimated GHG emission intensities. The top four are presented in the current paper.
5. Genetic trends for 2021 based on genomic estimated selection values for Norsvin Landrace and Duroc selection candidates for the four parameters were calculated. The effects the four parameters had on GHG emissions were estimated based on data in commercial farms, see point 3.
6. Progress measured in the field from 2014 to 2019 (Ingris, 2014, 2015, 2016, 2017, 2018, 2019) was estimated based on the GHG calculator combined with readily available production parameters from Norwegian pig farms, as described in point 1 and 2. The data used are obtained from a different time period than the data used for genetic gain estimations, described in point 5. However, we have chosen to focus on the reduction in GHG emissions from expected genetic gain in the years to come.

Results

The model enabled estimation of the effects on net GHG emission intensities using available production data from pork production (Table 1).

Of the investigated parameters, the feed conversion ratio (FCR) for finishers had the highest importance. When FCR was improved with one standard deviation, from 2.45 to 2.29 kg feed per kg growth, the reduction in total estimated pork GHG emissions was shown to be of 4.5%. A larger number of parameters were investigated, and the four parameters with largest effect are presented in Table 2, ordered by importance.

Table 1. GHG emissions presented per kg carcass weight (CW), with percentage distribution for Norwegian pig production.

Sources and type of GHG emissions	Estimated GHG emissions, kg CO_2 eq./kg CW	Percentage distribution
Overall	2.33	
Feeds	1.87	80%
Electricity and transport	0.11	5%
Enteric CH_4	0.14	6%
Manure CH_4	0.11	5%
Manure N_2O	0.10	4%

Table 2. Results of the sensitivity test. Changes in estimated GHG emissions are shown for one standard deviation (STD) improvement in key production characteristics.

	Current values, 2019	STD	Estimated GHG emissions intensity, kg CO_2 eq./kg CW	Change in estimated GHG emissions
FCR[1] finisher (kg/kg)	2.45	0.16	2.23	-4.5%
Weaned piglets/sow/year	27.9	3.78	2.28	-2.1%
FCR[1] weaners (kg/kg)	1.54	0.20	2.30	-1.4%
Mortality finishers, %	1.7	1.33	2.30	-1.3%

[1] FCR = feed conversion ratio as kg feed divided on kg growth in live weight, with on overage 9.5 and 9.9 MJ net energy per kg feed for finisher and weaned piglets, respectively.

Estimated genetic gain for parameters from the sensitivity test were based on data from 2021 (Table 3). The corresponding improvements from estimated genetic gain in annual reduced GHG emission intensities (Figure 1) were 1.4 and 1.9% for Landrace and Duroc, respectively. Based on the field data, there was estimated a substantial annual decrease between 2014 and 2019 in the emissions of pork, resulting in 6% lower emissions in 2019 than in 2014 (Figure 2).

Table 3. Estimated genetic gain based on genomic estimated breeding values for selection candidates in Norwegian Norsvin nucleus herds in 2021.

	Dam line (Landrace)	Sire line (Duroc)
FCR[1] finisher (kg/kg)	-0.03	-0.06
Weaned piglets/sow/year	0.96	-
FCR[1] weaners (kg/kg)	-0.01	-0.02
Mortality finishers, %	-0.11	-0.20

[1] FCR: see Table 2.

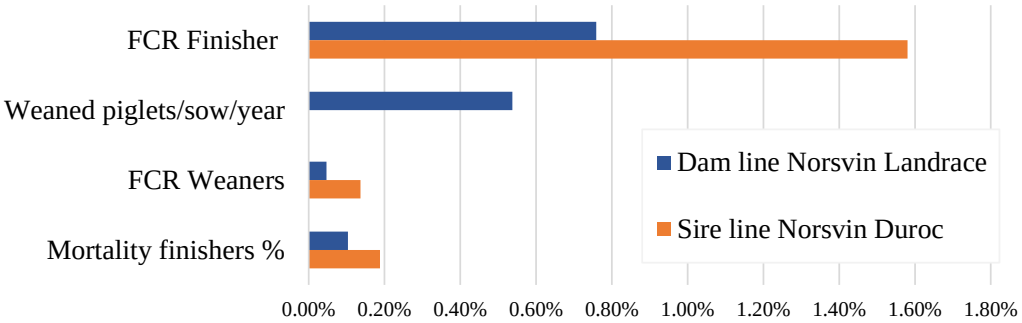


Figure 1. Percentage reduction in GHG emissions for the four most important parameters for GHG emissions in pig production, due to annual genetic gain.

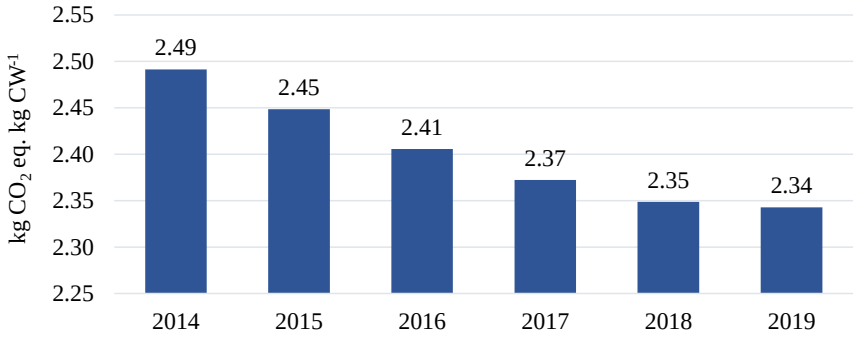


Figure 2. Estimated pork GHG emissions intensity from 632 Norwegian farms.

Discussion

So far, reductions in GHG emissions *per se* has not been included in the breeding programme for Norwegian pigs. However, as direct emissions from animals, *i.e.* enteric methane, is of minor magnitude, it is mainly feed efficiency traits that are of importance. About 80% of the GHG emissions per kg carcass weight comes from feed production (Table 1), and in an average Norwegian pig farm about 75% of the variable costs comes from the feed. It was therefore hypothesized that the genetic progress has led to reduced GHG emissions and that expected improvements will give a cumulative and permanent reduction in GHG emission intensities. In the current study, it was shown a yearly improvement in reduced GHGs of 1.4 and 1.9% for Landrace and Duroc, respectively, based on estimated genetic gain from breeding values. From 2014 to 2019 the average progress measured in the field is 1.2% annually. This encompasses more than genetic progress, there may also be effects from the increase in professionalism of management and feeding. Still, it is no doubt that the effects on net GHG emission intensities from breeding is profound, and the progress in pig genetics therefore contributes to an on-going strengthening of pork as a sustainable source for human food supply. Very few protein-rich foods have as low or lower GHG emissions than those we find for Norwegian-produced pork (Oort and Andrew, 2016; Bonesmo and Enger, 2021).

References

Bonesmo, H. and Enger, E. G. (2021). Livest Sci, (254): 104746. https://doi.org/10.1016/j.livsci.2021.104746

Falk, M. and Hofshagen, M. (2020). Dyrehelserapporten 2019. Veterinærinstituttet 19-2020.

Grøntvedt, C. A., Elstrøm, P., Stegger, M., Skov, R. L., Skytt Andersen, P. *et al.* (2016). Clin. Infect. Dis, 63(11): 1431-1438. https://doi.org/10.1093/cid/ciw552

Ingris, (2014, 2015, 2016, 2017, 2018, 2019). Årsstatistikk. Annual report. Available at: https://animalia.no

IPCC 2007. Forster, P., Ramaswamy, V., Artaxo, P., Berntsen, T., Betts, R. *et al.* (2007). Chapter 2. In Climate change 2007. The physical science basis. Available at:

http://www.ipcc.ch/pdf/assessment-report/ar4/wg1/ar4-wg1-chapter2.pdf

Nordbø, Ø., Gjerlaug-Enger, E., Aasmundstad, T. and Grindflek, E. (2014). Proc. of the 10th WCGALP, Vancouver, Canada

Oort, B. V. and Andrew, R. (2016). Climate footprints of Norwegian dairy and meat-a synthesis. CICERO report. Available at: http://hdl.handle.net/11250/2397086

18. An approach to the genetic parameters of methane emission in Latxa dairy sheep

C. Pineda-Quiroga*, I. Granado-Tajada, E. Ugarte, I. Goiri and A. García-Rodríguez

Animal Production Department, Neiker-BRTA, Campus Agroalimentario de Arkaute, 01080 Vitoria-Gasteiz, Spain; cpineda@neiker.eus

Abstract

Genetic selection is a promising approach to identify and breed lower-methane emitter animals. As a first step, this study aimed at estimating the genetic parameters of methane emission in Latxa breed using an experimental data set. For this, 295 methane concentration measurements were taken on 103 genotyped ewes, using a non-dispersive infrared methane detector. Methane intensity (MeI, g/l of milk) and production (MeP, g/day) were calculated. Results showed the existence of variability for both traits, but they are under low genetic control, with non-significant heritabilities of 0.11±0.13 for MeI and of 0.07±0.10 for MeP. Low repeatability (0.16±0.07 for MeI and 0.07±0.06 for MeP) suggests that the surrounding environment variables are the strongest controlling factors. Indeed, the body weight of ewes and the diet were defined as relevant non-genetic factors influencing emissions. A larger data set is needed in Latxa to properly make inferences.

Introduction

As widely known, there is a significant research effort to find ways to reduce enteric methane emissions produced by ruminants, due to their contribution to global warming. For this, animal breeding could make a valuable contribution to the whole set of methane mitigation strategies, not only because this trait is heritable, but also due to the advantage that results of animal selection are cumulative and last over time (Pickering *et al.*, 2015). Nevertheless, this approach requires time to have a noticeable effect, as selection is carried out over generations (Lassen and Difford, 2020).

Latxa breeding program has been successfully selecting for milk production, composition and udder morphology for the last decades, and the inclusion of new traits to the program scheme in the future is now being studied (Carabaño *et al.*, 2021; Pineda-Quiroga *et al.*, 2021). Before a novel trait, such as methane emission, can be added to any genetic program, a large number of records are needed. Furthermore, complementary information, such as genotyping data, is also valuable when available. Such information is used to appropriately quantify the phenotypic and genetic variability of the trait, as well as to identify non-genetic factors affecting the phenotype. However, obtaining individual methane measurements at the farm level is usually a time-demanding and costly task. In the sheep production context, the huge variety of facilities between breeds and systems represents an additional handicap. For this reason, pilot genetic studies carried out under experimental conditions to shed light about methane emissions in this species are welcome (Piñares-Patiño *et al.*, 2013; Robinson *et al.*, 2020).

In this context, the objective of the present study was to elucidate for the first time the genetic component of the methane emission in Latxa dairy sheep, combining phenotypic and genomic information, under experimental conditions.

Materials & methods

Records of methane emissions from 103 Latxa Cara Negra from Euskadi dairy ewes, from an experimental flock was collected using a portable non-dispersive infrared methane detector (Guardian NG infrared gas monitor; Edinburgh Instruments Ltd., Livingston, UK). Data were collected during different time points,

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_18, © C. Pineda-Quiroga *et al.* 2022

in which animals were housed indoor and were being fed forage and different concentrates. Briefly, a gas sampling tube was installed in the front of sheeps' nostrils and connected to the gas analyser. Individual methane concentration in the sheeps' breath was continuously measured at 1 s intervals during 10 minutes and stored in a datalogger. Each ewe underwent between two and three measurements (295 in total). Methane intensity (MeI, g/l of milk) and methane production (MeP, g/day) were then calculated.

Molecular information was available for all animals, which had been genotyped using 50K Illumina chip OvineSNP50 (n=28) or Affymetrix Axiom Ovine Genotyping Array (n=75). Markers in common, by chromosome and position between genotyping platforms, were kept before quality control. Monomorphic SNPs, SNPs with a call rate <90, with a minor allele frequency <0.05, and those markers located in the sex chromosomes were filtered out, resulting in a total of 30.074 effective markers for further analysis.

Initial exploratory analyses were undertaken using a general linear model to define the significantly different (P<0.05) and relevant effects to be included in the model for genetic estimations.

For variance component estimations, a genomic restricted expectation maximum likelihood, using the average information (AI) algorithm was carried out by using AIREMLf90 from the blupf90 family programs (Misztal *et al.*, 2002), following the model:

$$y = Xb + Zg + Zp + e \quad (1)$$

where y is the vector of phenotypic records (MeI and MeP), b is the vector of the fixed effects body weight (level 1: 45 to 53 kg, level 2: >53 to 64 kg, level 3: >64 to 95 kg), concentrate supplementation (linseed oil – LO, linseed cake – LC, rapeseed cake – RC, sunflower cake – SC, spent coffee grounds –SCG or without supplementation – CTR) and days in milk production fitted as covariate, with the respective incidence matrix X; g is the vector of direct animal genetic effect with incidence matrix Z, considering the variance/covariance matrix of the SNPs as the relationship matrix G in the mixed model equations, p is the vector of permanent effect with incidence matrix Z_p, and e is the vector of residuals.

Results

As shown in Table 1, the phenotypic values of methane varied depending on the trait used, being lower when it was expressed in function of the milk production (MeI). As regards the genetic components, the heritability showed slightly higher values when methane emission was referred as MeI compared to MeP, because the total variance was lower. The repeatability showed the same tendency as heritability.

Table 1. Phenotypic mean and additive (σ^2u), permanent (σ^2p), residual (σ^2r) variances, heritability (h^2) and repeatability (R) of methane intensity MeI (g/l of milk) and production MeP (g/day).

	MeI (g/l of milk)	MeP (g/day)
Mean ± sd	10.36±5.83	21.28±10.74
σ^2u	2.96	6.52
σ^2p	1.50	0.57
σ^2r	23.01	88.40
h^2 ± se	0.11±0.13	0.07±0.10
R ± se	0.16±0.07	0.07±0.06

Figure 1 shows the influence of non-genetic factors on methane emission. As shown, the emission increased with body weight (~13% for MeI and ~11% for MeP). For concentrate supplementation, LO, LC, RC and SC reduced the emissions in comparison with CTR (~8% for MeI and ~12% for MeP), whereas SCG increased them (~31% for both traits). For days producing milk, a steady increase for both traits was noted, although the slope was double in MeI than in MeP (0.10 vs 0.05).

Discussion

The findings of the present study confirm that methane emission is under moderately low additive genetic control, with the surrounding environment variables being the strongest controlling factor. As expected, Latxa's MeP heritability was low and close to the lowest reported value in the Merino sheep breed (0.10±0.06, Paganoni *et al.*, 2017). It was not possible to make a direct contrast of MeI with other sheep data because, to the best of our knowledge, no results have so far been reported for this. Nevertheless, our value is smaller than the lowest reported for Holstein dairy cows (0.20±0.16, Lassen *et al.*, 2016). These comparisons should be taken cautiously due to the fact that farm conditions, the number of sampled animals, the measurement techniques and the total records taken per individual varied between the studies. On top of that, our results suggest that MeI showed higher genetic variation over the total estimated variance than MeP, although it is evident that a larger set of measurements is needed in Latxa to properly make inferences.

The low repeatability values found in the present study, in line with those from Piñares-Patiño *et al.*, 2011 for methane emission (g/kg of DMI) measured on a hundred animals, might reaffirm that it is necessary to phenotype much more animals. Furthermore, these values may also confirm that other no-genetic effects influence the trait, such as those related to intrinsic factors of the animals' biology and their management on farm, as well as methods of recording (Lassen and Difford, 2020). Indeed, as presented, the body weight

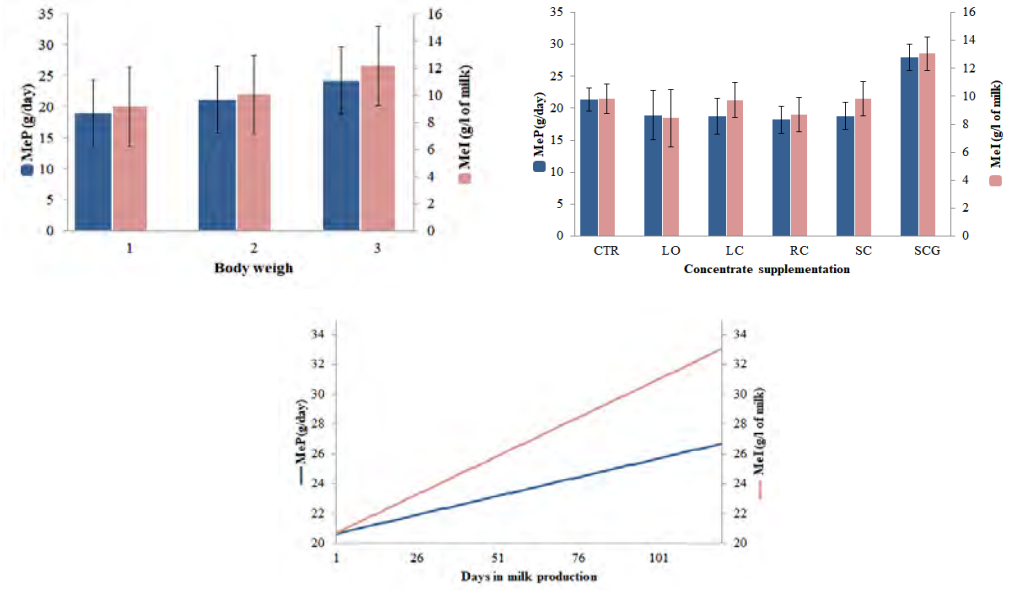


Figure 1. Non-genetic factors affecting the methane emission phenotype. Body weight; level 1: 45 to 53 kg, level 2: >53 to 64 kg, level 3: >64 to 95 kg; Concentrate supplementation – CTR: without supplementation, LO: linseed oil, LC: linseed cake, RC: rapeseed cake, SC: sunflower cake, SCG: spent coffee grounds; days in milk production – from 1 to 120 days.

and the type of concentrate largely influence the methane phenotype. The first factor because the bigger the animal, the larger the rumen capacity and the animals' feed intake, and the second factor because the diet composition alters the bacterial communities (Goiri *et al.*, 2020) and hence the production of methane precursors. Nevertheless, as proposed by Cottle and Conington (2012), it will be probably required to implement a combination of genetic and non-genetic measures to have short-term results, and for a cost-effective reduction in methane production.

All in all, there is potential for adopting genetic selection on methane emissions in ruminants and, in the future, genomic selection. Further investigations involving more phenotyped and genotyped animals, as well as measurements coming from sheep from diverse genetic backgrounds may also be needed.

References

Carabaño M.J., Pineda-Quiroga C., Ugarte E., Díaz C., Ramón M. (2021) J Dairy Sci 104(5):5755-5767. https://doi.org/10.3168/jds.2020-19503

Cottle D.J., Conington J. (2012) J Agric Sci 150:570-583. https://doi.org/10.1017/S0021859611000967

Goiri I., Díaz de Otálora X., Ruíz R., Rey J., Atxaerandio R., *et al.* (2020) Microorganisms 8(12):1961-1976. https://doi.org/10.3390/microorganisms8121961

Lassen J., and Løvendahl P. (2016) J Dairy Sci 99(3):1959-1967. https://doi.org/10.3168/jds.2015-10012

Lassen J., Difford G.F. (2020) Animal 14(S3):s473-s483. https://doi.org/10.1017/S1751731120001561

Misztal I. S., Tsuruta T., Strabel B., Auvray T., Druet D.H., *et al.* (2002). Proc of the 7th World Congress on Genetics Applied to Livestock Production, Montpellier, France.

Paganoni B., Rose G., Macleay C., Jones C., Brown D.J., *et al.* (2017) J Anim Sci 95:3839-3850. https://doi.org/10.2527/jas2017.1499

Pickering N.K., Oddy V.H., Basarab J., Cammack K., Hayes B., *et al.* (2015) Animal 9(9):1431-1440. https://doi.org/10.1017/S1751731115000968

Pineda-Quiroga C., Ugarte E., (2021) Proc. of the 72nd EAAP, Davos, Switzerland

Piñares-Patiño C.S., McEwan J.C., Dood K.G., Cárdenas E.A., Hegarty R.S. (2011) Anim Feed Sci Tech 166:210-218. https://doi.org/10.1016/j.anifeedsci.2011.04.068

Piñares-Patiño C.S., Hickey S.M., Young E.A., Doods K.G., MacLean S., *et al.* (2013) Animal 7(s2):316-321. https://doi.org/10.1017/S1751731113000864

Robinson D.L., Dominik S., Donaldson A.J., Oddy, V.H., (2020) Anim Prod Sci 60(7):880-892 https://doi.org/10.1071/AN18383

19. Ameliorating the contribution of livestock to global greenhouse gas production: dream or deliverable?

R. Hegarty[1*], J.C. McEwan[2] and S.J. Rowe[2]

[1]New Zealand Agricultural Greenhouse Gas Research Centre; [2]AgResearch, Invermay Agricultural Centre, Private Bag 50034, Mosgiel 9053, New Zealand; roger.hegarty@nzagrc.org.nz

Abstract

International agreements such as the Paris Agreement, the Global Methane Pledge and alliances such as Race to Zero, engage countries to constrain temperature rise by mitigating greenhouse gas emissions. Nationally Determined Contributions show mitigation ambitions are currently below that needed to reach these targets. While livestock methane need not be stopped completely to reach these targets, mitigation to reverse rising global methane emissions will be required, especially in developed countries. This paper addresses the vision for that mitigation, the sharing of that task across developed and developing nations, and how animal breeding and nutritional or management strategies can contribute. Delivering substantial absolute mitigation as well as reducing emission intensity and improving system efficiency are possible, thereby delivering on many of the Sustainable Development Goals.

Introduction to global greenhouse gases and growing human needs

Mitigation vison. While the nations of the world have committed to binding national targets to lower GHG emissions, progress to meet these targets is slow. Land-based measures (including livestock) have potential to mitigate approximately 10-15 $GtCO_2eq$/yr by 2050, corresponding to about 20-30% of the mitigation needed to achieve the 1.5 °C temperature target (Roe *et al.* 2021). The Nationally Determined Contribution (NDC) declares each nation's highest possible (but non-binding) ambition. (UNFCCC 2021) and possible contribution to the Paris target of 'holding the increase in the global average temperature to well below 2 °C above pre-industrial levels and pursuing efforts to limit the temperature increase to 1.5 °C above pre-industrial levels' (UNFCCC 2015). Currently, comparison of the target and the likely abatement from countries submitting an NDC shows 'the emissions gap remains large … the new pledges for 2030 reduce projected 2030 emissions by only 7.5%, whereas 30% is needed for 2 °C and 55% is needed for 1.5 °C' (UNEP 2021).

Global methane emissions continue to climb, with livestock CH_4 emissions alone increasing by 10-18 Gg CH_4/yr in 2000-2018. Most (73%) of the increase was due to increased livestock numbers and 27% due to increased emission/animal (Chang *et al.* 2021). Livestock CH_4 emissions are forecast to rise to 140 Gg CH_4 by 2050 under current policies but scope for 10 Mt/year of enteric CH_4 mitigation has been identified by Smith *et al.* (2021). There is no requirement to eliminate livestock methane emission for the Paris agreement temperature targets to be achieved. While net zero for CO_2 is pivotal, modelling of future mitigation pathways anticipates that the 1.5 °C Paris target is possible through zero net CO_2 with only a 38% reduction in livestock CH_4 emissions (Popp *et al.* 2017; Reisinger *et a*l. 2021). This reflects the shorter atmospheric lifetime of methane, meaning the warming due to past CH_4 emission diminishes quicker than it does for CO_2. This is important as it identifies scope for sustaining ruminant livestock as a basis of global food security despite a net zero CO_2 target.

Goals for the developing and developed world. Climate action is only one of the 17 Sustainable Development Goals (SDGs) and other goals may take priority. Developing countries face the greatest human population growth and associated food security threat. The 2020 global population of 7.80 billion people is forecast to rise to 8.6 billion by 2030 and 9.8 billion by 2050. Of the additional 2 billion people

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_19, © R. Hegarty *et al.* 2022

joining humanity between 2020 and 2050, 1.15 billion will be in Africa and 0.65 billion in Asia (UNDESA 2019). Hence, for most of the developing countries, adaptation to climate change outranks mitigation, and cross-cutting actions delivering on numerous SDGs, not just climate are desired.

The developing countries do not have the same obligation for mitigation as the developed countries. Article 4 of the Paris Agreement says 'developed country parties should continue taking the lead by undertaking economy-wide absolute emission reduction targets', while developing countries have their limited capability and need for support recognised. Of the 192 countries submitting NDCs, only 34% included livestock mitigation in their commitments (Rose *et al.* 2021). These authors grouped livestock commitments from all NDC submissions into 6 themes that could achieve mitigation and/or adaptation. These were: animal health, breed management, feed management, herd composition measurement, manure measurement, and silvopastoralism. This is a very loose coalition of strategies, but prospects for each of the leading animal mitigations are assessed below. For brevity and because enteric emissions are the largest livestock GHG source, comments are limited to enteric CH_4.

Direct enteric methane mitigation

Dietary interventions that can specifically inhibit methane production offer direct and readily quantifiable means to mitigate enteric emissions. Feed additives have been extensively reviewed (Smith *et al.* 2021) but a recent broader assessment identified that almost all research has been done using diets in which the additive was mixed in the compete ration. Consequently, after extensive testing and reviewing of methane mitigating feed additives, we are still in a stunningly poor position, with scant knowledge of inhibitor efficacy in grazing systems or which supplement system and frequency will be optimal (Hegarty *et al.* 2021).

Vaccination to harness the animal's immune system is the other direct means of mitigation. Factors affecting immunisation effectiveness in general and relevant to rumen methanogens have recently been reviewed (Baca-González *et al.* 2020) but mitigation results have been sparse since first results were published (Wright *et al.* 2004). A commercial vaccine is considered to be 7-10 years away (Reisinger *et al.* 2021) and despite some demonstrations of efficacy of salivary antibodies in-culture, repeatable mitigation *in vivo* has not been reported.

Animal management and efficiency in the farming system

There are multiple indirect ways of affecting daily methane production, from suppressing rumen protozoa to changing the proportions of starch and of oil in the diet, including breeding genetically modified ryegrass with increased lipid content (Winichayakul *et al.* 2020; Honan *et al.* 2021). Many of these affect fermentation products, feed intake and animal performance, enabling major shifts in livestock enterprise structure. Changes in nutrient supply can change maternal body condition and so reproductive performance, reduce the time to slaughter and the number of replacement females needing to be kept. It is these 'system-based' changes that offer most to the SDGs of developing countries as social, food production and emission outcomes change together. The opportunity to improve efficiency of conversion of feed to animal product (protein) on the global scale was highlighted by Herrero *et al.* (2013) reinforcing the opportunity for improved efficiency to underpin mitigation. Improving feed use efficiency by management or animal breeding not only reduces the emission intensity of product, but also provides farmers the flexibility to maintain production while running less stock, using less land area and so potentially running other enterprises.

Breeding for low methane livestock

The only strategy that has been implemented on a national basis and is applicable across all livestock systems is breeding (Rowe *et al.* 2021). Differences between individual animals in the amount of methane emitted in absolute terms, per unit of production or per unit of feed intake, occur across sheep, beef and dairy cattle. Reported heritabilities for absolute methane emissions (g/day) and methane yield (g CH_4/kg DMI) range from 0.1 to 0.3 (Brito *et al.* 2018). Furthermore, in New Zealand, sheep have been divergently selected for methane yield (g CH_4/kg DMI) for over 3 generations resulting in an average of 12% difference between the lines and up to 25% difference in methane yield amongst young sires. Rowe *et al.* (2019) showed sheep selected for lower methane yield were more profitable than their high emitting counterparts. On this basis, the breeding scheme has been expanded to the NZ national flock.

To date, the inclusion of methane within a selection index has been limited by its unknown economic value and a lack of information on genetic correlations with key production traits. However, the scope for lowering methane by including direct and indirect measures within current breeding goals is emerging. In a large dairy cattle study spanning multiple countries and systems, methane traits corrected for either productivity or intake were moderately correlated and inclusion within the breeding goal resulted in a selection response for lowered methane whilst improving feed efficiency and milk yield (Manzanilla-Pech *et al.* 2021). Authors noted that breeding for increased feed efficiency elicited a favourable correlated response so selection for efficiency could also result in decreased methane production. Conversely Johnson *et al.* (2018) reported an unfavourable correlation between residual feed efficiency and methane in sheep, suggesting that including feed efficiency in the index alone would not reduce methane emissions. De Haas *et al.* (2021) estimated that a putative reduction of 24% in methane intensity (g CH_4/kg milk) could be achieved by 2050 by including feed efficiency in the Dutch national dairy breeding goal. Taken together this is solid evidence that methane emissions can be lowered by breeding when included into an optimised selection index, but it is still unclear what metrics should be used. Rowe *et al.* (2021) described the incorporation of methane breeding values within the New Zealand national sheep breeding scheme giving a rate of genetic change of -0.58% per annum. The selection pressure required for this rate of gain would enable gains in productivity to continue. They showed that even with a lag phase of 6 years, that after 20 years, annual methane production was predicted to have reduced by 7.5% per annum saving a total of 4,490 kt of CO_2e. Whilst this will not meet the current targets set by global governments, it does offer substantial reduction for little input and remains the most cost effective and practical solution proposed to date.

A major issue for progress and uptake is that the value of breeding for lowered methane has yet to be determined. The true cost of emissions at the farm gate will determine the level of trade-offs required with improved production and efficiency. This approach necessitates either explicitly or implicitly placing an economic value on methane emissions that is related to their contribution to global warming. Metrics such as GWP100 and GWP* vary greatly in estimating the impact of methane emissions over the short- and long-term and the associated warming equivalents. Both the method of calculation (Cain *et al.* 2022) and future costs vary over an order of magnitude. It is unlikely therefore that the solution lies in a traditional bioeconomic model, and some form of desired gains index is likely to be preferable.

Although, livestock breeding alone is highly unlikely to deliver substantial mitigation towards Paris Agreement warming targets, it could play a substantial role both in reduction of livestock emissions and in adaptation. This will initially be in developed countries, but breeding is also consistent with SDG priorities of developing nations. Our expectation is that breeding will combine with other dietary and system-based mitigation approaches that are still being prepared for implementation. Together these can deliver decreases in absolute emissions and emission intensity and facilitate more efficient food and fibre production.

Acknowledgements

Financial support was provided by the New Zealand Agricultural Greenhouse Gas Research Centre (NZAGRC).

References

Baca-González, V., Asensio-Calavia, P., González-Acosta, S., Pérez de la Lastra, J.M., Morales de la Nuez, A., (2020) Vaccines, 8(3):460.

Brito, L.F., Schenkel S.F., Oliveira H., Cánovas A., Miglior F. (2018) WCGALP 11: 740

Cain M, Jenkins S, Allen MR, Lynch J, *et al.* 2022. Phil. Trans. Royal Soc. A.380 (2215):20200456.

Chang, J., Peng, S., Yin, Y., Ciais, P., *et al.*, (2021) AGU Advances 2(2), p.e2021AV000391.

De Haas Y., Veerkamp R.F., de Jong G., Aldridge M.N. (2021) Animal 15, Supplement 1,

Hegarty, R., Passetti, R., Dittmer, K., Wang *et al*., (2021) https://hdl.handle.net/10568/116489

Herrero, M., Havlík, P., Valin, H., Notenbaert, A., *et al.*, (2013) Proc.Nat. Acad. Sci. 110, 20888-20893.

Honan M, Feng X, Tricarico JM, Kebreab E. (2021) Anim. Prod. Sci. https://doi.org/10.1071 /AN20295

Johnson, P.L., Knowler, K., Wing, J., Hickey, S., Johnstone, P. (2018). WCGALP 11: 608.

Manzanilla-Pech C.I.V., Lovendahl P., Mansan Gordo D., *et al.* (2021) J. Dairy Sci. 104:8

UNFCCC 2015. Paris Agreement to the United Nations Framework Convention on Climate Change,. https://unfccc.int/sites/default/files/english_paris_agreement.pdf viewed 12/01/2022.

Popp, A., Calvin, K., Fujimori, S., Havlik, P. *et al.* (2017). Global Envir. Change, 42: 331-345.

Reisinger, A., Clark, H., Cowie, A.L., Emmet-Booth, J. *et al.*, (2021) Philos. Transact Royal Soc. A, 379(2210), p.20200452.

Roe, S., Streck, C., Beach, R., Busch, J., *et al.* (2021) Global Change Biology, 27: 6025-6058.

Rose S, Khatri-Chhetri A, Stier M, Wilkes A, *et al.*, (2021) https://cgspace.cgiar.org/bitstream/handle/10568/115885/CCAFS%20Info%20Note%20Livestock%202021%20NDCs.pdf viewed 12/01/2022.

Rowe S.J., Hickey S.M., Jonker A., Hess M.K., Janssen P., Johnson *et al.*, (2019) Proc. Assoc. Advmt. Anim. Breed. Genet. 23: 306.

Rowe S.J., Hickey S.M., Johnson P.L., Bilton T.P., *et al.* (2021) Proc. Assoc. Advmt. Anim. Breed. Genet. 23:306-309.

Smith, P., Reay, D. and Smith, J., (2021) Phil. Trans. Royal Soc. A: Mathematical, Physical & Engineering Sciences. 379 (2210), art. no. 20200451

UNDESA 2019 https://population.un.org/wpp/Download/Standard/Population viewed 12/1/22

UNEP (2021). https://www.unep.org/resources/emissions-gap-report-2021 viewed 12/1/2022

UNFCCC (2021) https://unfccc.int/documents/306848 viewed 12/1/2022

Winichayakul S, Beechey-Gradwell Z, Muetzel S, Molano *et al.* J Dairy Sci. (2020) 103(3):2405-2418.

Wright, A.D., Kennedy, P., O'Neill, C.J., Toovey, A.F *et al.*, (2004) Vaccine, 22: 3976-3985.

20. Genetically more efficient Australian dairy cows and sheep are higher emitters of methane per unit of food

B.J. Sepulveda[1,2*], S.K. Muir[3], S. Bolormaa[1], I.M. MacLeod[1], M.I. Knight[3], R. Behrendt[3], L.C. Marett[5,6], M.H. Deighton[4], J.B. Garner[5], P.J. Moate[5], W.J. Wales[5,6], R.O. Williams[5], H.D. Daetwyler[1,2], B.G. Cocks[1,2] and J.E. Pryce[1,2]

[1]Agriculture Victoria, AgriBio, Centre for AgriBioscience, Bundoora, VIC, 3083 Australia; [2]School of Applied Systems Biology, La Trobe University, Bundoora, VIC, 3083 Australia; [3]Agriculture Victoria, Hamilton Centre, Hamilton, VIC, 3300, Australia; [4]Cropmark Seeds Ltd. Canterbury, 7677, New Zealand; [5]Agriculture Victoria, Ellinbank, VIC, 3821, Australia; [6]Centre for Agricultural Innovation, School of Agriculture and Food, Faculty of Veterinary and Agricultural Sciences, The University of Melbourne, VIC, Australia; boris.sepulveda@agriculture.vic.gov.au

Abstract

Both the dairy cattle and sheep industries face the simultaneous challenge of improving feed efficiency and reducing methane emissions. Genomic selection is a valuable tool to reduce residual feed intake (RFI) and reduce methane yield (MeY), which are widely used traits for estimating efficiency and emissions. However, it is important to know how selecting one of these traits would affect the other and this relationship has been contentious in the literature. Here we estimated the genetic correlations between RFI and MeY in 584 Holstein dairy cattle and 445 Australian Maternal Composite ewes using bi-variate genomic best linear unbiased prediction models. In both datasets, negative genetic correlations between RFI and MeY were found, which means that selecting more feed efficient animals would increase the amount of methane emitted per kg of dry matter intake. Diet could play a role in this relationship.

Introduction

Feed is the main production cost for both dairy cattle and sheep production and improving feed efficiency is of economic importance to livestock producers. Residual feed intake (RFI) is an accepted measurement of feed efficiency obtained as the difference between actual and predicted dry matter intake (DMI) required for maintenance, growth, and production. Methane is a greenhouse gas emitted as a by-product of enteric fermentation in ruminants. Reducing methane emissions from dairy cattle and sheep is an important step to achieving the Australian goal of reducing greenhouse gas emissions by 26 to 28% from 2005 levels by 2030. On top of the environmental benefits, reducing methane emissions could increase productivity because methane represents a waste of energy that could be used for production.

Methane yield (MeY) is a methane emission trait calculated as the ratio of methane production (MeP) and dry matter intake (DMI). Both RFI and MeY are time-consuming or expensive to measure traits, making routine recording on commercial farms unviable. For these reasons, and given their socio-economic importance, RFI and MeY are ideal traits to be improved with genomic selection. However, it is necessary to know how selecting one of these traits would affect the other. Our hypothesis was that as methane emissions represent a waste of energy, more efficient sheep and cows were expected to be lower emitters of methane per kg of DMI (MeY).

Materials & methods

Animals. Feed intake and methane emissions were collected on 584 Holstein dairy cattle from 12 herd-year-season (HYS) cohorts between 2013 and 2017 and 445 Australian Maternal Composite ewes between 2013 and 2014. Cows were located at the Ellinbank SmartFarm (Ellinbank, Victoria, Australia), milked twice per day, and were between 55 and 127 (71±26) days in milk (DIM) and across parities 1 to 9. Daily

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_20, © B.J. Sepulveda *et al.* 2022

milk production was recorded in cows, and milk protein and fat were calculated using midinfrared spectroscopy. Energy corrected milk (ECM; kg/d) was calculated with the methodology of Visscher et al. (1994) as follows:

$$ECM = 0.1 \times \text{milk yield (kg)} + 5.2 \times \text{fat (\%)} + 2.6 \times \text{protein (\%)} \quad (1)$$

Ewes were located at the Hamilton SmartFarm (Hamilton, Victoria, Australia), were not in late pregnancy or lactating, and their phenotypes were measured at post-weaning, hogget, and adult age stages.

Phenotypes. Dairy cows were fed *ad libitum* for 3 hours after morning and afternoon milking with alfalfa hay or cubes, after which they were moved to an area without food but with water available *ad libitum*. One feeding bin was allocated per one or two cows. Additionally, cows consumed 6.2±0.25 kg of dry matter per day of crushed wheat in 2 portions at milking. The alfalfa hay and cubes provided to cows contained 8.7 and 10.2 MJ of ME/kg of dry matter (DM), 18.2 and 22.7% crude protein (CP), and 46.3 and 37.3% Neutral detergent fibre (NDF), respectively. The crushed wheat fed to cows had 13.8 and 13.8 MJ of ME/kg of DM, 13.1 and 13.9% CP, and 10.2 and 14.3% NDF, respectively. Ewes were fed *ad libitum* (24 hours per day) with a cereal straw-based pelleted diet for 10-14 days before recording feed intake. The pellets provided to ewes had 9.6±0.58 MJ of ME/kg of DM, 9.8±1.6% CP, 48±3.18% NFD. Ewes were spread across 10 pens, with less than 24 ewes per pen and two automated feeders. Pen number was constant within all experiments, but ewes could be assigned to different pens at different ages. Feed intake was measured between 32 and 37 days in cows and for a total of 53, 42, and 32 days for post-weaning, hogget, and adult ages in ewes, respectively. Ewes were weighed three times a week to calculate the average daily gain (ADG). RFI of cows was estimated by Bolormaa *et al.* (2021) as

$$RFI_{cows} = DMI - (\text{mean} + b_1 \text{ parity} + b_2 \text{ DIM} + b_3 \text{ HYS} + b_4 \text{ poly(age,-2)} + b_5 \text{ diet} + b_6 \text{ ECM} + b_7 \text{ BWT} + b_8 \Delta\text{BWT}), \quad (2)$$

where $b_1 - b_8$ are partial regression coefficients; BWT is mean body weight in the test; ΔBWT daily BWT change. RFI of ewes was calculated as

$$RFI_{ewes} = DMI - (\text{mean} + b_1 \text{ MMWT} + b_2 \text{ ADG} + b_3 \text{ EXP} + b_4 \text{ PEN:EXP} + b_5 \text{ AGE}) \quad (3)$$

where $b_1 - b_5$ are partial regression coefficients; MMWT is the metabolic mid-test weight obtained as the average weight of the animal during the feeding test to the power of 0.75; EXP is the feed intake experiment (n=5); PEN:EXP is the pen nested into the experiment (45 levels), and AGE is the age in days at the start of the experiment. RFI was the residual error of the expression, i.e. additional consumption after accounting for energy sinks, which were maintenance (MMWT) and growth (ADG).

Methane production (MeP; L/d) in cows was measured over 5 days with the sulphur hexafluoride (SF_6) tracer method described by Deighton *et al.* (2014) and averaged to obtain one observation per animal. Methane emissions were measured twice in ewes during each feed intake experiment with the portable accumulation chamber (PAC) technique described by Goopy *et al.* (2011). In the sheep population, the first methane measurement was obtained approximately four weeks after the start of the experiment and the second one approximately ten days later after the first measurement, and MeP (L/d) was estimated. MeY was calculated as MeP divided into DMI for both cattle and sheep.

Genotypes. The dairy cattle animals were genotyped for 54,609 single nucleotide polymorphism (SNP) from the BovineSNP50k v2 BeadChip (Illumina Inc., San Diego, CA). These SNP were mapped to the ARS-UDC1.2 bovine genome sequence assembly. SNP with call rate smaller than 0.9, minor allele frequency smaller than 0.025, and duplicate map positions were removed. After this filter, 36,951 SNP were retained for cattle. Ewes were genotyped with one of three chips – with 12,785, 15,000 or 54,241 SNP – and imputed with FImpute to 46,091 SNPs. Genomic relationship matrices (G) were created with the function Gmatrix of the R package AGHmatrix (Amadeu *et al.* 2016) following the method of Yang *et al.* (2010), and principal component analyses on G were completed. All animals had a genotype call rate greater than 0.9.

Genetic correlations. Genetic correlations between RFI and MeY were obtained with the R package ASReml-R (Butler *et al.* 2009) with bi-variate GBLUP models of the form

$$y = 1\mu + X\beta + Zg + e \quad (4)$$

where y is a vector of phenotypes; μ is the overall mean; β is a vector of fixed effects; g is a vector of random additive genetic effects; X and Z are the incidence matrices for β and g, respectively; and e is a vector of random residuals distributed as $N(0,I\sigma_e^2)$. The distribution of g is $N(0,G\sigma_g^2)$, where G is the genomic relationship matrix calculated from SNP genotypes following Yang *et al.* (2010). The fixed effects in cows were HYS, DIM, age and parity, and in sheep were EXP, PEN:EXP, and AGE.

Results

RFI had means of zero and standard deviations of 1.01 and 0.27 in dairy cows and sheep, respectively (Table 1). MeY means were 19.96±3.61 and 25.08±7.14 for cattle and sheep, respectively. In MeY, the range was wider, and the coefficients of variation and maximum values were lower in dairy cattle than in sheep. The coefficients of variation were 18.09 and 28.48, and maximum values were 32.89 and 54.19 for cattle and sheep, respectively. The coefficient of determination (R^2) of the observed DMI regressed on the predicted DMI in sheep was 0.79. The first two principal components of the G matrices together explained 11 and 12% of the variation in cattle and sheep; 14 and 3 distant animals in cattle and sheep were removed, respectively. The genetic correlations between RFI and MeY were strong and reasonably consistent and were -0.77±0.15 and -0.86±0.13 for dairy cattle and sheep, respectively.

Discussion

The range of phenotypic values of RFI and MeY and the coefficient of variation of MeY show variability in the dairy cattle and sheep populations, which can be used for the genetic improvement of these traits. The percentages of the G matrices explained by the first two principal components showed that population structures in the populations were minor.

Table 1. Descriptive statistics of residual feed intake (RFI) and methane yield (MeY) in dairy cattle and sheep.

Species	Trait	Animals	Mean	SD	Min	Max	CV
Dairy cattle	RFI	584	0.00	1.01	-3.49	4.69	NA
	MeY	424	19.96	3.61	11.11	32.89	18.09
Sheep	RFI	443	0.00	0.27	-1.08	1.05	NA
	MeY	458	25.08	7.14	6.72	54.19	28.47

The results from this study show that the selection of ruminants for improved feed efficiency (i.e. lower RFI), would lead to higher methane emissions per kg of DMI. This is somewhat unexpected, as it was hypothesised that lower RFI (more efficient animals) would have decreased MeY, as methane represents energy that could otherwise be used for production. In accordance with this approach, studies in dairy cattle and sheep fed with lower fibre and higher protein diets than our research found that selecting for low RFI would decrease MeY (Negussie *et al.* 2014; Paganoni *et al.* 2017). In contrast with these studies and with diets with lower protein and higher fibre, we found that MeY was higher for low RFI in dairy cattle and sheep. Similar to our results, heifers with low RFI had higher MeY values under a high-forage diet (Flay *et al.* 2019). Additionally, an inverse relationship between DMI and MeY has been reported in ruminants fed with forage diets (Hammond *et al.* 2013) and RFI as a portion of DMI could be expected as negatively correlated with MeY. The underlying biology of the genetic correlations between RFI and MeY is not clear yet and requires further research to elucidate.

References

Amadeu R.R., Cellon C., Olmstead J.W., Garcia A.A., Resende M.F. *et al.* (2016) Plant Genome 9(3):1-10. https://doi.org/10.3835/plantgenome2016.01.0009

Bolormaa S., Macleod I.M., Khansefid M., Marett L., Wales B. *et al.* (2021) Proc. of the 24th AAABG, Adelaide, Australia.

Butler D., Cullis B.R., Gilmour A. and Gogel B. (2009) ASReml-R Reference Manual. The State of Queensland, Department of Primary Industries and Fisheries.

Deighton M.H., Williams S.R.O., Hannah M.C., Eckard R.J., Boland T.M. *et al.* (2014) Anim. Feed Sci. Technol. 197:47-63. https://doi.org/10.1016/j.anifeedsci.2014.08.003

Flay H.E., Kuhn-Sherlock B., Macdonald K.A., Camara M., Lopez-Villalobos N. *et al.* (2019) J. Dairy Sci. 102(3):2708-2713. https://doi.org/10.3168/jds.2018-15234

Goopy J.P., Woodgate R., Donaldson A., Robinson D.L. and Hegarty R. (2011) Anim. Feed Sci. Technol. 166-167:219-226. https://doi.org/10.1016/j.anifeedsci.2011.04.012

Hammond K.J., Burke J.L., Koolaard J.P., Muetzel S., Pinares-Patiño C.S. *et al.* (2013) Anim. Feed Sci. Technol. 179(1):121-132. https://doi.org/10.1016/j.anifeedsci.2012.11.004

Negussie E., Mäntysaari P., Mäntysaari E.A. and Lidauer M.H. (2014) Proc. of the 10th WCGALP, Vancouver, Canada.

Paganoni B., Rose G., Macleay C., Jones C., Brown D. *et al.* (2017) J. Anim. Sci. 95(9):3839-3850. https://doi.org/10.2527/jas2017.1499

Visscher P.M., Bowman P.J. and Goddard M.E. (1994) Livestock Production Science 40(2):123-137. https://doi.org/10.1016/0301-6226(94)90042-6

Yang J., Benyamin B., Mcevoy B.P., Gordon S., Henders A.K. *et al.* (2010) Nat. Genet. 42(7):565-569. https://doi.org/10.1038/ng.608

21. The role of livestock genetics in addressing national GHG mitigation requirements

A.R. Cromie[1*], D.P. Berry[2], L. Shalloo[2], T. Kirk[3] and J. Crowley[3]

[1]*Irish Cattle Breeding Federation, Highfield House, Shinagh, Bandon, Co. Cork, P72 X050, Ireland;*
[2]*Animal & Grassland Research and Innovation Centre, Teagasc, Moorepark, Fermoy, Co.Cork, Ireland;*
[3]*AbacusBio Ltd., 442 Moray Pl., Dunedin 9016, New Zealand; acromie@icbf.com*

Abstract

Genetic improvement has the potential to mitigate some 1-1.5 MT of GHG from the Irish national cattle herd, over the next 10 years. This represents approximately 25% of the target for the sector, as defined by the Irish government within its National Climate Action Plan. The reductions will come through a combination of current genetic trends, coupled with innovative new approaches such as: (1) increased weighting on female fertility traits in our indexes; (2) selecting directly on new methane EBV's in future indexes; and (3) genotyping the national cattle herd. Given the major climate and environmental challenges that lie ahead, we are fortunate to have such a proven technology at our disposal.

Introduction

Irish agriculture is required to reduce its Greenhouse Gas (GHG) output from agriculture by 4.5-6.5 million ton by 2030; this represents a 21-30% reduction in gross emissions for the sector. Over the next number of months, these commitments will become part of a legally binding National Climate Action Plan, with each sector (e.g. agriculture, transport, construction) then having its own sectoral targets. These requirements are as a result of our country's commitment to global reductions in GHG output, through the United Nations Framework Convention on Climate Change (UNFCCC), the Paris Agreement, and the resultant EU Effort Sharing Decisions and Regulations.

Genetic improvement is well known to be permanent, sustainable, cumulative and highly cost effective with an excellent penetrance rate. As a result, it is a key (and proven!) vehicle to achieve year-on-year cumulative improvement in the profitability and sustainability of livestock (Simm, 2002). More recently, questions have been posed about the potential of genetic selection to address national GHG requirements, in line with the national, EU and global commitments. Indeed, a recently published communication from the European Commission on EU strategy to reduce methane emissions, has identified genetic improvement as a key tool through which member countries can reduce their GHG output nationally. As a result, the Irish government has asked the Irish Cattle Breeding Federation (ICBF), the organisation charged with responsibility for national genetic improvement in cattle in Ireland, to undertake an assessment of the potential role of genetic improvement in addressing our ambitious national GHG mitigation requirements.

The mitigation strategy presented here for cattle is also of global relevance, as other countries signed up to the Paris agreement are similarly being charged with the same effort sharing requirements; some 192 states represent 98% of the total global GHG output. It is therefore anticipated that the strategy proposed here will be of interest to both national and international audiences, as we collectively strive to reduce global GHG outputs from agriculture.

Materials & methods

Scenarios considered. Cattle genetic improvement strategies in Ireland are currently focused on maximizing profitability. This approach is also consistent with ongoing improvement in emissions intensity (EI, CO_2e/kg product) which are highly valued by the consumers of Irish livestock products, many of whom

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_21, © A.R. Cromie *et al.* 2022

are increasingly conscious of the GHG footprint associated with their diet and lifestyle choices. Within the analysis, the impact of current genetic trends on both emissions intensity and on gross emissions (in CO_2e) based on either a fixed level of production or a fixed population size (i.e. increased production) were investigated. This is critical, as traditionally genetic improvement has been viewed to deliver sustainability improvements through an emission intensity perspective, rather than gross emissions, with the latter now being the particular requirement of National and International GHG inventories.

Strategies applied. Various additional breeding strategies were also considered as part of the analysis, in an effort to identify options to yield the greatest GHG mitigation in the future. Each strategy was considered additive; so building on the current level of genetic gain within the National dairy and beef herds and then layering over additional enhancements, based on relevant national data, parameters and assumptions.

The scenarios considered, together with their impact under a fixed national output or a stable national herd size are summarised in Table 1 for the national dairy herd and Table 2 for the national beef herd, respectively.

Table 1. Impact of different strategies for GHG mitigation in the Irish dairy herd (KT CO_2e).

	Description	Fixed output (KT CO_2e)[1]	Stable herd (KT CO_2e)[1]
S1	Current genetic trends for dairy economic breeding index	658	69
S2	Addition of a new carbon sub-index within the EBI, with increased weighting on female fertility traits.	96	108
S3	Addition of direct methane EBVs in the breeding goal at the current economic values for methane.	67	77
S4	Increasing emphasis on new methane EBVs in the breeding goal;30% weighting in the overall index.	85	207
S5	Inclusion of a secondary selection requirement; that only top 40% sires on methane EBV are retained for breeding.	97	101
S6	National herd genotyped, resulting in a 20% increase in accuracy of selection.	200	112
	Total	1,203	674

[1] Based on current dairy cow population of 1.6 m dairy cows producing 8.61 M Tonnes milk.

Table 2. Impact of different strategies for GHG mitigation in the Irish beef herd (KT).

	Description	Fixed output (KT CO_2e)[1]	Stable herd (KT CO_2e)[1]
S1	Current genetic trends for beef replacement index	181	71
S2	Inclusion of methane EBV's in the breeding goal at current economic values for methane.	74	78
S3	Increasing emphasis on new methane EBV's in the breeding goal at a 30% weighting in the overall index.	91	110
S4	National herd genotyped, resulting in a 20% increase in accuracy of selection.	69	52
	Total	415	311

[1] Based on a current beef cow population of 0.9 m cows producing 168 KT Meat.

Modelling outcomes. Geneflow models were utilised where appropriate to quantify the benefits of genetic improvement to the overall GHG mitigation over a 10-year period from 2020 to 2030. When production levels were fixed, emissions mitigation were achieved through both reduced cow numbers and improved emissions intensity. Nonetheless, the mitigation achieved through reduced cow numbers was only counted once in the analysis.

Additional detail on the methodologies applied within the analysis is accessible in the final report and recommendations which were presented to the Irish government and cattle breeding industry in late 2021(Crowley *et al.*, 2021).

Results & discussion

Dairy. Results from Table 1 indicate that, in overall terms, genetic improvement has the potential to offset some 1,203 KT CO_2e (1.203 MT) of GHG under a fixed output scenario and 674 KT of GHG under a fixed population size scenario. Underpinning these potential improvements are the current genetic trends in the Irish national dairy cow breeding index (i.e. EBI), which is increasing at a rate of €10.8/year (or 0.19 genetic standard deviations/year), resulting in a reduction in gross emissions of 5.4 kg CO_2e/cow/year). These reductions are largely being driven by shorter calving interval and greater cow longevity, which are more than balancing out the productivity gains in terms of milk output, resulting in a reduction in gross emissions under a stable herd size scenario of 69 KT by 2030 (-0.7% in overall terms). A shorter calving interval in Ireland enables a greater proportion of the diet to be fulfilled from grazed pasture which has a lower carbon footprint than purchased in feeds and/or silage. The greater cow longevity manifests itself as a reduced requirement for replacement females (with their associated carbon cost) but also a greater proportion of the herd realizing their mature yield potential. Capping milk output would result in further reductions in gross emissions (-41 kg CO_2e/year), meaning that by 2030, cows would be producing the same amount of milk solids, but with 410 less kg CO_2e per lactation. Expressing this in overall terms equates to a reduction in CO_2e at the national level of 658 KT (8% reduction in total farmgate GHG).

Examining the various enhancements (i.e. S2, S3, S4, S5 & S6 in Table 1) indicates that each enhancement has the potential to mitigate additional GHG output, through placing further emphasis on female fertility (S2) and being prepared to select more directly and aggressively on the methane phenotypes within the breeding goal and overall breeding program (S3, S4 & S5). The impact of genomics (in terms of genotyping the national cattle herd and increasing selection accuracy) is also noteworthy (S6), especially when one considers the additional benefits that this strategy would bring from an industry standpoint, i.e. enhanced DNA-based traceability, GHG auditing based herd genetic merit, labor saving for farmers, future R&D, etc.

Beef. Results from Table 2 indicate that, in overall terms, genetic improvement has the potential to offset some 415 KT of GHG under a fixed output scenario and 311 KT of GHG in a scenario where the beef cow herd size is fixed. Fewer scenarios were considered for the beef sector due principally to the lesser use of AI in the national beef breeding program (25% in beef compared to 70% in dairy), which makes rapid implementation of any outcomes much more challenging (albeit it will eventually occur, but with a lag). Nevertheless, when combined with the reductions achieved from the dairy sector (Table 1), results indicate that genetic improvement has the potential to mitigate some 1.6 MT of GHG under a fixed output scenario and almost 1 MT under a stable herd scenario; this is from a total national target reduction of 4.5-6.5 MT.

Again, core to the potential beef improvements are the current genetic trends in the Irish national beef breeding index, the €uro-Star Replacement Index, which is increasing at a rate of €5.6/year (or 0.13 genetic standard deviations/year), resulting in a reduction in gross emissions of 7.5 kg CO_2e/cow/year). These reductions are largely being achieved by a lighter cow, lasting longer, resulting in a reduction in gross

emissions under a stable herd scenario of some 71 KT nationally (-2.1% of total beef farmgate GHG). Under a fixed output scenario (where we have the opportunity to capture increased productivity gains), this reduction increases to some 181 KT (-5.4% of total farmgate GHG). Examining further the various enhancements considered (i.e. S2, S3 & S4 from Table 2), the benefits of each are consistent with those from dairy and, again highlight the benefits of new programs and initiatives focused on methane phenotyping and increased genotyping at the national level.

Looking forward. Since the removal of EU milk quotas in 2015, the size of the Irish national cattle herd has grown steadily, from 6.19 m cattle in 2012 to 6.53 m cattle in 2021 (+5.4%). This, coupled with the continued importance of agriculture to the Irish economy (7.1% of total employment and 10% of total exports in value terms), means that there is little desire to impose strict constraints on output and/or cattle numbers. Instead, government, industry and farmers are now looking for innovative breeding solutions to the GHG (and broader environmental) challenges, through new initiatives such as: (1) national genotyping programs; (2) methane phenotype collection systems; and (3) a preparedness to place more emphasis on female fertility and methane traits in our breeding indexes. The costs of these initiatives will be significant, but as will be the prize; 1 MT CO_2e of potential mitigation and some 25% of our sectoral target at a national level. While many strategies are being routinely touted as the savior to reducing the environmental footprint of especially ruminants, no technology has proven itself time-and-time again to be a reliable, sustainable, cumulative and relatively low cost strategy to address grand challenges. We are fortunate to have such a proven technology at our disposal.

Given that climate change is a global issue, the urgency to cooperate internationally to accelerating genetic gain has never been so important. This must be in the form of more data and knowledge sharing, rather than simply building national and/or company specific solutions. Great strides have already been made at research level (e.g. initiatives such as global DMI, 1000 bulls etc) and international service level (e.g. ICAR, Interbull, Interbeef, etc), but the responsibility to grow this further is now compelling. Can we afford to miss this opportunity? The alternative is a future farming environment dominated by plant based proteins, as opposed to one where climate efficient livestock have the opportunity to interact with plants in the natural eco-systems that is required by both to co-exist, evolve and improve. Surely that is a goal worth aspiring to?!

References

Crowley J. *et al.*, (2021). GHG Mitigation Through Genetic Improvement and On-Farm Data Capture. Available at https://www.icbf.com/wp-content/uploads/2021/GHG_Mitigation.pdf

Simm G., (2002). The Genetic Improvement of Cattle and Sheep.

22. Can genomics cope with a 30% reduction of methane emission from livestock in 10 years?

O. González-Recio[1*], J. López-Paredes[2], A. Saborío-Montero[1,3], A. López-García[1], M. Gutiérrez-Rivas[1], I. Goiri[4], R. Atxaerandio[4], E. Ugarte[4], N. Charfeddine[2], J.A. Jiménez-Montero[2] and A. García-Rodríguez[4]

[1]Dpto. Mejora Genética Animal, INIA-CSIC, Ctra La Coruña km 7.5, 28040 Madrid, Spain; [2]CONAFE, Valdemoro, 28340 Madrid, Spain; [3]Centro de Investigación en Nutrición Animal y Escuela de Zootecnia, Universidad de Costa Rica, 11501 San José, Costa Rica; [4]Department of Animal Production, Neiker-BRTA, Campus Agroalimentario de Arkaute, 01192 Arkaute, Spain; gonzalez.oscar@inia.es

Abstract

Livestock will face an important challenge within the next decade to cope with the objective of cut by 30% methane emissions, as agreed in the COP26. This study summarises the latest genetic parameter estimates between methane, dry matter intake, microbiota composition, and production and body traits in Spanish dairy cattle. We evaluated the expected genetic progress after including methane into the breeding goal under different scenarios. Under the current trend in the cow population size, it is only possible to achieve the objective if methane is included with a large weight in the selection index and it is accompanied of other strategies. This may generate conflict with dairy producers and balanced strategies must be considered.

Introduction

As many as 104 countries have promised to cut their methane emissions by at least 30% by 2030 at the 26th session of the Conference of the Parties (CoP26) to the United Nations Framework Convention on Climate Change (UNFCCC) summit in Glasgow, United Kingdom. In 2019, methane in the atmosphere reached record levels, around two-and-a-half times above what they were in the pre-industrial era. Methane has a warming potential between 28-34 times that of CO_2 over a 100-year period. Over a 20-year period it is around 84 times as powerful as carbon dioxide per unit of mass. Ruminants are considered one of the most important sources of global methane (CH_4) emissions, producing around 4,623 Mt of CO_2 equivalents (CO_2e) per year, with dairy cattle being responsible for 2,128 Mt of CO_2e per year. Around 40% of total CH_4 comes from enteric fermentation in ruminants (Moss *et al.*, 2000). Eructated CH_4 is not only an important contributor to global warming, but also represents a loss of dietary energy because it is equivalent to between 2 and 12% of net energy intake (de Haas *et al.* 2011, Johnson and Johnson 1995). Levels of methane emissions are linked to the intake capacity of the animal. Consequently, reducing enteric CH_4 emissions in ruminants has become an important area of research, aligned with the commitment to reduce its greenhouse gases emissions by 2030. The short methane lifetime in the atmosphere (10 yrs) offers an interesting opportunity to reduce the global warming contribution from this gas in a shorter period. Decreasing enteric CH_4 from ruminants without altering animal production is desirable both as a strategy to decarbonize the livestock economy and to improve feed conversion efficiency. Selective breeding is an interesting strategy to tackle CH_4 emissions from ruminants. There is great potential in adopting genetic and genomic selection strategies, given that several studies have reported that enteric CH_4 is a heritable trait (Lassen and Løvendahl, 2016; Pszczola *et al.*, 2019). Enteric methane is mainly produced by microorganisms in the rumen during the feed fermentation. Investigating the ruminal microbiota and understanding their role during the methanogenesis will help to develop better breeding strategies. Besides, microbiota composition has also been reported to be heritable (Saborío-Montero *et al.*, 2020).

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_22, © O. González-Recio *et al.* 2022

The objective of this study is to evaluate the (co)variance parameter estimates between traits related to methane emissions, intake, conformation and production. Then, the genetic changes to reduce methane emissions in the next decade under 3 different selective breeding scenarios were evaluated.

Materials & methods

Data consisted of 11,042 weekly averaged records for dry matter intake (DMI) obtained from 551 dairy cows in 5 farms, as well as 4,624 methane emission measurements using a non-dispersive infrared methane detector installed within the feed bin of the automatic milking system during 14 to 21 days period from 1,501 cows in 14 farms. In addition, ruminal content was extracted from 437 cows with methane phenotypes. Data for milk yield, protein yield, fat yield, fertility, body depth, rump width, chest width and predicted live weight from 11,228 cows were included in the analysis. Heritability and genetic correlations were estimated from bivariate models within a single step framework using AIREML (Misztal *et al.*, 2002).

The large complexity of the ruminal microbiota was aggregated through principal component analysis (PCA), into few principal components (PC) that were used as proxies of the core metagenome. Bivariate animal models were applied using these PC and methane production as phenotypes.

The selection responses under the different scenarios was assessed using selection index theory, including methane production as an additional trait in the breeding goal. We assumed the current selection index (ICO) in the Spanish Holstein population as the benchmark. The other two scenarios included methane with either an economic weight (based on carbon tax) or desired genetic gains (20% reduction in ten years while maintaining positive genetic trends of protein and benefits per cow).

Results

Genetic parameters. Table 1 shows the genetic parameter estimates obtained in this study. The heritability estimates for methane production (0.17±0.05) and methane concentration (0.18±0.04) were moderate, and their genetic correlations with milk yield traits ranged between (-0.05±0.11 and 0.50±0.11). In general, higher genetic correlations were estimated with body capacity traits, ranging from 0.14 to 0.31. The estimated heritability for DMI was moderate (0.16±0.03), with a genetic correlation with milk yield of 0.41 (0.11), and with capacity index of 0.20±0.09. Genetic correlations between methane traits and DMI were positive and ranged from 0.20±0.48 to 0.27±0.43. The heritability estimates for the first PC was >0.30 at all taxonomic levels. The microbiome aggregated variables showed strong genetic correlation with methane traits (0.42±0.19 and 0.83±0.19), and dry matter intake (0.32±0.36).

Table 1. Heritability (diagonal) and genetic correlations (above diagonal) estimates[1] of traits of interest.

Trait	Protein (kg)	Fat (kg)	Milk (kg)	ICAP	DMI	Methane (ppm)	Methane (g/d)	Microbiome (PC1)
Protein (kg)	**0.33**	0.64	0.78	0.29	0.53	0.07	0.13	-0.02
Fat (kg)	0.05	**0.28**	0.87	0.32	0.50	0.30	0.32	0.18
Milk (kg)	0.07	0.08	**0.40**	0.30	0.41	-0.05	0.04	0.04
ICAP	0.01	0.02	0.01	-	0.20	0.31	0.14	0.12
DMI	0.19	0.11	0.11	0.09	**0.16**	0.20	0.27	0.32
Methane (ppm)	0.13	0.14	0.11	0.04	0.48	**0.17**	0.97	0.42
Methane (g/d)	0.12	0.13	0.11	0.07	0.43	0.03	**0.18**	0.83
Microbiome (PC1)	0.31	0.72	0.28	0.04	0.36	0.21	0.13	**0.42**

[1] Standard errors are shown below diagonal.

Expected genetic gain after including methane in the breeding objective. The inclusion of methane in the current ICO would slightly modify the genetic responses, with no relevant change in methane response, which would still increase per cow in future generations. A partial reduction of 20% would only be possible by applying ad-hoc weights for methane in the selection index to achieve such a desired genetic gain (Figure 1). This would come at an expense of a slower genetic progress in milk yield and its components (2 kg of protein per cow), and thus the economic response (5 €/cow/year).

Discussion

The results from our study suggest a correlated response in methane emissions when selecting for more ravenous animals, although these estimates showed large standard errors. These results highlight an important interrelationship between methane production and dry matter intake, with a significant role of the microbiome composition. The core microbiome composition condensed in aggregated PC variables was controlled by the cow genome. The PCA allowed us to condense the huge and fuzzy taxonomical and functional information from the metagenome into a few PC. We selected the PC1 (h^2=0.42±0.05) because it showed the largest association with methane, and separates bacteria from eukaryotes (protozoa and fungi) in the rumen. Protozoa and Fungi provides substrate to methanogenic archaea and their relative abundances represent an important driver of methane emissions (López-Garcia *et al.*, 2022).

Our results suggest that selecting cows that make a more efficient utilization of energy intake is feasible through selection on correlated type traits such as body capacity, stature, and body depth, without compromising high productive performance. Selection for lower methane emissions is expected to have similar results. The inclusion of the microbiome composition may assist on a more efficient selection on lower emissions and higher feed efficiency.

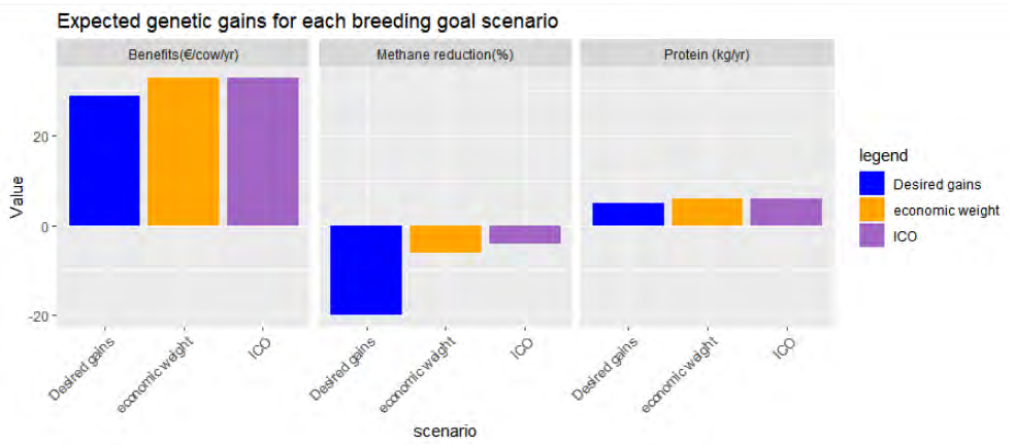


Figure 1. Expected genetic response on benefits, methane emissions, and protein yield according to each selection index scenario (current total merit (ICO) index, economic weight on methane, desired genetic gains) based on the current productive circumstances of Spanish dairy cattle.

Conclusions

The recent Global Methane Pledge aims to limit methane emissions by 30% in 10 years compared with 2020 levels. Selective breeding can contribute to reducing methane emissions from ruminants. Current selection trends will not contribute enough to achieve this goal. Selection indices will need to include methane related traits with weights calculated based on desired genetic gains. This would slow down the genetic gain of other important traits. Even with a slower genetic progress on milk yield traits, selective breeding could achieve a 20% reduction of methane emissions while increasing profitability of dairy cattle.

This study may serve as an example on how selective breeding may be applied to reduce methane emissions in other cattle and small ruminants populations. Selective breeding would contribute in a permanent and cumulative manner to such a reduction. However, other nutritional and technological measurements should be applied complementary to achieve the 30% reduction goal.

References

De Haas Y., Windig J.J., Calus M.P.L., Dijkstra J., de Haan M., *et al.* (2011) J. Dairy Sci. 94 (22118100): 6122-6134. https://doi.org/10.3168/jds.2011-4439.

Johnson K.A. and Johnson D.E. (1995) J. Anim. Sci. 73 (8567486): 2483-2492. https://doi.org/10.2527/1995.7382483x

Moss A.R., Jouany J.P., Newbold J. (2000) Ann. Zootech. 49: 231-253. https://doi.org/10.1051/animres:2000119

Lassen J. and Løvendahl P. (2016) J. Dairy Sci. 99 (26805978): 1959-1967 https://doi.org/10.3168/jds.2015-10012

López-García A., Saborío-Montero A., Gutiérrez-Rivas M., Atxaerandio R., Goiri I., *et al.* (2021) GigaScience. https://doi.org/10.1093/gigascience/giab088

Misztal I., Tsuruta S., Strabel T., Auvray B., Druet T., *et al.* (2002). BLUPF90 and related programs (BGF90). Proc. of 7th world congress on genetics applied to livestock production (Vol. 33, pp. 743-744).

Pszczola M., Calus M.P.L., Strabel T. (2019) J. Dairy Sci. 102 (30928263): 5342-5346. https://doi.org/10.3168/jds.2018-16066.

Saborío-Montero A., Gutiérrez-Rivas M., García-Rodríguez A., Atxaerandio R., Goiri I. *et al.*. (2020) J Anim Breed Genet. 137 (1): 36–48. https://doi.org/10.1111/jbg.12444

23. Low carbon emissions in future pig production – does a new breeding goal matter?

S. Zira[1*], H.F. Olsen[2] and L. Rydhmer[1]

[1]Department of Animal Breeding and Genetics, Swedish University of Agricultural Sciences, SLU Box 7023, 75007 Uppsala, Sweden; [2]Department of Animal and Aquaculture Sciences, Norwegian University of Life Sciences, PO Box 5003 NMBU, N-1432 Ås, Norway; stanley.zira@slu.se

Abstract

Breeding goals are based on the market and production expectations and we expect that the market will require pork with lower if not zero carbon emissions in the future. The influence of the breeding goal on the carbon footprint of pork was studied by comparing the carbon footprint of future scenarios (year 2040) with two different breeding goals; one reflecting the current breeding goal and one reflecting a future breeding goal aiming for increased feed efficiency and higher animal welfare. The diets in both scenarios included locally based feed ingredients such as silage and yeast protein, but no soybean. The carbon footprint of the pork production system was 6.5% lower with the alternative breeding goal including increased feed efficiency and better animal welfare, than with the current breeding goal.

Introduction

Pork is a food product with a high nutrient value and the average carbon footprint of pork today in Sweden is 3.3 kg CO_2 per kg pork retail weight (Landquist *et al.* 2020). Although global pork consumption per capita is not anticipated to increase between now and 2050 (Komarek *et al.* 2021), global pork production will increase with population growth. In contrast to ruminants, pigs cannot utilize large amounts of grass in semi-natural pastures but rely on feed crops. Therefore, an increase in pig production may translate to an increase in the amount of feed crops required. Besides feed-food competition, the increase in feed crop production may increase the absolute global warming impacts from pork and the arable land required for pork production. The climate impact of pig production depends on the feedstuffs used and the ability of the pigs to use feed efficiently.

There is an increasing interest in the use of silage from ley crops as an alternative protein source for pigs to reduce the dependency on soybean meal (Friman *et al.* 2021), and ley crops provide lower climate impacts of the following grain crop in the crop sequence through increasing yields and decreasing fertilizer requirements (Martin *et al.* 2020). A recent study indicated no significant difference in slaughter weights for pigs fed grass/clover silage and those fed diets based on soybean meal (Friman *et al.* 2021). Yeast protein produced from the hydrolysis of spruce trees can also be used to replace soybean meal for young pigs (Cruz *et al.* 2020). Yeast is an interesting protein source because it does not directly compete for land with crop production since it can be produced from forest waste, and grass/clover silage produced on annual leys can improve the crop sequence at pig farms.

Since feed production plays a major role in the climate impact of pig production (Landquist *et al.* 2020; Zira *et al.* 2021), individual feed efficiency is an important breeding goal trait. Feed efficiency on a farm level is also influenced by mortality and reproduction traits and these traits are also important for animal welfare. The aim of our simulation was to show the consequences of a breeding goal with high weight on feed efficiency and traits important for animal welfare on the climate impact of a future pig production system.

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_23, © S. Zira *et al.* 2022

Materials & methods

The diet was the same in both scenarios and included locally based feed ingredients such as grass/clover silage and yeast protein, but no soybean. The pigs had access to an outdoor veranda, silage and straw for enrichment. We compared the climate impact of pig production after 20 years of selection using a 'standard' breeding goal reflecting current breeding schemes or a new breeding goal.

A breeding program was simulated for a synthetic breed representing all three breeds commonly used in crossbred slaughter pigs ((Yorkshire × Landrace) × sire breed). The selection traits and their economic weights are shown in Table 1. The new breeding goal was based on results from studies aiming for improved sustainability in pig production through improved overall farm feed efficiency (Ottosen *et al.* 2020; Rauw *et al.* 2020; Soleimani and Gilbert, 2021) and improved animal welfare (Wallenbeck *et al.* 2016). The initial economic weights were adjusted in an iterative process, so that none of the traits in the new breeding program got an unfavourable trend (in spite of unfavourable genetic correlations).

A mating ratio of 1:40 was used with 2.5% of the boars and 20% for sows being selected to become parents using Best Linear Unbiased Prediction (BLUP) of breeding values simulated in a program i.e. SelAction (Rutten *et al.* 2002). The genetic evaluation of growth rates, feed efficiency, leg strength, and disease resistance was based on information from own performance of all pigs (i.e. the performance test). The genetic evaluation of meat percentage and meat quality was based on information from elder relatives (since selection candidates were not slaughtered). The genetic evaluation of sow traits (fertility, litter size, piglet survival, shoulder ulcers, and longevity) was also based on information from elder relatives. In the SelAction program, this was handled by simulating recording on a number of half-sibs that gives approximately the same accuracy as recording on elder relatives (400 half-sibs for meat percentage and meat quality, and 100 half-sibs for sow traits).

A population of 70 boars and 1000 sows was simulated in the SelAction program. Rounded values of 'standard' genetic parameters for pigs based on the literature were used as input for the simulation (Table 2). All traits were adjusted so that a higher value is more favourable (e.g. frequency of sows free from shoulder

Table 1. Selection traits, average values, standardised genetic standard deviation and relative economic weights used to calculate genetic change for pigs.

Selection traits	Mean 2020	Genetic stand dev	Current breeding goal		New breeding goal	
			Economic weights	Mean 2040	Economic weights	Mean 2040
Litter size, number of live born piglets	14.6	0.32	32.0	16.7	12.0	15.1
Piglet mortality, % of live born	18.0	0.32	-	20.2	8.0	16.7
Interval weaning-service ≤7 days, % of sows	90.0	0.32	-	88.8	7.0	90.4
Sows with shoulder ulcers, %	20.0	0.39	-	20.0	4.0	18.0
Productive life length (sow longevity), d	570	0.32	-	615	1.0	669
Growth rate, weaning to 35 kg, g/d	600	0.45	-	636	1.0	667
Growth rate, 35-120 kg, g/d	834	0.50	18.0	1,070	14.0	1,090
Feed conversion, MJ/kg growth	33.5	0.39	18.0	29.9	15.0	29.4
Leg strength, points 1-5 (5 is best)	3.5	0.45	20.0	4.2	15.0	4.3
Slaughter pigs treated for disease, %	20.0	0.39	-	20.0	7.0	18.0
Leanness, meat percent in carcass, %	58.6	0.71	12.0	62.8	5.5	59.3
Meat quality, drip loss, %	5.0	0.45	-	7.0	10.5	5.0

Table 2. Heritabilities (on the diagonal) and genetic correlations between selection traits (above the diagonal).

Trait	1. Litter size	2. Piglet survival	3. Interv wean-service	4. Shoulder ulcers	5. Sow life length	6. Early growth	7. Growth	8. Feed conv.	9. Legs	10. Pig health	11. Lean meat %	12. Meat quality
1	**0.10**	-0.30	0.0	0.00	0.00	0.00	-0.10	0.00	0.00	0.00	-0.10	0.00
2		**0.04**	0.0	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
3			**0.1**	0.00	0.10	0.00	-0.10	0.00	0.00	0.00	-0.10	0.00
4				**0.20**	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
5					**0.05**	0.00	0.00	0.00	0.20	0.00	0.00	0.00
6						**0.20**	0.30	0.20	0.00	0.00	0.00	0.00
7							**0.25**	0.50	-0.10	0.00	-0.10	-0.30
8								**0.20**	0.00	0.00	0.00	-0.10
9									**0.20**	0.00	0.00	0.00
10										**0.15**	0.00	0.00
11											**0.50**	-0.15
12												**0.16**

ulcers instead of frequency of sows with shoulder ulcers). Since genetic correlations vary between pig populations and over time, we carried out a sensitivity analysis with more unfavourable genetic correlations by adding -0.1 to all correlation estimates.

The carbon footprint method was used to estimate greenhouse gas emissions in the two scenarios. The carbon footprint calculation included emissions from feed production, animal housing, enteric fermentation of the pigs and manure management. The system boundary was from cradle to farm gate and the carbon footprint unit was kg carbon dioxide equivalents per kg retail weight pork.

Results

The pigs selected using the new breeding goal had a lower carbon footprint (2.25 kg CO_2 eq/kg pork) than pigs selected with the current breeding goal (2.40 kg CO_2 eq/kg pork) in 2040. The lower carbon footprint for pigs selected using the new breeding goal resulted from a combination of lower piglet mortality, higher growth rate, higher feed efficiency and higher sow longevity (leading to lower sow replacement rate). Using more unfavourable correlations in the simulation had a very small effect on the carbon footprint (2.42 kg CO_2 eq/kg pork with the current breeding goal). The carbon footprint estimated for 2040 was lower than the carbon footprint of pigs with today's genetic capacity (2020) in the same environment, which was 2.74 kg CO_2 eq/kg pork.

Discussion

Breeding is a useful tool in the reduction of the carbon footprint of pork. Twenty years of selection with the current breeding goal reduced the carbon footprint by 12.4%. A new breeding goal, aiming for improved sustainability in pig production through improved overall farm feed efficiency and animal welfare led to 6.5% lower carbon footprint than the current breeding goal, after twenty years of selection. This may seem like a small value but if a country like Sweden, producing an average of 245,000,000 kg of pork per year (FAOSTAT, 2021), would reduce the emissions by 0.15 kg CO_2 eq/kg pork that would equate to savings of 37,000,000 kg CO_2 eq on its carbon budget. It should be noticed that this reduction is possible while at the same time animal welfare is improved. The future breeding goal matters both for climate mitigation and animal welfare.

References

Cruz, A., Tauson, A.H., Matthiesen, C.F., Mydland, L.T. and Øverland, M. (2020) Livest. Sci. 231:103855 https://doi.org/10.1016/j.livsci.2019.103855.

Food and Agriculture Organisation Statistics (FAOSTAT), Livestock Primary. Available at: http://www.fao.org/faostat/en/#data/QL

Friman, J., Lundh, T. and Presto Åkerfeldt, M. (2021) Acta Agriculturae Scandinavica, Section A – Anim. Sci. 2021:1-10. https://doi.org/10.1080/09064702.2021.1993319

Komarek, A.M., Dunston, S., Enahoro, D., Godfray, H.C.J., Herrero, M., *et al.* (2021) Glob. Environ. Change 70:102343. https://doi.org/10.1016/j.gloenvcha.2021.102343

Landquist, B., Woodhouse, A., Axel-Nilsson, A., Sonesson, U., Elmquist, H., *et al.* (2020) Uppdaterad och utökad livscykelanalys av svensk grisproduktion. RISE, Lund, Sverige.

Martin, G., Durand, J.L., Duru, M., Gastal, F., Julier, B., *et al.* (2020) Agron. Sustain. Dev. 40(3). https://doi.org/10.1007/s13593-020-00620-9

Ottosen, M., Mackenzie, S.G., Wallace, M. and Kyriazakis, I. (2020) Int. J. Life Cycle Assess. 25:523-537. https://doi.org/10.1007/s11367-019-01686-8.

Rauw, W.M., Rydhmer, L., Kyriazakis, I., Øverland. M., Gilbert, H., *et al.* (2020) J. Sci. Food Agric. 100:3575-3586. https://doi.org/10.1002/jsfa.10338.

Rutten, M.J.M., Bijma, P., Woolliams, J.A. and van Arendonk, J.A.M. (2002) J. Heredity 93:456-458. https://doi.org/10.1093/jhered/93.6.456.

Soleimani, T. and Gilbert, H. (2021) Int. J. Life Cycle Assess. 26:455–469. https://doi.org/10.1007/s11367-020-01860-3.

Wallenbeck, A., Rydhmer, L., Röcklinsberg, H., Ljung, M., Strandberg, E., *et al.* (2016) Org. Agric. 6:171 -182. https://doi.org/10.1007/s13165-015-0125-3.

Zira, S., Rydhmer, L., Ivarsson, E., Hoffmann, R., Röös, E. (2021) Sustain. Prod. Consum. 28: 21-38. https://doi.org/10.1016/j.spc.2021.03.028.

24. Individual animal carbon footprints in a national beef database: AniPrint 1.0

J.J. Crowley[1*], L. Shalloo[2], R.D. Evans[3], J. Herron[2], P. Crosson[2], D.P. Berry[2], P. Owens[3] and A.R. Cromie[3]

[1]AbacusBio Ltd., 442 Moray Pl., Dunedin 9016, New Zealand; [2]Animal & Grassland Research and Innovation Centre, Teagasc, Moorepark, Fermoy, Co.Cork, Ireland; [3]Irish Cattle Breeding Federation, Highfield House, Shinagh, Bandon, Co. Cork, P72 X050, Ireland; jcrowley@abacusbio.com

Abstract

Irish agriculture is required to reduce its GHG output from agriculture by 4.5-6.5 million tons by 2030; this represents a 21-30% reduction in gross emissions for the sector. The Irish Cattle Breeding Federation, responsible there for genetic improvement in dairy, beef and sheep have a unique database on which animal carbon footprints can be estimated and extended to Irish farmers and other relevant third parties. Generating GHG totals at the animal levels allows for differences between animal types, systems, sires and other categories be generated. Additionally, procurement, slaughter age and management style can all be influenced by live animal carbon footprints. To this end, for beef, methods to generate per animal and per kg of product GHG totals (kg CO_2e) were developed and are currently deployed in the ICBF database under a module called AniPrint (current v1.0).

Introduction

Owing to the international concern over the nexus of agriculture and climate change, countries have set internal goals and mandates by which the contribution of agriculture to Greenhouse Gas (GHG) totals must be lowered. Irish agriculture is required to reduce its GHG output from agriculture by 4.5-6.5 million tons by 2030; this represents a 21-30% reduction in gross emissions for the sector. Many efforts are currently occurring and being developed simultaneously, ranging from improved manure storage, to improvements in artificial fertilizers, to measuring animals for enteric methane aiding selection decisions (Ryan *et al.*, WCGALP 2022)

The Irish Cattle Breeding Federation (ICBF) is responsible for genetic improvement of dairy, beef and sheep in Ireland and so have a large database of animal centric production data, pedigrees and genotypes. At the core of ICBFs responsibility to Irish farmers, detailed herd reports, summaries, genetic evaluations and mating allocations are some of the main outputs delivered. In addition to genetic improvement mitigating GHG (Quinton *et al.* 2018, Cromie *et al.*, WCGALP 2022), developing individual animal carbon footprint estimates based on animal data in the ICBF database can prove extremely valuable to analyse across industry differences in GHG per animal type, per production system, per sire, etc. Additionally, as data on carbon footprint values/animal can be available in real time, estimates can influence selection of animals for differing management styles, reproduction, purchase, sale or slaughter; both gross GHG and per unit of product produced. To this end, an initial framework to estimate per animal carbon emissions was drawn up and tested for Irish beef animals and is described herein.

Materials & methods

Energy demand of an animal inevitably drives its carbon footprint. Hence, growth profiles of all animals in the database were generated and subsequently, based on growth rate and animal weight, energy demand for growth and maintenance per day (of life) was calculated using equations set out in Rattray *et al.* (2017). The ICBF database contains liveweight data from many sources, including auction marts, farmer submitted records, progeny test centres and research stations. Carcass weights (Cwt) on all animals slaughtered in

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_24, © J.J. Crowley *et al.* 2022

Irish abattoirs are also collated by ICBF for the purposes of genetic evaluation and were available for this exercise. Figure 1. attempts to capture the concept of the stepwise process.

Predicting liveweights. While initially a random regression model was tested, a multiple regression approach was eventually favoured due to the predictive (rather than explanatory) need in this exercise, plus ease of deployment into a database. In the beef genetic evaluation for growth, ICBF adopt a multi-trait evaluation to utilise the plethora of liveweight data available. While only one growth PTA is published, all animals in the genetic evaluation receive solutions for approx. 8 liveweight traits (liveweight at different ages). It's these PTAs, plus other fixed effects such as animal type (steer, heifer, bull) and interactions with age, that are used as predictors of liveweights at certain key timepoints in an animal's life. Coefficients for these fixed effects were estimated by multiple regression on real liveweight observations. Data after 2005 was retained and almost 7M liveweights on 6.5M animals were available. This table of coefficients is now stored in the database and used when liveweight prediction is required. A current improvement to the current version of this tool is currently testing the usefulness of including the BLUE outputs from the weight evaluations which might give a more refined account of environmental effects.

Key time points. The aforementioned table of coefficients is easily deployable into a database calculation as there are key timepoints in a beef animal's life where we estimate a liveweight, and assume a constant gain between them. These key timepoints correspond to typical management and nutrition changes in the average beef animal's life in Ireland. Table 1. highlights these key timepoints. Once liveweights are estimated at these timepoints, we are able to estimate daily energy demand required (MJME) and subsequently estimate daily DMI (kg) depending on where that energy comes from and its density (MJME/ kg DM; Table 1). Diet types differ for spring and Autumn born animals.

Figure 1. Illustration of stepwise concept of estimating GHG per animal.

Table 1. An example of percentage of diet made up of either grass, silage or concentrate, across growth periods for a spring born animal.

Period	Start age (d)	Grass (%)	Silage (%)	Ration (%)
1	0	0	0	0
2	46	0.33	0	0
3	199	0.46	0	0.13
4	230	0.40	0.40	0.20
5	260	0	0.78	0.22
6	322	0.49	0.51	0
7	351	1	0	0
8	596	0.49	0.51	0
9	626	0	0.83	0.17
10	716	1	0	0
11	991	0.49	0.51	0
12	1,022	0	0.83	0.17
13	1,112	1	0	0

Greenhouse gases. Once daily DMI per animal is estimated, an emission factor (EF; g CH_4/kg DMI) is used to calculate daily enteric CH_4 and is subsequently summed across an animal's life. Different EFs are used for different feed types; 0.22 for grazed grass, 0.25 for silage and 0.13 for concentrate. Manure CH_4 is estimated using a Tier 1 IPCC method (Table 10.14, IPCC 2006) as slurry storage information per farm is absent from the ICBF database. This method attributes an average of 16.4 g CH_4/day or life/animal. Methane is converted to CO_2e (as with all GHG in this tool) using a global warming potential (GWP) of 25.

Using estimated lifetime DMI values per animal, direct and indirect N_2O dung and urine emissions can also be estimated using equations outlined in Ch. 11 of IPCC Guidelines for National Greenhouse Gas Inventories. Essentially these estimates will be a function of dietary N consumed, N retained and N excreted. Finally, N_2O from fertiliser use can also be estimated using equations described in the same IPCC chapter. While the ICBF database does not have data on fertiliser attributable to each animal, an estimated allocation per animal is derived based on an animal's nitrogen stocking rate (kg/ha organic N) and the recommended total N (kg/ha) to apply in consideration for that animal (Teagasc Beef Manual, 2016).

Four sources of GHG have been described as they make up approx. 95% of beef GHG. Currently these GHG is what version 1.0 of this tool is concerned with and additional GHG, e.g. CO_2 from off-farm concentrate production, is being considered for inclusion.

All animals and attribution. The aforementioned methods relate primarily to terminal animals, growing replacement heifers and working stock bulls. The cow, post age of first calving (AFC), will also have a carbon footprint that is attributable to the emissions intensity of the final product. For this framework, and as is common in most LCAs methods, a cow's carbon footprint pre AFC is attributable to her cull cow carcass, with all other lifetime emissions attributable to carcass of her progeny. In this version (1.0) of the framework, we assume a beef cow emits 2,800 kg CO_2e/yr (estimate from Teagasc Beef Research, personal communication). Thus, her AFC, age at slaughter and total number of progeny affect her gross emissions and attribution to emissions intensity. For beef out of the dairy herd (dairy-beef or dairy-dairy animals), a lesser amount of carbon from their dam is attributable to the final carcass (271.6 kg CO_2e) as milk solids are the majority of dairy cow product.

Results

When calculations are implemented, all beef animals in the ICBF database receive a lifetime GHG emissions estimate. This estimate is generated for the animal itself, and for the carcass it produces. Estimates are also available for live animals as a 'live ticker' of an animal's footprint. Table 2. is a sample summary of 2020 slaughtered animals. Additionally, Figure 2 provides an estimate of sources of different GHG by age category of beef-beef animals.

It is noteworthy that there is a large amount of variation within (as dictated by the standard deviations) categories described in Table 2. as well as between them. This highlights that there is potential variance to exploit when selecting for whole system carbon footprint.

Discussion

Initial insights into differences across the Irish beef cattle population from AniPrint 1.0, when it comes to GHG emissions and emissions intensity, has already highlighted avenues for GHG mitigation. Additional assessment of animal GHG totals will be carried out to help identify sires of benefit to the national carbon inventory. Third parties, such as abattoirs, also have interest in this tool in order to identify animals of desired finish characteristics along with a favourable carbon footprint. This may drive change at the farmer level to finish animals at an earlier age and focus on AFC, calving interval and cow longevity in their breeding objective.

Table 2. Summary statistics of gross GHG and emissions intensity by key animal groups; Beef-Beef, dairy-beef or dairy-dairy type animals slaughtered in different age category systems of 2020 slaughtered animals.

			End product (carcass)				Direct (terminal) animal			
Type	System	Cwt (kg)	CO_2e (kg)	CO_2e SD	CO_2e/ Cwt	CO_2e/ Cwt SD	CO_2e (kg)	CO_2e SD	CO_2e/kg Lwt	CO_2e/Lwt SD
Beef	20 mnth	338	6,154	1,120	18.6	3.8	3,532	613	5.8	0.4
Beef	24 mnth	351	6,961	1,669	20.2	5.1	4,157	561	6.6	0.4
Beef	30 mnth	374	8,463	1,196	23.0	3.9	5,308	604	7.9	0.5
Beef	30-36 mnth	379	9,538	1,497	25.6	4.6	6,164	908	9.1	0.8
Dairy	20 mnth	307	3,657	492	11.9	0.9	3,385	492	5.6	0.4
Dairy	24 mnth	326	4,297	428	13.3	3.5	4,026	428	6.3	1.5
Dairy	30 mnth	331	5,301	508	16.1	1.4	5,030	508	7.8	0.7
Dairy	30-36 mnth	341	6,098	763	18.0	1.8	5,826	763	8.7	0.9
DairyBf	20 mnth	300	3,511	477	11.8	0.9	3,239	477	5.7	0.4
DairyBf	24 mnth	322	4,178	447	13.1	1.0	3,907	447	6.5	0.5
DairyBf	30 mnth	344	5,313	504	15.6	1.3	5,042	504	7.8	0.6
DairyBf	30-36 mnth	353	6,118	815	17.5	1.8	5,846	815	8.9	0.9

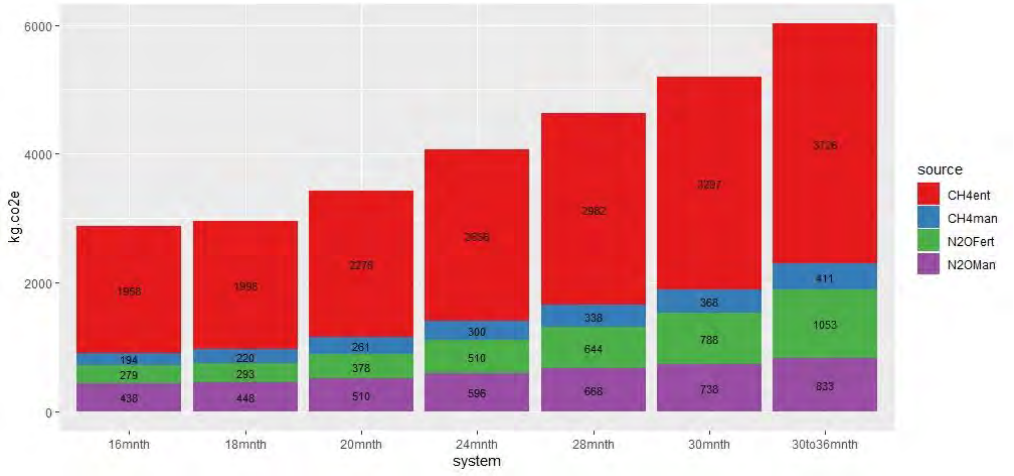


Figure 2. Direct terminal animal GHG sources (enteric CH_4, manure CH_4, artificial fertiliser N_2O & manure N_2O) per age group (system) of 2020 slaughtered animals.

References

Cromie, A.R. *et al.* (2022) Proc. of the 12th WCGALP, Rotterdam, Netherlands. Submitted.

International Panel on Climate Change (2006) 2006 IPCC Guidelines for National Greenhouse Gas Inventories

International Panel on Climate Change (2019) 2019 Refinement to the 2006 IPCC Guidelines for National Greenhouse Gas Inventories.

Quinton C.D. *et al.* (2018) Prediction of effects of beef selection indexes on greenhouse gas emissions. Animal. 12(5):889-897. https://doi.org/10.1017/S1751731117002373

Rattray, P.V., Brooks, I.M, & Nicol, A.M. (2017) Pasture and Supplements for Grazing Animals.

New Zealand Society of Animal Production Occasional Publication No. 14.

Ryan, C.V. *et al.* (2022) Proc. of the 12th WCGALP, Rotterdam, Netherlands. Submitted.

Teagasc Beef Manual (2016) At www.teagasc.ie/publications/2016/teagasc-beef-manual.php

Part 5

Novel traits: environment and greenhouse gas

25. Can ruminal microbial information help improve selection for low-methane emitting dairy cows?

L. Zetouni*, S. Roques, S.K. Kar, Y. de Haas, D. Schokker and M. Aldridge

Wageningen University & Research, P.O. Box 338, 6700 AH Wageningen, the Netherlands; larissa.zetouni@wur.nl

Abstract

Greenhouse gas emissions from livestock have been at the centre of a worldwide discussion. To achieve tangible results in mitigation strategies, all tools in the available toolset need to be used. One tool is the rumen microbiome, which are partly under the host genetics' control, and could help explain variation in methane emissions from ruminants. Data on methane emissions and rumen fluid samples were collected from 179 dairy cows from Dutch commercial herds. Genotypes from 70 cows were used to estimate heritability for methane intensity. Our goal was to assess the power of our pipeline, for subsequent application to a bigger dataset, by evaluating the value of adding microbial information to the analysis. Microbial estimates were highly confounded with the effect of farm. Our preliminary results show that our workflow is appropriate and that we will be able to estimate microbiability and heritability when our full dataset becomes available.

Introduction

The Dutch government has agreed on a set of guidelines, named the National Climate Agreement, with the goal to decrease by 2030 greenhouse gases emissions (GHG) by 49% from the 1990 levels. Five different sectors have been targeted, among them agriculture, where a reduction of 3.5 Mton CO_2eq is expected to be achieved. One source of GHG from agriculture is methane emissions from ruminants, which release methane into the atmosphere as a by-product of their digestion. A community of microorganisms produce, amongst other things, CO_2 and H_2 in the rumen while digesting the feed, which in turn are converted to methane and eructed by cattle (Janssen and Kirs, 2008).

Genetic selection is an important tool for breeders, as it brings cumulative and permanent progress for a given trait of interest (Wall *et al.*, 2010). Methane emissions from dairy cattle are known to be partly under genetic control (Difford *et al.*, 2020; Zetouni *et al.*, 2018), which means selecting for cows that emit less methane is possible, and could therefore be used as a tool to help GHG mitigation in livestock. Another feature that seems to be partly under host control is the vast population of microbes, i.e. the microbiome, found in the rumen (Malmuthuge and Guan, 2017; Roehe *et al.*, 2016). Variation in the relative composition of the microbiome, termed microbiability, has been shown to explain 13% of the variation in methane emissions from dairy cattle (Difford *et al.*, 2018). Therefore, microbial information could be used to help shed some light in understanding the genetic factors behind methane variation in cattle. Our goal was to evaluate the power of our sampling strategy and subsequent pipeline in this preliminary dataset, by evaluating to which extent microbial information could explain the variation in methane intensity from dairy cows from Dutch commercial herds. This paper focuses on preliminary data collection as an indicator on how to optimize sampling of a larger dataset for the estimation of methane microbiability and heritability, and in evaluating and developing our chosen methodology. Our study is part of an ongoing project with the aim to collect methane phenotypes, genotypes and rumen fluid samples from 1000 Dutch dairy cows.

Materials & methods

Data collection. Methane phenotypes and rumen samples were collected from 179 Holstein Friesian dairy cows, from 17 commercial herds spread across the Netherlands. Methane emissions were measured non-invasively using the GreenFeed system (C-Lock Inc., Rapid City, SD) depending on farm. Methane

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_25, © L. Zetouni *et al.* 2022

production per cow was expressed in g/day, which was used to estimate methane intensity (MI), defined here as the ratio between methane production and fat and protein corrected milk (FPCM). Genomic information from 70 cows was incorporated into the model to estimate heritability of MI. Cows were genotyped with the Eurogenomics 10K chip and imputation was routinely performed by CRV to 76,438 SNPs. The rumen fluid samples were collected on the last day of the seven-day methane measurement period using the oral stomach tube according to standard operating procedures developed at WUR (Muizelaar *et al.* 2020). The samples were immediately frozen on dry ice and subsequently stored in the freezer. Thereafter, microbial DNA was extracted, followed by library construction of the hypervariable region V4 (from the 16S rRNA gene). Sequencing was performed on an Illumina HiSeq platform at Genotypic Technology Pvt. Ltd. in Bangalore, India. Reads were pre-processed using QIIME2 suite v2020.8 and FAST QC v 0.11.9. The resulting Amplicon Sequence Variants (ASV) were identified using the SILVA v.138 classifier. Cows with less than 20 methane measurements during the measurement period were discarded due to unreliable methane results. The final dataset contained 48,478 ASVs for 179 samples and an average number of reads per sample of 431,476. The ASVs were rarefied to a uniform sampling depth of 81,771 reads, resulting in 42,799 ASVs. Results from the sequencing of individual rumen fluid samples were used as the microbial phenotype of each cow, in order to estimate microbiability.

Variance components estimation and statistical models. In order to test the power of our sampling strategy and chosen methodology when assessing microbial variance in our preliminary dataset, we first estimated microbiability using only the microbial data on all 179 cows – not including genotypes. A microbial relationship matrix was built with the relative abundance of ASVs as described in Ross *et al.* (2013). The relative abundance (RA_{ij}) of ASVs is considered for the elements $X_{ij} = \log(RA_{ij} - \overline{RA_j})$, where i is the rumen sample and j is the taxonomic unit. From that a metagenomic profile (X) is derived, with dimensions $n \times m$ (n is the number of rumen samples; m is the number of ASVs), which are then used to create the microbial relationship matrix, computed as M = XX'/m.

Microbiability was estimated by the following mixed model equation,

$$y = Xb + Zm + Uc + e \quad (1)$$

where y is a vector of the trait observations; X, Z, and U are incidence matrices associated with the vector of fixed effects b of parity (1 to 4+), grazing type (barn pasture or summer barn feeding), season the microbiome sample was taken (fall, spring, summer or winter), soil type (clay, peat or sand), and days in milk, the vector of ASV effects m ~ N(0, Mσ_m^2), the vector of random farm effects caused by a common environment c ~ N(0, Iσ_c^2); and e ~ (0, Iσ_e^2) is a vector of residuals. The terms σ_m^2, σ_c^2, σ_e^2, is for the microbial, farm, and residual variances respectively, and I is an identity matrix. Following Difford *et al.* (2018), microbiability was estimated as the proportion of phenotypic variance associated with differences in the microbiome, calculated as $\hat{m}^2 = \hat{\sigma}_m^2/(\hat{\sigma}_m^2 + \hat{\sigma}_c^2 + \hat{\sigma}_e^2)$. The proportion of variance caused by the common environment effect of farm was estimated as $\hat{c}^2 = \hat{\sigma}_c^2/(\hat{\sigma}_m^2 + \hat{\sigma}_c^2 + \hat{\sigma}_e^2)$.

For estimation of additive genetic variance for MI, a genomic relationship matrix was computed following the second version of the method proposed by VanRaden (2008) and based on the 70 genotypes available. The following mixed model was used,

$$y = Xb + Wg + Uc + e \quad (2)$$

where terms are the same as in (1), with the exception of W, which is an incidence matrix associated with the vector of genotype effects g ~ N(0, G σ_g^2).

In order to determine if microbial information could help refine the heritability for MI, a third model was evaluated, where both the genomic and microbial relationship matrices were fitted and farm was kept as common environmental effect, as follows,

$$y = Xb + Zm + Wg + Uc + e \quad (3)$$

where the terms are as described for (1) and (2). All variance components and model effects were estimated with ASReml 4.1 (Gilmour *et al.*, 2015). Our main objective behind testing three different models was to investigate how well they would perform on our preliminary dataset. With that, we wanted to assess our the sampling strategy, what could be improved in our workflow, and what lessons we could learn for the 1000 samples we aim to collect.

Due to the limited size of our dataset and the nature of our analysis, our expectation was that our estimates would present large standard errors.

Results

Estimates for MI heritability (h^2), microbiability (m^2) and common environmental effect of farm (c^2) for the models tested are shown in Table 1. Microbiability based on microbial information from all 179 (analysis 1_a) was of 0.10±0.13 (m^2 ± SE) and it was not significant. When we moved on to compute m^2 using data from cows with genotypes (analysis 1_b), the estimated value was zero. In analysis 1_a, c^2 was fairly small, 0.06±0.07 (c^2 ± SE); however, once we limited the model to use only data from cows with genotypes, c^2 increased to 0.20±0.19, while m^2 was zero. Heritability for MI, estimated based on 70 genotypes (analysis 2), was 0.06±0.37 (h^2 ± SE) and it was not significant, while c^2 was 0.19±0.19. When fitting both genomic and microbial relationship matrices (analysis 3), estimates for h^2, m^2 and c^2 were the same as the ones obtained for analysis 1_b and 2.

Discussion

There has been a plethora of literature over the past decade showing that there is a genetic component influencing methane production from cattle, regardless of the methane phenotype chosen (de Haas *et al.*, 2017). In this study, we looked at methane intensity. Heritability was low and not significant. High SE were expected due to the limited size of our preliminary dataset. Using methane data from a combined database, where methane was measured with different equipment but where GreenFeed data was present, Manzanilla-Pech *et al.* (2021) obtained a heritability of 0.38 for methane intensity; therefore, we can speculate that our estimates will improve as our dataset grows, and the lessons we have learned by conducting this study will increase the chances of obtaining significant heritability estimates for MI.

In our study, m^2 estimates were highly confounded with c^2. We believe that due to the limited dataset, the models are not able to properly separate the variance explained by the differences in microbiome from

Table 1. Heritability, microbiability and common environmental effect estimates for the different models[1].

Analysis	Group size	h^2 (SE)	m^2 (SE)	c^2 (SE)
1_a	179	NE	0.10 (0.13)	0.06 (0.07)
1_b	70	NE	0 (0.00)	0.20 (0.19)
2	70	0.06 (0.37)	NE	0.20 (0.20)
3	70	0.06 (0.37)	0 (0.00)	0.20 (0.20)

[1] h^2 = heritability for methane intensity; m^2 = microbiability for methane intensity; c^2 = common environmental effect of farm; SE = standard error; NE = not estimable.

the common environmental effect of farm. Data on the 70 cows with genotypes available came from eight different herds, and the number of cows per herd is not evenly distributed. When looking at m^2 based on 179 cows, the model more efficiently separated the microbial effect from the common environmental effect, indicating that data on more animals could help separating both effects.

With the experience of the confounding between farm and microbiome, we propose sampling on fewer farms and sampling as many cows as possible within a farm (here the number of cows per farm was smaller than the number of farms). Due to GreenFeed having a throughput limitation, we will need a method that allows for large scale phenotyping such as the sniffers, but that will mean compromising on the methane trait used. We were also limited by the ratio of cows that had both microbiome and genotype, therefore farms should be selected based on the proportion of animals in the herd already genotyped.

Conclusions

Even though more genotypes are needed in order to obtain significant h^2 estimates for MI, our results have highlighted the importance of modifying our sampling strategy for the next step of our project. The focus will be on data collection from fewer farms, increasing the number of cows per farm, in order to minimize the effect of c^2 on m^2, therefore improving our estimates of microbial effects. With that, we hope to get a deeper look into how the microbiome can be useful in explaining the variance we observe in methane emissions from dairy cows.

Acknowledgements

The research leading to these results was part of the Climate Approach project funded by the Dutch ministry for Agriculture, Nature and Food Quality (LNV) (project number BO-53-003).

References

De Haas Y., Pszczola M., Soyeurt H., Wall E., and Lassen J. (2017) J Dairy Sci 100:855-870. https://doi.org/10.3168/jds.2016-11246

Difford G.F., Plichta D.R., Løvendahl P., Lassen J., *et al.* (2018) PLoS Genet 14(10): e1007580. https://doi.org/10.1371/journal.pgen.1007580

Difford G.F., Løvendahl P., Veerkamp R.F., Bovenhuis H., *et al.* (2020) J Dairy Sci 103:2442-2459. https://doi.org/10.3168/jds.2019-16966

Gilmour A., Gogel B., Cullis B., Welham S., and Thompson R. (2015) VSN International Ltd, Hemel Hempstead, HP1 1ES, UK. www.vsni.co.uk

Janssen P.H. and Kirs M. (2008) Appl Environ Microbiol 74(12):3619–3625. https://doi.org/10.1128/AEM.02812-07

Malmuthuge N. and Guan L.L. (2017) Microbiome 8:8. https://doi.org/10.1186/s40104-016-0135-3

Manzanilla-Pech C.I.V., Løvendahl P., Gordo D.M., Difford G.F., *et al.* (2021) J Dairy Sci 104:8983–9001. https://doi.org/10.3168/jds.2020-19889

Muizelaar W., Bani P., Kuhla B., Larsen M., *et al.* (2020) Rumen fluid sampling via oral stomach tubing method. In: Mesgaran S.D., Baumont R., Munksgaard L., Humphries

D., Kennedy E., Dijkstra J., Dewhurst R., Ferguson H., Terré M., Kuhla B. (eds) Methods in cattle physiology and behaviour – Recommendations from the SmartCow consortium. PUBLISSO, Cologne, Germany, pp. 1-6.

Roehe R., Dewhurst R.J., Duthie C.A., Rooke J.A., *et al.* (2016) PLoS Genet 12(2):e1005846. https://doi.org/10.1371/journal.pgen.1005846

Ross E.M., Moate P.J., Marett L.C., Cocks B.G., and Hayes B.J. (2013) PLoS ONE 8(9): e73056. https://doi.org/10.1371/journal.pone.0073056

VanRaden P.M. (2008) J. Dairy Sci. 91:4414-4423. https://doi.org/10.3168/jds.2007-0980

Wall E., Simm G., and Moran D. (2010) Animal 4(3):366-376. https://doi.org/10.1017/S175173110999070X

Zetouni L., Kargo M., Norberg E., and Lassen J. (2018) J Dairy Sci 101:2273-2280. https://doi.org/10.3168/jds.2017-13402

26. Phenotypic differences and genetic parameters for methane concentration in BeefxDairy crossbred slaughter calves

K. Johansen[1*], M. Kargo[1], M. Bjerring[2], P. Løvendahl[1] and A.J. Buitenhuis[1]

[1]Aarhus University, Center for Quantitative Genetics and Genomics, Blichers Allé 20, 8830 Tjele, Denmark; [2]Aarhus University, Department of Animal Science, Blichers Allé 20, 8830 Tjele, Denmark; kresten.johansen@qgg.au.dk

Abstract

Genetic selection for reduced methane emission and BeefxDairy crossbreeding are strategies to lower the climate impact of beef. Combined they require knowledge of phenotypes on methane production for BeefxDairy offspring. Slaughter calves sired by Belgian Blue, Charolaise and Angus sires to Holstein dams were compared to purebred Holstein calves. Heritability for methane concentration (CH_4c) were estimated for Belgian blue crosses. Raw phenotypic means showed that CH_4c was lower for calves born after Charolaise and Angus sires compared to Belgian Blue. The heritability of log transformed CH_4c for Belgian blue crosses was estimated to be 0.44 (±0.08), and for log transformed CH_4c adjusted for background CH_4 it was 0.33 (±0.06). This is in line with heritabilities found for log transformed CH_4c in Holstein cows, and will be the basis for further analysis in genetic architecture of methane traits in crossbred slaughter calves.

Introduction

Reducing methane production from the dairy cattle population through genetic selection is seen as one of the ways to reduce the carbon footprint of cattle production. An increasing practice for Danish dairy farmers is insemination of low genetic merit dairy cows with beef semen to increase the value of the slaughter calf and overall income of the farm (Ettema *et al.* 2017). This reduces carbon footprints per kg carcass compared with purebred dairy calves (Mogensen *et al.*, 2015). The beef sires are currently selected on calving ease, youngstock survival and carcass traits (Davis *et al.*, 2020), but new traits could possibly lead to higher efficiency and lower climate impact.

Currently there is a lack of knowledge, whether the sire breed affects methane (CH_4) emission of BeefxDairy offspring, but also if there is variation between sires. Methane concentrations measured relative to CO_2 concentrations in the breath using the sniffer method has a high throughput and has been shown to be heritable in dairy cows (Lassen and Løvendahl, 2016). The aim of this paper is to investigate the distribution of CH_4 concentration in different crosses and estimate the heritability of CH_4 concentration in Belgian Blue (BBL) and Holstein (HOL) crossbred calves.

Materials & methods

Data were collected in the period June 2020 until November 2021. Calves were born on dairy farms in Denmark between July 2019 and June 2021. Six slaughter calf producers purchased these calves at the age of 3-8 weeks, and in these herds the animals were later phenotyped. 2,224 calves were crossbred, either a BBL (n=1,689), Charolaise (CHA) (n=306) or Angus (ANG) bull (n=229) on a Holstein dam. Purebred HOL bull calves (n=429) were also phenotyped, to match the capacity of the system. Calves were raised on a pelleted concentrate ration and offered *ad lib* straw, in five herds. One herd fed total mixed ratio (TMR). Feed intake on pelleted feed and TMR was measured in feed boxes, the feed was weighed as the calf entered and left the feed box. The calf entered the pens between 6 and 10 months of age and stayed in the pen for at least three weeks.

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_26, © K. Johansen *et al.* 2022

CO_2 and CH_4 concentrations were measured in feed boxes using the non-invasive infrared gas sensor (Guardian® NG, Edinburgh Instruments Ltd.). Three feed boxes were available for each pen and neighbouring pens shared a Guardian Sensor. The intake air tube for gas measurements were located in front of the calf when it was eating the feed. Feed intake was registered at every visit, whereas gas concentrations was measured in one of the six feed boxes connected to the Guardian Sensor. During visits with gas recording, CO_2 and CH_4 concentrations were registered every other second. Gas concentrations were adjusted for background concentrations and detection of eructations was made using double exponential smoothing (Difford *et al.* 2016). Gas visits where required to contain at least one eructation with methane concentration >500 ppm and be of length longer than 50 seconds. The average CH_4 concentration (ppm) (CH_4c) of the visit was used for estimating genetic parameters. Twice daily the gas sensors were running a reference measurement to measure background gas concentration, which was used to adjust the methane concentration. The adjusted phenotype (CH_4a) was calculated for all visits using the most recent reference measure. To ensure the residuals were normally distributed, both variables were transformed by the natural logarithm (logCH_4c, logCH_4a), to ensure normal distribution for the genetic parameter analysis.

Genetic analysis. Due to sample size, only phenotypes from BBLxHOL calves were used for estimating genetic parameters. For the variance component estimation, using the average information restricted maximum likelihood algorithm (AI-REML), the statistical software DMU was used (Madsen and Jensen, 2008). The univariate model for estimating genetic parameters for logCH_4c and logCH_4a was the following:

$$y= Xb+Z_1a+Z_2pe+e \quad (1)$$

where y is the vector of phenotypic observations, b is the vector of fixed effects (herd-sensor age in weeks, date of the visit, month-year of birth, sex, feed ration, hour of the day), a is the vector of additive genetic effects, pe is the vector of permanent random effects and e is the vector of random residual effect. The incidence matrix, X, assigns observations to fixed effects, Z_1 assigns the observations to random genetic effects and Z_2 assigns the observations to the permanent environmental effect. The distribution of the random effects was: var(a) = $A\sigma_a^2$, where A is the additive genetic relationship matrix and σ_a^2 is the additive genetic variance; var(pe) = $I\sigma_{pe}^2$ where I is an identity matrix and σ_{pe}^2 is the permanent environmental variance; var(e) = $I\sigma_e^2$ where σ_e^2 is the residual variance. Standard error of the estimates was estimated by calculating asymptotic standard errors. The repeatability was calculated as ratio between the sum of additive variance and permanent effect variance, and the phenotypic variance. The heritability was calculated as the ratio between additive variance and phenotypic variance.

Results

The logCH_4 concentration for each cross and herd is shown in Figure 1. Descriptive statistics of visits and methane concentration are presented in Table 1. The number of visits varied between breed groups. The slaughter calf producer contributed with different number of calves due to difference in capacity at sights (farm1=288, farm2=1,583, farm3=62, farm4=71, farm5 =554, farm6=105). The calves represented both sexes (bull=1,620, heifer=1,043). Phenotypic means and standard error of the mean of methane concentration are presented in Table 1.

The genetic parameters were estimated on the 1,689 BBL×HOL calves. The results from the parameter estimation are shown in Table 2. The heritability of the traits are moderate for logCH_4c and logCH_4a, and all significantly different from zero. Repeatability for logCH_4c and logCH_4a is equal to the heritability.

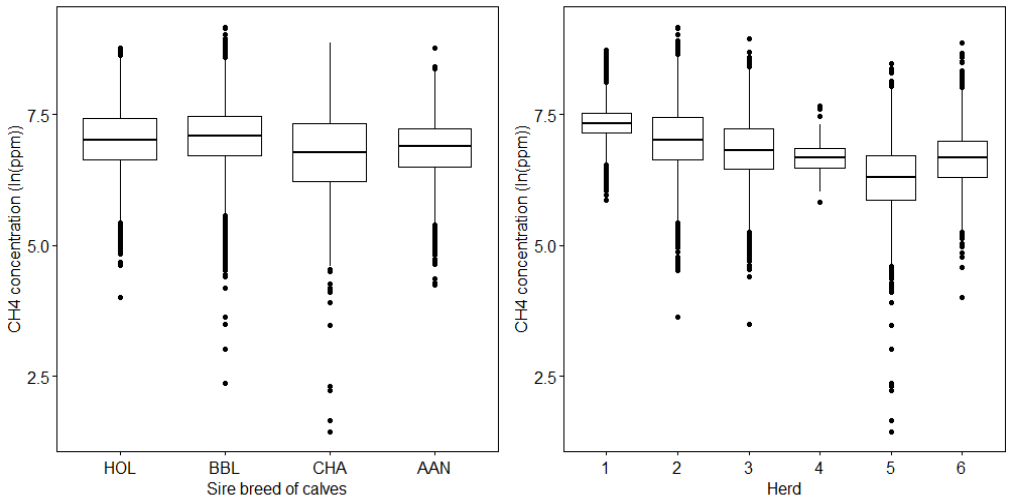


Figure 1. CH_4 concentration (ln transformed) for the breed combinations (left) and between the herds (right).

Table 1. Phenotypic means and standard error of the mean (SEM) for methane visits and traits[1] for each breed group[2] of slaughter calves.

Trait	BBL×HOL	CHA×HOL	ANG×HOL	HOL
Visits per animal (n)	45.4 (1.35)	21.0 (1.89)	23.8 (1.62)	40.0 (2.55)
Visit time (s)	43.8 (0.71)	42.5 (1.55)	49.1 (2.00)	39.7 (1.32)
CH4c (ppm)	1,404 (20.0)	1,066 (41.2)	1,091 (39.3)	1,361 (40.0)
CH4a (ppm)	1,029 (17.8)	729 (33.9)	768 (33.2)	1,076 (35.2)
logCH4c (ln(ppm))	7.09 (0.01)	6.74 (0.04)	6.85 (0.04)	7.05 (0.03)
logCH4a (ln(ppm))	6.73 (0.02)	6.34 (0.04)	6.47 (0.04)	6.79 (0.03)

[1] CH4c = Methane concentration, CH4adj = Methane concentration adjusted for background, logCH4c = ln transformed methane concentration, logCH4a = ln transformed methane concentration adjusted for background

[2] BBL×HOL= Belgian Blue cross, CHA×HOL= Charolaise cross, ANG×HOL = Angus cross, HOL = Holstein

Table 2. Estimated repeatability and heritability with asymptotic standard error for methane concentration traits[1]

Trait	R	h^2 (SE)
logCH4c	0.44	0.44 (0.08)
logCH4a	0.30	0.30 (0.06)

[1] $logCH_4c$ = ln transformed methane concentration, $logCH_4a$ = ln transformed methane concentration adjusted for background.

Discussion

This work is the first step on the road to achieving a methane phenotype for BeefxDairy crossbred calves. The level of CH_4 concentration is higher than what was measured in Danish Holstein cows, when methane was measured in the milking robot (Difford *et al.*, 2020; Manzanilla-Pech *et al.*, 2021). Mean CH_4 concentration CHAxHOL and ANGxHOL was lower compared to BBLxHOL and purebred HOL (Table 1). Variation within groups is large (Figure 1), so conclusions should be made with caution. No difference in CH_4 was seen between herds, and variation was not consistent across herds. This could be due to the use of TMR in one case (Herd 4).

The repeatability was equal to the estimated heritability. This is caused by the estimated permanent environmental effect being zero. This conflicts with other studies in dairy cows, where the repeatability was a lot higher than the heritability (Difford *et al.*, 2020). The standard error in the estimates in this study is however large. Increasing the sample size will provide a lower estimate of the standard error and a higher accuracy.

The heritabilities for CH_4 concentration was estimated with a large standard error, but is in range with CH_4 traits from other studies (Difford *et al.*, 2020; Hayes *et al.*, 2016; Lassen and Løvendahl, 2016, Manzanilla-Pech, 2021). Log transformed CH_4 concentration is used when estimating parameters on CH_4 concentration to meet the requirement of normal distributed residuals (Difford, *et al.*, 2016). Methane concentration can be extrapolated to a daily methane production which has a moderate heritability in lactating HOL cows (Manzanilla-Pech, *et al.* 2021). CHAxHOL and ANGxHOL had a lower phenotypic mean CH_4 concentration compared to BBLxHOL and purebred HOL. The heritability of CH_4 concentration was estimated to be moderate in BBLxHOL crosses. Further work will develop this phenotype further and investigate the genetic architecture of the trait.

References

Davis, R. B., Fikse, W. F., Carlén, E., Fogh, A., Pösö et al (2020). Available at https://www.nordicebv.info/wp-content/uploads/2020/10/BxD_final_report.pdf

Difford, G. F., Lassen, J., & Løvendahl, P. (2016). COMPUT ELECTRON AGR, *124*: 220-226. https://doi.org/10.1016/j.compag.2016.04.010

Difford, G. F., Løvendahl, P., Veerkamp, R. F., Bovenhuis, H., Visker, *et al.* (2020). J DAIRY SCI, *103*(3): 2442-2459. https://doi.org/10.3168/jds.2019-16966

Ettema, J. F., Thomasen, J. R., Hjortø, L., Kargo, M., Østergaard, S. *et al.* (2017). J DAIRY SCI, *100*(5): 4161-4171. https://doi.org/10.3168/jds.2016-11333

Hayes, B. J., Donoghue, K. A., Reich, C. M., Mason, B. A., Bird-Gardiner, T., *et al.* (2016). J ANIM SCI, *94*(3): 902-908. https://doi.org/10.2527/jas.2015-0078

Lassen, J., & Løvendahl, P. (2016). J DAIRY SCI, *99*(3): 1959-1967. https://doi.org/10.3168/jds.2015-10012

Madsen P., and Jensen J. (2008) A user's guide to DMU: A package for analysing multivariate mixed models. Version 6 release 4.7. Available at: http://dmu.agrsci.dk

Manzanilla-Pech, C. I. V., Difford, G. F., Sahana, G., Romé, H., Løvendahl, P., *et al.* (2021). J DAIRY SCI, https://doi.org/10.3168/jds.2021-20410

Mogensen, L., Kristensen, T., Nielsen, N. I., Spleth, P., Henriksson, M.,*et al.* (2015). LIVEST SCI, *174*: 126-143. https://doi.org/10.1016/j.livsci.2015.01.021

27. Genome-wide association study for mid-infrared methane predictions in Walloon dairy cows

H. Atashi[1,2*], H. Wilmot[1,3], S. Vanderick[1], X. Hubin[4], H. Soyeurt[1] and N. Gengler[1]

[1]Gembloux Agro-Bio Tech, University of Liège, 5030 Gembloux, Belgium; [2]Shiraz University, 71441-65186 Shiraz, Iran; [3]National Fund for Scientific Research (F.R.S.-FNRS), Rue d'Egmont 5, 1000 Brussels, Belgium; [4]Elevéo Asbl 5590 Ciney, Belgium; hadi.atashi@uliege.be

Abstract

This study aimed to identify genomic regions associated with two mid infrared-based CH_4 traits [predicted daily CH_4 emission (PME, g/d), and log-transformed predicted CH_4 intensity (LMI)] in Walloon dairy cows. The data consisted of 1,529,282 test-day records from 229,465 cows distributed in 1,530 herds collected from 2006 to 2021. Random regression test-day models were used to estimate variance components. The proportion of genetic variance explained by windows of 50 consecutive SNPs was calculated and regions accounting for at least 1.0% of the total genetic variance were identified. Mean (SD) daily h^2 estimated for PME and LMI were 0.14 (0.05) and 0.24 (0.05), respectively. Two regions on BTA14 (positions 1.86 to 2.12, and 1.48 to 1.68 Mb) were associated with both PME and LMI. A region between 144.38 to 144.46 Mb on BTA1 was associated with PME; and the region between 2.68 and 2.94 Mb on BTA14 was associated with LMI. Results showed potential for genome-enhanced advisory systems to reduce methane emissions.

Introduction

Emissions of greenhouse gases (GHG) have a considerable impact on climate change. Livestock production is a significant contributor to global human-induced GHG emissions in the form of CH_4, N_2O and CO_2 (De Haas *et al.*, 2021). Dairy cattle account for 20% of the global livestock sector's GHG, and over half of this is from enteric CH_4 emissions (Gerber *et al.*, 2013). Although a reduction in CH_4 emission could be achieved by using different CH_4 mitigation strategies like management or feeding (Cottle *et al.*, 2011), a genetic selection of low CH_4 emitting cows can be an effective approach to reduce GHG production from dairy cattle. The genetic variation on CH_4 emission is influenced by many genes with heterogeneous effects (Sypniewski *et al.*, 2021). However, there is currently a gap in the knowledge on the genetic architecture of this new trait (Kandel *et al.*, 2017; Sypniewski *et al.*, 2021). The short-term aim of this study was to estimate the genetic parameters of two MIR-based CH_4 features for the first-parity Walloon dairy cows and to identify their associated genomic regions. The long-term objective is to contribute to genome-enhanced advisory systems to reduce methane emissions.

Materials &methods

Data. Milk samples were collected on first parity cows from 2006 to 2021 by the Walloon Breeding Association (Ciney, Belgium). All milk samples were analysed using Milkoscan FT4000, FT6000, and FT+ (Foss-Electric A/S, Hillerød, Denmark) by the milk laboratory Comité du Lait (Battice, Belgium) to generate the MIR spectral data. Methane emissions (PME, g/d) were predicted from the recorded spectra using the equations developed by Vanlierde *et al.* (2021). To eliminate potential abnormal records, the PME values below the 0.1 percentile and above the 99.9 percentile were deleted (Kandel *et al.*, 2017). Methane emission intensity (PMI, g/ kg of milk) was defined as the ratio of PME divided by the daily milk yield (kg/d) recorded on the same test-day. This PMI was then log-transformed to be normally distributed and called log-transformed CH_4 intensity (LMI). Data were edited to include only cows with known birth date. Records from days in milk (DIM) lower than 5 d and greater than 365 d were eliminated. Age at the first calving (AFC) was restricted to the range of 540 to 1,200 d. Daily milk yield was restricted to the range of 3 to 99 kg. The final dataset consisted of 1,529,282 test-day records on 229,465 cows distributed in 1,530

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_27, © H. Atashi *et al.* 2022

herds collected from 2006 to 2021. More than 95% of cows included were Holstein. Genotypic data of 730,539 SNPs were available for 7,381 animals. SNPs with minor allele frequency less than 5%, and those that did not pass the Hardy-Weinberg equilibrium were excluded. In total, 565,049 SNPs were kept for genomic analyses.

Variance component estimation. The (co)variance components and genomic breeding values for PME and LMI were estimated based on the integration of the random regression test-day model (RR-TDM) into single-step GBLUP procedure (SS RR-TDM) using the following model:

$$y_{ijklm} = \mu + HTD_i + \sum_{b=0}^{4} AS_j \phi_b(t) + \sum_{b=0}^{2} HY_k \phi_b(t) + \sum_{b=0}^{2} pe_l \phi_b(t) + \sum_{b=0}^{2} a_l \phi_b(t) + e_{ijklm} \quad (1)$$

where y_{ijklm} is the test-day record (PME or LMI) on DIM m of cow l, belonging to i^{th} class of HTD, j^{th} class of AS, and k^{th} class of HY, HTD is the fixed effect of herd-test-day, AS is the fixed effect of age-season of calving defined as following: age at calving class (10 classes) × season of calving (four seasons), $\sum_{b=0}^{4} AS_j \phi_b(t)$ is the fixed regression coefficients of the age-season at calving modelled using Legendre polynomials of order 4, $\sum_{b=0}^{2} HY_k \phi_b(t)$, $\sum_{b=0}^{2} pe_l \phi_b(t)$ and $\sum_{b=0}^{2} a_l \phi_b(t)$ are, respectively, the random regression coefficients of herd-year at calving (HY), permanent environment, and additive effects modelled using Legendre polynomials of order 2, and e_{ijklm} is the residual effect.

The variance components were estimated through the Bayesian Gibbs sampling method using a single chain of 100,000 iterations after a burn-in period of 20,000. The vector of genomic estimated breeding values (GEBV) of the included traits for each animal i, was estimated by multiplying the vector of predicted regression coefficients of the additive genetic effect by the matrix of Legendre orthogonal polynomial covariates; that is, $GEBV_i = T\hat{g}_i$, where $\hat{g}_i$ is the vector of predicted regression coefficients of the additive genetic effect for animal i, and T is a matrix of orthogonal covariates associated with the Legendre orthogonal polynomial functions. The SNP effects were estimated individually for the included traits using the postGSf90 software (Aguilar *et al.*, 2014). The animal effects were decomposed into those for genotyped (a_g) and ungenotyped animals (a_n). The animal effects of genotyped animals are a function of the SNP effects, $a_g = Zu$, where Z is a matrix relating genotypes of each locus and u is a vector of the SNP marker effect. The variance of animal effects was assumed as:

$$Var(\mathbf{a_g}) = Var(\mathbf{Zu}) = \mathbf{ZDZ'}\sigma_u^2 = \mathbf{G}\sigma_a^2 \quad (2)$$

where D is a diagonal matrix of weights for variances of markers (D = I), σ_u^2 is the genetic additive genetic variance captured by each SNP marker when the weighted relationship matrix (G) was built with no weight. The SNP effects were obtained using following equation:

$$\hat{u} = \lambda \mathbf{DZ'}\,\mathbf{G^{-1}}\,\hat{a}_g = \mathbf{DZ'}[\mathbf{ZDZ'}]^{-1}\hat{a}_g \quad (3)$$

where λ was defined by VanRaden (2008) as a normalizing constant, as described below:

$$\lambda = \frac{\sigma_u^2}{\sigma_a^2} = \frac{1}{\sum_{i=1}^{M} 2p_i(1-p_i)} \quad (4)$$

The percentage of genetic variance explained by the i^{th} genomic region was estimated as:

$$\frac{Var(a_i)}{\sigma_a^2} \times 100\% = \frac{Var(\sum_{j=1}^{50} Z_j \hat{u}_j)}{\sigma_a^2} \times 100 \quad (5)$$

where a_i is the genetic value of the i^{th} region that consists of 50 adjacent SNPs, σ_a^2 is the total genetic variance, Zj is the vector of the SNP content of the j^{th} SNP for all individuals, and $\hat{u}_j$ is the marker effect of the j^{th} SNP within the i^{th} region. The results were presented by the proportion of variance explained by each window of 50 adjacent SNPs (with an average size of 165 kb) and windows explaining for at least 1.0% of the total additive genetic variance were identified.

Results

Means (SD) for PME, PMI and LMI were 326.8 (67.65) g/d, 14.97 (5.70) g CH_4/kg of milk yield, and was 2.64 (0.35), respectively. Predicted CH_4 and LMI reached a maximum of 341.6 g/d and 2.83 by DIM 80 and 344, respectively (Figure 1-A). Heritability estimates for PME and LMI were relatively stable across lactation, with a mean h^2 of 0.14 and 0.24, and peak h^2 of 0.21 (DIM 186) and 0.30 (DIM 211), respectively (Figure 1-B, Table 1). Two regions on BTA14 (positions 1.86 to 2.12, and 1.48 to 1.68 Mb) were associated with both PME and LMI. A region between 144.38 to 144.46 Mb on BTA1 was associated with PME; and the region between 2.68 and 2.94 Mb on BTA14 was associated with LMI (Figure 1-C, and 1-D). Functional genes including *LY6D*, *GML*, *CYHR1*, *PPP1R16A*, *ARHGAP39*, *ZNF7*, *OPLAH*, *MAF1*, and *SPATC1* were identified as potential candidate genes for PME and LMI (Assembly UMD3.1).

Table 1. Mean (SD) phenotypic values and heritability estimates for considered milk mid-infrared-based CH_4 features.

	Phenotypic values		Heritability		
Trait	**Mean**	**SD**	**Mean**	**SD**	**range**
PME[1]	326.8	67.65	0.14	0.05	0.06-0.21
LMI[2]	2.64	0.36	0.24	0.05	0.16-0.30

[1] Methane emission (PME, g/d) predicted from the recorded milk MIR spectra.

[2] Log-transformed CH_4 intensity (LMI) based on the ratio of PME divided by the daily milk yield (kg/d).

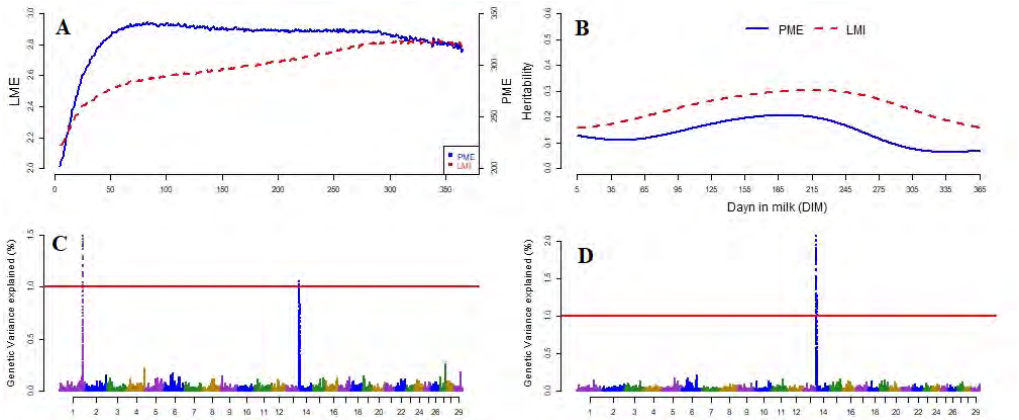


Figure 1. The lactation curves for PME (solid) and LMI (dashed) (A). Heritability estimates for PME (solid) and LMI (dashed) throughout lactation (B). Manhattan plots for PME (C) and LMI (D). The x-axis indicates the BTA number and the y-axis indicates the percentage of additive genetic variance explained by the window of 50 SNPs. The PME is the CH_4 emission (g/d) and the LMI is the log transformed of CH_4 intensity (g CH_4/kg of milk yield).

Discussion

We aimed to estimate genetic parameters and identify genomic regions associated with CH_4 emission in Walloon dairy cows. The mean PME found (326.8 g/d) was in the range reported by previous studies (Pickering *et al.*, 2015; Kandel *et al.*, 2017). Moderate h^2 were estimated for PME and LMI in agreement with the literature (Kandel *et al.*, 2017; Sypniewski *et al.*, 2021). Three genomic regions were found on BTA14 which, when combined, explained 2.10 and 3.35% of the total genetic variance of PME and LMI, respectively. The genomic regions and genes associated with PME and LMI have been previously reported to be associated with milk yield and composition in dairy cows (Bennewitz *et al.*, 2004; Atashi *et al.*, 2020). However, this is the first report on the association between these regions and CH_4 emission in cattle. In conclusion, the moderate h^2 found indicates that the studied traits can be introduced in the breeding programs. Further studies must be conducted to evaluate the relationship of our mid-infrared-based CH_4 traits to direct CH_4 traits and the effect of their inclusion in current breeding objectives. Moreover, large database could be used for genome-enhanced consulting systems to reduce methane emissions.

Acknowledgements

Project supported by the Walloon Government (SPW-DGARNE, Namur, Belgium) through the WALLeSmart Project (D31-1392). H. Wilmot, as Research Fellow acknowledges the support of the Fonds de la Recherche Scientifique – FNRS (Brussels, Belgium).

References

Aguilar I., Misztal I., Tsuruta S., Legarra A. and Wang, H. (2014). Proc. of the 10th WCGALP, Vancouver, Canada.

Atashi H., Salavati M., De Koster J., Ehrlich J., Crowe, M. *et al.* (2020). J. Anim. Breed. Genet. 137, 292-304. https://doi.org/10.1111/jbg.12442

Bennewitz J., Reinsch N., Paul S., Looft C., Kaupe B. *et al.* (2004). J. Dairy Sci. 87, 431-442. https://doi.org/10.3168/jds.s0022-0302(04)73182-3

Cottle D.J., Nolan J.V., and Wiedemann S.G. (2011). Anim. Prod. Sci. 51: 491-514. https://doi.org/10.1071/AN10163

De Haas Y., Veerkamp R., de Jong G., and Aldridge M. (2021). Animal 15: 100294. https://doi.org/10.1016/j.animal.2021.100294

Gerber P.J., Steinfeld H., Henderson B., Mottet A., Opio, *et al. (*2013). Tackling climate change through livestock – A global assessment of emissions and mitigation opportunities. Food and Agriculture Organization of the United Nations (FAO), Rome.

Kandel P., Vanrobays M.L., Vanlierde A., Dehareng F., Froidmont E. *et al.* (2017). J. Dairy Sci. 100: 5578-5591. https://doi.org/10.3168/jds.2016-11954

Pickering NK., Chagunda M.G., Banos G., Mrode R., McEwan J.C. *et al.* (2015). J. Anim. Sci. 93, 11-20.https://doi.org/10.2527/jas.2014-8302

Sypniewski M., Strabel T., and Pszczola M. (2021). Animals 11: 3175. https://doi.org/10.3390/ani11113175.

Vanlierde A, Dehareng F, Gengler N, Froidmont E, McParland S. et al (2021). J. Sci. Food Agric. 101(8), 3394-3403. https://doi.org/10.1002/jsfa.10969

VanRaden P.M. (2008). J Dairy Sci. 91: 4414-4423. https://doi.org/10.3168/jds.2007-0980

28. Genetic parameters for feed efficiency, gas emissions, oxygen consumption and wool traits in Australian Merino

C.B. Marques*, I. De Barbieri, J. Velazco, E.A. Navajas and G. Ciappesoni*

Instituto Nacional de Investigación Agropecuaria, INIA Las Brujas, 90100, Rincón del Colorado, Ruta 48 km 10, Canelones, Uruguay; cbalconi@inia.org.uy; gciappesoni@inia.org.uy

Abstract

The purpose of the present study was to estimate genetic parameters of growing globally important traits such as feed intake, residual feed intake, methane, carbon dioxide, oxygen consumption and greasy fleece weight in Uruguay. Data of three Australian Merino generations which integrate the Uruguayan National Genetic Evaluation for sheep were recorded (from 2019 to 2021). Heritabilities and genetic and phenotypic correlations were estimated for 930 animals sired by 19 rams. Results presented suggest that relevant genetic progress can be achieved in all traits (heritabilities between 0.23 and 0.41). Methane emission present moderate to high genetic correlations with the other traits (0.57-0.88). The most efficient animals will tend to consume less feed and emit less methane, without affecting wool production. The genetic correlation between wool production and the other traits were not significantly different from zero. The developing of this work will be the basis for the construction of agroecological breeding objectives.

Introduction

Ruminants emit methane (CH_4) through enteric fermentation in their digestive tract, which represents a loss of 6 to 10% of gross energy intake and is a significant source of anthropogenic CH_4 emissions globally (Schaefer *et al.*, 2016). Sheep emissions and their associations with feed intake and production traits are currently investigated as part of the implementation of greenhouse gas (GHG) mitigation strategies. Genetic selection of efficient and low-emitting animals is promising, as long as these traits are heritable. If a favourable correlation exists between traits such as feed efficiency and CH_4 emissions, a mitigation of CH_4 production could be obtained as a response of selecting for more efficient animals. The purpose of the present study was to estimate genetic parameters for feed efficiency, GHG emissions, oxygen consumption and wool production traits in Australian Merino lambs in Uruguay.

Materials & methods

Data and traits. Data of three Australian Merino generations which integrate the Uruguayan National Genetic Evaluation for sheep were recorded from 2019 to 2021. Information from 930 animals sired by 19 rams was available on the following traits: feed intake (FI), residual reed intake (RFI), CH_4, carbon dioxide (CO_2), oxygen consumption (O_2) and greasy fleece weight (GFW). The 2020 progeny was the first in which several of their dams (27%) or sires (15%) or both parents (4%) had also been recorded for feed intake and gas emissions as lambs in 2019. Descriptive statistics and data structure for ages (at trial and at shearing) and all the traits investigated are presented in Table 1.

Feed efficiency measurements. Three RFI tests were carried out per year (9 in total) where animals were allotted based on sex, birth type and age. After 14 days of acclimatization, each test period had a duration of 42 days, in which animals were allocated to one of five outdoor pens, with five automated feeding systems in each pen, according to body weight, sex and sire. Lambs were fed *ad libitum* with Lucerne haylage. Each pen had two automatic weighing platforms that allowed individual weight control.

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_28, © C.B. Marques *et al.* 2022

Table 1. Descriptive statistics and data structure for ages, residual feed intake (RFI), feed intake (FI), oxygen consumption (O_2), methane (CH_4), carbon dioxide (CO_2) and greasy fleece weight (GFW) traits.

Trait	n	Mean	Standard deviation	Min	Max
Age at trial (days)	930	295.33	45.79	204	378
Age at shearing (days)	930	413.03	8.97	387	432
RFI (kg/dry matter/day)	930	-0.001	0.15	-0.43	1.58
FI (kg/dry matter/day)	930	1.33	0.27	0.67	3.16
CH_4 (g/day)	930	22.43	5.61	8.91	44.66
O_2 (g/day)	930	981.37	191.96	427.33	2,114.66
CO_2 (g/day)	930	1,067.45	234.71	478.18	2,269.63
GFW (kg)	918	4.13	0.73	2.10	6.53

Portable accumulation chamber measurements. Sheep emissions were measured in portable accumulation chambers (PACs) made of polycarbonate sheet with an internal volume of 861.8 l (118.8-cm length × 120.7-cm height × 60.1-cm width), similar to those described by Goopy *et al.* (2011). The CH_4, O_2 and CO_2 concentration measurements were taken after confining the sheep in the PACs for a minimum time of 40 min. Total gas space inside the portable static chamber was estimated by assuming that the volume occupied by the sheep was equal to 1.01 L/kg liveweight (Jonker *et al.*, 2018) and subtracting the liveweight of the sheep from the internal volume of the chamber. Data analysed was the mean value of two measurements collected around day 30 and day 37 in trial. Using an analogous methodology as the described by Goopy *et al.* (2011), a calculation was performed to convert the CH_4, O_2 and CO_2 concentration from PACs to daily measurements.

Statistical analyses. A Bayesian analysis and an assessment of variance and covariance components using GIBBSF90 computer package (Misztal *et al.,* 2002) was performed. The number of iterations after burn-in used for posterior inferences was 300,000 with the model:

$y_{ijklm} = CG_i + BT_j + DA_k + age_l + a_m + e_{ijklm}$ where: y_{ijklm}: m^{th} performance record of animal l, CG_i: fixed effect of the contemporary group i (year-management group-sex-pen-trial, 134 levels), BT_j: fixed effect of birth type j (2 levels), DA_k: fixed effect of dam age k (3 levels), age: age at m measurement of animal l as a co-variable, a_l:random additive genetic effect of animal l (1,962 animals), and e_{ijklm}: random residual.

Results

Heritabilities, genetic and phenotypic correlations estimated are shown in Table 2. Moderate magnitude heritabilities between 0.23 and 0.41 were found for all traits. The three gases presented the lowest values. High genetic correlations were observed between RFI with FI and CH_4, FI with the three gases, and CO_2 with O_2 and CH_4 (range 0.64-0.97). The genetic relationship of RFI with CO_2 and O_2 with CH_4 were moderate (0.50-0.57). In this work, genetic correlations estimated between GFW and feed efficiency and greenhouse traits, as well as between RFI and O_2, were not significantly different from zero.

When analysing the phenotypic results, high correlation values were found between RFI and FI, and between CO_2 and other two gases (0.64-0.93). Phenotypic relationship was moderate between FI and the three gases, and also between O_2 and CH_4 (0.28-0.57). The phenotypic correlations of RFI with CH_4, CO_2 and GFW were not significantly different from zero.

Table 2. Genetic and phenotypic parameters for residual feed intake (RFI), feed intake (FI), oxygen consumption (O_2), methane (CH_4), carbon dioxide (CO_2) and greasy fleece weight (GFW) traits in Australian Merino breed[1].

	RFI	FI	O_2	CH_4	CO_2	GFW
RFI	**0.27 (0.09)**	0.79 (0.10)	0.37 (0.22)	0.64 (0.22)	0.50 (0.20)	0.05 (0.18)
FI	0.75 (0.05)	**0.38 (0.12)**	0.68 (0.14)	0.88 (0.10)	0.77 (0.11)	0.13 (0.18)
O_2	0.15 (0.08)	0.45 (0.08)	**0.26 (0.09)**	0.57 (0.20)	0.97 (0.02)	-0.13 (0.18)
CH_4	-0.02 (0.10)	0.28 (0.13)	0.57 (0.06)	**0.23 (0.12)**	0.66 (0.18)	0.06 (0.20)
CO_2	0.11 (0.08)	0.44 (0.09)	0.93 (0.01)	0.64 (0.05)	**0.27 (0.09)**	-0.15 (0.18)
GFW	0.01 (0.07)	0.24 (0.07)	0.26 (0.07)	0.18 (0.07)	0.27 (0.07)	**0.41 (0.03)**

[1] Heritabilities (in bold) on the diagonal, phenotypic and genetic correlations below and above the diagonal, respectively, posterior standard deviation in parenthesis.

In general, genetic correlations were of higher magnitude than phenotypic, except for GFW that also had genetic correlations close to zero with all other traits.

Discussion

To the authors' knowledge, this is the first study that reports genetic parameters for feed efficiency related traits together with gas emissions, oxygen consumption and wool production. The current study presents phenotypic values very similar and at least comparable to several others in literature, such as for FI (François *et al.*, 2002), O_2 (Jonker *et al.*, 2018), CH_4 (Jonker *et al.*, 2018; Pinares-Patiño *et al.*, 2013), CO_2, and GFW (Huisman *et al.*, 2008). In the case of GFW phenotypic parameters, Ciappesoni *et al.* (2010) reported a lower mean for the same breed and country which can be explained by the population structure. The current study used data from recent years of an animal selection nucleus, where a favourable genetic trend for heavier GFW is observed (www.geneticaovina.com.uy).

As genetic improvement by animal selection is possible as long as the selected traits are heritable, the moderate heritabilities presented in this work indicate that promising and faster genetic gains can be achieved in all traits. The heritability values presented for RFI and FI are very similar to those described by François *et al.* (2002), but higher than the heritability values reported by Cammack *et al.* (2005) of 0.11 and 0.25 for RFI and FI, respectively. The limited published literature for heritability of O_2 consumption in sheep is reported as 0.21 (Jonker *et al.,* 2018). In the same study, heritability estimates for CH_4 production and CO_2 were 0.19 and 0.16 for lambs, and 0.20 and 0.27 ewes, respectively. Similar CH_4 heritability estimates were also reported by Pinares-Patiño *et al.* (2013) and Robinson *et al.* (2010). Regarding GFW heritability, our estimate is in agreement with those found by Huisman *et al.* (2008) (0.46), and Ciappesoni *et al.* (2010) (0.45).

A slightly higher value for the genetic correlation between CH_4 and CO_2 of (0.84) and (0.86) was reported by Jonker *et al.* (2018) and Paganoni *et. al*, (2017), respectively. Thus, still in accordance with our high genetic correlation between CH_4 and CO_2.

When evaluating greenhouse gas mitigation strategies, the favourable genetic correlation between RFI efficiency and greenhouse gas emissions is promising in the search for more efficient animals. However, these preliminary data must be taken with caution until more information continues to record, and more precise estimates can be achieved.

Furthermore, GFW correlations were not significantly different from zero, showing that animal selection with emissions mitigation purpose should not impact wool production, which is one of the most important selection objectives in Merino breed. Therefore, the current selection for GFW does not seem to affect the efficiency of the animals or increase emissions. These favourable to zero genetic correlations would make possible to find more animals that face several objectives at the same time: productive, economic, and environmental.

The development of intensive phenotyping systems for traits related to feed intake, efficiency, and greenhouse gas mitigation in association with the estimation of genetic parameters gets on greater significance when is linked to stud-flocks and commercial populations. This will allow the identification of potential trade-off between traits and the implementation of genetic improvement programs for individual traits, or within a more holistic vision such as the agroecological breeding objectives.

Acknowledgements

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 772787 (SMARTER). Furthermore, by INIA – 'Instituto Nacional de Investigación Agropecuaria' (INIA_CL_38: RUMIAR), the SusAn, ICT-AGRI 2 and FACCE ERA-GAS funding bodies (GrassToGas project).

References

Ciappesoni, G., Gimeno, D., Ravagnolo, O. (2010). Proc. of the 9th *WCGALP,* Leipzig, Germany.

François, D., Bibé, B., Brunel, J. B., Weisbecker, J. L., & Ricard, E. (2002). Proc. of the 7th WCGALP, Montpellier, France.

Goopy, J. P., Woodgate, R., Donaldson, A., Robinson, D. L., & Hegarty, R. S. (2011). Anim. Feed Sci. Technol. 166-167: 219-226. https://doi.org/10.1016/j.anifeedsci.2011.04

Jonker, A., Hickey, S. M., Rowe, S. J., Janssen, P. H., Shackell, G. H, *et al.* (2018). J. Anim. Sci. 96(8): 3031-3042. https://doi.org/10.1093/jas/sky187

Huisman, A. E., Brown, D. J., Ball, A. J., & Graser, H.-U. (2008). Aust. J. Exp. Agric. 48(9): 1177. https://doi.org/10.1071/ea08119

Misztal, I., Tsuruta, S., Strabel, T. *et al.* (2002). Proc. of the 7th WCGALP, Montpellier, France.

Paganoni, B., Rose, G., Macleay, C., Jones, C., Brown, D. J, *et al.* (2017). J. Anim. Sci., 95(9), 3839-3850. https://doi.org/10.2527/jas.2017.1499

Pinares-Patiño, C. S., Hickey, S. M., Young, E. A., Dodds, K. G., MacLean, S., *et al.* (2013). Animal 7: 316-321. https://doi.org/10.1017/S1751731113000864

Robinson, D. L., Goopy, J. P., Hegarty, R. S., & Vercoe, P. E. (2010). Proc. of the 9th *WCGALP,* Leipzig, Germany.

Schaefer, H., S. E. Fletcher, C. Veidt, K. R. Lassey, G. W. Brailsford, *et al.* (2016) Science 352: 80-84. https://doi.org/10.1126/science.aad2705

29. Genetic selection of feed efficiency and methane emissions in sheep and cattle in Uruguay: progress and limitations

E.A. Navajas*, O. Ravagnolo, I. De Barbieri, M.I. Pravia, I. Aguilar, M.O. Lema, B. Vera, P. Peraza, C.B. Marques, J.I. Velazco and G. Ciappesoni

†*Instituto Nacional de Investigación Agropecuaria, INIA Las Brujas, 90100, Rincón del Colorado, Canelones, Uruguay; enavajas@inia.org.uy*

Abstract

Residual feed intake (RFI) and methane (CH_4) emissions are potential new selection objectives in beef and sheep breeding programmes to assist reaching greenhouse gas mitigation targets and to support economic, social, and environmental sustainability. Phenotyping platforms in both species are in place in Uruguay, linked with the genetically evaluated populations of main breeds (Hereford, Merino, Corriedale, Dohne and Texel), being the basis of the reference populations for genomic selection. Current progress and main findings are described here. New selection indexes and selection criteria, as residual CH_4, are needed given the potential unfavourable associations between CH_4 and performance. Larger reference populations imply higher genomic prediction accuracies, accurate estimations of genetic correlations among feed intake, RFI, CH_4 and performance and a comprehensive understanding of these associations. On-going rumen metagenomics and metatrascriptomics will also provide information about its value as predictor of genetic merit of these traits.

Introduction

Global concerns regarding mitigation of greenhouse gas (GHG) emissions and efficient use of limiting resources is intensifying, as well as pressures to achieve national targets in the context of international environmental agreements. The identification of alternatives to support sustainable development is particularly relevant for countries whose economy relay on sheep and cattle production, such as Uruguay. The incidence of enteric methane (CH_4) in the total GHG emissions tend to be high in these countries (i.e. 46%, Uruguay) and drastic alternatives, such as reducing national flocks and herds to decrease total CH_4 would have significant unfavourable economic and social implications. Emission intensity (CH_4 per unit of product) is one of the indicators proposed to assess progress of GHG mitigation keeping the balance among environmental, economic, and social sustainability.

Genetic selection can contribute to both components of emission intensity. Uruguay have had breeding programmes in beef cattle and sheep for 30 years, and genetic progress have been reported for individual traits and selection indexes in both species. Although this looks promising in terms of enhancing the contribution of genetic selection to reduce emission intensity, improving production accompanied with higher feed intake may jeopardise the expected progress (Barwick *et al.*, 2019). Feed intake, feed efficiency and GHG emissions are recorded in new phenotyping platforms with the aim of evaluating strategies to incorporate these traits into the national breeding programmes supporting sustainable development (Phocas *et al.*, 2016). Here we describe the research approach carried out in cattle and sheep in Uruguay since 2014, in connection with breeding programmes, results and limitations.

Feed intake and efficiency

Feed efficiency measurements rely on feed intake records (adjusted by dry matter, DMI), which is a difficult-to-measure trait in large number of animals. Commercial availability of automated feed intake recording systems contributed to partially overcome this limitation. Post-weaning feed efficiency tests of Hereford bulls and steers started in 2014 at the Central de Prueba Kiyú of the Hereford Breeders Society of

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_29, © E.A. Navajas *et al.* 2022

Uruguay, with the aim of implementing genomic selection of residual feed intake (RFI) (Ravagnolo *et al.*, 2018). Nowadays, the national database includes post-weaning feed efficiency information of 960 pedigree bulls and 560 steers. More recently, RFI at finishing (feedlot) is also evaluated in steers immediately after the postweaning test. A similar initiative is carried out in the main sheep breeds in Uruguay in INIA Experimental Station La Magnolia, with 981 Merino, 300 Corriedale, 367 Dohne and 77 Texel lambs recorded at the end of 2021. These animals belong to Selection Nucleus (Merino), Information Nucleus (Corriedale, Dohne) and Central Progeny Testing (Texel) and are strongly connected with population in the genetic evaluation (performance recorded).

Koch *et al.* (1963) defined RFI as the difference between actual and predicted DMI, based on average daily gain (ADG) and metabolic body weight (MBW). Improving feed efficiency by selecting for RFI is an appealing breeding objective because of the associated increase of farmers net income by reducing feed costs without compromising animal performance. At a production system or national level, improving RFI could also be interpreted as a contribution to optimise the use of limiting resource, such as land in grazing-based systems. The final impact will depend on its genetic variability and covariance with other breeding objectives.

Heritability of RFI and association with production traits

As previously reported in the literature, RFI is a moderately heritable trait in Uruguayan Hereford and Merino breeds, with estimates of 0.25 (Pravia *et al.*, 2022, unpublished) and 0.27 (Marques *et al.*, WCGALP 2022), respectively. The genetic correlations (r_G) of RFI and the traits used for its estimation in Hereford were similar to the phenotypic associations (r_p): positive correlations with DMI (r_P=0.63; r_G=0.40) and very low or zero with ADG (r_G=0.07; r_P=-0.003) and MBW (r_G=-0.12; r_P=-0.001).

Accurate estimations of genetic parameters depend on large databases, principally for genetic correlations, which is challenging for difficult-to-measure traits. Consequently, many of the first studies rely on phenotypic associations and comparisons between animals with extreme values for the traits of interest, particularly when aiming to establish the associations of RFI with other difficult-to-measure traits, such as carcass and meat quality. In agreement with other studies, no significant differences were found in pasture-based finishing performance, carcass quality, meat production and quality between high- and low-RFI steers, based on their post-weaning RFI (Pravia *et al.*, 2018). The analysis of post-weaning RFI and feedlot RFI in 100 Hereford steers indicated that both feed efficiencies were strongly correlated (0.72), whilst AGD and MBW at finishing were independent of post-weaning and feedlot RFI (P>0.05) (Silveira *et al.*, 2022, unpublished data). Although there is evidence indicating that low RFI (high feed efficiency) was associated with lower fatness, we did not find unfavourable phenotypic associations between RFI and fatness measured *in vivo* or post-slaughter, or in terms of genetic correlations (r_P=0.08; r_G=0.01, Pravia *et al.*, 2022, unpublished). This may be explained by the incorporation of subcutaneous fat depth in the multiple regression model used for RFI estimation, as proposed by Basarab *et al.* (2003).

Studies in Corriedale lambs also indicated that RFI tends to be independent of growth, wool production, and genetic resistance to gastrointestinal parasites (Navajas *et al.*, WCGALP 2022, Ferreira *et al.*, 2021), whilst r_P and r_G between RFI and fleece weight in Australian Merino sheep were not significantly different from zero (Marques *et al.*, WCGALP 2022).

Genetics of GHG emissions in beef cattle and sheep

Methane and carbon dioxide (CO_2) emissions are being measured in all lambs with feed intake and efficiency records using Portable Accumulation Chambers (PAC) (Paganoni *et al.*, 2017), leading to a very interesting database that is used to investigate the association between current breeding objectives and

potentially new ones, such as RFI with GHG emissions. Similarly, GreenFeed stations are now installed and recording CH_4 during the feed efficiency tests in Hereford.

The first analysis in Corriedale and Merino lambs showed no-significant phenotypic association between RFI and CH_4 and CO_2 (Navajas *et al.*, WCGALP 2022), although a favourable positive genetic correlation was estimated in Merino which indicates that less efficient lambs are also higher emitters (Marques *et al.*, WCGALP 2022). This is in agreement with other studies in sheep (Paganoni *et al.*, 2017) and beef cattle (Fitzsimons *et al.*, 2013), reinforcing the possibility of using RFI as an indirect selection criterion for reducing GHG emissions. The lack of consistency among studies relies on different factors as discussed by Cantalapiedra-Hijar *et al.* (2018).

A consistent result across studies is the positive association of DMI with RFI, CH_4, as well as with animal productive performance. As mentioned before, several studies indicates that RFI tend to be independent of productive trait (null or low correlation) and moderate to highly correlated with DMI. Associations of similar magnitude have been reported between CH_4 and performance traits, suggesting a potential antagonism that could limit the scope of GHG mitigation strategies. Nonetheless, new CH_4-derived traits such as residual methane emissions and the development of multi-trait selection including DMI and CH_4 are promising alternatives to reduce CH_4 emissions while maintaining animal production (Manzanilla-Pech *et al.*, 2021; Smith *et al.*, 2021).

As DMI is a main driver of animal productivity and CH_4 emissions, disentangling the relationships among these traits is key, and requires considering not only animal genetics but also the role of rumen microbiota. On-going characterization of rumen microbial community and metabolic pathways by metagenomics and metatranscriptomics will contribute to a better understanding of these associations. These analysis in the Uruguayan reference populations will also provide information about its value as predictor of genetic merit.

Improving feed efficiency and methane emission by genetic selection

The complexity of measuring GHG emission and DMI to estimate RFI, clearly indicates the benefits of implementing genomic selection. Because of this, RFI and GHG emissions phenotyping platforms in sheep and beef cattle in Uruguay were designed as the initial reference populations for genomic selection based on national and international research funding. DNA samples are stored at INIA animal DNA bank and have been genotyped mainly with medium density (50k) single nucleotide polymorphism (SNP) panels. An initial binational reference population of 731 Uruguayan Hereford bulls and steers and 1168 Canadian bulls was the basis for the estimation of genomic expected breeding values (GEPD) published since 2017 (Ravagnolo *et al.*, 2018). Research GEPD for RFI and CH_4 have also been estimated in 2021 for Australian Merino, as a first step before incorporating these traits in the breeding programme. Larger reference populations have favourable impact on high accuracies of genomic predictions, which also represent an incentive for breeders to invest in genotyping and RFI and CH_4 phenotyping that accelerates the growth of reference population (Berry *et al.*, 2016). Although the expected benefits of genomic selection are more relevant in difficult-to-measure traits, phenotyping costs and complexity are the main challenges for building the corresponding references populations. Successful implementation and uptake of genomic selection will increase the number of animals in the national genetic evaluation with GEBV for RFI and CH_4, with favourable implications in terms of the applicability of selection indices including these new traits and balancing potential antagonisms between productivity and GHG emissions.

In summary, improving feed efficiency and reducing CH_4 emission by animal selection is achievable as both traits are heritable and phenotyping larger number of animals is now possible, although still costly and limited to experimental research farms or nucleus herds/flocks. Further research is needed for a

more comprehensive understanding of the genetic and phenotypic relationships not only between these groups of traits, but also with productivity in order to optimise genetic progress supporting environmental, economic, and social sustainability. Expanding regional and international collaborations is one of the alternatives to build larger reference population for genomic selection with increased genomic accuracies.

Acknowledgements

The authors acknowledge the funding organizations INIA, Agencia Nacional de Investigación e Innovación (RTS_1_2012_1_3489), the European Union's Horizon 2020 research and innovation program under grant agreement No 772787 (SMARTER) and the SusAn, ICT-AGRI 2 and FACCE ERA-GAS funding bodies (GrassToGas project).

References

Barwick S.A., Henzell A.L., Herd R.M., Walmsley B.J and Arthur P.F. (2019) Genet. Sel. Evol. 51(18) https://doi.org/10.1186/s12711-019-0459-5

Basarab J.A., Price M.A., Aalhus J.L., Okine E.K., Snelling W.M. *et al.* (2003). Can. J. Anim. Sci. 83(2):189-204. https://doi.org/10.4141/A02-065

Berry D.P., Garcia J.F. and Garrick D.J. (2016) Anim. Front. 6(1):32-38.

Cantalapiedra-Hijar G., Abo-Ismail M., Carstens G.E., Guan, L.L., Hegarty R. *et al.* (2018). Animal 12:s321- s335 https://doi.org/10.1017/S1751731118001489

Fitzsimons C., Kenny D.A., Deighton M.H., Fahey A.G., and McGee, M. (2013) J. Anim. Sci. 91(12): 5789–5800. https://doi.org/10.2527/jas2013-695

Koch R., Swiger L., Chambers D., and Gregory K. (1963) J. Anim. Sci. 22(2):486–494.

Manzanilla-Pech C.I.V., Løvendahl, P., Mansan Gordo, D., Difford G.F., Pryce J.E., *et al.* (2021) J. Dairy Sci. 104(8): 8983-9001 https://doi.org/10.3168/jds.2020-19889

Paganoni B., Rose G., Macleay C., Jones C., Brown D.J., *et al.* (2017). J. Anim. Sci. 95(9): 3839–3850. https://doi.org/10.2527/jas.2017.1499

Pravia M.I., Navajas E.A., Macedo F.L., Clariget J., and Luzardo S. (2018). Proc. of the 11th WCGALP, Auckland, New Zealand.

Phocas F., Belloc C., Bidanel J., Delaby L., Dourmad J., *et al.* (2016). Animal 10(11):1749-1759. https://doi.org/10.1017/S1751731116000926

Ravagnolo O., Aguilar I., Crowley J.J., Pravia M.I., Lema M., *et al.* (2018) Proc. of the 11th WCGALP, Auckland, New Zealand.

Smith P.E., Waters S.M, Kenny D.A., Kirwan S.F., Conroy S. *et al.* (2021) J. Anim. Sci. 99(11)

30. Breeding for lowered methane emissions using rumen microbial profiles in sheep

T.P. Bilton[1*], R. Jordan[1], M.K. Hess[1], S.M. Hickey[2], N. Amyes[2], P. Johnson[1], H. Henry[1], H. Baird[1], T. van Stijn[1], B. Veenvliet[1], W. Bain[1], G. Pile[1], T. Watson[2], E. Sandoval[3], J. Peers-Adams[1], B. Bryson[4], A.S. Hess[5], K.G. Dodds[1], J.C. McEwan[1] and S.J. Rowe[1]

[1]AgResearch, Invermay Agricultural Centre, Private Bag 50034, Mosgiel 9053, New Zealand; [2]AgResearch, Ruakura Research Centre, Private Bag 3123, Hamilton 3240, New Zealand; [3]AgResearch, Grasslands Research Centre, Private Bag 11008, Palmerston North 4442, New Zealand; [4]AgResearch, Woodlands Research Farm, RD1, Woodlands, Invercargill 9871, New Zealand; [5]University of Nevada, Reno, Agriculture, Veterinary & Rangeland Sciences, 1664 N. Virginia St. mail stop 202, Reno, NV 89557, USA; timothy.bilton@agresearch.co.nz

Abstract

Reducing methane emission in ruminant animals via breeding requires estimating methane emissions using indirect or proxy measures, as direct measures are often impractical. Rumen microbial community (RMC) profiles have been shown to be associated with methane emissions and so may be a suitable proxy. Using a dataset of 3,640 sheep across 8 genetically linked flocks, we predicted methane emission from RMC profiles and obtained higher prediction accuracies compared to host genomic prediction. We estimated the correlations between methane traits measured directly using portable accumulation chambers (PAC) and predicted from RMC profiles. Moderate to high genetic correlations (0.66 and 0.77) were obtained, while phenotypic correlations were lower (0.32 and 0.35). These results suggest RMC profiles have potential to be used as a proxy for methane emissions in ruminants and that direct and indirect methane measures could be combined within a breeding scheme.

Introduction

A target of 30% reduction in methane emissions by 2030 has been agreed upon by governments and global organizations to combat climate change. Developing methods for reducing ruminant methane emissions while maintaining productivity is therefore vital for the future of the agricultural industry. This can be achieved through classical breeding practices (Rowe *et al.* 2021), but collecting direct measurements of methane is cumbersome and costly. This necessitates indirect or proxy measures of methane. These vary in applicability according to species and system (Negussie *et al.* 2017). One proxy that has utility across species and system is the collection and sequencing of rumen microbial community (RMC) profiles, which have been shown to be associated with methane emissions (Kittelmann *et al.* 2014; Difford *et al.* 2018). In this study, we assess the feasibility of RMC profiles as a proxy for methane emissions and its potential inclusion in commercial breeding scheme as a predictor of methane emissions.

Materials & methods

Experimental animals and protocols applied in this study were approved by the AgResearch Ruakura (Hamilton, NZ) Animal Ethics committee (approvals 14906, 14907 and 14981).

Animals & phenotypes. Data from 5 industry (n=1,281) and 3 research (n=2,359) genetically linked performance recorded sheep flocks were obtained and consisted of 2,136 ewes and 1,504 rams born between 2014 and 2020 (Table 1). Methane emissions were measured using the PAC protocol described in Jonker *et al.* (2018) for all animals when they were between 3 and 12 months of age, except for the born 2019 ewes in flock 4 which were measured at 18 months of age. Animals were measured in lots of 12. Animal information and methane phenotypes were downloaded from the Sheep Improvement Limited database

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_30, © T.P. Bilton *et al.* 2022

Table 1. Animal numbers by flock, sex and year of birth.

Sex	Flock	Year of birth						Total
		2014	2015	2016	2018	2019	2020	
Ewe	1	135	153	160				448
	2	63	146	148	144	94	117	712
	3		92	54		74	135	355
	4					245	231	476
	5						145	145
Ram	1				185	159	500	844
	6				161	168		329
	7				81			81
	8						250	250
Total		198	391	362	571	740	1,378	3,640

(Newman *et al.* 2000). Traits were methane emissions (CH_4 (g/d)) and Carbon Dioxide (CO_2 (g/d)). Scaled CH_4 (g/d) ($ScCH_4$) and the scaled molar $CH_4/(CO_2 + CH_4)$ ratio (ScRatio) traits (scaled by lot mean) were also calculated as described in Jonker *et al.* (2018) to account for variables associated with lot.

Rumen microbial profiles. Rumen samples were collected via stomach intubation immediately after animals were measured through the PAC units and preserved using either TNx2 solution described in Budel *et al.* (2022) (born 2020 animals and all animals in flock 4), or by fast freezing on dry ice and freeze-dried described by Kittelmann *et al.* (2014) and adapted by Hess *et al.* (2020). DNA was extracted using a phenol-chloroform with bead beating protocol (Rius *et al.* 2012) and sequenced using a restricted enzyme-reduced representation sequencing approach using *Pst*I with 384 samples per library as described by Hess *et al.* (2020). Samples were sequenced across multiple flowcells and run on either an Illumina HiSeq2500 or NovaSeq6000. The reference-free pipeline developed by Hess *et al.* (2020) was used to process raw sequence data into tags (unique 65 bp trimmed sequences present in 25% of samples in the training set) to generate individual RMC profiles, which were used to construct a microbial relationship matrix (MRM) by taking the correlation in the cohort adjusted RMC profiles.

Animal genotyping. Animals were genotyped on a variety of nested SNP arrays. SNPs with a call rate of 70% were retained, resulting in 14,923 SNPs in the combined data. A genomic relationship matrix (GRM) was computed in KGD (Dodds *et al.* 2015) using VanRaden method 1 with non-missing SNPs for each matrix entry and assuming missing data is at random.

Statistical models. Data was split into (a) a training set consisting of the 2,262 animals born before 2020 and (b) a validation set consisting of the 1,378 animals born in 2020. Using animals from the training set, univariate mixed models of the form:

$$y_{ijkl} = \mu + aod_j + brr_k + bdev_i + cg_l + a_i + e_{ijkl} \quad (1)$$

$$y_{ijkl} = \mu + aod_j + brr_k + bdev_i + cg_l + m_i + e_{ijkl} \quad (2)$$

$$y_{ijkl} = \mu + aod_j + brr_k + bdev_i + cg_l + a_i + m_i + e_{ijkl} \quad (3)$$

were fitted, where μ is the overall mean, aod_j is the effect of the j^{th} age of dam (1, 2, 3, 4+) group, brr_k is the effect for the k^{th} birth/rear rank (1/1, 2+/1, <3/2, 3+/2, 3+/3+) group, $bdev_i$ is the birth deviation from the flock and birth year mean for animal i, cg_l is the effect for the l^{th} contemporary group based on the

combination of flock, year of birth and sex, y_{ijkl} denotes the PAC methane trait ($ScCH_4$ or ScRatio), $m_i \sim N(0,\sigma_m^2 M)$, $a_i \sim N(0,\sigma_g^2 G)$, $e_{ijkl} \sim N(0,\sigma_e^2 I)$, M denotes the MRM, G denotes the GRM and I is the identity matrix. We refer to the microbial values, m_i, as the 'RMC methane trait' since it provides an estimate of the methane trait y_{ijkl} ($ScCH_4$ or ScRatio) from the RMC profiles. Predictions of the microbial values ($\hat{m}_i$) and the animals direct genomic breeding values ($\hat{a}_i$) were made for the born 2020 animals in the validation set. Prediction accuracies were computed as (a) the correlation between $\hat{m}_i$ or $\hat{a}_i$ and the adjusted phenotype (y_i^*) defined as the residuals from the linear model:

$$y_{ijkl} = \mu + aod_j + brr_k + bdev_i + cg_l + e_{ijkl} \quad (4)$$

fitted using both the training and validation sets, and (b) the accuracy of estimated breeding values (r_{eBV}) or the accuracy of estimated microbial values (r_{eMV}). To assess the heritability and genetic correlation of the PAC and RMC methane traits, a bivariate model of the form:

$$[\hat{m}_i, y_i^*] = \mu + a_i + e_i \quad (5)$$

was fitted using only the animals from the born 2020 validation set. All models were fitted in ASREML v4.2 (Gilmour *et al.* 2015).

Results

Prediction accuracies of the PAC methane traits from the RMC profiles and host genomics are given in Table 2. Higher accuracies were obtained from the RMC profiles for both traits compared to using host genomics. For the combined model (G+M), prediction accuracies were around zero for the host genomics and relatively unchanged for the RMC profiles. The accuracies of the estimated breeding and microbial values were similar across models ranging between 28 and 34%, although the standard errors of r_{eBV} were about twice those of r_{eMV}.

Table 3 gives the genetic parameter estimates from the bivariate analysis using the born 2020 validation animals. Heritability estimates for PAC methane traits were slightly higher than previously reported by Jonker *et al.* (2018) and approximately double the heritability estimates of the RMC methane trait. The genetic correlations between the PAC and RMC methane traits were moderate to high at 0.76 ($ScCH_4$) and 0.66 (ScRatio), while as expected, the phenotypic correlation estimates were lower at around 0.32 (ScCH4) and 0.35 (ScRatio). Given the variation in the data and modest number of animals used in this study, these results show clear potential for the use RMC profiles as a proxy for methane phenotypes.

Table 2. Born 2020 validation animals prediction accuracies for PAC methane traits predicted from RMC profiles (M), host genomics (G) or both combined (G+M).

Trait	Equation	Model	$\rho(\hat{a}_i, y_i^*)$[1]	$\rho(\hat{m}_i, y_i^*)$[1]	r_{eBV}^2	r_{eMV}^2
$ScCH_4$	1	G	0.092		0.32±0.09	
	2	M		0.335		0.31±0.06
	3	G+M	-0.042	0.334	0.30±0.11	0.31±0.06
ScRatio	1	G	0.215		0.34±0.09	
	2	M		0.313		0.28±0.06
	3	G+M	-0.039	0.310	0.34±0.12	0.29±0.06

[1] $\rho(x,y)$ denotes the Pearson correlation between x and y. $\hat{a}_i$ is the predicted breeding value, $\hat{m}_i$ is the predicted microbial value, and y_i^* is the adjusted phenotypes computed in Equation 4 for animal i.
[2] r_{eBV} is the accuracy of the estimated breeding values computed as $(1\text{-}s_i^2/\sigma_g^2)^{0.5}$ and r_{eMV} is the accuracy of the estimated microbial values computed as $(1\text{-}s_i^2/\sigma_m^2)^{0.5}$, where s_i^2 is the prediction error variance.

Table 3. Heritability, genetic correlations and phenotypic variances from bivariate analysis of measured PAC and predicted RMC methane traits using the born 2020 validation animals.

Parameter	$ScCH_4$		ScRatio	
	PAC	RMC	PAC	RMC
Heritability	0.39±0.03	0.19±0.07	0.31±0.05	0.18±0.05
Phenotypic variance	3.01±0.12	0.43±0.02	0.48±0.02	$5.4e^{-2}$±$5.1e^{-3}$
Genetic correlation	0.76±0.14		0.66±0.13	
Phenotypic correlation	0.35±0.03		0.32±0.03	

Discussion

RMC profiling has been shown to be predictive of methane emissions in sheep and cattle, however, direct methane measures of cattle grazing pasture have proven to be extremely difficult and expensive to acquire. These results offer direct evidence that microbial predictors are suitable as a proxy for predicting methane emissions. The method outlined here provides a potential approach to industry implementation, whereby animals will have different information (i.e. pedigree, genotypes, RMC profiles, methane traits) available for genetic evaluations. As shown in this study, implementation could be achieved through a 2-stage analysis, initially converting RMC profiles into predicted methane emissions with a subsequent ssBLUP step.

Acknowledgements

Financial support was provided by the Pastoral Greenhouse Gas Research Consortium (PGgRc), the New Zealand Agricultural Greenhouse Gas Research Centre (NZAGRC) via the programme 'Breeding low emitting ruminants MET5.1', and the Ministry of Business, Innovation and Employment (MBIE) via the 'Mapping the New Zealand ruminotype landscape'. We thank Focus Genetics, Beef and Lamb Genetics, Russell Proffit and all breeders involved for access to their animals, samples and data.

References

Budel J.C.C., Hess M.K., Bilton T.P., Henry H., Dodds K.G. *et al.* (2022) bioRxiv: 2022.2002.2010.480010. https://doi.org/10.1101/2022.02.10.480010

Difford G.F., Plichta D.R., Løvendahl P., Lassen J., Noel S.J. *et al.* (2018) PLoS Genetics 14(10): e1007580. https://doi.org/10.1371/journal.pgen.1007580

Dodds K.G., McEwan J.C., Brauning R., Anderson R.M., van Stijn T.C. *et al.* (2015) BMC Genomics 16: 1047. https://doi.org/10.1186/s12864-015-2252-3

Gilmour A.R., Gogel B.J., Cullis B.R., Welham S.J. and Thompson R. (2015) *ASReml user guide release 4.2 structural specification*. VSN International Ltd, Hemel Hempstead, HP1 1ES, UK.

Hess M.K., Rowe S.J., Van Stijn T.C., Henry H.M., Hickey S.M. *et al.* (2020) PLoS One 15(4): e0219882. https://doi.org/10.1371/journal.pone.0219882

Jonker A., Hickey S.M., Rowe S.J., Janssen P.H., Shackell G.H. *et al.* (2018) J. Anim. Sci. 96(8): 3031-3042. https://doi.org/10.1093/jas/sky187

Kittelmann S., Pinares-Patiño C.S., Seedorf H., Kirk M.R., Ganesh S. *et al.* (2014) PLoS One 9(7): e103171. https://doi.org/10.1371/journal.pone.0103171

Negussie E., de Haas Y., Dehareng F., Dewhurst R.J., Dijkstra J. *et al.* (2017) J. Dairy Sci. 100(4): 2433-2453. https://doi.org/10.3168/jds.2016-12030

Newman S.-A., Dodds K.G., Clarke J.N., Garrick D.J. and McEwan J.C. (2000) Proc. N. Z. Soc. Anim. Prod. 60: 195-197.

Rius A.G., Kittelmann S., Macdonald K.A., Waghorn G.C., Janssen P.H. *et al.* (2012) J. Dairy Sci. 95(9): 5024-5034. https://doi.org/10.3168/jds.2012-5392

Rowe S.J., Hickey S.M., Johnson P.L., Bilton T.P., Jonker A. *et al.* (2021) Proc. Assoc. Advmt. Anim. Breed. Genet. 24: 15-18.

31. Do we estimate dairy cows' methane production precisely enough?

M. Pszczola* and T. Strabel

Poznan University of Life Sciences, Department of Genetics and Animal Breeding, Wolynska 33, 60-637 Poznan, Poland; marcin.pszczola@puls.edu.pl

Abstract

For sniffers, quantifying methane (CH_4) production relies on milk production, body weight, and the stage of the cow's pregnancy. The relationship between these traits and CH_4 production is assumed to be constant. We assessed the impact of change in the relationships between body weight CH_4 production on selection. Data on 483 (330 genotyped) Polish Holstein-Friesian cows for CH_4 and CO_2 concentrations measured with sniffer and production data were available. We analysed four scenarios in which coefficients relating body weight to CH_4 production varied. Changes in the values of the coefficients had a substantial impact on the daily CH_4 production level, genetic parameters, and possible selection decisions.

Introduction

The sniffer method measures methane (CH_4) concentrations in the air sampled from dairy cows in the feeding bin of automated milking system (AMS), capturing only part of the exhaled or eructated air. Therefore, to quantify the amount of CH_4 produced by a cow, CO_2 production (Madsen *et al.*, 2010) is used. It allows converting CH_4 concentration to production. This conversion requires measuring CH_4 and CO_2 concentrations in the breath of the cows and estimating the daily production of CO_2. The daily CO_2 production is calculated based on converting heat production (HPU) to CO_2 according to Pedersen *et al.* (2008) by employing the formula of Madsen *et al.* (2010). This requires information on cow's milk production, body weight (BW), and stage of pregnancy. Therefore, the final phenotype, i.e. CH_4 production, relies on formula parameters (FP) relating the three mentioned cows' characteristics to heat production (Petersen and Sällvik, 2002). Most studies on methane production use the same values for FP (e.g. Lassen and Løvendahl, 2016; Pszczola *et al.*, 2017; Manzanilla-Pech *et al.*, 2021). However, the FP values may vary depending on the cow's age, production, physiological, or health status of an animal (Petersen and Sällvik, 2002). Using different parameters in FP may result in imprecise CH_4 production estimates and, consequently, impact selection decisions. This study aimed to assess the sensitivity of the CH_4 production to changes in FP related to BW.

Materials & methods

Data were collected at two commercial farms on 483 Polish Holstein-Friesian cows in the following periods: Farm 1: Dec 2, 2014, to Feb 3, 2016, and Jun 1, 2016, to Sep 17, 2016, and on Farm 2: Feb 5, 2016, to Mar 14, 2016. The gas measurements were recorded during milking in an AMS (Lely Astronaut A4; Lely Holding, Maassluis, the Netherlands) using an FTIR analyser (GASMET 4030; Gasmet Technologies Oy, Helsinki, Finland). The remaining data on cows were obtained from the AMS system. We analysed only data collected between 5 and 305 days of lactation. In total, we obtained 34,429 daily records. Cows were fed *ad libitum* with a partial mixed ration (PMR), and additional concentrate was provided in AMS depending on the animal's productivity. A detailed description of the study can be found in Pszczola *et al.* (2017). Out of all the phenotyped cows, 330 were genotyped with the Illumina BovineSNP50 v2.0 BeadChip (Illumina Inc., San Diego, CA, USA); see Pszczola *et al.* (2018) for details.

Predicting daily CH4 production. The daily CH_4 production (expressed in grams) was predicted from the formula presented in Madsen *et al.* (2010) as:

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_31, © M. Pszczola and T. Strabel 2022

$$predictedCH_4 = \frac{CH_4}{CO_2} * \frac{c_1 BW^{c_2} + 22MY + 1.6*10^{(-5)} PREG^3}{1000*180*24} * 0.668, \quad (1)$$

Scenarios. We analyzed four scenarios in which FP c_1 and c_2 in formula (1) were modified as follows: (1) *Base scenario* (as in Madsen *et al.*, 2010): c_1=5.6, and c_2=0.75; (2) *scenario C1*: c_1=7.64, and c_2=0.75; (3) *scenario C2*: c_1=5.6, and c_2=0.64; and (4) *scenario C1 & C2*: c_1=7.64, c_2=0.61. For each scenario CH_4 production was estimated.

Variance components estimation. Variance components for CH_4 production were estimated with BLUPF90 suite of programs (Misztal *et al.*, 2002) by combining pedigree and genotype information using the formula:

$$CH_{4_{ijkl}} = LAC_j \times \sum_{n=1}^{4} \beta_n DIM_{kn} + YK_l + \sum_{n=1}^{3} \boldsymbol{a}_{in} DIM_{kn} + \sum_{n=1}^{3} \boldsymbol{pe}_{in} DIM_{kn} + \boldsymbol{e_{ijkl}} \quad (2)$$

Where the $CH_{4_{ikl}}$ was the estimated and adjusted average daily CH_4 output of the i^{th} cow in the j^{th} lactation, the k^{th} day in milk (*DIM*), and the l^{th} year–week of the measurement, where year–week of the measurement (*YK*), where *YK* are farm-specific; β_n are fixed regression coefficients. Lactation (*LAC*) had two levels: 1 and 2+, in which records from up to lactation eight were included. The general lactation curve was modelled using third-order Legendre polynomials and random animal (*a*) and permanent environment (*pe*) effects with second-order Legendre polynomials and *e* is the residual term. The adjustment of $CH_{4_{ikl}}$ made prior to estimating variance components was the pre-correction for diurnal variation and AMS unit followed by averaging observations within one day as described in Pszczola *et al.* (2017). Next to variance components and heritability, we estimated breeding values using ssBLUP procedure (Aguilar *et al.*, 2010) and calculated the 305 breeding values (GEBV). The 305 GEBV were calculated only for genotyped cows with their records.

Results

Modification of the c_1 and c_2 substantially influenced the average daily CH_4 production level and its variation (Table 1). In scenario C1, cows produced, on average, 73 g of CH_4 per day more than in the base scenario and nearly 200 g more than in scenario C2.

Although variance components estimates had similar shapes across lactation for all scenarios (Figure 1), they differed by the levels. The Base scenario yielded intermediate values, while the highest values were obtained for scenario C1 and the lowest for C2.

Heritability of CH_4 production (Figure 2) had a similar pattern for all the scenarios forming a bell-shaped curve with the peak value around mid-lactation and corresponding to Pszczola *et al.* (2017). Although the differences at the beginning and the end of lactation were negligible, the broad range of heritabilities at the peak was obtained, ranging from 0.25 (scenario C1) to 0.35 (scenario C2) (Table 1).

Table 1. Daily methane production (g) and heritability levels across different scenarios.

Scenario[1]	Mean	SD	CV%	h^2
Base	427	83.3	19.5	0.21
C1	500	95.4	19.1	0.20
C2	311	65.4	21.1	0.27
C1 & C2	340	69.0	20.3	0.26

[1] Names of the scenarios indicate the parameter modified in the calculations.

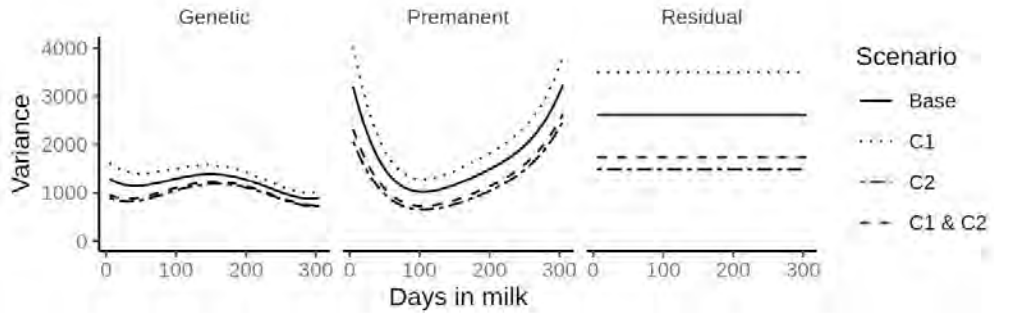


Figure 1. Variance components estimated for daily CH_4 production of dairy cows across days in milk for different parameters used for quantifying CH_4 production across different scenarios.

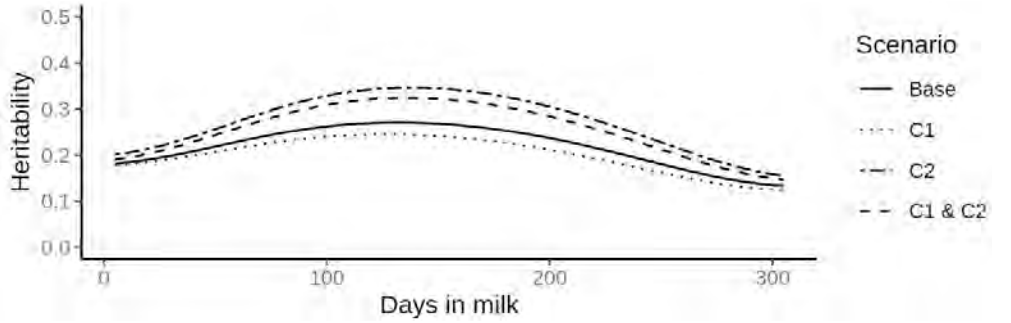


Figure 2. Heritability for daily CH_4 production of dairy cows across days in milk for different parameters used for quantifying CH_4 production across different scenarios.

The substantially different heritabilities had an unavoidable impact on CH_4 GEBV. The correlation between GEBV from the Base and other scenarios are shown in Table 2. The correlation between GEBV for the base scenario C1 scenarios was high (0.996). Whereas the correlation between GEBV for the base scenario and C2 was as low as 0.981, and between C1 & C2 was even lower (0.960) and led to re-ranking of animals which would impact the possible selection decisions. The lists of top 20 cows (lowest GEBV for CH_4 production) for the Base scenario and C1 shared two cows, and between base and C2 or base and C1 & C2 zero cows. For the top 100 cows, these numbers were 39, 34, and 36.

Table 2. Correlations between 305 GEBV for methane production of dairy cows for different parameters used for quantifying CH_4 production across different scenarios.

Scenario[1]	C1	C2	C1 & C2
Base	0.996	0.981	0.989
C1		0.960	0.972
C2			0.999

[1] Names of the scenarios indicate the parameter modified in the calculations.

[2] In the Base scenario, no parameters were modified.

Discussion

Obtaining accurate phenotypes for CH_4 production is challenging. Sniffer methods are promising technology allowing for gathering many observations for a high number of cows. Although sniffers are not exact, they provide phenotypes enabling genetic analyses that give the basis for selection for lower CH_4 production (e.g. Lassen and Løvendahl, 2016; Pszczola *et al.*, 2017; Manzanilla-Pech *et al.*, 2021). We showed that using alternative (to Madsen *et al.*, 2010) FP's for BW have a clear impact on CH_4 production with regard to: (1) assessment of the daily CH_4 production; (2) values of the estimated genetic parameters; and, in consequences, (3) potential selection decisions. Therefore, we suggest that the conversion of the CH_4 concentration to production based on the analysed equation should be employed with care. The parameters, which are sensitive to cow's age, production, physiological, or health status, should be precisely adjusted to analysed individuals to reflect the actual status of the population of interest.

References

Aguilar I., Misztal I., Johnson D.L., Legarra A., Tsuruta S. *et al.* (2010) J. Dairy Sci. 93(2):743-752. https://doi.org/10.3168/jds.2009-2730

Lassen, J., and Løvendahl P. (2016) J. Dairy Sci. 99(3):1959–1967. https://doi.org/10.3168/jds.2015-10012.

Madsen J., Bjerg B.S., Hvelplund T., Weisbjerg M.R., and Lund P. (2010) Livest. Sci. 129(1-3):223–227. https://doi.org/10.1016/j.livsci.2010.01.001.

Manzanilla-Pech C.I.V., Løvendahl P., Mansan Gordo D., Difford G.F., Pryce J.E., *et al.* (2021). J. Dairy Sci. 104(8):8983–9001. https://doi.org/10.3168/jds.2020-19889.

Misztal, I., Tsuruta S., Strabel T., Auvray B., Druet T., *et al.* (2002) Proc. of the 7th WCGALP, Montpellier, France.

Pedersen S., Blanes-Vidal V., Joergensen H., Chwalibog A., Haeussermann A., *et al.* (2008) E-journal CIGR 10:1–19.

Peters, R.H. (1983) The Ecological Implications of Body Size; Cambridge University Press: Cambridge, UK

Petersen S., and Sällvik K. (2002) Heat and moisture production at animal and house levels. In: S.K. Pedersen S. (eds) Climatization of animal houses. Research Centre Bygholm, Danish Institute of Agricultural Sciences, Denmark, pp. 10.

Pszczola M., Rzewuska K., Mucha S., and Strabel T. (2017) J. Anim. Sci. 95(11):4813–4819. https://doi.org/10.2527/jas2017.1842.

Pszczola M., Strabel T., Mucha S., and Sell-Kubiak E. (2018) Sci. Rep. 8:15164. https://doi.org/10.1038/s41598-018-33327-9.

32. Heritability of methane emission in Norwegian Red cows based on measures from GreenFeed in commercial herds

K.B. Wethal[1], G.F. Difford[2], K. Winnberg[2], E. Norberg[2] and B. Heringstad[1,2*]

[1]Geno Breeding and A.I. Association, Storhamargata 44, 2317 Hamar, Norway; [2]Department of Animal and Aquacultural Sciences, Faculty of Biosciences, Norwegian University of Life Sciences, P.O. Box 5003, 1432 Ås, Norway; bjorg.heringstad@nmbu.no

Abstract

The aim of this study was to estimate heritability of methane emissions for Norwegian Red dairy cows. Methane measured by GreenFeed was available from 12 commercial herds. Data was collected from February to October 2021 and the final dataset had a total of 62,057 records from 252 cows. Methane emission from individual visits was analysed. The average methane emission was 441 g/day. A linear animal repeatability model with fixed effects of parity and lactation week, and random effects of herd-testday, animal, and permanent environment was used to estimate variance components. The estimated heritability was 0.22 and repeatability 0.36.

Introduction

International agreements and commitments to reduce greenhouse-gas emissions imply that all sectors have to contribute to meet climate targets. One way to reduce the environmental footprint of dairy production is by breeding for reduced methane (CH_4) emissions. Implementing reduced CH_4 emission in the breeding program would have a direct and long-term environmental impact. Previous studies have reported genetic variation for methane emissions in Holstein dairy cattle (Lassen and Difford, 2020). In order to assess the genetic potential for reduced CH_4 emission in the Norwegian Red population, Geno breeding and A.I. association (www.norwegianred.com) implemented collection of measures of cows CH_4 emissions in commercial dairy herds. So far most genetic research on methane emission has been conducted in small and tightly controlled research herds which cannot encompass the full genetic variation of the population and tend to have multiple dietary treatments which complicate genetic evaluations (Tempelman *et al.*, 2015). For selection to be viable and genetic parameters to be relevant, large scale phenotype collection is needed in the commercial environments in which animals are expected to perform. To this end, Geno has installed GreenFeed (www.c-lockinc.com) equipment for CH_4 recording for individual cows at commercial dairy farms in the mid- and south-eastern part of Norway. The aim of this first initial genetic analyses was to estimate heritability of methane emissions for Norwegian Red dairy cows.

Materials & methods

The GreenFeed unit measure gas fluxes of CH_4 (and other gases) from individual cows. The cows visit the unit voluntary, where they are offered a small amount of concentrate to entice visits. Each cow may have multiple daily visits, but maximum five feedings per day. Each visit results in an estimate of daily CH_4 emission. Only visits of at least 2 minutes with 'good data', defined based on head position or proximity inside the crib, were defined as reliable and retained for further analysis.

Methane emission data was collected from February to October 2021 in 12 herds with loose housing and automatic milking systems (Lely or DeLaval). Additional information on the cows were from the Norwegian Dairy Herd Recording System. Norwegian Red cows that had at least 5 GreenFeed records were included, and data from the period from calving to 350 days in milk were used. The final dataset had a total of 62,057 records (1 record per visit) from 252 cows. The trait analyzed was daily CH_4 emission, gram per day (Each visit results in an estimate of daily CH_4 emission).

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_32, © K.B. Wethal *et al.* 2022

A linear animal repeatability model was used to estimate variance components using the DMU software (Madsen and Jensen, 2012). The model included fixed effects of parity (1, 2, 3, and >3) and lactation week (from 1 to 50) and random effects of herd-testday, animal, and permanent environment. The pedigree of cows with data was traced as far back as 6 generations and included a total of 3,791 animals.

Results

The average (standard deviation) methane emission was 441 (114) gram per day. The distribution of the phenotypic records followed an approximate normal distribution (Figure 1).

Methane emissions vary with stage of lactation. Figure 2 shows the solutions for fixed effect of lactation week and illustrates how CH_4 emissions by lactation stage follows the same shape as a lactation curve for milk yield.

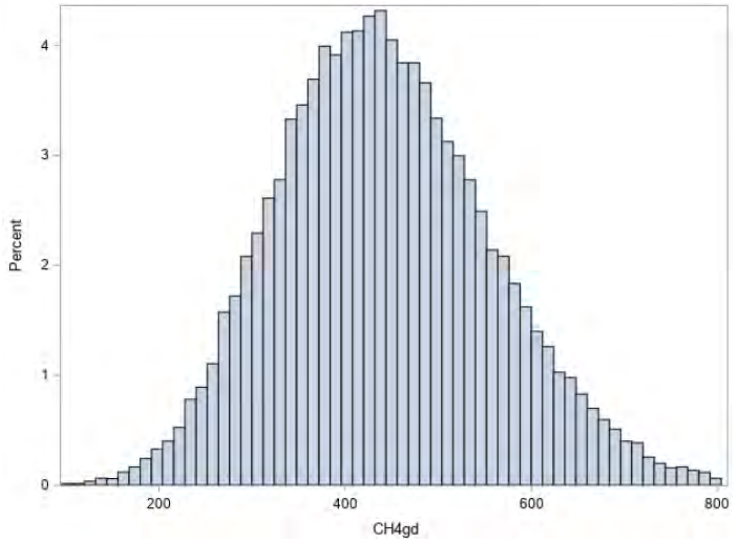


Figure 1. Distribution of phenotypic records of methane emission (CH_4gd) for Norwegian Red cows (1 record per visit at the GreenFeed unit, each visit gives an estimate of daily methane emission).

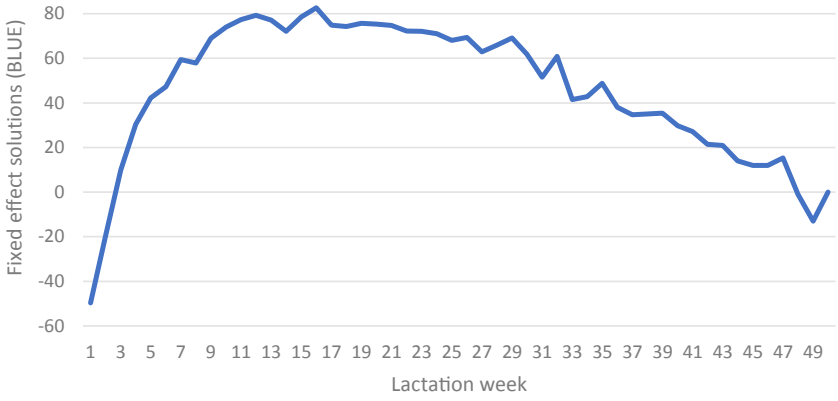


Figure 2. Fixed effect solutions for lactation week, effect on daily methane emission (gram per day) (standard errors were between 7 and 9).

The estimated variance components are given in Table 1. The corresponding heritability of CH_4 emission in Norwegian Red was 0.22 and repeatability 0.36.

Table 1. Estimated variance components with standard error (SE) for daily methane emission.

Variance component	Estimate	SE
Herd-testday	1,024.5	47.4
Additive genetic	2,693.7	1,002.1
Permanent environment	1,670.9	836.9
Residual	6,809.9	39.3

Discussion

Due to challenges related to large scale recording of CH_4 there are so far relatively few estimates of genetic parameters for direct CH_4 emission in dairy cows. This is, as far as we know, the first heritability estimate based on data from GreenFeed in commercial herds, and the first for the Norwegian Red breed. Our heritability estimate agrees well with previous published results, although not directly comparable due to differences in breed and phenotyping method/equipment. A review by Lassen and Difford (2020) showed that heritability estimates of CH_4 emissions measured by 'sniffer' in Holsteins varied from 0.11 to 0.45, with most estimates from 0.21 to 0.27. Our heritability estimate is larger than previous estimates of 0.15 and 0.11 for daily CH_4 emission in 100 Finnish Ayrshires and 409 grazing Italian Holsteins, respectively, both recorded using GreenFeeds in research herds (Wallace *et al.*, 2019). However, it is lower than 0.31±0.15 reported for weekly averaged CH_4 production in 305 Holstein cows from research herds in Canada and Switzerland in a joint evaluation (Manzanilla Pech *et al.*, 2021).

This initial analysis of CH_4 emissions in Norwegian Red cows shows that there is substantial genetic potential for reducing CH_4 emissions from dairy cows by breeding. With GreenFeed units installed in collaborating commercial herds and a system for data collection in place, we will soon have access to larger datasets for further genetic studies of CH_4 emissions. We will estimate genetic correlations between methane production and important traits in the breeding goal. Zetouni *et al.* (2018) found low genetic correlations between methane production and health, fertility and body type traits in Danish Holstein. However, they concluded that further studies based on larger dataset are needed to understand these relationships. Increased knowledge of the genetic relationship between traits related to climate, traditional efficiency traits, including feed efficiency, but also animal welfare and health traits is needed before reduced methane emission can be implemented in Norwegian Reds' future breeding goal.

Acknowledgements

We want to thank the dairy farmers hosting the GreenFeed units and providing data, Geno and Tine for access to data, and the Norwegian Research Council for funding of research project (project 309451).

References

Lassen, J. and Difford, G.F. (2020). Animal, Vol 14 (Suppl 3): 473-483. https://doi.org/10.1017/S1751731120001561

Madsen, P. and Jensen, J. (2012). DMU. Version 6, release 5.1.

Manzanilla-Pech, C.I.V., Mansan Gordo, D., Difford, G.F., Pryce, J.E., Schenkel, F. *et al.* (2021). J Dairy Sci 104 (8): 8993-9001. https://doi.org/10.3168/jds.2020-19889

Tempelman, R.J., Spurlock, D.M., Coffey, M., Veerkamp, R.F., Armentano, L.E. *et al.* (2015). J Dairy Sci 98 (3): 2013-2026. https://doi.org/10.3168/jds.2014.8510

Zetouni, L., Kargo, M., Norberg, E. and Lassen, J. (2018). J Dairy Sci 101(3): 2273-2280. https://doi.org/10.3168/jds.2017-13402.

Wallace, R.J, Sasson, G., Garnsworthy, P.C., Tapio, I., Gregson, E. *et al.* (2019). Science Advances 5 (7). https://doi.org/10.1126/sciadv.aav8391

33. Heritability and genetic correlations of enteric methane emissions of dairy cows measured by sniffers and GreenFeed

A.E. van Breukelen[1*], M.N. Aldridge[1], R.F. Veerkamp[1], L. Koning[2], L.B. Sebek[2] and Y. de Haas[1]

[1]Animal Breeding and Genomics, Wageningen University & Research, P.O. Box 338, 6700 AH Wageningen, the Netherlands; [2]Animal Nutrition, Wageningen University & Research, P.O. Box 338, 6700 AH Wageningen, the Netherlands; anouk.vanbreukelen@wur.nl

Abstract

Before methane (CH_4) emission can be mitigated with animal breeding, breath measurements have to be recorded on a large number of cows. Our aim was to estimate heritabilities for, and a genetic correlation between, CH_4 recorded by GreenFeed and sniffers. Repeated records were available for CH_4 production (g/cow/day) by GreenFeed and for CH_4 concentration (ppm) by sniffers. The data included 24,284 GreenFeed daily means from 822 cows, 172,948 sniffer daily means from 1,800 cows, and 1,787 daily means from both devices on the same day from 75 cows. Additionally, records were averaged per week. The datasets were analyzed using bivariate animal models. The results show that CH_4 emissions recorded by either device has a moderate heritability (0.18-0.37). Furthermore, the genetic correlation between weekly mean CH_4 recorded by GreenFeed or by sniffers was high (0.77). This suggest that the measurements can be used in the same genetic evaluations.

Introduction

Before enteric methane (CH_4) emissions of dairy cows can be mitigated through (genomic) selection, breath measurements have to be recorded on thousands of dairy cows to set up a reference population (de Haas *et al.*, 2017). To increase the number of phenotyped cows, it is beneficial if CH_4 records from common recording devices can be used in the same genetic evaluation. For genetics research, CH_4 is commonly recorded by spot-sampling devices called 'sniffers', because of its ability to phenotype a large number of cows cost-effectively (Madsen *et al.*, 2010). The devices use infrared spectroscopy to continuously measure enteric CH_4 and carbon dioxide (CO_2) concentrations (ppm) inside the feed bin of milking robots (AMS). For sniffer recorded CH_4 concentrations, heritabilities have been estimated for the mean CH_4 concentration, which ranged from 0.11 to 0.26 (Lassen and Difford, 2020). Spot-sampling devices that are commonly used in nutritional studies are GreenFeed units (GF). Unlike sniffers, the units record air flux, head position of the cow, wind speed and temperature, from which the CH_4 production in grams/cow/day can be estimated (Hammond *et al.*, 2016). However, this comes at a price and therefore the devices have a ten times higher investment, and also high running costs, which are limiting the number of cows that can be recorded. To date, heritability estimates for GF recorded CH_4 of dairy cows have not been published. Similarly, genetic correlations between CH_4 measured by sniffers and GF are not known yet. In this study, our aim was to: (1) estimate heritabilities for GF recorded CH_4 production; and (2) investigate if sniffer and GF measurements can be combined in genetic evaluations.

Materials & methods

Enteric CH_4 was recorded with GF units (C-lock Inc. Rapid City, SD, US) placed in the barn or in pasture, and by sniffers (WD-WUR v1.0, manufactured by Carltech BV), which were installed in close proximity to the AMS with an air inlet leading from the feed bin of the AMS. The GF units were exchanged between 16 farms and recorded for short periods of time on each farm, ranging from 14 days up to 90 days, between September 2018 and March 2020. Sniffers continuously recorded CH_4 emissions, for at least 64 days and up to 436 days (farm dependent), between March 2019 and September 2020, on 15 farms. A total of six farms were equipped with both devices, of which four farms had records from both devices on the same days. The

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_33, © A.E. van Breukelen *et al.* 2022

GF units, made 161,825 recorded visits on 1,184 cows, which were transformed to CH_4 emissions in g/cow/day (g/day) by C lock-inc. (using the method described by Huhtanen *et al.* (2015)). A detailed description of the general operation procedure of GF units is given by Velazco *et al.* (2017). The sniffers made 470,266 recorded visits on 2,413 cows. Sniffer records were an average of concentration (ppm) measurements taken every ten to thirty-five seconds (varying between devices), and during the second and up to fifth minute of an AMS visit. The recorded cows were of mixed parities and lactation stages. Genotype data, pedigree data, and other phenotypic data (breed composition and calving dates) were made available by the cooperative cattle improvement organization CRV (Arnhem, the Netherlands). In total 1,817 animals were genotyped with 76,438 SNPs.

Data processing. The CH_4 datasets were filtered to only include records for cows up to 305 days in milk, to correctly match the CH_4 records to calving dates and the corresponding parity. Furthermore, cows that were less than 75% Holstein, were removed from the dataset. The pedigree was pruned to only include animals with CH_4 records and their ancestors with the package 'optiSel' in R v3.6.1 (R Core Team, 2019; Wellmann, 2020). The pruned pedigree composed of 41,290 animals, of which 1,800 animals had a sniffer phenotype, 822 animals had a GF phenotype, and 75 animals had both.

The GF and sniffer CH_4 records were averaged and combined in one dataset with daily and in one dataset with weekly means, where records with less than three records per cow per week were discarded for the latter. The dataset with daily means composed of 24,284 GF records from 822 cows and 172,948 sniffer records from 1,800 cows. The combined daily dataset included 1,787 records from both devices on the same day from 75 cows. The dataset with weekly means composed of 4,345 GF records from 724 cows and 31,579 sniffer records from 1,744 cows. The combined weekly dataset included 333 records from both devices in the same week from 73 cows.

Statistical analysis. The pedigree and genotype data were combined in a genetic relationship matrix (H^{-1} matric), following the method of Aguilar *et al.* (2010) and Christensen and Lund (2010) using calc_grm version r1.143 (Calus and Vandenplas, 2016). The H^{-1} matrix comprised all 41,290 pedigreed animals. The CH_4 data was analyzed using animal models, including repeated records, using ASReml 4.2 (Gilmour *et al.*, 2015). Univariate models were used to derive starting values for bivariate models (results not shown). Thereafter, heritabilities, repeatabilities, genetic correlations, and phenotypic correlations were estimated using the following bivariate model:

$$\begin{bmatrix}\mathbf{y}_1\\ \mathbf{y}_2\end{bmatrix} = \begin{bmatrix}\mathbf{X}_1 & 0\\ 0 & \mathbf{X}_2\end{bmatrix}\begin{bmatrix}\mathbf{b}_1\\ \mathbf{b}_2\end{bmatrix} + \begin{bmatrix}\mathbf{Za}_1 & 0\\ 0 & \mathbf{Za}_2\end{bmatrix}\begin{bmatrix}\mathbf{a}_1\\ \mathbf{a}_2\end{bmatrix} + \begin{bmatrix}\mathbf{Zp}_1 & 0\\ 0 & \mathbf{Zp}_2\end{bmatrix}\begin{bmatrix}\mathbf{pe}_1\\ \mathbf{pe}_2\end{bmatrix} + \begin{bmatrix}\mathbf{e}_1\\ \mathbf{e}_2\end{bmatrix} \quad (1)$$

where y_i is a vector with records on trait *i* (GF CH_4 and sniffer CH_4, day/week); b_i is a vector containing fixed effects for trait *i*, which were farm×year×week of the measurement, days in milk, parity (from parity 1 to 4, where 4 is parity four or higher); a_i is a vector containing additive genetic effects for trait *i*; pe_i is a vector containing permanent environmental effects for trait *i*; and e_i is a vector with the residuals for trait *i*; X_i, Za_i and Zp_i are incidence matrices linking the records in y_i to the fixed effects, the additive genetic effects, and the permanent environmental effects, respectively. The additive genetic, permanent environmental and residual effects for all traits were assumed normally distributed with a mean of zero, a genetic variance of σ^2_{ji} for random effect *j* and trait *i*, and a genetic covariance between two traits of σ_{j1j2}:

$$\begin{bmatrix}j_1\\ j_2\end{bmatrix} \sim N\left[\binom{0}{0}, \mathbf{I} \otimes \begin{pmatrix}\sigma^2_{j_1} & \sigma_{j_1 j_2}\\ \sigma_{j_1 j_2} & \sigma^2_{j_2}\end{pmatrix}\right].$$

Results

The mean (±SD) CH_4 emission measured by GF was 436 g/day (±120) per day and 436 g/day (±97) per week. The mean for sniffers was 322 ppm (±251) per day and 325 ppm (±219) per week. The heritability and repeatability were higher for weekly mean CH_4 than for daily mean CH_4, both for CH_4 measured by GF and sniffers (Table 1), due to a reduction of residual variance after averaging per week. The heritabilities and the repeatabilities for GF measured CH_4 and sniffer measured CH_4 were similar.

Phenotypic and genetic correlations between GF and sniffer measured CH_4 were moderately positive, except for the genetic correlation between the weekly mean CH_4 emission measured by GF and sniffer, which was high (0.77±0.14) (Table 2). Both for daily and weekly mean measurements, the genetic correlations were higher than the phenotypic correlations. Furthermore, the genetic correlations between GF and sniffer measured CH_4 increased after averaging emissions to means per week.

Discussion

This study is the first to report heritabilities for CH_4 of dairy cows measured by GF units (g/day), and to report genetic correlations with CH_4 measured by sniffers (ppm). The in this study estimated heritabilities for CH_4 measured by GF or by sniffers were similar (0.19 vs 0.18 and 0.37 vs 0.36, for CH_4 as a mean per day or a mean per week respectively) (Table 1). The heritability and repeatability estimates for CH_4 measured by sniffers were in agreement with previous estimates that have been reported in the literature, where heritability estimates ranged between 0.11±0.02 and 0.26±0.11 (Lassen and Difford, 2020), and repeatability estimates ranged between 0.33 and 0.84 for different trait definitions and sniffer methods (Difford *et al.*, 2016). Similarly, the repeatability estimates for CH_4 measured by GF per day or per week (0.35 and 0.64, respectively) were in agreement with previous estimates for daily CH_4 production (0.47±0.12 (Sorg *et al.*, 2018)), and weekly CH_4 production (0.60 (Coppa *et al.*, 2021)).

The genetic correlation between weekly mean sniffer CH_4 and weekly mean GF CH_4 was high (0.77±0.14). The high genetic correlation shows that the two devices ranked the estimated breeding values of cows similarly from low to high emitting, and that therefore there is promise in using both devices in future genetic evaluations. In practice, the cheaper sniffers can collect CH_4 concentration records on a large number of cows, and the GF can be used to collect records on CH_4 production in gram/day when required for the breeding goal.

Table 1. Heritability (h^2) and repeatability (t) of methane emissions measured by the GreenFeed (g/day) or sniffers (ppm), as a mean per day or per week.

Trait		h^2	t
GreenFeed (g/day)	Day	0.19±0.02	0.35±0.01
	Week	0.37±0.04	0.64±0.01
Sniffer (ppm)	Day	0.18±0.01	0.34±0.01
	Week	0.36±0.02	0.61±0.01

Table 2. Phenotypic (r_p) and genetic (r_a) correlations for methane emissions measured by the GreenFeed (g/day) or sniffers (ppm), as a mean per day or per week.

	r_p	r_a
Day	0.41±0.03	0.53±0.16
Week	0.35±0.07	0.77±0.14

In conclusion, CH_4 by sniffers (ppm) and by GF units (g/day) are both moderately heritable and heritabilities and repeatabilities for the two devices are similar. The high genetic correlation suggests that measurements by sniffers and GF units can be combined in genetic evaluations, by increasing the number of cows with records on CH_4 emissions. However, before enteric CH_4 can be added to genetic evaluations, accurate genetic correlations with other breeding goal traits have to be estimated to ensure favorable changes in CH_4 emission without influencing production, feed efficiency, conformation, fertility, health or behavioral traits.

Funding

This study was part of the Climate Approach project funded by the Dutch ministry for agriculture, nature and food quality (LNV) (project number BO-53-003) and was supported by the TKI Agri & Food project LWV19155, and the partners CRV and FrieslandCampina.

References

Aguilar, I., Misztal, I., Johnson, D. L., Legarra, A., Tsuruta, S., *et al.* (2010). J Dairy Sci, 93(2), 743-752. https://doi.org/10.3168/jds.2009-2730

Calus, M. P. L., & Vandenplas, J. Calc_grm – a program to compute pedigree, genomic, and combined relationship matrices.

Christensen, O. F., & Lund, M. S. (2010). Genet Sel Evol, 42(1), 2. https://doi.org/10.1186/1297-9686-42-2

Coppa, M., Jurquet, J., Eugène, M., Dechaux, T., Rochette, Y., *et al.* (2021). Methods, 186, 59-67. https://doi.org/10.1016/j.ymeth.2020.11.004

Difford, G. F., Lassen, J., & Løvendahl, P. (2016). Comput Electron Agric, 124, 220-226. https://doi.org/10.1016/j.compag.2016.04.010

Gilmour, A., Gogel, B., Cullis, B., Welham, S., & Thompson, R. ASReml User Guide Release 4.1 Structural Specification. Available at https://asreml.kb.vsni.co.uk/wp-content/uploads/sites/3/2018/02/ASReml-4.1-Functional-Specification.pdf

de Haas, Y., Pszczola, M., Soyeurt, H., Wall, E., & Lassen, J. (2017). J Dairy Sci, 100(2), 855-870. https://doi.org/10.3168/jds.2016-11246

Hammond, K. J., Waghorn, G. C., & Hegarty, R. S. (2016). Anim Prod Sci, 56(3). https://doi.org/10.1071/an15631

Huhtanen, P., Cabezas-Garcia, E. H., Utsumi, S., & Zimmerman, S. (2015). J Dairy Sci, 98(5), 3394-3409. https://doi.org/10.3168/jds.2014-9118

Lassen, J., & Difford, G. F. (2020). Animal, 14(S3), s473-s483. https://doi.org/10.1017/S1751731120001561

Madsen, J., Bjerg, B. S., Hvelplund, T., Weisbjerg, M. R., & Lund, P. (2010). Livest Sci, 129(1-3), 223-227. https://doi.org/10.1016/j.livsci.2010.01.001

Sorg, D., Difford, G. F., Mühlbach, S., Kuhla, B., Swalve, H. H., *et al.* (2018). Comput Electron Agric, 153, 285-294. https://doi.org/10.1016/j.compag.2018.08.024

Team, R. C. R: A language and environment for statistical computing. Available at https://www.R-project.org/

Velazco, J. I., Herd, R. M., Cottle, D. J., & Hegarty, R. S. (2017). Anim Prod Sci, 57(4). https://doi.org/10.1071/an15111

Wellmann, R. optiSel: Optimum Contribution Selection and Population Genetics (Version 2.0.3.). Available at https://CRAN.R-project.org/package=optiSel

34. Methane emission has low genetic correlations to lamb growth traits in Norwegian White sheep

J.H. Jakobsen[1*], T. Blichfeldt[1], L.-B. Linneflaatten[1], M.O. Gløersen[1], L.E. Wallin[1] and J.C. McEwan[2]

[1]The Norwegian Association of Sheep and Goat Breeders, Box 104, 1431 Ås, Norway; [2]AgResearch Limited, Invermay Agricultural Centre, PB 50034, Mosgiel 9053, New Zealand; jj@nsg.no

Abstract

Selection for reduced methane (CH_4) emission is a mitigation option. Inclusion of liveweight adjusted CH_4 (CH_{4adj}) emissions as a trait in the breeding goal is expected to result in a cumulative and permanent reduction in sheep farms greenhouse gas emission. Fifty-minute CH_4 emissions were measured in portable accumulation chambers on 4,534 Norwegian White ewes from 44 breeding flocks and expressed as grams per hour. Ten years of observations for the lamb growth traits 42-day and 140-day weights were extracted from the database along with the corresponding pedigrees. Heritabilities and direct and maternal genetic correlations were estimated for CH_{4adj} and the lamb growth traits. The heritability for CH_{4adj} was 0.17, and heritabilities for the growth traits were in the range 0.07 to 0.13. The genetic correlation between CH_{4adj} and maternal genetic effect of 42-day weight was 0.32 and significantly different from zero, while other correlations between CH_{4adj} and growth were non-significant.

Introduction

In 2019, the Norwegian agriculture sector made an agreement with the Norwegian government to reduce its cumulated greenhouse gas (GHG) emissions over the 2020 to 2030 period by 11% from 45-million-ton to 40-million-ton (Regjeringen.no, 2021). The main source of GHG emission in lamb meat production comes from enteric methane (CH_4) emission produced as a by-product in the non-aerobic fermentation of feed. Other countries have shown that enteric CH_4 emission is heritable in sheep (Jonker *et al.*, 2018; Pinares-Patiño *et al.*, 2013) and in cattle (e.g. Lassen and Løvendahl, 2016; Pszczola *et al.*, 2017). Breeding is thus a mitigation option (Pickering *et al.*, 2015) and including CH_4 emission as a trait in the breeding goal will result in a cumulative and permanent reduction in the national GHG emission from lamb meat production. The aim of the current study was to investigate the heritability of CH_4 in Norwegian White sheep and compute the genetic correlations to direct and maternal genetic effect of 42-day and 140-day weights.

Materials & methods

Selection of sheep and flocks. Methane measurements in portable accumulation chambers (PAC) were collected on 4,534 Norwegian White sheep ewes that have lambed at least once or were expecting their first lambing. These sheep belong to forty-four (44) breeding flocks from 12 ram circles (Gjerdrem, 1969). Flocks from ram circles were chosen as they circulate rams for natural mating within the ram circle and in addition are required to use artificial insemination. The use of common rams thus results in genetic connectedness between flocks (Kuehn *et al.*, 2008).

Methane emission measurements. CH_4 emission (ppm) was measured in lots of 10 animals at a time. A hand-held Eagle2 instrument was used to capture accumulated 50 min CH_4 emissions following a measurement protocol developed in New Zealand (Jonker *et al.,* 2020). Sheep were either fed fresh grass or grass silage and were required to be off feed for at least one and less than four hours prior to entering the chamber and were in addition weighed prior to measurement. Fifty-minute CH_4 concentration was converted to CH_4 g/hr. Both CH_4 g/hr and ewe weights were scaled, they were divided by the mean of the lot and multiplied by the mean of all observations.

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_34, © J.H. Jakobsen *et al.* 2022

Growth traits in routine genetic evaluation. It is compulsory for the ram circle flocks to weigh lambs at birth and at weaning, while weighing during the main suckling period is voluntary. Weights for the main suckling period are recorded between 15 and 70 days of age and adjusted to a 42-day weight (42-d_{adj}), while weaning weights are recorded between 90 and 180 days of age and adjusted to 140 days (140-d_{adj}). Both 42-d_{adj} and 140-d_{adj} are expressed as weight gain from birth, and 42-d_{adj} thus represents a part of the growth period to weaning. The growth traits 42-d_{adj} and 140-d_{adj} are a part of the portfolio included for the routine genetic evaluations in Norwegian sheep.

Ten years of phenotypes of 42-d_{adj} and 140-d_{adj} were extracted for the 44 ram circle flocks with CH_4 emission measurements. Mean, standard deviation and number of observations of methane emission and weights are listed in Table 1.

Pedigree. Pedigree was extracted from the national sheep recording database and traced as far back as possible on animals with phenotypes. The pedigree included a total of 154,683 animals.

Genetic evaluation model. Data was analysed in a tri-variate animal model using the DMU-software (Madsen and Jensen, 2013).

The model for methane emission was:

$$Y_{ijk} = FL_i + A_j + b_1 \times W_{ijk} + a_{ijk} + e_{ijk} \quad (1)$$

Where Y_{ijk} is scaled CH_4 emission measured on animal k in Flock×Lot i of Age class j. W_{ijk} is scaled weight at measurement, b_1 is regression coefficient on Weight, a_{ijk} is random additive genetic effect of animal k, and e_{ijk} is random residual corresponding to Y_{ijk}.

The model for 42-d and 140-d weight was:

$$Y_{ijkl} = FY_i + ABKW_j + a_{ijk} + m_{ijl} + pe_{ijl} + e_{ijkl} \quad (2)$$

Where Y_{ijkl} is weight observation on lamb k in Flock×Year i and in rearing type $ABKW_j$, a_{ijk} is random additive genetic effect of lamb k, m_{ijl} is random additive genetic effect of the fostering dam l, pe_{ijl} is permanent environment effect of the fostering dam, and e_{ijkl} is random residual corresponding to Y_{ijkl}.

Results

Genetic standard deviations, heritabilities and genetic correlations for CH_{4adj}, 42-d_{adj} and 140-d_{adj} weights are shown in Table 2 along with standard errors. The heritability estimate for liveweight adjusted PAC CH_4 emission is similar to previously published sheep estimates (Goopy *et al.*, 2015; Dominik *et al.*, 2017) and

Table 1. Mean (sd) and number of observations for CH_{4adj}, live weight at measurement ($Weight_{adj}$), 42-day weight (42-d_{adj}), and 140-day weight (140-d_{adj}), above the diagonal is the number of observations with both traits.

	Mean (sd)	CH_{4adj}, g/hr	$Weight_{adj}$, kg	42-d_{adj}	140-d_{adj}
CH_{4adj}, g/hr	1.37 (0.28)	4,534	4,534	3,986	4,253
$Weight_{adj}$, kg	79.70 (9.88)		4,534	3,986	4,253
42-d_{adj}	14.34 (3.31)			91,873	84,496
140-d_{adj}	40.01 (7.69)				93,669

also similar to estimates where either feed intake or its proxies (liveweight, CO_2 emission, O_2 consumption) were used either as covariates or expressed in the denominator of a ratio of methane emission (Pinares-Patiño *et al.*, 2013; Jonker *et al.*, 2018). The only significant correlation between CH_{4adj} and growth traits was the maternal genetic effect of 42-day weight.

Genetic trends for CH_{4adj} and maternal genetic effect of 42-day weight are shown in Figure 1A along with number of ewes with CH_4 observations. Both traits follow the same trend until 2018, where the curve seems to flatten for CH_{4adj}. Most CH_4 observations are from ewes born 2017 through 2020. This may be contributing to the trend change, which in earlier years is largely driven by the maternal genetic correlation with $CH_{4adj.}$

Phenotypic trends for average carcass weight and number of days to slaughter for all 763 ram circle flocks are shown in Figure 1B for the same time period. The figure shows a decreasing trend in number of days to slaughter of one day per year and a simultaneously slight increase in average carcass weight. The increasing trend in methane emission is thus related to an increase in lamb growth during the main suckling period where the rumen only is partly functioning.

Discussion

Based on this study, including liveweight adjusted PAC CH_4 emission into the breeding goal for Norwegian White sheep imply a reduction in the CH_4 conversion factor Ym (% of gross energy intake partitioned into CH_4) Niu *et al.* (2021) and as such also the national emissions, given no additional feed stuffs are grown or imported.

Table 2. Genetic standard deviation (σ_a), heritabilities (in diagonal) and genetic correlations for CH_{4adj}, and direct (dir.) and maternal (mat.) genetic effect of 42-d_{adj} and 140-d_{adj} weights.

	σ_a	CH_{4adj}, g/hr	42-d_{adj} (dir.)	42-d_{adj} (mat.)	140-d_{adj} (dir.)	140-d_{adj} (mat.)
CH_{4adj}, g/hr	0.112	*0.17 (0.04)*				
42-d_{adj} (dir.)	0.897	-0.07 (0.11)	*0.11 (0.01)*			
42-d_{adj} (mat.)	0.776	0.32 (0.09)	-0.19 (0.05)	*0.08 (0.01)*		
140-d_{adj} (dir.)	2.291	0.19 (0.10)	0.71 (0.03)	0.05 (0.05)	*0.13 (0.01)*	
140-d_{adj} (mat.)	1.703	0.11 (0.10)	-0.18 (0.06)	0.82 (0.03)	0.00 (0.06)	*0.07 (0.01)*

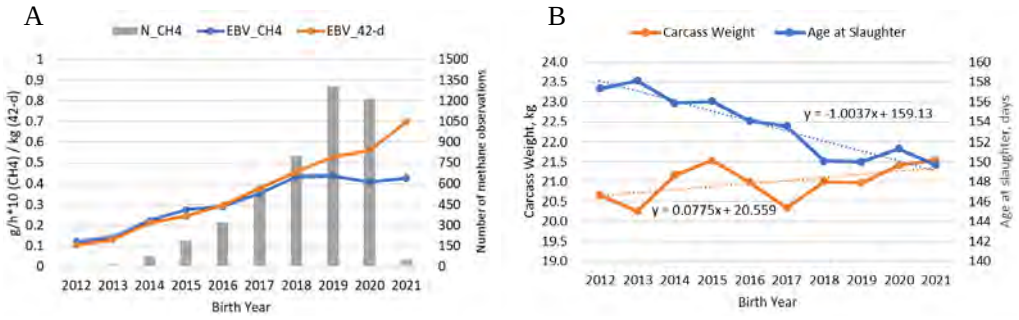


Figure 1. (A) Genetic trends for methane emission (g/hr ×10), maternal genetic effect of 42-day weight, and number of animals with methane measurements by age class. (B) Phenotypic trends for carcass weight and age at slaughter.

However, based on the correlations in Table 2, selection solely for reduced methane emission would result in a decrease in the maternal effect on lamb growth during main suckling period. This antagonistic relation should therefore be handled carefully when including methane emission in the breeding goal via appropriate weighting in the overall selection index.

Acknowledgements

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 772787 (SMARTER). The PAC chambers were designed and built by the Engineering Team at AgResearch New Zealand.

References

Dominik, S., Robinson, D.L., Donaldson, A.J., Cameron, M., Austin, K.L *et al.* (2017) Proc. of the 22nd AAABG, Townsville, Australia.

Gjedrem, T. (1969). Acta Agric. Scand. 19, 116–126. https://doi.org/10.1080/00015126909433399

Goopy, J.P., Robinson, D.L., Woodgate, R.T., Donaldson, A.J., Oddy, V.H., *et al.* (2015). Anim. Prod. Sci. 56, 116–122. https://doi.org/10.1071/AN18383

Jonker, A., Hickey, S.M., McEwan, J.C., and Waghorn G. (2020). Guidelines for estimating methane emissions from individual ruminants using: GreenFeed, 'sniffers', hand-held laser detector and portable accumulation chambers. Ministry for Primary Industries, Wellington, New Zealand. https://www.mpi.govt.nz/resources-and-forms/publications/

Jonker, A., Hickey, S.M., Rowe, S.J., Janssen, P.H., Shackell, G.H., *et al.* (2018). J. Anim. Sci. 96(8): 3031-3042. https://doi.org/10.1093/jas/sky187

Klimaplan for 2021–2030. (2021). Available at: https://www.regjeringen.no/no/dokumenter/meld.-st.-13-20202021/id2827405/

Kuehn, L.A., Notter, D.R., Nieuwhof, G.J., and Lewis, R.M. (2008) J. Anim. Sci. 86, 536–544. https://doi.org/10.2527/jas.2007-0256

Lassen J. and Løvendahl P. (2016). J. Dairy Sci 99, 1959–1967. https://doi.org/10.3168/jds.2015-10012

Madsen, P., and Jensen, J. (2013). A User's Guide to DMU. A Package for Analysing Multivariate Mixed Models. Version 6, release 5.2. Aarhus University, Aarhus, Denmark. Available at: https://dmu.ghpc.au.dk/dmu/DMU/Doc/Current/dmuv6_guide.5.2.pdf

Niu P., Schwarm A., Bonesmo H., Kidane A., Åby B.A., *et al.* (2021) Animals 11, 1891 https://doi.org/10.3390/ani11071891

Pickering N.K., Oddy V.H., Basarab J., Cammack K., Hayes B., *et al.* (2015). Animal 9, 1431–1440. https://doi.org/10.1017/S1751731115000968

Pinares-Patiño C.S., Hickey, S.M., Young, E.A., Dodds, K.G., MacLean, S, *et al.* (2013). Animal 7(Suppl 2), 316–321. https://doi.org/10.1017/S1751731113000864

Pszczola M., Rzewuska K., Mucha S., and Strabel T. (2017) J. Anim Sci 95, 4813–4819. https://doi.org/10.2527/jas2017.1842

35. Life cycle assessment to predict individual environmental impacts: towards selection for sustainable pig production

T. Soleimani[1], S. Hermesch[2] and H. Gilbert[1*]

[1]GenPhySE, Université de Toulouse, INRAE, ENVT, 31320 Castanet-Tolosan, France; [2]AGBU, a joint venture of NSW Department of Primary Industries and University of New England, 2351, Armidale, Australia; helene.gilbert@inrae.fr

Abstract

A life cycle assessment approach is proposed to quantify four environmental impacts at the individual pig level. The approach was applied to two lines divergently selected for feed efficiency, including simulations of performances with two diets. Significant line differences were obtained with least-cost diets for environmental impacts and profit ($P<0.001$), with a general advantage of the more efficient line. The optimised diets reduced the disadvantages of the less efficient line. Correlations between profit and environmental impact were high for feed efficiency (>0.85), and moderate (<0.45) for growth rate and body composition. Applied to pedigreed populations, this framework will provide genetic parameters of environmental impacts to improve the sustainability of livestock production.

Introduction

The Western pig production systems affect global warming potential (GWP), terrestrial acidification potential (AP), freshwater eutrophication potential (EP), and land occupation (LO) (McAuliffe *et al.*, 2016). However, early attempts to quantify environmental impacts (EI) in pigs essentially focused on nitrogen (N) and phosphorus excretion estimated at the animal level (Shirali *et al.*, 2012; Saintilan *et al.*, 2013). We propose a broader approach to predict the four EI at the individual level via a life cycle assessment model (LCA). It was applied to two pig lines with simulations of two nutritional strategies. Finally, a bio-economic model was used to evaluate the correlations between environmental impacts and profit.

Materials & methods

LCA model. A LCA model of typical pig farms in France was developed with a cradle-to-farm-gate system boundary (Soleimani and Gilbert, 2020) to evaluate GWP (kg CO_2 eq), AP (kg SO_2 eq), EP (kg P eq), and LO (m^2 a crop eq) from individual performance traits (Soleimani and Gilbert, 2021). The functional unit was 1 kg body weight of a pig at the gate. The ReCiPe Midpoint 2016 (H) V1.02 was used, together with the Ecoalim dataset of the AGRIBALYSE® database and the Ecoinvent inventory databases in SimaPro V8.5.4.0 on the MEANS platform, to obtain the impacts of diets and individual pig performance profiles.

Bio-economic model. The bio-economic model was developed from a linear profit model. The model mimicked the LCA model modules and ran from the same individual performances. It included the main costs of feed and water, labour, building and capital, energy, manure disposal, health, maintenance and repair (Soleimani *et al.*, 2021). Prices of diet ingredients (May 2020), energy consumption per kg pig and water to feed ratios were based on French averages. Revenue was from sales of live pigs, computed according to the French price grid from the baseline of a 100 kg carcass with lean meet percentage (LMP) of 56%, corrected for actual carcass weight and composition (July 2020).

Data. Experimental data were collected from the fifth generation of Large White pig lines divergently selected for residual feed intake (RFI, Gilbert *et al.*, 2017)). Fifty seven male pigs from each line (LRFI, more efficient pigs; HRFI, less efficient pigs) had *ad libitum* access to a one phase conventional diet during

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_35, © T. Soleimani *et al.* 2022

growth-finishing. Records for daily feed intake, average daily gain (ADG), feed conversion ratio (FCR) and back fat thickness (BFT) were available.

Diet formulations and performance simulations. Diet formulation was based on wheat, barley, corn, oats, peas, triticale, rapeseed meal, sunflower meal, soybean meal, sunflower oil, and synthetic L_lysine, L_threonine, L_tryptophan and DL_methionine. Compositions were obtained from the INRA-AFZ feed ingredients database (Sauvant *et al.*, 2004).

Individual nutritional requirements profiles were computed from individual performance imported to InraPorc® (Brossard *et al.*, 2014), using a Gamma function. Two diets were formulated within line, covering the average nutritional requirements of the line, to either minimize cost (least-cost diet) or minimize a combination of cost and environmental impacts (optimised diet). This cost and environment multi-objective was a linear function of the sum of the four environmental impacts of the ingredients, considering equal weights for the four EI (EI score), and the sum of the ingredients costs. The relative weights of the environmental and economic objectives was optimized for each line (Soleimani and Gilbert, 2021), using a Pareto-optimal front curve. Next, the individual performances of pigs offered one or the other diet formulated for their line were simulated until slaughter weight (120 kg live weight) with InraPorc®. The resulting fattening traits for each individual were used as input parameters to the LCA and bio-economic models, to compute individual environmental impacts and profit.

Statistical analyses. Analyses were computed in R (R Core Team, 2019). The line differences for performances, environmental impacts and profit were tested with linear models including the fixed effect of the line (lm procedure). To illustrate correlations between variables, principal component analyses (PCA, PCA procedure) were applied to performances, environmental impacts and profit for the full dataset before and after adjusting for line effects.

Results

Diets. Pigs from the LRFI line had higher average requirements than HRFI pigs ($P<0.05$). For instance, digestible lysine was 0.91±0.20 g/MJ net energy in LRFI and 0.86±0.18 in HRFI. As a result, diet compositions differed between lines and formulation objectives: the LRFI least-cost diet contained essentially triticale (55%) and barley (26%), the HRFI least-cost diet contained corn (50%), triticale (17%) and barley (15%), whereas the optimised diets were more diverse (LRFI: corn (38%), triticale (22%), barley (12%); HRFI: barley (36%), corn (17%), triticale (16%), wheat (11%)). Then, as compared to the least-cost diet, the optimized diets had lower EI scores (Table 1), but not all the environmental impacts were reduced. Their prices per MJ of net energy were 1.5 (HRFI) to 4.5% (LRFI) higher.

Table 1. Environmental impacts, environmental score (EI score) and price per unit of net energy of the least-cost (LC) and optimised diets for the low (LRFI) and high residual feed intake (HRFI) lines.

Diet		GWP	AP	EP	LO	EI score	Price
		kg CO_2 eq	g SO_2 eq	g P eq	m^2 a crop eq		0.01€
LRFI	Least-cost	0.541	0.613	0.0526	0.181	0.430	2.01
	Optimised	0.486	0.663	0.0505	0.152	0.394	2.10
HRFI	Least-cost	0.483	0.683	0.0599	0.141	0.399	2.03
	Optimized	0.490	0.643	0.0496	0.163	0.395	2.06

[1] GWP = global warming potential; AP = acidification potential; EP = eutrophication potential; LO = land occupation.

Environmental impacts and profit. As expected, the simulated performances with the two diets were very similar within line (Table 2), and corresponded to known differences between the lines (Gilbert *et al.*, 2017). The environmental impacts of producing 1 kg of body weight with the least-cost diets were lower in the LRFI line than in the HRFI line, by 3.5% for GWP to 12.2% for EP ($P<0.001$), except for LO (-8.7%), as expected from other studies (Monteiro *et al.*, 2021). With optimised diets, these impacts decreased for both lines by 2.2 to 10.6% compared to the least-cost diets, except for AP for LRFI and LO for HRFI. Line differences reduced in simulations with optimised diets, from 2.0% for AP ($P<0.05$) to 8.5% for LO ($P<0.001$)), but LRFI pigs still had reduced impacts compared to HRFI pigs.

Table 2. Mean (standard deviation) and *P*-values of line differences (*P*) for simulated production traits, environmental impacts for 1 kg live weight at the gate, and profit for low and high residual feed intake lines (LRFI, HRFI) fed least-cost or optimised diets.

	Least-cost diets		P^2	Optimised diets		P^2
Trait[1]	LRFI	HRFI		LRFI	HRFI	
Production						
ADG (kg/d)	0.77 (0.09)	0.80 (0.07)	*	0.78 (0.09)	0.82 (0.07)	*
FCR (kg /kg)	2.68 (0.17)	2.58 (0.17)	**	2.53 (0.18)	2.64 (0.19)	**
BFT (mm)	16.4 (1.1)	17.6 (1.0)	***	16.3 (1.1)	17.4 (1.0)	***
Environment						
GWP, kg CO_2 eq	2.02 (0.10)	2.09 (0.10)	***	1.96 (0.10)	2.02 (0.10)	***
AP, g SO_2 eq	33.1 (1.99)	37.1 (2.22)	***	34.6 (2.26)	35.3 (2.23)	***
EP, g P eq	1.39 (0.08)	1.56 (0.09)	***	1.36 (0.08)	1.40 (0.09)	*
LO, m^2 a crop eq	4.35 (0.25)	3.97 (0.22)	***	3.89 (0.23)	4.22 (0.25)	***
Profit, €	17.75 (5.56)	14.47 (7.01)	**	16.86 (5.68)	15.58 (5.64)	ns

[1] ADG = average daily gain; FCR = feed conversion ratio; BFT = back fat thickness; GWP = global warming potential; AP = acidification potential; EP = eutrophication potential; LO = land occupation.

[2] *** $P<0.001$; ** $P<0.01$; * $P<0.05$; ns = non-significant.

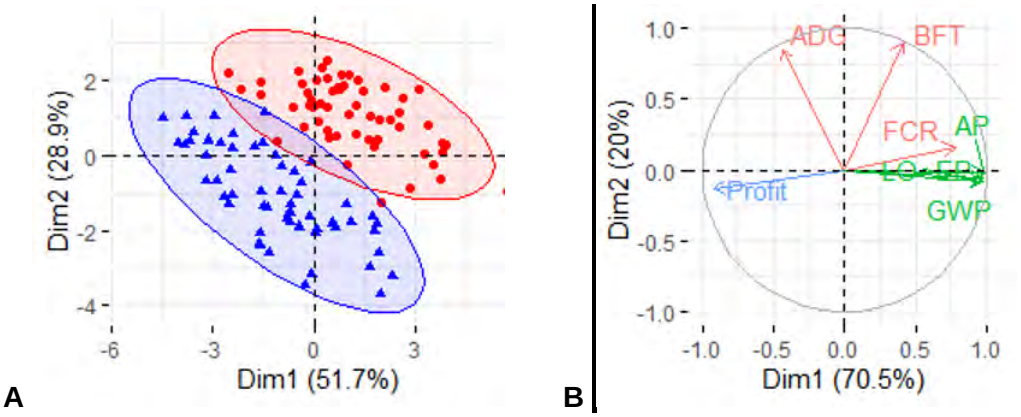


Figure 1. Principal components analyses of the least-cost diets dataset non-adjusted (A, red circles = HRFI pigs, blue triangles = LRFI pigs) and adjusted (B) for line effect.

With the least-cost diet, the LRFI pigs provided on average 3.28€ additional profit per pig, compared to the HRFI pigs ($P<0.01$). With performances and costs simulated with the optimized diets, the profit slightly decreased for the LRFI line (-0.89€/pig), whereas it was increased by 1.11 €/pig in the HRFI line, leading to a reduced and non-significant profit difference between lines with the optimised diets.

The correlations between variables did not differ between diets and Figure 1 only shows results for the least-cost diets. The PCA applied to non-adjusted variables mainly separated lines (Figure 1A.). After adjusting for line effects (Figure 1B.), environmental impacts had strong positive correlations with each other, and high negative correlations with profit (-0.84 to -0.86). FCR had a correlation of -0.76 with profit and ~0.65 with environmental impacts, whereas ADG and BFT had correlations <0.46 in absolute value with these variables.

Discussion

The LCA framework developed considers individual performances. This enables quantifying the main environmental impacts of individual pigs. In divergent lines, we showed how selection and diet formulation strategies result in different environmental impacts, but also that including environmental objectives in diet formulations could reduce the disadvantages of the less efficient line by using different feed ingredients. The proposed LCA framework, applied to large populations, will provide genetic parameters of environmental impacts to be used in breeding objectives, to contribute to the sustainability of livestock production.

Acknowledgements

This study is part of a collaborative project supported by the ERA-NET SUSAN programme via the ANR-16-SUSN-0005 contract (SusPig).

References

Brossard L., Vautier B., van Milgen J., Salaun Y., and Quiniou N. (2014) Anim. Prod. Sci. 54(12):1939-1945. https://doi.org/10.1071/AN14521

Gilbert H., Billon Y., Brossard L., Faure J., Gatellier P., *et al.* (2017) Animal 11(9):1427–1439. https://doi.org/10.1017/S175173111600286X

McAuliffe G.A., Chapman D.V., and Sage C.L. (2016) Environ. Impact Assess. Rev. 56:12–22. https://doi.org/10.1016/j.eiar.2015.08.008

Monteiro A.N.T.R, Brossard B., Gilbert H., and Dourmad J.Y. (2021) Front. Vet. Sci. 8:677857. https//doi.org/10.3389/fvets.2021.677857

R Core Team (2019). Vienna, Austria. URL https://www.R-project.org/

Saintilan R., Mérour I., Brossard L., Tribout T., Dourmad J.Y., *et al.* (2013) J. Anim. Sci. 91:2542–2554. https://doi.org/10.2527/jas.2012-5687

Sauvant D., Perez J.-M., and Tran G. (2004) Tables of composition and nutritional value of feed materials. Wageningen Academic Publishers, Wageningen, the Netherlands.

Shirali M., Doeschl-Wilson A., Knap P.W., Duthie C., Kanis E., *et al.* (2012) J. Anim. Sci. 90:1756–1765. https://doi.org/10.2527/jas.2011-4547

Soleimani T., and Gilbert H. (2020) Animal 14:2598–2608. https://doi.org/10.1017/S175173112000138X

Soleimani T., and Gilbert H. (2021) Int. J. Life Cycle Assess. https://doi.org/10.1007/s11367-020-01860-3

Soleimani T., Hermesch S., and Gilbert H. (2021) J. Anim. Sci. 99(3):1-14. https://doi.org/10.1093/jas/skab051

36. Genetic parameters for growth, feed intake and greenhouse gasses emissions in Italian Holstein-Friesian bulls

F. Tiezzi[1*], R. Finocchiaro[2], L. Benzoni[2], J.B.C.H.M. van Kaam[2], M. Marusi[2], R. Bozzi[1] and M. Cassandro[2,3]

[1]Department of Agriculture, Food, Environment and Forestry (DAGRI), University of Florence, Piazzale delle Cascine 18, 50144 Firenze (FI), Italy; [2]Associazione Nazionale Allevatori della Razza Frisona, Bruna e Jersey Italiana (ANAFIBJ), Via Bergamo 192, 26100 Cremona (CR), Italy; [3]Department of Agronomy, Food, Natural resources, Animals and Environment (DAFNAE), University of Padova, Viale dell'Università 16, 35020 Legnaro (PD), Italy; francesco.tiezzi2@unifi.it

Abstract

The objective of this study was to estimate genetic parameters for growth, feed intake and greenhouse gasses emission traits in Italian Holstein, using data collected on young bulls undergoing performance test. The phenotypic information was recorded on 189 bulls, for which genomic information was obtained using SNP chips. Data were analysed using a linear mixed model, which included an additive genetic effect using a genomic relationship matrix. Results show high heritability estimates for growth traits, moderate to high estimates for emission and feed intake traits. Selection for reduced emissions appears possible and reducing methane will also reduce carbon dioxide emissions (genetic correlation of 0.96). Feed intake traits could serve as indicators, because the genetic correlations range between 0.58 and 0.85. Growth traits showed strong positive correlations with emission traits (0.66 to 0.81), selection indices built at reducing emissions should carefully consider these traits as well.

Introduction

Dairy cattle is known to be impactful on greenhouse gasses (GHG) emissions, with its enteric emissions accounting for over ten percent of the emissions from the livestock sector globally (Gerber *et al.*, 2013). Methane emissions are also energetically expensive because of the fermentation process (Appuhamy *et al.*, 2016) but this inefficiency could be reduced by steering the fermentation process, redirecting the energetic resources to reproduction and milk production (Haque, 2018). Methane and carbon dioxide emission have been shown to be heritable, providing the basis for applying genetic selection for their reduction (Pickering *et al.*, 2015; Lassen and Løvendahl, 2016). Such selection could be applied by selecting directly for breath measurements, but also using indirect selection including indicator traits such as feed intake (De Haas *et al.*, 2017).

The objective of this study was to estimate genetic parameters for verifying the feasibility of (direct or indirect) selection for reduced GHG emissions in Italian Holsteins.

Materials & methods

Animals and data. The animals involved in this study were young Italian Holstein bulls undergoing performance test at the Italian National Breeders Association for the Holstein, Brown Swiss and Jersey dairy cattle breeds (ANAFIBJ). Phenotypic data was made available on 189 Holstein bulls between the age of 171 and 541 days. All bulls were genotyped using various SNP chips resulting in 98,586 SNP. Genomic data was edited using the preGSf90 software (Aguilar *et al.*, 2010), removing SNP with call rate below 0.9 and minor allele frequency below 0.05. After editing, 89,625 SNP were available.

The first set of traits included measures of body growth taken using electronic scales and stadiometers operated by qualified personnel. These included body weight (WEI), body condition score (BCS), heart

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_36, © F. Tiezzi *et al.* 2022

girth (HG) and height (HEI). The second group of traits included measures of feed intake and feeding behaviour assessed using the Roughage Intake Control system (RIC; Hokofarm Group, Marknesse, the Netherlands). Records were organized as daily measures: number of visits at the feeder per day (NVF), average intake at the feeder (AIF), average time at the feeder (ATF). The third group of traits included measures of greenhouse gasses emission and were obtained using the GreenFeed (C-Lock Inc., Rapid City, SD, USA). Again, the single-visit measures were converted to daily records and included: number of visits (NVG), carbon-dioxide daily emission (CO_2), methane daily emission (CH_4), average airflow (AIR) and average time (ATG). A single dataset was created, containing all the phenotypic information. Individuals showed an average of 4 records for the growth traits, 36 records for the RIC-derived traits, 19 records for the GreenFeed-derived traits. For all records, the age at recording was calculated. In addition, the date of birth of the individual was transformed to numerical values as the difference, in days, from a fixed arbitrary date.

Statistical analysis. Estimates of variance components and genetic parameters were carried out using a linear mixed model that include a genomic relationship matrix constructed using the SNP markers. The model was defined as follows:

$$y = Xb + Z_d d + Z_p p + Z_a a + e \quad (1)$$

where y is the vector containing phenotypic records, X and b are the incidence matrix and vector of solutions for the fixed effects (age at phenotyping, date of birth), Z_d and d are the incidence matrix and vector of solutions for the 'date of recording' uncorrelated random effect, Z_p and p are the incidence matrix and vector of solutions for the animal permanent environmental uncorrelated random effect, Z_a and a are the incidence matrix and vector of solutions for the animal additive genetic random effect (with genomic relationship matrix), e is the vector of random residuals. The model was implemented for single-trait and two-trait analyses in order to obtain estimates of heritability and genetic correlations, using the gibbs2f90 software (Misztal *et al.*, 2002). Estimates of variance components and genetic parameters were obtained as posterior means and their dispersion was obtained as the posterior SD and the 95% confidence intervals. The heritability was expressed as the ratio between the additive genetic variance and the sum of the four variance components.

Results

Descriptive statistics and heritability estimates for the studied traits are reported in Table 1. The growth and condition traits showed the largest estimates, all being above 0.50. While these traits are expected to be highly heritable, the estimates appear larger compared to those found in literature. This could be due to the relatively small sample size. On the other hand, heritability estimates for the RIC-derived traits were lower, ranging from 0.151 (ATF) to NVF (0.299). Heritability estimates for the emission traits were moderate to high, ranging from 0.189 for ATG to 0.530 for CO_2.

Genetic correlation estimates are reported in Table 2. Among the growth traits, correlations were strong for WEI-BCS and BCS-HG, but weak for all the other combinations of traits. Correlations between NVF and AIF was basically null (0.25), the genetic correlation between AIF and ATF was 0.97 (result not shown in table). Correlations among the GreenFeed-derived traits were moderate to strong, especially among CH_4, CO_2 and AIR, which were all above 0.85. NVG was moderately related to the other GreenFeed-derived traits, with correlations below 0.7. The genetic correlation between AIR and ATG was 0.95 (result not shown in table). Genetic correlations between the growth traits and the RIC-derived traits were all positive and strong, approaching unity (0.93 to 0.96). Genetic correlations between the growth traits and the GreenFeed-derived traits were also positive but moderate, with values ranging from 0.66 to 0.86. Similarly, the genetic correlations between the GreenFeed-derived traits and the RIC-derived traits were positive and moderate to strong, with values from 0.58 to 0.90.

Table 1. Descriptive statistics and heritability estimates (posterior means with posterior standard deviation) for the traits analysed.

Trait[1]	Metric	N	Mean	SD	h^2
WEI	kg	792	308.9	75.26	0.604 (0.250)
BCS	score	757	3	0.34	0.539 (0.165)
HG	cm	622	157.2	13.61	0.647 (0.232)
HEI	cm	621	125.6	7.47	0.569 (0.242)
NVF	count	6,402	49.7	24.25	0.299 (0.141)
AIF	kg	6,382	0.3	0.13	0.178 (0.123)
ATF	s	6,392	262.4	161.08	0.151 (0.106)
NVG	count	2,520	4	1.75	0.293 (0.110)
CO_2	g/d	2,520	6,154.10	1,066.37	0.530 (0.218)
CH_4	g/d	2,520	225	50.81	0.346 (0.191)
AIR	L/s	2,520	29.2	4.17	0.521 (0.112)
ATG	s	2,507	331.9	81.4	0.189 (0.111)

[1] WEI: body weight; BCS: body condition score; HG: heart girth; HEI: height; NVF: number of visits at the feeder; AIF: average intake at the feeder; ATF: average time at the feeder; NVG: number of visits at the GreenFeed; CO_2 daily carbon dioxide emissions; CH_4: daily methane emissions; AIR: average airflow at the visit; ATG: average time at the GreenFeed.

Table 2. Estimates of genetic correlations[1] among the growth-related traits[2], the RIC-derived traits[2] and the GreenFeed-based traits[2].

	WEI	BCS	HG	HEI	NVF	AIF	NVG	CO_2	CH_4	AIR
WEI		**0.80**	0.70	0.39	**0.93**	**0.96**	**0.87**	**0.79**	**0.79**	**0.85**
BCS	**0.80**		**0.67**	0.45	**0.93**	**0.96**	**0.90**	**0.81**	**0.80**	**0.86**
HG	0.70	**0.67**		0.26	**0.93**	**0.94**	**0.82**	**0.67**	**0.66**	**0.76**
HEI	0.39	0.45	0.26		**0.93**	**0.94**	**0.84**	**0.72**	**0.71**	**0.79**
NVF	**0.93**	**0.93**	**0.93**	**0.93**		0.25	**0.90**	**0.85**	**0.85**	**0.87**
AIF	**0.96**	**0.96**	**0.94**	**0.94**	0.25		**0.70**	**0.59**	**0.58**	**0.64**
NVG	**0.87**	**0.90**	**0.82**	**0.84**	**0.90**	**0.70**		**0.69**	**0.68**	**0.89**
CO_2	**0.79**	**0.81**	**0.67**	**0.72**	**0.85**	**0.59**	**0.69**		**0.91**	**0.85**
CH_4	**0.79**	**0.80**	**0.66**	**0.71**	**0.85**	**0.58**	**0.68**	**0.91**		**0.90**
AIR	**0.85**	**0.86**	**0.76**	**0.79**	**0.87**	**0.64**	**0.89**	**0.85**	**0.90**	

[1] Estimates of the genetic correlations are the posterior mean for the parameter. Values in bold indicate that the value '0' was not included within the confidence intervals, therefore are to be considered significant for $P<0.05$.

[2] WEI: body weight; BCS: body condition score; HG: heart girth; HEI: height; NVF: number of visits at the feeder; AIF: average intake at the feeder; NVG: number of visits at the GreenFeed; CO_2 daily carbon dioxide emissions; CH_4: daily methane emissions; AIR: average airflow at the visit.

Discussion

The estimates show substantial genetic variation for all the studied traits. The CO_2 and CH_4 daily emissions show large heritability with the possibility of selection, therefore reduction of GHG emissions. The estimated values for the heritability of CH_4 and CO_2 are larger than values found in literature (Lassen and Løvendahl, 2016; Brieder *et al.*, 2018), although this could be due to the involvement of growing bulls rather than lactating cows and the limited sample size in the current study. The genetic correlations indicate that there is the possibility to select for less impactful animals. The genetic correlation between CO_2 and

CH_4 emissions is close to unity (0.91), suggesting that selection for one component of the emissions would improve the other as well. Similarly, the AIR shows strong correlations with the GHG emissions, suggesting that this trait could be used as an indicator. Similarly, the NVG could serve as an indicator trait to reduce the emissions, although the achievable genetic progress would be limited due to the moderate genetic correlations (0.7-0.8). In fact, the value of 0.7 was indicated as the minimum acceptable value for achieving a relevant genetic progress in case of multiple-trait genomic predictions by Calus and Veerkamp (2011). Using this same criterion, NVF could reduce GHG emissions better than feed intake (AIF) given the stronger genetic correlations (0.85 vs ~0.60). Unsurprisingly, selection for larger animals will also lead to individuals that consume more feed with all correlations between growth traits and RIC-derived traits being above 0.9. Similarly, selection for larger animals could lead to more GHG emissions, although this relationship seems to be weaker because genetic correlations between growth trait and GHG emissions tend to be equal or smaller than 0.80. In general, selection indices could be built in order to reduce GHG emissions without compromising growth, condition, stature and feed intake. The upcoming research should involve the use of feed efficiency and relative GHG emission, as these components need to be adjusted by the growth and size of the animal.

References

Aguilar I., Misztal I., Johnson D.L., Legarra A., Tsuruta S., *et al.* (2010). J. Dairy Sci. 93:2 743–752. https://doi.org/10.3168/jds.2009-2730.

Appuhamy J.Adescriptive., France J. and Kebreab E. (2016). Glob. Chang. Biol. 22:9 3039–3056. https://doi.org/10.1111/gcb.13339.

Breider I.S., Wall E., and Garnsworthy P.C. (2019). J. Dairy Sci. 102:8 7277–7281. https://doi.org/10.3168/jds.2018-15909.

Calus M.P.L. and Veerkamp R.F. (2011). Genet. Sel. Evol. 43:26. https://doi.org/10.1186/1297-9686-43-26.

De Haas Y., Pszczola M., Soyeurt H., Wall E., and Lassen J. (2017). J. Dairy Sci. 100:2 855–870. https://doi.org/10.3168/jds.2016-11246.

Gerber P.J., Hristov, A.N., Henderson B., Makkar H., Oh J., *et al.* (2013). Animal 7:2 220–234. https://doi.org/10.1017/S1751731113000876.

Haque M.N. (2018). J. Anim. Sci. Tech. 60:15. https://doi.org/10.1186/s40781-018-0175-7.

Lassen J. and Løvendahl P. (2016). J. Dairy Sci. 99:3 1959–1967. https://doi.org/10.3168/jds.2015-10012.

Misztal I., Tsuruta S., Strabel T., Auvray B., Druet T., *et al.* (2002). Proc. of the 7th WCGALP. Montpellier, France.

Pickering N.K., Chagunda M.G.G, Banos G., Mrode R., McEwan J.C. *et al.* (2015). J. Anim. Sci. 93:1 11–20. https://doi.org/10.2527/jas.2014-8302.

37. Association of genetic resistance to internal nematodes and production traits on feed efficiency and methane emissions in Corriedale lambs

E.A. Navajas[1], G. Ciappesoni[1*], D. Gimeno[2], J.I. Velazco[1] and I. De Barbieri[1]

[1]Instituto Nacional de Investigación Agropecuaria, INIA Las Brujas, 90100, Rincón del Colorado, Canelones, Uruguay; [2]Secretariado Uruguayo de la Lana, Servando Gomez 2408, 11200, Montevideo, Uruguay; gciappesoni@inia.org.uy

Abstract

Potential trade-off among weaning (WWT) and yearling weights (YWT), greasy fleece weight (GWT) genetic resistance to gastrointestinal nematodes assessed by faecal worm egg account (FEC), residual feed intake (RFI) and methane (CH_4) and carbon dioxide (CO_2) emissions were investigated in 217 Corriedale female lambs. Improving RFI was strongly associated with lower dry matter intake (DMI) and independent of metabolic body weight (MWT) and average daily gain (ADG). Person correlation coefficients were also no significant (P>0.05) with WWT, YWT, GFW and FEC expected progeny differences (EPD). FEC EPD was not correlated with any of the traits investigated here. Both CH_4 and CO_2 were not associated with RFI, although the moderate correlations with DMI. Higher EPD for WWT, YWT and GFW were correlated with higher DMI, CO_2 and CH_4. Optimising higher productivity and environmental sustainability requires further research for a more comprehensive understanding of these associations.

Introduction

Sustainable sheep production requires considering not only traditional production traits but also those associated with resilience and greenhouse gas (GHG) mitigation. Genetic evaluation of Corriedale, as dual-purpose breed, includes body weight and wool production, which are relevant for economic income of commercial production. Improving genetic resistance to gastrointestinal nematodes (GIN), using faecal worm egg count (FEC) as selection criteria, is also included in the breeding programme (Ciappesoni *et al.*, 2014). Given the important losses in pastoral grazing extensive systems due to infestation by GIN, and the difficulties to control it only by grazing management and chemical treatments, genetic selection by FEC may also contribute to economic and environmental sustainability.

Improving feed efficiency and reducing enteric methane (CH_4) and carbon dioxide (CO_2) emissions by genetic selection represents an opportunity for reducing production costs (Tortereau *et al.*, 2020) and contributing to GHG mitigation strategies. Although a minimum detrimental effect on productivity is expected by improving residual feed intake (RFI) (Tortereau *et al.*, 2020) and FEC (Ferreira *et al.*, 2021), potential trade-off between production, efficiency, resilience or immune system, and GHG emissions need to be evaluated. Feed efficiency traits and GHG emissions are being recorded in lambs of the Corriedale Information Nucleus in Uruguay, providing the basis for investigating the associations among current and potentially new breeding objectives towards environmentally sustainable sheep production. The aim of this study was to investigate the association between growth, wool and FEC, based on the expected progeny differences (EPD) for these traits, and feed intake, feed efficiency and GHG emissions.

Materials & methods

Two hundred and seventeen Corriedale female lambs, from three cohorts (2018-2020) sired by 12 rams genetically linked breeding population were studied. The EPDs published in 2021 for weaning weight (WWT-E), body weight as hogget (YWH-E), greasy fleece weight (GFW-E) and FEC (FEC-E), which included animal own records, were provided by the National Sheep Genetic Evaluation System (www.geneticaovina.com.uy)

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_37, © E.A. Navajas *et al.* 2022

Feed efficiency and methane phenotypes. 42-day feed efficiency tests were carried out, after 14 days of acclimatisation to feed and facilities. Lambs were allotted to one of five pens according to initial body weight, type of birth and sire, and fed *ad libitum* with Lucerne haylage (22.0% of crude protein, 26.6% of acid detergent fibre, 32.2% of neutral detergent fibre, and 59.0% of dry matter). Average dry matter intake (DMI) and growth rate (ADG) were calculated based on daily feed intake and body weights recorded by automated feed bins and weighing platforms (Intergado®, Belo Horizonte, Brazil), respectively (Ferreira *et al.*, 2020). Portable accumulation chambers (PAC) were used to measure CH_4 and carbon dioxide (CO_2) emissions using protocol described by Paganoni *et al.* (2017). Two estimates per animal were performed in the last weeks of the feed intake test, with one week between them.

Data calculation and analysis. Residual feed intake (RFI, Koch *et al.*, 1963) was estimated as:

$$y = u + MBW + ADG + Test \times Pen + e(RFI)$$

where y = DMI (kg/day), MBW is the metabolic average body weight (kg, covariate), ADG is the average daily gain, previously calculated by linear regression (g/day, covariate), Test is the effect of the test (3 levels), Pen is the effect of the pen (5 levels) and RFI is the residual error.

Phenotypes of DMI, MBW and ADG, as well as CH_4 and CO_2, were adjusted by environmental effect using a lineal model that included Test × Pen as class effects, and age (days) as covariate. The corresponding residuals DMI-A, MBW-A, ADG-A, CH_4-A and CO_2-A were included in the analysis. Given the criteria used to allot animals to pens, adjusting by this factor also implies indirectly adjusting by initial body weight. The associations among traits were assessed by Pearson correlation coefficients based on the data described in Table 1. All analysis were carried out using CORR and GLM procedures in SAS program version 9.4 for Windows (SAS Inst., Cary, NC, USA).

Table 1. Mean, standard deviation (SD), minimum (Min) and maximum (Max) values of EPD for production traits, and feed efficiency and GHG emissions phenotypes of Corriedale female lambs (n=217).

Traits	Mean	SD	Min	Max
Weaning weight EPD (WWT-E, %)	7.80	3.70	-2.53	17.39
Yearling body weight EPD (YWT-E, %)	6.87	3.34	-2.51	15.84
Faecal worm egg count EPD (FEC-E, log FEC)	-0.11	0.10	-0.39	0.21
Greasy fleece weight EPD (GFW-E, %)	1.96	2.94	-5.63	11.82
Residual feed intake (RFI, kg DMI/day)	0.00	0.11	-0.26	0.49
Dry matter intake (DMI, kg/day)	1.21	0.22	0.49	1.80
Metabolic body weight (MWT, kg)	14.20	1.45	9.88	18.30
Average daily gain (ADG, kg/day)	0.17	0.04	0.05	0.27
Methane emissions (CH_4, g/day)	16.41	4.75	6.10	28.48
Carbon dioxide emissions (CO_2, g/day)	826.97	179.06	441.42	1,448.00
Adjusted DMI (DMI-A, kg/day)	0.00	0.15	-0.37	0.53
Adjusted MWT (MWT-A, kg)	0.00	0.77	-2.07	3.52
Adjusted ADG (ADG-A, kg/day)	0.00	0.04	-0.11	0.12
Adjusted CH4 (CH_4-A, g/day)	0.00	2.65	-7.93	7.91
Adjusted CO_2 (CO_2-A, g/day)	0.00	110.96	-319.50	340.64

Results

As expected, RFI was correlated with DMI (r=0.50) and independent of weight and growth. Similar associations were found between RFI and the adjusted traits, although the correlation coefficient with DMI-A was higher than with DMI (r=0.74). The associations among feed efficiency traits and GH emissions are in Table 2. High positive correlations were calculated for DMI and both GHG (~0.60), and the association was weaker for the adjusted phenotypes.

The correlations of growth and weights were also high with unadjusted and adjusted DMI, CH_4 and CO_2 (range 0.30 to 0.74). No significant association (P>0.05) were found between RFI and GHG emissions.

Body weights EPD had positive correlations with most of the traits included in this study. They were medium to high with DMI (0.57), MWT (0.66) and both GHG emissions traits (0.51-0.54). When these traits were adjusted, the correlation magnitude decreased to low-moderate (0.19-0.44) but remained significant (Table 3). The associations of GFW-E were positive and low with MWT and DMI, and negative with ADG. No association were found with between GFW-E and GHG emissions (P>0.05). On the other hand, any of the feed efficiency traits or GHG emissions were significantly correlated with FEC-E. RFI was not significantly associated with any of the EPD considered here.

Table 2. Correlation coefficients between feed efficiency and GHG emissions traits.[1]

Unadjusted[2]	**DMI**	**MWT**	**ADG**	**RFI**
DMI		**0.74**	**0.47**	**0.50**
CH_4	**0.58**	**0.52**	**0.38**	-0.04
CO_2	**0.59**	**0.70**	**0.51**	-0.03
Adjusted	**DMI-A**	**MWT-A**	**ADG-A**	**RFI**
DMI-A		**0.61**	**0.43**	**0.74**
CH_4-A	**0.26**	**0.35**	**0.30**	-0.04
CO_2-A	**0.42**	**0.54**	**0.45**	-0.03

[1] Coefficients significantly different from zero (P<0.05) are in bold.

[2] Abbreviations are described in Table 1.

Table 3. Correlation coefficients of feed efficiency and GHG emissions with EPD of production traits and FEC.[1,2]

	WWT-E	**YWT-E**	**FEC-E**	**GFW-E**
RFI	-0.05	-0.04	0.08	0.10
DMI	**0.57**	**0.57**	0.03	**0.21**
MWT	**0.66**	**0.66**	-0.02	**0.21**
ADG	**0.22**	**0.21**	-0.05	**-0.14**
CH_4	**0.54**	**0.53**	0.11	0.05
CO_2	**0.51**	**0.51**	0.06	-0.06
DMI-A	**0.19**	**0.20**	0.07	**0.23**
MWT-A	**0.43**	**0.44**	0.07	**0.35**
ADG-A	0.05	0.05	-0.08	-0.07
CH_4-A	**0.15**	**0.16**	0.05	0.07
CO_2-A	**0.24**	**0.24**	0.04	0.07

[1] Abbreviations are described in Table 1.

[2] Coefficients significantly different from zero (P<0.05) are in bold.

Discussion

Improving RFI by selection is an alternative to reduce production cost by lowering DMI without compromising animal performance (Tortereau *et al.*, 2020). The lack of statistically significant associations between RFI and genetic merits for growth and wool production is encouraging about the potential benefits of integrating this trait in the Corriedale breeding programme. However, we did not find a reduction in CH_4 emissions in more efficient lambs, as in other studies (i.e. Paganoni *et al.*, 2017). If genetic correlations follow these coefficients among adjusted phenotypes, RFI could not be considered as indirect selection criteria for GHG mitigation.

The positive associations between CH_4 and CO_2 and EPDs for growth and wool production indicate that reducing GHG emissions would imply a decrease of animal performance, and, therefore, of economic incomes derived from meat and wool. When GHG emissions were adjusted, correlation coefficients decreased from moderate-high to low magnitudes similar to those reported by Robinson *et al.* (2014).

One relevant driver of the associations among production, feed efficiency and GHG emission is DMI. Our results confirm DMI is highly correlated with RFI, as was reported at phenotypic and genetic level in the literature (Paganoni *et al.*, 2017; Tortereau *et al.*, 2020). High DMI is also associated with better animal performance, particularly when animals are fed at libitum like in the feed efficiency tests, as well as with CH_4 and CO_2. However, disentangling the biological interrelation among these traits is constrained given DMI remains as very difficult- to-measure trait, with limited available alternative for grazing conditions.

The non-significant correlations of FEC-EPD with feed efficiency traits indicate that improving FEC does not affect feed efficiency, productivity or GHG emissions. These results are in agreement with Ferreira *et al.* (2021), who did not find differences in these traits between lambs from Corriedale divergent FEC selection lines under natural and artificial GIN challenge.

In summary, these preliminary results indicate that improving feed efficiency by RFI would imply maintaining productivity with lower feed costs, with no unfavourable effect on genetic resistance to GIN and independently of GHG emissions. However, the positive correlations of DMI with GHG emissions and productivity implies a trade-off to be considered at the time of including GHG mitigation as new breeding objective, which will be checked by the estimation of the genetic correlations. As phenotyping continues, larger database will enable the estimation of these genetic parameters.

Acknowledgements

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 772787 (SMARTER), INIA-CL38- RUMIAR, the SusAn, ICT-AGRI 2 and FACCE ERA-GAS funding bodies (GrassToGas).

References

Ciappesoni G., Gimeno D., and Coronel F. (2014) Arch. Latinoam. Prod. Anim. 22:73-80.

Ferreira G.F., Ciappesoni G., Castells D., Amarilho-Silveira F., Navajas E.A., *et al.* (2021) Anim. Prod. Sci. 61(8): 754-760. https://doi.org/10.1071/AN20121

Koch R., Swiger L., Chambers D., and Gregory K. (1963) J. Anim. Sci. 22(2):486–494. Paganoni B., Rose G., Macleay C., Jones C., Brown D.J., *et al.* (2017). J. Anim. Sci. 95(9): 3839–3850. https://doi.org/10.2527/jas.2017.1499

Robinson DL, Goopy JP, Hegarty RS, Oddy VH, Thompson AN, *et al.* (2014) J. Anim. Sci. 92(10):4349-63. https://doi.org/10.2527/jas.2014-8042

38. Rumen microbiota and host gene expression associate to predisposed milk urea concentration in Holsteins

H. Honerlagen[1*], H. Reyer[1], D. Segelke[2], M. Oster[1], S. Ponsuksili[1], N. Trakooljul[1], B. Kuhla[1] and K. Wimmers[1,3]

[1]Research Institute for Farm Animal Biology (FBN), Wilhelm-Stahl-Allee 2, 18196, Dummerstorf, Germany; [2]IT-Solutions for Animal Production (vit), Heinrich-Schröder-Weg 1, 27283 Verden, Germany; [3]University of Rostock, Justus-von-Liebig-Weg 6, 18059, Rostock, Germany; honerlagen@fbn-dummerstorf.de

Abstract

The dairy sector has to reduce nitrogen (N) surpluses in the near future. In this context, the cow individual milk urea (MU) concentration has been suggested to reflect N-excretion in breeding strategies. The variability in MU levels is partly affected by ruminal processes attributed to microbes and host gene expression. To investigate the relevance of MU for breeding, we aimed to uncover ruminal mechanisms in cows with predisposed higher compared to lower MU (HMU vs LMU; n=20). Whereas the core microbiota at phylum level remained constant, the abundances of three microbial families differed between HMU and LMU cows. Transcriptome profiling of rumen tissue revealed 28 genes, including *MMP3*, *TNP2* and various genes of immune response that distinguished HMU and LMU cows. Our findings contribute to a deeper knowledge of MU predispositions in Holsteins and attribute to the optimisation of N-usage in conjunction with the reduction of N-emissions from dairy farms.

Introduction

The primary source of N-emissions from dairy cows is displayed by unutilised N, which is ingested by the cow with feed crude protein and excreted via urine, milk and faeces (Abdoun *et al.*, 2006). In this context, the ruminal conversion rate of N into valuable microbial protein, which is the result of rumen microbial growth, considerably contributes to the amount of N-utilisation and N-emissions (Tan *et al.*, 2021). Due to the high phenotypic correlation between milk urea (MU), urinary urea and blood urea concentration (Guliński *et al.*, 2016) the cow individual MU value has been suggested as substitute trait for breeding selection (Beatson *et al.*, 2019; Honerlagen *et al.*, 2021). In regard of successful breeding selection on N-use efficiency and reduced N-excretion from dairy cows, unravelling the molecular biological background of predisposed higher or lower MU concentrations (HMU – LMU) in dairy cows is required. Considering the large impact of ruminal processes on the N-metabolism in lactating ruminants, our study elucidated exclusively differences in the rumen microbial community and corresponding host-specific transcriptional processes between HMU and LMU cows.

Materials & methods

The predisposition for higher or lower MU concentration (HMU – LMU) was determined by deregressed proofs of a breeding value for MU, estimated by vit Verden (Vereinigte Informationssysteme Tierhaltung, Verden, Germany). The subsequent categorisation was further validated by the five recent milk records from each cow (HMU=278.1±15.8 mg/l, LMU 181.54±15.08 mg/l; mean ± SD). Based on this classification, 20 German Holsteins in late lactation were stalled in pairs, comprising one HMU and one LMU cow, treated as previously described (Müller *et al.*, 2021) and slaughtered after two weeks. Samples of rumen villi and rumen fluids (50 ml) were collected from each cow immediately after slaughter, snap frozen (-80 °C) and stored.

Laboratory analyses and sequencing pipelines. Microbial DNA was extracted from rumen fluids, quality checked and used to amplify the V4 hypervariable region of the 16S rRNA gene (Kozich *et al.*, 2013), which was sequenced on HiSeq2500 (Illumina, San Diego, CA). The reads were subsequently

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_38, © H. Honerlagen *et al.* 2022

filtered, trimmed, and removed from chimeras within the mothur software (version 1.44.1; Schloss *et al.*, 2009). Sequences were aligned to the Silva reference database (release 138; https://www.arb-silva.de/), aggregated into operational taxonomic units (OTU, sequence identity ≥97%) and annotated according to the Silva database. The count data was rarefied to the sample with the lowest read depth (745,235 reads). Statistical analyses were conducted at phylum (n=19) and family (n=116) levels. Total RNA was isolated from homogenized rumen villus epithelium, purified, qualitatively and quantitatively checked and utilized for mRNA library preparation. Libraries were paired-end sequenced on HiSeq2500. After quality control and filtering, sequences were aligned to the current reference genome (Bos_taurus.ARS-UCD1.2). The number of reads uniquely mapped to each gene were extracted using HTSeq (version 0.12.4).

Statistical analyses of rumen microbiome and transcriptome. Initially the core microbiome of HMU and LMU cows was investigated based on relative abundances at phylum level. Subsequently, DESeq2 was used to identify differentially abundant phyla and families between HMU and LMU cows (Love *et al.*, 2014). Similarly, the transcriptome dataset was analysed for differentially expressed genes (DEG) with DESeq2. The trial of each pair was included in all statistical models as fixed effect accounting for different crude protein levels in the feed rations. DEG were further assigned to their general biological functions using GeneCards (http://www.genecards.org).

Results

Bacteroidota und *Firmicutes* displayed the most dominant phyla in both, HMU and LMU cows by capturing 78.7% of relative abundances in all rumen fluids (43.4 and 35.3%, respectively) (Figure 1). The analysis at family level revealed *Monoglobaceae* ($P<0.01$) and *Ruminococcaceae* ($P=0.01$) being significantly more abundant in LMU compared to HMU cows. Microbes assigned to *Acetobacteraceae* sowed a significantly high relative occurrence in HMU compared to LMU cows ($P=0.04$).

In the filtered transcriptome dataset 28 DEG were identified between HMU and LMU animals. The highest fold changes were revealed for Matrix Metallopeptidase 3 (*MMP3*) and Transition Protein 2 (*TNP2*), with higher transcript abundance in HMU and LMU, respectively (Table 1). Five of the top ten DEG were identified as contributors of immune response and displayed higher expression in LMU cows.

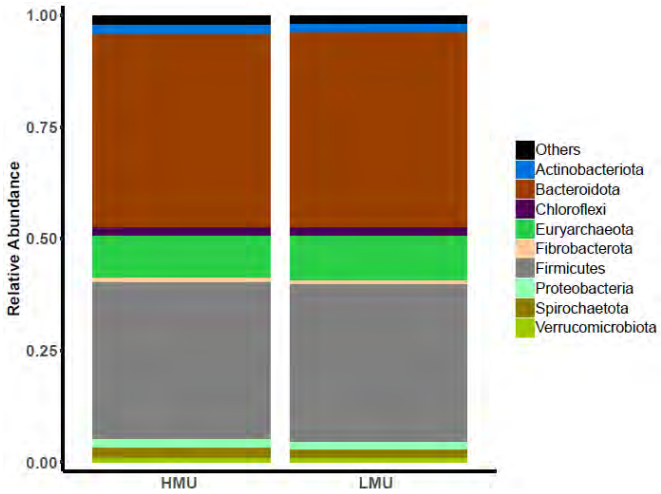


Figure 1. Microbial composition of rumen fluids from HMU and LMU cows at phylum level indicating a similar pattern with dominance of microbes assigned to *Bacteroidota* and *Firmicutes*.

Table 1. Most prominent DEG between HMU and LMU cows.

Gene symbol	Log 2 fold change[1]	Adjusted *P*-value[2]	General function[3]
TNP2	-1.40	0.0159	Fertility
MMP3	1.22	0.0356	Tissue remodeling
OAS1Y	-1.18	0.0316	Immune system
MX1	-1.09	0.0266	Immune system
IRF4	-0.99	0.0266	Immune system
PLA2G2D4	-0.95	0.0208	Immune system
SLAMF1	-0.94	0.0145	Immune system
GALNT8	0.91	0.0045	Protein metabolism
GPRC5A	0.91	0.0273	Cell development
CHI3L2	-0.84	0.0216	Cartilage biogenesis

[1] Derived from DESeq2 analysis and refer to HMU.

[2] Benjamini-Hochberg adjusted *P*-values.

[3] General functions were obtained from GeneCards database.

Discussion

Our study focused on the ruminal characterisation of Holsteins with predisposed higher or lower MU concentration (HMU – LMU). Both, HMU and LMU cows occupied the typical core microbiome of the bovine rumen (Wallace *et al.*, 2019) comprising mainly *Bacteroidota* and *Firmicutes*, which captured 78.7% of total rumen abundances (43.4 and 35.3% respectively). Apparently, MU predisposition was not substantially associated with distinguished rumen flora. However, three significantly differentially abundant families between HMU and LMU cows were identified. Interestingly, the *Ruminococcaceae* family, which displayed higher abundances in LMU cows has already been attributed a fundamental role in amino acidic and protein metabolism in the bovine rumen (Pacífico *et al.*, 2021). In addition, various *Ruminococcaceae* genera displayed higher abundances in goats with high N-efficient phenotypes (Wang *et al.*, 2019). Further analysis of the microbiome at genus level as well as the effect estimation of specific genera on N-utilisation and N-excretion might unravel substantial information of ruminal MU characterisation. The results of DEG analysis elucidated the enhanced expression of immune response genes in LMU cows. Interestingly, Zhang *et al.* reported the activation of immune response pathways in the rumen epithelium of calves with higher N-efficiency (Zhang *et al.*, 2021). Moreover, recent studies identified immune system activities of the host epithelia as key drivers of long- term host – microbial maintenance (Peterson *et al.*, 2007). In this context, the interplay between host gene expression, the microbial community and the rumen environment as well as the different partitioning of nitrogen to body compartments will be further investigated between HMU and LMU cows with regard to divergent N utilisation and N excretion mechanisms.

References

Abdoun K., Stumpff F., and Martens H. (2006) Anim Health Res Rev 7(1-2)**:**43-59. https://doi.org/10.1017/S1466252307001156

Beatson P. R., Meier S., Cullen N. G., Eding, H. (2019) Animal, 13(10): 2164-2171. https://doi.org/10.1017/S1751731119000235

Guliński P., Salamończyk E., and Młynek K. (2016) Anim Sci Pap Rep 34(1):5-23.

Honerlagen H., Reyer H., Oster M., Ponsuksili S., Trakooljul N., *et al.* (2021) Front Genet 12:1241. https://doi.org/10.3389/fgene.2021.699550

Kozich J.J., Westcott S.L., Baxter N.T., Highlander S.K., and Schloss P.D. (2013) Appl Environ Microbiol 79(17):5112-5120. https://doi.org/10.1128/AEM.01043-13

Love M.I., Huber W., and Anders S. (2014) Genome Biol 15(12):1-21. https://doi.org/10.1186/s13059-014-0550-8

Müller C.B.M., Görs S., Derno M., Tuchscherer A., Wimmers K., *et al.* (2021) Sci Total Environ 755(2):143-198. https://doi.org/10.1016/j.scitotenv.2020.143198

Pacífico C., Petri R.M., Ricci S., Mickdam E., Wetzels S.U., *et al.* (2021) Microorganisms 9(2):342. https://doi.org/10.3390/microorganisms9020342

Peterson D.A., Mcnulty N.P., Guruge J.L., and Gordon J.I. (2007) Cell Host Microbe 2(5):328-339. https://doi.org/10.1016/j.chom.2007.09.013

Tan P., Liu H., Zhao J., Gu X., Wei X., *et al.* (2021) Sci Total Environ 800:149-596. https://doi.org/10.1016/j.scitotenv.2021.149596

Wallace R.J., Sasson G., Garnsworthy P.C., Tapio I., Gregson E., *et al.* (2019) Sci Adv 5(7):83-91. https://doi.org/10.1126/sciadv.aav8391

Wang L., Liu K., Wang Z., Bai X., Peng Q., *et al.* (2019) Front Microbiol 10:239. https://doi.org/10.3389/fmicb.2019.00239

Zhang N., Teng Z., Li P., Fu T., Lian H., *et al.* (2021) PloS one 16(9):257-417. https://doi.org/10.1371/journal.pone.0257417

39. Predicting nitrogen use efficiency of individual dairy cows by mid-infrared spectra

R. Shi[1,2], W. Lou[1,2], B. Ducro[2], A. van der Linden[2], H.A. Mulder[2], S.J. Oosting[2], S. Li[1] and Y. Wang[1*]

[1]China Agricultural University, Beijing 100193, China; [2]Wageningen University & Research, P.O. Box 338, 6700 AH Wageningen, the Netherlands; wangyachun@cau.edu.cn

Abstract

Dairy farming may have negative impacts on the environment, such as ammonia emissions to air and nitrate leaching to water. By selecting more efficient cows, the amount of nitrogen excreted per cow and consequently the adverse environmental impacts will be reduced. The aim of this study was to predict nitrogen use efficiency (NUE) of individual dairy cows using mid-infrared (MIR) spectra of milk. A total of 600 feeding and MIR records of 56 Holstein cows were collected from a farm in Beijing, China. NUE was calculated as the ratio of nitrogen in milk to nitrogen intake. The coefficient of determination of the best model was 0.69, 0.62 and 0.70 for NUE, nitrogen loss and dry mater intake, respectively. The MIR wavenumbers around 981.00 and 1506.93 cm^{-1} and daily milk yield were the most important variables for prediction. These results show potential for large-scale genetic evaluations of nitrogen efficiency.

Introduction

As widely emphasised, nitrogen surpluses at dairy farms have negative impacts on the environment, such as ammonia emission and nitrate leaching to ground and surface water. Meanwhile, the increasing demand for animal products will result in higher production and thus results in more intensive dairy farming with higher total emissions of nitrogen (N) if N losses are not controlled. Therefore, a more efficient, and environmentally friendly production system is needed, in which NUE of dairy cows plays a key role. If efficient cows are selected, the fraction of intake that is not ending up in faeces or urine will be lower, which will contribute to lower N losses to the environment. Generally, this kind of resource allocation trait is difficult to measure individually (Lopez-Villalobos *et al.*, 2018). For instance, it requires assessment of daily feed intake, which is costly. To genetically improve NUE, routinely recording and cost-effective proxies are essential to initiate genetic evaluations. Recently, Grelet *et al.* (2020) obtained reasonable proxies for N related traits such as NUE and nitrogen loss (NL) by including MIR spectra of dairy cows in their prediction models. A maximum coefficient of determination (R^2) of 0.82 was observed in their validations, which indicated the proxies were reasonable for further genetic analysis.

However, these prediction models were only based on records from Holstein cows in early lactation. Individual NUE and NL in other lactation stages have not been predicted yet with MIR data. Literature also indicated that non-informative signals may exist in original MIR data, which will decrease the relation between phenotypes and MIR spectra (Bresolin and Dórea, 2020). Pre-treated MIR data may be beneficial for constructing robust prediction models. Therefore, different pre-processing methods were assessed in the current study. Additionally, it is of interest to quantify the contribution of different input variables. The objectives of this study were (1) to develop the best prediction model of NUE, NL and dry matter intake (DMI) for individual dairy cows in China, (2) to compare different prediction models of these traits, and (3) to investigate the importance of input variables in the prediction.

Materials & methods

A feeding trial was conducted in an experimental farm in Beijing, China between March 2017 and August 2017. A total of 56 Chinese Holstein cows in mid and late lactation stage (154-452 days in milk (DIM)) were offered the same diet during the experimental period in a free-stall design barn. Daily feed intake of

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_39, © R. Shi *et al.* 2022

individual cows was recorded by an automatic system (Roughage Intake Control System, Insentec B.V., Marknesse, the Netherlands). Daily DMI was calculated for each cow based on feed intake and dry matter content of the diet. Each cow was evaluated monthly for body condition scores (BCS) by two technicians. Additionally, the daily milk yield for each cow was recorded by the milking system.

Individual MIR spectra were obtained from Fourier transform spectrometers (Bentley, Chaska, USA) at the Beijing Dairy Cattle Centre. Fat, lactose, milk urea nitrogen (MUN) and total protein of milk samples were estimated from the MIR analysis.

NUE was defined as ratio of total N output in milk to total N intake from feed and NL was defined as total N intake from feed minus total N output in milk (Calsamiglia *et al.*, 2010). Two pre-processing methods were applied to the original MIR data to reduce the influence of noise in the MIR spectra, namely multiplicative scatter correction (MSC) and standard normal variate (SNV). Subsequently, wavenumbers induced by water were omitted, resulting in 215 data points for each record (Grelet *et al.*, 2016). The MIR spectra were then combined with feeding trial data, providing 600 records of 56 cows.

The selected MIR spectra, parity (1 and 2 or more), corresponding daily milk yield, BCS, DIM_g (actual DIM grouped by 5 days) and additional parameters generated by MIR analysis (protein, fat, lactose and MUN) were included for model development. MIR spectra were included in prediction either as original spectra or as result from pre-processing (MSC-spectra or SNV-spectra).

Furthermore, three different algorithms were applied for prediction: partial least squares (PLS), ridge regression (RR), and support vector machine (SVM) regression. Consequently, a total of 48 models were applied for predicting each trait (Table 1).

To validate the models, the dataset was randomly split in test and training sets in a ratio of 1 to 3, and a cow can only be in the test set or in the training set. Prediction models were constructed using the training sets and validated using the test sets. The performance metrics included R^2, root mean square error (RMSE), relative error (RE, calculated as RMSE/mean of the global data), and ratio of standard deviation of the global data to RMSE from the validation (RPD). Besides, the Spearman correlation coefficient (SpearR) between true values and predictions was included to test re-ranking of the best model. Importance of each input variable for the PLS and RR methods was defined as the absolute value of the corresponding regression coefficient.

Table 1. Prediction models for nitrogen use efficiency, nitrogen loss and dry matter intake.

Models	Predictors[1]	Number of input variables	Spectra	Algorithms	Count[2]
Model 1	MIR	215	Original, MSC, SNV	PLS, RR, SVM	9
Model 2	+ Milk yield, parity	217			9
Model 3	+ BCS	218			9
Model 4	+ DIM_g	219			9
Model 5	+ MUN	220			9
Model 6	+ Protein, fat, lactose	8	No MIR		3

[1] For model 2 to 5, the additional predictor of next model is based on previous model, and model 6 includes all additional predictors, except for MIR spectra.
[2] Number of prediction models developed using this set of predictors: 3 algorithms times 3 types of MIR spectra for models 1-5 = 9 models. See main text for abbreviations.

Results

Individual daily NUE was on average 26.4%, with a standard deviation of 8.6%, NL was on average 0.51 kg/day, with a standard deviation of 0.15 kg/day, and DMI was on average 25.15 kg/day, with a standard deviation of 5.95 kg/day.

Model 6 was the least accurate for NL and DMI regardless of the performance metrics (Figure 1). Model 2 generated the best prediction for NUE and DMI, with a relatively high R^2 and RPD, as well as a small RE. Model 1 had a better application to NL, from which the highest R^2 and RPD, as well as lowest RE were obtained. For most models with MIR, the pre-processing step resulted on average in a 12.3% higher R^2 and 8.6% higher RPD, in combination with a 11.2% smaller RE on average.

Model 2 using the SVM method, and model 1 and model 2 using the PLS method were the best models for predicting NUE, NL and DMI, respectively (Table 2). SNV-spectra were included in these models, and the best model for NL only included MIR spectra.

The most important variables were MIR wavenumbers 973.54 to 988.46 cm^{-1}, 1,480.80 to 1,574.07 cm^{-1}, and daily milk yield based on feature score.

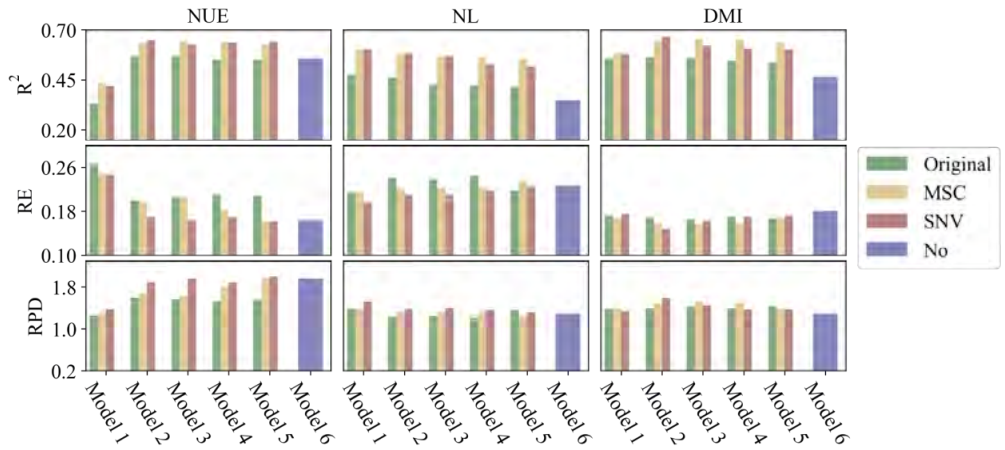


Figure 1. Statistics of different prediction models for nitrogen use efficiency, nitrogen loss and dry matter intake.

Table 2. Performance of best model for each trait[1].

Trait	Model information	R^2	RE	RPD	SpearR
NUE	Model 2, SNV-spectra, SVM	0.69	0.15	2.11	0.84
NL	Model 1, SNV-spectra, PLS	0.62	0.20	1.46	0.79
DMI	Model 2, SNV-spectra, PLS	0.70	0.14	1.70	0.81

[1] See main text for abbreviations.

Discussion

The current study developed accurate prediction models for daily NUE, NL and DMI of individual cows by comparing different prediction algorithms and precorrection methods for MIR data. Furthermore, the contribution of input variables was evaluated. The performance metrics of the best models for these three traits were comparable with those in the study of Grelet *et al.* (2020), which reported a range of R^2 from 0.59 to 0.68, RE from 0.14 to 0.23, and RPD from 1.57 to 2.07. It should be noted that the equations in Grelet *et al.* (2020) were based on three farms in three countries and therefore they might be more robust than prediction equations based on one farm. As long as feeding regimes are very similar on other farms, the prediction equations may facilitate genetic evaluation on a larger data set of MIR spectra. For instance, Chen *et al.* (2021) applied the prediction models of Grelet *et al.* (2020) to a large dataset for genetic evaluation of predicted NUE and NL in early lactation.

In this study, adding MIR spectra in the best models increased the R^2 and RPD by 10 to 30%, as well as reduced the RE by 0.03 to 0.10 compared to using model 6 (Figure 1). Meanwhile, this improvement was more obvious when MIR spectra were pre-treated. These results indicate that MIR includes additional information for better prediction of NUE, NL and DMI. Pre-treated MIR spectra were better than raw MIR spectra for developing accurate prediction models. Milk yield was an important variable in all models because it is highly correlated with N output (phenotypic correlation=0.96). The specific wave range of MIR that was important in prediction comprised C-H, C-N and N-N stretching (Grelet *et al.*, 2015; Xia *et al.*, 2018), which is related to proteins.

The results of current study indicate that the methodology is suitable to exploit MIR data for prediction of NUE, NL and DMI as reflected by a relatively high R^2. The resulting predictions can be applied to improve nutrition and management of dairy herds in China and to perform further genetic evaluations.

Acknowledgements

The China Scholarship Council (No. 201913043) is acknowledged for funding the PhD projects of Rui Shi and Wenqi Lou, and the Beijing Dairy Cattle Centre is acknowledged for providing data. This research is supported by China Agriculture Research System of MOF and MARA.

References

Bresolin T., and Dórea J.R.R. (2020) Front. Genet. 11:923. https://doi.org/10.3389/fgene.2020.00923

Calsamiglia S., Ferret A., Reynolds C.K., Kristensen N.B., and van Vuuren A.M. (2010) Animal 4:1184–1196. https://doi.org/10.1017/S1751731110000911

Chen Y., Vanderick S., Mota R.R., Grelet C., and Gengler N. (2021) J. Dairy Sci. 104:4413–4423. https://doi.org/10.3168/jds.2020-18849

Grelet C., Fernández Pierna J.A., Dardenne P., Baeten V., and Dehareng F. (2015) J. Dairy Sci. 98:2150–2160. https://doi.org/10.3168/jds.2014-8764

Grelet C., Bastin C., Gelé M., Davière J.-B., and Johan M. *et al.* (2016) J. Dairy Sci. 99:4816–4825. https://doi.org/10.3168/jds.2015-10477

Grelet C., Froidmont E., Foldager L., Salavati M., and Hostens M. *et al.* (2020) J. Dairy Sci. 103:4435–4445. https://doi.org/10.3168/jds.2019-17910

Lopez-Villalobos N., Correa-Luna M., Burke J.L., Sneddon N.W., and Schutz M.M. *et al.* (2018) Proc. of NZSAP 2018, Lincoln, New Zealand.

Xia Y., Ugarte C.M., Guan K., Pentrak M., and Wander M.M. (2018) Soil Sci. Soc. Am. J. 82:1415–1427. https://doi.org/10.2136/sssaj2018.05.0175

40. Clearing the air between methane and commercial beef cattle in Ireland

C.V. Ryan[1,2*], S.F. Kirwan[3], J.J. Crowley[4], D.C. Purfield[2], T. Pabiou[1] and R.D. Evans[1]

[1]Irish Cattle Breeding Federation, Highfield House, Shinagh, Bandon, Co. Cork, Ireland; [2]Munster Technological University, Bishopstown, Co. Cork, Ireland; [3]Animal Bioscience Research Centre, Teagasc Grange, Dunsany, Co. Meath, Ireland; [4]AbacusBio Ltd., 442 Moray Pl., Dunedin 9016, New Zealand; cryan@icbf.com

Abstract

The focus of this research was to analyse CH_4 data recorded by GreenFeed emission monitoring systems by assessing the repeatability of such phenotypic measures, to estimate genetic parameters and generate estimated breeding values (EBVs) for CH_4 related traits. Additionally, validation of EBVs in this research will aid in establishing a protocol for CH_4 measurement for the generation of national genetic evaluations. In total 979 animals have completed CH_4 recording to date, with 211,136 individual CH_4 observations. The repeatability and heritability of CH_4 increased as the number of the days of methane measurement per animal included in the phenotypic average increased however, EBV validation suggests shorter averaging periods result in better, more consistent EBVs.

Introduction

Measuring CH_4 is a costly exercise, but necessary to incorporate CH_4 traits into national breeding goals with the aim of reducing enteric CH_4 emissions. The GreenFeed emission monitoring (GEM) system estimates daily CH_4 production by measuring gas concentrations and airflow from cattle when they visit a GEM unit (Hammond *et al.*, 2015). While respiration chambers remain the gold standard for CH_4 measurement of livestock, strong correlations (0.81 (S.E.=0.10)) between respiration chambers and the GEM system have been reported (Garnsworthy *et al.*, 2019), suggesting GEM is a suitable lower-cost and scalable alternative to measure CH_4. However, there is large disparity between GEM protocols in existing CH_4 emission measurement studies as CH_4 concentrations (CH_4 gram/day) vary with visit number and duration to the GEM system which has consequences for data comparison across studies (Della Rosa, *et al.*, 2021). The objective of this study therefore was to develop a standardised protocol for routine enteric CH_4 data analysis with the goal of producing national EBVs.

Materials & methods

The data used in the present study were obtained from a pre-existing database managed by the Irish Cattle Breeding Federation (ICBF).

Data. CH_4 and carbon dioxide flux measurements using GreenFeed machines (C-Lock Inc., 2021) were recorded on 979 cattle destined for slaughter between 2018 and 2021 as part of the progeny performance testing in the Gene Ireland Progeny Performance Test Centre (Tully, Co. Kildare, Ireland). Measurements of feed intake, liveweight and ultrasound muscle measurements were also available on all animals. Animals ranged from 368 days to 910 days of age at the time of entry to test centre and were the progeny of AI bulls under evaluation of the Gene Ireland Breeding program. Animals were purchased in groups, referred to as intakes. All 26 intakes of animals were unisex and had an average intake size of 38 animals (S.D.=16.58). The final intake dataset was compromised of 510 steers, and 361 heifers and 90 bulls. On arrival to the test centre all animals underwent acclimatisation periods on both Insentec automatic feed stations for feed intake measurement and GEM machines. Full details of animal recruitment and acclimatisation management are detailed elsewhere by Kelly *et al.* (2019). Dry matter intake (DMI) was recorded as the sum of intake recorded via both the Insentec and GEM systems. To calculate Residual CH_4 (RCH_4) the

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_40, © C.V. Ryan *et al.* 2022

phenotypic relationship between each CH_4 measurement and feed intake at discrete intervals prior to the observation was determined. The resulting correlations with CH_4 were 0.55, 0.55, 0.49, 0.40, 0.31, 0.25, 0.24, 0.36 for the time periods 3, 6, 9, 12, 15, 18, 21, 24 hours respectively. Hence as each time period shows some relationship to CH_4, RCH_4 was estimated using multiple regression incorporating all DMI time periods in MiX99 (Luke, 2017) where RCH_4 is the methane corrected for DMI.

Data edits. From the initial data set of 979 animals, data from visits outside of the defined test period were excluded from all analysis. CH_4 measurements less than zero or in the top or bottom 1% of the data were also removed. For the estimation of genetic parameters, animals with unknown sires were excluded (n=14) and only CH_4 measurements where a minimum of three sires were represented in a given day were included. Animals were also required to have CH_4 records on five individual days. After edits, 712 animals were available for genetic parameter estimation, with 84,091 observations retained.

Trait definitions investigated. Seven phenotypic definitions of CH_4 and RCH_4 were investigated in this analysis. Phenotype 1 was each individual CH_4 observation, phenotypes 2-6 where CH_4 and RCH_4 observations were averaged every 1,5,10,15 and 30 days per animal and phenotype 7 where CH_4 and RCH_4 observations were averaged across the full test period per animal.

Phenotypic analysis. Phenotypic repeatability (for CH_4 only) was estimated for each phenotypic CH_4 averaging period using SAS Proc Mixed procedure, where contemporary group defined as GreenFeed machine, intake group number and day of observation and animal were included as fixed effects.

Genetic analysis. Univariate repeatability animal models in DMU (Madsen, P., and Jensen, J., 2013) were used to estimate variance components. The animal model applied to all seven traits can be described as:

$$y = Xb + Za + Wp + e \quad (1)$$

Where y was the methane trait (CH_4 and RCH_4), b is a vector of fixed class effects (contemporary group, time of day) and fixed regressions (age, breed proportion, general heterosis coefficient), a is the additive genetic effect of the animal, p is the random repeated environmental effect of animal (phenotype 1-6 only), and X, Z and W are the corresponding incidence matrices. Additionally, in phenotype 1, a random animal effect was included to account for within day repeatability per animal. EBVs were estimated using MiX99. Validation was performed by generating EBVs for whole and partial evaluation; partial evaluation excluded the most recent third of animals to enter the test facility, based on test start date. The full evaluation comprised of 961 animals with phenotypes and the validation evaluation had 658 animals. The linear regression method by Macedo, *et al.* (2020) was used to estimate the bias (both level and dispersion) and accuracy of both evaluations.

Results

Phenotypic repeatability of CH4 data. To ensure all animals were represented in each repeat period (Table 1), phenotypic analysis included all animals who were on test for a minimum of 60 days (n=340). Repeatability of CH_4 measure increased as the length of the averaging period increased (Table 1). The average number of visits per animal per day was considerably lower for the young bulls in comparison to the steers and heifers cohorts. However, diet differences between the young bull cohort means the differences in mean CH_4 measurement is not comparable to steer and heifer means.

Table 1. Phenotypic repeatability of CH_4.

	Steers		Heifers		Young bulls	
Number of animals	175		100		65	
Mean CH_4 (SD)	240.23 (84.59)		233.29 (77.74)		158.92 (73.42)	
Average visits per day (SD)	2.59 (1.03)		3.09 (0.70)		0.45 (0.23)	
Repeat period	n[1]	Repeatability	n[1]	Repeatability	n[1]	Repeatability
Within day	31,261	0.16	21,660	0.49	1,417	0.17
1 day period	10,204	0.16	6,649	0.28	725	0.24
5-day period	2,260	0.54	1,503	0.46	345	0.30
10-day period	1,203	0.60	794	0.54	259	0.34
15-day period	827	0.69	573	0.53	214	0.31
30-day period	498	0.62	314	0.49	128	0.45

[1] Number of observations in repeatability analysis.

Heritability estimates of methane traits. Heritability estimates of both CH_4 and RCH_4 increased as the duration of the averaging period increased (Table 2). Standard errors were high due to the limited nature of the dataset, but heritability was deemed significant.

EBV validation. Overall, RCH_4 was more consistent in the dispersion bias than CH_4 across the averaging periods, however the differences are small due to the limited dataset (Table 3). This suggests that RCH_4 may be a more suitable trait for inclusion in the national index.

Table 2. Heritability estimates for methane traits.

		CH_4			RCH_4		
	n[1]	Genetic variance	Residual variance	Heritability (SE)	Genetic variance	Residual variance	Heritability (SE)
Individual observations	84,091	338	3,372	0.08 (0.033)	316	2,527	0.09 (0.036)
1 day average	26,562	309	1,591	0.12 (0.057)	307	1,217	0.16 (0.063)
5-day average	6,390	281	624	0.19 (0.092)	234	524	0.20 (0.092)
10-day average	3,498	290	435	0.22 (0.108)	305	376	0.29 (0.114)
15-day average	2,550	255	404	0.20 (0.109)	262	359	0.26 (0.113)
30-day average	1,436	295	345	0.26 (0.124)	298	298	0.32 (0.131)
Full test average	712	324	732	0.31 (0.136)	352	545	0.39 (0.143)

[1] Number of observations in parameter estimation.

Table 3. Validation of CH_4 and RCH_4 EBVs.

	CH_4			RCH_4		
	Level bias	Dispersion bias	Ratio accuracy difference	Level bias	Dispersion bias	Ratio accuracy difference
Individual observations	-0.647	0.755	0.014	-0.545	0.851	0.024
1 day average	-0.661	0.730	0.008	-0.565	0.868	0.019
5-day average	-0.634	0.757	0.020	-0.634	0.757	0.009
10-day average	-0.692	0.694	0.023	-0.505	0.768	0.018
15-day average	-0.599	0.719	0.016	-0.599	0.719	0.012
30-day average	-0.712	0.707	0.028	-0.712	0.707	0.004
Full test average	-0.642	0.674	0.040	-0.624	0.707	0.003

Discussion

Methane is an expensive trait to record, and as such data utilisation needs to be maximised. Visit frequency of animals to the GEM system has an impact on the repeatability of phenotypic CH_4 measures and requires careful consideration when utilising CH_4 data. While increasing the length of the averaging period results in higher levels of heritability, EBV validation would suggest that the shorter averaging periods result in a more accurate EBVs due to the increased number of CH_4 measures included in the analysis. Ongoing work on the correct validation approach that correctly accounts for the non-homogenous nature of existing data and limited data size still needs to be completed. While the validation cohort in this analysis were the most recent third of animals, the number of individual CH_4 records excluded was 59.5% of total CH_4 records due to the heterogeneity of records between more recent animals and earlier animals. Other validation cohorts which have a less severe loss of records will be investigated. Further analysis will be required to establish the best trait for inclusion in the national index that is independent of traits in the current breeding objective and fits with industry requirements going forward.

References

C-Lock (2021). C-Lock – Products – GreenFeed – Large Animals. https://www.c-lockinc.com/researchers/products/greenfeed-large-animals (Accessed 16 May 2021).

Della Rosa, M.M., Jonker, A. and Waghorn, G.C. (2021). Anim. Feed Sci., 279, p.115018. https://doi.org/10.1016/j.anifeedsci.2021.115018

Garnsworthy, P.C., Difford, G.F., Bell, M.J., Bayat, A.R., Huhtanen, *et al.* (2019). Animals, 9(10), p.837. https://doi.org/10.3390/ani9100837

Hammond, K.J., Humphries, D.J., Crompton, L.A., Green, C. and Reynolds, C.K. (2015). Anim. Feed Sci., 203, pp.41-52. https://doi.org/10.1016/j.anifeedsci.2015.02.008

Kelly, D.N., Murphy, C., Sleator, R.D., Judge, M.M., Conroy, S.B., *et al.* (2019). J. Anim. Sci., 97(11), pp.4405-4417. https://doi.org/10.1093/jas/skz316

Macedo, F.L., Reverter, A. and Legarra, A. (2020). J. Dairy Sci, 103(1), pp.529-544. https://doi.org/10.3168/jds.2019-16603

MiX99 Development Team (2017). Biometrical Genetics, Natural Resources Institute Finland (Luke), FI-31600 Jokioinen, Finland.

Madsen, P., and Jensen, J. (2013). DMU Ver. 6, rel. 5.2 https://dmu.ghpc.au.dk/dmu/DMU/Doc/Current/dmuv6_guide.5.2.pdf

41. The impact of selection for divergence in methane emissions on alimentary tract measures of eight-month-old sheep

K.M. McRae[1*], S.M. Hickey[2], G.J. Greer[1], B. Bryson[3], K.J. Knowler[1], W.E. Bain[1], S.N. Elmes[1], E.A. Young[1], J. Wing[1], E. Waller[1], J.C. McEwan[1] and S.J. Rowe[1]

[1]AgResearch Invermay, Mosgiel, New Zealand; [2]AgResearch Ruakura, Hamilton, New Zealand; [3]AgResearch Woodlands, Woodlands, New Zealand; kathryn.mcrae@agresearch.co.nz

Abstract

Post-slaughter alimentary tract measures were collected from eight-month-old ram lambs during routine culling of sheep divergently selected for methane yield. Measures from 142 and 132 lambs from the high and low methane lines respectively were collected over five years. Low CH_4 animals had lighter reticulo-rumen full, empty and contents weights ($P<0.05$), but tended to have longer rumen papillae ($P=0.06$). Preliminary heritability estimates for reticulo-rumen components and papillae traits were significantly different from zero, ranging from 0.09 to 0.25. Results validate earlier findings that differences in the anatomical characteristics of the alimentary tract are associated with methane emissions and suggest that there is a significant host genetic component, warranting further investigation.

Introduction

Ruminant livestock production systems are under increasing scrutiny due to the release of enteric methane: a potent greenhouse gas associated with climate change. In New Zealand, enteric methane emissions (CH_4) from grazing livestock account for a third of total greenhouse gas emissions (MfE 2020). In ruminants, CH_4 comes primarily from the enteric fermentation process, where through a symbiotic relationship that has evolved over millennia, microbes in the reticulo-rumen break down and ferment plant material, providing energy for the host animal. Despite ruminants providing ~51% of animal protein consumed by humans, the complexity of ruminant digestion is poorly understood. Furthermore, there is little published literature on the plasticity of the alimentary tract despite huge variation across bovinae and ungulates. Research has shown that there is at least partial host genetic control of the symbiotic relationship as methane emissions are heritable and repeatable traits in sheep (Pinares-Patiño *et al.* 2013). Differences in methane production have been associated with variation in the alimentary tract and rumen microbial community structure (Kittelmann *et al.* 2014). Sheep with reduced CH_4 emissions have been shown to have differences in rumen physiology compared to their high emitting counterparts, including smaller rumen surface area and volume (Goopy *et al.* 2014; Bain *et al.* 2014; Lambe *et al.* 2019), and shorter retention times through the alimentary tract (Pinares-Patiño *et al.* 2003; Bond *et al.* 2017). There is also preliminary evidence from CH_4 yield selection lines for differences in lean muscle yield, dressing out percentage, and rumen and reticulum full weights (Elmes *et al.* 2014), as well as feed intake and feeding behaviour (Johnson *et al.* 2018). Breeding for reduced CH_4 emissions in sheep may therefore fundamentally change the morphology of the alimentary tract. Understanding this relationship could be informative for the selection of low methane emitting ruminants and the impact of such selection. The objective of this study was to determine the impact of selecting for divergent CH_4 emissions on anatomical characteristics of the reticulo-rumen.

Materials & methods

The use of animals for this study, including welfare and experimental procedures, was approved by AgResearch's Invermay and Grasslands Animal Ethics committees (Approvals 14907, 14603, 14491, 14370, 14455, 14007).

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_41, © K.M. McRae *et al.* 2022

Animals. Animals were from a flock selected for divergent CH_4 yield (g CH_4 / kg DMI), described fully in (Pinares-Patiño *et al.* 2013). Briefly, 1000 ewes were initially screened from four research progeny test flocks, with 100 ewe lambs retained based on CH_4 yield extremes. Animals were selected using respiration chamber measurements, repeated after 10-14 days (Pinares-Patiño *et al.* 2011). From the above flocks, 96 rams were selected from the born-2009 lambs. The lines were closed in 2012, and all subsequent animals used have been born within the flock, which is currently maintained at 100 ewes per line. Ram lambs that were surplus to breeding requirements were slaughtered between six and nine months of age through either a commercial processing plant or a research abattoir (Table 1).

Measurements. Live weight was recorded for each animal before slaughter. Animals that were processed through a commercial processing plant were slaughtered approximately 24 hours after removal from pasture, whereas those processed through a research abattoir were taken off pasture at least two hours before slaughter. Post-euthanasia, the gastrointestinal tract was removed and the full reticulo-rumen weight was recorded before thorough washing in tap water and re-weighing. Reticulo-rumen contents were estimated by subtracting the empty weight from the full weight. A 2 cm^2 sample was taken from the dorsal area of the rumen atrium, in the centre of the area with profuse papillae (Figure 1), pinned to a corkboard and stored in 10% paraformaldehyde for 10 days before being transferred 70% alcohol.

Table 1. Summary of animals with alimentary tract measures.

Birth year	High CH_4 yield	Low CH_4 yield	Average age (days)	Processing location
2012	37	23	261 (250-270)	Commercial
2013	30	33	255 (239-263)	Commercial
2014	28	29	239 (223-254)	Commercial
2015	30	29	223 (216-233)	Commercial
2017	17	18	274 (250-289)	Research abattoir

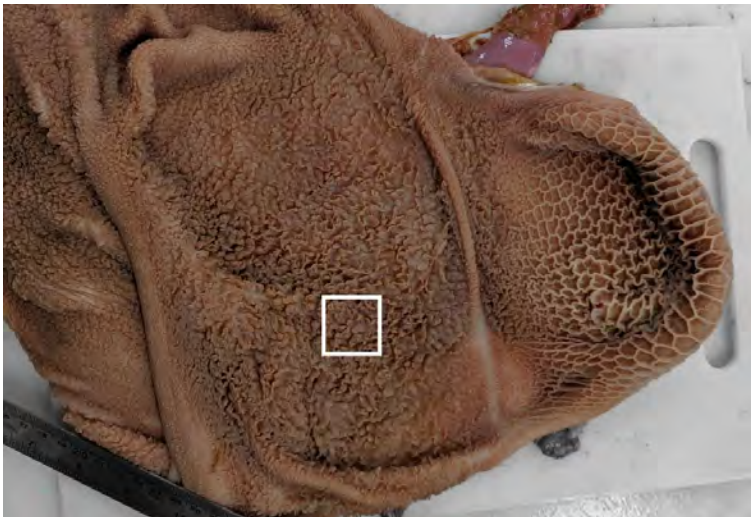


Figure 1. Ovine reticulo-rumen, showing the location for rumen papillae sampling.

Papillae density per cm^2 and length and width of 10 papillae per animal were measured using a dissecting microscope. Papillae length and width were used to estimate surface area (SA) per mm^2 using the following equation (Hill *et al.* 2005):

$$SA=2\times r\times \pi\times L+ \pi\times r^2 \quad (1)$$

where r = radius in mm and L = length in mm.

Statistical analysis. Alimentary tract measures were analysed separately using a standard least-squares model in SAS (SAS Institute Inc., Cary NC, USA) fitting sire as a random effect, selection line (high or low CH_4) and year of birth as fixed effects, and pre-slaughter weight as a covariate. Variance components were estimated using restricted maximum likelihood (REML) procedures fitting a sire model.

Results

The CH_4 selection lines differed in several reticulo-rumen measurements (Table 2). Low CH_4 animals had lighter reticulo-rumen full, empty and contents weights ($P<0.05$), and tended to have longer rumen papillae ($P=0.06$). Using a sire model, preliminary heritability estimates ranged between 0.16 and 0.19 for reticulo-rumen weight measures (Table 2). The heritability for papillae height was 0.25±0.07, however, papillae width was non-significant (0.10±0.06).

Discussion

The finding that low CH_4 emitting sheep have a smaller reticulo-rumen relative to body size is consistent with previous studies (Goopy *et al.* 2014; Bain *et al.* 2014; Bond *et al.* 2017). The mean full reticulo-rumen weight of low CH_4 yield animals was 6% lighter than their high CH_4 counterparts, which is in line with the 10% difference observed by Elmes *et al.* (2014) using a subset of the animals used in this study. Differences in rumen volume as estimated by CT scanning were larger (20%) in the 2009-born selection line ewes (Bain *et al.* 2014), however, that study screened in the phenotypic extremes for methane yield.

The weight of the entire gastrointestinal tract also has important implications for commercial farmers beyond the association with CH_4 emissions. To optimise carcass value, animals are weighed before selection for slaughter as a prediction of carcass weight. Dressing out percentage (carcass weight as a percentage of pre-slaughter weight) can vary, however, with one of the biggest contributing variables being the weight of the gastrointestinal tract ('gut fill') at the time of weighing. Gut fill can be between 10 and 22% of pre-fast live weight (Litherland, Dynes, and Moss 2010). Dressing out percentage was reported to vary in the CH_4

Table 2. Mean ± standard error and heritability of alimentary tract measures from high CH_4 (n=142) and low CH_4 (n=132) yield animals.

Trait	High CH_4	Low CH_4	*P*-value	Diff.[1]	Heritability
Reticulo-rumen full (g)	5,358±82	5,052±83	0.01*	6%	0.19±0.07
Reticulo-rumen empty (g)	988±8	956±8	0.01*	3%	0.16±0.07
Est. rumen contents (g)	4,370±81	4,096±82	0.02*	6%	0.20±0.07
Papillae count (per cm^2)	32±1	31±1	0.61	2%	0.09±0.05
Av. papillae height (mm)	3.69±0.09	3.92±0.09	0.06	-6%	0.25±0.07
Av. papillae width (mm)	1.94±0.04	1.95±0.04	0.85	0%	0.10±0.06
Av. papillae surface area	26.6±1.0	28.5±1.1	0.21	-7%	0.22±0.07

[1] Percentage difference between high and low CH_4 as a proportion of the high.

selection line animals, with low CH_4 animals having lighter reticulo-rumen weights and a higher dressing-out percentage (Elmes *et al.* 2014). Dressing out percentage is heritable (0.25±0.03) in New Zealand lambs (Brito *et al.* 2017), and the genetic relationship between dressing out percentage, CH_4 emissions and alimentary tract measures warrants further investigation.

While estimated in a small number of selection line animals, a significant genetic component was observed to multiple alimentary tract measures, including reticulo-rumen full and empty weights, rumen contents, and rumen papillae length. This finding is consistent with the moderate heritability of reticulo-rumen volume in Texel lambs (h^2=0.45±0.13), as estimated by computed tomography (CT) images (Lambe *et al.* 2022). Together, these results indicate that future research is needed in a larger cohort of animals to estimate accurate genetic parameters for alimentary tract measures, along with their relationship to production traits.

References

Bain W.E., Bezuidenhout L., Jopson N.B., Pinares-Patiño C.S., and McEwan, J.C. (2014) Proc. of the 10th WCGALP, Vancouver, Canada.

Bond J.J., Cameron M., Donaldson A.J., Austin K.L., Harden S. *et al.* (2017) Anim. Prod. Sci. 59(1):55–65. https://doi.org/10.1071/AN17141

Brito L.F., McEwan J.C, Miller S., Bain W.E., Lee M. *et al.* (2017) Small Rumin. Res. 154:81–91. https://doi.org/10.1016/j.smallrumres.2017.07.011

Elmes S.N., Bain W.E., Greer G.J., Hickey S.M., Young E.A. *et al.* (2014) Proc. NZSAP, Napier, New Zealand.

Goopy J.P., Donaldson A., Hegarty R., Vercoe P.E., Haynes F *et al.* (2014) Br. J. Nutr. 111(4):578–85. https://doi.org/10.1017/S0007114513002936

Hill S.R., Hopkins B.A., Davidson S., Bolt S.M., Diaz D.E. *et al.* (2005) J. Dairy Sci. 88(1): 324–26. https://doi.org/10.3168/jds.S0022-0302(05)72691-6

Johnson P.L., Knowler K., Wing J., Hickey S., and Johnstone P. (2018) Proc. of the 11th WCGALP, Auckland, New Zealand.

Kittelmann S., Pinares-Patiño C.S., Seedorf H., Kirk M.R., Ganesh S. *et al.* (2014) PloS One 9(7):e103171. https://doi.org/10.1371/journal.pone.0103171

Lambe N., Miller G., McLean K., Gordon J., and Dewhurst R. (2019) Proc. BSAS, Edinburgh, United Kingdom.

Lambe N., McLaren A, McLean K., Gordon J., Conington J. *et al* (2022) Anim. Prod. Sci. 'in press' https://doi.org/10.1071/AN21423

Litherland A.J., Dynes R.A., and Moss R. A. (2010) Proc. NZSAP, Palmerston North, New Zealand.

Ministry for the Environment (2020). New Zealand's Greenhouse Gas Inventory 1990-2018. Available at: https://environment.govt.nz/publications/new-zealands-greenhouse-gas-inventory-1990-2018/

Pinares-Patiño C.S., Hickey S.M., Young E.A., Dodds K.G., MacLean S. *et al.* (2013) Animal 7 (Suppl 2): 316–21 https://doi.org/10.1017/S1751731113000864

Pinares-Patiño C.S., McEwan J.C., Dodds K.G., Cárdenas E.A., Hegarty R.S. *et al.* (2011) Anim. Feed Sci. Technol. 166–167:210–18. https://doi.org/10.1016/j.anifeedsci.2011.04.068

Pinares-Patiño C.S., Ulyatt M.J., Lassey K.R., Barry T.N., and Holmes C.W. (2003) J. Ag. Sci. 140(2):205–14. https://doi.org/10.1017/S0021859603003046

42. Economic values for climate change responsive traits to develop a sustainable dairy cattle breeding goal in the tropics

C.K. Chepsiror[1], E.D. Ilatsia[2] and T.O. Okeno[1,3*]

[1]Animal Breeding & Genomics Group, Department of Animal Science, Egerton University, P.O Box, 536 Egerton, Njoro, Kenya; [2]Dairy Research Institute, Kenya Agricultural Research Organization (KALRO), P.O. Box 25, 20117, Naivasha, Kenya; [3]African Animal Breeding Academy, CoELIB, Centre, 18 Maziwa Road, Egerton University, Njoro, Kenya; tobias.okeno@egerton.ac.ke

Abstract

The objective of this study was to derive economic values for climate change responsive traits for inclusion in a dairy cattle breeding goal in the tropics. Using the Kenyan dairy cattle breeding program as the source of input parameters, a bio-economic model was developed to derive economic values for climate change responsive traits which included feed efficiency (FE), resilience (RE) and adaptability (AD) while economic values for disease resistance (DR) was derived using selection index methodology. Residual feed intake, variance deviations in milk yield, influence of increased temperature-humidity index on milk yield and brucellosis disease resistance were used as indicator traits for FE, RE, AD and DR, respectively. The economic values for FE, RE, AD and DR were; $ -0.14, 0.41, 1.55 and -0.58/cow/year, respectively. These economic values indicate that inclusion of climate change responsive traits in a breeding goal would lead to improved response to selection of dairy cattle in the tropics.

Introduction

Resource poor countries in the tropics are expected to bear the brunt of climate change effects, thus agricultural productivity of livestock systems will be negatively affected. This is because most agricultural activities in the tropics are rain-fed, hence livestock feed resources are expected to reduce as heat waves increase. Dairy cattle production costs are therefore expected to rise as farmers will have to spend more on feeds and cooling systems to manage increasing heat. Fortunately, these challenges could be overcome by breeding for dairy cattle which are adapted to the tropical production conditions as opposed to the current blind importation of genetic materials. This would require development of a robust dairy cattle breeding goal that accounts for both production, reproduction and climate change responsive traits which is currently lacking.

In the tropics, dairy cattle breeding has mainly focused on improvement of production, functional and reproductive traits (Kahi and Nitter, 2004), hence ignoring climate change responsive traits. With the effects of climate change, there is need for a paradigm shift to breed for dairy cattle that are not only productive but are feed efficient, resilient to dynamic environmental changes, adaptable and resistant to diseases. Breeding for such animals require inclusion of traits such as; feed efficiency, resilience, adaptability and disease resistance in the breeding goal. Feed efficiency is the relative ability of cows to turn feed nutrients into milk or milk components. Residual feed intake (RFI) which is the difference between actual feed intake and predicated feed intake on the basis of requirements and body weight maintenance has been used as an indicator for feed efficiency (Hietala *et al.,* 2014). Resilience (RE) is the capacity of the animal to be minimally affected by disturbances (Berghof *et al.,* 2019). Variance of deviations in milk yield provides an indication of the impact of a disturbance. Heat stress is the sum of external forces acting on an animal that causes an increase in body temperature. Temperature-Humidity Index (THI) is an indicator to assess high ambient temperature and humidity. Disease resistance is the ability of the host animal to exert control over a disease. Disease incidences result in high cost of production attributed to veterinary costs, low productivity and loss in products attributed to withdrawal periods when animals are under veterinary care.

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_42, © C.K. Chepsiror *et al.* 2022

Breeding for disease resistance, especially zoonotic disease such as Brucellosis which causes abortion at the end of gestation period, reduction in milk yield and infertility (Musallam *et al.,* 2019) would be a promising approach towards achieving good health for dairy cows and human consumers. Inclusion of these traits in a breeding goal require estimation of their economic values. Economic value is a unit change in the trait, while keeping all the other traits in the breeding goal constant (Kahi and Nitter, 2004). The objective of the current study was to derive economic values for climate change responsive traits for their inclusion in the breeding goal. Using the Kenyan dairy cattle breeding scheme, bio-economic modeling was used to derive the economic values feed efficiency, resilience and adaptability while selection index methodology was used to derive economic value for disease resistance.

Materials & methods

Procedure. Deterministic simulation approach was used to derive economic values for climate change responsive traits for their inclusion in the breeding goal of dairy cattle in the tropics. The current Kenya dairy cattle breeding goal which account for production, reproduction and milk quality traits was adopted.

Estimation of economic values. The bio-economic model was developed in a python computer program to derive economic values for feed efficiency (FE), resilience (RE) and adaptability (AD). On the other hand economic value for disease resistance (DR) was derived using selection index methodology (Hazel, 1943). All the economic values were in USD. The indicator traits for FE, RE, AD and DR were residual feed intake, variance deviation on milk yield, temperature-humidity index (THI) and brucellosis resistance, respectively. The input parameters on costs were obtained from the current dairy cattle production systems in Kenya. The main sources of revenues included sales of milk, in calf heifers, manure and culled cows and bull calves. The unit prices were adjusted for the current inflation rates. The economic values for FE, RE and AD were computed as:

$$EV = {}^{\Delta P}/_{\Delta T} \tag{1}$$

where, ΔP is change in profit after a unit increase in the trait of interest, while ΔT is marginal change in trait of interest after a unit increase.

Feed efficiency was measured as a function of Residual Feed Intake (RFI) according to Koch *et al.* (1963):

$$Y_i = \beta_0 + \beta_1 ADG + \beta_2 MWT + \varepsilon_i \tag{2}$$

where Y_i is daily feed intake (FI) of animal *i*, β_0 is regression intercept, β_1 *and* β_2 are partial regression coefficients of FI on average daily gain *ADG* and mid-weight *MWT*, ε_i is residual error of FI on animal *i*.

Resilience was measured using variance deviations on the lactation curve of 305-day milk yield according to Berghof *et al.* (2019):

$$\sigma^2 = \frac{x-\mu}{n} \tag{3}$$

where, σ^2 is the variance deviation in 305-day milk yield, $x - \mu$ is the deviation of daily milk yield from the overall mean and n is lactation length in days.

Adaptability was calculated using heat stress as an indicator and temperature-humidity index (THI) as the reference point. The equation by Hahn (1999) was used to calculate THI:

$$THI = 0.81.T + \left(\frac{RH}{100}\right).(T - 14.4) + 46.6 \quad (4)$$

where, T is the mean hourly ambient temperature (°C) and RH is mean hourly relative humidity (%).

The economic value for DR was computed using selection index methodology (Hazel, 1943). This method matches the breeding objective to expected responses in production traits and responses in these traits are maximized relative to overall gains. In the current study economic value for DR was estimated relative to milk yield (MY).

Since dairy farmers in Kenya and others in the tropics are risk averse due to fluctuation in weather conditions, outbreak of diseases and market prices, computation of risk rated economic values are most suitable. The risk rated economic values were computed as described by Kulak *et al.* (2003):

$$\text{REV} = \frac{\mu po\, f(\boldsymbol{g},\boldsymbol{e})}{\boldsymbol{e}\mu pi} - 0.5\ \lambda\ \text{E}\ \{\text{EE}_t - \text{E}\ (\text{EE}_t)^2\} \quad (5)$$

where, E (EE_t) is the expected values of economic efficiencies, λ the Arrow-Pratt coefficient of absolute risk aversion (0.02) and μ_{po} and μ_{pi} are the expected values of output and input prices, respectively, g, the vector of the variables determined by the genotype and e the vector of the variables associated with inputs.

Results

The bio-economic model developed and selection index methodology were able to derive risk rated economic values for climate change responsive traits. The findings indicated that RE and AD had positive economic values of $ 0.14 and $ 1.55/cow/year, respectively. On the other hand, FE and DR had negative corresponding economic values $ -0.14 and $ -0.58 per/cow/year (Table 1).

In this study, accounting for risks in the economic values did not yield errors larger than ±5% for any trait on the basis of evaluation. This indicates that both traditional and risk-rated models can be satisfactorily used to predict profitability and develop selection index criteria applicable in large-scale and smallholder dairy cattle breeding programmes in the tropics.

Discussion

The findings of the current study indicate that inclusion of climate change responsive traits in the current breeding goal would increase response to selection. This could be explained by the fact that the economic values for RE and AD were positive while those for FE and DR were negative (Table 1). The positive economic value for RE and AD indicate that selecting for these traits would result in dairy cattle with better adaptability and potential to produce milk even in the face of environmental fluctuations such as drought and heat stress. This in turn would result in high returns for the dairy farmers in the tropics where productivity of dairy cattle is dependent on weather fluctuations. Positive economic values for RE and AD have also been observed in previous studies of Bergohf *et al.* (2019). The negative economic values for FE

Table 1. Risk-rated economic values in US dollars (USD) for traits in the fixed herd size production system (Arrow Patt coefficient of absolute risk aversion (0.02) /cow/year.

Trait	Economic value
Resilience (RE)	0.14
Adaptability (AD)	1.55
Feed efficiency (FE)	-0.14
Disease resistance (DR)	-0.58

and DR in the current study should be good news to dairy farmers in the tropics. This is because some of the effects of climate change included reduced production of feed resources due to long drought and faster spread of livestock diseases due to migration of nomads with their livestock from one place to another. The negative economic value for FE indicate that selection for FE would result to dairy cattle which can efficiently utilize low quantities local feed resources and convert it to milk. This would reduce the cost of production as feed resources account for over 50% of total dairy cattle production costs in the tropics (Kahi and Nitter, 2004). The negative economic value for FE has also been obtained in the previous studies in the temperate countries Hietala *et al.* (2014). The negative economic value for DR was expected due to negative correlation between DR and MY which it was derived relative to. Similar findings were obtained with mastitis resistance in dairy cattle as reported by Sagwa *et al.* (2019). The negative economic value for DR indicate that selection for resistance to brucellosis would result in cows that are resistance to the disease. This could automatically, reduce production costs attributed to treatment, withdrawal and risk to human health.

Conclusions

Climate change responsive traits; feed efficiency, resilience, adaptability and disease resistance are important traits that should be included in the current dairy cattle breeding goal in the tropics. The positive economic values for resilience and adaptability and negative values for feed efficiency and disease resistance indicate feasibility of including these traits in the current dairy cattle breeding goal in the tropics. This would result in a robust dairy cattle breeding goal with higher returns to both large-scale and smallholder dairy farms in the tropics in the face of climate change.

References

Berghof T.V., Poppe M., and Mulder H.A. (2019) Front. genet. 9:692. https://doi.org/10.3389/fgene.2018.00692.

Hazel L. N. (1943) Genet. 28(6):476-490. https://pmc/articles/PMC1209225/pdf/476.pdf

Hietala P., Wolfová M., Wolf J., Kantanen J., and Juga J. (2014) J. dai. Sci. 97(2): 1092-1106. https://doi.org/10.3168/jds.2013-7085

Kahi A.K., and Nitter G. (2004) Livest. Prod. Sci. 88(1-2):161-177. https://doi.org/10.1016/j.livprodsci.2003.10.008

Kulak K., Wilton J., Fox G., and Dekkers J. (2003) Livest. Prod. Sci. 79(2-3):183 191. https://doi.org/10.1016/S0301-6226(02)00155-0

Musallam I., Ndour A. P., Yempabou D., Ngong C. A. C., Dzousse M. F. *et al.* (2019) Act. tropi. 197:105-042. https://doi.org/10.1016/j.actatropica.2019.105042

Sagwa C.B., Okeno T.O., and Kahi A.K. (2019) Anim. Sci. J. 49(6): 1148-1157. https://doi.org/10.4314/sajas.v49i6.18

43. Phenotypic and genetic trends in American Angus associated with climate variability

G. Rovere[1*], B.C.D. Cuyabano[2], B. Makanjuola[1], S. Kelly[3] and C. Gondro[1]

[1]*Department of Animal Science, Michigan State University, 474 S Shaw Ln, East Lansing, MI 48824, USA;* [2]*Université Paris Saclay, INRAE, AgroParisTech, GABI, Domaine de Vilvert, 78350 Jouy-en-Josas, France;* [3]*BEACON Center, East Lansing, MI 48824, USA;* roverega@msu.edu

Abstract

Climate change is expected to not only increase the frequency of extreme weather events, but to also increase the variability of climate conditions within geographical locations. Within the context of climate, livestock genetics research has mostly focused on G×E interactions and on selection for specific environments. This approach, however, does not address the sensitivity of animals to unstable environmental conditions, a likely scenario due to climate change. In this study, we explored the association between the variation in climate variables and the historical trends of growth traits in American Angus beef cattle. Analyses were performed using historical climate data, and phenotypic records of birth, weaning and post-weaning weights from American Angus beef cattle over the last three decades, from herds covering a wide area of the United States of America (USA). Results suggest a greater variation of both phenotypic and genetic trends in regions with high climate variability.

Introduction

One of the most pressing issues that humanity faces is the challenge of feeding a continuously growing population in an environment that is being disrupted by climate change. Due to the current concerns related to climate change there is an immediate need for breeding programs to reduce the environmental footprint of animal production (e.g. methane emission) and at the same time increase the resilience of animals to environmental changes. This is particularly relevant for beef cattle production systems which are largely extensive, and animals are constantly exposed to a broad range of environments (e.g. heat stress, which has a negative effect on commercial traits in beef cattle (Bradford *et al.*, 2016)). In terms of climate, livestock genetic research has mostly focused on estimation of G×E interactions (Fennewald *et al.*, 2018) and on selection for specific environments, viz recent studies that evidenced within-breed genetic diversity associated with different ecoregions in the USA that were defined based on the historical means of several climate variables and topography (Blackburn *et al.*, 2017; Braz *et al.*, 2021; Rowan *et al.*, 2021). However, one of the frequently noted outcomes of climate change is the increasing climate variability and frequency of extreme events (Trenberth *et al.*, 2015; Stott, 2016). It may be the case that subregions within these ecoregions suffer differently the effect of climate variability over the years. The objective of this study was to explore the association between the variability of climate conditions and the historical growth traits trends in American Angus beef cattle.

Materials & methods

Animal performance data. Phenotypic and pedigree information was provided by the American Angus Association. We performed the analysis on phenotypes recorded between 1990 and 2019 from 448 herds, all herds with at least 25 years of records and a minimum of 10 records per year-season (two seasons defined). Birth weight (BW), weaning weight (WW) and post-weaning weight (PWW) were recorded for a total of 1,754,310 Angus cattle.

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_43, © G. Rovere *et al.* 2022

Climate data. We used historical time-series climate data from 2,334 weather stations spread across the continental USA (Menne *et al.*, 2012). Eight variables[1] based on temperature and precipitation were selected from weather stations with at least 25 years of records between 1990 and 2019. A Principal Component Analysis (PCA) was carried out using the standard deviation and skewedness of each of the eight variables for each station during the period of interest. From this PCA, we used the squared cosine (COS2) of the components to calculate the distance of a given observation (station) from the center of the system. COS2 was then used as a proxy for each station's climate variability (ClimVar). Finally, we classified the weather stations into five ClimVar groups using the values of COS2 as: low [0-0.2], medium-low (0.2-0.4], medium (0.4-0.6], medium-high (0.6-0.8] and high (0.8-1]. Using GPS coordinates, the herds used in this study were associated to each weather station and their corresponding ClimVar group.

Genetic evaluation model. Variance components and estimated breeding values (EBV) were obtained using a single trait model $y = Xb + Za + e$, where y are the phenotype records (of BW, WW or PWW); b are the coefficients of the fixed effects (herd-year-season – 16,047 levels; sex – two levels, male or female; age of the dam at calving – four levels, 2yo, 3-5yo, 6yo, or 7-12yo; for WW and PWW the linear and quadratic effects of the age of the animal was included as well) and X is their design matrix; $a \sim N(0, A\sigma_a^2)$ are the breeding values and Z is their design matrix, with A as the numerator relationship matrix based on the pedigree; $e \sim N(0, I_n\sigma_e^2)$ are n random residuals. A was built tracing three generations back from the animal with records and comprised 2,208,105 animals. The analyses were implemented using airemlf90 which is part of the BLUPF90 family of programs (Misztal *et al.*, 2014).

Results

Figure 1 shows (A) the geographical location of the herds used in the analysis and their ClimVar grouping; and (B) the contribution of the skewness and standard deviation of each of the eight climate variables used to classify the herds into a ClimVar group. We noticed that herds located in the central-north region of the country were closer to locations with higher ClimVar during the analyzed period. These locations mostly correspond to the ecoregion denominated 'high plains' by Braz *et al.* (2021). Approximately 40% of the phenotypic records were from herds classified as medium to high ClimVar. The variables with a higher contribution in the PCA analysis were the standard deviations of the average and maximum temperature (sdTAVG and sdTMAX), and the variables related to precipitation, i.e. the variability of the total amount of precipitation per year (sdPRCP) and the number of days with ≥25.4 mm of rain in the year (sdDP1X). We also observed that the skewness of the average, maximum and minimum temperatures (skTAVG, skTMAX, skTMIN) were relevant, although mostly associated with the second component of the PCA analysis. Figure 2 shows the average phenotypic, genetic, and residual trends by year of birth, for herds located in the five ClimVar groups. We observed a decreasing trend for BW, while WW and PWW showed an increasing trend along the years. These trends are in line with the general emphasis given to selection traits in beef cattle breeding programs. Herds associated with high ClimVar showed a more irregular trend, especially for phenotypes.

A greater difference between the extreme ClimVar groups (low and high) was observed for phenotypic and genetic trends of BW, a rather unexpected result since BW is a phenotype modulated by the maternal environment and should be less directly influenced by changing weather conditions. The residual trends of the herds classified as high ClimVar were distinct for all three traits, however in the three traits residuals were more unstable for herds in the highest ClimVar locations.

[1] Average mean, maximum and minimum temperature in the year, total precipitation per year, number of days with maximum temperature ≥21.1 °C in the year, number of days with maximum temperature ≥32.2 °C in the year, number of days with ≥2.54 millimeters precipitation in the year, number of days with ≥25.4 millimeters precipitation in the year.

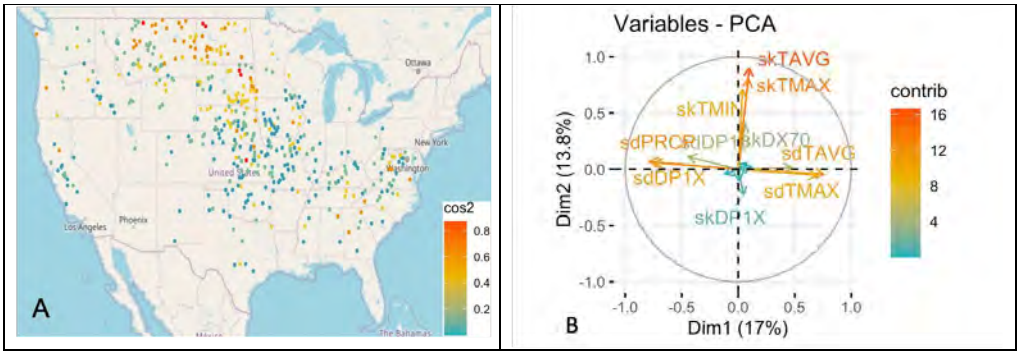


Figure 1. (A) Location of Angus herds classified by cosine square (COS2) distances; (B) contribution of the skewness and standard deviation of each climate variable used to build COS2.

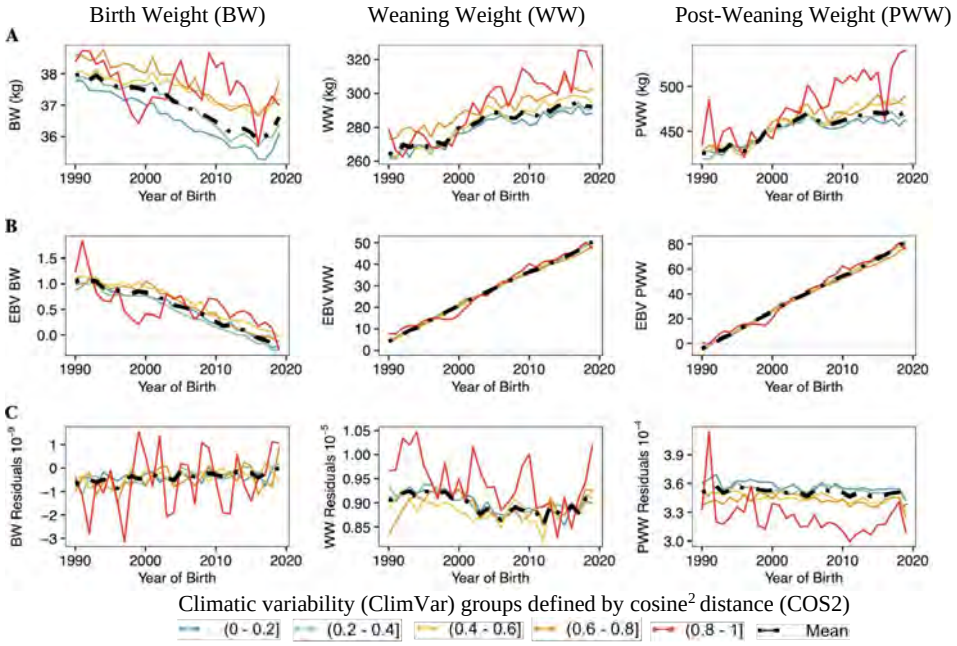


Figure 2. Average (A) phenotypic, (B) genetic, and (C) residual trends by birth year for herds located in the five ClimVar groups, for birth weight (BW), weaning weight (WW) and post-weaning weight (PWW).

Discussion

The USA uses ~45% of its area for agriculture and livestock production (The World Bank, 2021), which spans a large area with very different climate conditions. Previous studies suggested that the information from different ecoregions based on historical climate data averages can be useful for beef cattle production (e.g. Blackburn *et al.*, 2017; Braz *et al.*, 2021). Here, instead of considering these different ecoregions, we considered the climate variability over the last 30 years to geographically distinguish herds by classifying them in five climate variability (ClimVar) groups. High ClimVar was detected in several regions of the USA. Our objective was to examine the trends of Angus beef cattle growth traits in herds with 25 to 30 years of records when the environmental conditions are unstable (i.e. high ClimVar). These trends showed higher

variability across the years when compared to herds in more stable weather conditions, although it should be noted that a smaller number of herds were in these high ClimVar locations. The phenotypic trajectory of the higher ClimVar groups were above the mean trajectory, but it was less obvious for the genetic trends. Then, it is possible that the unstable climate condition biased mostly to beneficial conditions or that the management system adopted effectively neutralized the climate variability. The model used to obtain EBVs included the environmental effect of the contemporary group, however ClimVar groups showed different patterns indicating that more environmental effects could be accounted for in the model. These results reinforce the idea that the inclusion of climate variables and the development of novel methodologies can be of value for traditional genetic and phenotypic prediction models. Additionally, analyses that account for climate variability may be an alternative method to evaluate resilience to environmental changes instead of the use of specific environmental stressors.

References

Bertrand, J.K., J.D. Hough, and L.L. Benyshek (1987) J. Anim. Sci. 64:77–82. https://doi.org/10.2527/jas1987.64177x.

Blackburn, H.D., B. Krehbiel, S.A. Ericsson, C. Wilson, A.R. Caetano *et al.* (2017) PLoS One 12(5):e0176474. https://doi.org/10.1371/ journal.pone.0176474

Bradford, H.L., B.O. Fragomeni, J.K. Bertrand, D.A.L. Lourenco, and I. Misztal (2016) J. Anim. Sci. 94:4143–4150. https://doi.org/10.2527/jas.2016-0707.

Braz, C.U., T.N. Rowan, R.D. Schnabel, and J.E. Decker (2021) Sci. Rep. 11:13335. https://doi.org/10.1038/s41598-021-92455-x.

Fennewald, D.J., R.L. Weaber, and W.R. Lamberson (2018) J. Anim. Sci. 96:422–429. https://doi.org/10.1093/jas/skx080.

Misztal, I., S. Tsuruta, D.A.L. Lourenco, Y. Masuda, I. Aguilar *et al.* (2014) Available at: http://nce.ads.uga.edu/wiki/lib/exe/fetch.php?media=blupf90_all7.pdf

Stott, P. 2016. Science 352:1517–1518. https://doi.org/10.1126/science.aaf7271.

Trenberth, K.E., J.T. Fasullo, and T.G. Shepherd (2015) Nat. Clim. Chang. 5:725–730. https://doi.org/10.1038/nclimate2657.

The World Bank (2021) Available at: https://data.worldbank.org/indicator/.

44. Genetic variation in resilience to climate effects on beef carcass traits

H. Bunning* and E. Wall

SRUC, Edinburgh, EH9 3JG, United Kingdom; harriet.bunning@sruc.ac.uk

Abstract

Breeding climate resilient cattle is an important strategy to mitigate the effects of climate change. Resilience of beef cattle is less well understood compared to dairy, especially in temperate climates. This study aims to compare resilience to climate, both average weather and extreme weather events, for carcass weight, age at slaughter and carcass growth rate. Linear models were used to analyse these traits in over 1.7 million records from UK abattoirs. Resilience was estimated using regression coefficients for interactions between breed type and weather parameters. Results show variation between breeds for resilience to climate with British breeds showing the greatest resilience. Our results suggest an increase in heat waves by 1 per 100 days of life would reduce cold carcass weights of dairy and continental animals by about 150 g but increase that of British breed animals by about 360 g.

Introduction

Climate change is a challenge facing livestock production globally and mitigation strategies are needed to reduce its impact (Rojas-Downing *et al.*, 2017). The impact of climate change is already seen in tropical climates, but we also expect a significant impact in temperate regions (Wreford and Topp, 2020). Climate resilience is the capacity of the animal to be minimally affected by changes in the climate and there is good evidence of genetic variation for this trait (Doeschl-Wilson *et al.*, 2019).

Between-breed variation in climate resilience has been shown for a range of weather parameters and traits in cattle. Variation in resilience to heat stress is well documented, particularly in the tropics between tropically adapted *Bos indicus* breeds and less adapted *Bos taurus* breeds. For example, extreme heat led to reductions in dry matter intake in taurine Angus cattle, but no significant change in indicine Brahman cattle (Gaughan *et al.*, 2010). In temperate climates, studies have shown variation in the effect of climate on dairy cattle traits, where high producing dairy breeds, particularly Holsteins tend to be less resilience to heat stress than other more robust breeds, such as Simmentals (Gantner *et al.*, 2017) or Jerseys (Bryant *et al.*, 2007).

Climate resilience for beef traits, particularly in temperate systems, is less well understood. This study aims to compare how carcass traits are affected by weather across beef and dairy breeds. For traits expressed repeatedly over time, the effect of weather can be seen within an individual, allowing estimation of individual lifetime resilience by, for example, fitting reaction norm functions (Sánchez-Molano *et al.*, 2019). However, carcass traits are typically measured at the end of life so this is not possible. We also investigate grouping offspring by sire to produce a slope for each sire, which could be used as phenotypes in genetic analysis.

Materials & methods

We analysed data of over 1.7 million slaughter records (2000-2019) from UK abattoir companies (Pritchard *et al.*, 2021). Traits included cold carcass weight (CCW) and age at slaughter (AAS). A measure of daily carcass growth rate was calculated by dividing CCW by AAS (CGR). Breeds were grouped into 3 types: British beef, continental beef and dairy. Other data about animals were extracted to use as factors and covariates in the model. Sex was defined using data from the abattoir as castrated male, female or entire male. We also included age of the dam and the proportion of dairy breed in the dam's pedigree (Pritchard

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_44, © H. Bunning and E. Wall 2022

et al., 2021). We accounted for varying management practices, which might be regionally distributed and linked with weather, by including two contemporary groups in our model: Birth herd-year-season (BHYS) and Finishing (FHYS). Only animals in groups containing more than 4 were included. Location of death (i.e. abattoir) was also included.

We used weather data from the Met Office HadGrid-UK database, a data set of gridded climate variables derived from the network of UK land surface observations (Perry and Hollis, 2004). This was combined with animal locations and dates of stay, allowing the calculation of the average daily maximum temperature (Tmax), minimum temperature (Tmin) and precipitation (Rain) for the lifetime of each animal. Daily weather was also used to define the occurrence of extreme weather events, including heat waves (at least 3 days maximum temperature over threshold of 28, 27, 26 or 25 °C, depending on region), coldwaves (at least 3 days maximum temperature under 0 °C), dry days (less than 0.12 mm rainfall) and wet days (more than 7.65 mm rainfall). The frequency of each extreme weather event for an animal was calculated by dividing by AAS.

Analyses were carried out using linear fixed effect models using AS-REML (Butler *et al.*, 2018) and R. Two models were produced for each trait to assess the resilience of each breed group: (1) to average weather; and (2) to frequency of extreme weather events. Interaction effects between breed group and the weather parameter were fitted. For CCW, AAS was included as a covariate and for AAS, CCW was included as a covariate. We expected interactions between weather to be important so an interaction between Tmax and Tmin and another between Tmin and Rain were included in the average weather models. The generalised model was therefore:

$$\textit{Trait} \sim \textit{weather parameters} \times \textit{breed group} + \textit{other traits} + \text{sex} + \text{breed} + \text{BHYS} + \text{FHYS} + \text{death location} + \text{dam age} + \text{dam \% dairy} \quad (1)$$

To compare the genetic potential for resilience of different sires, we selected Limousin animals with a sire with more than 4 offspring. After edits, the data consisted of 1,945 animals with 237 sires. 94% of records were lost for lack of sire information and small contemporary group sizes. We fitted an interaction between sire and number of heat wave days in our model for CCW. The solutions give an indication of the genetic value for climate resilience of each sire, where sires with values closer to zero can be considered more resilient.

Results

Average weather. CCW increased for animals which experienced higher Tmin and Tmax and lower Rain. For CCW, continental breeds were least affected by changes in Tmin but most affected by changes in Tmax and Rain (Table 1). Animals which experienced higher Tmin and Rain but lower Tmax were younger at slaughter. For AAS, the most affected breed group varied across the different weather parameters. British breeds were the most affected by changes in Tmin, dairy breeds the most affected by Tmax and continental breeds the most affected by Rain. Animals which experienced higher Tmin, lower Tmax and lower Rain tended to have a greater CGR. For this trait, dairy cattle were affected more by changes in all three parameters (Tmin, Tmax and Rain) than the other breed groups.

Extreme weather frequency. CCW tended to be greater for animals which experienced lower extreme weather frequencies, although there were some differences between breed groups (Table 2). An increased frequency of heat waves had negative effects on the CCW of both continental and dairy breeds, but a positive effect in British breeds. Conversely, an increased frequency of extreme dry days had positive effects on both continental and dairy but a negative effect on British breeds. An increased frequency of extreme wet days had negative impacts on CCW of all breed types, but continental and dairy breeds were significantly more affected than British breeds. Animals which experienced more frequent extreme weather tended to be older

Table 1. Regression coefficients (and standard errors) for each level of breed group with lifetime average daily weather parameters.[1]

		British	Continental	Dairy
CCW (kg)	Tmin	2.44 (0.37)^{ab}	2.22 (0.37)^{a}	2.99 (0.37)^{b}
	Tmax	1.16 (0.16)^{ab}	1.46 (0.15)^{a}	1.02 (0.16)^{b}
	Rain	-1.14 (0.45)^{a}	-2.26 (0.44)^{b}	-0.90 (0.45)^{a}
AAS (days)	Tmin	-2.59 (0.54)^{a}	-2.04 (0.53)^{ab}	-1.19 (0.54)^{b}
	Tmax	10.18 (0.23)^{ab}	9.94 (0.22)^{a}	10.50 (0.23)^{b}
	Rain	-19.56 (0.65)^{a}	-21.19 (0.64)^{b}	-19.15 (0.65)^{a}
CGR (kg/day)	Tmin	0.000944 (0.00067)^{a}	0.000169 (0.00066)^{a}	0.00242 (0.00067)^{b}
	Tmax	-0.00549 (0.00028)^{ab}	-0.00499 (0.00027)^{a}	-0.00577 (0.00029)^{b}
	Rain	0.00819 (0.00080)^{a}	0.00732 (0.00079)^{a}	0.0100 (0.00081)^{b}

[1] Values with difference letters within a row are significantly different from one another ($P<0.05$).

Table 2. Regression coefficients (and standard errors) for each level of breed group with frequency of extreme weather days.[1]

		British	Continental	Dairy
CCW (kg)	Heat wave	36.48 (9.52)^{a}	-14.95 (8.402)^{b}	-15.47 (10.48)^{b}
	Dry	-2.585 (2.152)^{a}	25.86 (1.945)^{b}	12.52 (2.293)^{c}
	Wet	-15 (1.369)^{a}	-26.19 (1.159)^{b}	-25.62 (1.402)^{b}
AAS (days)	Heat wave	38.3 (13.8)^{a}	54.3 (12.2)^{a}	172.2 (15.2)^{b}
	Dry	133.0 (3.1)^{a}	142.6 (2.8)^{b}	108.9 (3.3)^{c}
	Wet	20.6 (2.0)^{a}	35.5 (1.7)^{b}	14.5 (2.0)^{c}
CGR (kg/day)	Heat wave	-0.0019 (0.0171)^{a}	-0.0974 (0.0151)^{b}	-0.0728 (0.0188)^{b}
	Dry	-0.0647 (0.0039)^{a}	-0.0348 (0.0035)^{b}	-0.0290 (0.0041)^{b}
	Wet	-0.0225 (0.0025)^{a}	-0.0339 (0.0021)^{b}	-0.0346 (0.0025)^{b}

[1] Values with difference letters within a row are significantly different from one another ($P<0.05$).

at slaughter, although British breeds were less affected by heat waves and wet days, whereas dairy breeds were least affected by dry days. For CGR, increased growth rates were seen when animals experienced lower frequencies of extreme weather across breed types. British breeds were less affected by heat waves and wet days, but more affected by dry days than continental and dairy breeds.

Sire resilience. Results showed that the interaction between sire and number of heat wave days had a significant effect on CCW. The solutions varied between -18.24 (s.e. 14.70) and +12.67 (s.e. 9.01) although the standard errors were large. A sample of slopes for three sires are shown in Figure 1. Here, results suggest that the CCWs of offspring of sires A and B are more affected (although in opposite directions) than sire C. Sire C has a higher genetic value for resilience as the slope is closer to zero.

Discussion

Results show variation between breeds for resilience to climate for CCW, AAS and CGR, both for average weather across the lifetime of the animal and frequency of extreme weather events. We might expect the British breeds to show more resilience as they are thought to be more robust and typically less strongly selected for production traits, which may lead to higher resilience. There is evidence in dairy cattle that

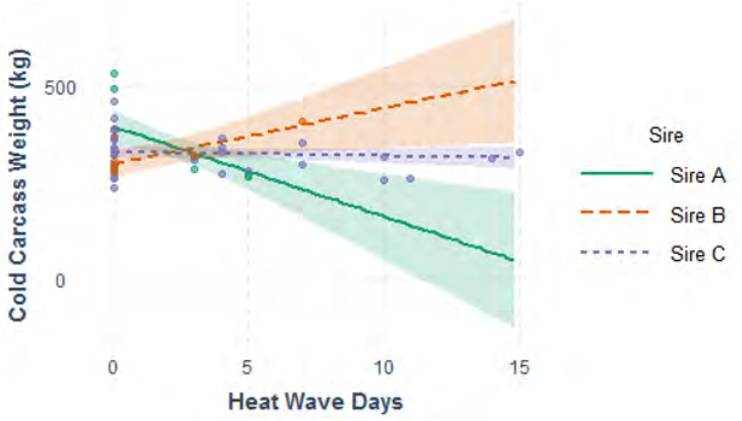


Figure 1. The effect of heat wave days on CCW for offspring of three sires. Solutions for the interaction term are used to produce the slopes of the lines with 95% confidence intervals illustrated by the shaded regions.

animals with lower production levels may be more resilient (Gantner *et al.*, 2017). Our results broadly follow this expectation, with British breeds coefficients being closer to 0. In particular, the positive correlation between heat wave frequency and CCW in British breeds suggests that these breeds may be less affected by heat stress compared to continental and dairy breeds. Results suggest an increase in heat waves by 1 per 100 days of life will reduce CCWs of dairy and continental animals by about 150 g but increase that of British breed animals by about 360 g.

In addition to physiological differences between these breeds, its important to consider that these differences may also reflect varying management strategies. For example, the positive correlation seen between dry day frequency and CCW for dairy and continental breeds (rather than the negative seen for British breeds), may show that dairy animals are less likely to have a forage-based diet, rather than having a genetic resilience. To account for this, we can compare offspring of sires within a breed. Our preliminary results suggest there is variation between the genetic value for resilience of sires. However more offspring per sire may be needed to test if resilience is significantly different between sires within and across breeds.

References

Bryant J.R., López-Villalobos N., Pryce J.E., Holmes C.W. and Johnson D.L. (2007) New Zealand Journal of Agricultural Research 50:327–338.

Butler D.G., Cullis B.R., Gilmour A.R., Gogel B.J. and Thompson R. (2018) ASReml-R Reference Manual Version 4.

Doeschl-Wilson A.B., Song J., Dekkers J., Berghof T.V.L., Poppe M. *et al.* (2019)

Frontiers in Genetics 9:692.

Gantner V., Bobic T., Gantner R., Gregic M., Kuterovac K., *et al.* (2017) International Journal of Biometeorology 61(9):1675–1685.

Gaughan J.B., Mader T.L., Holt S.M., Sullivan M.L. and Hahn G.L. (2010) International Journal of Biometeorology 54:617–627.

Perry M. and Hollis D. (2004) The generation of monthly gridded datasets for a range of climatic variables over the United Kingdom. Exeter, UK.

Pritchard T.C., Wall E. and Coffey M.P. (2021) Animal 15:100090.

Rojas-Downing M.M., Nejadhashemi A.P., Harrigan T. and Woznicki S.A. (2017) Climate Risk Management 16:145–163.

Sánchez-Molano E., Kapsona V.V., Ilska J.J., Desire S., Conington J., *et al.* (2019) BMC genetics 20:84.

Wreford A. and Topp C.F.E. (2020) Agricultural Systems 178:102737.

45. Genetic parameters for rumination time and traits related to sustainable dairy production

L.S.F. Lopes[1*], K. Houlahan[1], L. Alcantara[1], G.A. Oliveira Jr.[1], D. Tulpan[1], F. Miglior[1,2], F.S. Schenkel[1] and C.F. Baes[1,3]

[1]Centre for Genetic Improvement of Livestock, Department of Animal Biosciences, University of Guelph, 50 Stone Rd E, N1G 2W1, Guelph, Ontario, Canada; [2]Lactanet Canada, Guelph, Ontario, Canada, N1K 1E5; [3]Institute of Genetics, Vetsuisse Faculty, University of Bern, Bern, Switzerland; llopes@uoguelph.ca

Abstract

Rumination time was assessed as an indicator trait to improve sustainability in dairy cattle. This study contained 6,203 records on 611 mid-first lactation Canadian Holstein cows. Animal models for rumination time and methane emissions, and repeatability animal models for milk yield and dry matter intake were utilized in bivariate analyses to estimate genetic parameters including heritability, genetic and phenotypic correlations between traits. Results showed that rumination time had a moderate heritability (0.41±0.15) and genetic correlations of -0.37 (±0.30) with methane emissions, 0.19 (±0.15) with dry matter intake, and 0.51 (±0.13) with milk yield. These findings provide the basis for future work investigating the use of rumination time as an additional indicator trait for selecting for more sustainable dairy production.

Introduction

As the global population increases, so does the demand for sustainably produced dairy products. One main focus to improve the sustainability of the dairy industry has been to enhance feed efficiency (FE) and reduce methane (CH_4) emissions of dairy cattle (Hailemariam *et al.*, 2020). In addition to global warming concerns, methane emission is also responsible for 2 to 12% of the dietary energy losses in dairy (González-Recio *et al.*, 2020). Since efficient cows eat less while producing the same amount of milk as their counterparts (Houlahan *et al.*, 2021), elevated levels of ruminal CH_4 represent animals of lower sustainability and efficiency. Therefore, improving FE and reducing CH_4 emissions while increasing milk production levels are key for an efficient and sustainable dairy industry worldwide (Richardson *et al.*, 2020). However, it is difficult and expensive to measure phenotypes for FE and CH_4, making it important to study indicator traits. Rumination time (RT) has been suggested as a potential indicator for traits such as CH_4 yield and dry matter intake (DMI) (Schirmann *et al.*, 2009). One advantage of RT is that it can be automatically recorded on a large-scale, in commercial dairy herds, by sensor-based systems (Bikker *et al.*, 2014). However, the use of RT as an indicator trait for FE and CH_4, as well as genetic parameters for RT have not yet been fully explored. Therefore, the objective of this study was to evaluate RT as an indicator trait for efficient dairy cows, by estimating genetic parameters, phenotypic and genetic correlations between RT, DMI, CH_4, and MY in Canadian Holstein cows.

Materials & methods

Data was provided by the Efficient Dairy Genome Project (https://genomedairy.ualberta.ca) and the Resilient Dairy Genome Project (http://www.resilientdairy.ca). Data comprised of 6,203 records for MY (kg), CH_4 (g), DMI (kg), and RT (min) on 611 first lactation Holstein cows between 110 and 210 days in milk from March 2017 to December 2021. Milk samples were taken weekly and analysed by Lactanet Canada (Guelph, ON). Daily CH_4 measurements were obtained using the GreenFeed® System (C-Lock Inc., Rapid City, SD, USA). This system consists of a mobile feeding station that uses airflow, integration of gas concentrations measurements, temperature, humidity, and recognition of muzzle position to allow for the precise measurement of CH_4 (Huhtanen *et al.*, 2015). Daily RT was obtained from Heatime® neck collars, (SCR Hi Tag, Allflex, Netanya, Israel). Sensors in these collars detect and process the sound of rumination and regurgitation every two hours. Finally, daily records for DMI were obtained as described in Seymour *et al.* (2020).

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_45, © L.S.F. Lopes *et al.* 2022

Data quality control. Rumination time with fewer than 12 measurements within a day, and total daily RT equal to zero were removed. Traits measured daily (DMI, RT, and CH_4) needed a minimum of two records per week of lactation to be considered for analysis, where the week of lactation was defined using the week of MY test day. Traits were averaged to obtain weekly average measures for DMI, RT, and CH_4. All traits with values greater than three standard deviations from the mean were removed.

Statistical models. Repeatability animal models were fit for MY and DMI, and animal models were fit for CH_4 and RT in bivariate analyses to estimate the genetic parameters. All (co)variance components were estimated using the average information residual maximum likelihood algorithm in the ASREML 4.0 software (Gilmour *et al.*, 2015). Heritabilities were the average of the results of all bivariate combinations for each trait. In general, the models used in this study can be described as:

$$Y_{ijklm} = \mu + CM_i + WOL_j + YS_k\ a_l + pe_l + e_{ijklm}$$

where Y_{ijklm} represents the m^{th} phenotype (for MY and DMI) of the l^{th} animal; μ is the overall mean of the trait; CM_i is the fixed effect of the i^{th} age at calving class months (eight classes); WOL_j is the fixed effect of the j^{th} week of lactation (thirteen levels); YS_k is the fixed effect of the k^{th} year and season of calving class (sixteen classes); a_l is the random additive genetic effect of the l^{th} cow; pe_l is the random permanent environmental effect (for MY and DMI) of the l^{th} cow, and e_{ijklm} is the random residual error term.

Results

Descriptive statistics of the data are presented in Table 1. The observed phenotypic ranges were in line with current literature for first lactation cows at similar lactation stages (Hare *et al.*, 2018; López *et al.*, 2015; Zetouni *et al.*, 2018).

The estimated heritabilities, repeatabilities, phenotypic and genetic correlations for all traits are shown in Table 2. Except for DMI (0.29±0.07), all traits had estimated heritabilities slightly above 0.40, MY and DMI had repeatabilities of 0.79 (±0.01) and 0.55 (±0.02), respectively. In general, RT was genetically positively correlated to MY (0.51±0.13) and DMI (0.19±0.15), while being negatively correlated to CH_4 (-0.37±0.30).

Table 1. Number of records, number of animals with records, average number of weeks with records (Weeks), mean, standard deviation (SD) and coefficient of variation (CV) for milk yield (MY), methane yield (CH_4), dry matter intake (DMI) and rumination time (RT).

Trait	Records	Animals	Weeks	Mean (SD)	Min	Max	CV (%)
MY (kg)	5,734	609	10.69	31.65 (4.63)	17.50	45.50	14.65
CH_4 (g)	409	409	1.00	494.40 (84.02)	263.80	719.30	16.99
DMI (kg)	4,308	446	8.78	20.63 (2.95)	8.13	30.14	14.31
RT (min)	328	328	1.00	445.70 (51.33)	273.60	580.80	11.51

Table 2. Genetic correlation (above diagonal), heritability (diagonal), phenotypic correlation (below diagonal), and repeatability (t) for milk yield (MY), methane yield (CH_4), dry matter intake (DMI), and rumination time (RT).

Trait	MY (SE)	CH_4 (SE)	DMI (SE)	RT (SE)	t (SE)
MY	**0.44 (0.07)**	0.32 (0.12)	0.46 (0.16)	0.51 (0.13)	0.79 (0.01)
CH_4	0.16 (0.06)	**0.43 (0.12)**	0.81 (0.10)	-0.37 (0.30)	NA
DMI	0.31 (0.03)	0.50 (0.05)	**0.29 (0.07)**	0.19 (0.15)	0.55 (0.02)
RT	0.36 (0.06)	-0.10 (0.06)	0.23 (0.07)	**0.41 (0.15)**	NA

However, with the large SE for the genetic correlations, only the correlation with MY was statistically significant.

Discussion

Heritability estimates from this study were in line with those previously reported (Byskov *et al.*, 2017; Moretti *et al.*, 2018). Positive genetic correlations between DMI, MY, and RT, and the negative genetic correlation between CH_4 and RT suggest that selecting for RT could lead to higher milk production and lower emissions for the animals, even if animals of higher intake were obtained in the process. Nonetheless, it is important to note the high standard errors of all genetic correlations, given the dataset size available and the fact that a single herd provided all the data. Previous works (Marino *et al.*, 2021; Zetouni *et al.*, 2018) have shown that RT and CH_4 may not be directly related on a genetic level. However, the significant genetic correlation between MY and RT, and the point estimate between DMI and RT, suggest direct relationships between rumination time, milk production, and a possible relationship between intake and rumination time. Additionally, the moderate negative point estimate for the genetic correlation between RT and CH_4 may indicate a degree of association between RT and CH_4 emission, which would potentially allow for indirect selection for lower emissions. Rumination is a complex trait, mediated by several aspects of the life of the animal, and acquiring more RT data for short and specific periods of the lactation may not be enough to evaluate its relationship with production and efficiency traits since RT may vary according to the stage of lactation (Moretti *et al.*, 2018; Zetouni *et al.*, 2018). Ideally, the evaluation of sensor-based RT as an indicator trait for FE should be carried over for the whole lactation instead. Thus, to consider RT as an indicator trait for FE or CH_4, random regression models for RT over the whole lactation, such as realized for feed efficiency by Houlahan (2021), could be considered. Accounting for other sources of variation over time, and additional data collection for all the involved traits is essential for accurately determining the potential of RT to be used as an indicator trait in sustainable breeding programs.

Conclusions

Rumination time had a moderate heritability and was found genetically associated with MY. No significant relationship between RT, CH4, and DMI was found. A larger dataset is needed to determine whether RT might have the potential to be used in selection programs.

Acknowledgements

The authors acknowledge partners and collaborators of the efficient dairy genome project (https://genomedairy.ualberta.ca/) and the resilient dairy genome project (http://www.resilientdairy.ca).

References

Bikker, J. P., van Laar, H., Rump, P., Doorenbos, J., van Meurs, K., *et al.* (2014). J. Dairy Sci. 97(5):2974–2979. https://doi.org/10.3168/jds.2013-7560.

Byskov, M. V, Fogh, A., and Løvendahl, P. (2017). J. Dairy Sci. 100(12):9635–9642. https://doi.org/10.3168/jds.2016-12511.

Gilmour, A. R., Gogel, B. J., Cullis, B. R., Welham, S. J., Thompson, R. (2015). ASReml user guide release 4.1 functional specification. Hemel Hempstead, UK: VSN International Ltd. Available at: https://www.vsni.co.uk

González-Recio, O., López-Paredes, J., Ouatahar, L., Charfeddine, N., Ugarte, E., *et al.* (2020). J. Dairy Sci. 103(8):7210–7221. https://doi.org/10.3168/jds.2019-17598.

Hailemariam, D., Manafiazar, G., Basarab, J., Stothard, P., Miglior, *et al.* (2020). Can. J. Anim. Sci. 101(1):177–190. https://doi.org/10.1139/cjas-2019-0085.

Houlahan, K., Schenkel, F. S., Hailemariam, D., Lassen, J., Kargo, M., *et al.* (2021). Animals 11(4):1157. https://doi.org/10.3390/ani11041157.

Houlahan K. (2021). Ph.D. Thesis. University of Guelph. https://atrium.lib.uoguelph.ca/xmlui/handle/10214/26420.

López, S., France, J., Odongo, N., McBride, R., Kebreab, E., *et al.* (2015). J. Dairy Sci. 98(4):2701–2712. https://doi.org/10.3168/jds.2014-8132.

Marino, R., Petrera, F., Speroni, M., Rutigliano, T., Galli, A., *et al.* (2021). Animals. 11(6):1583. https://doi.org/10.3390/ani11061583.

Moretti, R., de Rezende, M. P. G., Biffani, S., and Bozzi, R. (2018). J. Anim. Breed. Genet. 135(4):293–299. https://doi.org/10.1111/jbg.12346.

Richardson, C. M., Baes, C. F., Amer, P. R., Quinton, C., Martin, P., *et al.* (2020). Animal 14(1):171–179. https://doi.org/10.1017/S175173111900154X.

Schirmann, K., Keyserlingk, M. A. G. Von, Weary, D. M., Veira, D. M., and Heuwieser, W. (2009). Technical note: Validation of a system for monitoring rumination in dairy cows. J. Dairy Sci. 92(12):6052–6055. https://doi.org/10.3168/jds.2009-2361.

Seymour, D. J., Cánovas, A., Chud, T. C. S., Cant, J. P., Osborne, V. R., *et al.* (2020). J. Dairy Sci. 103(2):1528-1540. https://doi.org/10.3168/jds.2019-17414.

Zetouni, L., Difford, G. F., Lassen, J., Byskov, M. V., Norberg, E., *et al.* (2018). J. Dairy Sci. 101(12):11074–11085. https://doi.org/10.3168/jds.2017-14280

46. Genotype by temperature-humidity index interactions on milk production and udder health traits in the Montbeliarde cows

A. Vinet[1*], S. Mattalia[1,2], R. Vallée[1,2], C. Bertrand[3], P. Bertuzzi[4], B.C.D. Cuyabano[1] and D. Boichard[1]

[1]Université Paris Saclay, INRAE, AgroParisTech, UMR1313 GABI, Domaine de Vilvert, 78350 Jouy-en-Josas, France; [2]Institut de l'Elevage, UMT eBIS, 75012 Paris, France; [3]INRAE, US310 CTIG, Domaine de Vilvert, 78350 Jouy-en-Josas, France; [4]INRAE, US1116 AGROCLIM, Domaine de Saint Paul, 84914 Avignon, France; aurelie.vinet@inrae.fr

Abstract

This study aimed to estimate the effect of temperature-humidity index (THI) and the magnitude of genotype-by-THI (G×THI) interactions on milk production traits and on somatic cell score in Montbéliarde cows. Test-day records from first and second lactations from 2016 to 2020 were associated with the average THI of the three days before the test-day record. A total of 301,078 test-day records from 39,469 cows in first lactation and 292,648 test-day records from 37,593 cows in second lactation were analyzed. The average THI effect, estimated with an animal mixed model, indicated that the optimal THI was below 55 (i.e. ~12-13 °C). Individual responses to THI were estimated simultaneously with the time-dependent response by random regression models. Regardless of the stage of lactation, genetic correlations along the THI gradient were >0.80 and most of them were >0.90, suggesting that G×THI interactions were weak for production and for udder health traits.

Introduction

Heat stress negatively influences cattle welfare, health and productivity (Lees *et al.*, 2019). To cope with global warming and forecasted increases in temperature and heat waves frequency, identifying high-producing animals tolerant to heat is of capital importance to satisfy the need for animal-sourced foods. Several measures exist to assess heat stress, one of them consists in determining the meteorological conditions that lead to a decrease in performance. The most widespread indicator for heat stress in cattle is a temperature-humidy index (THI) combining the effects of temperature and humidity. The first objective of the present study is to estimate the critical temperature marking the onset of heat stress through the decline in performance and to quantify the amplitude of this decline in the range of THI currently encountered in the French Montbeliarde breeding regions. The second objective is to perform genotype-by-environment studies to find out if there is individual variability in heat tolerance, and therefore if some animals may be better adapted to produce despite warm temperatures.

Materials & methods

Performance and meteorological data. Performances, along with pedigree information, were available for this study from the French national database. Data were selected from first (L1) and second (L2) lactation test-day records of Montbéliarde cows from five French regions (Franche-Comté, Rhône-Alpes, Auvergne, Bretagne and Pays-de-Loire), gathered between August 2016 and December 2020. The phenotypes analyzed were milk yield (MY), fat content (FC), protein content (PC) and somatic cell score (SCS) obtained as a log-transformation of somatic cell concentration ($SCS=\log_2(SCC/100{,}000)+3$). Only animals with known parents and age at calving between 23-42 months and 35-60 months for first and second parity, respectively, were considered. Lactation length was limited to 305 days. Finally, a total of 301,078 test-day performances from 39,469 cows (94,476 animals in pedigree files) in L1 and 292,648 performances from 37,593 cows (98,015 animals in pedigree files) in L2 from herds were randomly selected in the whole population.

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_46, © A. Vinet *et al.* 2022

Meteorological data were provided by the Safran database from Météo-France. This database contains daily estimated temperature and humidity values on a grid of 9,892 squares of 8×8 km.

Performance and meteorological databases were merged according to the zip code of each farm, and test-day records were associated with the average THI of the three days before the test-day. The daily THI was calculated (1.8×T+32)-(0.55-0.0055×RH)×(1.8×T-26), with T being the 24h-average temperature (Celsius), and RH the 24h-average relative humidity (%).

Estimation of THI effect on phenotypic performances. The model used to estimate the average effect of THI on MY, FC, PC, and SCS was:

$$y = X\beta + Z_1 a + Z_2 p + e \quad (1)$$

where y, β, a, p, and e are the vectors of recorded phenotypes, fixed effects, additive genetic effects, permanent environmental (pe) effects and the residuals of the model, respectively. X, Z_1 and Z_2 are the incidence matrices for β, a and p respectively. For each trait, the fixed effects considered were contemporary group (herd-year combination), days in milk (DIM), age at calving (in months), days carried calf (days from successful insemination until test-day, converted to months), month of calving (12 classes), and THI by classes.

To study the G×THI interactions, the random regression model used was:

$$\mathbf{y} = \mathbf{X\beta} + \sum_{r=0}^{2} \mathbf{Z_r}(d)\, \mathbf{a_r} + \sum_{s=1}^{3} \mathbf{Z_s}(t)\, \mathbf{a_s} + \sum_{m=0}^{2} \mathbf{Z_m}(d)\, \mathbf{p_m} + \sum_{n=1}^{3} \mathbf{Z_n}(t)\, \mathbf{p_n} + \mathbf{e} \quad (2)$$

where y, β, and e are the same as described for model (1). X is the incidence matrices for β. The contemporary group was defined by herd-test-day with at least three cows. The other fixed effects considered were as for model (1) excluding THI. The effect of THI was considered using polynomials along with DIM for the genetic and pe effects. Thus, for the genetic effect, Z_r and Z_s are vectors of covariates of the r^{th} and s^{th} polynomials, respectively, evaluated at DIM d or THI t; a_s and a_r are the genetic random regression coefficients. For the pe effect, the model is the same with Z_m and Z_n, the vectors of the polynomials depending respectively on DIM or THI and p_m and p_n, the pe random regression coefficients. For all traits we used Legendre polynomials of order 3 for THI and order 2 for DIM with an overall mean. The residual variances were considered heterogeneous for the 55 possible combinations of five THI classes and 11 DIM classes. These random regression models allowed to estimate the variances components and individual breeding values (EBV) for all combinations of THI and DIM.

For the analyzes MY, FC, PC, and SCS were considered separately. The pedigree was traced back three generations and trimmed. Variance components and all effects were estimated using the software WOMBAT (Meyer, 2007).

Results & discussion

Effect of THI on performances. The least-squares solutions of each classes of THI are presented in Figure 1. For MY, THI was optimal in the range of 50-55 (i.e. 7 to 13 °C) with both cold and heat stress effects, whereas FC, PC and SCS tended to present only heat stress effect, declining for THI above 35 (i.e. -3 to 2 °C). These results indicate a lower comfort THI than commonly found in high-producing dairy cows (Carabaño *et al.*, 2017). However, the fact that a large part of the studies on high-producing dairy cows concerns cows intensively raised in countries warmer than France may mitigate the effect of high outdoor temperatures due to controlling inside barns, sometimes with cooling systems, and a probable acclimatization to warmer temperatures from a young age. Most of the French Montbeliarde cows are kept outdoor in spring and

summer and thus probably more faced to hot weather events. In terms of phenotypic standard deviations (sd), the decrease in performances due to heat stress was more prominent for FC and PC (respectively 0.8 for FC in all parities, 1.2 and 0.6 phenotypic sd for PC in L1 and L2) than for MY (0.2 and 0.3 phenotypic sd for L1 and L2 respectively) and SCS (<0.2 phenotypic sd for all parities).

Estimation of G×THI interactions. Figure 2 illustrates genetic correlations between all combinations of THI for MY at DIM 150d. These correlations, as well as those for all other traits (Table 1) were very high, mostly higher than 0.9, suggesting that there is no G×THI interactions and therefore few differences between animals in terms of individual response to heat stress.

Even weak, G×THI interactions allow slight variability in individual response to heat stress. An illustration of this variability in tolerance to heat is given in Table 2 presenting the EBV across THI and the corresponding rank for the MY in L2 for the seven sires with at least 500 daughters in the dataset. Although ranking of sires was rather stable and the best sires were always the same, some changes in ranking could be observed. For example, the sire S4 was intermediate at cold and comfort THI but appeared to be one of the least tolerant to high THI (THI=75).

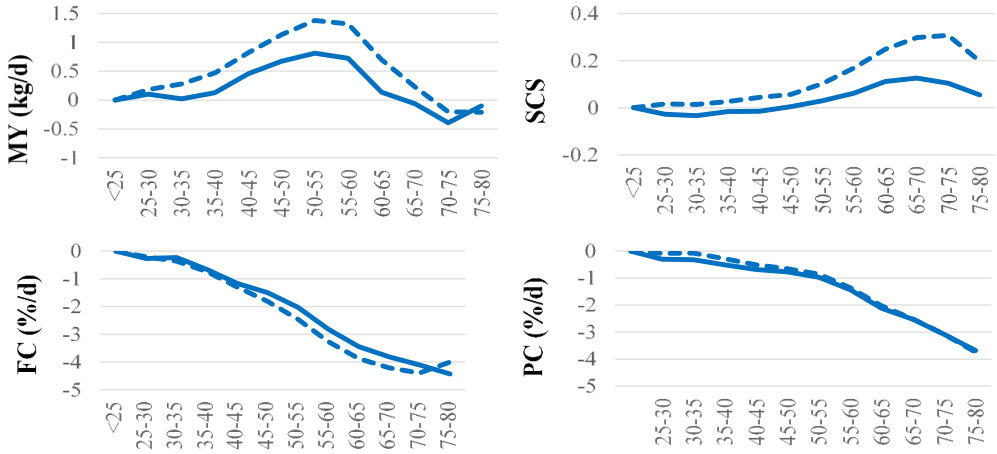


Figure 1. Least-squares solutions for MY, SCS, FC and PC in first (solid line) and second (dotted line) lactation.

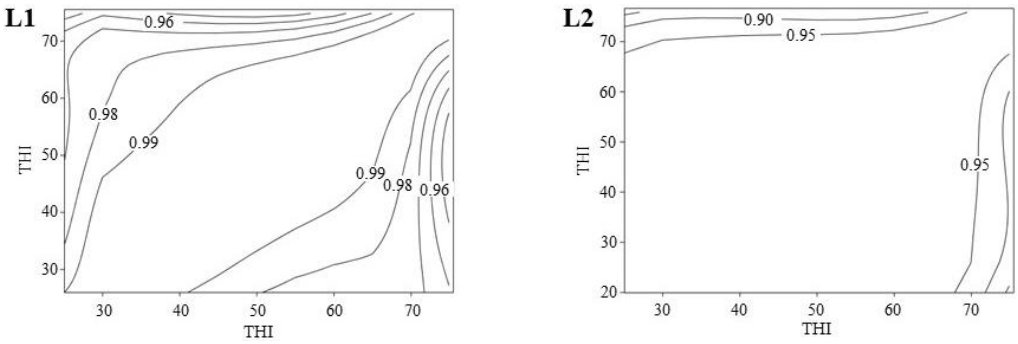


Figure 2. Genetic correlation estimates for MY at DIM 150d in L1 (left) and L2 (right) between all combinations of THI.

Table 1. Genetic correlations from random regression model analyses for MY, FC, PC and SCS between THI=40 and THI=70 at different stages of lactation (60d, 120d and 210d) and lactations rank (L1 or L2).

DIM (d)	MY		FC		PC		SCS	
	L1	L2	L1	L2	L1	L2	L1	L2
60	0.97	0.96	0.96	0.92	0.96	0.95	0.98	0.89
120	0.97	0.96	0.97	0.93	0.97	0.96	0.98	0.91
210	0.98	0.97	0.98	0.95	0.98	0.97	0.98	0.93

Table 2. EBV and rankings for MY in L2 of the sires with at least 500 daughters with performances. Results obtained at DIM=150d.

Sire	THI=30		THI=55		THI=75	
	EBV[1]	Rank	EBV	Rank	EBV	Rank
S1	0.95	1	0.88	1	1.74	1
S2	0.52	2	0.50	3	0.42	3
S3	0.33	3	0.58	2	0.42	2
S4	0.02	4	0.08	4	-0.67	6
S5	-0.27	5	-0.32	5	0.02	4
S6	-0.63	6	-0.65	6	-0.33	5
S7	-1.14	7	-0.94	7	-1.30	7

[1] EBV: expressed in genetic standard deviation unit, respectively 1.40 kg/d, 1.41 kg/d and 1.24 kg/d for MY in L2 at DIM=150d.

Conclusions

These first results confirm that the increase in temperature and in the frequency of heat peaks expected in next decades will lead to a decrease in milk production and to an increase of udder health problems in Montbéliarde cows raised in France. The selection of animals more heat-tolerant within high-producing cattle is one of the solutions presently studied to limit this decline in performance. Further studies on this topic are needed, especially concerning the reproduction of animals under heat stress. Nevertheless, although we observed only weak G×THI interactions, it seems that it is still possible to identify the animals that are more tolerant to heat stress and therefore, more adapted to global warming.

Acknowledgements

This study received funding from the European Union's Horizon 2020 research and innovation program under grant number 101000226 (Rumigen). This project adheres to EuroFAANG (https://eurofaang.eu). The CAICalor project was funded by APIS-GENE. The authors thank Meteo-France for the Safran database.

References

Carabanõ M.J., Ramon M., Diaz C., Molina A., Pérez-Guzmán M.D. *et al.* (2017) J. Anim. Sci. 95:1813-1826. https://doi.org/10.2527/jas.2016.1114

Lees M.L., Sejian V., Wallage A.L., Steel C.C., Mader T.L. *et al.* (2019) Animals 9(6):332. https://doi.org/10.3390/ani9060322

Meyer K. (2007) J. Zhejiang Univ. Sci. B 8:815–821. https://doi.org/10.1631/jzus.2007.B0815

Part 6

Challenges – resource allocation and genetics of feed intake and efficiency

47. Modelling multiplicativity in feed efficiency by regression on expected feed intake

M.H. Lidauer[1*], E. Negussie[1], T. Mehtiö[1], P. Mäntysaari[1], T. Kokkonen[2] and E.A. Mäntysaari[1]

[1]Natural Resources Institute Finland (Luke), Myllytie 1, 31600 Jokioinen, Finland; [2]University of Helsinki, Viikinkaari 1, 00014 Helsinki, Finland; martin.lidauer@luke.fi

Abstract

Factors involved in feed efficiency (FE) have multiplicative effect which should be accounted for in genetic evaluation models. We present the conceptual difference between regression on expected feed intake and residual feed intake applying both models on estimation of variance components using simulated and real FE data. Our results showed, when FE is simulated with multiplicative effects, that applying a residual feed intake model underestimates additive genetic variance. Results of the real data analyses were in line with those of the simulated data, which indicates that the regression on expected feed intake approach may be suitable and is preferred for genetic evaluation of FE in dairy cattle.

Introduction

Presently, FE traits are included into dairy cattle breeding programmes and measuring of feed intake on commercial farms is expected to increase. This will require that models for FE traits can account for the nutritional quality of the different diets. It is known that with same amount of dry matter intake (DMI), a cow's production may drop or raise due to the quality of the feed. Feed quality, for instance digestibility, has a multiplicative effect on production, i.e. the effect is less pronounced at low intake level and more pronounced at high intake level. Furthermore, in situations where intake is measured by its volume (Lassen *et al.*, 2018), metabolizable energy (ME) content per intake unit may change considerably across farms. Cow-specific efficiency factors like the individual metabolic efficiency (Martin *et al.*, 2021) have also a multiplicative effect. However, modelling FE by residual feed intake (RFI) (Koch *et al.*, 1963) offers no ways to account for it. Therefore, we present a modelling approach (Lidauer *et al.*, 2021) that is based on regression on expected feed intake (ReFI). The aim of this study was to describe the conceptual difference between ReFI and RFI and to highlight its implications with simulated and real data.

Materials & methods

For developing simple ReFI and RFI models, we assumed that there are FE records that have been collected repeatedly from cows that performed during different feed years and that the quality of diet differs across feed years.

The regression on expected feed intake approach. Under the assumption that FE factors have a multiplicative effect, it is appropriate to model FE by a random regression model where DMI is regressed on expected DMI (eDMI), which for the assumed FE records could be described as:

$$DMI_{ijk} = \beta_i \times eDMI_{ijk} + \psi_j \times eDMI_{ijk} + \alpha_j \times eDMI_{ijk} + \varepsilon_{ijk} \quad (1)$$

where DMI_{ijk} is a DMI observation k recorded in the year i for the cow j; β_i is the fixed regression coefficient for the feed year i; ψ_j is the random regression coefficient for the permanent environmental effect of cow j [$\psi \sim N(0, I\sigma^2_\psi)$]; α_j is the random regression coefficient for the additive genetic effect of cow j [$\alpha \sim N(0, A\sigma^2_\alpha)$, where A is the relationship matrix]; ε_{ijk} is the random residual [$\varepsilon \sim N(0, I\sigma^2_\varepsilon)$]; and $eDMI_{ijk}$ for record k is calculated based on energy requirements (ER) and realized production. The $eDMI_{ijk}$ covariable can be

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_47, © M.H. Lidauer *et al.* 2022

calculated by applying ER recommendations and an average ME content of the diets, which was assumed to be 11.0 MJ ME / kg DMI. Applying the Finnish ER values (Luke, 2021) results:

$$eDMI_{ijk} = 1.0/11.0\times(5.15\times ECM_{ijk}+0.515\times MBW_{ijk}+28.0\times BW\text{-}loss_{ijk}+34.0\times BW\text{-}gain_{ijk}), \quad (2)$$

where the coefficients represent the expected amount of ME needed to produce 1 kg of energy corrected milk (ECM), maintain 1 $kg^{0.75}$ metabolic body weight (MBW), gain 1 kg of body weight (BW), as well as the ME that becomes available by mobilizing 1 kg body tissue.

The residual feed intake approach. On the contrary, when assuming FE factors have an additive effect, it is reasonable to model RFI observations by a classical repeatability model:

$$RFI_{ijk} = C_i + p_j + a_j + e_{ijk} \quad (3)$$

where a RFI observation for record *k*, $RFI_{ijk} = DMI_{ijk} - eDMI_{ijk}$, is calculated by applying Equation 2; C_i is a fixed contemporary group effect of feed year *i*; p_j is the random permanent environmental effect of cow *j* [$p\sim N(0,I\sigma^2_p)$]; a_j is the random additive genetic effect of cow *j* [$a\sim N(0,A\sigma^2_a)$]; and e_{ijk} is the random residual [$e\sim N(0,I\sigma^2_e)$].

Conceptual difference between the approaches. The conceptual difference between the ReFI and RFI approaches can be demonstrated by assuming four FE records from four cows (A, B, C, D) that have been made in two contemporary groups, and that cows are equally efficient. Further, cow A is the dam of cow C and both have a daily DMI capability of 10 kg, and cow B is the dam of cow D and both have a daily DMI capability of 20 kg. Cows A and B are contemporaries and get a diet with 10% lower digestibility (L), and cows C and D are contemporaries and get a diet with 10% higher digestibility (H) compared to average diet digestibility. Consequently, under the assumption that digestibility has a multiplicative effect, the differences in the diets will cause cows A and C to have proportionally lower and cows B and D to have proportionally higher production. In the ReFI approach the effect of the diet is accounted by the fixed regression (β_i) on eDMI nested within contemporary groups, which results for these four records β_L=1.11 and β_H=0.91. In the RFI approach the effect of the diet is accounted by a fixed contemporary group effect (C_i), which results C_L=1.5 and C_H=-1.5. Then, the ReFI approach will yield no differences among the corrected observations whereas the RFI approach will yield differences within dam-cow pairs as shown in Table 1.

Simulation study. The demonstrated example shows that ReFI would be an appropriate approach if FE factors have a multiplicative effect. On the contrary, RFI would result inappropriate fit, which may cause an underestimation of additive genetic variance. To test this hypothesis, we performed a simulation study

Table 1. Conceptual differences in correcting for the diet effect.

			Regression on expected feed intake		Residual feed intake	
Cow	CG[1]	$eDMI_k$	DMI_k	$DMI_k-\beta_i\times eDMI_i$	$RFI_k=DMI_k-eDMI_k$	RFI_k-C_i
A	L	9.0	10.0	0.0	1.0	-0.5
B	L	18.0	20.0	0.0	2.0	0.5
C	H	11.0	10.0	0.0	-1.0	0.5
D	H	22.0	20.0	0.0	-2.0	-0.5

[1] Contemporary group with low (L) and high (H) digestible diet; eDMI = expected dry matter intake, DMI = dry matter intake; β_i = fixed regression coefficient effect; C_i = fixed contemporary group effect.

where DMI observations were simulated for realized productions that were recorded in FE trials. Therefore, from the Finnish FE research data a suitable data set was extracted that included 5,236 FE records from the lactation period 100£ days in milk £200 of 537 primiparous Nordic Red cattle dairy cows that were recorded during a period of 13 years at three research herds. The pedigree of the cows included 3,540 animals. A record consists of a DMI, ECM, MBW and BW change observation.

In a first step, for each record the eDMI was calculated applying Equation 2. In a second step, for model (1), when fitted to the data set, simulated values were generated for the model effects from normal distributions with standard deviations (SD) of the effects in FE percent units of 10% for the fixed feed year effects, 5% for the permanent environmental effects, and 3% for additive genetic effects. Thus, the fixed regression coefficients β_i were generated with an expectation of 1.0 and a SD of 0.1. The random regression coefficients for the permanent environmental and additive genetic effects were sampled by $\psi \sim N(0, I \times 0.0025)$ and $\alpha \sim N(0, A \times 0.0009)$ where a Cholesky decomposition of A was used for generating α, and the residuals were sampled by $\varepsilon \sim N(0, I \times 1.0)$. The simulated heritability (h^2) was 0.155 and the simulated repeatability (r) was 0.586 and can be calculated as: $h^2=\varphi\sigma^2_\alpha/(\varphi\sigma^2_\psi+\varphi\sigma^2_\alpha+\sigma^2_\varepsilon)$ and $r=(\varphi\sigma^2_\psi+\varphi\sigma^2_\alpha)/(\varphi\sigma^2_\psi+\varphi\sigma^2_\alpha+\sigma^2_e)$ with $\varphi=(d'd)/N$, where d is a vector of size n=5,236 that contains all eDMI covariables. For the used data φ was 416.3. In a final step, simulated effects and model (1) were used to calculated simulated DMI and RFI ($RFI_{ijk} = DMI_{ijk} - eDMI_{ijk}$) observations. Ten independent data replicates were simulated. Each data set was used to estimate the variance components by applying the ReFI model (1) and the RFI model (3).

Estimating variance components from real data. To get an indication of whether the ReFI approach is more appropriate for real data, the originally extracted data was used to estimate variance components by applying the ReFI model (1) and the RFI model (3). The data was recorded at three herds and therefore the feed year classification for the fixed effects in model (1) and (3) was replaced by a herd × feed year classification which comprised 20 classes.

Results

Simulation study. On average, same variance components, heritability, and repeatability as applied for the simulation were obtained (Table 2) when the simulated data sets were analysed with the ReFI model (1). However, when applying the RFI model (3), then this was only the case for the residual variance and the repeatability, but the estimated permanent environmental variance was on average larger and the estimated additive genetic variance was on average smaller compared to those used for the simulation. Thus, this resulted for the RFI model on average 20% lower heritability estimates. Furthermore, the SD of the r and h^2 estimates were larger when the RFI model was applied.

Real data. The estimated variance components when applying the ReFI model were 0.00364±0.00056 and 0.00112±0.00059 for the permanent environmental and additive genetic regression coefficients, respectively, and 1.695±0.03 for the residual effect. Thus, the estimated permanent environmental and additive genetic variation of FE were 6.0 and 3.3%. The estimated heritability and repeatability were 0.13±0.065 and 0.54±0.019. When applying the RFI model estimated variance components were 1.658±0.21, 0.246±0.21 and 1.668±0.04 for the permanent environmental, additive genetic and residual effects, respectively, which resulted a heritability and repeatability of 0.07±0.057 and 0.53±0.018. The estimated β_i fixed regression coefficients by model (1) for the herd × feed year classes had a mean of 1.01 and a SD of 3.9%, and the estimated C_i fixed effects of model (3) had a mean of 0.26 kg and a SD of 0.81 kg.

Table 2. Estimated variance components[1] from ten simulated data sets by two models.

Data sample	Regression on expected feed intake					Residual feed intake				
	σ^2_y	σ^2_α	σ^2_e	r	h^2	σ^2_p	σ^2_a	σ^2_e	r	h^2
1	0.00239	0.00090	0.97	0.584	0.160	1.04	0.37	0.99	0.588	0.155
2	0.00287	0.00065	1.01	0.592	0.109	1.41	0.06	1.02	0.589	0.024
3	0.00230	0.00083	1.00	0.565	0.151	1.11	0.17	1.01	0.559	0.075
4	0.00260	0.00096	0.98	0.602	0.163	1.03	0.47	0.99	0.602	0.190
5	0.00255	0.00091	1.00	0.591	0.156	1.07	0.38	1.01	0.591	0.155
6	0.00181	0.00162	1.04	0.577	0.273	0.81	0.58	1.06	0.567	0.237
7	0.00262	0.00082	1.02	0.585	0.139	1.25	0.19	1.02	0.583	0.076
8	0.00225	0.00074	1.05	0.543	0.135	1.03	0.18	1.06	0.533	0.080
9	0.00262	0.00097	0.99	0.601	0.162	1.10	0.45	1.00	0.606	0.175
10	0.00240	0.00056	0.96	0.563	0.107	1.10	0.15	0.97	0.565	0.069
Mean	0.00244	0.00090	1.00	0.580	0.155	1.09	0.30	1.01	0.578	0.124
SD	0.00029	0.00029	0.03	0.019	0.046	0.15	0.17	0.03	0.022	0.068

[1] Estimated variance components, repeatability (r) and heritability (h^2) for regression on expect feed intake model (1) and residual feed intake model (3) and mean and standard deviation (SD) of estimates across samples.

Discussion

We elaborated the conceptual difference between the ReFI and RFI approaches. The results of the simulation study indicate that ReFI fits better the data than RFI when FE factors have multiplicative effects. Both models were able to estimate the simulated repeatability. This was most likely because of seasonal calving and therefore having most of the repeated observations nested within same fixed effect classes. Referring to the explanation of the example in Table 2, it can be expected that both models will adjust the feed year effect similarly for average observations but not for observations deviating from the average. For the simulated data this causes reranking of cows within contemporary groups when RFI is applied, which might be the reason that RFI underestimated the additive genetic variance.

The findings from the simulation study were supported by the estimates from the real data analyses. With both models we estimated almost same repeatability for the real data, but the estimated heritability was clearly lower when applying RFI. When studying the fixed effect estimates it showed that for an average observation (eDMI=20.3 kg, RFI=0.35 kg) both models adjusted almost equally for the herd × feed year effect, i.e. the SD of the differences between the effects of both models was only 0.04 kg, whereas for observations with eDMI of 30 kg the SD of the differences increased to 0.37 kg, which caused reranking of cows. If it is valid that more precise ranking of cows will result higher heritability estimates, then this would give an indication that the multiplicative ReFI model was better suitable for the analysed FE data.

References

Koch R.M., Swiger L.A., Chambers D., and Gregory K.E. (1963) J. Anim. Sci. 22:486–494.

Lassen J., Thomasen J.R., Hansen R.H., Nielsen G.G.B., Olsen E., *et al.* (2018) Proc. of the 11th WCGALP, Auckland, New Zealand.

Lidauer M.H., Negussie E., Mäntysaari E.A., Mäntysaari P., Kajava S., *et al.* (2021) Proc. of the 72nd Annual Meeting of the European Federation of Animal Science, Davos, Switzerland.

Martin P., Ducrocq V., Faverdin P., and Friggens N.C. (2021) J. Dairy Sci. 104(6):6329–6342. https://doi.org/10.3168/jds.2020-19844.

Natural Resources Institute Finland (Luke), Feed tables 2021. Available at: http://www.luke.fi/feedtables.

48. Alternative models to predict residual feed intake in Hereford breed and effects on their breeding values accuracy

M.I. Pravia*, E.A. Navajas, I. Aguilar and O. Ravagnolo

Instituto Nacional de Investigación Agropecuaria, Las Brujas, Rincón del Colorado, Canelones, Uruguay; mpravia@inia.org.uy

Abstract

Residual feed intake (RFI) is a relevant trait, but expensive to measure so few candidates are phenotyped and have accurate GEBVs. Genetic evaluation corresponds to a single trait evaluation, where RFI are the residual from a lineal regression on feed intake on performance traits during feed-test. This study evaluates two approaches to predict a GEBV for RFI. First combing feed performance traits in a multitrait model to calculate a single measure of RFI using selection index, and 2nd the inclusion of weaning weight (WW) as an indicator trait combined with previous model. The GEBVs for the alternatives proposed were equivalent to conventional RFI. Using WW as a predictor trait on reference animals and candidates could be an inexpensive way to increase the accuracy of GEBVs. It allows to remove possible bias due to preselection of animals that participate on feed test and to extend prediction to more selection candidates.

Introduction

Residual feed intake (RFI) has been included in breeding programs with the aim of reducing feed intake without compromising animal performance. Because it is an expensive and labor-intensive to measure trait, only a small percentage of selection candidates are phenotyped Traditional genetic predictions for the rest of candidates relies on pedigree relationships with animals with records and have low accuracies. Although genomic selection is a promising tool for difficult to measure traits, accuracies depend on reference populations sizes. Using easily recordable predictor traits on both reference animals and candidates without RFI phenotypes could be another inexpensive way to increase accuracy. This approach also allows to remove possible bias due to preselection of animals that participate on feed test and to extend predictions to other selection candidates.

Conventional genetic evaluation of RFI is a two-step process. First, RFI is obtained as the residual from a lineal regression for DMI that includes metabolic liveweight (MWT), average daily gain (ADG), and ultrasound back fat (UBF) measured in feed intake tests. Breeding values are then estimated based on this phenotypic RFI in a genetic univariate model (Ravagnolo *et al.*, 2018). An alternative approach to calculate genetic merit for RFI would be by combining feed intake, growth and backfat in a multi-trait model to calculate a single measure of RFI using selection index theory (Kennedy *et al.*, 1993). Another approach proposed in this work is to use weaning weight (WW) as an indicator trait, which is already available and easily recordable in all animals, combined with DMI, ADG, MWT and UBF in a multi-trait model to calculate RFI using selection index as in previous model. The objective of this work is to analyze and compare the new approaches with the conventional methodology used in the Hereford breed in Uruguay.

Materials & methods

Animals and data. Phenotypic DMI, ADG, MWT and UBF data of 780 Hereford bulls were recorded in 9 postweaning 70-day feed intake tests, using an automated system (Growsafe) for DMI recording and following Beef Improvement Federation guidelines (BIF, 2010). Three generations of pedigree (26,572 animals), as well as genomic information of 5,439 animals imputed to the 50k SNP (BovineSNP50, Illumina San Diego, CA), were included in the analysis.

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_48, © M.I. Pravia *et al.* 2022

Methods. Three models were used for the estimation of genomic breeding values (GEBV) for RFI. In Model 1, RFI was first obtained as the residual of the following regression model:

$$DMI = CG + b_1 ADG + b_2 MWT + b_3 UBF + RFI \quad (1)$$

where CG is the feed intake test contemporary group (Test×Pen, n-15), DMI, ADG, MWT and UBF were defined previously, and b_1, b_2 y b_3 the partial regression coefficient of each trait on DMI. Then, GEBV for RFI (RFIg1) were obtained using GBLUP as implemented in BLUPF90 (Misztal *et al.*, 2016).

The genetic model was:

$$Y = Xb + Z_u + e \quad (2)$$

where Y is RFI obtained from (1), X is the incidence matrix for fixed effects (age of the animal in days at start of test as a covariable, age of dam as five class variable), Z is the matrix linking animals to phenotypes, u is the vector of genetic breeding values, and e is the vector of residual effects. It was assumed that $u_{RFI} \sim N(0, H\sigma^2_u)$ and $e \sim N(0, I\sigma^2_e)$, where H is the combination of pedigree-based relationships and differences between pedigree-based and genomic-based relationships (Aguilar *et al.*, 2010).

Model 2 is a 4-multi-trait model which included all traits used for RFI calculation (DMI, ADG, MWT and UBF):

$$\begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \\ Y_4 \end{bmatrix} = \begin{bmatrix} X_1\beta_1 \\ X_2\beta_2 \\ X_3\beta_3 \\ X_4\beta_4 \end{bmatrix} + \begin{bmatrix} Z_1u_1 \\ Z_2u_2 \\ Z_3u_3 \\ Z_4u_4 \end{bmatrix} + \begin{bmatrix} e_1 \\ e_2 \\ e_3 \\ e_4 \end{bmatrix} \quad (3)$$

where Y_{1-4} are DMI, ADG, MWT, UBF, and X and Z are the incidence matrices previously described. Solutions for GEBV for each trait were obtained, and the estimated breeding value for RFI (RFIg2) was calculated using selection index theory as follows:

$$RFI_{g2} = u_1 - b_2u_2 - b_3u_3 - b_4u_4 \quad (4)$$

where 1-4 are the GEBVs solutions for DMI (u_1), ADG (u_2), MWT (u_3), UBF(u_4) predicted from the model [3], and the RFI index coefficients calculated as:

$$\mathbf{V^{-1}C} = \begin{bmatrix} b_2 \\ b_3 \\ b_4 \end{bmatrix}, \ \mathbf{V} = \begin{bmatrix} \sigma^2_2 & \sigma_{u2u3} & \sigma_{u2u4} \\ \sigma_{u2u3} & \sigma^2_3 & \sigma_{u3u4} \\ \sigma_{u2u4} & \sigma_{u3u4} & \sigma^2_4 \end{bmatrix}, \ \mathbf{C} = \begin{bmatrix} \sigma_{u1u2} \\ \sigma_{u1u3} \\ \sigma_{u1u4} \end{bmatrix} \quad (5)$$

where V are the genetic variances and covariances for ADG (σ_2^2), MWT(σ_3^2), and UBF (σ_4^2), C the genetic covariances between DMI with ADG ($\sigma_{u1,u2}$), with MWT ($\sigma_{u1,u3}$), and with UBF ($\sigma_{u1,u4}$) (Table 1a), which were estimated using AIREML (Misztal *et al.*, 2016)

For Model 3, WW of the 780 animals with DMI, ADG, MWT and UBF data and 4,930 animals presented in their 235-contemporary groups (only with WW) at weaning were included in a 5-multi-trait model to estimate genetic (co)variances. Representation matrix is similarly to model (5). Genetic variances and covariances estimates are presented in Table 1(b). Solutions for the five traits were obtained, but only GEBVs for DMI, ADG, MWT and UBF were combined as in (4) for the estimation of RFIg3.

Table 1. Genetic co-variance components and genetic correlations obtained for 4-trait model used in $RFIg_2$ (a), and 5-trait model used in $RFIg_3$ (b).

a)	DMI (kg)	ADG (kg/d)	MWT (kg)	UBF (mm)	b)	WW (kg)	DMI (kg)	ADG (kg/d)	MWT (kg)	UBF (mm)
					WW	**249.6**	6.65	1.47	72.46	3.24
DMI	**0.36**	0.06	2.67	0.19	DMI	0.54	**0.61**	0.11	4.63	0.24
ADG	0.68	**0.02**	0.35	0.03	ADG	0.57	0.83	**0.03**	0.88	0.04
MWT	0.78	0.44	**32.66**	2.33	MWT	0.66	0.85	0.78	**48.43**	2.22
UBF	0.39	0.23	0.51	**0.64**	UBF	0.40	0.61	0.43	0.62	**0.26**

Additive genetic variance in diagonal, genetic covariances above and genetic correlations below the diagonal.

Accuracy. Individual GEBV accuracies were calculated using prediction error variances (PEV) from the inverse of the mixed model equations, as: $r = \sqrt{1 - PEV/(1+F)\sigma_g^2}$, where σ_g^2 is the genetic variance of RFIg1 ($\sigma_g^2 = 0.08$). For multi-trait analysis, the accuracies for RFI_2 and RFI_3 were calculated using PEV and PEC (prediction error covariances) (Tier and Meyer, 2004).

Bias and re-ranking. Slopes of regression and Spearman rank correlations between the alternative GEBV obtained from the different models were computed to investigate bias and changes in the ranking of animals.

Results & discussion

The high correlation coefficient and the regression slope of 1 between RFIg1 and RFIg2 (Figure 1a) indicated that Models 1 and 2 are equivalent, as described by Kennedy *et al.* (1993). The comparison of GEBVs by Model 1 and Model 3 showed some changes in ranking (Figure 1b), although the correlation between both was high. The regression slope in this case was above one indicating an underestimation of GEBVs when conventional Model 1 is used, which is corrected by Model 3. Based on Pszczola *et al.* (2013), we can assumed that RFIg3 are less biased than RFIg1 due to: (1) it enables the model to more accurately predict the level of the breeding value of an evaluated animal relative to the reference animals because records of own phenotypes for predictor traits on the evaluated animals give information about the Mendelian sampling, and (2) it accounts for differences between the evaluated animals (animals with scarcely trait records) and the reference population where all animals had the predictor trait record.

Higher standard deviations were observed for RFIg2 (0.18) and RFIg3 (0.23), in comparison with RFIg1 (0.17). This should be consequence of the genetic variances components of all traits considered in the multi-trait model. The mean of RFIg2 (0.016), and RFIg3 (0.021) where higher than RFIg1 (0.014), as result of the underestimation of predicted GEBVs mentioned above.

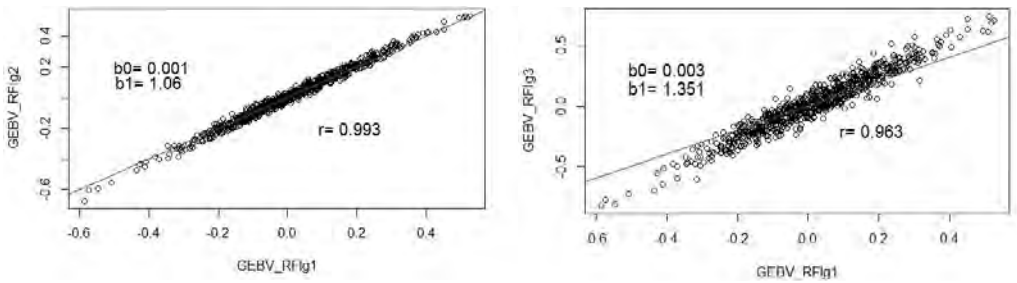


Figure 1. Correlations (r), intercept (b0) and slope of regression (b1) between alternatives GEBVs for RFI for the group of animals with phenotypes. (a) RFIg1 vs RFIg2, (b) RFIg1 vs RFIg3.

A small increase of GEBV accuracy was observed in Figure 2a for RFIg2 in relation with RFIg1, but when WW was considered in the multi-trait analysis (RFIg3, Figure 2b) the mean accuracy improved 39%. The mean accuracies for RFIg1, RFIg2 and RFIg3 for animals with phenotypes were 0.33, 0.36 and 0.46 respectively, meanwhile for contemporary animals without phenotypes nor genotypes the mean accuracies were 0.08, 0.09 and 0.11 for the three approaches, respectively. An improvement in accuracy was expected as WW is highly correlated with DMI and the other traits considered in the calculation of RFIg3.

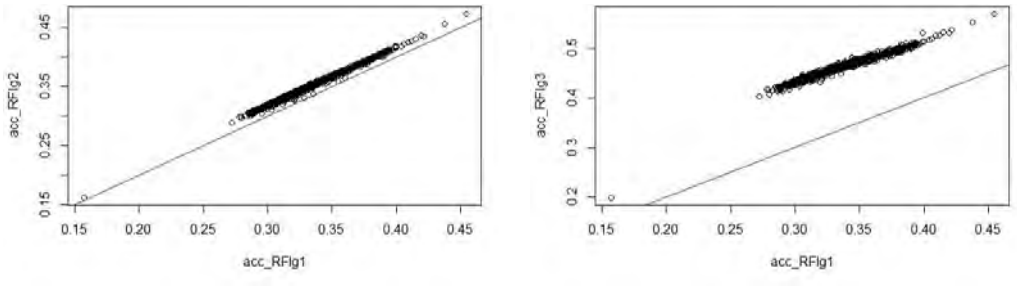


Figure 2. Correlations between accuracy of alternatives GEBVs for RFI for the group of animals with phenotypes. (a) acc_RFIg1 vs acc_RFIg2, (b) acc_RFIg1 vs acc_RFIg3.

Our results show the benefits of considering a multi-trait model including a predictor trait such as WW to predict breeding values for RFI, which were less bias and more accurate than considering a univariate model. This study will be complemented with validations analysis.

References

Aguilar I., Misztal I. Johnson D. L., Legarra A., Tsuruta S., *et al.* (2010). J. Dairy Sci. 93(2):743–752. https://doi.org/10.3168/jds.2009-2730

Kennedy B.W., Van der Welf J.H.J., Meuwissen T.H.E. (1993). J. Anim. Sci. 71(12):3239-3250. https://doi.org/10.2527/1993.71123239x

Misztal I., Tsuruta S., Lourenco D., Aguilar I. *et al.* (2016). Available at: http://nce.ads.uga.edu/wiki/lib/exe/fetch.php?media=tutorial_blupf90.pdf

Pszczola M., Veerkamp R.F., de Hass Y., Wall, E., Strabel T., *et al.* (2013) Animal. 7(11):1759-1768. https://doi.org/10.1017/S175173111300150X

Ravagnolo, O., Aguilar, I., Crowley, J.J., Pravia, M.I., *et al.* (2018). Proc. of the 11th WCGALP, Auckland, New Zealand, 11:723. http://www.wcgalp.org/proceedings/2018/

Tier, B. and Meyer, K. (2004). Anim. Breed. Genet. 121(2):77–89. https://doi.org/10.1111/j.1439-0388.2003.00444.x

49. Feed intake can be predicted as quantitative or qualitative traits

F. Tortereau[1*], C. Marie-Etancelin[1], D. Marcon[2] and J.L Weisbecker[1]

[1]GenPhySE, Université de Toulouse, INRAE, ENVT, 31326 Castanet-Tolosan, France; [2]INRAE, Experimental Unit P3R, 18390 Osmoy, France; flavie.tortereau@inrae.fr

Abstract

Feed intake (FI) of ruminants is of high interest to study feed efficiency, greenhouse gas emissions and health traits. However, FI is a very expensive trait to record. Here, we investigated the genetic correlations between true average daily feed intake (ADFI) considered either as a continuous or categorical trait, and the correlations between simple predictions of FI based on body weights and growth with true ADFI. ADFI is genetically the same trait whether considered as a continuous or categorical trait. RFI calculated from categorical ADFI was genetically the same trait as RFI calculated from true ADFI. However, estimated ADFI from a simple model accounting for zootechnical traits, although positively correlated with true ADFI, led to reduced genetic correlations between categorical traits and ADFI: a maximum of 25% of misclassification of lambs is expected if qualitative predictions of ADFI or RFI have to be considered in future analyses.

Introduction

Feed efficiency (FE) is a key trait to improve in small ruminants, from an environmental to an economic point of view. Different criteria exist to define FE, the more popular being feed conversion ratio (FCR) and residual feed intake (RFI) that are considered the reference/gold standards. However, both require feed intake (FI) to be recorded. Different tools exist to record individual FI, but most of them cannot be used in commercial farms, either because they do not comply with animal management (e.g. individual crates) or because they are too expensive (automatic feeders or individual troughs). Many studies aim at identifying proxies of FI or Average Daily Feed Intake (ADFI, over a given period of time) such as metabolites or microbiota, with models also including feed characteristics. Pittroff and Kothmann (2001) reviewed models of quantitative predictions of FI applied to sheep, and concluded that additional studies and datasets were required. Predicting FI is still a challenge in livestock, as shown by recent studies on the relationships between FI and body weights (Lewis and Emmans, 2020). In physiological studies, animals are commonly grouped based on their ADFI to carry out different analyses, but to our knowledge, no genetic study looked at ADFI as a categorical trait; this could enable the application of other prediction models. The objective of this work is to study the genetic parameters of true and predicted ADFI (over an eight-week period) considered either as a quantitative or categorical trait, as well as RFI.

Materials & methods

Animals and traits. Animals and traits have already been presented in Tortereau *et al.* (2020). Briefly, a total of 951 Romane male lambs were phenotyped under a 100% low-energy concentrate diet for body weights, feed intake and body composition traits during 8 weeks, starting from 3 months of age. Daily feed intake was averaged over the 8-week testing period to estimate the average daily feed intake (ADFI). The individual RFI was estimated as the residual of the multiple linear regression of ADFI on Average Daily Gain (ADG), BackFat Thickness and Muscle Depth at the end of the test to account for production requirements, and on the metabolic BW (mBW) at the end of the test to account for maintenance requirements (proc reg; SAS Institute Inc., Cary, NC, USA).

Definition of alternate ADFI traits and RFI. We defined new traits based either on true ADFI or on estimated ADFI (eADFI). Estimated ADFI was defined based on the coefficients retrieved from the multiple

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_49, © F. Tortereau *et al.* 2022

linear regression of ADFI on ADG, mBW and age at the beginning of the test (startAge), on a subset of 465 lambs tested from 2009 to 2012. Then, this equation was applied to all individuals (2009-2016) to calculate eADFI. Mean (m) and standard deviation (sd) of ADFI and eADFI were calculated to define the thresholds between five categories of the new categorical traits ADFI_5 and eADFI_5 (Figure 1).

The average ADFI and eADFI were estimated for each of the five categories, and each lamb was assigned this average as a new trait value (ADFI5 and eADFI5, respectively).

An additional RFI trait (RFI5) was calculated with the same multiple linear regression as presented above applied to ADFI5. To summarize, each lamb had eight traits: ADFI, eADFI, RFI, ADFI_5gr, eADFI_5gr, ADFI5, eADFI5, and RFI5.

Estimation of genetic parameters. Heritabilities and genetic correlations between continuous traits were estimated with univariate and bivariate analyses between each pair of traits under an animal mixed model using the REML methodology implemented in the ASReml software (Gilmour *et al.*, 2009). Genetic correlations involving categorical traits (ADFI_5gr and eADFI_5gr) were estimated with TM software developed by Legarra *et al.* (2011) which handles categorical traits. The fixed effects in the models fitted to each trait were the year (n=8) and the pen (n=7). Body weight at the beginning of the test was included as a covariate for ADFI, but not for RFI.

Results

With the subset of lambs born from 2009 to 2012, we defined the prediction equation of feed intake as eADFI = -550 + 0.90×ADG + 114×mBW – 1.48×startAge (R^2=0.80).

With the full dataset, the phenotypic correlation between ADFI and eADFI was 0.78±0.01. A total of 74% of lambs were classified in the same ADFI and eADFI categories. Among the other 26% of lambs classified in different categories, all individuals but three were assigned to the group just before or just after (Table 1).

The average differences between ADFI and ADFI5 or eADFI5 (average values of ADFI or eADFI) of the category the animal belongs to) were 5.2% (±3.7) and 6.8% (±5.0), with a maximum difference of 14.6 and 30.2%, respectively.

Heritabilities of traits related to ADFI and eADFI ranged from 0.10±0.06 for eADFI5 to 0.33±0.10 for ADFI_5gr (Table 2).

Genetic correlations indicate that ADFI, ADFI_5gr and ADFI5 are genetically the same trait. Similarly, RFI is the same trait whether estimated from true ADFI or from averaged values (RFI5) (Table 2). The genetic correlation between true and estimated ADFI was positive (0.62±0.12). The categorical trait eADFI_5gr and the corresponding continuous trait eADFI5 showed significant positive genetic correlations with ADFI (0.63 and 0.54, respectively).

Figure 1. Definition of ADFI_5gr and eADFI_5gr categorical traits, with m (mean) and sd (standard deviation), for ADFI (m=1,962 g/d; sd=249 g/d) and eADFI (m=1,929 g/d; sd=231 g/d).

Table 1. Distribution of lambs into five ADFI and eADFI categories.

		eADFI_5gr				
		1st categ.	2nd categ.	3rd categ.	4th categ.	5th categ.
ADFI_5gr	1st categ.	11	10			
	2nd categ.	20	85	33		
	3rd categ.	2	77	540	41	
	4th categ.			47	66	9
	5th categ.			1	3	6

Table 2. Predicted traits: heritabilities and genetic correlations with true average daily feed intake and true residual feed intake.

	ADFI	ADFI_5gr	ADFI5	RFI5	eADFI	eADFI_5gr	eADFI5
type of trait[1]	cont.	categ.	cont.	cont.	cont.	categ.	cont.
h^2	0.28 (±0.08)	0.33 (±0.10)	0.16 (±0.07)	0.20 (±0.07)	0.22 (±0.08)	0.27 (±0.10)	0.10 (±0.06)
genetic corr. with ADFI		1.00	1.00	0.68 (±0.16)	0.62 (±0.12)	0.63 (±0.16)	0.54 (±0.21)
genetic corr. with RFI	0.78 (±0.08)	not different from 0	0.87 (±0.11)	1.00	not different from 0	not different from 0	not different from 0

[1] Traits are either categorical (categ.) or continuous (cont.).

Discussion

The link between feed intake and body weight, growth and age is well known, and many studies aimed at predicting quantitatively feed intake (Pittroff and Kothmann, 2001). With Romane male lambs, we highlighted that ADFI is genetically the same trait when expressed as a categorical trait (ADFI_5gr) or as average values of feed intake (ADFI5). This makes it possible to consider a wider variety of prediction methods including qualitative prediction methods. Recently, in pigs, machine learning procedures were successfully applied on transcriptomic data to predict either categories of FE (with a 100% success of assignation) or FE values (with R^2=0.80 and RMSE=8%) (Messad *et al.*, 2021). In our study, we highlight that RFI and RFI5 are genetically the same trait despite the 5% difference between ADFI and ADFI5. Moreover, RFI5 was less heritable than RFI (0.20±0.07 vs 0.45±0.08; Tortereau *et al.*, 2020). Additional studies of the impact of feed intake accuracy on RFI estimates and RFI genetic parameters will be of high interest in order to better consider the phenotyping strategies of feed intake.

In our study, we tested different numbers of categories (from two to five based on ADFI and eADFI distributions) and results were similar to the ones we presented with five categories (results not shown). However, the genetic correlation between eADFI_5gr and ADFI is only 0.63, which is significantly different from 1.00: misclassification of a quarter of the individuals would have consequences for the genetic correlation with ADFI.

With a simple estimation of ADFI from age, mBW and ADG, we obtained a new trait, eADFI, significantly correlated with ADFI (Table 2). The distribution of lambs within five classes, eADFI_5gr, was still positively genetically correlated to ADFI. Even if the objective of this paper was not to analyse the prediction accuracy of ADFI from classical zootechnical traits, this result highlights that such a simple prediction might be proposed as a first prediction of ADFI. However, with such a prediction, RFI cannot be calculated given that eADFI represents the maintenance and requirements of ADFI, whereas it might be suitable for prediction

of FCR. Most of the models considered in Pittroff and Kothmann (2001) rely on animal requirements and thus do not consider the residual part of feed intake. New proxies such as transcriptome, metabolites and microbiota have to be considered in order to better predict ADFI and RFI (Cantalapiedra-Hijar *et al.*, 2018). With PLS-DA approaches carried out on ruminal microbiota data only, it was possible to discriminate high vs low efficient animals, and high vs low eating lambs, but with a poor prediction accuracy, particularly for RFI groups (Le Graverand *et al.* 2021). Therefore, future models combining new proxies and zootechnical traits might help in predicting ADFI or RFI in sheep, whether quantitatively or qualitatively.

Acknowledgements

This work received funding from the French funding agency (ANR) (GrassToGas project (ERA-GAS no. 39413)) and from the European Union's Horizon 2020 research and innovation program under the Grant Agreement No. 772787 (SMARTER).

References

Cantalapiedra-Hijar G., Abo-Ismail M., Carstens G.E., Guan L.L. *et al.* (2018) Animal 12(S2):s321-s335. https://doi.org/10.1017/S1751731118001489

Gilmour AR, Gogel BJ, Cullis BR, Thompson R and Butler D (2009) ASReml user guide release 3.0. VSN Int. Ltd Hemel Hempstead UK.

Le Graverand Q. Tortereau F., Meynadier A., Marcon D. and Marie-Etancelin C. (2021) Proc. of the 72nd EAAP, Davos, Switzerland.

Legarra, A., L. Varona & E. Lopez de Maturana, (2011) TM Threshold models – Available at: http://genoweb.toulouse.inra.fr/~alegarra/tm_folder/manualtm.pdf

Lewis R.M. and Emmans G.C. (2020) J Anim Sci. 98(4):skaa087. https://doi.org/10.1093/jas/skaa087.

Messad F., Louveau I., Renaudeau D., Gilbert H. and Gondret I. (2021) BMC Genomics 22:501. https://doi.org/10.1186/s12864-021-07843-4

Pittroff W. and Kothmann M. M. (2001). Livest. Prod. Sci. 71(2-3):131–150. https://doi.org/10.1016/S0301-6226(01)00218-4

Tortereau F., Marie-Etancelin C., Weisbecker J.-L., Marcon D. *et al.* (2020) Animal 14:681–687. https://doi.org/10.1017/S1751731119002544

50. Prediction of breeding values for feed intake in pigs using individual versus group records along with correlated traits

C. Zhang[1*], R.A. Kemp[1], J.C.M. Dekkers[2], G.S. Plastow[3] and H. Gao[4]

[1]Genesus Inc., Oakville, MB R0H 0Y0, Canada; [2]Department of Animal Science, Iowa State University, Ames, IA 50011, USA; [3]Department of Agricultural, Food and Nutritional Science, University of Alberta, Edmonton, AB T6G 2R3, Canada; [4]Natural Resources Institute Finland (Luke), Myllytie 1, 31600 Jokioinen, Finland; czhang@genesus.com

Abstract

Individual average daily feed intake (ADFI), age at 120 kg (AGE), backfat (FAT), loin muscle depth (LMD), and group feed intake (sumFI) were used to investigate estimated breeding values (EBV) for feed intake (FI) using individual versus group FI data, along with individual data on AGE, FAT, and LMD, which had moderate to high genetic correlations with ADFI. The correlation of FI EBV using only ADFI versus sumFI was <0.25, which increased to 0.78-0.93 when adding data of the correlated traits in 4-trait models. The correlation between FI EBV from the 4-trait model with/without sumFI data was 0.99. The correlation between EBV from the 4-trait model with/without ADFI ranged from 0.79-0.94. Thus, group FI has negligible value when using correlated traits in pedigree-based analyses, while individual ADFI provides benefit, especially for young selection candidates.

Introduction

Feed intake (FI) is an economically important trait for pig production. It is very costly to collect feed intake data on individual pigs compared with other production traits and thus the number of pigs with individual FI data is limited. A more cost-effective way would be to collect feed intake at the group level as pigs are normally housed in groups and fed by a shared feeder. This group data may be valuable in increasing the accuracy of individual feed intake estimated breeding values (EBV) when combined with individual feed intake data. To date, several studies have investigated this approach using simulated data (Olson *et al.*, 2006; Su *et al.*, 2018; Gao *et al.*, 2019; Ma *et al.*, 2021) but very few studies have used real data (Piles and Sánchez, 2019; Gao *et al.*, 2021). The general conclusion from these studies is that group records can be used to obtain EBV for individual animals, and the EBV accuracy can be increased by adding genomic information and individual data on genetically correlated traits. Therefore, the objective of this study was to use previously collected individual feed intake and production trait phenotypes to evaluate the potential of using group feed intake data and correlated traits to enhance prediction of individual feed intake EBV.

Materials & methods

Animals and data collection. A total of 12,021 Duroc nucleus pigs (9,215 males and 2,806 females) tested during 2004-2021 were used for this study. For feed intake, each group started with 22-24 pigs in a single test pen fitted with two electronic feeders (IVOG-feeding-station, Hokofarm Group). Details about animal management and feed intake recording are in Zhang *et al.* (2019). Before 2015, the individual feed intake was recorded from the start of the grow-finish phase (~25 kg), and after 2015 it was recorded from the midpoint (~70 kg), until the average weight within the pen was about 120 kg. Individual average daily feed intake (ADFI) was calculated as the FI in the last 6 weeks of the test period (about midpoint to the end of grow-finish phase) based on linear regression of daily FI on the days on test. Individual body weight was measured around 70 kg and considered as the initial weight. At the end of test, animals were individually weighed and backfat (FAT) and longissimus muscle (LMD) depth were measured by B-mode ultrasound. End test weight was adjusted to 120 kg and used to compute age at 120 kg (AGE). Both FAT and LMD were also adjusted to 120 kg. Group FI (sumFI) was computed as the sum of ADFI of all animals in a pen. Finally,

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_50, © C. Zhang *et al.* 2022

a total of 742 groups were generated from 12,021 animals. Group sizes ranged from 7 to 24, including 1 to 8 littermates. A pedigree of 18 generations with 16,274 animals, including 905 sires and 3,897 dams, was used to compute the numerator relationship matrix. It is important to note that sumFI was calculated as the sum of the ADFI for all animals in a group. Since ADFI collected on each animal is the only real source of data, ADFI and sumFI are expected to predict the same feed intake breeding value. Thus, any departure from 1 of the correlation between EBV for FI based on ADFI versus sumFI depends on which data is used for prediction.

Statistical methods. Five scenarios were used to obtain FI EBV using ADFI or sumFI, including EBV1: EBV of ADFI from a univariate model using ADFI data; EBV2: EBV of ADFI from a 4-trait model using data on ADFI, AGE, FAT, and LMD; EBV3: EBV of sumFI from a univariate model using sumFI data; EBV4: EBV of sumFI from a 4-trait model using data on sumFI, AGE, FAT, and LMD; EBV5: EBV of ADFI from the same model as EBV2 but with ADFI data being set as missing for all animals. As sumFI was calculated from ADFI, they were assumed to have the same variance components (VCs) as those for ADFI. Therefore, VCs obtained from a univariate analysis of ADFI were used for EBV1 and EBV3, while VCs from the 4-trait model with ADFI, AGE, FAT, and LMD were used for EBV2, EBV4, and EBV5. The model for the individual records was:

$$y = Xb + Z_a a + Z_l l + Z_c c + e \quad (1)$$

where y is the vector of phenotypes for the trait, b is a vector of fixed effects, which represents sex and pen (across off-test dates) and initial body weight as a covariate, a is a vector of random additive genetic effects [$a \sim N(0, A\sigma_a^2)$], where A is the pedigree based relationship matrix and σ_a^2 is the additive genetic variance, l is a vector of litter effects [$l \sim N(0, I\sigma_l^2)$], c is a vector of group effects (combination of off-test date and pen) effects with variance σ_c^2 [$c \sim N(0, I\sigma_c^2)$], where I is an identity matrix and e is a vector of random residuals with variance σ_e^2 [$e \sim N(0, I\sigma_e^2)$], and X, Z_a, Z_l, and Z_c are incidence matrices associating *b,* a, l, and c with y. Following Su *et al.* (2018), the model for sumFI corresponding to the model for individual records was:

$$T(y) = T(Xb + Z_a a + Z_l l + Z_c c + e) \quad (2)$$

where T is an indicator matrix that links individual records to groups. Element (T_{ij}) of T is 1 if the j^{th} animal belongs to the i^{th} group, or 0 otherwise. Matrix T also sums variables for each level of a particular factor within a group. All analyses were performed using the DMU package (Madsen and Jensen, 2013). The correlations of EBVs from different models were compared by sex and for young animals (those completing the test in the last 10 months) and for the remaining called 'old animals' in this study.

Results

Estimates of the genetic correlations between ADFI with AGE, FAT, and LMD were -0.80, 0.20, and -0.49, respectively. Moderate genetic correlations were also found among AGE, FAT, and LMD, with a range of absolute values from 0.25 to 0.42.

Correlations of EBV between different scenarios are presented in Table 1. Based on the univariate models, the correlation of EBV between ADFI (EBV1) and sumFI (EBV3) was less than 0.25. When individual data on AGE, FAT and LMD were included in a 4-trait model, the correlation of EBV based on ADFI (EBV2) versus sumFI (EBV4) increased to 0.78-0.82 for young animals and to 0.93 for old animals. These results indicate that, when using group feed intake records, the individual records on the feed intake correlated traits, such as AGE, FAT and LMD, should be included.

Table 1. Correlations of EBV with or without individual or group feed intake data and data on correlated traits.

			Correlation of EBV[2]								
Sex	Animal[1]	Number	1-3	2-4	5-1	2-5	3-5	4-5	1-2	3-4	4-1
F	old		0.16	0.93	0.68	0.94	0.17	0.99	0.87	0.18	0.67
F	young	595	0.12	0.82	0.42	0.83	0.20	0.99	0.83	0.23	0.42
M	old	8,656	0.24	0.93	0.68	0.93	0.25	0.99	0.88	0.26	0.67
M	young	538	0.09	0.78	0.40	0.79	0.16	0.99	0.85	0.21	0.40

[1] Young includes animals with end of test less than 10 months before the latest end of test date in this data set; the remaining tested animals are referred to as 'old'.

[2] 1-3 means correlation between EBV1 and EBV3, the same for all other comparisons. EBV1: EBV of ADFI from a univariate model using ADFI data; EBV2: EBV of ADFI from a 4-trait model using data on ADFI, AGE, FAT, and LMD; EBV3: EBV of sumFI from a univariate model using sumFI data; EBV4: EBV of sumFI from a 4-trait model using data on sumFI, AGE, FAT, and LMD; EBV5: EBV of ADFI from the same model as EBV2 but with ADFI data being set as missing for all animals.

To further evaluate the value of using group feed intake for individual FI EBV, the ADFI data was set as missing for all animals in the 4-trait model, which means that the EBV of feed intake for individual animals (EBV5) was indirectly predicted using the three correlated traits only. Compared with EBV1, which was based on ADFI only, the correlation between EBV4 and EBV1 was very similar to the correlation between EBV5 and EBV1. Furthermore, the correlation of EBV using the 4-trait model with EBV for sumFI (EBV4) and with EBV without any data on ADFI (EBV5) was 0.99. These results indicate that, if individual data on the correlated traits are already available, adding group feed intake data in the multi-trait model had little impact on FI EBV. When individual data on the correlated traits are available, collecting individual feed intake is beneficial for feed intake EBV, especially for young animals. The correlation between the EBV obtained from the 4-trait model with (EBV2) and without (EBV5) individual feed intake was 0.79 and 0.83 for young males and females, respectively. The correlation of EBV based on ADFI only (EBV1) with EBV based on the 4-trait model without ADFI (EBV5) was even lower, at 0.40 and 0.42 for young males and females, respectively.

Discussion

Our results show that single-trait EBV for feed intake based on group FI records had a very low correlation with single-trait EBV based on individual FI records. When individual data on the correlated traits AGE, FAT, and LMD were added in a multi-trait analysis, the correlation of EBV from group FI records with EBV from individual FI records greatly increased. Therefore, when using group feed intake, these correlated traits should be collected and used to predict EBV for individual FI. Using simulation, Ma *et al.* (2021) also reported that including individual data on a correlated trait (r=0.8) increased the accuracy of EBV for a trait with group records. It is usually assumed that, when using pedigrees, including correlated traits with individual records is a good approach to improve the accuracy of EBV for a trait with group records, as littermates within a group can obtain different EBV resulting from having different phenotypes for the correlated traits (Su *et al.*, 2018; Cooper *et al.*, 2010; Ma *et al.*, 2021). Our results also show that, when individual data on correlated traits are available, adding group feed intake in the multi-trait model had very little value to improve the accuracy of EBV for individual FI. Ma *et al.* (2021) also reported that including group records had little impact on the accuracy of EBV for the trait with group records when individual data of a correlated trait were included in a bivariate model. However, based on our results, adding individual feed intake in the 4-trait model resulted in a substantial change in the FI EBV, especially for young animals.

Several studies have shown that the accuracy of EBV based on group records is greater when group size is smaller and when individuals within a group are more closely related (Olson *et al.*, 2006; Peeters *et al.*, 2013; Su *et al.*, 2018; Ma *et al.*, 2021). For practical reasons, our data consisted of relatively large group sizes, with an average of 16 pigs per group, and a range of 7-24 pigs per group in the final analysis. Relationships within a group were also relatively small, as ~30% of the animals had no littermate within their group and ~30% of the animals had only 1 or 2 littermates. In addition, relationships used in our analyses were based on pedigree, which cannot separate littermates, either across or within groups. Genomic information can distinguish littermates and has the advantage of capturing both Mendelian segregation and genetic links through unknown common ancestors. Both Gao *et al.* (2021) and Ma *et al.* (2021) reported an increase in EBV accuracy from adding genomic information, especially when predicted using the bivariate model. About 60% of the animals in this study have been genotyped and this data will be used in subsequent analyses.

References

Cooper A.J., Ferrell C.L., Cundiff L.V., and Van Vleck L.D. (2010) J. Anim. Sci. 88:1967–72. https://doi.org/10.2527/jas.2009-2391

Gao H., Nielsen B., Su G., Madsen P., Jensen J., *et al.* (2019) G3 (Bethesda) 9(9):2935-2940. https://doi.org/10.1534/g3.119.400484

Gao H., Su G., Jensen J., Madsen P., Christensen O.F., *et al.* (2021) Genet. Sel. Evol. 53: 33. https://doi.org/10.1186/s12711-021-0062

Ma X., Christensen O.F., Gao H., Huang R., Nielsen B., *et al.* (2021) Heredity 126:206–217. https://doi.org/10.1038/s41437-020-0339-3

Madsen P., and Jensen J. (2013) A user's guide to DMU-A package for analysing multivariate mixed models. Version 6. http://dmu.agrsci.dk/ DMU/Doc/Current/dmuv6_guide.5.2.pdf.

Olson K.M., Garrick D.J., and Enns R.M. (2006) J. Anim. Sci. 84(1):88-89. https://doi.org/10.2527/2006.84188x

Peeters K., Ellen E.D., and Bijma P. (2013) Genet. Sel. Evol. 45:27. https://doi.org/10.1186/1297-9686-45-27

Piles M., and Sánchez J.P. (2019) J. Anim. Breed. Genet. 136(6):474-483. https://doi.org/10.1111/jbg.12395

Su G., Madsen P., Nielsen B., Ostersen T., Shirali M., *et al.* (2018) Genet. Sel. Evol. 50:42. https://doi.org/10.1186/s12711-018-0413-y

Zhang C., Kemp R.A., Stothard P., Wang Z., Boddicker N., *et al.* (2018) Genet. Sel. Evol. 50:14. https://doi.org/10.1186/s12711-018-0387-9

51. Phenotypic and genetic associations between feed efficiency and FTIR milk-spectra

H.O. Toledo-Alvarado[1*], R.J. Tempelman[2], M. Lopez-Cruz[3], M.J. VandeHaar[2], J.E.P. Santos[4], F. Peñagaricano[5], P. Khanal[2] and G. de los Campos[6,7,8]

[1]Department of Genetics and Biostatistics, National Autonomous University of Mexico, Ciudad Universitaria, P.O. Box 04510, Mexico City, Mexico; [2]Department of Animal Science, Michigan State University, Michigan, P.O. Box 48824, East Lansing, MI, USA; [3]Department of Plant, Soil and Microbial Sciences, Michigan State University, Michigan, P.O. Box 48824, East Lansing, MI, USA; [4]Department of Animal Sciences, University of Florida, Florida, P.O. Box 32608, Gainesville, FL, USA; [5]Department of Animal and Dairy Sciences, University of Wisconsin-Madison, Wisconsin, P.O. Box 53706, Madison, WI, USA; [6]Department of Epidemiology and Biostatistics, Michigan State University, Michigan, P.O. Box 48824, East Lansing, MI, USA; [7]Department of Statistics and Probability, Michigan State University, Michigan, P.O. Box 48824, East Lansing, MI, USA; [8]Institute for Quantitative Health Science and Engineering, Michigan State University, Michigan, P.O. Box 48824, East Lansing, MI, USA; h.toledo.a@fmvz.unam.mx

Abstract

Data involving 9,226 records from 483 US Holstein cows were used to study the genetic relationship between feed efficiency, measured as Residual Feed Intake (RFI) and the Fourier transform infrared (FTIR) spectra of milk samples. Estimates of heritabilities, genetic and phenotypic correlations, as well as the accuracy of an indirect selection using spectra were calculated. The prediction accuracy of RFI was also evaluated using different methodologies. Genotypes were used to fit univariate and bivariate (each individual wavenumber and RFI) GBLUP analyses. A genetic relationship between RFI and wavenumbers associated with fat, protein and urea were observed. Notable differences were detected in the patterns of phenotypic versus genetic correlations. The maximum indirect selection accuracy estimated using a particular wavenumber was 0.11. No notable differences in the accuracy of genetic prediction were observed between different methodologies that use the full spectra for RFI prediction, and the accuracies were generally higher (~0.15) than using the individual wavenumbers.

Introduction

Feed efficiency is an economically and environmentally relevant trait; however, measuring it in commercial dairy farms is not feasible. Therefore, there is an increased interest in using milk-spectroscopy to predict and use feed efficiency to improve farming management and genetic programs. Feed efficiency can be measured indirectly through the residual feed intake (RFI) of cows in experimental conditions; these data can be used to calibrate equations to predict RFI from phenotypes recorded in commercial conditions such as milk-spectra. The absorbance of any given Fourier transform infrared (FTIR)-spectra wavelength is a proxy phenotype that is potentially affected by genetic and environmental factors with a given heritability (Bittante and Cecchinato, 2013). Usually, competing prediction methods (e.g. partial least squares, LASSO, Bayesian methods) are compared based on the (phenotypic) correlation between predicted RFI and measured RFI in testing data sets. Such phenotypic correlations are affected by both genetic and environmental effects affecting RFI and the measured traits. From a genetic perspective, the accuracy of indirect selection may be a more appropriate measure of performance. Therefore, the objective of this study was to quantify the genetic prediction accuracy of individual wavelengths and of commonly used prediction models as predictors of genetic merit for RFI.

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_51, © H.O. Toledo-Alvarado *et al.* 2022

Materials & methods

Data from several experimental trials conducted at the University of Florida (7,205 records from 311 cows) and Michigan State University (2,021 records from 172 cows) were used in this study. Cows were randomly assigned to individual feeding gates used to evaluate daily feed intake. Cows had *ad libitum* access to their respective diets, which were fed as total mixed ration (TMR). Dry matter intake (DMI) for individual cows was based on the amount of TMR offered minus the amount refused multiplied by the dried matter content of the TMR in a particular week. Data collected each week included the body weights, milk yields and milk composition, and the FTIR spectral data associated with milk samples. All milk samples were analysed using Bentley FTS instruments with 899 spectral points from wavenumber 649 to 3,998 cm^{-1}. Animals were genotyped on various available chips and imputed. The genotypes were edited for quality control and 78,964 SNP were included in the analysis. A better description of the data can be found in Gaddis *et al.* (2021).

RFI calculation. An energy sink model (1) was used to estimate RFI using the net energy for lactation (NE_L) as determined in the NRC (2001). The weekly metabolic body weight (MBW) was calculated as the weekly average $BW^{0.75}$. The weekly body weight changes ΔBW was calculated using a linear regression of actual or projected BW on days from the beginning to the end of a week (Berry and Crowley, 2013). Thus, the model to measure residual feed intake included the fixed effects of parity (primiparous or multiparous) interacting with the specific fifth-order Legendre polynomial regressions on DIM, and linear regressions on NE_L, MBW and the ΔBW. The random effects included the effects of experiment-specific rations and test weeks. The estimated residuals $\hat{\varepsilon}$ from this model were determined to be the weekly RFI.

Prediction accuracy of individual wavelengths. The accuracy of indirect selection of an indicator trait (a wavenumber in our case) is the product of the square-root of the heritability of the indicator trait (h_I^2) times the genetic correlation between the measured trait (wavenumber) and the selection objective (RFI, r_{g_y,g_I}), that is $Acc(g_{y_i}, I_i) = r_{g_y,g_I}\sqrt{h_I^2}$ (Falconer and Mackay 1996), where I stands for each wavenumber. Therefore, we first fitted a sequence of 899 bivariate GBLUP models using RFI and one wavenumber (adjusted by the same effects included in the model to estimate the RFI) as the two traits. From these analyses we estimated h_I^2, r_{g_y,g_I}, and $Acc(g_{y_i}, I_i)$. All the analyses were performed using the Multitrait() function of the BGLR R-package (Pérez and de los Campos, 2014), with genomic relationship matrices derived from genotypes, including the random effects of the animal and the permanent environment.

Genetic prediction accuracy of prediction equations. To evaluate the genetic prediction accuracy that can be achieved when combining information from all the spectrum we calibrated prediction equations for RFI (with the 899 available wavelengths as predictors) using Partial Least Squares (PLS with 10 components) and five Bayesian methods which differed on the prior used. To evaluate accuracy, and following Lopez-Cruz *et al.* (2020), we partitioned the data into a training and a testing set. The testing set was formed by randomly selecting 50 cows and assigning the phenotypes of those cows to the testing set. The remaining data was used for model training. We then calibrated the prediction equations using the training data. We then applied the resulting prediction equations to the milk spectra data of the testing set to obtain a prediction of RFI phenotypes. Finally, we estimated the genetic prediction accuracy of each method by fitting bivariate GBLUP models with RFI and the predicted RFIs as traits, including the random effects of the animal and the permanent environment. To account for uncertainty due to sampling of training and testing sets we conducted 100 random training-testing partitions and applied the procedure above described to each of the partitions; from this, we present results averaged over 100 training-testing partitions.

Results

The estimated heritability for the RFI was 0.17±0.03, while for the spectra the estimated heritabilities (Figure 1a) ranged between 0.12 and 0.30. The maximum heritability was found at wavenumber 2,976 cm^{-1}; in this region, Fat-B [absorbance by the carbonhydrogen stretch (C–H)] is usually filtered (Kaylegian *et al.*, 2006). Genetic and phenotypic correlations varied with spectra region (Figure 1b) with estimated genetic correlations ranging between -0.24 and 0.22. From wavenumber 1,167 to 1,689 cm^{-1}, positive genetic and phenotypic correlations were estimated, this region is associated with the prediction of protein (1,546 cm^{-1}) (Kaylegian *et al.*, 2006), urea (1,469 cm^{-1}), fat (1,460 cm^{-1}) (Hansen, 1998), and acetone (1,238 cm^{-1}) (Hansen, 1999). The maximum value observed for the accuracy of the index was 0.11 at the wavenumber 1,536 cm^{-1}, note that protein is usually filtered from 1,546 cm^{-1} to 1,492 cm^{-1} (Kaylegian *et al.*, 2006).

The heritabilities of predicted phenotypes across different methodologies (Table 1) were similar to the estimated heritability of RFI. No major differences were found between the different methodologies for using milk-spectra in the prediction of RFI (ranging from 0.14 to 0.16). However, the Bayesian methods (A and Lasso) showed a genetic correlation and accuracy slightly higher than the PLS traditional methodology. An important point to note is that the accuracy was higher when using the entire spectrum compared to the analysis of the individual wavelengths.

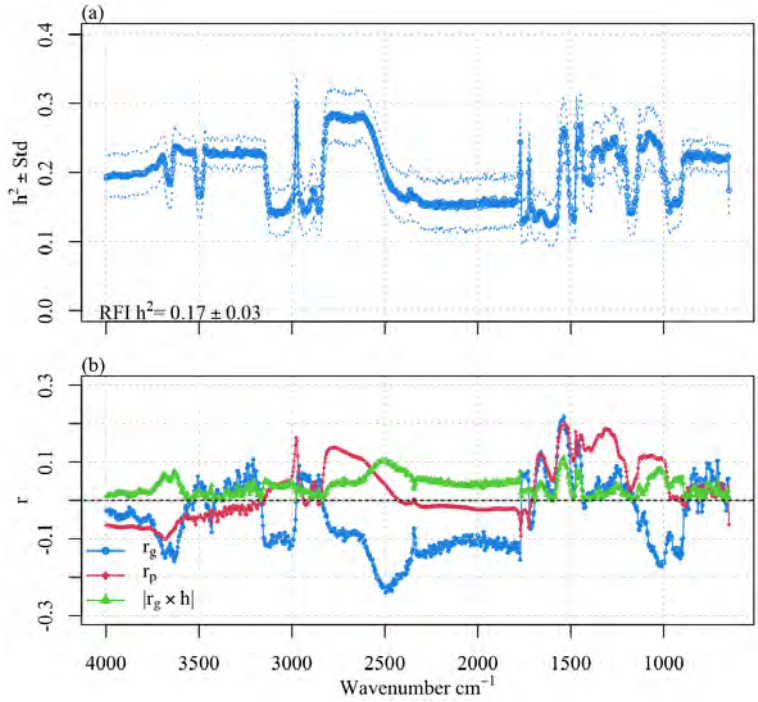


Figure 1. (a) Heritability (h^2) of FTIR wavelengths; (b) Genetic correlation (r_g), phenotypic correlation (r_p), and index accuracy $|r_g \times h|$ between each wavelength and RFI.

Table 1. Heritability (h^2) of predicted Residual Feed Intake, genetic correlation (r_g) and genetic prediction accuracy between predicted and observed RFI using different methodologies.

Method	$H^2 \pm$ Std	$r_g \pm$ Std	Accuracy ± std
PLS	0.15±0.04	0.35±0.19	0.14±0.08
Bayes A	0.16±0.05	0.38±0.19	0.16±0.09
Bayes B	0.19±0.06	0.34±0.20	0.15±0.09
Bayes C	0.18±0.06	0.31±0.20	0.14±0.09
Bayes Lasso	0.16±0.05	0.38±0.19	0.15±0.09
Bayes Ridge Regression	0.16±0.04	0.37±0.19	0.15±0.08

Discussion

The heritabilities for the spectrum were comparable with other studies. Bittante and Cechinato (2013) and Rovere *et al.* (2019) estimated the highest heritabilities from 1,582 to 930 cm^{-1}; whereas we observed several other regions with heritabilities above 0.25 and associated with fat, protein and urea (1,238 to 2,976 cm^{-1}). Regions of the spectra had modest r_g with RFI, this, together with the h^2 of individual wavenumbers, led to individual band accuracies of indirect selection ranging from 0 to 0.1. The wavenumbers most predictive of genetic merit for RFI were located in the fingerprint and in the mid-infrared region (from 3,048 to 1,701 cm^{-1}). When we used all milk-spectra data into prediction equations for RFI we found higher accuracies of indirect selection, with Bayes A achieving the highest accuracy. Since the accuracy of direct selection is $\sqrt{h^2_{RFI}}$, this is a relative efficiency of indirect selection of more than 1/3rd (0.36 using $0.15/\sqrt{0.17}$). This information could be used to derive predictions of RFI at large scale that could be considered as indicator traits for selection to improve feed efficiency.

References

Berry, D.P., and Crowley J.J. (2013) J. Anim. Sci. 91(4):1594-1613. https://doi.org/10.2527/jas.2012-5862

Bittante G., and Cecchinato A. (2013) J. Dairy. Sci. 96(9):5991-6006. https://doi.org/10.3168/jds.2013-6583

Falconer D.S., and Mackay T.F.C. (1996) Introduction to Quantitative Genetics. 4th ed. Pearson Education.

Gaddis K., VanRaden P., Tempelman R., Weigel K., White H., *et al.* (2021) Interbull Bulletin (56), 147-152.

Hansen P.W. (1998) Milchwissenschaft 53(5):251-255.

Hansen P.W. (1999) J. Dairy Sci. 82(9):2005-2010. https://doi.org/10.3168/jds.S0022-0302(99)75437-8.

Kaylegian K.E., Houghton G.E., Lynch J.M., Fleming J.R., and Barbano D.M. (2006) J. Dairy Sci. 89(8):2817-2832. https://doi.org/10.3168/jds.s0022-0302(06)72555-3

Lopez-Cruz, M., Olson, E., Rovere, G., Crossa, J., Dreisigacker, S., *et al.* (2020). Sci. Rep. 10(1), 1-12. https://doi.org/10.1038/s41598-020-65011-2

National Research Council (NRC) (2001). Nutrient Requirements of Dairy Cattle. Natl. Acad. Press, Washington DC, USA.

Pérez P., and de los Campos G. (2014) Genetics 198(2):483-495. https://doi.org/10.1534/genetics.114.164442

Rovere G., de los Campos G., Tempelman R.J., Vazquez A.I., Miglior F., *et al.* (2019) J. Dairy Sci. 102(2): 1354-1363. https://doi.org/10.3168/jds.2018-15109

52. Genetic relationships between weight loss in early lactation and daily milk production until 305 days in Holstein cows

T. Tribout[1*], S. Minéry[2], R. Vallée[2], S. Saille[3], D. Saunier[4], V. Ducrocq[1], P. Faverdin[5] and D. Boichard[1]

[1]Université Paris Saclay, INRAE, AgroParisTech, GABI, 78350 Jouy-en-Josas, France; [2]Institut de l'Elevage, 75012 Paris, France; [3]INNOVAL, CS 10040, 35538 Noyal sur Vilaine, France; [4]FCEL, France Conseil Elevage, 75009 Paris, France; [5]INRAE, AgroCampus Ouest, PEGASE, 35590 Saint-Gilles, France; thierry.tribout@inrae.fr

Abstract

Body mobilization of high-yielding dairy cows is intense after calving and can generate health and fertility troubles. This study aimed to estimate the genetic correlation between body weight loss during the five first weeks of lactation (BWL) and daily milk production (MY24a) in first (L1) and second (L2) lactations Holstein cows. The dataset included 859,020 MY24a and 570,651 body weight (BW) daily phenotypes from 3,989 L1 cows, and 665,361 MY24a and 449,449 BW daily phenotypes from 3,060 L2 cows, recorded in 36 French commercial farms equipped with milking robots including an automatic weighing platform. The genetic parameters of BW and MY24a in L1 and L2 were estimated using a 4-trait random regression model. Genetic correlations between BWL and MY24a ranged from -0.26 to 0.05 in L1, and varied from -0.11 to 0.10 in L2, suggesting that selecting to reduce early body mobilization while maintaining milk production is possible.

Introduction

Dairy cows experience negative energy balance after calving, and mobilize fat and protein reserves to compensate. This mechanism, if excessive, leads to health problems and low fertility and hampers cow longevity and economic profitability. In recent years, a growing proportion of French commercial dairy farms have been equipped with milking robots that record daily milk production and body weight of lactating cows. The objective of the study was to exploit these data to estimate the genetic relationship between early body weight loss, reflecting body mobilization, and daily milk yield, in order to evaluate the possibility of decoupling production and mobilization.

Materials & methods

The data used for this study originated from 36 commercial dairy farms located in Western France. Average herd size was 92 cows [range: 46-189]. These herds are members of the *INNOVAL* DHIS network (http://www.innoval.com).

Data were recorded on 5,372 Holstein cows (4,464 and 3,364 cows in first (L1) and second (L2) lactation, respectively), calved from July 2014 to July 2018. Barns were equipped with Lely milking robots including an automatic weighing platform. At each visit k of a cow in the milking unit, the robot recorded the date and time of the visit, the weight of collected milk ($MILK_{elem,k}$) and the body weight of the cow. In total, more than 4.5 million visits were collected for the 5,372 cows.

For each cow i in lactation j (1 or 2) and each test day t:

- the sum of the $MILK_{elem,i,j,t,k}$ was calculated and standardized into a 24 hours milk yield (MY24) as $MY24(i,j,t)=\Sigma_{k=1,kt}\ (MILK_{elem,i,j,t,k})\times 1440/\Delta_{time}$, where Δ_{time} is the time interval (in minutes) between the last milking of the current day and the last milking of the previous day. Then MY24 was standardized to 4% fat (FC) and 3.3% protein (PC) contents (=MY24a). Daily FC and PC were estimated from monthly measures of each cow with Wilmink's lactation model.

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_52, © T. Tribout *et al.* 2022

- Body weight is affected by change in digestive tract content and therefore by change in feed intake. Feed intake increase in the beginning of the lactation partially hides initial body weight loss. To account for this potential bias, BW was corrected for the change in feed intake between day in milk t and calving (t=0), assuming a 4.5 kg digestive content change per kg dry matter intake change (Faverdin, personal communication):

$$BWci(i,j,t) = BW(i,j,t) - 4.5\times(DMI(i,j,t) - DMI(i,j,0)) \quad (1)$$

where DMI at day in milk t of cow i in lactation j was predicted as:

$$DMI(i,j,t) = \mu(j,t) + \alpha(j,t)\times BW(i,j,t) + \beta(j,t)\times MY24(i,j,t). \quad (2)$$

The prediction equations for $\mu(j,t)$, $\alpha(j,t)$ and $\beta(j,t)$ were previously modelled as second order polynomials on a similar experimental Holstein population for which daily records of BW, MY and DMI were available (Martin *et al.*, 2021).

For each cow and lactation, a 3rd order polynomial was fitted to the lactation MY24a and BWci data, and the outlier phenotypes deviating from prediction by more than 3 standard deviations were discarded.

Pedigree (3 generations) and dates of inseminations of the animals were extracted from the French national database. For each test date of a cow, the corresponding numbers of days in milk (DIM= test date – calving date) and of days carried calf (DCC = test date – date of conception (if necessary) were calculated. The test dates between calving date and successful insemination date were assigned the DCC=0 specific level.

Genetic parameters estimates were obtained using WOMBAT software (Meyer, 2007). Phenotypes in L1 and L2 were considered as different traits. Due to the high number of records it was not possible to directly perform 4-trait random regressions on daily data for MY24a and BWci in L1 and L2. Therefore, a 2-step approach was applied.

Single trait analyses on daily data for BWci and MY24a in L1 and L2. Only data with DIM lower than 305 days and DCC lower than 242 days were considered. Records from animals with unknown parents, older than 3.5 years at first calving, or with less than 30 daily records available were discarded. Herd-test date contemporary groups with less than 4 records were discarded. Numbers of performances and cows after filtering are presented in Table 1.

The model used for single-trait analyses included the fixed effects of herd-test date (HTD), number of milkings per day (1, 2, 3, >3 – only for MY24a) (NMD), DCC as 4-node (DCC 0, 99, 174, 242 days) cubic splines, calving month × DIM interaction as 6-node (DIM 1, 15, 40, 90, 200, 305 days) cubic splines, age at calving (AaC) × DIM interaction as 6-node (DIM 1, 15, 40, 90, 200, 305 days) cubic splines. Five classes of AaC were considered: <25, 25-27, 27-30, 30-33, >33 months of age (moa) in L1, and <37, 37-39, 39-42, 42-45, >45 moa in L2.

Table 1. Numbers of daily phenotypes and cows considered in the single-trait analyses.

	Lactation 1		Lactation 2	
Trait	Daily phenotypes	cows	Daily phenotypes	cows
MY24a	859,020	3,988	665,361	3,060
BWci	570,651	3,439	449,449	2,660

The random permanent environment effect and the additive genetic effect were modeled as second order Legendre polynomials. Finally, the residual variance was considered heterogeneous throughout the lactation period. Seven periods were defined, during which the residual variance was supposed to be constant: DIM 1-7, 8-20, 21-50, 51-100, 101-150, 151-230, and 231-305.

Multi-trait analyses on pre-adjusted weekly means. Daily phenotypes for MY24a and BWci in L1 and L2 were adjusted for DCC, calving month and AaC, as well as NMD for MY24a, and the average performance per week in milk (WM) was calculated for each cow from WM 1 to 44 (wMY24a and wBWci). Cows with less than 9 weekly performances, as well as phenotypes of herd-test week with less than 10 cows were discarded. Numbers of performances and cows remaining after filtering are presented in Table 2.

Then, 4-trait random regression analyses were performed on the weekly averages with models including the fixed effect of herd-test week, the random effects of permanent environment and additive genetic value as second order Legendre polynomials, and a residual whose heterogeneous variance was supposed to be constant on seven periods: WM 1, 2-3, 4-7, 8-13, 14-21, 22-33, and 34-44.

Genetic correlations between early body weight loss and milk production. After adjusting BWci data for environmental effects, the maximum weight loss of cows was achieved on average 31 days after calving in L1 (-33.9 kg) and L2 (-37.3 kg), i.e. in the 5th week of lactation. The genetic variances and covariances between the random regression coefficients of wMY24a and wBWci estimated in the 4-trait analysis were therefore combined together to derive the genetic correlation in L1 and L2 between body weight loss from WM1 to WM5 (BWL=BWci at WM5 – BWci at WM1) and wMY24a across lactation.

Results & discussion

The heritability for wMY24a showed similar trends in L1 and L2, increasing from 0.06 at WM1 to 0.20 at WM12, and then remaining stable. These values are in the low range of the literature for test-day milk yield of Holstein cows (e.g. Druet *et al.*, 2005). BWL showed very small genetic standard deviations compared to BWci (e.g. 5.1 kg for BWL vs 30.2 kg for BWci at DIM 31 in L1). The resulting heritability was low (0.04 in L1 and 0.02 in L2), but consistent with the heritability estimates obtained by Berry *et al.* (2003) for body weight changes from DIM 5 to DIM 60 (0.06) on multiparous Holstein cows.

The genetic correlations (r_g) between wMY24a at adjacent WM were very high, and decreased as the interval increased. However, r_g between early and later lactation (e.g. before WM10 vs after WM15) ranged from 0.3 and 0.8 in L1 and from 0.6 and 0.85 in L2.

The estimates of genetic correlations between BWL and wMY24a along lactations 1 and 2 are presented in Figure 1.

A moderate negative genetic correlation was found on L1 cows between BWL and wMY24a before WM17, being minimum in WM1 (-0.26), and then becoming progressively closer to 0. This suggests that stronger body mobilization in early lactation is genetically associated with higher milk production. This is consistent

Table 2. Numbers of weekly phenotypes and cows considered in the 4-trait analysis.

	Lactation 1		Lactation 2	
Trait	Weekly phenotypes	cows	Weekly phenotypes	cows
wMY24a	122,720	3,728	91,594	2,887
wBWci	102,582	3,253	78,137	2,553

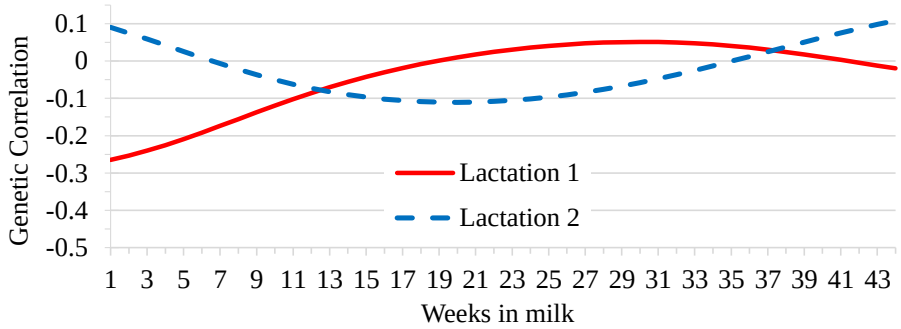


Figure 1. Genetic correlation estimates between body weight loss during the first 5 weeks after calving and adjusted daily milk yield along lactations 1 and 2.

with the results of Berry *et al.* (2003) reporting moderate unfavourable genetic correlations between milk yield and body weight change and body condition score change during the first 2 months of lactation. After ~WM17, BWL and wMY24a appear, however, genetically independent on L1 cows. The genetic correlation between BWL and wMY24a in L2 cows was weaker, declining from 0.1 to -0.1 from WM1 to WM20, and then re-increasing up to 0.1 throughout the second half of lactation.

Milking robots are a valuable source of data. It is unfortunate that we did not have individual daily feed intake of the cows to adjust their BW, leading us to predict it under some assumptions. These predictions seem consistent, even if only approximations. However, keeping this in mind, our results suggest that it is possible to select cows with high milk yield after lactation peak while controlling body mobilization.

Acknowledgements

This work was performed in the framework of the DEFFILAIT project (ANR-15-CE20-0014), funded by the French Agence Nationale de la Recherche (ANR) and ApisGene. INNOVAL and FCEL are gratefully acknowledged for managing the data from milking robots.

References

Berry D.P., Buckley F., Dillon P., Evans R.D., Rath M. *et al.* (2003) J. Dairy Sci. 86:2193-2204. https://doi.org/10.3168/jds.S0022-0302(03)73809-0

Druet T., Jaffrézic F., Ducrocq V. (2005) Genet. Sel. Evol. 37:257-271. https://doi.org/10.1186/1297-9686-37-4-257

Faverdin P., Charrier A., Fischer A. (2017) Proc. of the 8th ECPLF, Nantes, France.

Martin P., Ducrocq V., Gordo D.G.M., Friggens N.C. (2021) Animal 15(2):1001001. https://doi.org/10.1016/j.animal.2020.100101

Meyer, K. (2007) J. Zhekiang Uni. SCIENCE B 8: 815-821. https://doi.org/10.1631/jzus.2007.B0815

53. Influence of body condition score genomic index on performance trajectories over the lactation period in Holstein cows

R. Lefebvre[1*], P. Faverdin[2], S. Barbey[3], J. Jurquet[4], T. Tribout[1], D. Boichard[1] and P. Martin[1]

[1]Université Paris-Saclay, INRAE, AgroParisTech, GABI, 78350, Jouy-en-Josas, France; [2]INRAE, AgroCampus Ouest, UMR 1348 PEGASE, 35590 Saint Gilles, France; [3]INRAE, UE326 Domaine Expérimental du Pin, 61310 Gouffern en Auge, France; [4]Institut de l'Elevage, 49071 Beaucouzé, France; rachel.lefebvre@inrae.fr

Abstract

Body condition score (BCS) is a good estimate of body reserves and could serve as a proxy to prevent some metabolic disorders. Based on 686 lactations from three experimental farms, we investigated to what extent BCS direct genomic values (DGV) from the national evaluation, estimated independently from the phenotypes of this study, reflect profiles of body reserves along lactation, and how it affects body weight, feed intake and milk production. A model including different regressions of BCS DGVs was applied to each trait over the lactation period. Differences in BCS DGV resulted in differences in observed BCS and body weight during the entire lactation. Differences in daily milk production reached 1 kg per day all along lactation. Interestingly, feed intake increased more rapidly with BCS DGV in early lactation but was similar thereafter. The BCS index is a promising tool to limit intense mobilization in early lactation but still needs further investigation.

Introduction

Body reserves are used as an adjustment variable for feed intake during the lactation period. In early lactation, the increase in feed intake is indeed not sufficient to cover the energy requirements for milk production (Banos *et al.*, 2005) and cows therefore enter a negative energy status that leads them to mobilize their body reserves (Friggens *et al.*, 2004). Conversely, later in the lactation, when milk production decreases while intake is still high, cows can recover and build new body reserves. Body condition score (BCS) is a good estimate of the amount of stored fat on the body (Broster and Broster, 1998), despite being based on the human eye which makes it difficult to identify small changes on a short time interval. It has been shown that positive genetic correlations exist between BCS and feed efficiency components such as feed intake and body weight (Manzanilla-Pech *et al.*, 2016). However, if cows with high genetic merit for milk production are reported to mobilize more body reserves than cows with lower genetic merit (Pryce *et al.*, 2001), a too intense mobilization is not favourable as it can lead to metabolic disorders (Weber *et al.*, 2013) and poor fertility. In France, estimated breeding values (EBV) are calculated based on one measure of BCS recorded during the first lactation. In this article, we investigate to what extent this genomic evaluation based on a single phenotypic record per cow may reflect different dynamics of body reserves along the lactation period and how it could be used to limit mobilization in early lactation. We also investigated how differences in BCS evaluation would affect levels and trajectories of other traits such as feed intake or milk production.

Materials & methods

Population resources and feeding management. This experiment was carried out from 2014 to 2020 using Holstein cows from three French experimental farms: 169 Holstein cows from the Le-Pin-Au-Haras INRAE facility, 145 from the Mejusseaume INRAE facility and 121 from the Trinottières experimental farm, representing a total of 686 lactations with parity ranging from 1 to 7. All animals were handled with care in accordance with the French Ministry of Agriculture guidelines for animal research and the applicable European Union guidelines and regulations on animal experiments. Cows were fed individually using an electronic gate feeding system and ear-tag identification. A total mixed ration (TMR) was distributed *ad libitum* (approximately 10% left overs). For Mejusseaume, the TMR was based on maize

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_53, © R. Lefebvre *et al.* 2022

silage and concentrates with an energy value of 0.93 forage unit by kg of dry matter (UFL/kg of DM, 1 UFL equating to 7.12 MJ of net energy for lactation, INRA, 2018). In Trinottières, two TMR were used (one half of the cows for each diet): one based on maize silage and concentrates and the other one including grass silage, with an energy value of 0.96 and 1.00 UFL/kg of DM respectively. Finally, in Le Pin, the TMR was based on maize silage, with grass silage, rapeseed meal, minerals and concentrates, and an energy value of 0.95 UFL/kg of DM.

Phenotyping and trait definition. Cows were milked twice daily and milk yield (MY) was individually recorded at each milking. Body weight (BW) was recorded at each milking using an automatic weighing system. Individual daily feed intake was measured as the difference between distributed and next morning refusals weight (Mejusseaume and Trinottières) or as the sum of intake weighed at each visit of the cow (Le Pin). Dry matter intake (DMI) was calculated daily, based on dry matter content of TMR. Body condition was scored on a 5-point scale with 0.25 steps by the same 2 or 3 trained scorers, monthly or every 2 weeks, depending on the farm. Average BCS of the scorers was used. The study included data from 5-250 days in milk (DIM).

Genotyping and genomic evaluation. All animals were genotyped with the Illumina SNP50k or EuroG10k chip (Illumina Inc., San Diego, CA) from blood samples or ear punches. Genotyping was performed at LABOGENA, Jouy-en-Josas, France. Missing 50k genotypes were imputed with FImpute (Sargolzaei *et al.*, 2014). Direct genomic values (DGV) of experimental animals were obtained by combining their genotypes with the SNP effects estimated by the French national evaluation system. It is worth noting that phenotypes from these three experimental farms are not used in the national evaluation and therefore DGVs used in this study were fully independent of any phenotypic information from the animals. BCS index was expressed in genetic standard deviations, with a zero mean.

Statistical analysis. For each of the four considered traits (MY, DMI, BW and BCS), effect of BCS index (iBC) on trait trajectory over lactation was analysed with a model including: (1) a lactation curve to take DIM into account; and (2) different regressions on BCS index in order to estimate its impact on the trait.

At each day in milk, the following model was applied:

$$y_{it} = \mu + x_i \times f + WIL_t(\text{parity}) + reg_{it}(BC) + \varepsilon_i \quad (1)$$

$$WIL_t(\text{parity}) = a + b \times t + c \times e^{-0.06*t} \quad (2)$$

$$reg_{it}(BC) = \alpha \times iBC + \beta \times iBC \times st + \gamma \times iBC \times st^2 \quad (3)$$

where y_{it} is the performance of animal i at DIM t, μ is the overall mean, f is the vector of fixed effects of farm-date and, for MY and BW, the corresponding DGV as a covariate, x_i is the incidence vector relating cow i to fixed effects, WIL_t(parity) is the intra parity Wilmink model of the lactation curve (primiparous vs multiparous) as described in Equation 2 (Wilmink, 1987) and reg_{it}(BC) is the regression on BCS index as described in Equation 3 with st the standardized DIM ((t-125)/250) and ε_i is the residual.

Analyses were performed with the GLM procedure of the SAS software (SAS institute Inc., 2008).

To illustrate BCS DGV effects, results were averaged by class of DGV values. Five classes were defined as follows: 'BC--' if iBC≤-1; 'BC-' if -1<iBC≤-0.5; 'BC0' if -0.5<iBC<0.5; 'BC+' if 0.5≤iBC<1; 'BC++' if iBC≥1. BC classes included 29, 69, 184, 90 and 63 cows, respectively.

Results

The descriptive statistics regarding MY, DMI, BW and BCS for overall lactations showed no differences in mean DMI, BW and MY between farms. Some slight differences in BCS existed between farms with the highest value for the Trinottières farm (2.7 points), an intermediate value for the Le Pin farm (2.3 points) and the lowest value for the Mejusseaume farm (2.0 points). Trajectories of MY, DMI, BW and BCS are presented in Figure 1. Daily milk rapidly increased by 12 kg between 5 and 50 DIM, then slowly decreased. On the contrary, DMI increased by 8 kg between 5 and 50 DIM but much more slowly than milk. BW decreased by 14 kg during the first two weeks of lactation, and then increased constantly with animals gaining on average 80 kg in total. Regarding BCS, the tendency curve shows a small condition loss of 0.25 points at the beginning of lactation, then a recovery of 0.4 points at the end of lactation.

The effects of BCS index on trait trajectories are presented in Figure 2. Regarding BCS (Figure 2a), positive BC cows always had a greater score than negative BC cows, independent of the stage of lactation. This difference increased along the first 6 months with a 0.4 and 0.8 score difference between extremes in the beginning and mid lactation, respectively. Negative BC cows had a greater milk production (Figure 2b) with barely no difference between extremes at the beginning of lactation, and 1.2 kg per day difference in the middle of lactation. Regarding DMI (Figure 2c), positive BC cows had a greater intake at the beginning of lactation (+1.2 kg a day). After 100 days, we observed no more difference between cows. Negative BC cows were lighter than positive BC cows (Figure 2d) with an increasing difference over lactation from 40 to 65 kg.

Discussion

The small difference observed in the average BCS among the farms is more likely due to slight differences of scale interpretation among the technicians rather than real differences of body reserves, as there was no cross-validation of scores nor common training of technicians. However, these differences were accounted for in the fixed effects of the models.

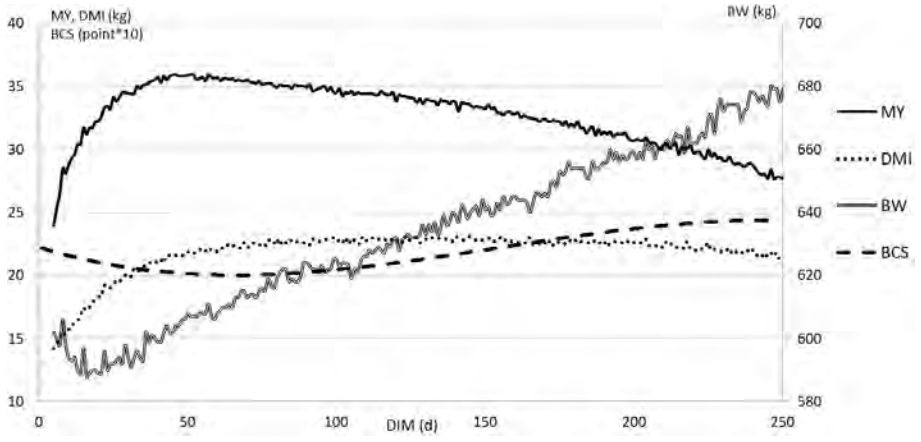


Figure 1. Trajectories of milk yield (MY), dry matter intake (DMI), body condition score ×10 (BCS) and body weight (BW) over days in milk (DIM).

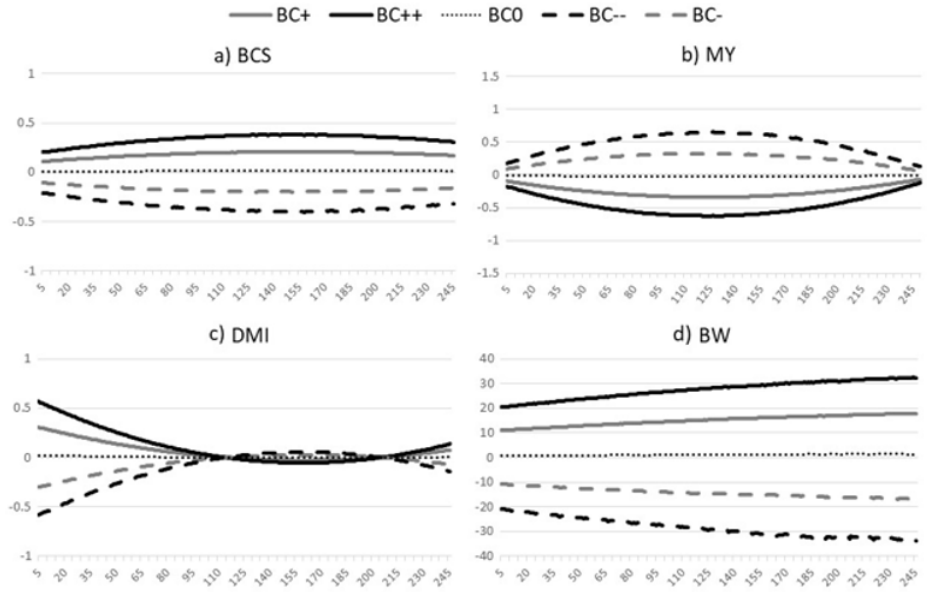


Figure 2. Effect of body condition score (BCS) index on trajectories of (a) BCS (points), (b) milk yield (MY, kg), (c) dry matter intake (DMI, kg) and (d) body weight (BW, kg) from 5 to 250 days in milk.

Observed differences in BCS were in agreement with BCS DGV but a bit lower at the beginning of the lactation. These differences in BCS DGV also translated into moderate differences of BW increasing all along the lactation. Changes of MY per day are minimal (1 l maximum per day between extreme groups) but cumulative. DMI is influenced only in early lactation. Therefore, the BCS index is a promising tool to limit intense mobilization in early lactation but still needs further investigation.

References

Banos G., Brotherstone S., and Coffey M.P. (2005) J. Dairy Sci. 88:2616–2623. https://doi.org/10.3168/jds.S0022-0302(05)72938-6.

Broster W.H., and Broster V.J. (1998) J. Dairy Res. 65:155–173. https://doi.org/10.1017/s0022029997002550.

Friggens N.C., Ingvartsen K.L., and Emmans G.C. (2004) J. Dairy Sci. 87:988–1000. https://doi.org/10.3168/jds.S0022-0302(04)73244-0.

INRA (2018) Alimentation des ruminants. Editions Quae. Versailles, France.

Manzanilla-Pech C.I.V., Veerkamp R.F., Tempelman R.J., van Pelt M.L., Weigel K.A., *et al* (2016) J. Dairy Sci. 99:443–457. https://doi.org/10.3168/jds.2015-9727.

Pryce J.E., Coffey M.P., and Simm G. (2001). J. Dairy Sci. 84:1508–1515. https://doi.org/10.3168/jds.S0022-0302(01)70184-1.

SAS institute Inc. (2008) SAS/STAT 9.2 User's Guide. SAS institute Inc., Cary, NC.

Sargolzaei M., Chesnais J. and Schenkel F. (2014) BMC Genom. 15:478. https://doi.org/10.1186/1471-2164-15-478.

Weber C., Hametner C., Tuchscherer A., Losand B., Kanitz E., *et al* (2013) J. Dairy Sci. 96:165–180. https://doi.org/10.3168/jds.2012-5574.

Wilmink J.B.M. (1987) Livest. Prod. Sci. 16:335–348. https://doi.org/10.1016/0301-6226(87)90003-0.

54. Feed efficiency and resource allocation trade-offs: theory, evidence and prospects

F. Douhard*, R. Rupp and H. Gilbert

GenPhySE, Université de Toulouse, INRAE, ENVT, 31326 Castanet-Tolosan, France; frederic.douhard@inrae.fr

Abstract

The resource allocation theory has long been of interest in livestock as a framework to explain the correlated responses to selection for high productivity. Increasing limitations on feed resources and genetic improvement of feed efficiency (FE) should further promote allocation constraints. Yet, predicting those consequences for breeding critically depends upon the possibility to get testable predictions from the current framework. Here we assessed the changes in energy allocation when selecting for FE, and the consequences of those changes on health and reproduction. For this, we used two approaches; a survey of selection experiments among various species and a case study in sheep focusing on a potential trade-off between FE and response to an infectious challenge (parasite infection). Overall, results provide weak evidence for the current allocation framework. We outline several directions to better appraise the consequences of breeding for FE in increasingly challenging environments.

Introduction

Global pressures on feed resources require livestock that consume less feed to produce under controlled conditions but also emphasize the challenge of producing when feed quantity or quality can drop unexpectedly. Although those issues well translate into separate researches on new breeding objectives for FE and for robustness, the links between those two types of traits remain unclear whereas most of their mechanisms should share a common nutritional basis. Moreover, those links may have key implications in terms of trade-offs and genotype-by-environment interactions. A framework representing how nutrition affects multiple traits responses to selection would be thus insightful. For this purpose, the evolutionary theory of resource allocation has become of interest in animal breeding since it provides a simple way to consider the responses to selection according to the feed availability and to the competing demands for nutrient among the different functions of an organism (Beilharz *et al.*, 1993; Rauw, 2009). Yet, this framework only had a limited impact on the study of FE so far. Moreover, some responses to selection for FE have called into question the occurrence of resource allocation trade-offs (Rauw, 2012; Gilbert *et al.*, 2017). Although the resource allocation theory has been introduced in the livestock context since the 90's, its applicability to these conditions remains seldom questioned and tested. The fast-growing body of research on FE thus provides new opportunities to assess the mechanistic bases of resource allocation trade-offs and how they might shape selection responses.

Why should selection for feed efficiency lead to energetic trade-offs?

Recently, we have reconsidered the theoretical background to predict trade-offs, focusing on energy allocation (Douhard *et al.*, 2021). Our framework is based on the classic 'Y model' of a resource acquisition and allocation (Van Noordwijk and De Jong, 1986) applied to animals typically selected under non-limiting resources. For each animal, the daily rate of metabolizable energy intake is split into one part allocated to productivity ('P'; energy retained as new biomass such as muscle, fat, milk, egg or wool plus the associated energy of biosynthesis dissipated as heat), and the other part allocated to activity and self-maintenance. Allocation to maintenance can be approximated by resting metabolic rate (RMR) and is supposed to maximize the chance of survival and future reproduction (Figure 1a). Most undesirable side effects of selection for high productivity are thus thought to stem in the extent to which RMR is reduced in the total energy budget (once its covariation with body size has been accounted for). However, a same selective

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_54, © F. Douhard *et al.* 2022

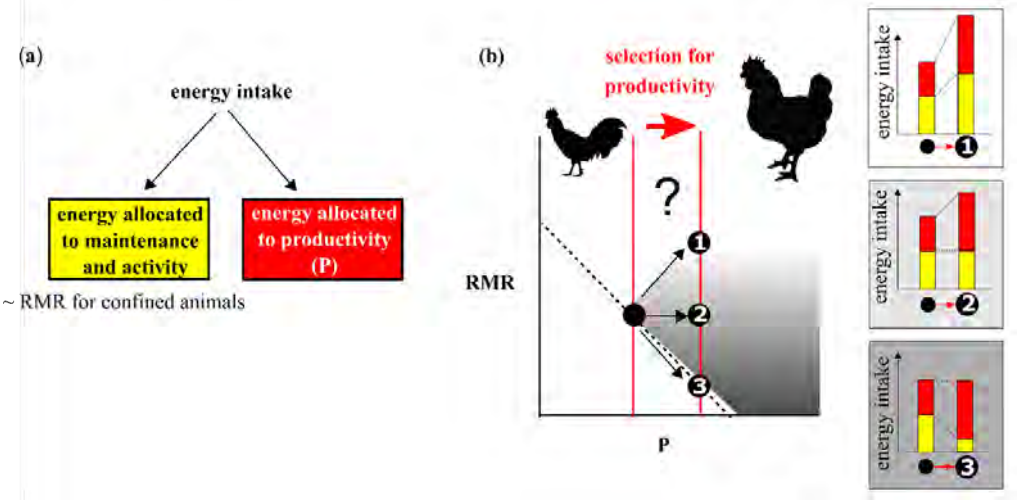


Figure 1. The 'Y model' of energy allocation applied to livestock (a) and the different paths through which increased productivity could translate in terms of changes in energy allocated to productivity or maintenance (b). Maintenance can reasonably be approximated by resting metabolic rate (RMR) for confined animals. Grey shade reflects the trade-off intensity. Adapted from Douhard *et al.* (2021).

increase in productivity could be mediated through different changes in energy allocation depending on the relationships between productivity, maintenance and intake (Figure 1b). A higher rate of biosynthesis could require higher maintenance costs (path no.1 where P and RMR are positively associated) or could be simply fuelled through higher intake (path no.2 where RMR and P are independent). Alternatively, constraints on intake could favour a reduced energy allocation to maintenance (path no.3 where RMR and P are negatively associated). All those cases can lead to an increase in P relatively to energy intake and thus improve FE. However, if an energy allocation trade-off occurs then selection strategies that promote a reduction in RMR should increase the risk of an impaired health or reproduction.

What is the evidence for energy allocation trade-offs?

Review of comparisons between breeds or lines selected in favourable environments. First, to assess the relationships between maintenance, productivity and feed intake in the context of livestock and artificially selected laboratory mammals and birds, we surveyed studies that compared RMR between breeds or lines of contrasting yields when feed intake was not constrained (Figure 2a). Only few studies provided some support for an energy allocation trade-off (path n°3 in Figure 1b), in particular during growth in poultry. In contrast, during reproduction, increased maternal productivity seemed often associated with an increased RMR (path n°1 in Figure 1b). Then, to explore the correlated effects of a reduced RMR on different aspects of health or reproduction, we reviewed selection experiments on FE or RMR (Figure 2b). Indeed, responses to selection for FE or for RMR were consistently and positively related which suggests that selection for FE generally leads to some genetic decrease in RMR (but also in some behavioural activities), even though feed intake is not restricted. Despite this constraint on metabolism, we found little evidence for unfavourable effects on components of health (e.g. oxidative stress, neonatal survival, humoral or cell-mediated immunity) and reproduction (e.g. sperm traits, litter size, milk output).

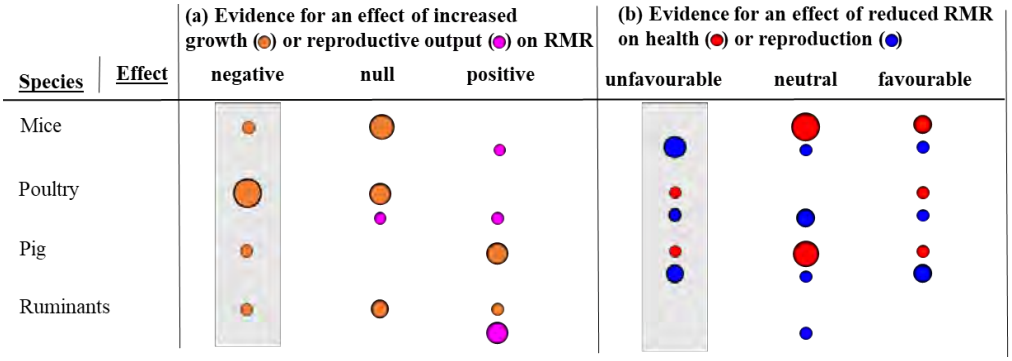


Figure 2. Evidence for association between resting metabolic rate (RMR) and selection for growth or reproductive output when feed intake is not restricted (a), and effects of a genetically reduced RMR on health or reproduction (b). The size of the circle indicates the number of supporting comparisons (1≤n≤5) between selected lines or breeds. In (b) when different components of health or reproduction were investigated within a same pair of lines there were considered as different comparisons. The scenario corresponding to an energy allocation trade-off (path n°3 in Figure 1b) is highlighted in grey.

Case study in sheep selected for response to an infectious challenge or for feed efficiency. To better detect a potential energy allocation trade-off, it may be necessary to go beyond the simple Y model (Figure 1a) notably to focus on a component of maintenance that can be specifically recruited under challenging conditions, e.g. a specific immune response to infection. For this, we studied the potential trade-off between FE and resistance to gastrointestinal parasitism based on an infectious challenge (single-dose infection with *Hæmonchus Contortus*) in growing lambs from Romane meat-sheep (experimental approval D18-174-01). In a first trial, lambs were from lines divergently selected on parasite resistance (resistant (R) vs susceptible (S); Sallé *et al.*, 2021), whereas in another trial they were from lines divergently selected for FE (low Residual Feed Intake (RFI-) vs high (RFI+); Tortereau *et al.*, 2020) from the same flock. Within each trial, direct responses to selection were consistent: fecal worm egg count of was much lower in the R line than in the S line, and FE was better in the RFI- line than in the RFI+ line. Contrary to our expectation, no evidence for a trade-off was found. Between the RFI lines, the difference in FE was maintained during infection, and no difference in parasite resistance was detected. In the R and S lines we even found some evidence for a synergy between FE and parasite resistance. Voluntary feed intake was lower for the R line than for the S line during infection, which also led to a lower RFI. Later on, we re-infected the R and S lines during the first peripartum and tested the effect of a feed restriction (protein). This restriction led to a similar decrease in body weight and condition between lines, with no interaction with parasite resistance.

Limitations and future directions

Based on our findings, the consequences of breeding for FE look much more difficult to predict than one could anticipate from the energy allocation framework alone. First, although the genetic antagonisms observed between productivity and health or fertility are often considered as a result of resource allocation trade-offs, we did not find strong support that those trade-offs can be underpinned by changes in energy allocation. Second, even when energy allocation to maintenance is reduced during selection for FE, the expected unfavourable effects on health or reproduction were far from clear. Lastly, our case study neither indicates that selection for a particular health component (parasite resistance) deteriorates FE during a challenge, nor vice-versa. Overall, we thus confirm a lack of direct or mechanistic support for the energy allocation framework when applied to livestock.

However, there were at least three main limitations when assessing this framework. First, the energy allocation to maintenance remains difficult both to estimate and to interpret. In particular, dissecting the biological basis of FE or RMR does not simply translate into the resource acquisition-allocation framework as various maintenance processes also contribute to feed acquisition or to production. Second, the aspects of energy budgets that fundamentally constrain selection are still very elusive. Third, while the consequences of a reduced allocation to maintenance on fitness-related traits should be considered on the long-term, most studies of FE have focused on reproduction or health traits in young adult animals, which may have led to inadequate measurements to correctly assess the resource allocation framework.

Beyond those limitations, it becomes increasingly clear that resources allocation constraints do not provide a sufficient explanation for the occurrence of trade-offs and that other mechanisms underpinning unfavourable genetic correlations should be considered (e.g. nutrient-sensing signalling). Those considerations should contribute to a more integrative framework to orientate breeding objectives according to the agroecological context.

Acknowledgements

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 772787 (SMARTER).

References

Beilharz R.G., Luxford B.G., and Wilkinson J.L. (1993) J. Anim. Breed. Genet. 110(1-6):161–170. https://doi.org/10.1111/j.1439-0388.1993.tb00728.x

Douhard F., Douhard M., Gilbert H., Monget P., *et al.* (2021) Evol. Appl. 14:2726–2749. https://doi.org/10.1111/eva.13320

Gilbert H., Billon Y., Brossard L., Faure J., Gatellier P., *et al.* (2017) Animal 11(9):1427–1439. https://doi.org/10.1017/S175173111600286X

Rauw W.M. (2009) Resource allocation theory applied to farm animal production. CABI International, Wallingford, UK.

Rauw W. M. (2012). Feed efficiency and animal robustness. In: R. A. Hill (ed) Feed Efficiency in the Beef Industry. John Wiley & Sons, Ames, Iowa, USA, pp. 105–122.

Sallé G., Deiss V., Marquis C., Cortet J., Serreau D., *et al.* (2021) Evol. Appl., 14(11): 2591-2602. https://doi.org/10.1111/eva.13294

Tortereau F., Marie-Etancelin C., Weisbecker J.-L., Marcon D., Bouvier F., *et al.* (2020) Animal 14(4), 681–687. https://doi.org/10.1017/S1751731119002544

Van Noordwijk, A.J., and De Jong, G. (1986). Am. Nat. 128(1):137-142. https://doi.org/10.1086/284547

55. Mechanistic-based prediction of selection response on resilience and feed efficiency traits in dairy cattle

A. Bouquet[1*], M. Slagboom[1], J. Thomasen[2], N.C. Friggens[3], M. Kargo[1] and L. Puillet[3]

[1]QGG Center, Aarhus University, Blichers Allé 20, 8830 Tjele, Denmark; [2]VikingGenetics, Ebeltoftvej 16, 8960 Randers SØ, Denmark; [3]Université Paris-Saclay, INRAE, AgroParisTech, UMR MoSAR, 75005 Paris, France; alban.bouquet@qgg.au.dk

Abstract

The objective of this study was to present a new methodology to predict selection response on resilience and feed efficiency (FE) in dairy cattle. This approach combines genetic and mechanistic modelling to describe the biological mechanisms underlying these traits. A dairy cattle breeding scheme was simulated considering a non-limiting nutritional environment and two different breeding goals focusing either on milk production or FE. Selection response was predicted within the non-limiting environment but also for a prospective low-input system (LS). Predictions obtained with conventional and mechanistic-based methods were consistent for milk production, body weight and FE within the non-limiting environment. However, genetic trends predicted for fertility were different. Selection response achieved on milk production was much smaller in the LS than in the breeding nucleus due to the increased nutritional constraint. The breeding goal with emphasis on FE enabled a better transfer of genetic gain to the LS environment.

Introduction

Breeding for resilience and feed efficiency (FE) in dairy cattle is difficult because these traits result from a subtle balance between biological functions aiming at maintaining a cow's ability to produce and reproduce. This balance is also highly dependent on the nutritional environment: trade-offs between production and fitness traits are exacerbated when energy intake is limiting (Puillet *et al.*, 2021). With climate change and societal demands for expanding grass-based production systems, nutritional environments might be more uncertain and diverse in the future. When defining novel breeding strategies on resilience and FE, it is critical to anticipate these changes and select cows that can cope with various production systems. This paper describes a new mechanistic-based method (MM) combining genetic and mechanistic modelling to predict selection response on cow milk production and FE traits. This approach can be used within a given breeding environment but also facilitates predicting selection response that would be expected in more challenging environmental scenarios.

Materials & methods

General approach to predict selection response. In this study, we capitalize on the AQAL bioenergetic model to simulate phenotypic trajectories of milk production and FE traits over the lifetime of dairy cows depending on energy available in the environment (Puillet *et al.*, 2016; 2021). This dynamic mechanistic model of a single cow is grounded on the resource acquisition and allocation (AA) theory to describe cows' responses to nutritional challenges. Variability in the response observed at the population level is induced by assuming genetic and phenotypic variance for a set of four input parameters describing energy AA strategies of cows. Acquisition parameters correspond to the maximal intake of a non-reproducing cow (basal acquisition: BasAcq) and the increase of intake during lactation (lactation acquisition: LactAcq) expected in a non-limiting environment. Allocation parameters correspond to the rate of transfer of energy from growth to survival (RTGS) defining the trade-off between structural mass and body reserves, and the allocation to lactation (LactAll) defining the energy investment in milk production. Mean values of these four parameters were determined by a calibration procedure using real data as described by Puillet *et al.* (2021). The effect

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_55, © A. Bouquet *et al.* 2022

of the nutritional environment on the variability of traits was modelled through stochastic simulations. To predict genetic gain, it was assumed that selection on usual production and functional traits changes the mean genetic level of energy acquisition and allocation variables. Thus, updating AA parameters considering this correlated response is a way to simulate new performance datasets mimicking the effects of selection with AQAL. The method comprises three steps. First, datasets with a pedigree structure were simulated with AQAL considering the same AA parameters as Puillet *et al.* (2021) and a non-limiting nutritional environment typical of breeding herds. This environment corresponded to the 'high & stable' (HS) scenario in Puillet *et al.* (2021). Simulated data were used to estimate genetic parameters between simulated traits and AA input traits. Second, the ADAM breeding scheme simulation tool (Pedersen *et al.*, 2009) was used to estimate the correlated selection response expected on AA traits for a given breeding goal and a typical dairy cattle breeding scheme structure. Finally, milk production and FE traits were simulated by updating the AQAL AA input parameters with the correlated selection response to estimate the change in performance due to selection.

Considered traits and genetic parameters. Four traits were analyzed in this study and were estimated for primiparous cows: milk production (MP), body weight at calving (BW), energy conversion efficiency (ECE) and the interval between first mating and conception (IFC). The four AA input traits were assumed to have a heritability of 0.35, a phenotypic coefficient of variation of 10% and to be genetically uncorrelated as assumed by Puillet *et al.* (2021). Genetic parameters were estimated between AA traits and simulated traits using restricted maximum likelihood applied to bivariate linear animal mixed models (Table 1). Genetic correlations were close to 1 between BasAcq and BW, as well as between LactAll and ECE. Hence, we considered these pairs of traits as single traits.

Simulation of breeding scheme. A large-scale breeding nucleus was simulated over a 30-year period with ADAM (Pedersen *et al.*, 2009). It comprised 20,000 females equally distributed in 200 herds and mated with 100 sires each year. True breeding values (TBV) and phenotypes were sampled using genetic parameters in Table 1. Milk production, BW and IFC were recorded on all cows. Genomic selection was simulated using the pseudo-genomic selection approach (Buch *et al.*, 2012). The accuracy of genomic breeding values (GEBV) was estimated using selection index theory to combine information from bull and cow reference populations (Buch *et al.*, 2012). We assumed a reference population of 10,000 bulls with 100 daughters each, 50,000 cows with performance for MP, BW, IFC, and 5,000 cows with ECE phenotypes. The GEBV accuracy of candidates without performance was 0.68 for MP and BW, 0.60 for ECE and 0.58 for IFC. Using simulated traits, pseudo-genomic phenotypes and pedigree, GEBV were predicted using DMU software (Madsen and Jensen, 2013).

Each year, the best 4,000 male and 4,000 female calves were assumed to be genotyped based on parental breeding values. After genotyping, the best 100 1-year old males were selected to be used as sires for one

Table 1. Genetic parameters for acquisition, allocation and simulated traits (heritability on the diagonal, genetic and phenotypic correlations above and below the diagonal).

	BasAcq / BW	LactAcq	RTGS	LactAll / ECE	MP	IFC
BasAcq / BW	**0.35**	0.00	0.00	0.00	0.50	-0.23
LactAcq	0.00	**0.35**	0.00	0.00	0.38	-0.12
RTGS	0.00	0.00	**0.35**	0.00	0.24	-0.64
LactAll / ECE	0.00	0.00	0.00	**0.35**	0.72	0.56
MP	0.52	0.39	0.23	0.73	**0.34**	0.12
IFC	-0.05	-0.02	-0.13	0.11	0.06	**0.01**

year. Within each herd, the best 100 females aged from 1 to 5 years were selected using estimated breeding values. Selection decisions were based on a total merit index. Two breeding goals were defined to reflect a different balance between production and functional traits. In the first breeding goal, called M-FER, equal weight was given to MP and IFC. In the second breeding goal, called EFF-FER, equal weight was given to ECE and IFC. The first 20 years were considered as burn-in period and were discarded. Annual genetic gain was predicted for each trait by regressing mean TBV of selection candidates on their birth year over the last ten years. All presented results were averaged based on 30 replicates.

Genetic gain predictions based on the mechanistic model. Mean values of the four AA parameters were updated considering the correlated selection response expected over a 20-year period. Changes in heritability and phenotypic variance of AA traits due to selection were also estimated as observed in ADAM stochastic simulations. Changes in genetic covariance between AA traits were small (≤0.04 in absolute value) and were ignored. Selection response was estimated for each trait on an annual basis as the change in performance simulated with the initial and updated sets of AA parameters within the non-limiting HS environment.

Selection response that would be expressed in a low nutritional environment following selection in the HS environment was estimated in the same way but simulating datasets with AQAL considering the 'low & stable' (LS) environment in Puillet *et al.* (2021). In this environment, the nutritional constraint was substantial: dry matter intake was on average 14.7% lower than total potential acquisition.

Results & discussion

Correlated selection responses on AA traits predicted for the two breeding goals are presented in Table 2. With the M-FER breeding goal, high genetic gain was achieved on basal acquisition because this trait was positively correlated to MP and IFC. Indeed, when access to feed is non-limiting, increasing feed intake by selection is an option to improve both MP and IFC. Therefore, although a high weight was given to MP, genetic gain on LactAll was only limited. Selecting on the EFF-FER breeding goal led to a more balanced genetic gain on the four AA traits, with a shift in genetic gain from BasAcq to LactAll. In both scenarios, similar positive genetic gain was achieved on RTGS meaning that selection favors cows allocating more energy for body reserves accretion than growing structural mass.

Selection responses predicted for simulated traits are presented in Table 3. With the conventional method, high genetic gain was estimated on MP, BW and IFC for the M-FER breeding goal in the HS environment. Genetic gain on ECE was more limited. With the EFF-FER breeding goal, genetic gain was higher on ECE and more limited on BW and IFC. This was consistent with genetic correlations between AA input traits and simulated traits.

Selection response predictions were consistent across methods and breeding goals within the HS environment for MP, BW and ECE.

Table 2. Mean annual genetic gain[1] (in genetic standard deviation units) estimated for acquisition and allocation traits in the breeding environment for the two breeding goals.

Breeding goal	BasAcq	LactAcq	RTGS	LactAll
M-FER	0.34	0.13	0.20	0.09
EFF-FER	0.22	0.14	0.21	0.23

[1] Standard error of the mean ranged from 0.002 to 0.003 for all predictions.

Table 3. Annual genetic gain (expressed in genetic standard deviation units) predicted for simulated traits with the conventional (CONV) and mechanistic-based (MM) methods in the breeding environment (HS) and the low nutritional scenario (LS).

	CONV prediction – HS		MM prediction – HS		MM prediction – LSI	
	Breeding goal		Breeding goal		Breeding goal	
Trait	M-FER	EFF-FER	M-FER	EFF-FER	M-FER	EFF-FER
MP	0.33	0.38	0.34	0.41	0.15	0.28
BW	0.34	0.22	0.30	0.19	0.23	0.16
ECE	0.09	0.23	0.08	0.22	0.09	0.23
IFC	-0.21	-0.11	-0.01	0.06	0.00	0.06

Indeed, differences in annual genetic gain were lower than 0.04 genetic standard deviation units (σ_g). Differences were larger on IFC. Annual genetic gain was reduced from -0.21 to -0.01 σ_g in the M-FER scenario and from -0.11 to 0.06 σ_g in the EFF-FER scenario, i.e. a zero or even unfavorable trend was predicted on fertility in the M-FER and EFF-FER scenarios with the MM approach.

Genetic gain that would be expressed in the LS environment given selection in the HS environment was estimated with the MM approach (Table 3). In the LS environment, selection response predicted on MP and BW with the M-FER breeding goal was reduced by 56 and 23% relative to the HS environment. The reduction in genetic gain was more limited in the EFF-FER scenario (-32% and -16% for MP and BW). Indeed, increasing basal acquisition by selection in the HS environment led to further increase the nutritional constraint in the LS scenario due to increased maintenance requirements. This also explains the larger reduction in selection response on BW and MP with the M-FER breeding goal due to faster genetic progress on basal acquisition. Hence, a larger proportion of genetic gain was transferred from the high to the low nutritional environment with the EFF-FER breeding goal.

In conclusion, this study showed the feasibility of combining mechanistic and genetic modelling to predict selection response on milk and FE traits in dairy cattle. Predictions were consistent for MP and FE traits within the breeding environment. Compared to the standard approach, the new method also casts light on the AA strategies used to achieve genetic gain. Furthermore, it enabled the estimation of genetic gain that could be transferred to a prospective environment considering both genetic trends and the magnitude of physiological trade-offs on production and functional traits.

Acknowledgements

This work has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 727213 (GenTORE).

References

Buch L.H., Kargo M., Berg P., Lassen J., and Sørensen, A. C. (2012) Animal 6(06):880-886. https://doi.org/10.1017/S1751731111002205

Madsen P., and Jensen J. (2013) DMU: A package for analyzing multivariate mixed models. User's guide – Version 6, release 5.2. QGG Center, Aarhus University, Tjele, Denmark.

Pedersen L.D., Sørensen A.C., Henryon M., Ansari-Mahyari S., and Berg P. (2009) Livest. Sci. 121(2-3):343-344. https://doi.org/10.1016/j.livsci.2008.06.028

Puillet L, Réale D., and Friggens N.C. (2016) Genet. Sel. Evol. 48:72. https://doi.org/10.1186/s12711-016-0251-8

Puillet L., Ducrocq V., Friggens N.C., and Amer P.R. (2021) J. Dairy Sci. 104(5):5805–5816. https://doi.org/10.3168/jds.2020-19610

56. Towards the characterisation of animal robustness by dynamic energy allocation indicators in fattening pigs

G. Lenoir[1,2,3*], R. Muñoz-Tamayo[1], L. Flatres-Grall[2], I. David[3] and N.C. Friggens[1]

[1]Université Paris-Saclay, INRAE-AgroParisTech, UMR Modélisation Systémique Appliquée aux Ruminants, 75005 Paris, France; [2]AXIOM, 37310 Azay-sur-Indre, France; [3]GenPhySE, INRAE, ENVT, Université de Toulouse, 31326 Castanet-Tolosan, France; glenoir@axiom-genetics.com

Abstract

The objectives of this study were to investigate the possibility of characterising animal robustness by using indicators based on the dynamics of energy allocation of the animal and to determine their genetic parameters. A total of 2,140 pigs, from the Piétrain NN Français line, were raised at the AXIOM boar testing station. This farm was equipped with automatic feeding system, recording individual weight and feed intake at each visit. We used a dynamic linear regression model to characterize the evolution of the allocation factor (α_t) between cumulative net energy available, estimated from feed intake, and cumulative weight gain during fattening period. The variance of α_t, that could be interpreted as an indicator of the response of the animal to perturbations/stress, showed moderate heritability (0.27±0.08). Our perspective is to further decompose the allocation factor into components to better characterise the robustness phenotype.

Introduction

Livestock farming faces new challenges related to climate change and societal concerns, e.g. animal welfare and use of antibiotics. These challenges require having animals able to adapt to these new conditions, which implies an improvement of robustness while maintaining a high level of production. There is no real consensus on the definition of robustness as well as on the ways to phenotype it. Following the definition of robustness adapted to the context of artificial selection of Knap (2005), we have recently proposed robustness scores estimated from phenotypes commonly available on farm (Lenoir *et al.*, 2021). However, these scores had low heritability indicating the need to find other indicators that can be used for genetic improvement. Development of new technologies in livestock production, such as automatic feeding system, allow recording of longitudinal data over a period (weight, feed intake, feeding duration). Several studies have used such data to quantify robustness and resilience indicators based on deviation between potential of production of an individual and its observed production (Nguyen-Ba *et al.*, 2020; Revilla *et al.*, 2022). Definition and modelling of individual potential are challenging issues in these approaches. The first objective of this study was to investigate the possibility of characterising animal robustness by using indicators based on the dynamics of energy allocation of the animal. Our rationale is that these indicators should reflect the ability of an animal to express or adapt its production potential in the face of changes in the environment relative to other animals that have been raised under the same conditions. To identify these indicators, we followed a modelling approach applied on longitudinal measurements of body weight and feed intake of fattening pigs (35 to 110 kg live weight). The second objective was to estimate the genetic determinism of the resulting indicators and to compare them with other robustness traits.

Materials & methods

Animals. Pigs from Piétrain NN Français paternal line, free from halothane-sensitivity, of the Axiom company were used in this study. The animals considered in this study were 2,140 entire males raised from January 2019 to April 2021 at the AXIOM boar testing station and born from two different farms. They entered the boar testing station after weaning and were raised in quarantine rooms and in post-weaning rooms for 7 weeks. Then, animals were transferred to fattening rooms when they were 75.3±3.4 days of age (34.5±6.2 kg BW). They were kept in fattening rooms during 74.8±4.0 days until the individual

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_56, © G. Lenoir *et al.* 2022

candidate test at around 149.7±4.1 days of age (108.8±11.5 kg BW). Fattening rooms were equipped with automatic feeding system (AFS) Nedap pig performance testing feeding station (Nedap N.V.; Groenlo, the Netherlands).

Data. During the fattening period, BW (kg) and feed intake (FI; kg) were recorded each time the animal went into the AFS. Other measurements made during the individual test were: average ultrasonic backfat thickness (BF100) and ultrasonic longissimus dorsi thickness (LD100), both adjusted to 100 kg liveweight. Individual average daily gain (ADG) and daily feed intake (DFI) were calculated. The feed conversion ratio (FCR) was calculated as the ratio between the total FI and the weight gain during the fattening period. A phenotype to characterize the robustness (R2) of the candidates was determined from the visual observation performed during the individual test based on Lenoir *et al.* (2021). The binary trait R2 differentiated animals that were selectable (score 1), from those that were dead or not selectable (score 0). The ABC index developed by Revilla *et al.* (2021) was calculated using weight measured by AFS for each animal alive at the end of the fattening period. The trait ABC was the accumulated difference (area) between a theoretical unperturbed growth curve and the perturbed curve.

After pre-treatment process of data recorded by AFS, the weight (W_t; kg) and the feed intake (F_t; kg) for each fattening day were estimated, where t was time in days since the transfer to fattening room. Then, F_t was converted in net energy intake (EI_t) by using a factor of 9.85 MJ/kg of NE. The net energy available at day t (NEA_t) was the difference between EI_t and the net energy maintenance requirements at day t (MR_t). The value of MR_t was estimated according to Noblet *et al.* (2016), $MR_t = 1.05 \times W_t^{0.6} \times 0.74$.

Modelling growth allocation. We could represent the link between cumulative NEA (CNEA) and cumulative weight gain (CW) by a standard linear regression for each animal (1).

$$CW_t = \alpha\, CNEA_{t-1} + \varepsilon_t, \qquad \varepsilon_t \sim N(0,\sigma^2) \tag{1}$$

Where CW_t is the time series of cumulative weight gain (kg) at day t; $CNEA_{t-1}$ is the cumulative net energy available (MJ) at day t-1; α is an allocation factor of energy to weight gain. We assumed in this study that a perturbation is linked with a change in the allocation of energy available to the growth. Therefore, the relationship between CW and CNEA evolves over time. To characterize this evolution, we used a dynamic linear regression model (Petris *et al.*, 2009) built with two equations: an observation Equation 2, relating cumulated weight gains and cumulated NEA, and a system Equation 3, describing the changes in α_t (unobserved state variable) from day to day according to a stochastic process.

$$CW_t = \alpha_t CNEA_{t-1} + v_t, \qquad v_t \sim N(0,\sigma_v^2) \tag{2}$$

$$\alpha_t = \alpha_{t-1} + w_t, \qquad w_t \sim N(0,\sigma_w^2) \tag{3}$$

The model has been built using the function dlmModReg of the package dlm of R (Petris *et al.*, 2009). It included two unknown parameters (σ_v^2 and σ_w^2) that were estimated by maximum likelihood with the function dlmMLE. The values of α_t were calculated independently for each animal with a Kalman smoother algorithm (function dlmSmooth). For each individual, using the estimated α_t across time, we further calculated two traits: the mean (MA) and the variance (VA).

Genetic parameters estimation. Genetic parameters were estimated using ASReml 3.0 software (Gilmour *et al.*, 2009) using the restricted maximum likelihood method (REML). The fixed effect was the fattening group (33 levels). In addition, two random effects were included in the model: common litter

effect and genetic additive effect of the animal. Traits ABC, MA and VA were standardized on a mean of 0 and a standard deviation of 1. Firstly, variance and covariance components were estimated with a 4-trait linear animal model including traits under selection (ADG, FCR, BF100 and LD100). Secondly, to estimate heritability for each non-selected trait (MA, VA, DFI, ABC and R2) and their genetic correlations with the traits under selection, 5-trait linear animal models including the 4 traits under selection and 1 trait to be estimated were used. R2 was considered as a continuous phenotype. Thirdly, to estimate genetic correlations between MA and VA with the other non-selected traits, 2-trait linear animal models were performed. The pedigree contained 3,944 animals across 24 generations.

Results

For traits ABC, MA and VA, descriptive values are presented in Table 1 (raw values before standardization). For the traits under selection (ADG, FCR, BF100 and LD100) and DFI, heritability estimates were moderate, from 0.31±0.07 to 0.41±0.07, expect for FCR with a lower value (0.15±0.05). Precisions of estimates were quite low, given the sample size. Heritability of MA (0.16±0.05) was similar to those estimated for FCR. For the 3 robustness traits, heritabilities were low for ABC and R2, 0.06±0.03 and 0.08±0.04 respectively, and moderate for VA (0.27±0.08). The trait MA had moderate to high negative genetic correlations with production traits (-0.52 to -0.79). It was positively correlated with DFI (0.28). Genetic correlations between VA and ADG, BF100 and LD100 were not significantly different from 0. The trait VA was moderately correlated with ABC (0.52±0.29) and negatively with FCR and DFI. The three genetic correlations between the traits MA, VA and R2 (ranging from -0.15 to -0.20) were not significantly different from 0. Several estimates of genetic correlations had large standard errors and should be interpreted with caution.

Discussion

The traits MA and FCR were quite similar, with close heritability estimates and a strong and favourable correlation. This correlation looks different from 1, which could imply that the trait MA captures other elements of energy allocation than FCR. The strong and unfavourable genetic correlation with ADG could be related on the way these two traits were estimated. They were measured over an identical period for all individuals but were not standardized between starting and finishing weights (ADG 30-110 kg). Some of the animals tested reached their maturity weight before the end of testing period, which leaded to a drop

Table 1. Descriptive statistics (Mean and *SD*: standard deviation), heritability (h^2 ± standard error) for studied traits and genetic correlations (r^2a ± standard error) of mean of α_t (MA) and variance of α_t (VA) with studied traits.

	ADG kg/d	BF100 mm	LD100 mm	FCR kg/kg	DFI kg/d	R2 no unit	ABC no unit	MA kg/MJ	VA kg^2/MJ^2
Mean	0.992	6.6	68.0	2.26	2.240	0.80	30,726	0.094	0.0005
SD	*0.108*	*0.8*	*5.3*	*0.19*	*0.287*	-	*24,764*	*0.019*	*0.0046*
h^2	0.36[1] ±0.05	0.32[1] ±0.07	0.41[1] ±0.07	0.15[1] ±0.05	0.31[2] ±0.07	0.08[2] ±0.04	0.06[2] ±0.03	0.16[2] ±0.05	0.27[2] ±0.08
r^2a MA	-0.64[2] ±0.14	-0.52[2] ±0.16	-0.59[2] ±0.17	-0.79[2] ±0.13	0.28[3] ±0.10	-0.17[3] ±0.26	0.57[3] ±0.32	-	-0.15[3] ±0.26
r^2a VA	-0.04[2] ±0.19	0.09[2] ±0.41	-0.02[2] ±0.18	-0.41[2] ±0.21	-0.36[3] ±0.16	-0.20[3] ±0.27	0.52[3] ±0.29	-0.15[3] ±0.26	-

ADG = average daily gain; BF100= backfat thickness estimated at 100 kg; LD100= longissimus dorsi thickness estimated at 100 kg; DFI= average daily feed intake; R2 = robustness trait; ABC= resilience index.

[1] Estimates from a 4-traits multiple trait model (ADG, FCR, BF100, LD100).

[2] Estimates from a 5-traits multiple trait model (ADG, FCR, BF100, LD100 and the trait under consideration).

[3] Estimates from a bivariate model (2 traits under consideration).

in feed efficiency. The trait MA described the average allocation of net energy in growth during fattening period. This trait seems to be an interesting way to phenotype feed efficiency but the relationship between evolution of α_t over time and degree of maturity, regardless maintenance requirements, needs further investigation. The trait MA was also unfavourably correlated to the resilience indicator ABC, suggesting that an increase of energy allocation increases the risk of deviation of potential ADG in case of an environmental perturbation. In contrast, VA was favourably correlated with robustness trait and moderately with ABC trait. The trait VA, variation of α_t over testing period, could be interpreted as an indicator of the response of the animal to perturbations/stress. Indeed, more robust animals are less impacted by perturbations, we assumed that these animals would have a 'more stable' α_t and consequently a lower value of VA. Heritability of this trait was higher than the other robustness traits. This makes it possible to consider a selection on the trait VA. A selection to reduce the variance of α_t would have a negative impact on FCR genetic potential, due to the moderate negative correlation with FCR, and no (or slightly positive) effect on the mean of α_t.

These preliminary results show the value of using the dynamic linear regression method in order to estimate time-trends in allocation (α_t) and to define robustness indicators on the basis of energy allocation. Our perspective is to further decompose the allocation factor into components such as degree of maturity, sensitivity to perturbations, etc in order to better characterise the robustness phenotype and propose useful indicators of robustness for genetic selection.

References

Gilmour A.R., Gogel B.J., Cullis B.R., and Thompson R. (2009) ASREML user guide release 3.0. Available at: https://www.vsni.co.uk/downloads/asreml/release3/UserGuide.pdf

Knap P.W. (2005) J Exp Agric 45: 763–773. https://doi.org/10.1071/EA05041

Lenoir G., Flatres-Grall L., Friggens N.C., and David I. (2021) Preprint Zenodo. https://doi.org/10.5281/zenodo.5572681

Nguyen-Ba H., Milgen J. van, and Taghipoor M. (2019) Animal 14(2): 253-260.
https://doi.org/10.1017/S1751731119001976

Noblet J., Dourmad J.-Y., and Milgen J. van (2016) Proc. of the 48th Journ. Rech. Porcine, Paris, France.

Petris G., Petrone S., and Campagnoli P. (2009) Dynamic Linear Models with R, Use R. Springer, New York, NY. https://doi.org/10.1007/b135794_1

Revilla M., Lenoir G., Flatres-Grall L., Muñoz-Tamayo R., and Friggens N.C. (2022) PC Journal 2(e9). https://doi.org/10.24072/pcjournal.82

57. Milk metabolite profiles in goats selected for longevity support link between resource allocation and resilience

M. Ithurbide[1*], H. Wang[2], C. Huau[1], I. Palhière[1], T. Fassier[3], J. Pires[4], T. Larsen[5], J. Cao[2], N.C. Friggens[6] and R. Rupp[1]

[1]GenPhySE, Université de Toulouse, INRAE, Castanet Tolosan, France; [2]Department of Statistics and Actuarial Science, Simon Fraser University, Burnaby BC, Canada; [3]Domaine de Bourges, INRAE, Osmoy, France; [4]INRAE, Université Clermont Auvergne, Vetagro Sup, UMR Herbivores, Saint-Genès-Champanelle, France; [5]Aarhus University, Faculty of Agricultural Sciences, Research Centre Foulum, P.O. Box 50, 8830 Tjele, Denmark; [6]UMR 0791 Modélisation Systémique Appliquée aux Ruminants, INRAE, AgroParisTech, Université Paris-Saclay, 75005 Paris, France; marie.ithurbide@inrae.fr

Abstract

The aim of this study was to assess the use of milk metabolites as an indicator of resilience. Hyper-selection based on functional longevity of AI bucks successfully created two groups of goats with different lifespan. Furthermore, the higher body weight of high longevity line goats and lower milk fat:protein ratio during the beginning of the first lactation suggested different resource allocation profiles between lines. Finally, the difference between the two lines of goats in milk metabolites evolutions during an imposed underfeeding challenge supported this hypothesis. These preliminary results support the value of milk metabolites as a promising noninvasive proxy related to resilience in dairy ruminants.

Introduction

Resilience is the ability of a goat to cope with environmental disturbances, such as pathogens or feed restriction. Genetic selection on resilience remains a challenge. Because resilience translates into enhanced functional longevity, we created two lines of Alpine goats using hyper-selected bucks with the most extreme estimated breeding values for functional longevity. Recent literature showed evidence of a relationship between indicators of body fatness dynamics and resilience (Savietto *et al.*, 2015). In the present study, we explored the evolution of metabolic pathways related with lipomobilisation during feed restriction among goats selected for extreme functional longevity. Accordingly, an experiment was carried out in which time-series measurements of 14 milk metabolites were made in dairy goats exposed to an underfeeding challenge during early lactation. The aim of this study was to validate the use of milk metabolites as an indicator of resilience.

Materials & methods

Experiment. Following the method developed by (Palhière *et al.*, 2018), we made a hyper-selection of dairy goats based on the genetic evaluation for functional longevity of 8,787 alpine AI bucks. From 2017 to 2021, two lines of goats were bred in an INRAE experimental farm (Bourges, France; Ethics approval C18-174-01) comprising 212 low longevity line (low_LGV) and 228 high longevity line goats (high_LGV). The following measures were regularly registered by trained technicians: serum immunoglobulin G concentration (IgG) at birth, monthly weight, health treatment, milk production, milk somatic cell scores (SCS), death/culling date.

In 2020, 49 primiparous dairy goats were exposed to a 2-d underfeeding challenge during early lactation (mean days in milk=39, SD=11). The challenge consisted of a 7-d control period (day -7 to 0) with a standard diet composed of forage and concentrate, 2 d of straw only feeding (day 1 and 2), and a 4-d recovery period on the standard diet (day 2 to 6). Prior to the start of each challenge period, the goats received the standard diet from parturition onwards. The goats were milked twice per day. From the beginning of

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_57, © M. Ithurbide *et al.* 2022

the prechallenge control period until the end of the postchallenge recovery period, concentrations of 14 milk metabolites were measured daily from the morning milking: glucose-6-Phosphate (Glu6P), glucose (Glu), galactose (Gal), beta-hydroxy-butyrate (BOHB), isocitrate (Isoc), glutamate (Glut), NH2, lactate deshydrogenase (LDH), Urea, Cholin, Malate, urate, triacylglycerol (TAG), cholesterol (Chol).

Animals were cared and handled in accordance with the French legislation on animal experimentation and European Convention for the Protection of Vertebrates Used for Experimental and Other Scientific Purposes (European Directive 86/609) under specific authorization number APAFIS#8613-2017012012585646.

Survival analysis and linear models. The R package 'survival' was used to compare lifespan between the two lines. Survival analysis was performed using a Cox model. The model included both fixed covariate (line, year of birth, IgG at birth), and time-dependent covariates: weight, health treatment events (Zhang *et al.* 2018). Moreover, we ran linear regressions using the lm and glm procedures in R in order to compare between the two lines: weight at birth, monthly weight, IgG at birth, first lactation milk yield at DIM 90 and 250 (MY90, MY250), milk fat: protein ratio and SCS.

Functional principal component analysis. We characterized milk metabolite profiles upon challenge using a functional PCA (fPCA) for each milk metabolite (Yao *et al.*, 2005).The daily concentration of each metabolite during the whole challenge period was included in the fPCA. We used the R package 'fdapace'. This performs local linear smoothing on the covariance matrix and then estimates the principal components with conditional expectation method. Each individual milk metabolite curve was then characterized by a PCscore for each component. We chose the number of components as the minimum needed to explain 80% of the variance of the dataset.

Sparse PLS DA. We then used Sparse PLS-Discriminant Analysis (sPLS-DA) to evaluate the ability of milk metabolites to distinguish the longevity lines. This is a linear multivariate model which performs classification tasks and is able to predict the class of new samples (Lê Cao *et al.*, 2011). The method integrated a continuous data matrix comprising the individual PCscores and a categorical outcome variable: the line of the goat (high/low_LGV). sPLS-DA seeks the components that best separate the sample groups, and also selects variables that best discriminate between groups using lasso penalization. We chose the number of components using cross validation on a non-sparse model (comprising all variables) and then tuned the number of variables to select on each component. We assessed the final performance of the model using a 5-fold cross validation.

Results & discussion

Comparison of the two longevity goat lines. Weight and growth during the first year of life were similar in both goat lines. In contrast, the low_LGV goats had a lower weight during the beginning of first lactation than high_LGV goats. The milk fat to protein ratio was also significantly higher in low_LGV goats during first lactation. The overall survival (from birth to death) of high_LGV goats was significantly greater than low_LGV goats (hazard ratio=0.63, CI=0.47;0.86) even after we included treatment, growth, IgG at birth, and year effects, in the model (Figure 1). The line effect was not constant over time: no significant effect was found during the first year of life and the difference emerged after first kidding. This suggests that survival at an early stage of life and during later productive life are under different genetic regulation. These results support the hypothesis that the two lines of goats present different resilience levels and that this may be linked to different resource allocation profiles between lines. The analysis of milk metabolites profiles should contribute to validate this hypothesis.

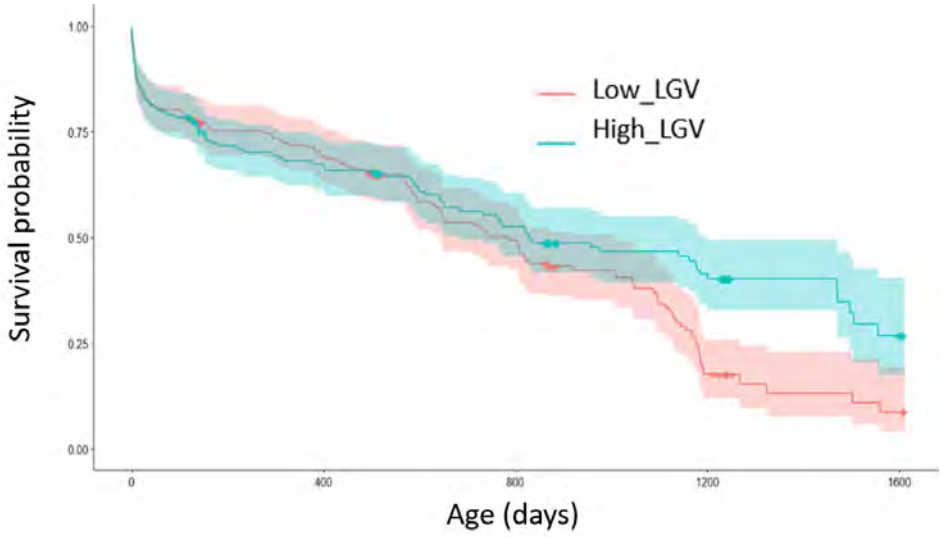


Figure 1. Plot of Kaplan–Meier curve showing survival against time for high longevity line (high_LGV) and low longevity line (low_LGV). Shaded areas represent the 95% confidence intervals calculated by Cox analysis.

Profiling of milk metabolites upon underfeeding challenge. Using functional principal component analysis, between 2 and 4 components were required to explain at least 80% of variation resulting in 50 variables for the Sparse PLS-DA (named Chol1, Chol2 and Chol3 for variables describing cholesterol profiles). Figure 2 details the construction of the fPCA for Glu.

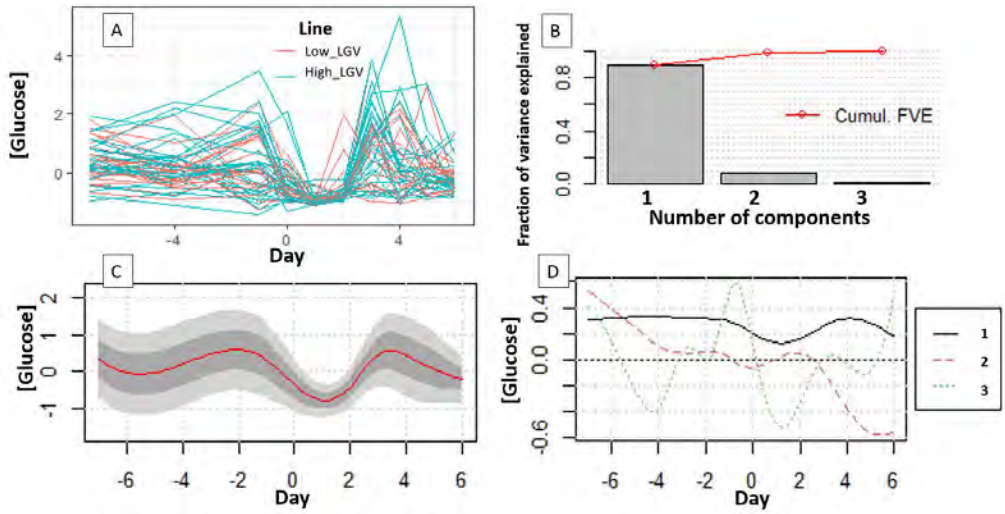


Figure 2. Functional principal component analysis results for one milk metabolite (concentration of milk free glucose) through the underfeeding challenge. (A) Individual raw data, each line represents 1 goat. Colors represent lines. (B) Fraction of variance explained by the first 3 PCA components. (C) Estimation of the mean curve by functional smoothing. (D) Underlying Eigenfunctions.

Discriminating longevity goat lines with milk metabolites. The optimal number of components in the sPLS-DA was 3. The lasso penalization selected 30, 2 and 20 variables for components 1, 2 and 3, respectively. The balanced error rate estimated for the model was 30.2% (respectively 30.0 and 30.3 for low_LGV and high_LGV lines). As shown in Figure 3 the most discriminating parameters between lines were Chol3, Gal2, LDH4, and TAG3. These results indicate that differences in functional longevity, which are seemingly due to differences in resilience, can be detected from milk metabolite responses to a short-term underfeeding challenge. Further, this preliminary analysis suggests that the physiological mechanisms underpinning differences in resilience are related in part to lipid metabolism.

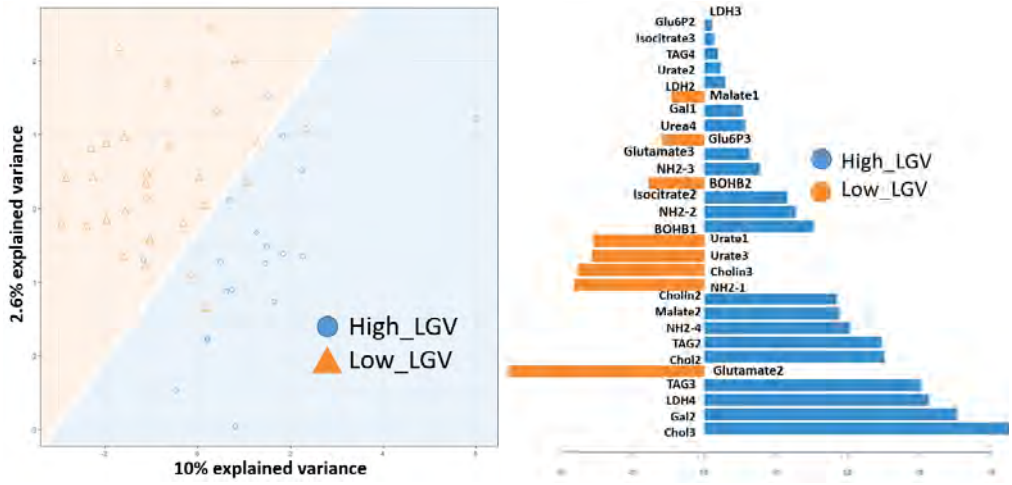


Figure 3. (A) Sample plot from the sPLS-DA applied to the fPCscores of milk metabolites to discriminate lines of goats for functional longevity. (B) Loading plot from the sPLS-DA. Colors indicates the line in which the median is maximum for each fPCscore.

Acknowledgements

The authors thank the staff of the INRAE experimental unit at Bourges for technical support in producing, raising, and monitoring the animals. This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 772787 and the national APIS-GENE (Paris, France) funded projects ActiveGoat and Resilait.

References

Lê Cao K.A., Boitard S., and Besse P. (2011) BMC Bioinformatics 12 (1): 253. https://doi.org/10.1186/1471-2105-12-253.

Palhière I., Oget C., and Rupp R. (2018) Proc. of the 11th WCGALP, Auckland, New Zealand

Savietto D., Friggens N.C., and Pascual J.J. (2015). Genetics, Selection, Evolution : GSE 47 (1): 2. https://doi.org/10.1186/s12711-014-0073-5.

Yao F., Müller H.G., and Wang J.-L. (2005). Journal of the American Statistical Association 100 (470): 577–90. https://doi.org/10.1198/016214504000001745.

Zhang, Z., Reinikainen J., Adeleke K.A., Pieterse M.E., and Groothuis-Oudshoorn G.M.C. (2018). Annals of Translational Medicine 6 (7): 121. https://doi.org/10.21037/atm.2018.02.12

58. Lifetime feed efficiency and deep phenotypes from scarce feed intake records using the mechanistic LiGAPS-Dairy model

A. van der Linden[1], G.C.B. Schopen[1], B. Gredler-Grandl[1], G. de Jong[2], M. Sol[2], S. van der Beek[2,3] and R.F. Veerkamp[1*]

[1]Wageningen University & Research, P.O. Box 338, 6700 AH Wageningen, the Netherlands;[2]CRV, Postbus 454, 6800 AL Arnhem, the Netherlands; [3]Global Layer Genetics GmbH, Hogenbogen 1, 49429 Visbek, Germany; roel.veerkamp@wur.nl

Abstract

Ideally, selection for feed efficiency requires deep phenotyping of net efficiency, or lifetime recording of intake and all energy sinks across environments. However, recording of feed intake is scarce. Therefore, net efficiency is often defined as a simplistic linear equation, e.g. RFI. We tested the use of the mechanistic LiGAPS-Dairy model to derive nine deep phenotypes with a dataset for 1,228 dairy cows, combining feed intake, yield and liveweight data, with ration, weather, cow and farm data. Mismatch between data recording and model assumptions made this process time consuming, but allowing for missing parities and further automation should improve this quickly. We managed for 206 cows to estimate the deep phenotypes. Heritability and phenotypic correlations between the nine traits were estimated. When the pipeline is finished, the mechanistic LiGAPS-Dairy model will enable us to derive a more comprehensive breeding goal, more closely resembling net efficiency, whilst utilising scarce records.

Introduction

Extending breeding goals require more and more scarcely recorded data to be included in animal breeding. For example, breeding values for feed efficiency should be based on a full lactation or even a full lifetime of many cows, and across many different environments. However, in practice feed intake has been recorded for limited numbers of cows, during part of their lifetime, and on cows that are mostly kept indoors in intensive feeding systems. Also, feed efficiency is a composite trait containing many elements: energy uses for milk production, gestation, maintenance and growth, all with their own net efficiency (Veerkamp and Emmans, 1995). Ideally, animal breeding should target these net efficiencies (e.g. digestibility or converting net energy to yield) rather than partitioning energy away from essential functions towards production. However, net efficiencies are even more difficult to measure than feed intake, and therefore, often simple approximations, like residual feed intake (RFI), are used. A potential solution to make the step from scarcely recorded traits, with limited data per animal in selective circumstances, towards more plausible breeding goals is to use mechanistic models to standardise or derive new traits that can be used in breeding values estimation.

The mechanistic model LiGAPS-Dairy can assess the performance of cows in different environments and in different farm types and uses recorded phenotypes to simulate the performance of dairy cows over their entire lifetime with a time step of one day. Inputs required for this model are parameters that represent the animals' traits, outside weather temperature, information on feed quality and available feed quantity, and livestock management. The feeding strategy implemented can be precisely set and the model has been evaluated under Dutch conditions (van der Linden *et al.*, 2021).

The model genotype (in contrast to the DNA genotype that animal breeders use) of an animal is included in the LiGAPS-Dairy model as a set of different parameters that represent the individual animal's traits, i.e. traits that otherwise need deep phenotyping. Parameter values of individual cows can be determined by model calibration with measured data for milk production, feed intake and liveweight. Using these deeper

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_58, © A. van der Linden *et al.* 2022

phenotypes, the model simulates milk production, feed intake and liveweight across the lifetime of a cow and therefore allows for new definitions of feed efficiency for dairy cows outside the space where data were recorded, e.g. a full lifetime efficiency or in different environments, on top of selecting for the deeper phenotypes directly. For example, a recent pilot study used data from indoor-housed cows at a commercial farm of CRV (milk production, feed intake and live weight) with input model parameters (feed quality and feed availability, weather conditions and livestock management) to calibrate nine parameters (or deep phenotypes) for 101 individual cows, and simulated feed efficiency of these cows in an outdoor grazing system.

The objective of this study was to get estimates for nine deep phenotypes, and to assess their heritability and correlations. We report on work in progress to expand the LiGAPS-Dairy pipeline on a more complex (more herds and years and different recording strategies) and larger dataset (~1,228 dairy cows).

Materials & methods

A dataset was available on 1,228 cows from Dairy Campus and other experimental farms with weekly records on dry matter intake (DMI), fat- and protein-corrected milk (FPCM) and liveweight (LW) recorded in 3,701 lactations and 605 different experimental treatments between 1991 and 2021. For the weekly records there were at least 25 records for DMI, 9 records for kg of milk and 9 records for LW per cow. A cow should have records for at least 2 lactations (the first lactation often differs from the second or third). All cows in the feed efficiency experiments were housed indoors. Birth and calving dates were used as model input to resemble the life of cows in the experiments.

Weather data were obtained for de Bilt, the Netherlands (52.2°N, 5.2°E) and were assumed to be representative for the locations of the experimental farms. Feed quality was not available for individual experiments. The average feed quality of the diets was used from the study by Zom *et al.* (2012), where the average proportion of concentrates was 37% of the dry matter (DM) intake; Crude protein (CP) of concentrates was 188 g/kg DM, digestible Organic Matter (dOM) was 829 g/kg DM, and CP of roughage was 136 g/kg DM, and dOM was 700 g/kg DM.

The LiGAPS-Dairy model, written in R, was subsequently ran for each individual cow with Azure Databricks on the Microsoft Azure platform with inputs from the cow and farm. The model produced outputs that included FPCM, DMI, and LW. During calibration of the deep phenotypes, the simulated FPCM, LW and DMI were compared with the measured data for individual cows, and the difference was expressed as the root mean squared error (RMSE). The RMSEs for FPCM, DMI, and LW were added up and used as input for a genetic algorithm from the R-package 'GA'. This algorithm selected the combination of the nine deep phenotypes that resulted in the lowest sum of RMSE for each cow by iterating the model.

The deep phenotypes of the LiGAPS-Dairy model resulted in nine newly derived feed efficiency traits, together with DMI, FPCM and LW. Heritabilities and genetic correlations were estimated with ASReml (Gilmour *et al.*, 2015) and the model included only the mean and the additive genetic effect using the pedigree relationship matrix for each animal.

Results

The dataset used at the start is described in Table 1. Since it was the first time that the LiGAPS-Dairy model was used with this large dataset, optimising the model was very time consuming. At the time of writing, missing lactations and cows with parity above 5 created model failures, and needed to be excluded. After this selection, data on nearly all cows tested converged, and we were able to estimate the nine deep phenotypes for a cow in about 50 minutes, a task that could be parallelized by using Databricks. In total,

data on 206 cows converged and a small percentage failed due to mismatches between cow data and model assumptions. Solving these mismatches, and further automation and parallelisation is needed to make this process more efficient, and eventually get data on more animals converged. The summary statistics for the nine deep phenotypes are given in Table 2.

Using the 206 animals available just now, only five of the nine heritability estimates converged, albeit with large standard errors (Table 3). The low number of generations used in GA (n=2) might have reduced accuracy of the phenotypes and therefore the heritably. Phenotypic correlations were large between maintenance requirements and milk efficiency, contrary to what was expected when gross efficiency was used. Also, the correlations between the two milk efficiency traits and two Wood curve parameters were high.

Discussion

We reported research in progress on the development of the LiGAPS-Dairy model to derive nine deep phenotypes for scarcely recorded data on cows. Time needed to fit the sparse data was more than expected, and most effort was, and is, on automation of parallelisation and model adjustments to fit the data.

When efficient data fitting is resolved, in the near future, further steps required are to add more animals for the genetic parameter estimation, and to investigate the model fit per cow. Especially how model fit is affected by the records available for a cow, and the number of generations and base population size assumed in the genetic algorithm. These are likely to improve estimation of the heritability, since minimal

Table 1. Summary statistics of the data used.

# cows	1,228		
# lactations	3,701		
Average parity (sd)	2.99 (1.67)		
	FPCM (kg/day)	**LW (kg/cow)**	**DMI (kg/day)**
# records	50,364	58,491	62,375
Mean	18.7	609.5	21.3
stdev	7.0	112.5	4.2

Table 2. Average (Avg), standard deviation (sd), minimum (min) and maximum (max) values of nine deep phenotypes in Holstein-Friesian cows (initial n=206).

Description deep phenotypes	Units	Min	Avg	Max	sd
Energy requirements for maintenance	kJ NE/kg$^{0.75}$/day	256.9	360.3	420.9	49.4
Maximum intake capacity	Fill units/kg$^{0.75}$/day	105.7	120.9	134.4	6.8
Maximum body weight	kg	550.2	651.5	812.1	72.5
Woods curve parameter 1[1]	-	2,387	3,109	3,562	255
Woods curve parameter 2[1]	-	11.1	16.5	30.0	2.4
Woods curve parameter 3[1]	-	0.19	0.28	0.40	0.04
Net energy efficiency for milk	%	0.77	0.89	0.97	0.06
Protein efficiency for milk	%	0.42	0.62	0.70	0.08
Protein requirements per unit NE	g per kJ NE	1.88	2.03	2.20	0.10

[1] Woods curve parameter 1, 2 and 3 indicate the first, second, and third parameter of Woods curve multiplied by 100, describing jointly deeper phenotypes for the genetic potential for milk production in LiGAPS-Dairy.

Table 3. Heritability (diagonal), and phenotypic correlations, with standard errors in brackets for the nine deep phenotypes in Holstein-Friesian cows (initial n=206).

	1	2	3	4	5	6	7	8	9
1 Energy requirements maintenance	0.10	0.34	0.19	0.33	-0.04	-0.05	0.68	0.40	-0.30
	(0.14)	(0.06)	(0.07)	(0.06)	(0.07)	(0.07)	(0.04)	(0.06)	(0.06)
2 Maximum intake capacity		0.28	0.18	-0.16	0.16	0.11	0.12	-0.01	-0.18
		(0.23)	(0.07)	(0.07)	(0.07)	(0.07)	(0.07)	(0.07)	(0.07)
3 Maximum body weight			0.14	0.11	-0.02	0.19	-0.08	-0.20	-0.08
			(0.19)	(0.07)	(0.07)	(0.07)	(0.07)	(0.07)	(0.07)
4 Woods curve 1				0.00	-0.62	-0.31	0.39	0.55	-0.21
				(0.12)	(0.04)	(0.06)	(0.06)	(0.05)	(0.07)
5 Woods curve 2					0.00	-0.12	-0.29	-0.08	-0.12
					(0.15)	(0.07)	(0.06)	(0.07)	(0.07)
6 Woods curve 3						0.00	-0.36	-0.15	-0.36
						(0.09)	(0.06)	(0.07)	(0.06)
7 Net energy efficiency milk							0.00	0.45	-0.13
							(0.14)	(0.06)	(0.07)
8 Protein efficiency milk								0.73	-0.12
								(0.18)	(0.07)
9 Protein per unit NE									0.10
									(0.17)

criteria (two generations and the base population of 40 animals) were used. With a large dataset, the genomic predictions for the nine deep phenotypes can be subsequently used through the mechanistic LiGAPS-Dairy model to derive a more comprehensive feed efficiency breeding goal, more closely than RFI resembling net efficiency or lifetime efficiency, whilst still utilising scarce records for feed intake, FPCM and LW. Furthermore, through this mechanistic model the impact of selection can be estimated in different environments, i.e. indoor versus outdoor systems, or different climates.

References

Gilmour A.R., Gogel B.J., Cullis B.R., Welham S.J. and Thompson R. (2015) ASReml User Guide Release 4.1 Functional Specification, pp. VSN International Ltd., Hemel Hempstead, HP1 1ES, UK

Van Der Linden A., Oosting S.J., Van De Ven G.W.J., Zom R., Van Ittersum M.K. *et al.* (2021) J. Dairy Sci. 104(5)**:** 5689-5704. https://doi.org/10.3168/jds.2020-19078

Veerkamp R.F., and Emmans G.C. (1995) Livest. Prod. Sci. 44(2)**:** 87-97. https://doi.org/10.1016/0301-6226(95)00065-0

Zom R.L.G., Andre G. and Van Vuuren A.M. (2012) Livestock Science 143(1)**:** 43-57. https://doi.org/10.1016/j.livsci.2011.08.014

59. Genetic analysis of feed efficiency in Atlantic salmon

A.K. Sonesson[1*], B. Hatlen[1], B. Gjerde[1], G.F. Difford[1], P. Brunsvik[1], T.H.E. Meuwissen[2] and A. Norris[3]

[1]*Nofima AS, P.O. Box 210, 1432 Ås, Norway;* [2]*NMBU-IHA, P.O. Box 5003, 1432 Norway;*
[3]*MOWI AS, Sandviksboder 77AB, Bergen, Norway;* anna.sonesson@nofima.no

Abstract

Families (35) of Atlantic salmon were duplicated in 70 tanks and entered into a feed efficiency trial. Cumulative Feed intake (FI), weight gain (WTGAIN), feed efficiency (FE) and final energy content of the fish (ENERGY), which depends on its lipid content, were recorded at tank level and analysed at the family level using ~55000 SNP markers. To the best of our knowledge, genomic analysis of feed efficiency traits at the family level have not been estimated before, highlighting new possibilities for analysing family level group data with genomics. Results showed medium-high heritability for the four traits. ENERGY, which is a direct effect of body lipid content, was negatively correlated to FE (r=-0.60), showing its potential as a selection criterion for feed efficiency. As expected, the genetic correlation between WTGAIN and FI was extremely high in this dataset, which indicates that gross WTGAIN is a very good predictor of FI.

Introduction

Feed efficiency (g body growth per g feed eaten) is the holy grail in aquaculture production, because of its large economic importance and its relation to resource efficiency. Feed is responsible for 50% of operational costs and 70% of the carbon foot prints are related to feed (Iversen *et al.*, 2020; Winther *et al.*, 2020) in the Norwegian Atlantic salmon production.

Feed intake is difficult to record, especially at individual level, and therefore it is difficult to select directly for feed efficiency in breeding programmes. For selective breeding, feed intake is therefore typically measured at family level (Kolstad *et al.*, 2004), but this is costly and, so far, only small experiments have been performed with low numbers of families and relatively small fish. At individual level, the number of feed pellets eaten can be obtained by X-raying fish after using a feed with X-ray opaque beads (Kause *et al.*, 2006; G.F. Difford, personal communications). Alternatively, protein utilisation efficiency, which is strongly related to FE, can be assessed by measuring isotope retention in the body from feed (Dvergedal *et al.*, 2019). Finally, feed intake of fish in an experimental set up with one fish per aquarium has given promising results in European seabass (Besson *et al.*, 2019), but not in Atlantic salmon (B. Gjerde, personal communications).

Due to the difficulty to measure individual feed intake, feed efficiency has mainly been improved by indirect selection for growth; 62% of the between-tank (family) variation in feed intake was explained by growth and sampling day in Dvergedal *et al.* (2019). With a given feed composition, feed efficiency is proportional to the ratio between energy efficiency and energy in the body. Thus, an alternative route to a more feed efficient fish, avoiding tedious and costly measurements of feed intake, may be via direct selection for both increased growth and decreasing body energy (Kause *et al.*, 2016). In the case of tank level phenotypes, genomic analysis can be used to evaluate genomic differences between families that are linked to phenotypic differences between families.

The aim of this study was to estimate genetic parameters for feed efficiency traits in Atlantic salmon juveniles using family mean records and individual level genomic data.

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_59, © A.K. Sonesson *et al.* 2022

Materials & methods

Fish. The fish experiment took place at the Nofima Research Station for sustainable aquaculture at Sunndalsøra. Thirty-five full-sib families of MOWIs elite population were split into two duplicated freshwater tanks with 25 individuals per tank, i.e. 1,750 fish in total in 70 tanks that were located in two different rooms with 20 and 50 tanks, respectively. They were grown over 50 days from about 50 g to 110 g average body weight.

Recording. Four relevant traits for feed efficiency were recorded. The mean weight gain (WTGAIN; g) measured on the fish in each of the 70 tanks; Daily feed intake (FI; g) measured at the tank level by means of given feed and feed not eaten on wasted feed collectors; final body energy (ENERGY; kJ/g) measured by bomb calorimetry from a minced pool of all fish per tank; feed efficiency (FE) at the tank level as WTGAIN/FI. The fish were individually genotyped by a customised SNP array of MOWI with 55,735 SNPs.

Genomic analysis. Estimates of (co)variances of the random effects in the applied statistical multitrait SNP effects model was obtained using ASReml software (Butler *et al.*, 2017) with room of the tank as a fixed effect for the four traits FI, WTGAIN, FE and ENERGY. The model was:

$$y_i = H_l + (Z_i \otimes I_4)b + e_i \quad (1)$$

where y_i is a vector of average trait values (over 25 fish) in tank *i* for the traits FI, WTGAIN, FE and ENERGY; e_i is a vector of within-tank residuals for the 4 traits; H_l is the vector of room *l* effects to which tank i belongs; b is a vector of multitrait SNP effects (4 effects per SNP); Z_i is the vector of the average SNP genotypes of the fish in tank *i*, as a deviation from the overall genotype and standardised, i.e. $Z_{ij}=2(q_{ij}-p_j)/s$ where q_{ij},is the frequency of SNP j within tank i, p_j is the overall allele frequency for SNP *j*, and s is the square root of the sum of the heterozygosities ($2p_j(1-p_j)$) of the SNPs. The variance of the trait-effects of SNP *j* is $Var(b_j)=G$, where G is the (4×4) genetic covariance matrix of the traits. Similarly, $Var(e_i) = R$ where R is the (4×4) residual covariance matrix of the traits. Since the trait values are averaged over 25 fish, the actual environmental (co)variances are a factor 25 larger than R, i.e. 25×R.

Results

Figure 1 shows the feed efficiency of the duplicated families; mean was 1.28±0.005. The family mean feed efficiencies range from about 1.1 to 1.4, i.e. a difference of ~30%.

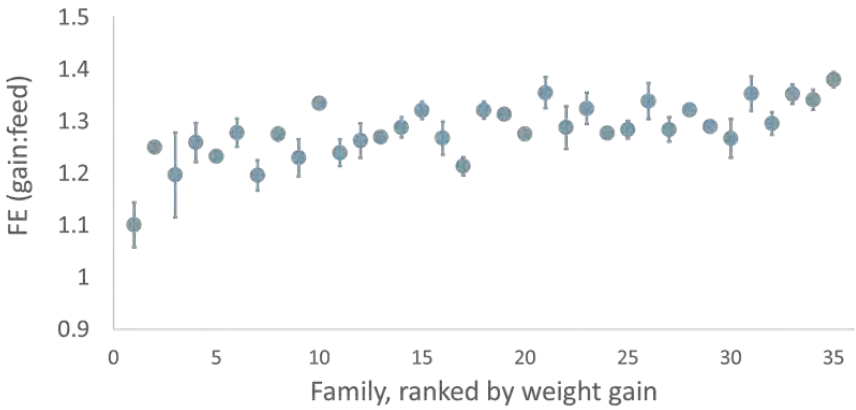


Figure 1. Feed efficiency (FE) of the 35 duplicated families of Atlantic salmon (mean ± s.e).

Table 1. Heritability (diagonal) and genetic (above diagonal) and residual (below diagonal) correlations between FI, WTGAIN, FE and ENERGY in Atlantic salmon.

	FI	WTGAIN	FE	ENERGY
FI	0.35	0.98	0.56	0.09
WTGAIN	0.92	0.42	0.69	-0.24
FE	-0.20	0.20	0.17	-0.60
ENERGY	0.55	0.43	-0.36	0.27

The four studied traits had medium heritability ranging from 0.17 for FE to 0.42 for WTGAIN. The genetic correlation of WTGAIN with FI was very high (0.98) while that with FE (0.69) was much lower. The negative genetic correlation between FE and body ENERGY (-0.60) shows that high FE is accompanied by low fat content.

Discussion

The chosen experimental design of family/tank data proved useful to estimate genetic parameters for the traits when a genomic analysis was used with SNP array data of ~50,000 SNPs. Although genetic parameters are expected to be estimated without bias, the use of tank-means implies that tank-residuals cannot be split into environmental effects of the tank and individual (within tank) environmental effects. This affects the estimates of heritability as shown below. The estimated genetic correlations between the traits were very similar to those obtained from a model with room as a fixed effect, sire and dam as random effects where variance components for sire and dam were assumed to be equal, and with tank as the additional random residual effect, and with the additive genetic numerator relationship matrix added to the sire and dam effects in the mixed model equation. From this sire and dam model, the heritability of the traits cannot be calculated as this requires estimates of the environmental variance within the tanks for each of the traits. The estimated residual (tank) variance from this model accounted for about 3.5% of the estimated total phenotypic variance, i.e. the sum of the estimates of the sire, dam and residual variances.

The heritability of FI, WTGAIN, FE and body ENERGY were moderate, where FE had unfortunately the lowest heritability (0.17). The environmental variances (σ_e^2), which are needed for the denominator in the calculation of heritability, where assumed here to equal $\sigma_e^2=25\times\sigma_r^2$, which assumes that the variance due to tank effects is 0, i.e. environmental effects between the same trait in replicated tanks are uncorrelated. In the presence of tank effects, the environmental effects show an intra-class correlation of t, and the environmental variance is $\sigma_e^2 = \sigma_r^2/[t+(1-t)/25]$, which varies between $25\times\sigma_r^2$ and $1\times\sigma_r^2$, for t=0 to t=1, respectively, where t=1 (no individual environmental variance) is very unrealistic. Hence, we may have used an over-estimate of the environmental variance for the heritability estimation, which implies that the heritability in Table 1 are conservative estimates.

The genetic correlation between WTGAIN and FI was very high (0.98). This shows that gross WTGAIN (not accounted for body composition) is a good predictor of FI and that there is very little room for genetic improvement of residual feed intake, i.e. feed intake not related to WTGAIN, although small improvements of this trait may yield high economic benefits as feed accounts for a large proportion of the total cost. Thus, precise measurements of feed intake is necessary for direct selection for FE. However, the fish used here were only tested at the freshwater stage, and results might differ in seawater-phase fish, which is where most of the growth takes place. Our experimental design may also be applied to larger fish in the seawater-phase fish, but at a much higher cost as larger full-sib family-replicated units (cages) are required.

The medium and thus much lower genetic correlation between WTGAIN and FE (0.69) shows that high growth rate is favourably correlated to FE and that WTGAIN is a relatively good predictor of FE. The negative genetic correlation between FE and body ENERGY (-0.60) shows that high FE is accompanied by low fat content and thus ENERGY and WTGAIN. Together this may provide a very promising indicator trait for genetic selection for improved feed efficiency (as also found by Kause *et al.* 2016), as it is easy to measure compared to feed intake, given that body ENERGY can be reduced without negative effects on consumer preferences.

Acknowledgements

This study was financed by the EU project NewTechAqua- H2020-grant no 862658. We would also like to acknowledge the good work done at MOWI to produce and genotype the fish, and technicians at the Nofima Research Station for sustainable aquaculture at Sunndalsøra for performing the fish experiment in an excellent way.

References

Besson M., Allal F., Chatain B., Vergnet A., Clota F. *et al.* (2019) Front. Genet. Article 219 https://doi.org/10.3389/fgene.2019.00219

Butler D.G., Cullis B.R., Gilmour A.R., Gogel B.G., Thompson, R. (2017) ASReml-R Reference Manual Version 4. VSN International Ltd, Hemel Hempstead, HP1 1ES, UK.

Dvergedal H., Ødegård J., Øverland M., Mydland L.T., Klemetsdal G. (2019) Genet. Sel. Evol. 51: Article 13. https://doi.org/10.1186/s12711-019-0455-9

Iversen A., Asche F., Hermansen Ø., Nystøyl R. (2020) Aquaculture 522: Article 735089. https://doi.org/10.1016/j.aquaculture.2020.735089

Kause A., Tobin D., Dobly A., Houlihan D., Martin S., *et al.* (2006) Genet. Sel. Evol. 38:389–409. https://doi.org/10.1051/gse:2006011

Kause A., Kiessling A., Martin S.A.M., Houlihan D., Ruohonen K. (2016) Br. J. Nutr. 116:1656–65. https://doi.org/10.1017/S0007114516003603

Kolstad, K., Grisdale-Helland, B., Gjerde, B. (2004) Aquaculture 241: 169-177. https://10.1016/j.aquaculture.2004.09.001

Winther U., Hognes E.S., Jafarzadeh S., Ziegler F. (2020) Report No. 2019:01505. SINTEF Ocean AS

60. Validation and genetic analysis of a feed efficiency criterion in French Lacaune ewes

C. Machefert[1*], C. Robert-Granié[1], P. Hassoun[2], S. Parisot[3], C. Allain[3], D. Portes[3], G. Lagriffoul[4], J.M Astruc[4], and H. Larroque[1]

[1]GenPhySE, Université de Toulouse, INRAE, ENVT, 31326 Castanet-Tolosan, France; [2]SELMET, INRAE, CIRAD, Montpellier SupAgro, Univ Montpellier, 34060 Montpellier, France; [3]INRAE, Domaine de La Fage, 12250 Saint-Jean et Saint-Paul, France; [4]Institut de l'Elevage – CNBL, 75595 Paris, France; coralie.machefert@inrae.fr

Abstract

Improving feed efficiency have environmental, economic and social benefits for dairy ewes farming systems. Selection of animals using efficiently their feed resources while maintaining a quality milk production is therefore a major objective for breeders. However, phenotyping feed efficiency remains difficult to achieve on-farm on a large scale. For this purpose, a criterion of lactation feed conversion ratio measurable on commercial farms was proposed. First, accuracy of this criterion was evaluated on experimental data with precise phenotypes including individual intake. This criterion seemed to be, with some recommendations, a good approximation of feed efficiency, with the main limitation of collective feeding. Then, the criterion was applied to 6 test-days from 6,425 lactations of Lacaune ewes from production farms. In these conditions, the estimation of the genetic parameters of this feed efficiency criterion showed a low heritability (0.12±0.02) and a strong genetic correlation with the daily milk yield (0.80±0.03).

Introduction

In addition to its economic importance, improving feed efficiency in dairy sheep could also help to reduce environmental impacts of sheep production and to improve the sustainability of farming systems through food autonomy (Cantalapiedra-hijar *et al.*, 2020). Thus, having animals capable of using their feed as efficiently as possible, of mobilizing their body reserves at key moments of the production cycle, while maintaining a milk production in quantity and quality, as well as all the physiological functions (mating, resistance to diseases) is a strong challenge for the dairy sheep farmers. However, feed efficiency is difficult to phenotype on a large scale in dairy sheep farms. Indeed, for this it is necessary to record individual performances as milk production and its components, body weight (BW), body condition score (BCS) and feed intake (FI). One of the purposes of the European project SMARTER (SMAll RuminanTs breeding for Efficiency and Resilience, 2018-2022) is to evaluate novel phenotypes for feed efficiency, measurable in commercial farms.

The objective of our study was to propose a new criterion to approximate feed conversion ratio (FCR) of lactating ewes in a collective feeding context of commercial farms named lactation feed conversion ratio (LFCR). First, accuracy and reliability of the proposed criterion LFCR were evaluated. For that, using data from Lacaune ewes of La Fage INRAE experimental unit, this criterion was estimated and compared to a 'reference' one calculated with individual performances. Then, phenotypic and genetic analyses were performed in order to evaluate environmental and genetic factors (heritability) that explained variability of this feed efficiency trait as well as to estimate its genetic correlation with daily milk yield (DMY).

Materials & methods

Herds and ewes involved in the European SMARTER project. In France, the monitoring of the zootechnical performances of dairy ewes was carried out on commercial farms during two dairy years: 2019-2020 and 2020-2021. Overall, 8 French Lacaune dairy sheep herds were involved in SMARTER, located in the Roquefort area with around 4,500 ewes phenotyped by year.

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_60, © C. Machefert1 *et al.* 2022

Data collected. Phenotypes recorded for each ewe during a dairy year were related to: body reserve dynamics (5 BCS with target physiological stages: 1 month before lambing, end of suckling, 1st test-day (TD), 1 month before and after mating); milk production (milk yield, fat and protein contents on 6 TD after suckling and each month after); quantity and quality of feed offered (collective feed and automatic concentrate feeder (ACF)) at each TD and reproduction traits. After quality control and selection of data of ewes with at least 2 BCS, 2,529 ewes in 2020 and 3,896 in 2021 were retained for the analysis, corresponding to 6,425 lactations, with 1,744 ewes phenotyped over the two dairy years.

Feed efficiency criterion. Classical formula of FCR that divides the milk production by the feed intake does not consider the other requirements. LFCR makes the ratio of lactation requirements and energy inputs used to produce milk from feed and body reserves. BCS_{Δ} represented the variation between two successive BCSs. Energy for maintenance and growth not used for milk was subtracted from the inputs. All variables were first converted to net energy per day (Hassoun and Bocquier, 2007*)*. Lactation requirements were calculated from standardized milk yield (SMY) as proposed by Bocquier *et al.* (1993). Based on SMARTER data, FI represented the estimated average amounts of feed distributed collectively per ewe (pasture, forages and concentrates), plus individual concentrates in the ACF according to the dairy level. For the grazing part, Boissieu *et al.* (2019) proposed an estimation of the ingestion at pasture according to the time of presence per ewe. A parity-dependent reference BW was assigned to all ewes to calculate maintenance and growth requirements. For the growth of the primiparous, a standard average daily gain of 80 g/d was set. LFCR was calculated at each TD. Efficient animals had upper LFCR values.

Validation of the LFCR. Within the framework of the European iSAGE project at the La Fage INRAE experimental unit, Lacaune dairy ewes were individually monitored for their dairy performances (SMY), BW, BCS, and FI (based on weighing of forages and concentrates distributed and refused in individual troughs). All these measures were performed individually once a week (4 times per week for FI) over a period of 4 months during their 2nd and/or 3rd lactation, corresponding to 70 lactations available (46 ewes). LFCR was calculated once at the 2nd TD. BCS_{Δ} represented the variation in BCS between the end of suckling and the 2nd TD. The time step was 20 days. In order to mimic the on-farm SMARTER available data (collective feed, missing data), each variable: SMY, BW, BCS_{Δ} and forage intake, one at a time, was set for all ewes to the mean value of the batch (n=70). The four simplified batch LFCR results were successively compared to the individual LFCR. Statistical parameters as R^2, RMSE and bias were computed with R software to assess reliability of the approximated criterion.

Statistical analysis. In order to determine environmental factors affecting SMARTER feed efficiency results, analyses of variance were performed with the packages car, lme4 and lmer implemented in the R software. First, at each TD, fixed environmental effects were included into a linear model and selected using Fisher's tests. Genetic parameters were estimated with WOMBAT software (Meyer, 2007) using the REML method. A five-generation pedigree was used containing 17,451 animals. For each TD, the single-trait genetic model was:

$$y = Xb + Za + Wp + e \quad (1)$$

where y is LFCR measured at each TD; b the vector of fixed effects: parity (1/2/3/4+), lambing period (start/end) according to parity, litter size (single/multiple), lactation month crossed to TD number (2 groups of lactation months by TD) interacting with dairy year (2020/2021), herd (n=8) interacting with dairy year and herd interacting with mating mode (animal insemination/return/natural breeding); a is the vector of random additive genetic effects ($N(0, \sigma_a^2 A)$); p is a vector for random permanent environmental effects ($N(0, \sigma_p^2 I)$); e is the vector of residual random effects; X, Z and W are incidence matrices. Then merging all

TD data, a multi-trait animal model was retained in order to estimate the genetic correlation between LFCR and DMY. The models used for these two traits were the same as above adding a TD number fixed effect, interactions of all fixed effects with TD number (except lactation month crossed to TD number interacting with dairy year).

Results

Validation of the LFCR. Statistical parameters to assess the reliability of LFCR computed with SMARTER-like phenotypes (Table 1) showed that phenotypes contributing the most to the accuracy of this criterion were: (1) forage intake; (2) SMY; (3) BCS_{Δ} and the less was; (4) BW of the animal.

The absence of individual values for BW has little impact on the accuracy (R^2=0.97). For a small sample, replacing the variation in BCS with the mean value gives a relatively good estimate of feed efficiency (R^2=0.82). The lack of individual values for milk performances or feed intake has more impact and can leads to reclassification of ewes.

Considering these results, calculation of LFCR was carried out for ewes of the SMARTER project with the following principles: (1) Only ewes with complete dairy performances were kept; (2) feed intake was calculated from the feed distributed in the collective feed (with a 10% refusal rate for dry and wet forages) and from the concentrates distributed individually with ACF available in all farms; (3) an imputation of missing body condition scores (at least 2 BCS to register is imposed) was performed per farm using the copyMean method (Genolini and Falissard, 2011); (4) in the absence of BW, a reference value according to parity (in agreement with the technicians and similar to the literature) was applied (Hassoun and Bocquier, 2007).

Genetic parameters. Heritabilities and repeatability of LFCR were calculated at each TD (Table 2). Heritability and repeatability (between the two dairy years) seemed to be lower at the beginning than at the end of lactation.

Table 1. Statistical prediction parameters comparing initial and the four simplified (with a value set to the average of the batch) LFCR results.

	BW (kg)	BCS_{Δ} (over 20 d)	SMY (L/d)	Forage dry matter intake (kg/d)
Average value of the batch (= set value)	76	0.1	2.7	2.5
Standard deviation	8	0.1	0.5	0.4
Minimum	58	-0.4	1.8	1.4
Maximum	96	0.3	3.9	3.6
Coefficient of determination (R^2)	0.97	0.82	0.59	0.37
Root mean square error (RMSE)	0.05	0.11	0.21	0.21
Average bias	0	0.02	-0.03	0.03
Range of individual bias	[-0.26 ; 0.13]	[-0.22 ; 0.39]	[-0.58 ; 0.36]	[-0.24 ; 1.09]

Table 2. Heritabilities and repeatability of LFCR calculated at each test-day.

	TD 1	TD 2	TD 3	TD4	TD 5	TD 6
Heritability ± SE	0.07±0.02	0.11±0.02	0.11±0.02	0.12±0.02	0.12±0.03	0.20±0.03
Repeatability	0.26	0.28	0.28	0.29	0.35	0.40

Combining all TD records, genetic analyses were also performed with a multivariate model for LFCR and DMY. The heritabilities were low for these two traits: 0.12±0.02 for LFCR and 0.20±0.02 for DMY. There was a high genetic correlation of 0.80±0.03 between the two traits. The repeatability of the LFCR was 0.40, and it was 0.53 for DMY.

Discussion

There is a lack of information on the heritability of feed efficiency and its genetic correlation with production traits in dairy ewes. We can compare our results to studies carried out in dairy cows according to the feeding system used. With an estimated FI on pasture Hurley *et al.* (2018) found an heritability for a daily energy conversion efficiency ratio similar to our result (0.11±0.02). With individual daily FI measurements the heritability of a lactation feed conversion ratio was higher (0.25±0.03) and the genetic correlation with SMY was also high (Manafiazar *et al.*, 2016). Close to our results, heritability of daily milk yield of Lacaune ewes is around 0.28 (Barillet and Boichard, 1987). It has been shown that selecting a population on a ratio is difficult to implement because of the difficulty to predict the response to selection of ratio traits (Gunsett, 1984). However, our criterion can be measurable on commercial farms, and reflects an overall feed efficiency of ewe in a collective feeding context.

Acknowledgements

The authors thank the breeders and technicians of the breeding organizations that collected data. This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 772787 (SMARTER). It also benefited from data collected with the financial support of EU project H2020-679302 (iSAGE).

References

Barillet F., and Boichard D. (1987) Genet. Sel. Evol. 19(3):459-474. https://doi.org/10.1186/1297-9686-19-4-459

Bocquier E., Barillet F., Guillouet P., and Jacquin M. (1993) Ann Zootech 42(1):57-66.

Boissieu C., Fança B., and Hassoun P. (2019). La mise à l'herbe et le pâturage. In: L'Institut de l'Elevage (eds) L'alimentation des brebis laitières. L'Institut de l'Elevage, Paris, France, pp. 81-98.

Cantalapiedra-Hijar G., Faverdin P., Friggens N., and Martin P. (2020) INRAE Prod. Anim 33(4):235-248. https://doi.org/10.20870/productions-animales.2020.33.4.4594

Genolini C., and Falissard B. (2011) Comput. Methods Programs Biomed. 104(1):112-121. https://doi.org/10.1016/j.cmpb.2011.05.008

Gunsett F.C. (1984). J. Anim. Sci. 59(5):1185–1193. https://doi.org/10.2527/jas1984.5951185x

Hassoun P., and Bocquier F. (2007). Alimentation des ovins. In: Inra Editions (eds) Alimentation des bovins, ovins et caprins. Editions Quae, Paris, France, pp. 123-138.

Hurley A.M., Lopez-Villalobos N., McParland S., Lewis E., Kennedy E. *et al.* (2018) J. Dairy Sci. 101:1267-1280. https://doi.org/10.3168/jds.2017-12841

Manafiazar G., Goonewardene L., Miglior F., Crews D.-H., Basarab J.-A. *et al.* (2016) Animal 10(3):381-389. https://doi.org/10.1017/S1751731115002281

Meyer K. (2007) J Zhejiang Univ Sci B 8(11):815– 821. https://doi.org/10.1631/jzus.2007.B0815

61. Gene co-expression networks contributing to variation in residual feed intake in bovine hepatic tissue

K. Keogh[1*], M. McGee[2] and D.A. Kenny[1]

[1]Animal and Bioscience Research Department, Teagasc, Grange, Dunsany, Co. Meath, Ireland; [2]Livestock Systems Research Department, Teagasc, Grange, Dunsany, Co. Meath, Ireland; kate.a.keogh@teagasc.ie

Abstract

Gene co-expression network analyses were performed on hepatic transcriptome data of two breed types (Charolais (CH) and Holstein-Friesian (HF)), divergent for feed efficiency status through measurement of residual feed intake (RFI). The relationship between co-expressed genes and RFI was further assessed through correlation of co-expression networks with trait data including RFI, average daily gain (ADG) and dry matter intake (DMI). Across both breeds examined, one network of co-expressed genes was positively correlated with DMI (r=0.48, P=0.03), with a tendency also identified between the same network and RFI (r=0.42, P=0.07). Functional annotation of genes within this network revealed enrichment (P<0.05) of pathways related to both amino acid and fatty acid metabolism. Our results suggest a role for these processes in the hepatic tissue of cattle divergent for RFI. Moreover, genes within this network may represent potential biomarkers for the selection of improved feed efficiency in cattle across varying breed types.

Introduction

Provision of feed in beef production systems is a major determinant of overall profitability. For example, feed costs can account for up to 75% of the variable costs associated with beef cattle production (Kenny *et al.*, 2018). Thus, research related to the genomic control and subsequent selection of cattle with improved feed efficiency has received attention in recent years. In particular, RFI, defined as the difference between an animal's actual and predicted feed intake, has become the feed efficiency measure of choice due to its independence of growth and body size (Kenny *et al.*, 2018). Research efforts aimed at uncovering the underlying biology governing RFI in cattle, primarily through global transcriptomic evaluations, have established a role for genes involved in fatty acid metabolism and immune function towards variation in RFI. However, despite these findings, key robust genes or genomic regions contributing to the trait remain to be identified. Of the transcriptomic studies conducted to date few or no genes have been identified as commonly affecting RFI across studies. This is undoubtedly due to the multifaceted nature of the RFI trait as well as the varying experimental parameters employed across different studies, such as breed types, dietary sources and stage of development evaluated, ultimately confounding the subsequent outcome. Moreover, the majority of studies to date have focused on the identification of genes that are differentially expressed between animals divergent for RFI. However, such studies do not allow for an explanation of the interactions between genes governing the expression of a particular trait. The identification of highly co-expressed genes that contribute to the expression of complex traits may offer greater insights into the molecular control of a particular trait. Moreover, examining the relationship between genes through co-expression analyses, also allows for the identification of key hub genes, which are important in regulating the expression of several other genes within a network of co-expressed genes, and which ultimately may hold potential as biomarkers for the selection of a trait of interest. Thus, the objective of the current study was to identify networks of co-expressed genes in the liver tissue of CH and HF steers, divergent for RFI phenotype as well as the component traits of RFI including ADG and DMI. Specifically we sought to identify networks of co-expressed genes that were commonly associated and contributing to the RFI trait across the two contrasting breeds utilised. The liver was chosen as a target tissue as it is a central metabolic organ that has shown to be responsive to variation in RFI in cattle (Kenny *et al.*, 2018).

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_61, © K. Keogh *et al.* 2022

Materials & methods

Animal model. This experiment was conducted as part of a larger research programme designed to examine the within-animal repeatability of intake, growth and feed efficiency between varying stages of development in CH and HF beef steers (Coyle *et al.*, 2016). In total, 167 steers (90 CH and 77 HF) were used for this study. Following a dietary adaptation period, DMI and growth rates were measured over a 70-day feeding trial, during which all steers were offered the same high concentrate diet consisting of *ad libitum* concentrates plus a restricted allowance of grass silage daily. The residuals of the regression of DMI on ADG, and mid-test metabolic body weight within each breed were used to compute individual RFI coefficients for each steer (GLM procedure of SAS9.3). Residual feed intake was calculated for each animal as the difference between actual and predicted DMI. Within each breed, steers were ranked for high-RFI (feed-inefficient; n=12) and low-RFI (feed-efficient; n=12). Liver tissue samples were collected through percutaneous punch biopsy from all high-RFI and low-RFI steers following completion of the 70-day dietary trial. Tissue samples were washed with sterile DPBS and immediately snap frozen in liquid nitrogen before subsequent storage at -80 °C.

RNAsequencing and co-expression analyses. RNA isolation, cDNA library preparation, RNAsequencing and bioinformatics analysis have been described previously in Higgins *et al.* (2019). The weighted gene co-expression network analysis (WGCNA) software package (Langfelder and Horvath, 2008) was used to derive networks of co-expressed genes from liver transcriptomic data. RNAseq read count data were firstly filtered for lowly expressed genes and then normalised in EdgeR by read counts per million mapped reads, (adjusted for sample library size). Normalised count data were then $Log_2(x+1)$ transformed in R. In order to identify networks of co-expressed genes common across both breeds examined, a WGCNA consensus analysis was undertaken. The consensus analysis was undertaken using the automatic one-step function, blockwise consensus analysis of WGCNA, incorporating unsigned, weighted correlation network construction and module detection. The resultant networks of co-expressed genes, denoted by different colour identifiers, were then correlated with trait data. Specifically the relationship between networks of co-expressed genes and trait data including RFI, ADG and DMI were calculated by Pearson correlation. Networks which were commonly statistically significantly ($P<0.05$) associated with selected trait data across the two breeds evaluated were then selected for further analysis as potentially biologically interesting networks associated with RFI in beef cattle. Pathway analysis was performed on genes from each network identified as significantly correlated with trait data using DAVID gene ontology program. Gene ontology terms were considered significant if the adjusted *P*-value was less than 0.05.

Results

Within each breed, differences in ADG were not apparent ($P>0.05$), however high-RFI steers consumed more feed compared with low-RFI steers across both CH and HF breeds ($P<0.05$). Gene co-expression analysis resulted in the construction of 18 separate networks of co-expressed genes. Within the CH dataset, a number of separate networks were significantly associated with both RFI and DMI. These included the greenyellow network which was negatively associated with RFI (r=-0.44; $P=0.04$). Similarly both DMI and RFI were negatively associated with the tan network (DMI: r=-0.48; $P=0.03$; RFI: r=-0.51; $P=0.02$). Conversely, positive associations were evident within the CH dataset between the pink, blue and yellow networks and both RFI and DMI (pink network: DMI: r=0.59; $P=0.004$; RFI: r=0.47; $P=0.03$; blue network: DMI: r=0.45; $P=0.04$; RFI: r=0.49; $P=0.02$; yellow network: DMI: r=0.52; $P=0.01$; RFI: r=0.54; $P=0.01$). Within the HF dataset only one network (pink network) was associated with the traits evaluated (DMI: r=0.48; $P=0.03$). The pink network also displayed a tendency towards a significant association with RFI for the HF steers (r=0.42; $P=0.07$). Results from the WGCNA consensus analysis indicated significant commonality between CH and HF data for the pink network, which was positively associated with DMI across both breeds (r=0.48; $P=0.03$), whilst a tendency also existed between this network and RFI value

for the consensus analysis (r=0.42; *P*=0.07). Enriched biological pathways for significant networks are presented in Table 1, no pathways were significantly enriched for the greenyellow network.

Discussion

Residual feed intake has become a measure of choice for feed efficiency evaluations in cattle, due to its independence from growth and body size (Kenny *et al.*, 2018). However, despite RFI being a favourable measure of feed efficiency, knowledge of the underlying biology governing the trait remains to be elucidated fully. Studies evaluating the transcriptional control of RFI in cattle have identified biological processes related to fatty acid metabolism and the immune system as contributing to variation in RFI (Keogh *et al.*, 2018). However, despite commonality in biological processes, key genes contributing to RFI have not been identified across studies. This may be due to the use of varying experimental parameters across studies, for example employing different breed types and dietary regimes, which ultimately confound experimental results. Thus, the current study aimed to overcome that limitation through the evaluation of the molecular control of RFI across two varying breed types through gene co-expression analyses. A role for fatty acid metabolism was also evident in the current study through the association between the pink network of co-expressed genes with RFI and DMI across both breed types. Commonality of genes within the pink network across both CH and HF breeds, suggests a potential role for genes within this network towards the RFI trait irrespective of breed. Moreover, specific genes within the pink network were previously implicated towards variation in RFI in cattle, in studies utilising diverse breeds. These included *CYP1A1* and *CES1*, which are involved in fatty acid and lipid metabolism and which were also differentially expressed between high- and low-RFI cattle in the data of Weber *et al.* (2016) and Mukiibi *et al.* (2018). Genes involved in

Table 1. Biological pathways enriched in networks associated with trait data.

Enriched pathway	*P*-value
Tan network	
-Cell cycle	<0.0001
-DNA replication	0.0001
Pink network	
-Metabolic pathways	<0.0001
-Tyrosine metabolism	<0.0001
-Tryptophan metabolism	<0.0001
-Fatty acid degradation	<0.0001
-Biosynthesis of amino acids	0.0005
-Glycine, serine and threonine metabolism	0.027
-Fatty acid metabolism	0.047
Blue network	
-RNA transport	0.0059
-Metabolic pathways	0.011
-Proteasome	0.039
-mRNA surveillance pathway	0.039
-Oxidative phosphorylation	0.045
Yellow network	
-Ribosome	<0.0001
-Oxidative phosphorylation	<0.0001
-Proteasome	0.0013
-Protein processing in endoplasmic reticulum	0.032

amino acid processing including the aminopeptidase, *ANPEP* and the transaminase gene *TAT* were also differentially expressed in Higgins *et al.* (2019), Weber *et al.* (2016) and Mukiibi *et al.* (2018). Finally, *AHSG* which encodes a glycoprotein implicated in insulin resistance was differentially expressed in the liver based studies of Chen *et al.* (2011) and Weber *et al.* (2016). Results from this study further implicate biological processes related to fatty acid metabolism towards variation in RFI. Furthermore, genes within the pink network of co-expressed genes warrant further evaluation for their potential role as robust biomarkers for the genomic selection of RFI across varying breed type.

References

Chen Y., Gondro C., Quinn K., Herd R.M., Parnell P.F., *et al.* (2011) Anim. Genet. 42(5):475-490. https://doi.org/10.1111/j.1365-2052.2011.02182.x

Coyle S., Fitzsimons C., Kenny D.A., Kelly A.K., and McGee M. (2016) J. Anim. Sci. 94(5):719. https://doi.org/10.2527/jam2016-1481

Higgins M.G., Kenny D.A., Fitzsimons C., Blackshields G., Coyle S., *et al.* (2019) BMC Genomics 20(1):525. https://doi.org/10.1186/s12864-019-5906-8

Kenny D.A., Fitzsimons C., Waters S.M., and McGee M. (2018) Animal 12(9):1815-1826. https://doi.org/10.1017/S1751731118000976

Keogh K., Waters S.M., and Kenny D.A. (2018) Proc. of British Society of Animal Science, Dublin, Ireland.

Langfelder P., and Horvath S. (2008) BMC Bioinformatics 29(9):559. https://doi.org/10.1186/1471-2105-9-559

Mukiibi R., Vinsky M., Keogh K., Fitzsimons C., Stothard P., *et al.* (2018) Sci. Rep. 8(1):7303. https://doi.org/10.1038/s41598-018-25605-3

Weber K.L., Welly B.T., Van Eenennaam A.L., Young A.E., Porto-Neto L.R., *et al.* (2016) PLoS One 11(3):e0152274. https://doi.org/10.1371/journal.pone.0152274

62. Genetics parameters of feed efficiency in dairy goats, under commercial conditions

M. Chassier[1*], F. Mosnier[1], R. Rupp[2], B. Bluet[3], A. Bailly-Salins[4] and I. Palhière[2]

[1]Idele, Chemin de Borde Rouge, 31326 Castanet-Tolosan Cedex, France; [2]GenPhySE, Université de Toulouse, INRAE, INPT, ENVT, 31326 Castanet-Tolosan, France; [3]Idele, Chambre Régionale D'agriculture de Poitou-Charente CS 45002, 86550 Mignaloux-Beauvoir, France; [4]Capgènes, Agropôle 2135, Route de Chauvigny, 86550 Mignaloux-Beauvoir, France; marjorie.chassier@idele.fr

Abstract

The economic importance of genetically improving feed efficiency has been recognized in dairy production. Feed efficiency is a genetically complex trait that can be described as units of product output (e.g. milk yield) per unit of feed input. Previous work showed that feed efficiency is heritable in many species. This study aimed to investigate the heritability for Residual Energy Intake in two breeds of dairy goats (Alpine and Saanen) under commercial conditions. Heritabilities were moderate for both breeds (0.18±0.08 in Alpine and 0.20±0.07 in Saanen).

Introduction

Feeding costs are one of the largest expenditures in dairy production. For reducing their production costs and having a positive impact on environmental sustainability of dairy industry, breeders have increased interest in improving the balance between output (milk production) and input (feed intake). Feed efficiency has been introduced into pig and poultry breeding programmes and begins to be introduced in dairy cattle (Brito *et al*., 2021). In small ruminants (sheep and goat), feed efficiency is poorly documented and need to be evaluated. The European SMARTER (SMAll RuminanTs breeding for Efficiency and Resilience) project aims to develop innovative strategies to improve feed efficiency and farm resilience in small ruminants. This study has been conducted as part of this project and aims to estimate genetic parameters of feed efficiency under commercial conditions, for a large diversity of breeding systems in two breeds of dairy goats (Alpine and Saanen).

Materials & methods

Animals and housing. The experiment was performed in 14 commercial farms and at the Experimental Farm of La Sapinière (INRAE, Bourges), between 2019 and 2021. To date, 1,636 (663 Alpine and 973 Saanen) primiparous dairy goats were phenotyped for feed efficiency. Feed intake was recorded 4 times during the lactation: 2 times at the beginning of the lactation (between 0 and 60 DIM and between 60 and 90 DIM), around the reproduction (between 210 and 260 DIM) and at the end of the lactation (between 240 and 280 DIM). A total of 4,827 records (1,879 and 2,948 for Alpine and Saanen, respectively) were included in the dataset.

Animals were fed with different forages and concentrates, depending on the breeder. At each test day, feed intake was determined by weighing the total ration distributed and that wasted, by trained staff from the milk recording organisms. The forage quantity was measured by weighing all the offered forage, with a scale, at the batch or farm level (not individually). For concentrates, the quantity was measured either individually with automatic feeders or manually in milking parlour, or at the batch level by weighing all the offered concentrates, depending on the farm. Thus, for farms without individual distribution of concentrates, the individual feed intake was the average feed intake of the batch to which the animal belongs (83% of the dataset). For farms with individual distribution of concentrates, the individual feed intake was the average feed intake of the batch to witch the animal belongs for forage plus the individual intake of concentrates

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_62, © M. Chassier *et al.* 2022

(17%). Dry matter intake (DMI) was thus estimated from the information indicated on the concentrate labels and from forage analysis, for each animal and each test day. Energy Intake (EI) was estimated by multiplying DMI and energy concentration. Nutritional feed quality was recorded for each forage and each concentrate and energy content was given by INRAE (Agabriel, 2010). Test day milk recording data (milk yield, fat and protein contents) were also measured, at the same time than the feed intake control.

The chest width (CW) was used as a proxy of the body weight and was measured one time during the lactation (about 150 DIM). No body condition scores were performed.

REI estimation. To estimate feed efficiency, Residual Energy Intake (REI), was estimated as the residual of a linear regression model (1):

$$EI= \beta_0 + \beta_1 \times MY + \beta_2 \times FC + \beta_3 \times PC + \beta_4 \times CW + REI \quad (1)$$

Where, EI is the energy intake (expressed in Unité Fourragère Lait unit (UFL), 1 UFL=1.7 Mcal), β_0 is the intercept, β_1 is the regression coefficient for milk yield (MY), β_2 and β_3 are the regression coefficients of fat and protein contents (FC and PC) and β_4 is the regression coefficient for chest width (CW).

We classified the animals in 3 groups of REI, using standard deviation (sd_REI): inefficient (REI>0.5 × sd_REI), intermediate (-05 × sd_REI≤REI≤0.5 X sd_REI), efficient (REI<-0.5 × sd_REI).

Estimation of genetic parameters. The traits analysed were REI and MY. The genetic parameters were estimated, for each breed separately, using WOMBAT software (Mayer, 2007), with the following animal linear models:

$$Y=Flock + Camp + Htd + PhSt + a_n + perm_p + e \quad (2)$$

$$Y=Age + Camp + Htd + PhSt + a_n + perm_p + e \quad (3)$$

Where, Y is the observation vector for REI (2) or MY (3), Flock is the fixed effect of the flock, Age is the effect of the age at kidding, Camp is the fixed effect of the lactation campaign, Htd is the fixed effect of the herd test day, PhSt is the fixed effect of the physiological stage. The random effects included in the model were the additive genetic effect of the animal (a_n), the permanent environmental effect ($perm_p$) and the residual (e).

Only animals with two or more test day records are kept. The final data set comprised 1,331 and 2,414 test day records of 455 and 785 Alpine and Saanen goats, respectively. Animals were the progeny of 74 and 94 sires and 355 and 576 dams and the pedigree contained 7,484 and 8,652 animals for Alpine and Saanen breeds, respectively.

Results

A general description of the data is presented in Table 1. DMI was on average 2.7 kg with a standard deviation of 310 and 280 g in Alpine and Saanen breed, respectively. EI was on average 2.5 UFL for both breeds, with a moderate variability (CV of 13 and 11%). The residual energy intake (REI) was zero on average by definition. The milk yield was on average 3.40 l in Alpine breed and 3.04 l in Saanen breed with a standard deviation of 940 ml and 770 ml respectively. The mean chest width was 88 cm in the two breeds, with a low variability (CV of 5% for both breeds).

Table 1. Descriptive statistics of REI (residual energy intake), DMI (dry matter intake), energy intake (EI), milk yield (MY) and chest width (CW) per goat breed.

Breed	Trait	n	Min	Mean	Max	Sd[1]	CV[2]
Alpine	REI	1,879	-1.02	0.00	0.71	0.27	11%
	DMI (kg)	1,879	2.01	2.68	3.63	0.31	12%
	EI (UFL)	1,879	1.66	2.48	3.41	0.31	13%
	MY (L)	1,879	0.90	3.40	6.50	0.94	28%
	CW (cm)	1,879	76.00	88.24	103.00	4.29	5%
Saanen	REI	2,948	-0.87	0.00	1.01	0.28	11%
	DMI (kg)	2,948	2.02	2.82	3.78	0.28	10%
	EI (UFL)	2,948	1.78	2.62	3.85	0.29	11%
	MY (L)	2,948	0.30	3.04	6.60	0.77	25%
	CW (cm)	2,948	76.00	88.51	103.00	4.64	5%

[1] Standard deviation.

[2] Coefficients of variation.

Dry matter intake was higher in the inefficient group, and lower in the efficient group, with a difference of 0.49 kg of DMI per day in Alpine breed and 0.57 kg in Saanen breed between both groups (Table 2). This difference was explained both by a lower CDMI (concentrates DMI) and a lower FDMI (forage DMI) for the inefficient group. The ratio of milk output to DMI (DMI/MY) is lower for the efficient group. MY (results not shown) is the same in the three groups

Variance components, heritabilities and repeatability of the traits analysed are shown in Table 3. The heritability of test day milk yield was 0.19 and 0.20 in Alpine and Saanen breeds, respectively. Estimated heritabilities for REI, in both breeds, were moderate (0.18 and 0.20), with higher repeatability for Alpine breed (0.31) than Saanen breed (0.12).

Discussion

Feed efficiency is a difficult trait to measure under commercial conditions due to the diversity of feeding systems and the lack of individual feed intake data, due to the expensive cost of automatic feeders for breeders. Our results show that feed efficiency can be estimated, and that a ranking on the REI could be set up to classify goats: efficient *vs* inefficient. The heritability of test day milk yield were around 0.20 in

Table 2. Descriptive statistics of DMI (dry matter intake), energy intake (EI), ratio of milk output to DMI (DMI/MY), concentrates DMI (CDMI) and forage DMI (FDMI) per group of REI and per goat breed.

	Mean (sd)					
	Inefficient		Intermediate		Efficient	
Trait	Alpine (n=526)	Saanen (n=724)	Alpine (n=879)	Saanen (n=1,317)	Alpine (n=474)	Saanen (n=907)
DMI (kg)	2.86 (0.26)	3.11 (0.15)	2.73 (0.26)	2.86 (0.16)	2.37 (0.20)	2.54 (0.21)
EI (UFL)	2.71 (0.15)	2.98 (0.16)	2.55 (0.15)	2.64 (0.11)	2.08 (0.26)	2.32 (0.21)
DMI/MY	0.95 (0.25)	1.13 (0.44)	0.79 (0.23)	0.99 (0.25)	0.83 (0.30)	0.89 (0.37)
CDMI (kg)	1.29 (0.17)	1.02 (0.22)	1.24 (0.17)	0.95 (0.09)	1.03 (0.17)	0.79 (0.19)
FDMI (kg)	1.57 (0.38)	2.08 (0.19)	1.47 (0.36)	1.89 (0.14)	1.33 (0.20)	1.74 (0.21)

Table 3. Variance components: additive genetic variance (V_A), permanent environmental variance (V_E) and residual variance (V_R), heritabilities (h^2) and repeatability (R) for residual energy intake (REI) and milk yield (MY), with SEs in brackets.

Breed	Trait	V_A	V_E	V_R	H^2	R
Alpine	REI	0.004 (0.001)	0.006 (0.002)	0.009 (0.001)	0.18 (0.08)	0.31 (0.08)
	MY	0.096 (0.049)	0.186 (0.044)	0.227 (0.011)	0.19 (0.09)	0.37 (0.09)
Saanen	REI	0.003 (0.001)	0.002 (0.001)	0.011 (0.0001)	0.20 (0.07)	0.13 (0.06)
	MY	0.097 (0.036)	0.185 (0.031)	0.202 (0.007)	0.20 (0.07)	0.38 (0.07)

both Alpine and Saanen breeds. Arnal *et al.* (2019) found higher heritabilities of 0.27 and 0.28 on test day milk yield for the same breeds. The estimated heritability of REI were slightly lower than the feed efficiency heritability reported by Desire *et al.* (2017) in mixed-breed (Saanen, Alpine and Toggenburg) population (around 0.25). In Smarter project, feed efficiency was also investigated in dairy ewes, under commercial farms. Machefert *et al.* (2022) found a lower heritability (0.12) for a feed efficiency conversion ratio, in Lacaune ewes. According to Berry and Crowley (2013), heritability varies between studies depending on the type of model used to estimate feed efficiency (ratio traits or regression/residual traits) and on the type of animal involved (growing and adult animals). Further analysis needs to be conducting before considering a selection on REI in dairy goats.

Acknowledgements

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 772787 (SMARTER). The authors would like to thank the technical staff of milk recording, FCEL for providing us the data and the breeders involved in the SMARTER project.

References

Agabriel J., (2010) Alimentation des bovins, ovins et caprins – Besoins des animaux – Valeurs des aliments – Tables inra 2010. édition remaniée. Quae, France. https://www.quae.com/produit/753/9782759210114/alimentation-des-bovins-ovins-et-caprins.

Arnal M., Larroque H., Leclerc H., Ducrocq V., Robert-Granié C. (2019) Genet Sel Evol 51, 43. https://doi.org/10.1186/s12711-019-0485-3

Berry D.P., and Crowley J.J. (2013) J Anim Sci 91:1594-1613. https://doi.org/10.2527/jas.2012-5862

Brito L.F., Oliveira R. H., Houlahan K., Fonseca P.A.S., Lam S *et al.* (2020) Can J Anim Sci 100(4):587-604. https://doi.org/10.1139/cjas-2019-0193

Desire S., Mucha S., Mrode R., Coffey M., Broadbent J., Conington J. (2017) Proc of the 68th EAAP, Tallin, Estonia.

Köck A, Ledinek M, Gruber L, Steininger F, Fuerst-Waltl B,et al. (2018) J Dairy Sci 101(1):445-455. https://doi.org/10.3168/jds.2017-13281

Machefert C., Robert-Granié C., G. Lagriffoul, Astruc J.M, Parisot S., *et al.* (2022) Proc. 12 th World Cong. Appl. Livest. Prod. Submitted

Meyer K (2007) WOMBAT – A tool for mixed model analyses in quantitative genetics by restricted maximum likelihood (REML) Journal of Zhejiang University Science, 8:pp. 815-821. https://doi.org/10.1016/j.animal.2021.100202

63. Genetic analysis of predicted dry matter intake and gross feed efficiency in South African Holstein cows

M.A. Madilindi[1,2*], O.T. Zishiri[1], B. Dube[2] and C.B. Banga[2,3,4]

[1]Discipline of Genetics, College of Agriculture, Engineering and Science, University of KwaZulu-Natal, Private Bag X54001, Durban, 4000, South Africa; [2]ARC-Animal Production, Private Bag X2, Irene, 0062, South Africa; [3]Department of Animal Sciences, Faculty of Science, Tshwane University of Technology, Private Bag X680, Pretoria, 0001, South Africa; [4]Department of Agriculture and Animal Health, University of South Africa, Private Bag X6, Florida 1710, South Africa; matomemadilindi@gmail.com

Abstract

Genetic parameters were estimated for dry matter intake and gross feed efficiency predicted from milk components in the first three parities of South African Holstein cattle. Test-day data from the South African National Milk Recording Scheme were used to predict dry matter intake (pDMI) from milk yield. Predicted gross feed efficiency (pGFE) was then calculated from pDMI and energy-corrected milk (ECM). Variance components and genetic parameters were estimated by a repeatability animal model using the ASReml program. Heritability estimates ranged from 0.05±0.02 for pGFE in mid lactation to 0.13±0.03 for pDMI in late lactation. Overall estimates of heritability across parities were 0.08±0.02 and 0.13±0.02 for pGFE and pDMI, respectively. Genetic correlations between pDMI and pGFE were moderate and negative in early (-0.42±0.24) and mid lactation (-0.20±0.24), and low and positive in late lactation (0.05±0.17). There was a decrease in genetic correlation between pDMI and pGFE with increase in parity, from 0.26±0.16 to -0.09±0.17. The low heritability estimates for pDMI and pGFE indicate low accuracy of selection for these traits in South African Holstein cattle. This can, however, be improved through multi-trait analyses including both traits and others that are correlated.

Introduction

Feed is a major cost in dairy production; hence, increased feed efficiency is key to achieving sustainable improvements in herd profitability. Better efficiency of feed utilisation is also associated with a reduction of major greenhouse gas emissions and, therefore, may contribute towards mitigating climate change (Connor, 2015). Thus, genetic improvement of feed efficiency could be a solution to increasing environmental sustainability and herd profitability of the dairy industry (Berry *et al.*, 2015).

Genetic selection for feed efficiency is a challenging task that requires the phenotypic measurement of many animals in order to accurately predict the genetic merit of each animal for the trait. Generally, it is costly and difficult to measure dry matter intake (DMI) directly in lactating cows, in order to calculate gross feed efficiency (GFE). This challenge may, however, be overcome through selection on DMI and GFE predicted from routinely measured milk production traits (Zhang *et al.*, 2020; Madilindi *et al.*, 2021). A major prerequisite to the effective use of such predicted traits as selection criteria is that they should exhibit genetic variation. Hence, the primary aim of this study was to estimate genetic (co)variances and related parameters for predicted dry matter intake and gross feed efficiency in South African Holstein cattle. Such parameters could provide the basis for achieving genetic improvement of feed efficiency in this population.

Materials & methods

Data. Data were from 568 herds participating in the South African National Milk Recording Scheme, and were extracted from the Integrated Registration and Genetic Information System of South Africa. The original data set consisted of 1,739,261 test-day records of 140,345 South African Holstein cows that calved

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_63, © M.A. Madilindi *et al.* 2022

between 2009 and 2019. Only the first three parities were considered, consisting of 440,062 test-day records of 62,695 cows, after data editing.

Prediction of dry matter intake and gross feed efficiency. Firstly, test-day milk yield (kg/day) was used to calculate daily-predicted dry matter intake (pDMI) (kg/day) per cow, based on the simple regression prediction Equation 1, developed by Lindgren *et al.* (2001). Preliminary analysis using South African Holstein data had found this equation to be reliable, as indicated by an R^2 value of 0.76.

$$pDMI\ (kg/day) = 5.7 + 0.43 \times Milk\ (kg/day) \quad (1)$$

Next, energy-corrected milk (ECM) yield (kg/day) per cow was calculated from test-day milk yield (kg/day), protein percent (%), butterfat percent (%) and lactose percent (%), using Equation 2, according to Kirchgeβner (1997).

$$ECM\ (kg/day) = \text{milk yield (kg) x } \frac{(0.39 \text{ x butterfat \%} + 0.24 \text{ x protein \%} + 0.17 \text{ x lactose \%})}{3.17} \quad (2)$$

Predicted gross feed efficiency (pGFE) per cow was subsequently calculated, using Equation 3:

$$pGFE = \frac{\text{ECM (kg/day)}}{\text{pDMI (kg/day)}} \quad (3)$$

Statistical analysis. Variance components, heritability, and repeatability for pDMI and pGFE were estimated by ASReml 3.0 (Gilmour *et al.*, 2009) using a repeatability animal model (Equation 4):

$$y_{jklmn} = \mu + HTD_j + LS_k + P_l + \beta ACC_{jklmn} + a_{jklmn} + p_{jklmn} + e_{jklmn,} \quad (4)$$

where y_{jklmn} is an observation of pDMI or pGFE; μ is the overall mean; HTD_j is the fixed effect of herd-test-day contemporary group; LS_k is the fixed effect of lactation stage (k = early lactation (10-100 days in milk (DIM)), mid lactation (101-200 DIM), and late lactation 201-305 DIM); P_l is the fixed effect of parity (l =1 to 3); ACC_{jklmn} is age of cow at calving; β_{jklmn} is a linear regression coefficient on age of cow at calving; a_{jklmn} is the random animal additive genetic effect with var (a) ~distribution $N(0,A\sigma^2_a)$, where σ^2_a is the additive genetic variance and A is the relationship matrix; p_{jklmn} is the random permanent environmental effect of the animal with distribution $N(0,I\sigma^2_p)$, where σ^2_p is the permanent environmental variance; e_{jklmn} is the random residual effects with distribution $N(0,I\sigma^2_e)$, where σ^2_e is the residual variance and I is an identity matrix.

Bivariate analyses were carried out to estimate genetic correlations between pDMI and pGFE within and across lactation stages as well as parities.

Results

Estimates of heritability and genetic correlations for pDMI and pGFE by stage of lactation are presented in Table 1. Heritabilities ranged from 0.05±0.02 for pGFE in mid lactation to 0.13±0.03 for pDMI in late lactation. The heritability estimates of pDMI tended to increase with stage of lactation. No trend of heritability estimates for pGFE was observed among lactation stages. Genetic correlations between pDMI and pGFE varied from -0.42±0.24 in early to 0.05±0.17 in late lactation.

Table 2 contains estimates of heritability and genetic correlations for pDMI and pGFE within parity. Heritability estimates ranged from 0.07±0.02 for pGFE in both first and third-parity to 0.14±0.03 for pDMI in third-parity. There was a slight increase in heritabilities of pDMI with increase in parity. No trend of heritability estimates for pGFE was observed across parities. Overall estimates across parities

Table 1. Heritability estimates (diagonals) and genetic correlations (off diagonals) for daily-predicted dry matter intake and gross feed efficiency within lactation stage in South African Holstein cattle.[1]

	Early		Mid		Late	
Traits	**pDMI**	**pGFE**	**pDMI**	**pGFE**	**pDMI**	**pGFE**
pDMI[1]	**0.07±0.02**	-0.42±0.24	**0.09±0.02**	-0.20±0.24	**0.13±0.03**	0.05±0.17
pGFE[2]		**0.08±0.02**		**0.05±0.02**		**0.09±0.02**

[1] pDMI[1] = predicted dry matter intake; pGFE = predicted gross feed efficiency.

were 0.08±0.02 and 0.13±0.02, respectively, for pGFE and pDMI. Repeatability estimates of 0.26±0.01 and 0.41±0.01 were obtained across lactations, respectively, for pGFE and pDMI. Genetic correlations between pDMI and pGFE varied from -0.09±0.17 in third-parity to 0.26±0.16 in first-parity.

Discussion

Heritabilities for pDMI and pGFE were generally low within lactation stages. The heritability estimates of pDMI slightly increased with stage of lactation, implying that accuracy of selection for this trait may depend on the lactation stage in which selection is applied. Tarekegn *et al.* (2021) similarly reported an increase in heritability estimates for measured daily DMI with lactation week. No trend of heritability estimates for pGFE was observed among lactation stages in the current study. However, a decrease in the heritability estimate for GFE with increase in lactation stage was observed previously by Spurlock *et al.* (2012). These discrepancies may be a reflection of differences in the methods used to measure the traits. The genetic correlations between pDMI and pGFE was inexplicably negative in early and mid-lactations, and positive in late lactation. There were, unfortunately, no estimates available in the literature with which to compare these results. The relatively high genetic correlation in early lactation may assist in increasing accuracy of selection, if the two traits are included in a multi-trait analysis.

Heritability estimates for pDMI and pGFE were low within first-three parities. A slight increase in heritabilities of pDMI with increase in parity was observed. This indicates that accuracy of selection for pDMI may depend on the parity in which selection is applied. Recently, Zhang *et al.* (2020) reported heritability estimates comparable with those obtained in the current study for pDMI in the first-two parities, although the estimate for the first parity was higher than for the second. No trend of heritability estimates for pGFE was observed across parities in our study. Spurlock *et al.* (2012) however reported higher heritability estimates for actual GFE in primiparous (0.47) compared to multiparous (0.43) cows, although their estimates were much larger than those found in the current study. The inconsistency in estimates of heritability for DMI and GFE among studies could be attributable to discrepancies in the measurement/prediction methods as well as stages of lactation during which measurements were taken. Moderate repeatability estimates were obtained across lactations, implying that first lactation records on

Table 2. Heritability estimates (diagonals) and genetic correlations (off diagonals) for daily-predicted dry matter intake and gross feed efficiency within first-three parities in South African Holstein cattle.

	Parity 1		Parity 2		Parity 3	
Traits	**pDMI**	**pGFE**	**pDMI**	**pGFE**	**pDMI**	**pGFE**
pDMI[1]	**0.11±0.03**	0.26±0.16	**0.12±0.03**	0.11±0.16	**0.14±0.03**	-0.09±0.17
pGFE[2]		**0.07±0.02**		**0.10±0.02**		**0.07±0.02**

[1] pDMI[1] = predicted dry matter intake; [2] pGFE = predicted gross feed efficiency.

pDMI can fairly predict pDMI in later lactations. Genetic correlation between pDMI and pGFE decreased with increase in parity in the current study. This may mean that increased accuracy of selection through multiple-trait analysis of the two traits would be more effective with first lactation data. Genome-wide association studies may be warranted to validate these results.

Conclusions

Results of the current study show that there is low genetic variation in DMI and GFE predicted from milk components, which implies low accuracy of selection. Moderate genetic correlations between the traits, however, may be used to increase accuracy of selection in multi-trait analyses. Variation in pDMI appears to slightly vary with parity and stage of lactation; thus, lactation stage-specific and parity-specific selection may be required. Further research may be required to identify genomic regions, and subsequently genes, associated with these predicted feed efficiency traits, for possible marker-assisted selection, which might accelerate genetic improvement.

References

Berry D.P., Kennedy E., and Crowley J.J. (2015) Genetics of feed intake and reproduction. Pages 502-522 in The Genetics of Cattle. D.J. Garrick and A. Ruvinsky, ed. CABI, Wallingford, UK.

Connor, E.E. (2015) Animal 9:395-408. https://doi.org/10.1017/S1751731114002997

Gilmour A.R., Gogel B.J., Cullis B.R., and Thompson R. (2009) ASREML User's Guide Release 3.0. VSN International Ltd., Hemel Hempstead, U.K.

Lindgren E., Murphy M., and Andersson, T. (2001) Värdering av foder. Lantmännen Foderutveckling AB, Nötfor. Almqvist and Wiksell. Uppsala, Sweden.

Kirchgeβner M. (1997) Tierernährung. 10th ed. DLG-Verlag, Frankfurt, Germany.

Madilindi M.A., Banga C.B., Dube B., and Zishiri O.T. (2021) Proc. of the 11th UKZN PRIS, Durban, South Africa.

Spurlock D.M., Dekkers J.C.M., Fernando R., Koltes D.A., and Wolc, A. (2012) J. Dairy Sci. 95:5393-5402. https://doi.org/10.3168/jds.2012-5407

Tarekegn G.M., Karlsson J., Kronqvist C., Berglund B., Holtenius K. *et al.* (2021) J. Dairy Sci. 104:4424-4440. https://doi.org/10.3168/jds.2020-19224

Zhang L, Gengler N., Dehareng F., Colinet F., Froidmont E. *et al.* (2020) Animals 10(873):1-13. https://doi.org/10.3390/ani10050873

64. Single-step genomic prediction for metabolic body weight of Nordic Red dairy cattle

R.M. Kempe[1*], T. Mehtiö[1], A.-M. Leino[1], M. Koivula[1], T. Pitkänen[1], R.B. Stephansen[2], E.A. Mäntysaari[1], E. Negussie[1], J. Pösö[3], U.S. Nielsen[4], W.F. Fikse[5], G.P. Aamand[6] and M.H. Lidauer[1]

[1]Natural Resources Institute Finland (Luke), Myllytie 1, 31600 Jokioinen, Finland; [2]Aarhus University, QGG, Blichers Alle 20, 8830 Tjele, Denmark; [3]Faba co-op, P.O. Box 40, 01301 Vantaa, Finland; [4]SEGES Cattle, Agro Food Park 15, 8200 Aarhus N, Denmark; [5]Växa, Box 288, 75105 Uppsala, Sweden; [6]Nordic Cattle Genetic Evaluation, Agro Food Park 15, 8200 Aarhus N, Denmark; riitta.kempe@luke.fi

Abstract

The aim of this study was to predict genomic breeding values for metabolic body weight (MBW) by a single-step model and to compare obtained breeding values to those from a multi-step genomic prediction model and a pedigree-based BLUP model. Data contained over 3 million observations from Danish, Finnish and Swedish Red dairy cattle born in the years 1990 to 2017. The applied multiple-trait animal models included six traits: 1st, 2nd, and 3rd parity MBW, and first parity stature, chest width and body depth. Breeding values from the single-step model were in good agreement with those from the multi-step model (correlations were 0.98-0.99), but BLUP breeding values differed from genomic breeding values slightly for bulls and clearly for cows. The single-step model gave higher validation reliabilities for bulls and cows than the multi-step model. The results support using a single-step approach for the genomic evaluation of MBW.

Introduction

In 2020, Nordic Cattle Genetic Evaluation (NAV) included the Saved Feed Index, which contains the metabolic body weight (MBW) and residual feed intake in the Nordic Total Merit index for the Danish, Finnish, and Swedish Holstein (HOL), Red dairy cattle (RDC) and Jersey populations. MBW is the core trait for the genetic evaluation of maintenance efficiency as the feed requirement for maintenance is assumed to have a linear relationship with MBW. The MBW evaluation (Lidauer *et al.*, 2019) is based on three core traits; MBW in 1st, 2nd, and 3rd parity, and three correlated conformation traits; stature (ST), chest width (CW), and body depth (BD). Currently, genomic breeding values for MBW are predicted by a multiple-step approach but a single-step genomic prediction model (Aguilar *et al.*, 2010; Christensen and Lund, 2010) is expected to improve predictions. The aim of this study was to develop a single-step genomic prediction model for MBW in Nordic RDC. Cross-validation statistics and regression of recent breeding values on earlier breeding values (Legarra and Reverter, 2018) were applied to compare the developed single-step model with the currently used multiple-step model, and with the pedigree-based BLUP model.

Materials & methods

Phenotypic data and pedigree. The data contained all available body weight (BW) and conformation traits observations of Danish, Finnish and Swedish RDC cows born in the years 1990 to 2017. This included BW observations from Finnish cows, which were measured by tape (heart girth), BW observations from Danish cows, which were measured by scale, and first parity conformation observations from cows of all countries. The raw BW observations were pre-processed as explained by Lidauer *et al.* (2019) to obtain for each lactation, that had BW records, one MBW observation ($BW^{0.75}$) with an associated weighting coefficient that accounted for the type and number of the raw BW observations. The first parity conformation observations were taken from the NAV routine conformation evaluation (NAV, 2019) with the following trait definition: ST (linear, cm), CW (linear score from 1 to 9) and BD (linear score from 1 to 9). After editing, the data consisted of 1.28 million MBW observations and 2.94 million conformation observations.

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_64, © R.M. Kempe *et al.* 2022

The pedigree of the cows with observations included 2.99 million animals and the base population was described with 182 unknown parent groups (UPG).

Best linear unbiased prediction model. The BLUP model developed for the NAV routine MBW evaluation was the basis for building the single-step genomic predication model. The BLUP model is a multiple-trait animal model with six traits, which are MBW in 1st, 2nd, and 3rd parity, and first parity ST, CW, and BD. The applied heritabilities for these traits were 0.46, 0.51, 0.56, 0.60, 0.18, and 0.26, respectively. A detailed description of the model and the applied variance components is given in Lidauer *et al.* (2019). The correlated information from the conformation traits is important because there is no BW information available from Swedish cows. Moreover, recording of BW by tape is reducing in Finland and measuring BW by scale is slowly becoming more common. Genetic correlations between the three MBW traits and the conformation traits were on average 0.67, 0.55, and 0.49 with ST, CW, and BD, respectively. Furthermore, to improve predictions of Finnish herds the observations from HOL cows were also included to increase contemporary group sizes. The estimated breeding values (EBV) for MBW were used to form for each animal one combined MBW EBV by weighting 1st, 2nd, and 3rd parity EBVs with the weights 0.30, 0.25, and 0.45, respectively, and changing the sign of the index values so that a higher index value refers to a lower MBW.

Marker data. Genotype information from RDC animals only was included for developing the single-step model. Bulls were genotyped using the Illumina Bovine SNP50 Bead Chip (Illumina, San Diego, CA) and the cows using the lower-density EuroG MD chip (http://www.eurogenomics.com/). The unobserved genotypes with the lower-density chip were imputed to the 50K density. After editing there were 46,914 single nucleotide polymorphisms (SNP) markers available for 51,417 animals.

Multi-step genomic prediction model. For building the multi-step genomic prediction model, which was based on SNP-BLUP (msSNPBLUP), all genotyped bulls which had at least 20 daughters and a MBW EBV reliability of at least 0.5 and all genotyped cows that had observations were included. This resulted in 5,554 bulls and 43,276 cows and in total 48,830 genotyped animals. De-regressed proofs (DRP) based on the combined EBV for MBW of 1st, 2nd, and 3rd parity were used as observations for the SNP-BLUP evaluation and each animal's observation was weighted by the animal's effective record contribution (ERC). In addition to marker effects, the SNP-BLUP model included a polygenic effect explaining 10% of the genetic variation.

Single-step genomic prediction model. Relationships between animals in the single-step genomic prediction model (ssGBLUP) were described by the H matrix (Aguilar *et al.*, 2010; Christensen and Lund, 2010) which was implemented as described by Koivula *et al.* (2021). The VanRaden method I (VanRaden, 2008) was used for building the genomic relationship matrix by blending the G matrix with the residual polygenic effect that accounts for 10% of the genetic variation. The QP transformation was applied to account for the UPG. A combined genomic enhanced breeding value (GEBV) for MBW was formed in the same way as the combined EBV for MBW.

Model validation. The ssGBLUP model was validated by forward prediction cross-validation, which was done for all the three models. For the cross-validation, we selected 1,721 cows and 354 bulls that had no own information in the reduced data set. For the evaluation with the reduced data, observations from most recent four years of bulls and their progenies were excluded. The same pedigree and genomic information were used as for the full data set evaluation to obtain breeding values for candidates (BV_c). Thus, in the SNP-BLUP evaluation for the reduced data all candidate animals were included but had no own data. The cross-validation reliability (r^2_{cv}) was calculated as:

$$r^2_{cv} = \text{corr}(DRP, BV_c)^2 / r^2_{DRP} \quad (1)$$

where DRP were calculated separately for both the bull validation candidate, and the cow validation candidate group, to ensure proper calculation of ERC weights used for the de-regression. The second statistic we applied was the regression of full data breeding values on reduced data breeding values (Legarra and Reverter, 2018), which has an expectation of b_1=1.0.

Results

Correlations between ssGBLUP and msSNPBLUP breeding values of genotyped animals were 0.98-0.99, indicating that the models gave almost the same genomic breeding values (Table 1). The correlation between BLUP and ssGBLUP or msSNPBLUP breeding values was 0.97 in bulls and 0.84 to 0.86 in cows, respectively. Thus, BLUP breeding values were slightly different in bulls and clearly different in cows compared with genomic breeding values. The ssGBLUP GEBV had 5 to 12% higher standard deviation compared with msSNPBLUP or BLUP breeding values in cows and 0.2 to 5.5% higher in bulls.

The correlations between candidates' BV_c and their future DRP are presented in Table 2. Correlations were the highest when BV_c were estimated with ssGBLUP for both bulls and cows. This resulted also the highest r^2_{cv} with ssGBLUP being as high as 0.91 for the bull candidates. The r^2_{cv} of BV_c obtained with the msSNPBLUP model was higher than that when using BLUP in candidate bulls, but not in cows. For all three models, r^2_{cv} was higher for bulls compared with cows. Mean DRP reliability was for bulls 0.71 and for cows 0.61.

Table 1. Standard deviation (Std) and correlation between MBW breeding values and those with de-regressed proofs (DRP) for reference population animals.

	Prediction model[1]	Std	Correlations		
			msSNPBLUP	ssGBLUP	DRP
Cows	BLUP	4.11	0.86	0.84	0.88
n=35,075	msSNPBLUP	4.36		0.98	0.75
	ssGBLUP	4.59			0.74
Bulls	BLUP	4.56	0.97	0.97	0.95
n=5,539	msSNPBLUP	4.33		0.99	0.90
	ssGBLUP	4.57			0.91

[1] BLUP = best linear unbiased prediction; msSNPBLUP = multi-step genomic BLUP, ssGBLUP = single-step genomic BLUP.

Table 2. Cross-validation estimates correlation ($r_{(DRP,BVc)}$), validation reliability (r^2_{cv}), regression coefficient (b_1) for the bull and cow candidate groups with different models.

	Prediction model[1]	Cross-validation reliability		Legarra-Reverter	
		$r_{(DRP,BVc)}$	r^2_{cv}	b_1	R_2
Cows	BLUP	0.67	0.74	0.96	0.51
n=1,721	msSNPBLUP	0.65	0.69	1.09	0.70
	ssGBLUP	0.73	0.86	1.05	0.84
Bulls	BLUP	0.74	0.77	0.96	0.56
n=354	msSNPBLUP	0.75	0.80	1.10	0.74
	ssGBLUP	0.81	0.91	1.01	0.85

[1] BLUP = best linear unbiased prediction; msSNPBLUP = multi-step genomic BLUP, ssGBLUP = single-step genomic BLUP.

The b_1 estimates were closer to the expectation for ssGBLUP in candidate bulls than those for the BLUP model and msSNPBLUP model (Table 2). In cows, the *b1* estimate for BLUP was slightly closer to expectation than for the ssGBLUP or msSNPBLUP models. The b_1 estimates for msSNPBLUP were larger than 1.0 indicating that BV_c underpredicted the future breeding values, whereas the b_1 estimates for BLUP were below 1.0 indicating that BV_c overpredicted future breeding values.

Discussion

Breeding values from the single-step model were in good agreement with those from the multi-step model and reasonably different from those of the BLUP model. Correlations between traditional BLUP and genomic models were clearly lower than correlations among genomic models. Also, country-wise (Denmark, Finland, Sweden) investigation of the breeding values confirmed that the single-step model is modelling the data appropriately. The obtained cross-validation reliabilities were higher than those for RDC yield traits ssGBLUP GEBVs (Koivula *et al.* 2018). This can be explained by the high heritability of the MBW traits and ST, and the high genetic correlation between MBW and ST. The reason why we obtained a lower r^2_{cv} value for the msSNPBLUP evaluation compared with the BLUP evaluation for the cow candidate group was unclear and requires further assessments.

The regression of candidates' full data GEBV on candidates reduced data BV_c were close to the expectation when applying ssGBLUP, and therefore ssGBLUP is preferable over msSNPBLUP for obtaining GEBV for candidates. The results of this study support the implementation of ssGBLUP for the genomic evaluation of MBW in Nordic RDC and to develop ssGBLUP also for the Nordic HOL and Jersey MBW evaluation.

References

Aguilar I., Misztal I., Johnson D.L., Legarra A., Tsuruta S., *et al.* (2010) J. Dairy Sci. 93(2):743–752. https://doi.org/10.3168/jds.2009-2730

Christensen O.F., and Lund M.S. (2010) Genet. Sel. Evol. 42:2. https://doi.org/10.1186/1297-9686-42-2

Koivula M., Strandén I., Aamand G.P. and Mäntysaari, E.A. (2018) J. Anim. Breed. Genet. 135(2):107-115. https://doi.org/10.1111/jbg.12318

Koivula M., Strandén I. Aamand G.P. and Mäntysaari E.A. (2021) J. Dairy Sci. 104(9):10049–10058. https://doi.org/10.3168/jds.2020-19821

Legarra A., and Reverter A. (2018) Genet. Sel. Evol. 50:53. https://doi.org/10.1186/s12711-018-0426-6

Lidauer M.H., Leino A.-M., Stephansen R.S., Pösö J., Nielsen U.S., *et al.* (2019) Interbull Bulletin no. 55:21-25. https://journal.interbull.org/index.php/ib/issue/view/78

NAV (2021) NAV routine genetic evaluation of dairy cattle – data and genetic models (12th ed). Available on: https://nordicebv.info/wp-content/uploads/2021/10/NAV-routine-genetic-evaluation_EDITYSS-08102021.pdf

VanRaden P. M. (2008) J. Dairy Sci. 91: 4414–4423. https://doi.org/10.3168/jds.2007-0980

65. Application of imputed data and single-step GBLUP in the study of residual feed intake trait of Nellore cattle

J.A.II.V. Silva[1,2*], M.H.V. Oliveira[1], A.M. Castilhos[1], A.M. Toro-Ospina[3], H. Srihi[4] and L. Varona[4]

[1]School of Agricultural and Veterinarian Sciences, Unesp, 14884-900, Jaboticabal, Brazil; [2]National Council for Scientific and Technological Development, CNPq, Brasília, 71605-001, Brazil; [3]Universidad de la Amazonia, Sede Porvenir, 180001, Florencia Caqueta, Colombia; [4]Zaragoza University, 50013, Zaragoza, Spain; augusto.ii@unesp.br

Abstract

The availability of genome high density (HD) data provides new opportunities for combined population genomic prediction. An imputation from Nellore animals has been implemented to analyze, the impact of imputed data on heritability estimation (h^2) and breeding values for residual feed intake (RFI), by single-step GBLUP. Data sets of 607 Nellore bulls were used to impute a genotype bank composed of 1,466 Nellore animals, resulting in a dataset of 735,964 SNPs for all animals. The BLUPF90 package was used to perform quality control, resulting in 471,301 SNPs. Analysis to obtain h^2 and breeding values for RFI were performed, using an animal model considering the single-step GBLUP. The BLUPF90 was efficient in processing the imputed file. The higher h^2 of the RFI suggests an influence of the imputed data. The correlation of the breeding values indicates that the use of HD and 56K data result in similar breeding values.

Introduction

Zebu breeds contribute considerably to the agricultural economy (milk and beef) in tropical countries. The Nellore breed (*Bos indicus*) is the main research focus for the beef sector in Brazil and has several genetic improvement programs with private and public coordination.

The scientific advancement and productive research of the taurine beef breeds far surpasses the Zebu breeds. In addition, currently available SNP matrices for genotyping are constructed from dbSNP data, therefore, are biased towards taurine animals limiting genomic studies of indicus breeds (Iqbal *et al.*, 2019). To overcome the taurine preference, some genotyping companies have recently introduced low- and medium-density markers for indicus breeds, but still suffering from verification bias (Boison *et al.*, 2015, Ospina *et al.*, 2021). Additional studies that resemble the entire genome of indicus races from different geographic regions are needed to discover and deposit more indicative SNPs (Choi *et al.*, 2014).

Few studies were performed in Zebu cattle (Stafuzza *et al.*, 2017, Fernandes Junior *et al.*, 2021). The search for alternatives for the imputation implementation data in genomic analysis is essential for its use and, consequently, a reduction in costs.

Several studies have studied residual feed intake (RFI) using genomic data, identifying important regions for selection. However, larger studies at the genome level, mainly in Zebu breeds, are of importance for tropical countries. Research on Nellore breed has increased, but better knowledge of the breed is required to improve traits of economic interest. Thus, studies with Nellore data will contribute for understanding of Zebu breeds genome.

The aim of the study was to implement genotype imputation from Nellore animals, and to analyse by single-step GBLUP the impact of imputed data on the results of h^2, and breeding values for RFI.

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_65, © J.A.II.V. Silva *et al.* 2022

Materials & methods

The RFI has been implemented as a selection criterion since 2010 by the Nellore Qualitas Program (group of farms in seven states in Brazil). From 15,000 males produced annually, 120 animals are selected, based on an index (based on breeding values of the traits weight at weaning, post gain weaning, muscle score and scrotal circumference), to participate in a feed efficiency test. At the end of the test, the young bulls, with an average age of 23 months are selected, have the semen collected and the breeders participating in the program use it in a progeny test. RFI is defined as the difference in observed feed consumption minus predicted and is measured in 56 days of testing, after 21 days of adaptation.

Data from 607 Nellore bulls genotyped with HD (~777k) were used to impute a genotype bank composed of other 1,466 Nellore animals from the Nellore Qualitas Program. The imputation was performed using the FImpute v3 software (Sargolzaei *et al.*, 2014), being performed considering a deterministic approach for the haplotypes phase and imputing all the absent genotypes, considering the pedigree information.

Analysis to obtain h^2, GWAS, and breeding values for the RFI trait were performed with the BLUPF90 package (Misztal *et al.*, 2001), using an animal model considering the single-step GBLUP (SSGBLUP, Aguilar *et al.* 2010) method as follows:

$$y = X\beta + Za + e, \quad (1)$$

where y is the vector of RFI; β is a vector with the fixed effects of CG (defined by animals participating in the test in the same year and pen) and age at the beginning of the test as a linear co-variable; a is a vector of additive genetic effects assumed normally distributed N (0, H ⊗ σ_a^2), in which σ_a^2 is the additive genetic variance and H is the pedigree-genomic relationship matrix ⊗ is the Kronecker product and e is a residual vector assumed N (0, I ⊗ σ_e^2), where I is the identity matrix, and σ_e^2 is the residual variance. The X and Z are known as incidence matrices related to fixed and additive genetic effects, respectively.

The H is a matrix that combines pedigree and genomic information, and its inverse (H^{-1}) is given by: $H^{-1} = A^{-1} + \begin{bmatrix} 0 & 0 \\ 0 & G^{-1} - A_{22}^{-1} \end{bmatrix}$, where A^{-1} is the inverse of the pedigree relationship matrix, A_{22}^{-1} represents the inverse of the pedigree relationship matrix for genotyped animals, and G^{-1} is the inverse of the genomic relationship matrix. The pedigree file contained 7,745 animals, generated from animals with phenotypes and up to the fourth generation.

After imputation, and the BLUPF90 family programs were used to perform a new quality control, removing 0 SNPs with a call rate less than 0.90, with MAF less than 0.05 were removed 264,663 SNPs, monomorphic 8,212 SNPs were dropped and there were no animals with a call rate less than 0.90, resulting in 471,301 SNPs of and 2,073 individuals. The correlation between diagonal, all, and off-diagonal elements of G and A were equal to 0.027, 0.766 and 0.821, respectively. From this phase, the estimation of h^2 of the RFI was obtained (AiremlF90 program), the GWAS analyses (results not shown), all SNPs were ranked based on their *P*-values (results not shown) and obtained the genetic values and accuracies of the animals.

Results

The data set with 735,964 SNPs to be imputed had mean MAF equal to 0.566 (minimum of 0.0 and maximum of 1), and after quality control, the file contained 471,301 SNPs, with average MAF equal to 0.539 (minimum of 0.05 and maximum of 0.95).

The RFI h^2 estimate using the imputed data for WGS was equal to 0.29±0.08, a value higher than that obtained with the pedigree alone (0.23±0.07).

The distribution of breeding value for RFI is reported in Figure 1 for three groups (animals genotyped with the HD, 56K and 29K panel) from using only the pedigree file (BV_BLUP), and the pedigree and the imputed data for HD (BV_HD). Low correlation was observed between the breeding values for the lower density panel. The behaviour of all panels of higher density was like those shown (Figure 1) with lower density.

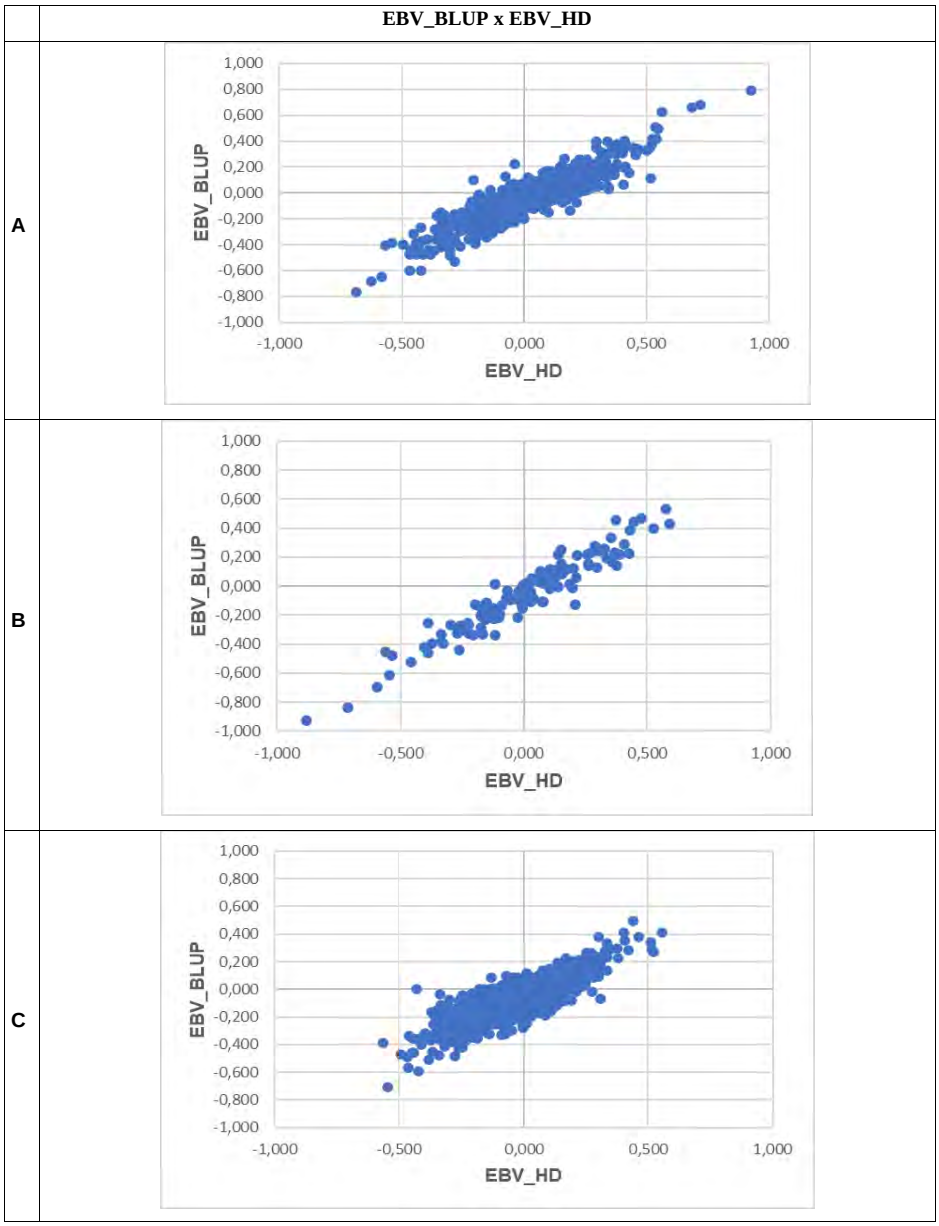


Figure 1. Distribution of expected breeding values for residual feed intake from animals genotyped with the HD (A), 56K (B) and 29K (C) panel from using only the pedigree file (BV_BLUP) and the pedigree and the imputed data for HD (BV_HD).

Discussion

The BLUPF90 package was efficient and fast in processing the imputed file for HD with 471,301 SNPs, and building the H^{-1}, as well as estimating the h^2, GWAS, and breeding values for the RFI trait.

The high MAF mean in the imputed set for WGS indicates that the number of SNPs increased with the high MAF, that is, the number of rare variants decreased.

The h^2 of the RFI higher than that obtained with only the pedigree suggests an influence of the imputed data for HD. Wainschtein *et al.* (2021) suggest that rare variants, in regions with low LD, are important in estimating h^2 for complex traits.

The correlation of the breeding values indicates that the use of HD and 56K data result in similar breeding values (Figure 1). However, the results showing a larger dispersion of EBV from genotyped animals with low-density panel (29K) suggest low quality of imputation.

Acknowledgements

This work was supported by Sao Paulo Research Foundation (FAPESP grants: #2017/10630-2, # 2009/16118-5, and 2017/23633-0). We also thank the Nellore Qualitas Program and group of farms that maintains the phenotype and genotypes data.

References

Aguilar I., Misztal I., Johnson D.L., Legarra A., Tsuruta S., *et al.* (2010). J. Dairy Sci. 93(2), 743–752. https://doi.org/10.3168/jds.2009-2730

Boison S.A., Santos D.J.A., Utsunomiya A.T.H., Carvalheiro R., Neves H.R., *et al.* (2015) J. Dairy Sci. 98(7), 4969-4989. https://doi.org/10.3168/jds.2014-9213

Choi J.W., Liao X., Stothard P. Chung W.H., Jeong H.J., *et al.* (2014) PLoS One. 9, e101127. https://doi.org/10.1371/journal.pone.0101127.

Fernandes Júnior G.A., Carvalheiro R., Oliveira H.N., Sargolzaei M., Costilla R. *et al.* (2021) Genet Sel Evol. 53: 27. https://doi.org/10.1186/s12711-021-00622-5

Iqbal N., Liu X., Yang T., Huang Z., Hanif Q., *et al.* (2019) PLoS One. 4, e0215065. https://doi.org/10.1371/journal.pone.0215065

Larmer S.G., Sargolzaei M., Brito L.F., Ventura R.V., Schenkel F.S. (2017) BMC Genet 18, 120. https://doi.org/10.1186/s12863-017-0588-1.

Misztal I., Tsuruta S., Strabel T., Auvray B., Druet T., *et al.* (2002). In: 7th World Congress on Genetics Applied to Livestock Production. Montpellier, France, 2002. http://wcgalp.org/system/files/proceedings/2002/blupf90-and-related-programs-bgf90.pdf

Ospina A.M.T., Aguilar I., Oliveira M.H.V., Correia L.E.C.S, Vercesi Filho A.E. *et al.* (2021) Genome. 64: 893–899 https://doi.org/10.1139/gen-2020-0081

Sargolzaei M, Chesnais JP, Schenkel FS. (2014) BMC Genomics 15, 478 (2014). https://doi.org/10.1186/1471-2164-15-478.

Stafuzza N.B., Zerlotini A., Lobo F.P., Yamagishi M.E.B., Chud T.C.S., *et al.* (2017) PLoS One. 12, e0173954.. https://doi.org/10.1371/journal.pone.0173954.

Wainschtein P., Jain D., Zheng Z., TOPMed Anthropometry Working Group, NHLBI Trans-Omics for Precision Medicine Consortium. (2021) bioRxiv 588020, https://doi.org/10.1101/588020

66. Relationship between proxies of energy states and nitrogen use efficiency for Holstein cows in early lactation

Y. Chen[1*], C. Grelet[2], S. Vanderick[1], H. Hu[1] and N. Gengler[1]

[1]University of Liège, Gembloux Agro-Bio Tech (ULiège-GxABT), Passage des Déportés 2, 5030 Gembloux, Belgium; [2]Walloon Agricultural Research Center (CRA-W), Chaussée de Namur 24, 5030 Gembloux, Belgium; yansen.chen@doct.uliege.be

Abstract

The purposes of this study were to estimate the genetic parameters of the energy status indicator (C18:1 *cis*-9) of Holstein cows in early lactation and its relationship with nitrogen use efficiency (NUE) (predicted NUE and milk urea concentration). After editing, the data included 143,517 records within 5 to 50 days in milk from 52,198 cows, and 3,546 animals in the pedigree with 28,427 SNPs. Two multiple-trait repeatability models were used in this study. In early lactation, the average C18:1 *cis*-9 was gradually decreasing and was highest in May. The heritabilities of C18:1 *cis*-9 for primiparous and multiparous cows were 0.12 and 0.09, respectively. The C18:1 *cis*-9 had positive genetic correlations with predicted NUE (from 0.28 to 0.67), and weak genetic correlations with milk yield, milk urea concentration (from -0.15 to 0.14). This study suggests that breeding for NUE alone may enhance energy troubles in early lactation.

Introduction

Most dairy cows will be in a negative energy balance (NEB) in early lactation due to energy intake that cannot meet the needs of milk production (Churakov *et al.*, 2021). If the dairy cow is in NEB state for too long, the risk of the dairy cow suffering from metabolic diseases (such as ketosis) will increase (Zachut *et al.*, 2020), but it is difficult to directly detect the energy status of dairy cows. Usually, people use indirect prediction of the NEB of dairy cows as an alternative method. Although using milk mid-infrared (MIR) spectra to indirectly predict the NEB of dairy cows is a good method, the coefficient of determination of its prediction model is varying in different studies (from 0.48 to 0.77) (Ho *et al.*, 2019; Smith *et al.*, 2019).

Studies have shown that oleic acid (C18:1 *cis*-9) is a good proxy for identifying NEB in dairy cows (Churakov *et al.*, 2021, Guinguina *et al.*, 2021). When the cow is in NEB, its body fat reserves will be mobilized, leading to an increase in C18:1 *cis*-9 in milk (Zachut *et al.*, 2020). The C18:1 *cis*-9 can be accurately and quickly predicted using MIR. The determination coefficients for cross-validation of its prediction model were 0.95 (Soyeurt *et al.*, 2011), which was confirmed by the study of Grelet *et al.* (2014). The genetic parameters of C18:1 *cis*-9 have been reported but they only focus on the first parity (Bastin *et al.*, 2011; Mehtiö *et al.*, 2020). As the nitrogen use efficiency (NUE) of dairy cows has attracted attention (Chen *et al.*, 2021b) in recent years, we guess that its addition to breeding may increase the NEB of dairy cows. In this study, the new proxy (predicted NUE, PNUE) and traditional proxy (milk urea concentration, MU) of NUE were used as representatives to evaluate their relationships with C18:1 *cis*-9 (Chen *et al.*, 2021a, 2021b). At the same time, this research has added milk yield (MY) as a representative of the main economic traits.

The purposes of this study were to estimate the genetic parameters of C18:1 *cis*-9 in the early lactation of the first five parities of dairy cows and its relationship with PNUE, MU, and MY.

Materials & methods

Data. All data were collected from 2012 to 2019 during the official milk recording in the Walloon Region of Belgium. The MIR analysis of milk samples was performed by MilkoScan FT 6000 and FT+ spectrometers (Foss Electric, Hillerød, Denmark) to provide MIR spectra and milk composition. Edits excluded irregular

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_66, © Y. Chen1 *et al.* 2022

data for daily MY (<1 and >99 kg) and MU (<2 and >70 mg/dl). The filtering method for C18:1 *cis*-9 (g/dl of milk) and PNUE was the same as Chen *et al.* (2021b). Furthermore, the range of days in milk (DIM) was from 5 to 50, parity within 5, and the ranges of accepted calving age were: from 23 to 37, 35 to 52, 47 to 67, 59 to 83, 71 to 96 months for the first to fifth parity, respectively. After filtering, 134,517 records from 52,198 Holstein cows on 774 farms were used in this study. The pedigree related to the dataset comprised 120,933 animals, and 3,546 (397 bulls) of them had genotypes with 28,427 SNPs. The genotype data were provided by the cattle breeding system in the Walloon Region of Belgium. Besides, each trait was divided into primiparous and multiparous (including second to fifth parity) classes, except the MU (based on the research results of Chen *et al.*, 2021a). Hereafter, C18:1 *cis*-9, PNUE, and MY data were divided following parities and identified as C18:1 *cis*-9_1, PNUE1, MY1, C18:1*cis*-9_2+, PNUE2+, and MY2+. At the same time, based on its trend with DIM (Figure 1), C18:1 *cis*-9 was also divided into 3 traits, C18:1 *cis*-9_1 (first parity), C18:1 *cis*-9_2 (second parity), C18:1 *cis*-9_3+ (from third to fifth parity). All milk MIR spectra were standardized according to Grelet *et al.* (2015). The C18:1 *cis*-9 and PNUE of each cow were predicted by the equations based on the models established by Grelet *et al.* (2014, 2020). The determination coefficients and root mean square errors of cross-validation were 0.95 and 0.06% for C18:1 *cis*-9, and 0.68 and 5.01% for the PNUE model, respectively.

Models. The three-trait and seven-trait repeatability models were used to estimate the variance components for C18:1 *cis*-9 (C18:1 *cis*-9_1, C18:1 *cis*-9_2, and C18:1 *cis*-9_3+) and seven traits (C18:1 *cis*-9_1, PNUE1, MY1, MU, C18:1 *cis*-9_2+, PNUE2+, and MY2+). The models were the same as Chen *et al.* (2021b), expect the number of traits.

Genetic parameters. The variance components, heritability (h^2), and genetic correlations of the three-trait model were estimated by AIREMLF90 (version 1.149). The variance components of the seven-trait model were estimated by Gibbs sampling in THRGIBBS1F90 (version 2.119), and posterior convergence was analyzed by POSTGIBBSF90 (version 3.15). Among them, the posterior means of variances were obtained using 100,000 samples after a burn-in of 20,000 samples. Using the mean of the posterior variance components, the h^2, genetic, and phenotypic correlations were calculated. Approximate standard errors (SE) of parameters were obtained from POSTGIBBSF90 (version 3.15) program.

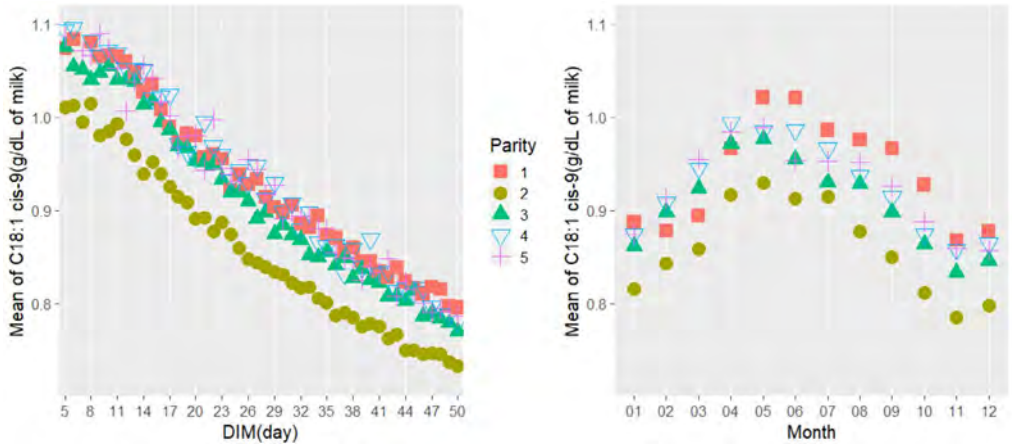


Figure 1. The curves of the mean of C18:1 *cis*-9 (g/dl of milk) with days in milk (DIM) and month for Holstein cows in the first five parties.

Results & discussion

Descriptive phenotype. The average (standard deviation, SD) C18:1 *cis*-9 in primiparous and multiparous cows was 0.92 g/dl (0.25) and 0.89 g/dl (0.24), respectively. This was in line with other studies (Bastin *et al.*, 2011; Mehtiö *et al.*, 2020; Churakov *et al.*, 2021). The trends of the average C18:1 *cis*-9 of dairy cows in early lactation were shown in Figure 1. The average C18:1 *cis*-9 gradually decreased from DIM 5 to 50, ranging from 0.73 (SD: 0.15) to 1.11 (SD: 0.28), which is consistent with the report by Churakov *et al.* (2021). Within the months of the year, the average C18:1 cis-9 ranged from 0.78 (SD: 0.20) to 1.02 (SD: 0.25), and each parity reached its maximum in May, except for the fourth parity. It may be that C18:1 *cis*-9 was very sensitive to heat stress in dairy cows (Hammami *et al.*, 2014), and cows got fresh pasture in May (Frétin *et al.*, 2019).

Heritabilities. In the seven-trait model, the h^2 (SE) of C18:1 *cis*-9 in primiparous and multiparous were 0.12 (0.01) and 0.09 (0.01) (Table 1), which were consistent with our results in the three-trait model (not shown). The h^2 of C18:1 *cis*-9 in this study was similar to the results obtained in the same area of the previous study (Bastin *et al.*, 2011), and was also consistent with the results obtained in the Nordic Red Dairy cattle study in Finland (Mehtiö *et al.*, 2020). Moreover, the h^2 of PNUE and MU were consistent with the results of our previous studies (Chen *et al.*, 2021a, 2021b).

Correlations. The genetic and phenotypic correlations of the seven traits were described in Table 1. From the three-trait model analysis, the genetic correlations of the first and second with third parity+ traits were 0.70 and 0.93, respectively. Then we changed it to two traits, namely primiparous as the first and multiparous as the second one. The genetic correlations between C18:1 *cis*-9 and PNUE were positive (from 0.28 to 0.67), which indicates that increasing NUE in the early lactation may be detrimental to its energy balance. Another reason is that in early lactation, cows both mobilize energy and protein body reserves, so NUE is artificially high. This was similar to the results of Hurley *et al.* (2018) who showed that improving feed efficiency was not conducive to the energy balance of cattle. The C18:1 *cis*-9 had weak genetic correlations with MY and MU, ranging from -0.14 to 0.14, and from -0.15 to 0.03, respectively. The phenotypic correlations between C18:1 *cis*-9 and PNUE were moderately positive in the same parity class (0.33 and 0.34). However, the C18:1 *cis*-9 had low phenotypic correlations with MY and MU, ranging from -0.04 to 0.10.

Table 1. Heritabilities[1] (diagonal, bold), genetic correlations (above the diagonal) and phenotypic correlations (below the diagonal) among the four features[2] for primiparous (n=41,035) and multiparous (n=93,482) Holstein cows in early lactation.

	C18:1 *cis*-9_1	**PNUE1**	**MY1**	**MU**	**C18:1 *cis*-9_2+**	**PNUE2+**	**MY2+**
C18:1 *cis*-9_1	**0.12**	0.67	0.14	0.03	0.67	0.36	0.00
PNUE1	0.33	**0.17**	0.71	0.12	0.28	0.70	0.50
MY1	0.10	0.47	**0.18**	0.01	-0.14	0.51	0.78
MU	0.00	0.03	0.01	**0.12**	-0.15	0.04	-0.02
C18:1 *cis*-9_2+	0.09	0.03	-0.04	0.02	**0.09**	0.48	-0.05
PNUE2+	0.07	0.15	0.13	0.03	0.34	**0.12**	0.74
MY2+	0.02	0.15	0.21	0.03	0.06	0.58	**0.13**

[1] The standard errors of all heritabilities, genetic correlations, and phenotypic correlations are 0.01, less than 0.07, and less than 0.01, respectively.

[2] The four features were divided into seven traits, namely C18:1 *cis*-9 (g/dl of milk) (C18:1 *cis*-9_1), predict nitrogen use efficiency (%) (PNUE1), daily milk yield (kg) (MY1) in primiparous cattle, and milk urea concentration (mg/dl) (MU) in the first five parities, and C18:1 *cis*-9_2+, PNUE2+, MY2+, corresponding to multiparous cattle.

Conclusions

The C18:1 *cis*-9 had low h^2 in the early lactation of the first five parities for Holstein cows. In addition, the C18:1 *cis*-9 had positive genetic correlations with PNUE (unfavorable), and low genetic correlations with MY and MU (favorable). This study suggests that breeding for NUE alone may enhance energy troubles in early lactation.

Acknowledgements

The China Scholarship Council (Beijing) is acknowledged for funding the PhD project of Yansen Chen. The authors would like to thank the Walloon Region and European Union's INTERREG NWE program for their financial support of the OptiMIR project, and the European Union's Seventh Framework Program for funding the GplusE project under grant agreement no. 613689.

References

Bastin C., Gengler N., and Soyeurt H. (2011) J. Dairy Sci. 94(8):4152-4163. https://doi.org/10.3168/jds.2010-4108

Chen Y., Atashi H., Vanderick S., Mota R.R., Soyeurt H., *et al.* (2021a) J. Dairy Sci. 104(12): 12741-12755. https://doi.org/10.3168/jds.2021-20659

Chen Y., Vanderick S., Mota R.R., Grelet C., GplusE Consortium*, *et al.* (2021b) J. Dairy Sci. 104(4): 4413-4423. https://doi.org/10.3168/jds.2020-18849

Churakov M., Karlsson J., Edvardsson Rasmussen A., and Holtenius K. (2021) Animal 15(7): 100253. https://doi.org/10.1016/j.animal.2021.100253

Frétin M., Martin, B., Buchin, S., Desserre, B., Lavigne, R., *et al.* (2019 J. Dairy Sci. 102(2): 1131–1143. https://doi.org/10.3168/jds.2018-15534

Grelet C., Fernández Pierna J.A., Dardenne P., Baeten V., and Dehareng F. (2015) J. Dairy Sci. 98(4): 2150-2160. https://doi.org/10.3168/jds.2014-8764

Grelet C., Fernández Pierna J.A., Soyeurt H., Dehareng F., Gengler N., *et al.* (2014) Proc. of the 65th EAAP, Copenhagen, Denmark.

Grelet C., Froidmont E., Foldager L., Salavati M., Hostens M., *et al.* (2020) J. Dairy Sci. 103(5): 4435-4445. https://doi.org/10.3168/jds.2019-17910

Guinguina A., Yan T., Trevisi E., and Huhtanen P. (2021) J. Dairy Sci. 104(6): 6701-6714. https://doi.org/10.3168/jds.2020-19591

Hammami H., Vandenplas J., Vanrobays M.L., Rekik B., Bastin C., *et al.* (2015) J. Dairy Sci. 98(7): 4956-4968. https://doi.org/10.3168/jds.2014-9148

Ho P.N., Marett L.C., Wales W.J., Axford M., Oakes E.M., *et al.* (2019) Anim. Prod. Sci. 60(1): 164-168. https://doi.org/10.1071/AN18532

Hurley A.M., Lopez-Villalobos N., McParland S., Lewis E., Kennedy E., *et al.* (2018) J. Dairy Sci. 101(2): 1267-1280. https://doi.org/10.3168/jds.2017-12841

Mehtiö T., Mäntysaari P., Negussie E., Leino A.M., Pösö J., *et al.* (2020) Animal 14(8): 1588-1597. https://doi.org/10.1017/S1751731120000439

Smith S.L., Denholm S.J., Coffey M.P., Wall E. (2019) J. Dairy Sci. 102(12): 11169-11179. https://doi.org/10.3168/jds.2018-16112

Soyeurt H., Dehareng F., Gengler N., McParland S., Wall E., *et al.* (2011) J. Dairy Sci. 94(4): 1657-1667. https://doi.org/10.3168/jds.2010-3408

Zachut M., Šperanda M., De Almeida A.M., Gabai G., Mobasheri A., *et al.* (2020) J. Dairy Res. 87(1): 4-13. https://doi.org/10.1017/S0022029920000084

67. Genetic variation exists in nitrogen use efficiency for lactating cows in grass-grazed dairy systems

E. Tavernier[1,2*], I.C. Gormley[2], L. Delaby[3], S. McParland[1], M. O'Donovan[1] and D.P. Berry[1]

[1]Teagasc, Animal & Grassland Research and Innovation Centre, Moorepark, Fermoy, Co. Cork, Ireland; [2]School of Mathematics and Statistics, University College Dublin, Ireland; [3]INRAE, UMR Pegase, 16 Le Clos 35590 Saint-Gilles, France; elia.dufosse@teagasc.ie

Abstract

Improving nitrogen efficiency in dairy cattle is of interest to reduce the environmental burden of dairy production. One strategy is to breed cows for improved nitrogen use efficiency and less nitrogen excretion. Two nitrogen efficiency statistics were computed: nitrogen use efficiency (NUE), which was defined as the ratio of the nitrogen in the saleable output relative to nitrogen intake, and nitrogen balance (Nbalance) defined as the difference between the nitrogen intake and the nitrogen output in saleable product. Using a dataset of 2,262 lactations from grazing dairy cows, considerable genetic variability was detected with heritability estimates for NUE and Nbalance of 0.06 (0.02) and 0.13 (0.03), respectively. Nbalance was both strongly phenotypically correlated (r_p=0.98 (0.001)) and strongly genetically correlated (r_g=0.98 (0.01)) with nitrogen intake.

Introduction

Nitrogen excess in grazing dairy systems generally originates either from fertiliser or animal excreta and contributes to eutrophication as well as nitrous oxide pollution. Excreted nitrogen, termed nitrogen loss, can be an economic cost for farms. Improving nitrogen use efficiency (NUE) of the animal can therefore potentially reduce nitrogen excretion, representing both an environmental and economic saving. Nitrogen use efficiency can be defined as a function of nitrogen input and output; nitrogen input is essentially nitrogen intake. Nitrogen output includes both the crude protein (CP) and urea in the milk, as well as nitrogen used for growth and body tissue accretion, along with the nitrogen in the excreted urine and faeces. Milk protein and milk urea nitrogen (MUN) phenotypes are routinely available from milk samples and have been demonstrated to be heritable (Lopez-Villalobos *et al.*, 2018), with MUN having being proposed as an indicator of NUE or nitrogen excretion. Feed intake (and by extension nitrogen intake) is also known to be heritable in dairy cows. No routine measures of nitrogen content of the faeces and urine exist but, within a balanced system, the nitrogen excreted should be a function of what was available less that used in other bodily functions (ignoring differences in the efficiency of nitrogen use for the different bodily functions). The objective of the present study was to quantify the genetic variation in NUE and nitrogen excreted in lactating dairy cows and to estimate its correlation with performance traits.

Materials & methods

Data. A total of 4,979 milk, live-weight and intake measures from 1,324 grazing Irish dairy cows of 4 different breeds (i.e. Holstein-Friesian, Jersey, Norwegian Red, and crosses) were available from experiments conducted between 2008 and 2018 on four Teagasc experimental farms. All the cows were milked twice daily. The chemical composition (protein, fat, casein, lactose, and urea concentration) was computed from mid-infrared analysis of consecutive evening and morning milk samples once weekly. Cow live-weight was recorded every 1 to 3 weeks. Individual grass dry matter intake (GDMI) was recorded using the n-alkane technique as described by Dillon and Stakelum (1989) up to four times per lactation. The measure of GDMI was an average of a 6-d measurement period. Concentrate dry matter intake (CDMI) was also recorded during this measurement period. Grass and concentrate crude protein content were available.

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_67, © E. Tavernier *et al.* 2022

Nitrogen use efficiency. Total nitrogen intake (Nintake) was computed as the sum of (GDMI × CP content of grass) plus (CDMI × CP content of concentrate) all divided by 6.25. The nitrogen output in the milk (Nmilk) was computed as the sum of (milk yield × CP content)/6.38 and (milk yield × MUN content); the weekly mean Nmilk coinciding with the week of GDMI measurement was then calculated. The nitrogen fixed in muscles (Nmuscles) was estimated from fitting a mixed model to individual live-weight records of all cows separately from 20 days after first calving until either culling, the date of sale or date of data extraction; a random regression mixed model with quadratic fixed effects and linear random effect for cow was used. The nitrogen fixed in the conceptus (Nconceptus) was computed using a formula of INRA (2018). The nitrogen mobilised or put into reserve was computed as (33/6.25) × the energy balance (according the French Net Energy system) (INRA, 2018), if it was positive, it was considered as nitrogen put in reserve (Nreserve) and if it was negative, its absolute value was consider as nitrogen mobilised from the reserve (Nmob). The nitrogen fixed in product (Noutput) was defined as the sum of Nmilk, Nmuscles, Nconceptus and Nreserve. The nitrogen available (Navailable) was the sum of Nintake and Nmob. Nitrogen balance (Nbalance) was subsequently calculated as Navailable less Noutput. The NUE was defined as the ratio of Noutput to Navailable. (Co)variance components for all traits were estimated using a series of univariate and bivariate animal models in ASREML (version 4.0), with contemporary group (concatenation of farm, treatment and date of milking), the interaction of stage of lactation (9 classes) with parity, heterosis and recombination loss all considered as fixed effects; a direct additive genetic effect along with a within- and across-parity permanent environmental effect of the cow were all included as random effects with a pedigree relationship matrix used to account for relationships among cows.

Results

Cows, on average, ingested 551 grams of nitrogen per day changing across lactation (Figure 1) using, on average, 23.2% of the Nintake for milk, growth and gestation; cows, therefore, on average, excreted 429 g of nitrogen per day in urine and faeces combined (Table 1). The heritability of the Nbalance was 0.13 (0.03) which is less than for the milk traits while the heritability of NUE was 0.06 (0.02) (Table 1). There was a strong genetic correlation between Nbalance and Nintake (0.98), a moderate to strong genetic correlation of 0.71 between milk yield and Nintake and a correlation of 0.64 between milk yield and Nbalance (Table 2). No genetic correlation existed between MUN and either of the two computed nitrogen efficiency metrics (Table 2).

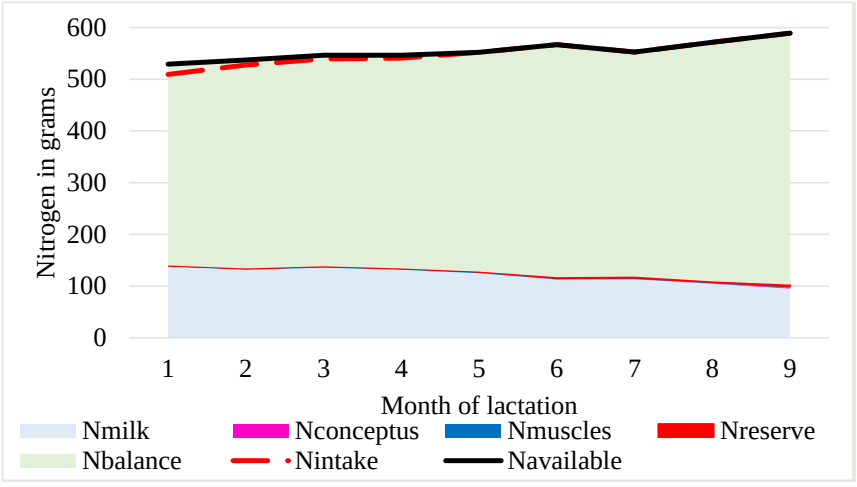


Figure 1. Nitrogen intake and partitioning by month of lactation.

Table 1. Mean, genetic standard deviation (SD), heritability and repeatability with standard errors (SE) of the traits studied.

Trait	Mean	Genetic SD	Heritability (SE)	Repeatability (SE)
Milk yield (kg)	20.2	1.43	0.23(0.04)	0.52 (0.02)
Protein content (%)	3.67	0.16	0.49 (0.05)	0.62 (0.02)
Fat content (%)	4.44	0.39	0.43(0.05)	0.50 (0.02)
MUN (mg/dl)	32.7	2.10	0.20 (0.03)	0.28 (0.02)
Nintake (g)	551	24.05	0.16 (0.04)	0.43 (0.02)
Nbalance (g)	429	17.75	0.13 (0.03)	0.38 (0.02)
NUE (%)	23.2	0.06	0.06 (0.02)	0.25 (0.02)

Table 2. Phenotypic (below) and genetic correlation (above) of the traits studied with their standard errors in parenthesis.

	Milk yield (kg)	Protein content (%)	Fat content (%)	MUN (mg/dl)	Nintake (kg)	Nbalance (kg)	NUE (%)
Milk yield (kg)		-0.58 (0.09)	-0.70 (0.08)	0.10 (0.14)	0.71 (0.10)	0.64 (0.12)	0.49 (0.15)
Protein content (%)	-0.33		0.68 (0.06)	0.10 (0.10)	0.003 (0.12)	-0.03 (0.13)	0.13 (0.18)
Fat content (%)	-0.37	0.47		0.24 (0.11)	-0.08 (0.13)	0.01 (0.13)	-0.45 (0.14)
MUN (mg/dl)	0.03	0.01	0.07		0.24 (0.14)	0.27 (0.15)	-0.07 (0.20)
Nintake (kg)	0.40	0.04	-0.04	0.07		0.98 (0.01)	0.20 (0.23)
Nbalance (kg)	0.23	0.02	0.04	0.06	0.98		-0.01 (0.24)
NUE (%)	0.63	0.05	-0.25	0.04	-0.24	-0.42	

[1] SE range of phenotypic correlation is 0.001-0.02.

Discussion

Almost half of Irish rivers had an unsatisfactory nitrate concentration in 2020 (EPA, 2020). Cows potentially contribute to this since over 70% of the nitrogen ingested by cows is excreted in urine or faeces (Ryan *et al.*, 2011); the mean NUE from the data in the present study was 23.2% so 76.8% of Nintake was excreted by the cows on this study. The relatively high heritability of Nbalance (0.13) means it is more amenable to achieving a higher accuracy of selection thus further improving the potential for genetic gain.

Based on a mean daily Nbalance of 429 g in the present study, a cow is expected to excrete, on average, 131 kg of nitrogen per lactation (i.e. 0.429 N kg/day × 305 day/lactation). Moreover, using the estimated genetic SD of 17.75 g for Nbalance, cows in the top 10% for genetic merit for Nbalance are expected to excrete 9.5 kg less nitrogen per lactation (i.e. 305 days).

No estimates of genetic parameters exist in the scientific literature for either NUE or Nbalance as defined in the present study. Nonetheless, crude protein efficiency, defined as crude protein output in milk divided by the crude protein intake, is commonly used as a nitrogen efficiency indicator. Using a population of 970 American Holstein dairy cows fed in confinement, Vallimont *et al.* (2011) computed a heritability of 0.21(0.06) for this trait while Lopez-Villalobos *et al.* (2018) reported a low heritability of 0.11 (0.08) using 468 New Zealand mixed-breed grazing dairy cows. The present study and the study of Lopez-Villalobos *et al.* (2018) both had low heritability estimates for their nitrogen efficiency approximation, NUE and crude protein efficiency, respectively, which may be linked to the fact that cows were grazing in both studies.

Grazing cow Nintake may be more variable as the grass intake and composition vary during the year, increasing the part of variation explain by the environment, and therefore deflating the heritability.

MUN is commonly used as an indicator of nitrogen efficiency for dairy cows fed in confinement. In the present study, no genetic correlation existed between either Nbalance or NUE with MUN. A positive relationship between MUN and the proportion of nitrogen intake that ended up in milk is often reported but, according to Aizimu *et al.* (2021), this correlation weakens when MUN in high. As grazing cows tend to have a higher MUN than confinement cows, owing to the imbalance between energy and protein intake, and Nmilk is the main part of the Noutput (97% of the Noutput), MUN seemed less sensitive to change of NUE. Thus, MUN does not seem to be a suitable trait to select as a means of trying to reduce the nitrogen excreted by grazing dairy cows.

Conclusions

The results of the present study clearly demonstrate the presence of exploitable genetic variation in several measures of nitrogen intake and efficiency metrics; Nintake seems a good trait to select to reduce Nbalance but Nbalance and Nintake were nonetheless genetically correlated with milk yield implying that genetic selection to reduce nitrogen intake would impact genetic gain for milk yield.

References

Aizimu W., Al-Marashdeh O., Hodge S., Dewhurst R.J., et al (2021) Dairy 2(3):435–451. https://doi.org/10.3390/dairy2030035

Dillon P., and Stakelum G. (1989) Irish J. Agric. Res. 8:104.

Environmental Protection Agency (EPA), Water Quality in 2020: An Indicators Report. Available at: https://www.epa.ie/publications/monitoring--assessment/freshwater--marine/water-quality-in-2020.php

INRA (2018). Alimentation des ruminants: Apports nutritionnels – Besoins et réponses des animaux – Rationnement – Tables des valeurs des aliments. Editions Quae, Versailles, France.

Lopez-Villalobos N., Correa Luna M., Burke J.L., Sneddon N.W., et al (2018) N. Z. J. Anim. Sci. Prod. 78:56–61.

Ryan W., Hennessy D., Murphy J.J., Boland T.M., and Shalloo L. (2011) J. of Dairy Sci. 94(2):1032–1044. https://doi.org/10.3168/jds.2010-3294

Vallimont J.E., Dechow C.D., Daubert J.M., Dekleva M.W. et al (2011) J. Dairy Sci. 94(4):2108–2113. https://doi.org/10.3168/jds.2010-3888

68. Study of the causal relationship between the rabbit cecal microbiota and host's growth and feed efficiency

M. Piles*, M. Mora, J.P. Sánchez, Y. Ramayo-Caldas, O. González-Rodríguez and M. Velasco-Galilea

Institute of Agrifood Research and Technology (IRTA), Animal Breeding and Genetics, Torre Marimon C-59, km 12,1, 08140 Caldes de Montbui, Barcelona, Spain; miriam.piles@irta.cat

Abstract

We proposed a structural equation model to provide a better understanding of how genetic effects directly affect average daily gain on *ad libitum* (ADG_{AL}) and restricted (ADG_R) feeding regimens and how these effects affect those same traits through the microbiota (indirect effects). 138 OTUs (operational taxonomic unit) with structural coefficients statistically different from zero were identified, 104 and 34 had a positive and negative effect, respectively. These OTUs belonged, mostly, to the family *Lachnospiraceae, Ruminococcaeae* and *Clostridiales*. The highest heritability was found for New.Reference369 that belongs to genus *Desulfovibrio* (0.234 (0.091)). This OTU was identified as an important OTU with total negative effect on ADG_R. Members from orders *Bacteroidales* and *Clostridiales* and family *Lachnospiraceae* had a positive and relevant indirect effect on ADG. The contribution of indirect effects to total covariance was greater than 25% for 28 and 17% of the OTUs with a significant effect on ADG_{AL} and ADG_R, respectively.

Introduction

In several studies, the effect of the gut microbiome (M) on phenotypic performance has been studied. The term effect carries a causal meaning but in the case of analysis of the impact of the microbiome on a phenotype, it can be confounded because of the genetic association between the microbiome and the phenotype since both, the relative abundance of microbiota across individuals and the phenotype, can be attributed to host genetic effects. The importance of host genetic effects on growth and feed efficiency (FE) that is exerted through microbiome effects (indirect genetic effects) relative to direct genetic effects is still unknown. Direct and indirect effects could be opposite in sign leading to a null effect of the genotype on the trait if they are of similar magnitude. Structural equation models (SEM) can help to disentangle the complex interplay between the host genome and the microbiome as it allows the decomposition of the total effect on a trait into direct and indirect contributions.

Average daily gain on a restricted feeding (ADG_R) regimen can be considered as a FE trait and used as selection criteria since variation in this trait is directly related to variation in FE because of constant feed intake (FI) (Nguyen *et al.*, 2005; Drouilhet *et al.*, 2016). Both average daily gain on *ad libitum* feeding (ADG_{AL}) and ADG_R are heritable implying that host genetics play an important role in regulating feed utilization and growth (Piles and Sánchez, 2019). On the other hand, it is well known that gut microbiota influences host metabolism (Maltecca *et al.*, 2019) and genome-wide association studies (GWAS) have identified several host genetic variations that affect the gut microbiome composition (Psifidi *et al.*, 2018) and several traits related with growth and feed efficiency (Sánchez *et al.*, 2020). However, it remains unknown whether host genetics affect feed utilization through their ability to promote a favourable and stable microbial community in the gut or whether microbiome and host genetics interact to affect FE and growth.

The objectives of this research are: (1) To identify microbial features with a relevant impact on growth and feed efficiency; and (2) To quantify the relative importance of direct and microbial-mediated genetic effects on total variance and covariances of the same traits.

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_68, © M. Piles *et al.* 2022

Materials & methods

Animals and data. Animals came from the Caldes rabbit sire line. At weaning, the animals were randomly assigned to one of the two different feeding regimes: *ad libitum* (AL) or restricted (R) to 75% of the AL feed intake. The individual body weight (BW) was weekly recorded for all animals in both feeding regimes. From BW raw records, individual average daily gain (ADG) was computed as the slope of the within animal regression of all BW measurements on their respective ages in days. This trait was individually computed for each feeding regime, thus obtaining ADG on AL (ADG_{AL}) or R (ADG_{R}). Cecal samples of 425 rabbits of the two feeding regimes were collected. Details about the process of sample processing, DNA extraction and sequencing processes are in Velasco-Galilea *et al.* (2020). After filtering, the OTU (operational taxonomic unit) table contained the sequence counts of 951 OTUs. The analysis was conducted by normalizing the OTU table with the cumulative sum scaling (CSS) method.

Structural equation model. A Bayesian analysis was performed for each combination of both phenotypes (ADG_{AL} and ADG_{R}) and one microbial feature (M, expressed as the relative abundance of a taxon). The SEM for ADG_{AL}, ADG_{R} and the m^{th} microbial feature can be written in the following way:

$$y=\Lambda y+Xb+Z_1l+Z_2c+Z_3u+e \quad (1)$$

Where y is the vector containing the observations for traits ADG_{AL}, ADG_{R} and M (with ADG_{R} as missing value for animals on AL and ADG_{AL} as missing value for animals on R), Λ is a (3×3) matrix with zeros in the diagonal and with structural coefficients (λ) in the off diagonals. Two structural coefficients were considered in this study, $\lambda_{AL\leftarrow M}$ and $\lambda_{R\leftarrow M}$, representing the effect of M on ADG_{AL} and ADG_{R}, respectively; b is the vector of systematic effects with incidence matrix X; this vector included the effects of the factors: the combination of batch and farm (5 levels) and the body weight at weaning (2 levels: big and small); l and c are the vectors of permanent environmental effects of litter (195 levels) and cage (189 levels), respectively; u is the vector of genotypic effects (412 levels) and e the vector of model residuals.

Results

After implementing the SEM, 138 OTUs with structural coefficients statistically different from zero were identified, where 82 OTUs had a statistically significant effect on ADG_{AL} and 80 on ADG_{R}. Of those 138 OTUs, 104 had a positive effect on ADG, where only 14 matched ADG_{AL} and ADG_{R}. Conversely, 34 OTUs had a negative effect on ADG, where 10 matched between the two feeding regimes. However, according to the employed model, neither of the common OTUs had an important effect on the phenotype. The most important OTUs were different for the two feeding regimes, being 860192, a member of genus *Coprobacillus*, the OTU with the most relevant effect on ADG_{AL} (-1.91 g/Normalized OTU count) and New.ReferenceOTU369, taxonomically classified as *Desulfovibrio*, the one with the largest effect on ADG_{R} (-1.93 g/Normalized OTU count), both having a negative effect on ADG. The OTUs with the highest positive effect on this trait belonged to the order *Bacteroidales* (857827) (1.87 g/Normalized OTU count) for ADG_{AL} and to the family *Ruminococcaceae* (New.ReferenceOTU1337) (1.86 g/Normalized OTU count) for ADG_{R}. These structural coefficients indicate the expected change in ADG_{AL} or ADG_{R} when the relative abundance of M increases by 1 unit (in our case in g/ Normalized OTU count).

In general, the estimates of heritabilities for the most important OTUs that affect ADG_{AL} and ADG_{R} were low to medium. Regarding the OTUs that had a meaningful effect on ADG_{AL}, the posterior mean (posterior sd) of heritability ranged from 0.08 (0.05) for New.ReferenceOTU1122 to 0.171 (0.07) for 857827, members of genus *Ruminococcus* and order Bacteroidales, respectively. On the other hand, for OTUs with an important effect on ADG_{R} this parameter ranged from 0.05 (0.03) for a member of genus *Phascolarctobacterium* (New.ReferenceOTU570) to 0.23 (0.09) for New.ReferenceOTU369, which belongs to genus *Desulfovibrio*.

In SEM, variance components correspond to direct effects on ADG_{AL} or ADG_{R}, whereas in MTM they correspond to total effects, i.e. direct plus indirect effects mediated by the microbiota. In general, distributions of total genetic variances and direct genetic variance were slightly different because of the small contributions of the indirect effects. However, for members of order *Clostridiales*, and families *Ruminococcaceae* (New.ReferenceOTU1337) and *Lachnospiraceae* (New.ReferenceOTU381) (Figure 1, panel A) the total indirect effects had an important negative effect on the total genetic variance of ADG_{R} despite the positive value of $\lambda_{R \leftarrow M}$ because of the negative covariance between direct additive genetic effects. The contribution of total indirect effects to total genetic variance was small (less than 10%) for all OTUs.

Regarding associations between ADG_{AL} and ADG_{R} with M, total and direct additive genetic covariance were very similar for members of order *Clostridiales*, OTUs 849147 and New.ReferenceOTU1863, due to the almost null contribution of the indirect effects. Indirect effects of OTU 857827 from order *Bacteroidales* and OTU 825939 taxonomic assigned to order *Clostridiales* had an important positive effect on the covariance with ADG_{AL}, while indirect effects of a member of genus *Desulfovibrio* (New.Reference369) (Figure 1, panel B) and family *Lachnospiraceae* (New.ReferenceOTU381) had an important negative and positive effect on the covariance with ADG_{R}, respectively. The contribution of indirect effects to total genetic covariance was greater than 25% for 28% of the OTUs with a significant effect on ADG_{AL}, and for 17% of the OTUs with a significant effect on ADG_{R}.

Discussion

Feed efficiency and the influence that gut microbiota has on it have been widely studied in the last years in different species like rabbits (Drouilhet *et al.*, 2016), chickens (Wen *et al.*, 2021) and pigs (Bergamaschi *et al.*, 2020). Despite this, the importance of host genetics effects on feed efficiency exerted through microbiome effects is largely unknown. This study provides a better understanding of how genetic effects directly affect ADG_{AL} and ADG_{R} (direct effects) and how these effects affect ADG_{AL} and ADG_{R} through the microbiota (indirect effects). To assess this, a SEM was implemented to determine the causal relationship between the microbiome and the two phenotypes. Knowledge of these effects could help in the design of sustainable approaches to improve growth and feed efficiency in livestock.

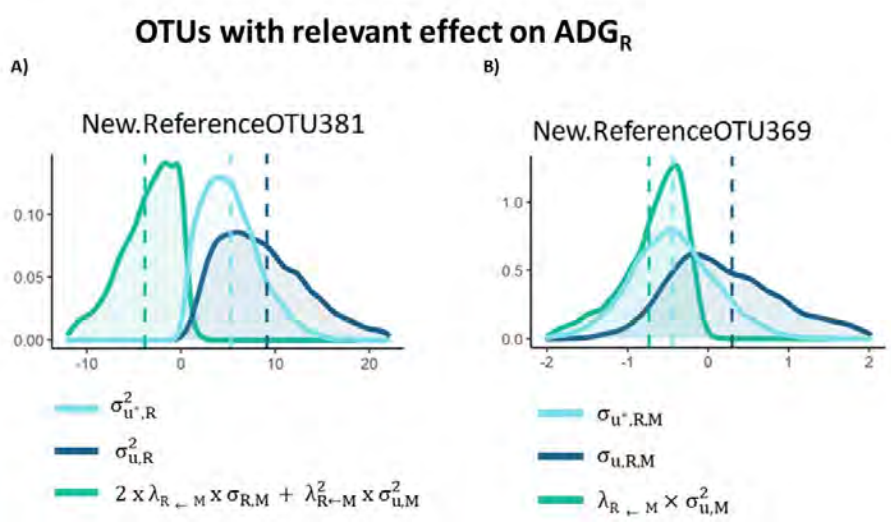


Figure 1. Components of total (light blue) genetic variance (panel A) and covariance (panel B) for ADG_{R}: direct effects (dark blue) and indirect effect (green).

OTUs with an important effect (i.e. structural coefficient) seems to be different for growth and for efficiency. Regarding the contribution of direct and microbial-mediated genetic effects to total genetic variance, only the indirect effects of the family *Ruminococcaceae* (New.Reference1337) and the family *Lachnospiraceae* (New.Reference381, Figure 1 panel A) have a relevant contribution to the total genetic variance of ADG_R. These OTUs had a positive effect on ADG_R but the covariance between M and ADG_R was negative making the contribution of total indirect effects negative.

Concerning the contribution of indirect effects to total genetic covariance between ADG_{AL} and ADG_R with the microbiota, the member from the genus *Desulfovibrio* (New.ReferenceOTU369, Figure 1 panel B) had a negative meaningful effect on ADG_R and the contribution of indirect effects to total covariance is negative and opposite to covariance due to direct effects. This result is in agreement with previous findings in pigs (Bergamaschi *et al.*, 2020; Aliakbari *et al.*, 2021) where members of genus *Desulfovibrio* were identified as being negatively correlated with FE. Members from order *Bacteroidales*, order *Clostridiales* and family *Lachnospiraceae* had a positive and relevant contribution of indirect effect to total covariance with growth and they have been identified in previous studies as members associated with FE in different species (Liu *et al.*, 2021; Velasco-Galilea *et al.*, 2021). For the first time, in this study, the effects that the host genetics exerts on growth and FE (direct effects) and the effects that the host genetics exerts on those traits through the gut microbiota (indirect effects) were evaluated. Knowledge of how direct and indirect effects act on the phenotype is relevant to fully describe the processes of mediation and understand the response in the phenotype if the microbial composition changed.

References

Aliakbari A., Zemb O., Billon Y., Barilly C., Ahn I., *et al.* (2021) J. Anim. Breed. Genet. 138(4):491-507. https://doi.org/10.1111/jbg.12539

Bergamaschi M., Tiezzi F., Howard J., Huang Y.J., Gray K.A, *et al.* (2020) Microbiome 8(1):110. https://doi.org/10.1186/s40168-020-00888-9

Drouilhet L., Achard C.S., Zemb O., Molette C., Gidenne T., *et al.* (2016) J Anim Sci 94(1):38-48. https://doi.org/10.2527/jas.2015-9402

Liu J., Stewart S.N., Robinson K., Yang Q., Lyu W., *et al.* (2021) J Animal Sci Biotechnol 12:22. https://doi.org/10.1186/s40104-020-00542-2

Maltecca C., Lu D., Schillebeeckx C., McNulty N.P., Schwab C., *et al.* (2019) Sci Rep 9(1):6574. https://doi.org/10.1038/s41598-019-43031-x

Nguyen N.H., McPhee C.P., and Wade C.M. (2005) J. Anim. Breed. Genet. 122(4):264-270. https://doi.org/10.1111/j.1439-0388.2005.00531.x

Piles M., and Sánchez J.P. (2019) J. Anim. Breed. Genet. 136(6):474-483. https://doi.org/10.1111/jbg.12395

Psifidi, A., M., Crotta R., Pandit B., Fosso P., Koringa G., *et al.* (2018) Proc. WCALP, 11 (584)

Sánchez J.P., Legarra A., Velasco-Galilea M., Piles M., Sánchez A., *et al.* (2020) Anim. Genet. https://doi.org/10.1111/age.12988

Velasco-Galilea M., Guivernau M., Piles M., Viñas M., Rafel O., *et al.* (2020) Anim. Microb. 2(1):40. https://doi.org/10.1186/s42523-020-00059-z

Velasco-Galilea M., Piles M., Ramayo-Caldas Y., and Sánchez J.P. (2021) Sci Rep 11(1):19495. https://doi.org/10.1038/s41598-021-99028-y

Wen C., Yan W., Mai C., Duan Z., Zheng J., *et al.* (2021) Microbiome 9(1):126. https://doi.org/10.1186/s40168-021-01040-x

69. Part of the functional rumen core microbiome is influenced by the bovine host genome and associated with feed efficiency

M. Martínez-Álvaro[1*], J. Mattock[2], Z. Weng[3], R.J. Dewhurst[1], M.A. Cleveland[3], M. Watson[2] and R. Roehe[1]

[1]Scotland's Rural College, Edinburgh, United Kingdom; [2]The Roslin Institute and the Royal (Dick) School of Veterinary Studies, University of Edinburgh, United Kingdom; [3]Genus plc, DeForest, WI, USA; marina.alvaro@sruc.ac.uk

Abstract

Our research focuses on exploring the possibility of improving feed efficiency in beef through genomic selection of the functional rumen microbiome. Using an environmentally standardized commercial population of 383 crossbred British Blue × Holstein cattle, we demonstrate that the host genome influences microbiome function. A 23% of the 3,594 additive log-ratio transformed microbial gene abundances (*alr*-MGs) identified from whole metagenome sequencing showed significant host genomic effects (heritabilities ranging from 0.15 to 0.66). We also found that 118 of these *alr*-MGs were host-genomically correlated (|0.47| to |0.83|) with feed efficiency. This highlights the strength of correlated host genomic effects of specific microbial processes and feed efficiency and suggests the possibility of improving feed efficiency through a microbiome-driven breeding strategy.

Introduction

Feeding accounts for ~70% of the expenses in commercial beef companies. Any effort to reduce the ratio of feed intake to growth rate will contribute to the economic profit of the industry, but will also help to mitigate the current environmental challenge facing food production systems. Microbial metabolism in the rumen plays a key role in cattle growth performance, as they derive ~70% of their energy and >50% of their aminoacid requirements from microbial fermentation; many microbiome-derived metabolites also act as regulatory signals to the host. The influence of host genetics on the rumen microbiome is now widely recognised, and can likely be explained by genetically influenced host features that favour specific microbial colonisation and their functions in the gut. The main hypothesis of this study is that a substantial part of the host's polygenic control over its feed efficiency is mediated by host-genomic influence on the metabolism of its microbiome, and therefore subjected to genomic selection. In our previous study using an experimental population (Martínez-Álvaro *et al.* 2021), we found that ~18% of the functional core microbiome in the rumen was influenced by the host genome and termed it as host-genomically influenced functional core microbiome (HGFC). In this study, we aim to validate these results by examining the HGFC of a new commercial population of 383 cattle. We also aim to investigate host-genomic correlations between HGFC and feed efficiency, which are critical to obtain animals with more efficient feed fermentation and lower environmental impact by microbiome-driven breeding.

Materials & methods

Animals. Data are from 383 British Blue and Holstein crossbred cattle (230 females and 153 males) balanced along 16 batches reared in Warwickshire (UK) during 2018-2020. At 233±30 days of age, cattle were assessed for feed intake during 35 days using Growsafe (GrowSafe Systems Ltd., Airdrie, Canada) and Integrado (Betim, Brazil) electronic feeding equipment, after a 7-10 day adaptation period. All animals were fed a concentrate-based diet (52.3 g dry matter/kg) and weighed at the beginning (average weight=283±52 kg) and at the end (345±51 kg) of the feeding trial. Their growth rate was calculated as the slope between the weight gain and the days elapsed during the feed intake assessment. Feed efficiency (FCR) was then calculated as the ratio between growth rate (kg/day) and average daily dry matter intake (kg/day).

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_69, © M. Martínez-Álvaro *et al.* 2022

Genomic and metagenomic data. For host DNA analysis, an ear tissue sample was collected using an Allflex Livestock Intelligence Tissue Sampling Tag (Durham, UK) and stored at -80 °C. Samples were processed and genotyped using the GGP Bovine uLD 35K array (Neogene, Cambridgeshire, UK) and then imputed up to 70K using Eagle (v2.4.1) and Impute (v2). All animals had a call rate>90%. SNPs were filtered (PLINK version 1.09b) for call rates<95%, deviation from Hardy-Weinberg proportions (χ^2 test $P<10^{-8}$), and minor allele frequency<0.05; and 47,206 SNPs remained. For microbial DNA analysis, 5 ml of rumen fluid was collected through an esophageal tube at the end of the feeding trial (at 279±30 days of age), mixed with 10 ml of PBS containing glycerol (87%), and stored at -80 °C. DNA extraction from the rumen samples was performed by bead beating with column filtration, and DNA quality was assessed using Nanodrop ND-1000. DNA Illumina TruSeq libraries were prepared from genomic DNA and sequenced on NovaSeq by Edinburgh Genomics (Edinburgh, UK). Paired-end reads (2 × 150 bp) were generated, resulting in between 15.7 and 58.9 Gigabases per sample. The abundance of 6,963 functional microbial genes (MGs) was determined by comparison with the Kyoto Encyclopedia of Genes and Genomes database (2020-10-04) using KofamScan. To exclude non-core microbiome functions, we used only MGs that were present in >70% of the samples (3,595 MGs). Zero abundances were imputed based on a Bayesian-multiplicative replacement using zCompositions R package. Abundances were then additive log-ratio transformed (3,594 *alr*-MGs) using the abundance of MG NAD+ kinase (*ppnK*) as a reference, selected using the criteria described in Greenacre *et al.* (2021).

Estimation of genomic parameters. Genomic heritabilities (h^2) of FCR and 3,594 *alr*-MG were estimated by fitting 3,595 univariate GBLUP animal models including batch and sex as fixed effects, age as a covariate (for FCR only), and a host genomic effect, assumed to be normally distributed with mean equal to 0 and variance equal to the genomic relationship matrix (VanRaden method 2) multiplied by the genomic variance. Bayesian statistics were used assuming bounded flat priors for all unknowns. Results were based on Markov chain Monte Carlo chains with 1,000,000 iterations (burn-in=200,000, lag=100). The significance of host genomic effects on each *alr*-MG was tested by running models without the host-genomic effect and computing the Deviance Information Criterion difference (DIC_{diff}) between the models, and the Bayes Factor (BF) as the ratio between the means of their posterior likelihood distribution. We accounted for the false discovery rate by setting a BF multitest adjusted threshold to reject the null hypothesis>17.02, as described in Wen *et al.* (2017) assuming an alpha value of 0.01. A joint condition of BF >17.2 and $DIC_{diff} \leq -10$ was required to consider the genomic model more plausible that the model without host-genomic effect. Genomic correlations between FCR and *alr*-MGs (rg_{FCR}) with significant host-genomic effects (n=841) were calculated by fitting 841 bivariate GBLUP animal models, with the same effects as for univariate models and the same assumptions, extended for two traits. To examine their confidence level, we estimated the posterior probability that rg_{FCR} was positive or negative and considered it significant when ≥90%.

Results

In our commercial population, 24% of the phenotypic FCR variation among cattle was explained by host genomic variation (s.e. 13%), and the host-genomic effects on FCR were highly evident ($BF=1.19\times10^{+3}$ and DIC_{diff} = -27.3). Host-genomic effects were significant in ~23% of the functional core microbiome (841/3,594 *alr*-MGs tested), which cumulatively accounts for 6.6% of the total relative abundance in the rumen. The h^2 estimates of this HGFC microbiome, which ranged from 0.15±0.09 to 0.66±0.18 (Figure 1a), indicate that the rumen microbiome is exploitable by genomic selection. These *alr*-MGs are involved in a variety of metabolic functions that could play a key role modulating the nutrient flux available for host growth: e.g. sulfur metabolism (h^2 of *cysU, cysD,* and *cysN* ranged from 0.37±0.16 to 0.46±0.16), synthesis of cysteine, methionine, and arginine (*speE, metN, metC, urea, ureC*) or production of propionate and butyrate (*ACAT*, *alsD*, *cat2*, *pduP*, *lcdA*, *prpF*), all with $h^2 \geq 0.20\pm0.11$. Others could be important

determining microbial fermentation by regulating H_2 intraluminal pressure; e.g. *mtrB*, *mvhD*, *fdhB*, *hdrA2* in methane metabolism presented $h^2 \geq 0.28 \pm 0.13$. Importantly, HGFC also contained *alr*-MGs involved in cell-to-cell microbial communication processes such as quorum sensing (e.g. ABC.PE-*S* and ABC.PE.*P*) and also in genetic processes such as ribosomal biosynthesis (e.g. *RP-L7*, *RP-L2*). Of the 841 *alr*-MGs comprising the HGFC, 118 had rg_{FCR} estimates with $P_0 \geq 90\%$; 48 of them were positive (rg_{FCR} from 0.5 to 0.8), and 70 negative (rg_{FCR} from -0.47 to -0.83) (Figure 1b), suggesting that a part of the HGFC is under common host genomic influence with FCR. Interestingly, the abundance of *mtrB* and *mvhd* in methane metabolism was associated with less efficient animals (rg_{FCR}=0.53 and 0.56, P_0=0.90 and 0.91), whereas the abundance of those involved in the production of aminoacids (*metC, metN, ureA, ureB, speB*) or volatile fatty acids (*ACAT*), sulfur metabolism (*cysD*) or quorum sensing (ABC.PE.*S*) were associated with more efficient hosts (rg_{FCR} -0.54 to -0.73, $P_0 \geq 0.90$).

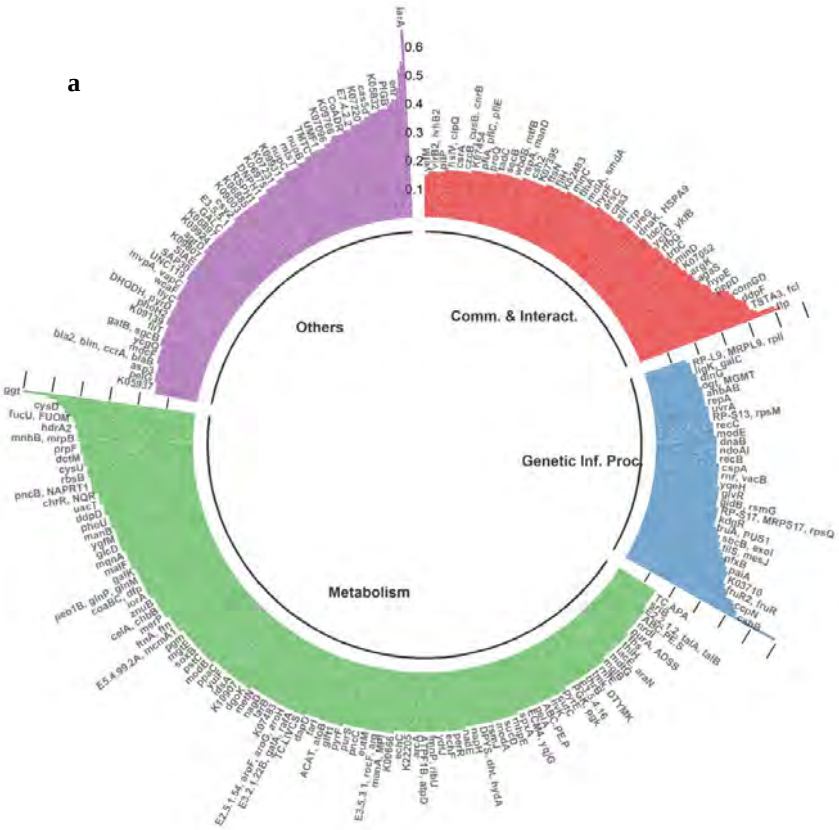


Figure 1. Genomic parameters of additive log-ratio transformed abundances of microbial genes (*alr*-MGs) in the rumen of beef cattle. (a) Heritability estimates of 841 *alr*-MGs with significant host genomic effects based on Bayes Factor>17.2 and Deviance Information Criterion difference≤-10, comprising the host-genomically influenced functional core microbiome (HGFC).

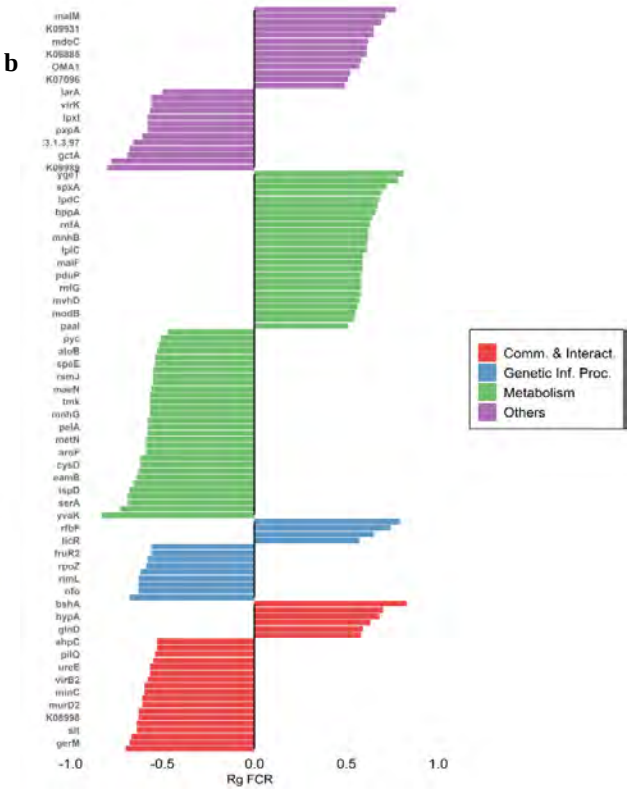


Figure 1. Continued. (b) Genomic correlations between feed efficiency (rg_{FCR}) and 118 *alr*-MGs from the HGFC for which rg_{FCR} was ≥90% likely to be positive or negative. Microbial genes are grouped according to microbial biological processes: Microbial communication and host-microbiome interaction (Comm. & Interact.), Genetic information processes (Genetic Inform. Proc.), Metabolism, and Others.

Discussion

Our study suggests that differences among animals in complex functions of the microbiome are partly (~23%) under the control of the host genome, which is consistent with our previous findings in other experimental populations (~18%), and raises the possibility of exploiting a new source of genetic variation in ruminants. Interestingly, some of the *alr*-MGs with positive rg_{FCR} are involved in methane metabolism (*mtrB and mvhd*), and others with negative rg_{FCR} (*metN*) were estimated to have favourable r_g with methane emissions in a previous study (unpublished data). This suggests that genomic selection on the functional microbiome could be a cost-effective permanent solution to achieve a ruminal balance that optimises fermentation of feed into essential nutrients utilised by the host, while favouring other traits as mitigation of emissions of the greenhouse gas methane.

References

Greenacre, M, Martínez-Álvaro M. and Blasco A. (2021) Front. Microbiol. 12:727398. https://doi.org/10.3389/fmicb.2021.727398

Martínez-Álvaro M., Duthie C.A., Mattock J., Dewhurst R., Cleveland M.A., *et al.* (2021) Proc. of the 72nd EAAP, Davos, Switzerland

Wen, X.(2017) Stat. Biosci. 9:28–49. https://doi.org/10.1007/s12561-016-9153-0

70. Identifying diversity in cattle performance using Bayesian inference and a model

J.F. Ramirez*, L. Puillet and N.C. Friggens

Université Paris-Saclay, INRAE, AgroParisTech, UMR Modélisation Systémique Appliquée aux Ruminants, 75005 Paris, France; johnfredy.ramirezagudelo@inrae.fr

Abstract

Quantitative description of individual animal performance is a key element of breeding. In addition to measurements of traits associated with productivity, dynamic-mechanistic models can be used for identifying diversity of complex traits like feed efficiency. The objective of this work was to use Bayesian inference to quantify differences in a model's 'genetic-scaling' parameters (related to resource acquisition and resource allocation) within- and between-breeds. We first describe the experimental data processing before calibration, the function used for the error calculation, and the Bayesian algorithm used for parameter inference. Two databases were used as case studies. The results showed that this approach allowed us to identify variations in the genetic-scaling parameters within- and between-breeds. This approach is a useful tool to quantify complex traits for future selection of efficient and resilient animals.

Introduction

In breeding programs, most of the emphasis has been made on traits directly associated with productivity, such as milk yield or daily weight gain. However, the biological mechanisms underlying productivity are difficult to be taken into account for selection. When direct measurements of biological mechanisms are difficult or even impossible, mechanistic models allow estimation of these underlying traits by individual parametrization of a model. Puillet *et al.* (2016) developed a dynamic-mechanistic model (AQAL) to generate genetically-driven trajectories of energy acquisition and allocation in dairy cows lifespan, with four genetic-scaling parameters (GSP). For a given individual, the simulated performance is the result of the combination of a capacity to acquire resources and how these acquired resources are allocated to different life functions. This model was used to evaluate the extent of genotype by environment interaction for feed efficiency traits (Puillet *et al.*, 2021), demonstrating the usefulness of mechanistic modelling in breeding strategies evaluation. In this type of model, Bayesian inference can be used to quantify the uncertainty in the parameter values, because the parameters are treated as random variables with an underlying probability distribution that describes them (Miles and Smith, 2019). This work presents a protocol to use Bayesian inference for within- and between-breeds characterization of resource acquisition and allocation parameters. The protocol presented here aims to provide a comprehensive description for the individualized calibration of AQAL in Python. First, we describe the observed data processing before calibration. Then, we describe the general structure of the calibration process. We illustrate the application of the protocol for two case studies.

Materials & methods

Time-series observations, data handling before calibration. For calibrating AQAL, experimental data is necessary to run the model and to compute the error. The model uses a vector of values to represent the increase in intake per lactation through the cow's lifespan. This vector must be defined by using the relationship among the averages of the maximum values (upper quartile) of milk yield observed in each lactation. The energy value of the diet (GE or ME) is another model input. At least three traits must be used to calculate the differences between simulated and experimental data: milk yield (MY), body weight (BW), and body condition score (BCS, 1-5 scale). MY must be expressed in kilocalories of metabolizable energy (ME) assuming a metabolic efficiency of 0.64, and gross energy content (GE, kcal/kg) of 5,700, 9,200, 3,950 for protein, fat, and lactose.

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_70, © J.F. Ramirez *et al.* 2022

Formulation of the objective function. In each iteration of the calibration process, the objective function uses the set of parameters generated by the Bayesian algorithm and returns the mismatch between the simulation and the experimental values of MY, BW, and BCS. We used the Mean Absolute Percentage Error (MAPE) to quantify the mismatch because this metric avoids the effect of the trait scales on the average error calculation. One MAPE is calculated for each variable × lactation, then the overall average MAPE was calculated from these individual variable × lactation MAPEs. To minimize the effect of unbalances among the individual errors, the overall error variance was added to the overall MAPE. After calibration, the minimum value of this sum was used to identify the best parameter set for a given individual animal.

Parameter optimization. In the Bayesian approach, the goal of the calibration is to infer the parameter posterior distributions given a set of observations (Miles and Smith, 2019). For parameter generation, we adopt an approximate-inference technique based on Monte Carlo methods, which aims at generating a sequence of random parameters from a Markov chain whose distributions approach the posterior distributions of each parameter. In particular, we have used the delayed rejection adaptive Metropolis (DRAM) algorithm (Haario *et al.*, 2006) implemented for Python in the 'pymcmcstat' library (Miles and Smith, 2019). The adaptive Metropolis (AM) component in DRAM means that each parameter is generated following the distribution calculated with all of the previous states in the chain. The delayed rejection (DR) mechanism is used to stimulate mixing by sampling from a narrower proposed distribution when a parameter sample is rejected (Haario *et al.*, 2006). Information known a priori about the parameter distributions is defined in the prior function; however, the default program behaviour is to use a uniform prior which is a common approach for these types of problems. (Miles and Smith, 2019). The parameters optimized were related to acquisition: *acq_bas_pot*, *acq_bas_lac* (basal and an additional lactation acquisition), and allocation of energy resource: *init_G*, *s2pc_pot*, (growth and lactation allocation).

Case studies. Two databases were used to test our parameters inference procedure. The first database (One-Breed) contains data from 26 Holstein cows. This database reported at least four lactations for all cows. The second database (MEMO, see Friggens *et al.* 2007) contains data from three breeds (18 Holstein, 17 Red Danish, and 19 Jersey cows). This database reported at least three lactations for all cows. Both databases report measures of MY and composition, BW, and BCS. The first and second databases were used to examine within- and between-breed diversity of the four GSP of AQAL, respectively.

Results

Figure 1 shows an example of simulation, using the best set of parameters for a specific cow, and observed data of MY, BW, and BCS. Figure 2 shows an example of the Markov chains for a specific cow. The points in charts of Figure 2 correspond to the accepted parameters values in each iteration. The first half of each chain served to tune the sampler and was discarded (burn-in period). To approximate the posterior distribution for each parameter, parameter sets with low error from the second half of each chain were used. Figure 3 compares the parameter distributions for a specific cow and the breed (One-Breed database). Figure 4 shows the parameter distributions for the three breeds (Holstein, Red Danish, and Jersey) in the MEMO database. The parameters *acq_bas_pot* and *acq_bas_lac* are expressed in kilograms of dry matter, the parameters init_G and s2pc_pot are dimensionless. Table 1 presents the average parameters values and the associated variance.

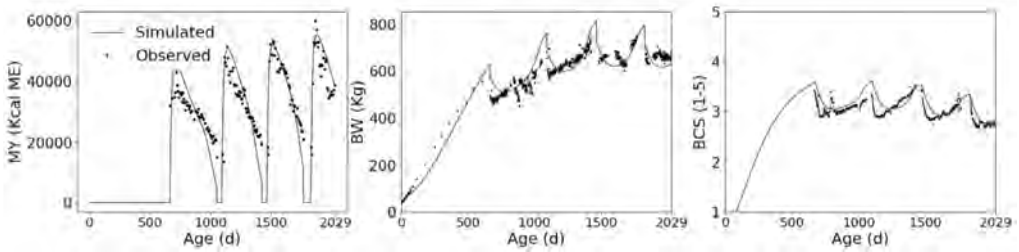


Figure 1. Example of simulation results with the best set of parameters for an individual.

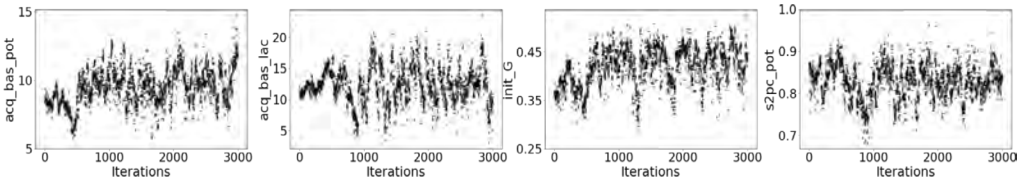


Figure 2. Example of the Markov chains of genetic-scale parameters.

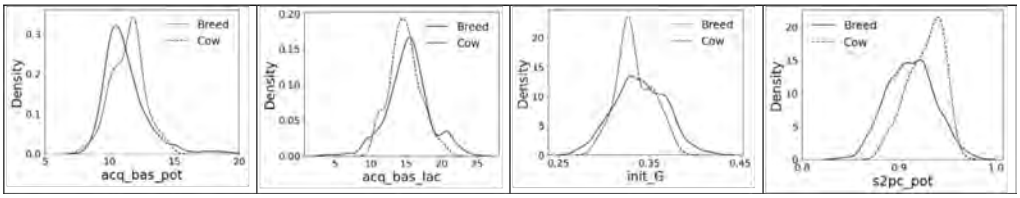


Figure 3. Posterior distributions of the genetic-scale parameters for a breed and an example of one individual of that breed.

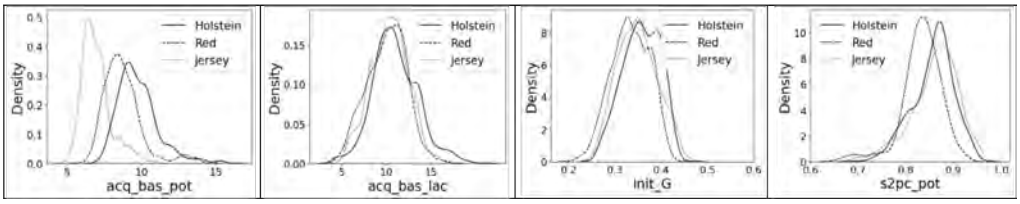


Figure 4. Posterior distributions of the genetic-scale parameters between breeds.

Table 1. Parameter values per database and breed.

Parameters	One-breed	MEMO		
		Holstein	Danish Red	Jersey
	Median (Std)	Median (Std)	Median (Std)	Median (Std)
acq_bas_pot	11.1 (1.3)	9.9 (1.1)	8.8 (1.3)	7.2 (1.3)
acq_bas_lac	16.1 (2.9)	12.0 (2.0)	10.9 (1.4)	10.4 (1.8)
init_G	0.34 (0.025)	0.36 (0.039)	0.33 (0.040)	0.35 (0.037)
s2pc_pot	0.92 (0.022)	0.86 (0.030)	0.84 (0.029)	0.86 (0.042)

acq_bas_pot, *acq_bas_lac:* basal and an additional lactation acquisition (kg of DM).
init_G, *s2pc_pot:* growth and lactation allocation (Dimensionless).

Discussion

We chose to use a Bayesian method for parameter estimation because we expected to explore different local minima within the response surface of the objective function (since the model contains multiple nonlinear functions). In this context, the Bayesian inference used (DRAM) was an efficient method to explore the parameter space, showing the power of the proposed method for this type of task. The results allow us to demonstrate the ability of a dynamic-mechanistic model to identify variations in GSP both within- and between-breeds. This is very interesting for the future identification and selection of efficient and resilient animals because dynamic-mechanistic models (e.g. AQAL) can simulate the consequences of different combinations of acquisition and allocation parameters on animal performance under different nutritional environments (see Puillet *et al.*, 2021).

Acknowledgements

This work has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 727213 (GenTORE). We thank Trouw Nutrition R&D for sharing part of the data used for the work presented in this manuscript.

References

Friggens N.C., Berg P., Theilgaard P., Korsgaard I.R., Ingvartsen K.L., *et al.* (2007) J Dairy Sci 90(11):5291-5305. https://doi.org/10.3168/jds.2007-0173

Haario H., Laine M., Mira A., and Saksman E. (2006) Stat Comput 16(4):339-354. https://doi.org/10.1007/s11222-006-9438-0

Miles P.R., and Smith R.C. (2019) Proc. of the 18th Python in Science Conference (SCIPY 2019), Texas, USA.

Puillet L., Ducrocq V., Friggens N.C., and Amer P.R. (2021) J Dairy Sci 104(5):5805-5816. https://doi.org/10.3168/jds.2020-19610

Puillet L., Réale D., and Friggens N.C. (2016) Genet Sel Evol 48(1):1-16. https://doi.org/10.1186/s12711-016-0251-8

71. Feed efficiency of crossbreds of Montbéliarde, Viking Red, and Holstein with Holstein cows fed two alternative diets

B.J. Heins[1,2*], G.M. Pereira[1,2] and L.B. Hansen[1]

[1]University of Minnesota, 1364 Eckles Ave, 55108, St. Paul, MN, USA; [2]West Central Research and Outreach Center, University of Minnesota, 46352 State Hwy 329, 56267, Morris, MN, USA; hein0106@umn.edu

Abstract

The objective of this study was to compare three-breed rotational crossbred (CB) cows of the Montbéliarde, Viking Red, and Holstein (HO) breeds with HO cows fed two alternative diets for dry matter intake (DMI), fat plus protein production (CFP), body weight (BW), body condition score (BCS), feed efficiency and residual feed intake (RFI) from 46 to 150 days in milk during first lactation. Cows were fed either a traditional total mixed ration (TMR) diet (TRAD) or a higher fibre, lower starch TMR diet (HFLS). The CB cows fed HFLS had less DMI (-12%) compared with the HO cows fed TRAD. For CFP, CB and HO cows were not different when fed TRAD or HFLS. The mean RFI from 46 to 150 days was lower and more desirable for CB cows fed HFLS (-120.0 kg) compared with HO cows fed TRAD (85.3 kg).

Introduction

Feed costs impact the profitability of dairy farms and represent 47% of the total costs of U.S. milk production. In recent years, dairy producers have experienced volatility of cost when purchasing grain for dairy cow diets. Therefore, dairy producers have explored alternative diets for dairy cattle that contain more forage and less concentrate to improve feed efficiency and profitability. Research of crossbred (CB) cows has predominantly focused on production, fertility, and health, but little research has compared CB cows and Holstein (HO) cows for feed efficiency. The dry matter intake (DMI) and feed efficiency of CB and HO cows fed the same TMR have been compared in confinement herds (Heins *et al.*, 2008; Hazel *et al.*, 2013). Shonka-Martin *et al.* (2019) reported 3-breed CB cows of Montbéliarde (MO), Viking Red (VR), and HO breeds had a lower mean RFI compared with HO cows fed a similar TMR during the first 150 days of first lactation. Dairy producers have hypothesized three-breed rotational CB cows of MO, VR, and HO are more feed efficient than HO cows fed a TMR with more fibre and less starch. Currently, a TMR based on corn silage is fed to most cows in confinement herds in the U.S., and dairy producers have explored alternative diets to improve nutrient digestibility, feed efficiency and profitability of cows. Therefore, the objective of this study was to assess the effects of feeding two alternative diets to three-breed rotational CB cows of MO, VR, and HO versus HO cows for DMI, fat plus protein production, body weight, BCS, feed efficiency, and RFI from 46 to 150 days in milk during first lactation.

Materials & methods

Cows were housed at the Dairy Cattle Teaching and Research Facility of the University of Minnesota in St. Paul, USA. Primiparous cows calved from September 2019 to March 2020. In total, 36 cows (17 CB, 19 HO) were unimpaired experimental units with data available from 46 to 150 days in milk.

Collection of feed intake and diets. Cows were assigned to an individual tie stall in a confinement barn for the duration of the study, and the feed mangers were partitioned to provide each cow an exclusive feeding space. The feed delivered and the refusals were electronically recorded with Feed Supervisor (Supervisor Systems, Dresser, WI) software. Diets fed to cows were formulated by a professional nutritionist to meet or exceed dairy NRC requirements. The two diets were formulated to have equivalent net energy content of 1.6 Mcal/kg and CP of 16.8%. Cows were paired by expected calving date and subsequently assigned randomly to a traditional diet (TRAD) or a higher fibre, lower starch diet (HFLS) that were both fed as

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_71, © B.J. Heins *et al.* 2022

a total mixed ration (TMR). Diets were composed of corn silage, chopped alfalfa hay, corn gluten feed, cottonseed, ground corn, molasses, concentrate supplements and vitamins and minerals

The daily feed intake multiplied by the average monthly dry matter for TRAD and HFLS provided daily DMI for individual cows, and these were smoothed using the mean of seven days (the actual day, the third days before and the third days after). Legendre polynomial random regression function was used to estimate 105-day lactation curves for DMI from 46 to 150 days in milk with four parameters from PROC HPMIXED of SAS. Daily milk production was the sum of the morning and evening milking. Fat, protein, and lactose percentage were from twice monthly Dairy Herd Improvement Association samples (Minnesota DHIA, Buffalo, MN). The BW and body condition score (BCS) were recorded once weekly. The BW was recorded on individual cows with an electronic scale as cows exited the milking parlour during the morning milking. The BCS was recorded by one person throughout the study on a five-point scale with one = thin and five = fat.

Feed efficiency and residual feed intake. The regression coefficients (energy sinks) to estimate RFI of cows were energy expended in milk (MilkE), metabolic BW (MBW) and the energy for BW change (BWCE; Hardie *et al.*, 2015; Shonka-Martin *et al.*, 2019). The MilkE was estimated according to the NRC (2001) as MilkE = 9.29 × fat (kg) + 5.63 × protein (kg) + 3.95 × lactose (kg). Legendre polynomial random regression function was used to estimate 105-day lactation curves for MilkE from 46 to 150 days in milk with three parameters from PROC HPMIXED of SAS. Estimates for daily MilkE were summed from 46 to 150 days in milk to determine total MilkE for each cow. The MBW reflected the maintenance requirements for a cow and was calculated as $BW^{0.75}$ with the mean BW of each cow from 46 to 150 days in milk. The BWCE was estimated as BWCE (Mcal/d) = body weight change (BWC; kg/d) × (2.88 + 1.026 × BCS) × 0.85 (NRC, 2001; Hardie *et al.*, 2015; Shonka-Martin *et al.*, 2019). The BCS was the mean BCS of each cow from 46 to 150 days in milk. The MilkE, MBW, and BWCE were energy sinks in the statistical model to estimate RFI. The RFI was defined as the error term (ε_i) from the regression model:

$$DMI_i = \beta_0 + \beta_1 \times MilkE_i + \beta_2 \times MBW_i + \beta_3 \times BWCE_i + \varepsilon_i \quad (1)$$

where DMI_i is the total DMI from 46 to 150 days in milk for the i^{th} cow; β_0 is the intercept and β_1, β_2, β_3 are the regression coefficients for MilkE, MBW and BWCE from 46 to 150 days in milk, respectively. All CB and HO cows were analysed simultaneously.

Statistical analysis. Independent variables for the statistical analysis of DMI, CFP, BW, BCS, feed efficiency and RFI were the fixed effects of diet (TRAD or HFLS) and breed group (CB or HO) as well as the interaction of diet and breed group. The PROC GLM of SAS was used to conduct the ANOVA and obtain solutions.

Results

The least squares means and standard errors of means from 46 to 150 days in milk of DMI, CFP, BW, BCS, Milk energy, metabolic body weight, body weight change in energy, and RFI are in Table 1. The HO cows fed TRAD (2,592 kg) had more (P<0.05) DMI than HO cows fed HFLS (2,229 kg) and CB cows fed HFLS (2,278 kg). The CB and HO cows that were fed TRAD and HFLS were not different for DMI. The CB and HO cows were not different across diets for CFP. The BW did not differ for CB and HO cows that were fed TRAD. The BCS was higher (P<0.05) for CB cows fed either TRAD or HFLS compared with HO cows fed either TRAD or HFLS. The CB cows were not different for BCS whether fed TRAD or HFLS, and the HO cows were not different for BCS across the two diets.

The CB cows that were fed HFLS (121 kg) had more (P<0.05) MBW compared with HO cows that were fed HFLS (114 kg). The CB cows that were fed either TRAD or HFLS had more (P<0.05) BWCE than HO cows that were fed HFLS. The HO cows fed TRAD had more (P<0.05) BWCE compared with HO cows that were fed HFLS. The RFI was lower (P<0.05) for CB cows that were fed HFLS (-120.0 kg) compared with the HO cows that were fed TRAD (+85.3 kg), and less RFI is favourable.

Discussion

The lower DMI, higher BCS, and similar CFP for CB cows fed HFLS compared with HO cows fed TRAD in this study may suggest improved profitability when cows are fed alternative diets. Perhaps, the CB cows used DMI beyond their needs for production and maintenance for higher BCS. This study only investigated 46 to 150 days in milk of first lactation, and additional research is needed to compare multiple lactations of CB and HO cows across 305 days. Dairy producers initially moved to CB cows to improve health and fertility; however, dairy producers may improve profitability of dairying with CB cows fed higher forage and lower starch diets.

The partial regression coefficients were similar to those of Shonka-Martin *et al.* (2019) who reported coefficients of 0.75, 8.45, and 0.62 for MilkE, MBW, and BWCE respectively for 3-breed rotational CB cows of MO, VR and HO and HO cows that were fed the same diet. The CB cows fed HFLS had lower mean RFI than HO cows fed TRAD but did not differ for MilkE, MBW and BWCE. Furthermore, CB cows fed HFLS consumed less DMI than HO cows fed TRAD. The more favourable RFI for the CB cows fed HFLS met expectations because the CB cows that were fed HFLS had less DMI (-314 kg) and had lower DMI/BW (-0.63) than HO cows that were fed TRAD. Similarly, Shonka-Martin *et al.* (2019) reported three-breed rotational CB cows of MO, VR and HO had a lower mean RFI (-65.5 kg vs +68.8 kg) than HO cows during the first 150 day of first lactation.

In the current study, the CB cows fed HFLS may have had more maintenance requirement for BW; however, CFP and MilkE were not different for the CB and HO cows. The RFI does not consider change in BCS, as well as body measurements (stature, dairy strength, and body depth), fertility and longevity. Therefore,

Table 1. Least square means and their SE for DMI, CFP, BW, BCS, and feed efficiency from 46 to 150 days in milk for primiparous Holstein and 3-breed rotational crossbred cows fed a traditional and a higher fibre lower starch diet.[1,2]

Variable	Traditional				Higher fibre lower starch			
	Holstein		Montbéliarde/Viking Red/ Holstein		Holstein		Montbéliarde/Viking Red/ Holstein	
	LSM	SE	LSM	SE	LSM	SE	LSM	SE
DMI	2,592^{a}	99.2	2,436^{ab}	109.7	2,229^{b}	116.3	2,278^{b}	116.3
CFP	269	7.2	256	7.9	256	8.4	255	8.4
BW	586^{ab}	14.8	576^{ab}	16.3	550^{a}	17.3	600^{b}	17.3
BCS	3.19^{a}	0.1	3.64^{b}	0.1	3.02^{a}	0.1	3.64^{b}	0.1
MilkE	2,808	74.8	2,651	82.7	2,645	87.8	2,610	87.8
MBW	119^{ab}	2.3	117^{ab}	2.5	114^{a}	2.7	121^{b}	2.7
BWCE	342^{a}	24.8	447^{b}	27.4	264^{c}	29.0	377^{ab}	29.0
RFI	85.3^{a}	56.7	0.8^{ab}	62.7	1.8^{ab}	66.5	-120.0^{b}	66.5

[1] Means within a row with different superscripts differ P<0.05.

[2] BWCE = the energy for body weight change; MBW = metabolic body weight; MilkE = milk energy.

RFI may be underestimated in the current study for CB cows compared with HO cows. Dairy producers may feed either TRAD or HFLS to CB cows without loss of CFP. For HO cows, the DMI was less for the cows that were fed HFLS compared with the cows that were fed TRAD. Less RFI for the CB cows fed HFLS compared with the HO cows that were fed TRAD indicated that the CB cows were more feed efficient than HO cows. Genetic selection of feed efficient cows should not be based solely on reducing DMI, but should consider BW, BCS, production, health, fertility, and longevity.

References

Hardie L.C., Armentano L.E., Shaver R.D., VandeHaar M.J., Spurlock D.M., *et al.* (2015) J. Dairy Sci. 98:2727–2737. https://doi.org/10.3168/jds.2014-8580

Hazel A.R., Heins B.J., Seykora A.J., and Hansen L.B. (2013) J. Dairy Sci. 96:1915–1923. https://doi.org/10.3168/jds.2012-5667

Heins B.J., Hansen L.B., Seykora A.J., Hazel A.R., Johnson D.G., *et al.* (2008) J. Dairy Sci. 91:3716–3722. https://doi.org/10.3168/jds.2008-1094

NRC (2001) Nutrient Requirements of Dairy Cattle. Natl. Acad. Press, Washington, DC.

Shonka-Martin B.N., Heins B.J. and Hansen L.B. (2019) J. Dairy Sci. 102:3661–3673. https://doi.org/10.3168/jds.2018-15682

72. Genetic and phenotypic associations among production, feed efficiency and immune traits in indigenous chicken of Kenya

S.A. Miyumo[1*], C.B. Wasike[2], E.D. Ilatsia[3], J. Bennewitz[4] and M.G.G. Chagunda[1]

[1]Department of Animal Breeding and Husbandry in the Tropics and Sub-tropics, University of Hohenheim, Garben str. 17, 70599 Stuttgart, Germany; [2]Livestock Efficiency Enhancement group (LEEG), Department of Animal and Fisheries Sciences, Maseno University, P.O. private bag 40105 Maseno, Kenya; [3]Kenya Agricultural and Livestock Research Organization, P.O. Box 25, 20117, Naivasha, Kenya; [4]Department of Animal Breeding and Genetics, University of Hohenheim, Garben str. 17, 70599 Stuttgart, Germany; sophie.miyumo@uni-hohenheim.de

Abstract

This study aimed at determining genetic associations between production traits with feed efficiency and immune traits pre- (eight to 20 weeks) and post (90 days from on-set of lay) maturity in indigenous chicken. Production traits were body-weight at week 12 (BW_{12}) and age at first egg (AFE). Feed efficiency included residual feed intake (RFI) and feed conversion ratio (FCR). Immunity included natural antibodies binding to keyhole limpet hemocyanin (KLH-IgM) and specific antibodies binding to Newcastle disease (NDV-IgG). Heritability estimates in the two periods across traits varied from 0.10 to 0.51. Genetic correlations showed that BW_{12} was negatively correlated with RFI but positively with FCR, KLH-IgM and NDV-IgG in the growing period. Post maturity, AFE was positively correlated with RFI but negatively with FCR, KLH-IgM and NDV-IgG. Results from this study indicate that selection for productivity may have genetic consequences on feed efficiency and immune-competence and this relationship should be considered.

Introduction

Indigenous chicken (IC) play significant roles in nutrition, food security and economic growth in many rural households in most countries in the tropical regions. In Kenya, these chicken are produced under challenging environments (scarce feed resources and high exposure to myriad pathogens) limiting their optimal utilization (Lwelamira *et al.*, 2009). Nevertheless, the Kenyan IC is a highly diverse population with a large plasticity in genomic regions that confer adaptability to scavenging conditions, poor nutrition and tolerance to pathogens (Ngeno *et al.*, 2015). Consequently, efforts are underway to exploit the genetic diversity of IC population to improve their productivity. Selection for production is likely to influence resource partitioning. With resource scarcity, preferential partitioning of resources to specific functions is inevitable. This leads to trade-offs between different functions in the body e.g. high growth rate and low antibody response upon disease challenge (Van der Most *et al.*, 2010). This demonstrates competitive nutrient partitioning between growth and immune functions, and also suggests possible involvement of pleiotropic gene action. Considering IC ability to produce and survive in poor conditions, studies on the impact of improving their growth or egg production on resource use efficiency and immunity are sparse. The current study aimed at estimating genetic and phenotypic correlations among production, feed efficiency and immune traits pre- and post-maturity in indigenous chicken.

Materials & methods

The study was conducted at the poultry research unit of Kenya Agriculture and Livestock Research Organization (NRI-KALRO), Naivasha. The IC population under the chicken breeding program at NRI-KALRO are under continuous selection to develop meat (ML) and egg (EL) lines. The ML is selected for body weight at 12 weeks of age (BW_{12}) while EL is selected based on age at first egg (AFE) (Ilatsia *et al.*, 2017). Five generations of these chicken comprising of a base generation/control line (CL), and four generations of both ML and EL were used. Management of the experimental population followed the standard operating

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_72, © S.A. Miyumo *et al.* 2022

procedures of the breeding program at NRI-KALRO (Ilatsia *et al.*, 2017). From week eight, the birds were transferred to individual feeding units for individual weight and feed intake data collection. Routine health management involved vaccination against Marek's disease, infectious bursal disease, Newcastle disease, Fowl pox and Fowl Typhoid which are endemic, deworming and disinfection.

Data collection. Data on BW_{12} was recorded while date of hatch and first egg were used to compute AFE in the ML and EL. On both lines, daily feed intake and monthly weight records were collected and used to compute average daily feed intake (ADFI), average daily gain (ADG) and metabolic body weight (MBW), residual feed intake (RFI) and feed conversion ratio (FCR) in the growing period (eight to 20 weeks) (Miyumo *et al.*, 2016). At onset of lay, initial body weight, daily feed intake, egg number and weight recording commenced and continued until day 90, in which final body weight was measured. Using these traits, estimation of ADFI, ADG, MBW, average daily egg mass (EM), RFI and FCR in the laying period was done as described by Yuan *et al.* (2015). To assess antibody traits during the growing and laying periods, birds were blood sampled at 16 and 28 weeks of age, respectively. Natural antibodies (Nabs-IgM isotype) binding to keyhole limpet hemocyanin (KLH) and specific antibodies (IgG isotype) binding to Newcastle disease virus (NDV) were evaluated; KLH antigen was used based on its suitability to test Nabs (Van der Klein *et al.*, 2015) while NDV was considered due to its impact on IC populations (Lwelamira *et al.*, 2009). Antibodies KLH-IgM and NDV-IgG were measured using indirect ELISA (Enzyme-linked Immunosorbent assay). In the growing period, 1,279 records on BW_{12}, 721 records on each of the feed efficiency traits and 1,162 records on each of the antibody traits were available for analyses. In the laying period, 652 records on AFE, 578 records on each of the feed efficiency traits and 652 on each of the antibody traits were used in the analyses.

Statistical analysis. A series of bi-variate mixed animal models were fitted to estimate trait heritabilities as well as genetic and phenotypic correlations among BW_{12}, AFE, RFI, FCR, KLH-IgM and NDV-IgG by restricted maximum likelihood using the WOMBAT software (Meyer, 2007). The model included sex, generation and line as fixed effects and a random animal and a random residual effects. Due to the various combinations of the bi-variate models, the resulting variance components and variance ratios were pooled by weighting each estimate by the inverse of its sampling variance to obtain global estimates.

Results

Genetic parameters on production, feed efficiency and immune traits during the growing and laying periods are presented in Table 1. Heritability estimates across the traits ranged from 0.10 to 0.44 in the growing period and from 0.25 to 0.51 in the laying period. Genetic correlations (r_g) among the traits varied between the two production periods. During the growing period, BW_{12} was negatively (r_g=-0.96) associated with RFI but positively correlated with FCR (r_g=0.29) and KLH-IgM (r_g=0.47) with low association with NDV-IgG (r_g=0.06). There were negative correlations between RFI and FCR (r_g=-0.99), and between KLH-IgM and NDV-IgG (r_g=-0.85). Between efficiency and immunity, RFI was positively correlated with KLH-IgM (r_g=0.92) and NDV-IgG (r_g=0.78). In the laying period, AFE was positively (r_g=0.35) correlated to RFI but negatively associated with FCR (r_g=-0.44) and the immune traits (r_g=-0.12 to -0.39). On the other hand, there were negative correlations between RFI and FCR (r_g=-0.41), and between KLH-IgM and NDV-IgG (r_g=-0.99). Between efficiency and immune measures, RFI was positively correlated to KLH-IgM (0.40) and NDV-IgG (0.50). Genetic correlations of traits between the growing period and laying period are presented in Table 2. Between the two production traits, BW_{12} and AFE were negatively (r_g=-0.88) correlated. For RFI, KLH-IgM and NDV-IgG, the correlations were positive (r_g=0.62 to 0.99) between the two production phases.

Table 1. Estimates of heritability[1] (diagonal), genetic (lower diagonal) and phenotypic (upper diagonal) correlations on production, efficiency and immunity traits.

	Growing period (9 to 20 weeks of age)				
Traits[2]	**BW_{12}**	**RFI**	**FCR**	**KLH-IgM**	**NDV-IgG**
BW_{12}	**0.44±0.03**	0.06	-0.27	0.04	-0.05
RFI	-0.96	**0.39±0.02**	0.22	0.02	-0.04
FCR	0.29	-0.99	**0.42±0.04**	-0.18	0.01
KLH-IgM	0.47	0.92	-0.17	**0.22±0.03**	0.01
NDV-IgG	0.06	0.78	0.03	-0.85	**0.10±0.02**
	Laying period (from onset of lay to 90 days into egg production)				
Traits[2]	**AFE**	**RFI**	**FCR**	**KLH-IgM**	**NDV-IgG**
AFE	**0.31±0.04**	-0.06	-0.08	-0.04	-0.02
RFI	0.35	**0.30±0.05**	0.30	-0.05	-0.15
FCR	-0.44	-0.41	**0.34±0.04**	0.03	0.07
KLH-IgM	-0.39	0.40	-0.19	**0.51±0.04**	-0.16
NDV-IgG	-0.12	0.50	0.14	-0.99	**0.25±0.03**

[1] ±SE = standard errors.

[2] BW_{12} is body weight at week 12; AFE is age at first egg; RFI is residual feed intake; FCR is feed conversion ratio; KLH-IgM is natural antibody IgM isotype binding to KLH antigen; NDV-IgG is specific antibody IgG isotype binding to NDV antigen.

Table 2. Genetic (r_g) and phenotypic (r_p) correlations on production, efficiency and immunity traits between the growing and laying periods.

Growing period[1]	**Laying period[1]**	**r_g**	**±SE**	**r_p**	**±SE**
BW_{12}	AFE	-0.88	0.38	-0.08	0.05
RFI_1	RFI_2	0.99	0.12	0.04	0.01
FCR_1	FCR_2	0.19	0.18	0.08	0.04
$KLH\text{-}IgM_1$	$KLH\text{-}IgM_2$	0.62	0.21	0.05	0.03
$NDV\text{-}IgG_1$	$NDV\text{-}IgG_2$	0.82	0.17	0.09	0.03

[1] BW_{12}, AFE, RFI, FCR, KLH-IgM and NDV-IgG are as described in Table 1: RFI_1, FCR_1, $KLH\text{-}IgM_1$ and $NDV\text{-}IgG_1$ are traits measured in the growing period; RFI_2, FCR_2, $KLH\text{-}IgM_2$ and $NDV\text{-}IgG_2$ are traits measured in the laying period.

Discussion

Heritability estimates across the traits in both periods (apart from NDV-IgG in the growing period) suggest that adequate additive genetic variance exist, which could be exploited for selection. Genetic associations between production and efficiency traits indicated that genetic improvement for BW_{12} and AFE would be accompanied with lower RFI values (high efficiency) but higher FCR values (low efficiency). Although both are measures of feed efficiency, the disparity in correlation patterns with BW_{12} and AFE could be due to RFI accounting for maintenance function while FCR does not (Yuan *et al.*, 2015). This is further supported by the negative correlations between RFI and FCR in both production periods, indicating that chicken deemed efficient based on RFI would be less efficient when using FCR. This observation, however, contradicts results reported by Yuan *et al.* (2015). Aggrey and Rekaya (2013) found that MBW had more influence on efficiency than production and that MBW was positively and negatively correlated with FCR and RFI, respectively. It is possible that the current population was diverting more resources into maintenance than production and hence, the negative associations between RFI and FCR. Considering

that KLH-IgM has been previously linked to higher survival rates (Sun *et al.*, 2013), the correlations imply that improvement of production traits BW_{12} and AFE may not have negative effects on survivability in IC. On the other hand, selection for BW_{12} and AFE would have little to no influence on adaptive humoral immunity (NDV-IgG). Although RFI reflected the efficiency of feed use into production, its correlations with immune traits suggested inefficiencies such that chicken with higher productivity and immune-competence consumed more feed to support both functions. These observations indicate that an indirect relationship between feed efficiency and immune function may be present (Van der Most *et al.*, 2011). The high and negative association between the two antibody traits in both production periods could be attributed to the different T helper-cell (Th) pathways induced by KLH and NDV (Sun *et al.*, 2013). The observed negative correlation between BW_{12} and AFE implies that selection for high BW_{12} would lead to delayed sexual maturity. This may be undesirable for the egg lines given the negative associations between AFE and egg numbers (Van der Klein *et al.*, 2015). Feed efficiency measured pre- and post-maturity shared a similar genetic background suggesting that efficient chicken in the growing period would likely rank as efficient in the laying period. Similar pattern was also observed in the immune traits such that chicken with higher antibodies pre-maturity maintained their immune-competence post-maturity.

Conclusions

The current study confirms presence of genetic correlations between production traits with feed efficiency and immune traits in the IC population. This indicates that possible consequences on feed efficiency and immune function should be considered when selective breeding for improved productivity in IC is applied.

References

Aggrey S.E., and Rekaya R. (2013) Poult. Sci., 92(10):2600-2605. https://doi.org/10.3382/ps.2013-03302.

Ilatsia, E.D., Magothe, T.M., Okeno, T.O., Muasya T.K., Ng'eno K., *et al.* (2017) Proc. of the ASAL-APRP End of Project Conference and Exhibition, Nairobi, Kenya.

Lwelamira J., Kifaro G.C., and Gwakisa P.S. (2009) Trop. Anim. Health Prod. 41(1):51-59. https://doi.org/10.1007/s11250-008-9153-2.

Meyer K. (2007) J. Zhejiang Univ. Sci. B 8(11):815-821. https://doi.org/10.1631/jzus.2007.b0815.

Miyumo S., Kahi A.K., and Wasike C.B. (2016) Trop. Anim. Health Prod. 48(8):1569-1575. https://doi.org/10.1007/s11250-016-1129-z.

Ngeno K., Van der Waaij E.H., Megens H.J., Kahi A.K., Van Arendonk J.A.M., *et al.* (2015) Anim. Genet. Resour. Inf. 56:1-7. https://doi.org/10.1017/s2078633614000484.

Sun Y., Biscarini F., Bovenhuis H., Parmentier H.K., and Van der Poel J.J. (2013) Anim. Genet. 44(4):413-424. https://doi.org/10.1111/age.12014.

Van der Klein S.A.S., Berghof T.V.L., Arts J.A.J., Parmentier H.K., Van der Poel J.J., *et al.* (2015) Poult. Sci. 94(5):875–882. https://doi.org/10.3382/ps/pev052.

Van der Most P.J., de Jong B., Parmentier H.K., and Verhulst S. (2011) Funct. Ecol. 25(1):74-80. https://doi.org/10.1111/j.1365-2435.2010.01800.x.

Yuan J., Dou T., Ma M., Yi G., Chen S., *et al.* (2015) Poult. Sci. 94(7):1470-1475. https://doi.org/10.3382/ps/pev122.

73. Novel genetic parameters to improve gRFI in dairy cattle using big data from multiple lactations and countries

R.B. Stephansen[1*], P. Martin[2], B. Gredler-Grandl[3], C.I.V. Manzanilla Pech[1], G. Sahana[1], V. Ducrocq[2], C. Baes[4,5], F. Schenkel[4], R.F. Veerkamp[3] and J. Lassen[1,6]

[1]QGG, Aarhus University, Blichers Alle 20, 8830 Tjele, Denmark; [2]Université Paris-Saclay, INRAE, AgroParisTech, UMR GABI, 78350 Jouy-en-Josas, France; [3]ABG, Wageningen University&Research, P.O. Box 338, 6700 AH Wageningen, the Netherlands; [4]CGIL, University of Guelph, Guelph, ON, N1G 2W1, Canada; [5]Vetsuisse Faculty, University of Bern, 3012 Bern, Switzerland; [6]VikingGenetics, Ebeltoftvej 16, Assentoft, 8960 Randers, Denmark; rasmus.stephansen@qgg.au.dk

Abstract

To optimize breeding for feed efficiency, dry matter intake data from 3,967 Holstein cows with 131,234 weekly records from Denmark, France, the Netherlands, and Canada were combined. A random regression model was fitted to 1st and 2nd parity data for dry matter intake, energy corrected milk and body weight. Results showed that genetic residual feed intake is heritable in both parities, with an average heritability of 0.21. Genetic correlations across lactation and parities showed moderate to high estimates between mid and late lactation within parity. Early lactation had low genetic correlation with mid and late lactation. The genetic correlations between parities were around zero at all lactation stages. These findings show the importance of data collection and modelling at all lactation stages and parities.

Introduction

Genetic improvement of feed efficiency (FE) is a hot research topic to increase the environmental and economical sustainability in the dairy cattle industry. However, expensive recording schemes for large-scale feed intake registration and appropriate trait definition for FE in the breeding goal are main limitations to implement effective genetic evaluation of FE. We address the trait definition by studying genetic parameters for genetic residual feed intake (gRFI) and lack of feed intake data by international collaboration. Most evaluation centers use the Feed saved definition (Pryce *et al.*, 2015), where residual feed intake (RFI) is an important part. Residual feed intake is typically seen as an output of phenotypic regression of an animal's feed intake on energy sinks (yield, maintenance, etc.). Kennedy *et al.* (1993) proposed gRFI, assuming that the genetic correlations between feed intake and energy sinks are zero, which is important to avoid double counting in breeding goals. To overcome limited data size, collaboration on feed intake data around the world is crucial to achieve genetic progress for the economical important gRFI trait. Martin *et al.* (2021b) proposed a RFI model that was able to better describe the dynamic nature of the lactation, but this model did not include a genetic part. The aim of this study was to estimate variance components for a dynamic RFI model using time series data from multiple lactations and countries.

Materials & methods

Phenotypes, data editing & pedigree. Data from 870 Danish Holstein cows were collected from 2003-2019, at the Danish Cattle Research Center (DKC; Foulum, Denmark). Feeding data were described by Martin *et al.* (2021b). Data from 329 French Holstein cows were collected from 2014-2020, at the Le Pin and Méjusseaume research herds (Martin *et al.*, 2021a; Lefebvre *et al.* submitted to WCGALP 2022). Data from 1,953 Dutch Holstein cows were collected from 1991-2019, in different research herds at Wageningen University (Heida *et al.*, 2021). The Danish, French and Dutch data were shared as a part of the Gentore project (https://www.gentore.eu). Additionally, data from 815 Canadian Holstein cows were collected from 2015-2020, at the Sunalta, DRTC and Elora research farms (https://genomedairy.ualberta.ca/; http://www.resilientdairy.ca/).

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_73, © R.B. Stephansen *et al.* 2022

The combined dataset consisted of 5,470 lactations and 131,234 weekly records for feed intake. Various editing and plausibility checks were undertaken to ensure high-quality data, and requirements to cohesive lactations with a length of minimum 100 days in milk. All traits had their variance standardized to the average variance across herd, year, parity, and week of lactation. The pedigree obtained from the collaborators was pruned for eight generations. Missing parents were assigned to genetic groups with effects of country, breed, and birth year. Data editing were performed in the SAS software version 9.4.

Statistical analysis. The statistical software DMU (Madsen and Jensen, 2013) was used for variance component estimation, using the AI-REML algorithm. The random regression models were as follows:

$$y_{ijclmnop} = \mu + HYS_i + WOL_j + \beta_1 ACC_c + \beta_2 ACC_c^2 + \sum_{l=0}^{m} a_{cl}\Phi_{cjl} + \sum_{l=0}^{m} pe_{cl}\Phi_{cjl} + EXP_n + YM_o + e_{ijclnop}$$

where y were the phenotype DMI, energy corrected milk (ECM; Sjaunja (1990)) or BW for cow c on week of lactation j (j=1, 2, …, 44), μ were the intercepts, *HYS* were the fixed effects of herd × year × season i (670 levels; season were defined as quarters from date of calving), *WOL* were the fixed effects for week of lactation j, β_1 and β_2 were the regressions on age at calving (*ACC)* and ACC^2 for cow c respectively, a were the random additive effects with the m^{th} order of Legendre polynomials (LP), pe were the random permanent environmental effects with the m^{th} order of LP, *EXP* were the random effects of trial n (376 levels), *YM* were the random effects of year × month defined from the record date o (320 levels) and e were the residuals p (11 levels, one every 4 weeks assuming residual covariance to be 0). It is assumed that var(a) = A⊗m, var(pe) = I⊗m and var(e) = $I\sigma_e^2$ = R, where A is the numerator relationship matrix, ⊗ is the Kronecker product and m is the order of LP fitted for a and pe effects (m was 1st order for DMI and ECM, and 2nd order for BW). To derive genetic and phenotypic output, the notation by Islam *et al.* (2020) and Martin *et al.* (2021b) was followed. The final variance component matrix was made positive definite using the bending method of Schaeffer (2014).

Results

The descriptive statistics show small differences in mean and variance levels between the countries (Table 1).

Genetic RFI is heritable (avg. 1st parity: 0.18, 2nd parity: 0.24) with additive variance ranging between 0.2 and 2.8 in 1st parity and between 0.6 and 3.5 in 2nd parity. The highest levels of additive variance were seen during early and late lactation. The genetic correlations across lactation are presented in Figure 1. For 1st and 2nd parity, gRFI in early lactation was not correlated to gRFI in mid and late lactation. The genetic correlation between gRFI in mid and late lactation was moderate to high. The genetic correlation between 1st and 2 parity was around zero.

Table 1. Descriptive statistics for phenotypes by the different countries.

	Denmark	France	The Netherlands	Canada
No. DMI records	65,793	14,370	52,869	26,741
Avg. DMI (SD)	21.5 (3.7)	21.2 (3.7)	20.7 (3.8)	22.1 (5.0)
Avg. ECM (SD)	34.3 (7.8)	30.4 (6.7)	31.3 (7.2)	35.3 (7.5)
Avg. BW (SD)	653.4 (73.3)	626.4 (71.8)	593.3 (76.8)	667.7 (80.0)
Avg. WOL (SD)	20.7 (12.0)	20.0 (11.7)	17.0 (10.5)	20.8 (11.8)
Avg. Parity (SD)	1.7 (0.8)	1.5 (0.7)	1.6 (0.8)	1.6 (0.7)

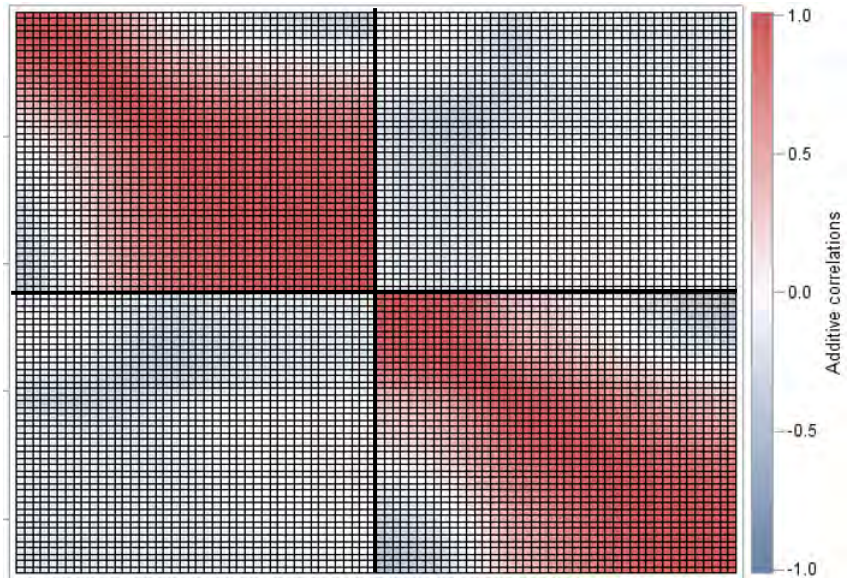


Figure 1. Genetic correlations across lactation in the first two parities. 1st parity is seen in upper left box and 2nd parity in lower right box. The genetic correlations between parities are in the off-diagonal boxes. Each pixel represents one week of lactation, with a total of 88 weeks of lactation when combining 1st and 2nd parity.

Discussion

Within parity, it is expected that gRFI in early lactation is different from the rest of the lactation, because of mobilization not properly accounted for. This was also seen in Li *et al.* (2017) using Danish data from 1st parity Holstein cows.

Absence of correlation between 1st and 2nd parity could be expected in early lactation due to higher mobilization of body reserve in 2nd parity animals compared with 1st parity; however, a higher correlation could be expected between parities in mid and late lactation. This could be an artifact of bending the variance component matrix. These results suggest it is important to register phenotypes for DMI, ECM and BW through the whole lactation and in multiple lactations. This is necessary to achieve the highest possible genetic improvement for feed efficiency. Furthermore, it is important to avoid that selection for gRFI leads to an increased level of metabolic disorders in dairy cattle. The low correlation between early lactation and the rest of the lactation indicates that selection for gRFI possibly leads to an increased degree of mobilization. Future research should focus on ensuring that selection for gRFI do not lead to increased mobilization in dairy cattle. A possibility is to include body condition scores in the gRFI equation, to account for the change in adipose tissue.

Conclusions

We demonstrated that gRFI is heritable and the genetic correlation structure shows that within parity, mid and late lactations are moderately to highly correlated. However, the genetic correlation between parities is around zero at all lactation stages. These results suggest that collection of data in all lactation stages and parities is important for genetic gain in feed efficiency.

Acknowledgements

The research leading to these results has received funding from European Union's Horizon 2020 research and innovation programme – GenTORE – under grant agreement No 727213. The authors would also like to acknowledge the efficient dairy genome project (https://genomedairy.ualberta.ca/) and the resilient dairy genome project (http://www.resilientdairy.ca). The authors thank Prof. Just Jensen and Per Madsen, QGG, Aarhus University for their guidance and valuable advices in the analysis part.

References

Heida M., Schopen G.C., Te Pas M.F., Gredler-Grandl B., and Verkamp R.F. (2021) J. Dairy Sci. 104(8): 8966-8982. https://doi.org/10.3168/jds.2020-19533

Islam M., Jensen J., Løvendahl P., Karlskov-Mortensen P., and Shirali M. (2020) J. Dairy Sci. 103(10): 9150-9166. https://doi.org/10.3168/jds.2019-17137

Kennedy B.W., van der Werf J.H., and Meuwissen T.H. (1993) J. Ani. Sci. 71(12):3239-3250. https://doi.org/10.2527/1993.71123239x

Lefebvre R., Faverdin P., Barbey S., Jurquet J. and Tribout T. *et al.* (2022) WCGALP 2022

Li B., Berglund B., Fikse W.F., Lassen J., and Lidauer M.H. *et al.* (2017) J. Dairy Sci. 100(11):9076-9084. https://doi.org/10.3168/jds.2017-12775

Madsen P., and Jensen J. (2013) Aarhus Res Cent. Foulum Box 50, 8830 Tjele Denmark.

Martin P., Ducrocq V., Fischer A., and Friggens N.C. (2021a) Animal 15(12):100412. https://doi.org/10.1016/j.animal.2021.100412

Martin P., Ducrocq V., Gordo D.G.M., and Friggens N.C. (2021b) Animal 15(2):100101. https://doi.org/10.1016/j.animal.2020.100101

Pryce J.E., Gonzalez-Recio O., Nieuwhof G., Wales W.J., and Coffey M.P. *et al* (2015) J. Dairy Sci. 98(10):7340-7350. https://doi.org/10.3168/jds.2015-9621

Schaeffer, L. 2014. Making covariance matrices positive definite. Available at: https://animalbiosciences.uoguelph.ca/~lrs/ELARES/PDforce.pdf

Sjaunja, L. (1990) Proc. of the 27th Session International Committee for Recording and Productivity of Milk Animals; 2-6 July 1990, Paris, France.

74. Genetic parameters of feeding behaviour traits in parental lines of a mule duck

H. Chapuis[1*], M. Lague[2] and H. Gilbert[1]

[1]GenPhySE, Université de Toulouse INRAE ENVT, 31326 Castanet Tolosan, France; [2]UEPFG INRAE, 1076 route de Haut-Mauco, 40280 Benquet, France; herve.chapuis@inrae.fr

Abstract

Automatic feeders were used to characterize the feeding behaviour traits of two populations of ducks (White Pekin and Muscovy). In addition to feed conversion ratio, six phenotypes were derived at the scale of the day and the meal and genetic parameters were estimated. Feeding rate in the Pekin population was twice as large as in the Muscovy population (19 vs 9 g/min). Feeding traits exhibited moderate (0.38) to high (0.67) heritabilities. A similar pattern of correlations could be observed between these traits in the two populations. Feed conversion ratio did not show any remarkable correlation with feeding traits in the Pekin line, except with daily feed intake, but was negatively correlated with the number of meals and meal feed intake in the Muscovy population (0.68). Further studies will be needed to decipher how these traits can be incorporated in breeding programs without impairing ducks' abilities to produce fatty liver.

Introduction

The mule duck contributes to more than 90% of the production of French fatty liver. It is a sterile hybrid obtained by crossing common duck females (*Anas platyrynchos*) and Muscovy drakes (*Cairina moschata*). Selection to improve fatty liver production is carried out on its parental populations, and only limited knowledge is available about the relationships between parental and crossbred traits. In addition, as recording individual feed intake in both parental populations is still a challenge, feed efficiency remains poor. We set up a design of purebred ducks to investigate detailed feeding traits to improve feed efficiency. We computed feeding behaviour traits using automated feeders (Bley and Bessei, 2008) and estimated genetic parameters for these traits in the parental populations, and their relationships with feed efficiency.

Materials & methods

The present study was conducted in agreement with the 2010/63/EU regulation for use of animals for research purposes. Animals were bred at the INRAE Duck farm (UEPFG, Benquet, France) which has been approved for animal experimentation (C40-037-1).

Birds and housing. Three successive generations of animals were produced in a dam White Pekin (PEK) common duck line and in a sire Muscovy (MUS) line. Electronically tagged animals were reared up to the age of 7 (PEK) or 8 (MUS) weeks in a 96 m^2 pen equipped with 6 automatic feeders similar to those used by Basso *et al.* (2014), including animal weighing at each visit. To facilitate duck management, the area was divided in two parts with a barrier, so each bird had effective *ad libitum* access to 3 feeders. In the PEK line, we reared 273 males and 467 females, while only males (n=460) were studied in the MUS line.

Data acquisition. The date and time at the beginning and end of each visit to the feeder was recorded. Animal weight and feed consumption were recorded for each visit. Feeding behaviour traits were then computed. Daily Feed Intake (DFI) and Daily Feeding Time (DFT) definitions are straightforward. Feeding Rate (FR) was computed as the ratio of feed intake over feeding time for each visit and averaged over the whole period. Following Howie *et al.*'s (2009) method 3 to compute the so-called 'meal criterion', visits were aggregated into meals in order to allow for reliable comparisons of feeding behaviours between the two populations. For each duck, visits were defined as part of a single meal if the interval between visits was

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_74, © H. Chapuis *et al.* 2022

lower than the meal criterion (1,616 s in PEK, 2,113 s in MUS). We therefore recorded the number of meals per day (NM), Meal Duration (MD) and Meal Feed Intake (MFI). Feed efficiency was addressed through computation of feed conversion ratio (FCR), as the ratio of total feed intake over the observed weight gain. Average Daily Gain was also computed. We planned to collect and analyse early feed and growth data over a 5-week span (i.e. 35 days). In order to limit the occurrence of misidentified visits, which were found numerous at early age in the Muscovy population (Cobo *et al.*, 2017), the 35-day control period spanned from Day22 to Day56 in the MUS line, while it started at Day15 and ended at Day49 in the PEK line.

Statistical analysis. The phenotypic comparison of performances between lines was obtained through a linear model accounting for the effects of batch and sex in the PEK line. Then, genetic parameters were estimated using ASRemL (Gilmour *et al.*, 2015), with the model:

$$y = \mu + \beta_{batch} (+ \gamma_{sex}) + u + e \quad (1)$$

where y is the phenotype, μ the general mean, β_{batch} and γ_{sex} are the fixed effects of batch and sex (PEK line only), respectively, u the additive genetic effects, and e the residual effects. In order to normalize the data, DFT and MD were log-transformed prior to analyses. Values more than three standard deviations from the mean were considered as outliers and removed from the analysis. First, univariate analyses were run to compute heritabilities, and next multivariate models were applied with the same effects to obtain correlations.

Results

The least squares means of the linear model applied for phenotypic line comparisons are given in Table 1. MUS ducks spent daily almost twice as much time feeding as PEK ducks, while the amount of feed they consumed was 12% lower. This resulted in a feeding rate more than twice as great for the PEK ducks, as reported in Cobo *et al.* (2017) where only male data were considered. Compared to the MUS line, the number of meals per day was greater and meal duration was roughly 2.5 times shorter in the PEK line (Table 1). FCR was better (i.e. lower) in the Muscovy line. The average daily gain was greater in the MUS line. Yet, it should be remembered that the cut-off points occurred one week earlier in the PEK population.

The genetic parameters are displayed in Table 2. In both lines, feeding traits were highly heritable; most heritabilities were higher than 0.5, such as NM and MFI in both populations. DFT and FR exhibited remarkably high heritability values (~0.6) in the PEK line and around 0.4 in the MUS line.

Table 1. Mean values (least squares means[1] ± standard error) of feeding behaviour traits.

Trait	PEKIN dam line	MUSCOVY sire line	*P* line
Daily feed intake (g/d)	225.00±0.72	200.00±1.16	<0.0001
Daily feeding time (s/d)	811.0±12.1	1,595.0±19.5	<0.0001
Feeding rate (g/min)	18.84±0.15	8.86±0.24	<0.0001
Number of daily meals/d	4.46±0.03	3.54±0.05	<0.0001
Meal feed intake (g)	53.2±0.5	59.1±0.8	<0.0001
Meal duration (s)	191.00±3.98	470.00±6.41	<0.0001
Feed conversion ratio (kg/kg)	3.25±0.01	2.56±0.01	<0.0001
Average daily gain (g/d)	69.1±0.24	79.7±0.38	<0.0001

[1] Least squares means from a linear model including the fixed effect of batch (PEK and MUS) and sex (MUS).

In PEK, DFI was moderately correlated with MFI and FCR (~0.35), while these correlations were much higher in MUS (~0.70). DFI and MD were positively but not significantly correlated in PEK, while the correlation was positive in MUS. In both lines, DFT and FR were strongly and negatively correlated.

Discussion

The lower FCR in the MUS line compared to the PEK line was expected, as Muscovy duck is known to produce a leaner meat than the White Pekin and is popular in France for this very reason. Here we studied only males in the MUS population, which is known to exhibit a larger sexual dimorphism for body weight than the PEK line (Tai and Rouvier, 1998). This may contribute to the difference in FCR observed between lines.

We computed a meal criterion within each batch using Howie *et al.*'s (2009) method 3. This method is based on analyses of the probability for an animal to start feeding given the time since the last visit. This requires finding the nadir of a potentially flat curve, which may be not precisely estimated. For the PEK line, the meal criterion was close to the value of 1,725 seconds found by Howie *et al.* (2009) in another Pekin duck population.

Our genetic parameters for feeding behaviour traits were in good agreement with those obtained on Pekin broilers by Le Mignon *et al.* (2017) and Thiele (2017). With a heritability of 0.33±0.11 for feed intake from Day15 to Day42, Zhang *et al.* (2017) obtained a slightly lower value than ours. Some common features can be noted between our study and that of Le Mignon *et al.* (2017): (1) Strong negative correlations between the number of meals and the meal size, reflecting that an increase in the meal number does not lead to increased feed intake. If the animal has reached satiety, an increase in the number of meals does not respond to an urge to eat more feed but simply illustrates a variability in feeding behaviour; (2) Moderately negative

Table 2. Genetic parameters of feeding traits[1] in two duck lines.

	DFI	DFT	FR	NM	MFI	MD	FCR
PEKIN							
DFI	**0.49±0.08**	0.13±0.13	0.12±0.13	-0.15±0.13	0.36±0.12	0.24±0.14	0.33±0.12
DFT		**0.57±0.07**	-0.95±0.01	0.44±0.10	-0.41±0.11	0.61±0.08	-0.05±0.13
FR			**0.61±0.07**	-0.43±0.10	0.48±0.09	-0.57±0.08	0.06±0.13
NM				**0.57±0.07**	-0.95±0.08	-0.42±0.08	-0.06±0.08
MFI					**0.63±0.07**	0.44±0.08	0.10±0.08
MD						**0.48±0.07**	0.01±0.08
FCR							**0.48±0.07**
MUSCOVY							
DFI	**0.54±0.12**	0.23±0.20	0.18±0.20	-0.20±0.18	0.69±0.10	0.50±0.19	0.73±0.15
DFT		**0.43±0.11**	-0.88±0.05	0.52±0.17	-0.10±0.20	0.68±0.12	0.32±0.24
FR			**0.39±0.11**	-0.49±0.18	0.41±0.18	-0.60±0.14	-0.06±0.26
NM				**0.54±0.11**	-0.90±0.04	-0.31±0.18	-0.44±0.21
MFI					**0.67±0.11**	0.44±0.16	0.68±0.16
MD						**0.38±0.11**	0.01±0.08
FCR							**0.25±0.09**

[1] DFI = daily feed intake; DFT = daily feeding time; FR = feeding rate; NM = number of daily meals; MFI = meal feed intake; MD = meal duration; FCR = feed conversion ratio.

correlations between number of meals and feeding rate. Birds exhibiting a large FR may swiftly reach satiety and not feel the need to start another feeding sequence; (3) A large and positive, but significantly different from unity, correlation between meal duration and total time spent feeding, emphasizing an exploitable variability of feeding behaviours. As expected, the latter is also strongly negatively correlated with FR. These three points are also common to our two populations, which might suggest some generality about ducks' feeding behaviours.

For feed efficiency, we focused on FCR. Feed efficiency could also be profitably addressed through the fraction of feed intake that is not explained by maintenance and production requirements, *aka* Residual Feed Intake (RFI). Yet, in the case of fatty duck production, attention should be paid to an adequate derivation of RFI, i.e. accounting for body lipid composition, as in Drouilhet *et al.* (2014). With estimates of 0.24 and 0.31, FCR was less heritable in the two Pekin lines studied by Le Mignon *et al.* (2017) than in our study. In the PEK line, our estimates of the correlation between FCR and feeding behaviour traits are in accordance with those obtained by Thiele (2017) and Le Mignon *et al.* (2017): they are of low magnitude (<0.20), except for DFI. Conversely, in the MUS line, FCR exhibited strong correlations with DFI and MFI, which may suggest a larger part of feed allocated to growth compared with the PEK line. In addition, in MUS, the negative correlation between NM and FCR, though not accurately estimated, could indicate that selecting for increased number of meals would improve FCR, which contradicts the literature in which efficient animals generally show less feeding activity.

These first genetic analyses of a large design focusing on feeding characteristics of the parental lines of the mule duck show that feeding traits are heritable in these lines, and some are correlated to feed efficiency. Further studies will be conducted to decipher how these traits can be incorporated in breeding programs without impairing the ducks' abilities to produce fatty liver.

This study was funded by the French interprofessional association for fatty liver production (CIFOG), and INRAE animal genetic division. The authors thank UEPFG technical staff for animal care and data collection.

References

Basso B., Lagüe M., Guy G., Ricard E. and Marie-Etancelin C. (2014) J. Anim. Sci. 92:1639-1946. https://doi.org/10.2527/jas.2013-7110

Bley T.A.G. and Bessei W. (2008) Poult. Sci. 87:215-221. https://doi.org/10.3382/ps.2006-00446

Cobo E., Lagüe M., Cornuez A., Bernadet M.-D., Martin X. *et al.* (2017) Proc. of « Douzièmes Journées de la Recherche Avicole et Palmipèdes à Foie Gras », Tours, France.

Drouilhet L, Basso B., Bernadet M.D., Cornuez A., Bodin L. *et al.* (2014) J. Anim. Sci. 92:4287-4296. https://doi.org/10.2527/jas.2014-8064

Gilmour, A.R., Gogel, B.J., Cullis, B. R., Welham, S.J. and Thompson, R. (2015). ASReml

User Guide Release 4.1 Functional Specification, VSN International Ltd, Hemel Hempstead,

Howie J.A., Tolkamp B.J. Avendaño S. and Kyriazakis I. (2009) Appl. Anim. Behav. Sci. 116:101-109 https://dois.org/10.1016/j.applanim.2008.09.005

Le Mignon G., Blanchet M. and Chapuis H. (2017) Proc of Xth European Symposium on Poultry Genetics, Saint-Malo, France.

Tai C. and Rouvier R. (1998) Genet. Sel. Evol. 30:163-170

Thiele H.H. (2017) available at Feed Efficiency and Feeding Behaviour in Pekin Ducks - Lohmann Breeders (lohmann-breeders.com)

Zhang Y., Guo Z.B., Xie M., Zhang Z. and Hou S. (2017) Asian-Australas J. Anim. Sci 30:167-170. https://doi.org/10.5713/ajas.15.0577

75. Estimates of genetic parameters for feeding behavior traits and its association with feed efficiency in Holstein cows

L. Cavani[1*], W.E. Brown[1], K.L. Parker Gaddis[2], R.J. Tempelman[3], M.J. VandeHaar[3], H.M. White[1], F. Peñagaricano[1] and K.A. Weigel[1]

[1]University of Wisconsin, Department of Animal and Dairy Sciences, 1675 Observatory Drive, 53706, Madison, WI, USA; [2]Council on Dairy Cattle Breeding, 4201 Northview Drive, 20716, Bowie, MD, USA; [3]Michigan State University, Department of Animal Science, 474 S. Shaw Lane, 48824, East Lansing, MI, USA; cavani@wisc.edu

Abstract

The objective of this study was to estimate genetic parameters for feeding behavior and the genetic correlations with feed efficiency traits in Holstein cows. Data consisted of 75,877 daily feeding behavior records of 1,328 Holstein cows in 31 different experiments conducted from 2009 to 2020. The statistical model included lactation and days in milk as fixed effects, and experiment-treatment, animal, and permanent environment as random effects. Estimates of heritability ranged from 0.09 to 0.23 and repeatability from 0.23 to 0.52 for feeding behavior, with strong genetic correlations between traits. Feeding rate and intake per visit or meal showed positive and strong genetic correlations with residual feed intake, dry matter intake, and metabolic body weight. We conclude that cows that eat at a slower rate may be more feed efficient. Measures of feeding behavior could be useful indicators of dairy cow feed efficiency.

Introduction

Genetic selection is a powerful tool to improve livestock production, given that genetic gains are cumulative and permanent. Genomic technologies allow the inclusion of important economic traits that are difficult to measure into breeding programs. Automated feeders, for example, enable the evaluation of feed efficiency on an individual animal basis. In addition, individual measures of feeding behavior can be assessed using this technology. Interestingly, there is evidence that variation in feeding behavior explains part of the variation observed in residual feed intake (RFI) in dairy cattle (Connor *et al.* 2013; Green *et al.* 2013; Lin *et al.* 2013). Although this is a reliable system, it is too expensive for commercial farms. Therefore, feeding behavior may be used as an indicator trait for feed efficiency, using less expensive wearable sensors or computer vision systems that do not actually measure individual intakes. As such, the objective of this study was to estimate genetic parameters for feeding behavior traits and genetic correlations of feeding behavior with milk energy, metabolic body weight, change in body weight, dry matter intake (DMI), and RFI in US Holstein cows.

Materials & methods

Data consisted of 75,877 feeding behavior records of 1,328 (546 primiparous and 782 multiparous) Holstein cows in 31 experiments from 2009 to 2020 in a freestall facility at University of Wisconsin-Madison with a roughage intake control system (RIC; Hokofarm Group, Marknesse, Netherlands). Cows were provided *ab libitum* access to feed and water, with the diets formulated to meet or exceed nutrient requirements. The RIC system recorded, for each cow, the start and end time of each visit to the feeder and the weight of feed consumed during each visit. Using the feeding behavior variables, we also developed a meal concept. The meal criterion was calculated by fitting a mixture model of the $\log_{10}$ frequency distribution of the intervals between feeder visits (DeVries *et al.* 2003; Horvath and Miller-Cushon 2019). Any interval between feeder visits less than the meal criterion was defined as the same meal. The meal criterion derived from this dataset was 26.4 minutes. Descriptive statistics for daily records of feeding behavior traits and DMI are shown in Table 1. In addition, weekly records of the three energy sinks, namely milk energy, metabolic body weight and change in body weight, and DMI were assessed to calculate weekly RFI values. In addition to phenotypic records, pedigree information for the last three generations was considered in the analyses.

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_75, © L. Cavani *et al.* 2022

Genetic parameter estimation was performed in bivariate analyses using the following model:

$$y=X\beta+ Z_1b+ Z_2u+Wpe+e,$$

where y is a vector of phenotypic records, β is a vector of fixed effects, b is a vector of random block effects of experiment-treatment (60 levels), u is a vector of random additive genetic effects, pe is a vector of random permanent environmental effects, and e is the vector of random residual effects. Fixed effects included lactation number with 4 levels (1 to 4+) and days in milk with 9 levels (grouped by 19 days). Matrices X, Z_1, Z_2, and W are incidence matrices relating y to β, b, u, and pe, respectively. Estimates were obtained using restricted maximum likelihood method through the AIREMLF90 software (Aguilar *et al.* 2018).

Results

Descriptive statistics for daily records of feeding behavior traits, daily DMI, and their heritability and repeatability estimates are shown in Table 1. Heritability estimates for feeding behavior traits ranged from 0.09 to 0.23, and for dry matter intake was about 0.12. Most traits were highly repeatable across days, with repeatability estimates ranging from 0.23 to 0.52.

Genetic and phenotypic correlations for daily records are depicted in Figure 1. Standard errors (SE) of estimates ranged from 0.01 to 0.10 (data not shown). In general, feeding behavior measures were strongly genetically correlated. Dry matter intake showed positive and strong genetic correlations with feeding rate (0.71 and 0.68), and intake per visit and per meal (0.66 and 0.88).

Genetic and phenotypic correlations between feeding behavior traits and milk energy, metabolic body weight, change in body weight, dry matter intake, and RFI, measured as averages by week, are depicted in Figure 2. The SE of estimates ranged from 0.03 to 0.14 (data not shown). Our results showed that some feeding behavior traits were strongly genetically correlated with RFI, and/or with the traits considered in the equation to calculate RFI. Only body weight change was not significantly associated with feeding behavior.

Table 1. Descriptive statistics and estimates of heritability (h^2) and repeatability (r^2) for daily feeding behavior and feed dry matter (DM) intake in Holstein cows.

Traits	Mean	SD	$h^2 \pm SE$	$r^2 \pm SE$
Number of feeder visits per day	35.7	15.6	0.16±0.03	0.52±0.02
Number of meals per day	7.5	1.75	0.09±0.02	0.23±0.01
Duration of each feeder visit (minutes)	6.8	2.7	0.16±0.03	0.47±0.02
Duration of each meal (minutes)	29.4	8.9	0.14±0.02	0.31±0.01
Total duration of feeder visits per day (minutes)	212.4	55.4	0.16±0.03	0.38±0.02
Intake per visit (kg of DM)	0.9	0.4	0.16±0.03	0.49±0.02
Intake per meal (kg of DM)	3.8	1.1	0.13±0.02	0.28±0.01
Feeding rate per visit (kg of DM per minute)	0.15	0.06	0.11±0.02	0.32±0.02
Feeding rate per meal (kg of DM per minute)	0.14	0.04	0.23±0.03	0.49±0.02
Dry matter intake (kg/d)	21.9	5	0.12±0.02	0.31±0.01

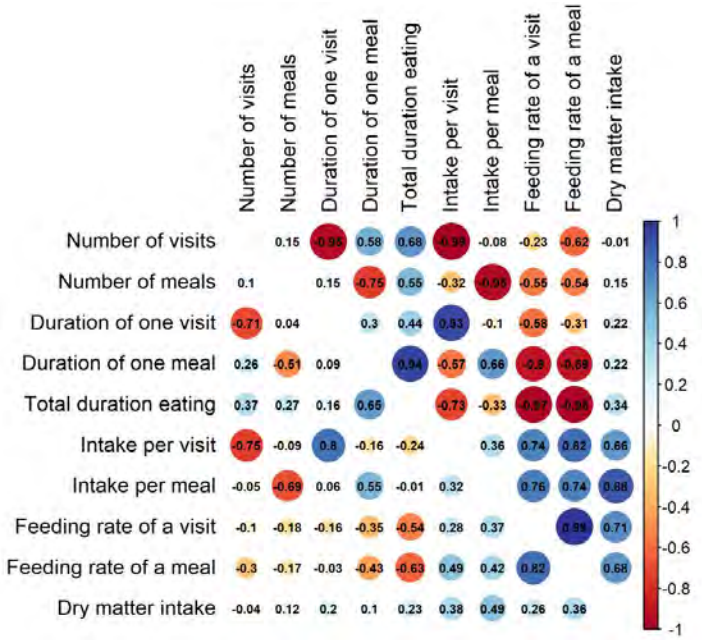


Figure 1. Genetic correlations (above diagonal) and phenotypic correlations (below diagonal) between daily feeding behavior traits and DMI in lactating Holstein cows.

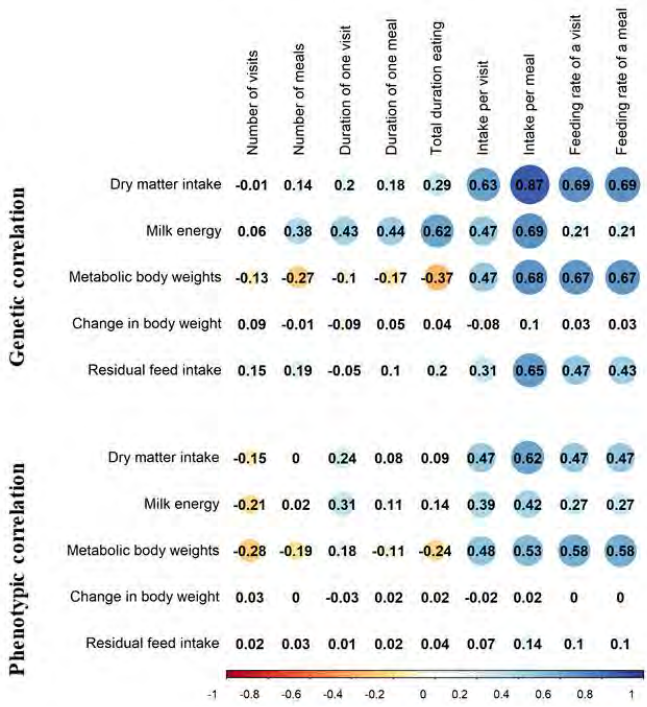


Figure 2. Genetic correlations and phenotypic correlations between weekly averages for feeding behavior and feed efficiency in lactating Holstein cows.

Discussion

Heritability estimates indicated that a significant proportion of phenotypic variance is explained by additive genetic effects, such that feeding behavior could be incorporated into genetic selection schemes. Higher heritability estimates have been reported for feeding behavior in dairy and beef cattle, but those studies used weekly averages (or longer), rather than daily records (Benfica *et al.* 2020.; Lin *et al.* 2013; Olson *et al.* 2020). Genetic correlations among feeding behavior traits showed that, with more visits or more meals per day, cows spend less time at each feeder or meal, with a lower intake per visit or meal. Genetic correlations between feeding rate and RFI (0.47 and 0.43) suggest that selection of animals that eat at a slower rate will also tend to select efficient animals with lower RFI. Feeding rate was strongly associated with DMI and metabolic body weight, both genetically and phenotypically. The same pattern was observed for intake per visit and intake per meal, indicating that cows with greater DMI per visit or meal tend to be less efficient genetically (greater RFI, milk energy, metabolic body weight, and total DMI). Studies reporting genetic correlations between feeding behavior and RFI are limited in lactating dairy cows, though Lin *et al.* (2013) observed a moderate genetic correlation between feeding duration and RFI (0.27) and a negligible correlation between feeding rate and RFI in dairy heifers. Overall, the results of this study suggest that dairy cows with slower feeding rates and smaller intakes per visit or meal are more feed efficient (lower RFI). Therefore, feeding behavior could be implemented into genetic selection programs as indicators of feed efficiency, which would increase selection accuracy for RFI.

References

Aguilar I., Tsuruta S., Masuda Y., Lourenco, D.A.L., Legarra A. *et al.* (2018) Proc. of the 11th WCGALP, Auckland, New Zealand.

Benfica L.F., Sakamoto L.S., Magalhaes A.F.B., Oliveira, M.H.V., Albuquerque L.G. *et al.* (2020) J. Anim. Sci. 98(11):1-9. https://doi.org/10.1093/jas/skaa350

Connor E.E., Hutchison J.L., Norman H.D., Olson K.M., Van Tassell C.P. *et al.* (2013) J. Anim. Sci. 91:3978-3989. https://doi.org/10.2527/jas.2012-5977

DeVries T.J., von Keyserlingk M.A.G., Weary D.M., and Beauchemin K.A. (2003) J. Dairy Sci. 86:3354-3361. https://doi.org/10.3168/jds.S0022-0302(03)73938-1

Green T.C., Jago J.G., Macdonald K.A., and Waghorn G.C. (2013) J. Dairy Sci. 96:3098-3107. https://doi.org/10.3168/jds.2012-6087

Horvath K.C., and Miller-Cushon E.K. (2019) J. Dairy Sci. 102:3421-3430. https://doi.org/10.3168/jds.2018-15460

Lin Z., Macleod I., and Pryce JE (2013) J. Dairy Sci. 96:2654-2656. http://dx.doi.org/10.3168/jds.2012-6134

Olson C.A., Li C., Block H., McKeown L., Basarab J.A. (2020) J. Anim. Breed. Genet. 138:300-313. https://doi.org/10.1111/jbg.12522

76. Protein efficiency is heritable but not genetically correlated with meat quality in pigs

E.O. Ewaoluwagbemiga[1,3*], G. Bee[2] and C. Kasper[1]

[1]Animal GenoPhenomics Group, Agroscope Posieux 1725, Switzerland; [2]Swine Research Unit, Agroscope Posieux 1725, Switzerland; [3]Animal Genomics, ETH Zurich, Zurich, Switzerland; esther.ewaoluwagbemiga@agroscope.admin.ch

Abstract

The improvement of protein efficiency (PE) is an essential component in the development of a sustainable pig production. This study aims to estimate the heritability of PE when dietary protein is reduced, and to estimate its genetic correlations with meat quality and production traits. A total of 682 pigs were used in estimating the heritability of PE, and 512 pigs in estimating genetic correlations. The results from this study showed potential for selective breeding of PE in Large White pigs with a heritability of 0.40. Meat quality, except meat redness, is not expected to be influenced by genetic improvement of PE, as there were no significant genetic correlations between PE and meat quality traits. Favourable relationships exist between PE and feed conversion ratio, and with average daily feed intake, which means that protein-efficient pigs fed under a reduced-protein diet consume less feed and efficiently convert the feed to lean mass.

Introduction

In the past, the focus of animal breeding had been on improving production traits, but today, sustainability concerns are gradually gaining importance. An essential component to consider in the development of sustainable pig production is the reduction of nitrogen excretion in fattening pigs. The inefficient use of dietary protein by the animals results in the excretion of excess nitrogen, thereby contributing to pollution (Notarnicola *et al.*, 2017). Therefore, considering the environmental and ecological impact of pig production, it is essential to improve protein efficiency (PE) in pigs. In addition to nutritional strategies to improve PE, a genetic solution can be sought. Feed conversion ratio (FCR) and residual feed intake (RFI) are used to characterize feed efficiency. Although FCR and RFI are expected to correlate with PE, selection for improved FCR and RFI would likely increase energy efficiency rather than PE (Kasper *et al.*, 2020; Millet *et al.*, 2018) since energy intake is the main factor driving feed intake. Moreover, it has been shown that selecting for improved FCR and RFI with the aim of reducing nutrient excretion may not be as efficient as selecting for the nutrient trait itself (de Verdal *et al.*, 2011). In contrast to FCR and RFI, very few studies have investigated the possibility to improve PE, which may be due to a number of factors such as difficulties in phenotyping animals for this trait and the lack of approved and validated proxies. The heritability of PE reported in the studies of Saintilan *et al.* (2013) and Kasper *et al.* (2020) indicate the potential to select genetically for improved PE in pigs. However, knowledge of genetic correlations of PE with other traits of importance such as meat quality and growth performance traits is important to account for possible trade-offs. The aim of this study was therefore to estimate the heritability for PE and its genetic correlations with meat quality and growth performance traits.

Materials & methods

A total of 682 Swiss Large White pigs were used in this study, which all had *ad libitum* access to feed and water. Once a pig reached a body weight (BW) of ≥20 kg (average ± standard deviation: 22.5±1.6 kg), it was allocated to a pen where it remained until slaughter. The automatic feeder dispensed the grower diet to pigs until an average BW of 63±2 kg, and thereafter a finisher diet until an average BW of 106±5 kg when the pigs were slaughtered. The levels of digestible protein and essential amino acids for the grower and finisher diets were 20% lower than the recommended levels, with the same digestible energy. The automatic feeders

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_76, © E.O. Ewaoluwagbemiga *et al.* 2022

(Schauer Maschinenfabrik GmbH & Co. KG, Prambachkirchen, Austria) recorded all daily visits to the feeder and the quantity of feed consumed per visit for each animal.

Protein efficiency. After slaughter, the left carcass including the whole head was scanned with dual-energy X-ray (DXA) absorptiometry to determine lean tissue content, which was used in the following prediction equation to estimate the protein content retained in the carcass (Kasper *et al.*, 2021).

$$\text{Protein content carcass (g)} = -482.745 + 0.23\ (\text{g lean tissue DXA} \times 2) \quad (1)$$

PE of the carcass was calculated as the ratio of protein deposited in the carcass minus the protein content in the carcass at 20 kg BW to the total dietary protein intake. The protein content of the carcass at 20 kg BW was estimated based on data from the experiment of Ruiz-Ascacibar *et al.* (2017).

Other traits in the study. Loin muscle area and backfat thickness were assessed on the left carcass side at the 10th to 12th ribs' level. A picture of the loin muscle was taken and the area was measured. Backfat thickness was measured with a ruler. Intramuscular fat content was determined by the Soxhlet method. A 2 cm thick chop was cut from the loin muscle at the 10th to 12th ribs' level and freed from adhering adipose and connective tissues. After 20 min in the cooling room, CIE Lab-values (lightness, redness and yellowness), were recorded using the Minolta CR-300 colorimeter. Subsequently, the chops were weighed, suspended in sealed plastic bags at 2 °C for 48 h and re-weighed to determine the drip loss. The same chops were then vacuum-sealed and cooked in a water bath at 72 °C for 45 min, then cooled in cold water for 15 min, rinsed to remove coagulated sarcoplasmic protein, dabbed dry and weighed to determine cooking loss. The cooked chops were frozen at -20 °C and shear force was measured on four cores of 13 mm diameter using a Warner Bratzler Shear Force machine.

Statistical Analysis. Bayesian animal models were performed using MCMCglmm (Hadfield 2010) in R. The genetic and common environmental variance components of PE were estimated using a univariate animal model (corrected for the effects of sex, slaughter weight, slaughter age, and farrowing series) and genetic correlations were estimated using bivariate models. Both univariate and bivariate models were performed with a Markov Chain Monte Carlo algorithm, using the model formula

$$y=\mu+Xb+Za+Z_c c+e \quad (2)$$

where y is a vector of observations of PE, b is a vector of fixed effects of farrowing series, sex, slaughter weight and slaughter age. X is an incidence matrix relating records to fixed effects. a is a vector of random additive genetic effects, and Z is the corresponding incidence matrix. c is a vector of random litter effects, and Z_c is the corresponding incidence matrix. Heritability was thereafter computed as the ratio of genetic variance to the phenotypic variance ($h^2 = V_A / V_P$), where phenotypic variance (V_P) is the sum of genetic variance (V_A), litter (common environmental) effect (V_{CE}) and residual variance (V_R). The litter effect was calculated as the ratio of the litter effect to the phenotypic variance ($CE^2 = V_{CE}/V_P$). A non-informative prior with an inverse Wishart distribution for the random effects and normal distribution for the fixed effects were used. The model was run with 2 million iterations, 100,000 burn-in and 1000 thinning intervals.

Results

The posterior distributions of the variance decomposition of PE is presented in Figure 1 (posterior distributions of the respective variance components are shown in the upper part, points representing single estimates are shown together with a box plot, where whiskers represent the 95% credible interval). The PE of the carcass is clearly heritable (h^2=0.40, 95% credible interval = (0.22, 0.59)). The common environment

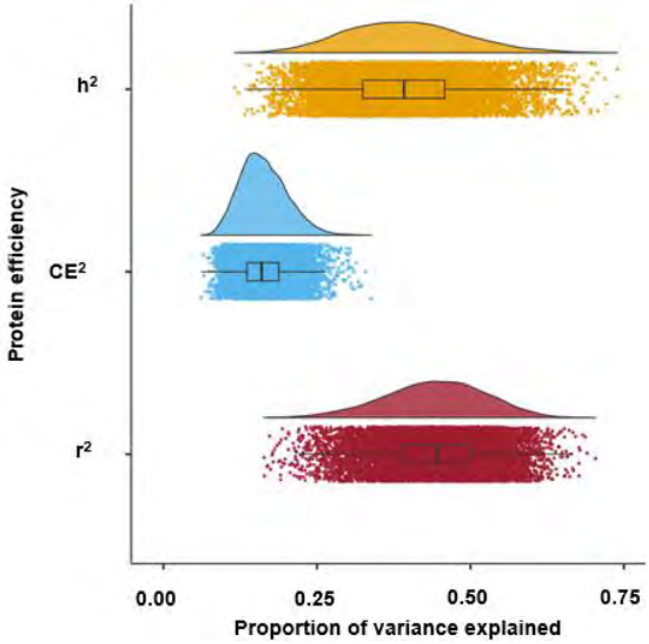


Figure 1. Posterior distributions of heritability (h^2), common environment effect (CE^2) and residual variance (r^2) of protein efficiency of the carcass.

(litter effect) also contributed to the phenotypic variation with a variance of 0.16 (0.09, 0.24), showing that pigs in the same litter performed similarly compared to others in different litters. The genetic correlations between PE and meat quality traits were generally non-significant with wide credible intervals that included zero (Table 1). However, meat redness trait showed a significant negative and moderate genetic correlation. There were high and negative genetic correlations between PE and production traits, with favourable relationships between PE and average daily feed intake and FCR (Table 1).

Table 1. Genetic correlations of protein efficiency with meat quality and production traits.

Group	Trait	r_g [1] [95% credible interval]
Meat quality traits	Loin muscle area	0.06 [-0.44, 0.60]
	Cooking loss	0.04 [-0.34, 0.39]
	Drip loss	0.03 [-0.17, 0.17]
	Shear force	0.05 [-0.20, 0.28]
	Intramuscular fat	-0.07 [-0.29, 0.16]
	Backfat thickness	-0.21 [-0.43, 0.08]
	Meat yellowness	-0.28 [-0.47, 0.03]
	Meat lightness	-0.42 [-0.88, 0.29]
	Meat redness	-0.35 [-0.59, -0.08]
Production traits	Average daily gain	-0.65 [-0.71, -0.58]
	Feed conversion ratio	-0.96 [-0.97, -0.95]
	Average daily feed intake	-0.93 [-0.95, -0-91]

[1] Genetic correlation.

Discussion

In this study, we showed that PE is heritable, and an improvement in PE when dietary protein is reduced is not expected to influence the quality of meat, except for the redness of meat. The favourable genetic correlations between PE and average daily feed intake (ADFI) and feed conversion ratio (FCR) show that protein-efficient pigs are expected to consume less feed and efficiently convert the feed to meat. This suggests that breeding for more protein-efficient, i.e. more sustainable pigs, could lower the cost of production, but special attention should be paid to the genetic correlation with growth rate to actually enable economic pig production. In addition, this could help decrease the competition between humans and livestock for plant-protein sources such as soybean. However, breeding for protein-efficient pigs, under a reduced dietary protein condition, may decrease average daily gain. The heritability estimate here was higher than that reported by Kasper *et al.* (2020), which may be due to higher sample size in the current study or to the experimental design. While all pigs in this study were slaughtered at 100 kg BW, pigs in the study of Kasper *et al.* (2020) were slaughtered between 40-100 kg BW, though differences in slaughter weight were corrected for statistically. A weak but significant effect of litter was observed on PE, in contrast to the study of Kasper *et al.* (2020), where the litter effect estimate was 0.002. This might be also a consequence of the larger sample size. Kasper *et al.* (2020) also reported the same relationship between PE and FCR. Although a strong negative (but favourable) genetic correlation was observed between PE and FCR, selection for FCR does not necessarily equate to selection for PE in terms of selection response. Unravelling this relationship should be the focus of future studies.

References

de Verdal H., Narcy A., Bastianelli D., Chapuis H., Même N., *et al.* (2011) BMC Genetics 12(1):1-10. https://doi.org/10.1186/1471-2156-12-71

Hadfield J.D. (2010) J. Stat. Softw. 33(2):1–22. https://www.jstatsoft.org/v33/i02/

Kasper C., Ruiz-Ascacibar I., Stoll P., and Bee G. (2020) J. Anim. Breed. Genet. 137(6):545-558. https://doi.org/10.1111/jbg.12472

Kasper C., Schlegel P., Ruiz-Ascacibar I., Stoll P., and Bee G. (2021) Animal 15(8):100307. https://doi.org/10.1016/j.animal.2021.100307

Millet S., Aluwé M., Van den Broeke A., Leen F., De Boever J., *et al.* (2018) Animal 12(5):1060-1067. https://doi.org/10.1017/S1751731117002610

Notarnicola B., Tassielli G., Renzulli P.A., Castellani V., and Sala S. (2017) J. Clean Prod. 140:753-765. https://doi.org/10.1016/j.jclepro.2016.06.080

Ruiz-Ascacibar I., Stoll P., Kreuzer M., Boillat V., Spring P., *et al.* (2017) Animal 11(3):394-404. https://doi.org/10.1017/S1751731116001634

Saintilan R., Merour I., Brossard L., Tribout T., Dourmad J.Y., *et al.* (2013) Anim. Sci. J. 6:2542-2554. https://doi.org/10.2527/jas.2012-5687

77. Genetic analysis of digestibility traits in pigs measured by near-infrared spectroscopy

K.H. Martinsen[1*], E.G. Enger[1], E. Grindflek[1], N.K. Afseth[2], L.T. Mydland[3], S. Wallen[3] and T.H.E. Meuwissen[3]

[1]Norsvin, Storhamargata 44, 2317 Hamar, Norway; [2]Norwegian Institute of Food, Fisheries and Aquaculture (Nofima), P.O. Box 210, 1432 Ås, Norway; [3]Norwegian University of Life Sciences, P.O Box 5003, 1432 Ås, Norway; kristine.martinsen@norsvin.no

Abstract

The aim of this study was to estimate genetic parameters for apparent total tract digestibility of nitrogen (ATTDn) and dry matter (ATTDdm) and to investigate their genetic relationship to other relevant production traits in pigs. Near infrared spectroscopy was used for prediction of nitrogen content in feces. The estimated heritabilities of ATTDdm and ATTDn were 0.22 and 0.20, respectively. The genetic correlations between ATTDdm and ATTDn and feed intake from 40 to 120 kg were -0.5 and -0.65, respectively. No significant genetic correlations were found between any of the digestibility traits and growth and carcass quality traits. The results suggested that the current selection strategy is improving the digestibility of nitrogen and dry matter, but that selection for improved digestibility would reduce the feed consumption from 40 to 120 kg.

Introduction

Protein is a costly nutrient in feed and efficient use of proteins from the feed is therefore important due to global issues such as increasing human population and need for proper nutrition, climate changes, greenhouse gas emissions and profit in pork production.

Less than 1% of the greenhouse gas emissions from Norway is from the pork production. The largest source of emissions from pig production systems is nitrous oxide from manure storage (Bonesmo and Enger 2021). This emission represents a nutrient loss from the animal and by improving nutrient efficiency in pigs, emissions and production efficiency can be improved.

Since 1958, Norsvin Landrace has been bred for improved efficiency (Gjerlaug-Enger *et al.*, 2012). This has led to a reduction in carbon dioxide equivalents per kg carcass weight (kg CO_2-eq./kg CW) from 3.5 to 2.3 kg CO_2-eq./kg CW. This reduction is a result of more weaned piglets, lower mortality, shorter growth period and improved feed efficiency (Bonesmo and Enger 2021). To further improve genetic selection for efficiency to reduce carbon footprint and improved efficiency, nitrogen digestibility can be a trait of interest. However, measuring digestibility on a large-scale can be challenging, as this previously has been measured by using respiratory chambers and total excreta collection. Nirea *et al.* (2018) concluded that near-infrared spectroscopy (NIRS) can be used for prediction of chemical composition in feces and is a cost-effective method for measuring digestibility in large-scale, and Déru *et al.* (2021) showed that additive genetic variation existed for digestibility of different nutrients in Large White pigs measured by NIRS. The aim of this study was therefore to estimate genetic parameters for apparent total tract digestibility (ATTD) of nutrients measured by NIRS.

Materials & methods

The data material for the study was collected at Norsvins boar testing station in Norway. The test period was between 40 to 120 kg live weight of the boars. The data recording was conducted from June 2019 to April 2021 on randomly selected Norsvin Duroc boars, fed *ad libitum* on conventional concentrate containing 9.15 g SID lysine/kg and 9.55 MJ NE/kg.

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_77, © K.H. Martinsen *et al.* 2022

Fecal sampling. To determine ATTD of nutrients, individual fecal samples were sampled at 120 kg live weight. The samples were stored at -20 °C before processing and chemical analysis. The samples were freeze dried, homogenized and ground (1 mm sieve) prior to chemical analyses and NIRS analyses. In total, fecal samples were collected from 1,895 boars.

Near-infrared spectroscopy. All NIR spectra were acquired using an XDS Rapid Content Analyzer (Foss, Hilleroed, Denmark) over the spectral region of 400-2,500 nm. For all samples, triplicate measurements were obtained. A total of 300 fecal samples were used to make a calibration model for nitrogen content (Kjeldahl-N) in the fecal samples. Partial least squares regression was used for calibration and prediction (cross-validation R^2=0.73, estimation error (RMSECV) = 1.6 g nitrogen/ kg dry matter). Subsequently, this calibration model was used to predict nitrogen contents in the remaining samples. Prior to regression and prediction, all NIR spectra were pre-processed employing 1-order derivatization and subsequent normalisation using the Standard Normal Variate approach (Barnes *et al.,* 1989).

Traits. All animals had individual records of the following traits: feed consumption from 40 to 120 kg (F40120), days from 40 to 120 kg (D40120), loin depth at 100 kg (LDP) and back fat at 100 kg (BF). Apparent total tract digestibility of nitrogen (ATTDn) and apparent total tract digestibility of dry matter (ATTDdm) was calculated by the model described in Maynard and Loosli (1969), using an indigestible marker (acid insoluble ash) occurring naturally in the feed. Descriptive statistics for the traits are shown in Table 1.

Genetic analyses. The traits were analysed using multivariate animal models, and estimation of variance components and genetic correlations were performed using the DMU software package (Madsen and Jensen, 2013).

The following models were used to analyse the different traits.

For ATTDdm and ATTDn:

$$y_{ijklmno} = MOB_i + OP_j + DAY_k + HYS_l + LITT_NO_m + animal_n + pen_o + e_{ijklmno} \quad (1)$$

For F40120, D40120, LDP and BF:

$$y_{ijklm} = MOB_i + HYS_j + LITT_NO_k + animal_l + pen_m + e_{ijklm} \quad (2)$$

Table 1. Descriptive statistics for feed consumption from 40 to 120 kg (F40120), days from 40 to 120 kg (D40120), loin depth at 100 kg (LDP) and back fat at 100 kg (BF). Apparent total tract digestibility of nitrogen (ATTDn) and apparent total tract digestibility of dry matter (ATTDdm).

	n[1]	Mean	SD[1]	Min[1]	Max[1]
ATTDdm (%)	1,895	73.2	4.6	47.2	84.4
ATTDn (%)	1,608	71.3	6.2	31.3	85.6
F40120 (kg)	13,922	171.4	13.7	100.3	274.7
D40120 (days)	14,030	75.4	8.7	50.9	114.8
BF (mm)	9,306	5.7	1.5	2.9	13.0
LDP (mm)	9,526	58.3	4.7	41.2	77.6

[1] Number of animals (n), standard deviation (SD), minimum value (min), and maximum value (max).

The sampler (OP), day of sampling (DAY), herd-year-season born (HYS), birth month (MOB) and dams litter number (LITT_NO) were included as fixed effects, whereas animal and pen were included as random effects. Pedigree information was traced 10 generations back and in total there were 24706 animals in the pedigree.

Results

Estimates of variance components and heritability for ATTDdm and ATTDn are presented in Table 2. Both additive genetic variance components were significantly different from zero. The estimated heritability for ATTDdm was higher than the heritability for ATTDn. The genetic correlation between ATTDdm and ATTDn was estimated to 0.80.

The genetic correlations between ATTDdm and ATTDn to production traits are presented in Table 3. The genetic correlations between the digestibility traits (ATTDdm and ATTDn) and F40120 were -0.50 and -0.65, respectively. The genetic correlations between the ATTDdm and ATTDn and D40120, LDP and BF were not significantly different from zero.

Discussion

This study has shown that genetic variation exists in apparent total tract digestibility traits, when chemical components were predicted by NIRS, and digestibility was estimated based on an internal marker in the feed. Improved digestibility of nutrients will increase the resource utilization of the animal. Selection for changed body composition, such as carcass leanness and reduced backfat has been a strategy to obtain feed efficient animals in pig breeding. Other traits, such as gross feed intake, feed conversion ratio and residual feed intake has also been used (Patience, 2012). Some of these traits might provide selection responses that are related to the allocation of feed resources to different tissues. This has ensured that feed is used for production. However, to improve feed efficiency further, it is important to know how well the intestines utilize the nutrients in the feed. The trait, ATTDn, can therefore further improve the efficiency of the animals, and hence reduce nitrogen loss in the environment, which is becoming an increasing global problem. To predict ATTDdm and ATTDn based on a spot fecal sample as an alternative to total collection method, acid insoluble ash was used as a marker. Brestenský *et al.* (2017) suggested that acid insoluble

Table 2. Variance components, heritabilities (h^2)[1] and the genetic correlation (r_g) for apparent total tract digestibility of dry matter (ATTDdm) and nitrogen (ATTDn).

	Additive genetic variance (σ_a^2)	Pen variance (σ_{pen}^2)	Residual variance (σ_e^2)	h^2	r_g
ATTDdm (%)	2.9	1.8	10.0	0.22	0.80
ATTDn (%)	5.5	3.2	22.0	0.20	

[1] $h^2 = (\sigma_a^2) / (\sigma_a^2) + (\sigma_e^2)$.

Table 3. Genetic correlations (SE) between apparent total tract digestibility of dry matter (ATTDdm) and nitrogen (ATTDn) and production traits.

	F40120[1]	D40120[2]	LDP[3]	BF[4]
ATTDdm	-0.50 (0.12)	-0.11 (0.15)	-0.18 (0.12)	-0.15 (0.15)
ATTDn	-0.65 (0.11)	-0.10 (0.15)	-0.22 (0.15)	-0.19 (0.15)

[1] Feed consumption from 40 and 120 kg live weight.

[2] Days from 40 to 120 kg live weight.

[3] Loin depth at 100 kg.

[4] Back fat at 100 kg.

ash was a suitable marker to use in digestibility studies for pigs. However, caution must be made with the sampling, to avoid contamination. In the current study, rooting material was held back for three days prior to sampling to avoid the pigs getting acid insoluble ash from other sources than the feed. The method for chemical analysis could influence the level of digestibility estimates, but not the variation. Bastianelli *et al.* (2015) proposed to use direct prediction of digestibility of different nutrients using NIRS. Then, the need for a marker would be minimized, as it is only needed for the fecal samples used for developing the calibration models. The variance components estimated in this study suggested that both digestibility traits are heritable and can be selected for. Few studies have previously carried out genetic analyses of digestibility traits in pigs. However, Déru *et al.* (2021) estimated a heritability for digestibility coefficients for nitrogen to be 0.27. This result is higher than the current study, but on a smaller dataset, different breed and conventional diet. The same study found an even higher heritability for nitrogen digestibility on a high-fibre diet (0.56), due to higher genetic variation. Increasing fibre content in the diet might increase the genetic variation in the trait but could at the same time reduce the digestibility of nutrients (Déru *et al.,* 2021; Zhang *et al.,* 2013). Genetic correlations between the digestibility traits and production traits were identified and suggested that the current selection strategy is improving the digestibility of nitrogen and dry matter. The results are in accordance with Déru *et al.* (2021), which estimated a high negative genetic correlation between nitrogen digestibility and daily feed intake (-0.59) in animals on conventional diet. The current study did not find any significant genetic correlations between ATTDdm, ATTDn and D40120, BF or LDP. This highlights that digestibility traits are heritable, but mainly related to feed intake and general function of the intestines, as opposed to allocation of feed resources to different tissues in the body. However, the results suggested that the current selection strategy is improving the digestibility of nitrogen and dry matter.

Acknowledgements

The authors would like to thank the Norwegian Research Council for funding through project number 268124, the technicians at Norsvin Delta for sampling and Labtek at NMBU for chemical analyses.

References

Barnes R.J., Dhanoa M.S., and Lister S.J. (1989) Appl. Spectrosc. 43(5):772-777. https://doi.org/10.1366/0003702894202201

Bastianelli D., Bonnal L., Jaguelin-Peyraud Y., and Noblet J. (2015) Animal. 9(5):781-786. https://doi.org/10.1017/S1751731114003097

Bonesmo H., and Enger E.G. (2021). Livest. Sci. 254:104746-104758. https://doi.org/10.1016/j.livsci.2021.104746

Brestenský M., Nitrayová S., Heger J., and Patrás P. (2017) J. Anim. Physiol. Anim. Nutr. 101(1):46-52. https://doi.org/10.1111/jpn.12503

Déru V., Bouquet A., Labussière E., Ganier P., Blanchet B. *et al.* (2021) J. Anim. Breed. Genet. 138(2):246–258. https://doi.org/10.1111/jbg.12506

Gjerlaug-Enger E.J., Kongsro J., Ødegård J., Aass L., and Vangen O. (2012) Animal. 6(1):9-18. https://doi.org/10.1017/S1751731111001455

Madsen P., and Jensen J. (2013) A User's Guide to DMU. Version 6, release 5.2. University of Aarhus, Center for Quantitative Genetics and Genomics Dept. of Molecular Biology and Genetics, Research Centre Foulum, Tjele, Denmark.

Maynard L.A., and Loosli J.K. (1969) Animal Nutrition 6th edition. McGraw-Hill Book Company, New York, US.

Nirea K.G., de Nanclares M.P., Skugor A., Afseth N.K., Meuwissen T.H.E. *et al.* (2018) J. Anim. Sci. 96(7):2826-2837. https://doi.org/10.1093/jas/sky182

Patience J.F. (2012) Feed efficiency in swine. Wageningen Academic Publishers, Wageningen, the Netherlands.

Zhang W., Defa L., Liu L., Zang J., Duan W. *et al.* (2013) J. Anim. Sci. Biotechnol. 4(1):17-23. https://doi.org/10.1186/2049-1891-4-17

Part 7

Society – contributions animal breeding to animal welfare

78. Genetic analysis of number of functional teats, piglet survival, and piglet weight in organic pigs

R.M. Zaalberg[1*], T.T. Chu[1,2], J. Jensen[1] and T.M. Villumsen[1]

[1]Aarhus University, Center for Quantitative Genetics and Genomics, 8830 Tjele, Denmark; [2]Vietnam National University of Agriculture, Faculty of Animal Science, Vietnam; roos.zaaalberg@qgg.au.dk

Abstract

Organic pig breeding could be improved by including welfare related traits into breeding goals. This study aimed to estimate genetic parameters for traits related to piglet welfare in organic pigs. Data were collected from an organic multiplier farm. Piglets were weighted at birth and at day ten, number of functional teats were counted at birth. For sows number of living and dead piglets were recorded at birth and on day eleven, and for each litter mean litter weight was determined. Pedigree based animal models were used to estimate genetic parameters. For piglet weight, common litter and maternal genetics were included in the model. Heritability estimates was high for number of functional teats and mean litter weight (0.25-0.48), but low for the other traits (<0.10). Piglet mortality and mean litter weight were negatively correlated (-0.55 to -0.72), whereas number teats was positively correlated to number of living piglets (0.40). In conclusion, there is significant genetic variation for traits related to piglet mortality in organic pigs.

Introduction

The organic pig sector in Europe is gradually expanding (Eurostat, 2021). However, few developments have been made in developing organic pig breeding programs and this may have negative consequences on the compliance of the organic principles within organic pig production (IFOAM, 2005). One of the pillars of organic pig farming is improved pig welfare. For example, sows should move freely, and piglets and sows be kept outdoors on pasture with individual huts during the pre-weaning period. On the other hand, these freedoms give rise to welfare issues, such as increased piglet mortality compared to conventional production (Prunier *et al.*, 2014). One reason for this issue could be that organic pig breeding depends on genetic material from conventional breeding populations. This would be reasonable if there are no genotype by environment interactions between organic and conventional pig production systems. Due to the significant differences between the systems, this seems unlikely, especially for traits such as piglet survival. Conventional breeding programs breed for sows that produce larger litters. Large litters are associated with lower average birth weight (Prunier *et al.*, 2014). Smaller piglets struggle with thermoregulation, and often consume less colostrum and, therefore, have lower chances of survival (Milligan *et al.*, 2002; Rangstrup-Christensen *et al.*, 2018). In outdoor organic systems, rapid intervention is not always possible. Consequently, sows and piglets need to be more independent. Sows, for example, need to have enough teats to support her piglets (Rangstrup-Christensen *et al.*, 2018), and piglets should be strong enough to get through the first risky period without intervention from farm personnel (Quiniou *et al.*, 2002). Not much is known about genetic relations between traits related to survival, piglet weight, litter size and number of functional teats for pigs in organic production environments. The aim of this study, therefore, was to evaluate genetic parameters for piglets traits associated to early piglet weight, and number of functional teats, and sow traits associated to piglet survival.

Materials & methods

Data collection. Data were collected from October 2018 through May 2021 at an organic multiplier farm in Denmark (Bovbjerg Økologi A/S). After removing outliers, 13,546 pigs with a record for number of functional teats at birth (NFT) remained. Individual piglet weight was available for 28,345 piglets at birth (W0), and 22,276 piglets at day ten (W10). Litter size traits were piglets alive at birth (A0), alive at

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_78, © R.M. Zaalberg *et al.* 2022

day 11 (A11), dead until day 4 (D4), and dead at day 11 (D11). These litter size traits were available for approximately 1,200 sows with 3,900 litters. Mean litter weight at birth (μLW0) and day 10 (μLW10) was available for approximately 750 sows with 1,800 litters. For piglet weight, 89% of the piglets were Yorkshire × Landrace (YL), and 11% Landrace (LL). For litter size, 6% of the litters was LL, 48% YL and 46% Duroc × YL (DYL). For mean litter weight, 11% of the litters was LL, 88% YL and 1% DYL.

Statistical analyses. Pedigree-based genetic analyses were performed in DMU (Madsen and Jensen, 2013). For estimating variance components, the following univariate linear regression equations were used.

$$NFT = Xb + Za + e \quad (1)$$

$$\text{Piglet weight} = Xb + Za + Z_m a_m + Wc + Sh + e \quad (2)$$

$$\text{Litter size} = Xb + Za + Cpe + Sh + e \quad (3)$$

$$\text{Mean litter weight} = Xb + Za + Cpe + Sh + e \quad (4)$$

Where b is a vector of fixed effects. For NFT, the fixed effect was breed. For piglet weight (W0, W10), the fixed effects were parity of sow, birth group, sex×breed, and a regression on age at the time of recording. For litter size (A0, A11, D4, D11), the fixed effects were litter breed, and parity of sow. For mean litter weight (μLW0, μLW10), the fixed effects were litter breed, parity of sow, birth group, and a regression on average age at the time of recording. Furthermore, a were vector of random additive direct genetic effects, a_m a vector of random maternal additive genetic effects, c a vector of random common litter effects, h a vector of random herd-year-season effects, pe a vector of random permanent environmental effects caused by the sow, and e a vector of residuals. X, Z, Z_m, W, S, and C were design matrices. Random effects were assumed to be independent, and to follow multivariate normal distributions: $a \sim N[0, A\sigma_a^2]$, $a_m \sim N[0, A\sigma_{am}^2]$, $c \sim N[0, I\sigma_c^2]$, $h \sim N[0, I\sigma_h^2]$, $pe \sim N[0, I\sigma_{pe}^2]$, and $e \sim N[0, I\sigma_e^2]$, where A is the relationship matrix, I is an identity matrix, and σ_a^2, σ_{am}^2, σ_c^2, σ_h^2, σ_{pe}^2, and σ_e^2 are the variance components for the random effects. Based on these variance components, heritability was estimated as:

$$h^2 = \sigma_a^2 / (\sigma_a^2 + \sigma_{am}^2 + \sigma_c^2 + \sigma_h^2 + \sigma_{pe}^2 + \sigma_e^2) \quad (5)$$

For estimating genetic correlations, bivariate models were used based on models (1)-(4). Because herd-year-season effects were too small, this random effect was excluded from bivariate models. In addition, for piglet weight only the maternal genetic component was considered.

Results

Means and standard deviations for the nine traits are presented in the top rows of Table 1. In the lower part of Table 1, heritability and correlation estimates for all traits are shown.

Discussion

The results show there is significant genetic variation for the studied traits measured under organic production conditions (Table 1). In addition, heritability estimates for litter traits such as A0 and μLW0 were comparable to those observed in conventional populations (Rydhmer, 2000). This indicates that even in a less controllable environment, such as outdoor, organic production systems, it is still possible to select for higher piglet weight, better piglet survival at sow level, and NFT.

Table 1. At the top, means and standard deviations for nine traits are shown. In the bottom, heritability (diagonal) and genetic correlations (off-diagonal) for three traits recorded in piglets, and seven traits recorded in sows are presented.

	W0	**W10**	**NFT**	**A0**	**A11**	**D4**	**D11**	**µLW0**	**µLW10**
Unit	kg	kg	#	#	#	#	#	kg	kg
Mean	1.57	3.59	16.17	14.61	12.92	1.26	1.69	1.60	3.64
SD	0.38	0.85	1.16	3.90	3.53	1.65	1.91	0.29	0.62
Weight birth/d0	**0.09**								
Weight d10	0.83[1]	**0.08**							
NFT	0.20	NS[2]	**0.49**						
Alive d0	-	-	0.36	**0.07**					
Alive d11	-	-	0.40	0.95	**0.06**				
Dead d4	-	-	NS	0.53	NS	**0.04**			
Dead d11	-	-	NS	0.73	NS	NC[3]	**0.03**		
µ litter weight d0	-	-	NS	-0.63	-0.53	-0.66	-0.72	**0.36**	
µ litter weight d10	-	-	NS	-0.60	-0.59	-0.55	-0.63	0.82	**0.25**

[1] Genetic correlation based on maternal genetic component.

[2] NS = not significantly different from 0, because |estimate| <2×SE. For significant parameters, SE ranged from 0.01 to 0.21.

[3] NC = model did not converge.

To increase litter size and reduce piglet mortality, current breeding programs for pigs in Denmark select for number of piglets alive at day 5 (Nielsen *et al.*, 2013; SEGES, 2015). Based on consideration of organic production principles, it would be more suitable to select directly for improved welfare (Nielsen *et al.*, 2011), such as fewer dead piglets. To improve survival among organic piglets, one could select for heavier, and thus, more robust piglets (Milligan *et al.*, 2002; Quiniou *et al.*, 2002). Our results suggest that selection for sows that produce heavier piglets will result in fewer dead piglets, but also, fewer piglets alive at day 11 (Table 1). Also, weighing piglets individually is labour intensive compared to counting piglets for litter size traits. One option would be to develop a breeding goal that includes both number of piglets alive and piglet survival. Yet, it is important that selection for larger litters is accompanied by selection for NFT, to ensure the sow can accommodate her piglets (Rangstrup-Christensen *et al.*, 2018). The heritability and correlation estimates presented in the current study provide a basis for developing a welfare-focused breeding goal for organic pigs. In conclusion, the results of this study show there is significant genetic variation for piglet weight, litter weight, number of dead or living piglets, and number of functional teats for pigs reared in an organic outdoor production system. The genetic parameters found in this study can be used to develop a breeding goal for organic pigs that can be used to select directly for fewer dead piglets, and thus, improved piglet welfare.

References

Eurostat (2021) Organic Livestock. Available at: https://ec.europa.eu/eurostat/databrowser/view/org_lstspec/default/table?lang=en

IFOAM (2005) Principles of Organic Agriculture. Available at: https://www.ifoam.bio/en/organic-landmarks/principles-organic-agriculture

Madsen, P., J. Jensen (2013). A User's Guide to DMU. Available at: https://dmu.ghpc.au.dk/dmu/DMU/.

Milligan B.N., Fraser D., Kramer D.L. (2002) Livest. Prod. Sci. 76:181-191. https://doi.org/10.1016/S0301-6226(02)00012-X

Nielsen H.M., Olesen I., Navrud S., Kolstad K., Amer P. (2011) J. Agric. Environ. Ethics. 24:309-330. https://doi.org/10.1007/s10806-010-9264-4

Nielsen B., Su G., Lund M.S., Madsen P. (2013) J. Anim. Sci. 91:2575-2582. https://doi.org/10.2527/jas.2012-5990

Prunier A., Dipper S., Bochicchio D., Edwards S., Leeb C., Lindgren K., Sundrum A., Dietze K., Bonde M. (2014) Organ. Agr. 4:163-173. https://doi.org/10.1007/s13165-013-0040-4

Quiniou N., Dagorn J., Gaudré D. (2002) Livest. Prod. Sci. 78:63-70. https://doi.org/10.1016/S0301-6226(02)00181-1

Rangstrup-Christensen L., Schild S.-L.A., Pedersen L.J., Sørensen J.T. (2018) Res. Vet. Sci. 118:171-180. https://doi.org/10.1016/j.rvsc.2018.02.010

SEGES (2015) Differentietet avl af grise - via Danavl. Available at: https://www.landbrugsinfo.dk/-/media/landbrugsinfo/public/0/0/9/diff_avl_af_oeko_grise_via_danavl.pdf

Rydhmer, L. (2000) Livest. Prod. Sci. 66:1-12. https://doi.org/10.1016/S0301-6226(99)00170-0

79. Genetic analysis of feeding behaviour and tail lesions

K.H. Martinsen[1*], Ø. Nordbø[1], E.G. Enger[1], D. Olsen[1] and E. Grindflek[1]

[1]*Norsvin, Storhamargata 44, 2317, Hamar, Norway; kristine.martinsen@norsvin.no*

Abstract

To improve production efficiency and ensure good animal welfare, novel phenotypes are developed for minimizing damaging behaviour in pigs. The aim of this study was therefore to estimate genetic parameters for two novel feeding behaviour phenotypes and investigate their genetic relationship with other traits recorded at the nucleus boar testing station. The genetic analyses were performed using multivariate animal models. The estimated heritability of interrupter (IntER) and interrupted (IntED) at the feeding station was 0.58 and 0.61, respectively, and the genetic correlation between the traits was 0.89. The genetic correlation between the feeding behaviour traits (IntER and IntED) with tail lesions and feed consumption from 40 to 120 kg live weight was not significantly different from zero. The results suggested that the feeding behaviour traits in this study could be treated as the same trait. However, as the frequency of tail lesions in this study was low, the genetic relationship between feeding behaviour and tail lesions remains unclear.

Introduction

Swine breeding companies strive to develop novel traits for desired pen behaviour to ensure genetic material which is production efficient, have high health status and good animal welfare. This involves minimising antagonistic behaviours such as aggression and tail biting and stimulating positive behaviours such as play and social behaviour. Individual behaviour data is difficult to collect, as it requires continuously recording individual pigs over time, rather than use one single point of observation. Feeding stations with radio frequency identification system (RFID) provide continuous high resolution individual recordings of feeding data. Every time an animal enters the feeding station, weight, feed intake and occupation time are recorded. When such data is aggregated over time, the individual patterns of feeding behaviour become clearly distinguishable, and when combined with pedigree or genomic data, they could be used as a source to find genetic variation in these traits. The feeding station data could be used to measure the feed intake related traits directly on the individual (Cheng *et al.*, 2021; Putz *et al.*, 2019). Alternatively, the timing between subsequent visits of feeding station could be used to improve the interaction matrices in estimation of indirect social genetic effects (Angarita *et al.,* 2021). If an animal is interrupting other animals during feeding, this could affect the feed intake and growth of the pen mates. In this paper, we used the feeding station data to develop novel traits to quantify, on an individual level, interruptions around feeding. These traits were further investigated to identify if the traits were heritable and if there were genetic relationships between these traits, feed consumption, number of visits in the feeding station and tail lesions.

Materials & methods

The data in this study were from purebred Norsvin Landrace boars with feeding data collected at one of the nucleus boar testing stations. In total, 4,281 boars from the period 2018 to 2021, were included in the study.

Feeding data. The feeding data were recorded by feeding stations (FIRE; Osborne Industries Inc., Osborne, KS). Individual amount of feed consumed at each visit and time spent per visit were recorded daily throughout the test period, from 40 to 120 kg live weight. Only one animal can enter the feeding station at the time, but interruptions can occur. Therefore, two traits (events) from feeding stations were generated for this study. These events were when the animal was an interrupter at the feeding station (IntER) and when the animal was interrupted in the feeding station (IntED). In addition, median of daily number of visits through the test period (MVIS) was included in the study. An interruption in the feeding station

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_79, © K.H. Martinsen *et al.* 2022

was defined as the occurrence of two subsequent feed recordings, when the first animal had spent a short period of time in the feeding station (time <2 min), and it was a short time interval (time <1 min) from the animal leaving the feeding station until the next animal entered. The animal leaving the station was defined as IntED and the subsequent animal was defined as IntER. The number of 'interrupted' and 'interrupter' events were summarized per day and averaged over number of days in the test period. Feed consumption from 40 to 120 kg (F40120) was also included in the analyses. Descriptive statistics are presented in Table 1.

Tail biting. Recording of tail lesions (TL) was done throughout the whole test period. The trait was recorded as a binary trait, stating whether the animal had a tail lesion or not. In total, 149 animals (3.5%) in the data set had tail lesions, recorded from 2019 to 2021. Descriptive statistics of the trait is presented in Table 1.

Genetic analyses. The traits were analysed using multivariate animal models, and estimation of variance components and genetic correlations were performed using the DMU software package (Madsen and Jensen, 2013).

The following two models were used to analyse the different traits.

For IntER, IntED, MVIS and F40120:

$$y_{ijklm} = HY_i + MOB_j + PAR_k + animal_l + pen_m + e_{ijklm} \quad (1)$$

For TL:

$$y_{ijk} = YS_res_i + animal_j + pen_k + e_{ijk} \quad (2)$$

Herd-year at birth (HY) *i*, birth month (MOB) *j*, dam's parity number (PAR) *k*, where included as fixed effects in Model 1 and year-season when the animal started the test (YS_res) *i* was included as a fixed effect in Model 2. The effects of animal *l* and pen *m* were included as random effects in Model 1 and animal *j* and pen *k* were included as random effects in Model 2. Litter effect was not significant, as few litter mates were present in the data. Pedigree information was traced 10 generations back and in total there were 12,483 animals included in the pedigree.

Results

Estimates of variance components and heritabilities for feeding behaviour traits and tail lesion are presented in Table 2. Both additive genetic variance components were significantly different from zero. The estimated heritability for the feeding behaviour traits (IntER, IntED and MVIS) were high (>0.5), and the heritability for TL was low (0.05).

Table 1. Descriptive statistics for interrupter (IntER), interrupted (IntED), median daily visits in the feeding station (MVIS), tail lesions (TL) and feed consumption in kg from 40 to 120 kg live weight (F40120).

	n[1]	Mean	SD[1]	Min[1]	Max[1]
IntER	1,926	1.72	0.87	0.34	6.94
IntED	1,926	1.71	1.16	0.11	10.23
MVIS	1,926	9.71	2.92	4	23
TL	4,281	1.03	0.18	1	2
F40120	1,884	171.8	12.1	117.6	239.7

[1] Number of animals (n), standard deviation (SD), minimum value (min), and maximum value (max).

Table 2. Variance components and heritabilities (h^2)[1] for interrupter in feeding station (IntER), interrupted in feeding station (IntED), median of number of visits in the feeding station (MVIS) and tail lesions (TL).

	Additive genetic variance (σ^2_a)	Pen variance (σ^2_g)	Residual variance (σ^2_e)	h^2
IntER	0.27	0.08	0.24	0.53
IntED	0.82	0.02	0.52	0.61
MVIS	6.02	0.02	2.80	0.68
TL	0.0009	0.01	0.018	0.05

[1] $h^2 = (\sigma^2_a) / (\sigma^2_a + \sigma^2_e)$.

The genetic correlations between the traits are presented in Table 3. The genetic correlation between IntER and IntED was 0.89. Similar high genetic correlations were found between IntER and MVIS (0.94) and IntED and MVIS (0.94). The genetic correlations between the feeding behaviour traits and TL and F40120 were low and not significantly different from zero (Table 3).

Discussion

This study has shown that genetic variation existed in feeding behaviour traits recorded in feeding stations and tail lesions in Norsvin Landrace. Other studies have also reported high heritabilities on a wide range of different feeding behaviour traits recorded by feeding stations (Herrera-Cáceres *et al.* 2020; Kavlak *et al.,* 2021). Kavlak *et al.* (2021) reported slightly lower heritabilities for feeding behaviour traits recorded by feeding stations. For number of visits in the feeding station per day, the heritability ranged from 0.26-0.33 depending on age. However, the study included two breeds and crossbred animals in the same analysis and corrected for breed. This might have influenced the results of Kavlak *et al.* (2021). The current study only investigated purebred animals from the same breed. Breuer *et al.* (2005) found significant genetic variation in clinical tail biting in Landrace and estimated a heritability of 0.05 for the trait tail biter. The current study also had information about tail biting animals, but the data material was too small for the analysis. The current study found a heritability of 0.05 for tail lesion in the animal, in accordance with Canario and Flatres-Grall (2018) in a synthetic line (0.06). Scientific studies of tail biting are challenging, as recording of tail biting incidences is difficult. It appears suddenly, is unpredictable and spreads throughout the barn, which makes it difficult to record information pre- or post-outbreak (D'Eath *et al.,* 2014). Using new technology (such as cameras), might make it possible to detect more and a diversity of tail biting incidences, as well as more genetic variation in such traits.

Several studies have tried to investigate potential proxy traits for tail biting, such as pig activity and chewing on objects (Statham *et al.,* 2010: Ursinus *et al.,* 2014). Statham *et al.* (2009) found that activity level in the pen were statistically higher in groups with severe tail biting outbreaks, four days prior to the outbreak. Similarly, Wallenbeck and Keeling (2013) concluded that electronic feeding stations can be

Table 3. Genetic correlations (SE) between feeding behaviour traits, tail lesion (TL), median of number of visits per day (MVIS), and feed consumption from 40 to 120 kg live weight (F40120).

	MVIS	TL	F40120	IntER[1]
IntER[1]	0.94 (0.02)	-0.09 (0.2)	-0.12 (0.13)	-
IntED[2]	0.94 (0.02)	0.05 (0.2)	-0.10 (0.13)	0.89 (0.03)

[1] Interrupor in feeding station.

[2] Interrupted in feeding station.

used for surveillance of tail biting outbreaks in pigs. The current study did not find any significant genetic correlations between the feeding behaviour traits and tail lesions. In this study, 149 animals had recorded a tail lesion (3.5%), and the rest of the animals were assumed healthy. This might underestimate the actual frequency of tail lesions**.** Further, it is reasonable to expect more accurate results if the frequency was higher. In the current study, only information about tail lesions is included. From a biological perspective, a trait like 'tail biter' might be more linked to feeding behaviour traits like IntER, as biting the tail of an animal inside the feeding station is a strategy for getting access to the feeding station. However, such a link was not identified in this study. As the frequency of tail lesions in this study was low, the results should be interpreted with caution. There is a need for more and accurate data to identify the genetic relationship between feeding behaviour and tail lesions. The genetic correlation between MVIS and TL (not shown) was not significantly different from zero, in accordance with the other genetic correlations between feeding behaviour and tail lesions (Table 3). However, the current study revealed high genetic correlations between IntED, IntER and MVIS (Table 3). These results suggested that the feeding behaviour traits in this study were similar, as they describe different quantifications of activity level at the feeding station area. Further, it seemed like IntER or IntED added little additional information to MVIS, which is routinely recorded in the breeding program (Table 3).

References

Angarita B.K., Han J., Cantet R.J.C., Chewning S.K., Wurtz K.E. *et al.* (2021) J. Anim. Sci. 99(5):042. https://doi.org/10.1093/jas/skab042

Breuer K., Sutcliffe M.E.M., Mercer, J.T., Rance K.A., O'Connell N.E. *et al.* (2005) Livest. Prod. Sci. 93:87-94. https://doi.org/10.1016/j.livprodsci.2004.11.009

Canario, L., and Flatres-Grall L. (2018) Proc. of the 11th WCGALP, Auckland, New Zealand.

Cheng J., Putz A.M., Harding J.C.S., Dyck M.K., Fortin F. *et al.* (2021) J. Anim. Sci. Biotechnol. 12(1):105. https://doi.org/10.1186/s40104-021-00622-x

D'Eath R.B., Arnott G., Turner S.P., Jensen T., Lahrmann H.P. *et al.* (2014) Animal. 8: 1479-1497. https://doi.org/10.1017/S1751731114001359

Herrera-Cáceres W., Regab M., and Sánchez J. (2020) Animal. 14(2): 233-242. https://doi.org/10.1017/S1751731119002179

Kavlak A.T., Strandén I., Lidauer M.H., and Uimari P. (2021) Animal. 15:100168 https://doi.org/10.1016/j.animal.2020.100168

Madsen P., and Jensen J. (2013). A User's Guide to DMU. Version 6, release 5.2. University of Aarhus, Center for Quantitative Genetics and Genomics Dept. of Molecular Biology and Genetics, Research Centre Foulum, Tjele, Denmark.

Putz A.M., Harding J.C.S., Dyck M.K., Fortin F., Plastow G.S. *et al.* (2019) Front. Genet. 9:660. https://doi.org/10.3389/fgene.2018.00660

Statham P., Green L., Bichard M., and Mendl M. (2009) Appl. Anim. Behav. Sci. 121:157-167. https://doi.org/10.1016/j.applanim.2009.09.011

Ursinus W., Van Reenen C.G., Kemp B., and Bolhuis J.E. (2014) Appl. Anim. Behav. Sci. 156:22-36. https://doi.org/10.1016/j.applanim.2014.04.001

Wallenbeck A., and Keeling L.J. (2013) J. Anim. Sci. 91:2879-2884. https://doi.org/10.2527/jas.2012-5848

80. Response to selection for social dominance including direct and indirect genetic effects

B. Tuliozi*, E. Mancin, G. Gomez Proto, R. Mantovani and C. Sartori

Department of Agronomy, Food, Natural Resources, Animals and Environment, University of Padova, Via dell'Università 16, 35020 Legnaro, Italy; beniamino.tuliozi@unipd.it

Abstract

This study aimed to estimate genetic parameters for social dominance and calculate the response to selection for the direct and indirect component of this trait in an Alpine cattle breed. The outcomes of dyadic interactions occurring in 18 years of ritualized cow contests were analysed using a quantitative model including indirect genetic effects. Response to selection was calculated via a multivariate approach, assigning different weights to direct and indirect components of social dominance. Variances for direct and indirect components were equivalent in magnitude to each other and to their covariance, which exhibited the opposite sign. This caused selection responses to be equal in magnitude but differing in verse whenever selection weights were slightly higher for one component. Our study provides evidence that minimal departures from a balanced selection can strongly select the ability to win (or lose) contests: we argue that this has might have wide-ranging consequences for crucial welfare issues.

Introduction

Social dominance, defined by Strauss and Shizuka (2022) as 'a dyadic social relationship that emerges from sequences of agonistic interactions, where one individual exhibits subordination' is a key aspect of the life of social species. This relationship is the basic component of the social hierarchies that are established within social groups, which in turn can determine resource access and greatly impact individuals' phenotype and welfare. Dominance hierarchies arise in cattle whenever unfamiliar individuals are grouped together, both under intensive rearing systems and in free-ranging conditions (Phillips, 2002). While dominance *per se* is not correlated with increased aggressiveness (Plusquellec and Bouissou, 2001), a consistent selection for dominant individuals can lead to welfare issues, as several key health and life history traits present negative genetic correlations with fighting ability (Sartori *et al.*, 2020). An effective strategy to study the genetic basis of social dominance is to use indirect genetic effects (IGEs), which model the influence of other individuals on the focal's phenotype. Cows of local Alpine breeds of cattle in Switzerland and Italy are employed in ritualized fighting contests that mirror those that happen at summer pastures when herds meet and interact freely. These contests have great importance in local culture and are structured as a series of dyadic agonistic encounters between cows. This setting has given researchers the rare chance to investigate the genetics of social dominance using models including IGE (Sartori and Mantovani, 2013). The response to selection of social dominance has not been yet studied: in this study our aims are thus: (1) assess the variance components for social dominance including IGEs; (2) describe how to calculate the response to selection for direct and indirect components of dominance under different selective conditions. Our results will allow to lay the necessary groundwork for considering social dominance into future welfare decisions.

Materials & methods

Study subjects and data collection. All 12,016 study subjects (23,591 individuals in the pedigree) belonged to the Aosta Chestnut-Black Pied cattle. Information on dominance (42,938 contests) was collected during ritualized cow contests (Batailles des Reines; 20-21 day-long weekly contests per year over 18 years). This local tradition finds its roots centuries ago, as a way of mimicking the same contests for dominance occurring at pasture. Each contest consists of a series of individual confrontations among cows belonging the entire population, where two cows coming from different herds are led to face each other and

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_80, © B. Tuliozi *et al.* 2022

are free to interact until one of the two cows shows a submissive posture. Cow interactions are not cruel (most interactions are not physical) and welfare of animals is guaranteed by qualified expert veterinarians monitoring the competitions. Each cow is allowed to compete in multiple contests per year, each including up to seven confrontations per cow. For further details see Sartori and Mantovani (2013).

Modelling social dominance trait. We considered the dyadic interactions as a trait, and to avoid data duplication each interaction was present only once in the dataset, with one individual being randomly chosen as focal while the other was recorded as opponent. Phenotype was 2 for victory and 1 for defeat. Fixed factors retained in the analysis were year, day of tournament, weight category, difference in age and difference in weight between focal and opponent. Under the variance partitioning approach (Bijma, 2010) for IGEs, the genetic model can be written as: $y = X\beta + W_h h + W_D Pe_D + W_C Pe_C + Z_D a_D + Z_C a_C + e$; where y is an N×1 vector including the binary outcomes, β is the vector of fixed effects, h is the vector for the herd of focal, Pe_D is the vector of direct permanent environmental (Pe) components, Pe_C is the vector of Pe effects provided by the opponent, a_D is the vector of direct genetic component, a_C is the vector of the additive indirect effect, and e is the vector of error terms. Moreover, X, W_h, W_D, W_C, Z_D and Z_C are incidence matrices with proper dimensions. With A being the additive relationships matrix, genetic components can be written as C⊗A, with C as a 2×2 matrix including direct and indirect genetic (co)variances, that are σ^2_{aD}, σ^2_{aC} and $2{\cdot}\sigma_{aDC}$. A similar structure can be assumed also for Pe, that is CPe⊗I, including an identity matrix I and a 2×2 matrix for direct and indirect Pe (co)variances σ^2_{PeD}, σ^2_{PeC} and $2{\cdot}\sigma_{PeDC}$. The model was run under a Bayesian approach using the program THRGIBBS1f90 (Misztal 2008). We ran 550,000 Gibbs sampling iterations with a burn-in of 50,000 samples. We computed a direct heritability for social dominance as $h^2_D = \sigma^2_{aD}/\sigma^2_P$, where the total phenotypic variance σ^2_P sums all variance estimates. We also calculated an indirect heritability $h^2_C = \sigma^2_{aC}/\sigma^2_P$, a total heritable variance $h^2_{TOT} = (\sigma^2_{aD}+\sigma^2_{aC}+2\sigma_{aD,C})/\sigma^2_P$, and the genetic correlation between direct and indirect components, $r_{aDC} = \sigma_{aDC}/(\sigma^2_{aD}{\cdot}\sigma^2_{aC})^{0.5}$ (Bijma 2010). The variation of individual estimated direct and indirect breeding values as well as of the estimated total breeding value (written as TBV = DBV+SBV, Bijma 2007) were plotted in a 20 years' time-window (Figure 1).

Predicted response to selection. When selecting for social dominance, typically there is direct selection for the capability to win contests, i.e. being dominant. At the same time, this brings a correlated response on the capability to lose contests, i.e. being subordinate. Direct response to selection for a trait x (R_x; Falconer and Mackay, 1996) is: $R_x = ih_x\sigma_{ax}$ (1), where i is the selection intensity (deviation of selected individuals

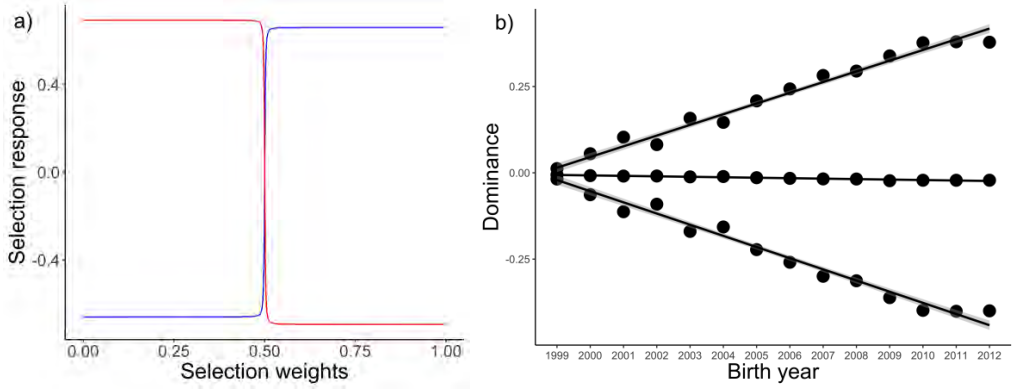


Figure 1. (a) Influence of selection weights of the direct component on selection responses. (b) Genetic trend of direct, indirect, and total estimated breeding values.

from population mean on the phenotypic standard deviation of the trait (σ_{Px}), h is the square root of the narrow-sense heritability of the trait ($h^2_x = \sigma^2_{ax}/\sigma^2_{Px}$), and σ_{ax} is the genetic standard deviation of the trait. The correlated response to selection on a trait y (CR_y; Falconer and Mackay, 1996) is: $CR_y = ih_x r_a \sigma_{ay}$ (2), where r_a is the genetic correlation between traits x and y and σ_{ay} is the genetic standard deviation of the trait y. In social dominance, an individual showing the direct trait 'good at winning' also holds the information 'bad at losing' as indirect effect: we thus expect a genetic correlation between these two components close to -1: r_a = -1, and equal variances but opposite covariances for direct (D) and indirect (C) components (Bijma, 2011): $\sigma^2_{aC} = \sigma^2_{aD}$; $\sigma_{aDC} = -\sigma^2_{aD}$. Empirical evidence has been already provided (e.g. Wilson *et al.*, 2011; Sartori and Mantovani, 2013). Moving from these assumptions and formulas (1) and (2), the direct selection for D is $R_D = ih_D\sigma_{aD}$ and the correlated response on C is $CR_C = ih_D(-1)\sigma_{aD}$, that is: $CR_y = -ih_D\sigma_{aD}$ (3). We therefore expect for the subordinate condition a correlated response to selection that is equal in magnitude but opposite in direction with respect to the direct selection response on the dominant status. Similar relationships may also link the permanent environmental variance of direct and indirect effects when present as in Sartori and Mantovani (2013). We can make similar considerations also moving to a multivariate context and applying the formulas of multivariate response to selection (R; Lande, 1979), here written as in Kause *et al.* (2015): $R = (i/\sigma_i)b'P^{-1}$ (4), where σ_i is the standard deviation of the selection index $\sigma_i = (b'Pb)^{0.5}$, b is the vector of weights for the selection index, being $b = P^{-1}Ga$, P the phenotypic (co)variance matrix and G is genetic (co) variance matrices. Here, the genetic matrix considered is C, and P sums C, CPe and the random residual term. Vector a includes the standardized economic weights of direct and indirect components, obtained as a_0/σ_a, being each a_0 the economic weight assigned to the target component and σ_a the respective square root of the genetic variance (either direct or indirect). We assigned different economic weights varying from 0 to 1 to direct and indirect components, mimicking the possible contexts in which dominance can occur and evolve. Yearly selection rate was calculated by dividing R by the population generation length t=4.32 computed from pedigree.

Results & discussion

Posterior means are reported in Table 1. As expected, σ^2_{PeD} and σ^2_{aD} are close to equal respectively to σ^2_{PeC} and σ^2_{aC}. σ_{PeDC} and σ_{aDC} are also similar in magnitude to their respective components, but with opposite sign; in fact, we found a negative correlation close to -1 between direct and indirect genetic component of dominance. Heritability of both direct and indirect component was in line with previous studies (Sartori and Mantovani 2013), with total heritability close to 0, because of the strong negative covariance (Moore

Table 1. Variance components, heritabilities, response to selection of social dominance.

	σ^2_{PeD}	σ_{PeDC}	σ^2_{PeC}	σ^2_{aD}	σ_{aDC}	σ^2_{aC}	σ^2_e	
Variance[1]	0.178	-0.156	0.144	0.235	-0.243	0.260	1.001	
(±sd)	(±0.018)	(±0.014)	(±0.017)	(±0.020)	(±0.017)	(±0.019)	(±0.010)	
	h^2_D	h^2_C	h^2_{Tot}				r_{aDC}	
Heritability[2]	0.224	0.247	0.008				-0.984	
(±sd)	(±0.019)	(±0.018)	(±0.006)		Genetic correlation (±sd)		(±0.012)	
Economic weights (direct-indirect)			1-0		0.5-0.5		0-1	
Response to selection			Gen.[3]	Year	Gen.	Year	Gen.	Year
- for direct component			0.660	0.153	-0.035	-0.008	-0.660	-0.153
- for indirect component			-0.693	-0.160	0.044	0.010	0.693	0.160

[1] σ^2_{PeD}: direct environmental effect, σ^2_{PeC}: indirect permanent environmental effect, σ_{PeDC}: covariance between these two effects, σ^2_{aD}: direct additive genetic effect, σ^2_{aC}: indirect genetic effect (IGE), σ_{aDC}: covariance between these two effects, σ^2_e: residual variance.

[2] direct heritability for social dominance, h^2_C: indirect heritability, h^2_{Tot}: total heritability, r_{aDC}: genetic correlation for direct and indirect genetic effects.

[3] Gen: generational.

et al., 2002). Strength of response to selection (R) varied according to the economic weights assigned (Figure 1a): as expected, direct and indirect components showed roughly opposite R, apart when the same economic weight (0.5) was assigned. In this case, both direct and indirect selective pressures are active on both components, constraining R to values close to zero. This mimics a situation in which both being dominant and subordinate status are advantageous for individuals (the selection for indirect component of being dominant is the same to select for being subordinate). It is interesting to note that minimal departures from a balanced selection for direct and indirect components lead to the same R we have when only one component is directly selected, meaning that just scarce selective pressures are able to trigger an empirical selection favouring the ability to dominate, or vice versa. This situation is occurring in Aosta Chestnut-Black Pied cattle: genetic trends (Figure 1b) show a weak but constant divergence of the estimated breeding values of the direct and indirect component and a null variation of the total breeding value, which suggests the occurrence of an active empirical selection for being dominant, as if economic weight higher than 0.5 was assigned to this component. Social dominance itself does not carry a negative connotation for the individuals' welfare: breeds selected for social dominance do not show increased aggressiveness (Plusquellec and Bouissou, 2001). However, the genetic trend hereby reported might underscore a genuine risk for the welfare of the breed due to the worsening of health traits genetically correlated to social dominance, such as somatic cells count (Sartori *et al.*, 2020). Given our evidence in support of the strong selection response for social dominance, both direct and indirect, we argue that an awareness of possible ongoing selective pressures for social dominance at the breed or herd level may help to guarantee an adequate animal welfare by informing management decisions – for example, by choosing breeding individuals showing favourable features for both social dominance and welfare-related traits.

References

Bijma P. (2010) Genetics 186(3):1013-1028. https://doi.org/10.1534/genetics.110.120493.

Bijma P. (2011) Genetics 189(4):1347-1359. https://doi.org/10.1534/genetics.111.130617.

Bijma P., Muir W.M., and Van Arendonk J.A.M. (2007) Genetics 175:277-288. https://doi.org/10.1534/genetics.106.062711.

Falconer D.S., and MacKay T.F.C. (1996) Introduction to Quantitative Genetics. Addison Wesley Longman, Harlow, UK.

Kause A., Mikkola L., Strandén I., and Sirkko K. (2015) Animal 9(1):35-42. https://doi.org/10.1017/S1751731114001992.

Lande R. (1979) Evolution 33:402-416. https://doi.org/10.2307/2407630.

Misztal (2008) J. Anim. Breed. Gen. 125:363-370. https://doi.org/10.1111/j.1439-0388.2008.07744.x

Moore A.J., Haynes K.F., Preziosi R.F., and Moore P.J. (2002). Am. Nat. 160: S186–S197. https://doi.org/10.1086/342899.

Sartori C., and Mantovani R. (2013) Heredity 110(1):3-9. http://doi.org/10.1038/hdy.2012.56.

Plusquellec P., and Bouissou, M. F. (2001). Appl. Anim. Behav. Sci. 72(1):1-21. https://doi.org/10.1016/S0168-1591(00)00198-2.

Sartori C., Guzzo N., and Mantovani R. (2020) Animal 14(1):13-21. https://doi.org/10.1017/S175173111900168X.

Strauss D.E., and Shizuka D. (2022) Philos. Trans. R. Soc. 377:20200445. http://doi.org/10.1098/rstb.2020.0445.

81. Empowerment of the Italian small ruminant sector: new traits and tools toward a sustainable management

A. Negro[1,2*], M. Cortellari[2], A. Bionda[2], S. Grande[1], A. Cesarani[3], A. Carta[4], N. Macciotta[3], S. Biffani[5] and P. Crepaldi[2]

[1]Ufficio Studi, Associazione Nazionale della Pastorizia (Asso.Na.Pa.), Via XXIV Maggio 44, 00187, Rome, Italy; [2]Dipartimento di Scienze Agrarie e Ambientali – Produzione, Territorio, Agroenergia, Università degli Studi di Milano, Via Celoria 2, 20133, Milan, Italy; [3]Dipartimento di Agraria, Università degli Studi di Sassari, Viale Italia 39, 07100, Sassari, Italy; [4]Unità di Ricerca di Genetica e Biotecnologie, Agris Sardegna, Loc. Bonassai S.S. 291 Sassari-Fertilia – Km. 18,600, 07100, Sassari, Italy; [5]Istituto di Biologia e Biotecnologia Agraria (IBBA), Consiglio Nazionale delle Ricerche, Via Edoardo Bassini 15, 20133, Milan, Italy; a.negro@assonapa.it

Abstract

This work summarizes the main results of the CHEESR project, a national initiative with the aim of introducing innovative tools for biodiversity conservation and genetic improvements of the Italian small ruminant populations. Pedigrees from 81 breeds and genomic data from 11 case-study breeds have been analyzed. These, together with the collection of new phenotypes, led to the development of welfare- and health-related genomic tools for the first time in Italian sheep and goats. The introduction of these new instruments, which will continue with the SHEEP&GOAT project, will contribute to devise new strategies for sustainable development of the Italian small ruminant sector.

Introduction

Livestock genetic variability has a pivotal role in the ecosystem conservation, the land management, and the mitigation of the effects of climate change (Cao *et al.* 2021). Thanks to its history and variety of environments, climates, and traditions, Italy is characterized by a great biodiversity of all livestock species (Cortellari *et al.* 2021b), with over 80 small ruminant breeds. The scientific research about the Italian biodiversity has been conducted by the *Italian Goat Consortium* for goats (Nicoloso *et al.* 2015; Cortellari *et al.* 2021a; b) and by the Bi.Ov.*Ita Consortium* for sheep (Ciani *et al.* 2014; Mastrangelo *et al.* 2018). The European Agricultural Fund for Rural Development (EAFRD) financed the Italian National Rural Development Plan (PSRN) – sub-measure 10.2 – 'Conservation, Health and Efficiency Empowerment of Small Ruminant' (CHEESR project), managed by the Italian Sheep and Goat Breeders Association (Asso.Na.Pa.). This project aims to introduce in the Italian sheep and goat sector several innovative actions related to: the monitoring and conservation of the biodiversity, the improvement of the breeds with the collection of new phenotypes, and the development of new genetic and genomic breeding values.

The present work summarizes the main results of the CHEESR project, focusing on the application of innovative genomic tools in the Italian small ruminant sector to support biodiversity management and the genetic improvement of the populations under selection, eventually providing information about future developments.

Materials & methods

Genealogic data of 35 goat breeds and 46 sheep breeds were analyzed through *optiSel* R package: pedigree depth and completeness and the main genetic and demographic parameters were estimated. Furthermore, effective population size (Ne) was estimated using Wright's formula, based on the male (M) and female (F) ratio (4MF / M+F), calculated in 2020 for each breed.

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_81, © A. Negro *et al.* 2022

The OvineSNP50 and GoatSNP50 bead chips were used to genotype 2,820 sheep of seven breeds (Sarda, Istriana-Carsolina, Comisana, Massese, Delle Langhe, Fabrianese, and Gentile di Puglia) and 3,320 goats of four breeds (Saanen, Camosciata delle Alpi, Garganica, and Nicastrese). Runs of homozygosity (ROH) and the related inbreeding coefficient (F_{ROH}) were analyzed using Plink software. These data were also used, together with phenotypic information collected from 3,680 goats and 766 sheep, to develop new genomic breeding values with the BLUPF90 family softwares. The Single Step Genomic Best Linear Unbiased Prediction (ssGBLUP) method was applied to estimate the variance components of the welfare score (Wel), locomotion score (Loc), body condition score (BCS), and somatic cell score (SC). Wel is composed of 2 traits (chest circumference and BCS) for goats and 4 traits (chest circumference, BCS, presence of abscesses, and *lupìa*) for sheep. Loc is composed of 9 traits for goat and 8 for sheep; these morphological traits and morpho-functional defects are related to limbs, dorsal line and rump, and udder as well as hoof care.

Results

Considering the pedigrees of all the 81 Italian breeds, in median, we could trace up to 2 and 3 full generations and up to 5 and 7 maximum generations for goat and sheep, respectively. The pedigree completeness varied among breeds, with a median percentage of sheep with a complete first generation higher than goats (34 vs 14%). According to the effective population size (N_e), 18% of the breeds appear to be at risk of extinction in the short term (N_e<50).

Pedigree and genomic data for the 11 genotyped breeds are reported in Table 1. The correlation between the pedigree-based (F_{PED}) and the genomic (F_{ROH}) inbreeding coefficients was significant for Sarda, Massese, Comisana, Delle Langhe, Camosciata delle Alpi and Saanen breeds. At species level, the correlation was higher in sheep (r=0.82, *P*<0.0001) than in goats (r=0.28, *P*<0.0001). Differ0ences in the recent inbreeding, derived from long ROH, can depend on the management and structure of the populations.

All the traits presented here (Wel, Loc, BCS, and SC) are not under selection and have been analyzed for the first time in the Italian populations within this project. The results of the two composed scores (Wel and Loc) are reported in Table 2. The great majority of animals do not present welfare-related problems (Wel = 0). Instead, about only one third of the sheep presents the lowest possible Loc score. The heritability of all the analysed traits was calculated. The wide highest probability density (HPD) interval of the estimated h^2 suggests that, increasing the sample size and selecting the most representative animals, there is room for improvement (Table 3).

Table 1. Results of the pedigree and genomic analyses for the genotyped breeds.[1]

Breed	FullGen	MaxGen	N_e	F_{PED}	F_{ROH}	r	% F_{ROH} for ROH>16 Mb
[S]Sarda	9	40	19,169	0.093	0.134	0.80*	19%
[S]Massese	8	28	1,561	0.038	0.078	0.32*	10%
[S]Gentile di Puglia	4	11	1,042	0.008	0.045	0.84	22%
[S]Comisana	8	20	726	0.018	0.051	0.38*	9%
[S]Delle Langhe	9	33	248	0.060	0.118	0.44*	17%
[S]Fabrianese	6	16	190	0.079	0.128	0.16	47%
[S]Istriana-Carsolina	3	8	126	0.036	0.141	0.16	36%
[G]Camosciata delle Alpi	6	22	1,891	0.016	0.065	0.26*	22%
[G]Saanen	7	23	1,450	0.020	0.071	0.31*	26%
[G]Garganica	3	5	793	0.021	0.077	0.02	40%
[G]Nicastrese	2	4	684	0.000	0.036	0.00	24%

[1] [S] = sheep; [G] = goat; FullGen = maximum number of fully traced generations; MaxGen = maximum number of generations traced; N_e = effective population size; F_{PED} = mean pedigree-based inbreeding; F_{ROH} = mean genomic inbreeding, r = correlation coefficient between F_{PED} and F_{ROH}; * = *P*<0.0001.

Table 2. Results of welfare and locomotion scores.[1]

Breed	Wel			Loc		
	0	1	2	0	1	2
[S]Comisana (n. 529)	88%	12%	0%	32%	49%	19%
[S]Massese (n. 219)	93%	6%	1%	31%	51%	18%
[G]Camosciata delle Alpi (n. 2662)	97%	3%	0%	87%	10%	3%

[1] [S] = sheep; [G] = goat; Wel = welfare score; Loc = locomotion score; the higher scores are related to animals with more welfare-or-locomotion-related problems.

Table 3. Heritability for the analyzed traits.[1]

Breed	Wel h^2 (HPD)	Loc h^2 (HPD)	BCS h^2 (HPD)	SC h^2 (SD)
[S]Comisana	0.10 (0.006-0.24)	0.09 (0.001-0.24)	0.11 (0.02-0.25)	0.03 (0.01)
[S]Massese	0.13 (0.004-0.34)	0.18 (0.005-0.43)	0.17 (0.004-0.38)	0.05 (0.02)
[G]Saanen				0.09 (0.02)
[G]Camosciata delle Alpi	0.30 (0.003-0.58)	0.13 (0.004-0.28)	0.26 (0.03-0.48)	0.10 (0.02)

[1] [S] = sheep; [G] = goat; Wel = welfare score; Loc = locomotion score; BCS = body condition score; SC = somatic cell score; HPD = high probability density interval; SD = sample standard deviation; h^2 = heritability.

Discussion

Our findings show that the Italian small ruminant populations present a wide variety of situations in terms of pedigree management. Overall, sheep breeds are characterized by a higher pedigree depth, completeness, and effective population size than goats. These parameters are fundamental for the management of the populations, both when pursuing their conservation and genetic improvement, because they can influence the accuracy of inbreeding and parentage estimates. The integration of the genomic tools with the more traditional ones represents an important upgrade that allows to overcome unreliable or incomplete pedigree data. Moreover, the collection of new phenotypes led to the development of breeding values that meet the demands for greater guarantees in terms of animal welfare and health, paving the way for a more sustainable small ruminant sector. Our results show a low variability of Wel and wide high probability density interval (HPD) of the hereditability estimates of Wel, Loc, and BCS. With the aim to develop an accurate welfare breeding evaluation we will review our results and the welfare-related traits collected and analyzed considering the difference between the two species and the diversified management system. All the activities undertaken within the CHEESR project, including the phenotypic and genotypic characterization for the study of biodiversity and the development of genomic breeding values, will continue with the *SHEEP&GOAT* project (supported by EAFRD for the period 2021-2023): through the genotyping of all the Italian small ruminant breeds present in at least three registered farms, this project will help to better portray the situation of the whole sector and will contribute to define national strategies to safeguard the Italian biodiversity.

References

Cao J., Baumung R., Boettcher P., Scherf B., Besbes B., *et al.* (2021) Sustainability 13(2):775. https://doi.org/10.3390/SU13020775

Ciani E., Crepaldi P., Nicoloso L., Lasagna E., Sarti F.M., *et al.* (2014) Anim. Genet. 45:256–266. https://doi.org/10.1111/age.12106

Cortellari M., Barbato M., Talenti A., Bionda A., Carta A., *et al.* (2021a) Sci. Rep. 11:10986. https://doi.org/10.1038/s41598-021-89900-2

Cortellari M., Bionda A., Negro A., Frattini S., Mastrangelo S., *et al.* (2021) Genet. Sel. Evol. 53:92. https://doi.org/10.1186/S12711-021-00685-4

Mastrangelo S., Ciani E., Sardina M.T., Sottile G., Pilla F., *et al.* (2018) Anim. Genet. 49:71–81. https://doi.org/10.1111/age.12634

Nicoloso L., Bomba L., Colli L., Negrini R., Milanesi M., *et al.* (2015) Genet. Sel. Evol. 47:62. https://doi.org/10.1186/s12711-015-0140-6

82. How can breeding for improved animal welfare meet societal expectations and challenges?

A. Wallenbeck

Swedish University of Agricultural Sciences, P.O. Box 7068, 750 07 Uppsala, Sweden; anna.wallenbeck@slu.se

Abstract

Animal production can play an important role in meeting many major societal challenges worldwide. Moreover, expectations from consumers and society on improved welfare in production animals relate to all parts of the production process, including breeding. The welfare of production animals is affected, both positively and negatively, by breeding techniques and by the genetic changes accomplished through breeding. Development of broad breeding goals, including animal welfare traits, taking the preferences of stakeholders in a broad sense into account, would improve the possibilities to meet societal expectations and challenges. A major challenge for future breeding for improved animal welfare meet societal expectations and challenges is to breed animals suitable for unknown future production environments. Examples of breeding options are inclusion of animal welfare-related traits in genomic selection using information from automatic surveillance and development of associative breeding.

Introduction

Public and societal interest in animal welfare is high (Eurobarometer, 2016) and consumer awareness of animal welfare in production animals is increasing (EconWelfare, 2011; Leonardsson, 2011). Animal welfare, i.e. the subjective experience of the individual animal, its biological functioning and its adaptation to its current environment (e.g. Fraser *et al.*, 1997), encompasses a combination of aspects relating to animal health, physiology and behaviour. Expectations from consumers and society on improved welfare in production animals affect all parts of the production process, including breeding. In a growing number of cases, consumer and public opinion are influencing the choice of breeding animals in modern animal production, e.g. the Belgium Blue cattle breed has been banned in Sweden and there are restrictions on the use of fast-growing broiler hybrids in the Netherlands and Norway. Thus, animal production with a high focus on improved welfare is no longer a small niche market for farmers, or for breeding companies.

Animal production, including the aspects of animal welfare and breeding, plays an important role some of the major societal challenges in the world today. Goals set by the United Nations in 2015, in response to the major global societal challenges, include a world without poverty and hunger, and safe from the worst effects of climate change and loss of biodiversity by 2030 (United Nations, 2015). These Sustainable Development Goals (SDGs) have a wide scope and production animals are rarely mentioned, while animal welfare is not mentioned at all. However, mutual benefits from improving animal welfare and achieving the SDGs have been identified, indicating a positive role of improved animal welfare in meeting the SDGs (Keeling *et al.*, 2019; Olmos Antillón *et al.*, 2021).

This paper describes the effects of breeding on the welfare of production animals and examines ways in which future breeding for improved animal welfare can meet societal expectations and challenges.

Breeding for improved animal welfare

Effects of breeding on animal welfare. Animal breeding involves designing a breeding programme, selecting the parents for the next generation and developing techniques related to breeding, such as artificial insemination (AI), embryo transfer (ET), sexing of semen, detailed analyses of DNA information and gene modification. Several of these techniques have consequences for animal welfare, both positive and

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_82, © A. Wallenbeck *et al.* 2022

negative. One example is the strategic use of sexed semen to improve animal welfare by avoiding culling of many young animals (e.g. male offspring with low economic value in egg and dairy production). Another example is the rapid decrease in the spread of diseases when AI was introduced in livestock breeding and mixing of animals from different herds for breeding purposes was no longer necessary. However, AI and ET may include matching of breeds with large size differences, which can result in birthing problems if the offspring is too large for the mother. Moreover, DNA information has been used for many decades for elimination of diseases, and thus improved animal welfare.

The traditional aim of modern breeding programmes for production animals is to increase production efficiency, e.g. more milk, eggs or lean meat growth per animal and/or feed unit. The direct effect has been a remarkable increase in production efficiency (Van der Steen *et al.*, 2005; NAV, 2018), and thus a reduction in the environmental load per unit of product. However, in cases with unfavourable genetic correlations between production and animal welfare-related traits and where animal welfare-related traits are not monitored, it can lead to deterioration in animal welfare.

Towards sustainable breeding – for improved animal welfare, production efficiency and low environmental load. During recent decades, there has been a general trend for broadening breeding goals to include animal welfare-related traits such as health, temperament, survival and longevity. These broader breeding goals have led to improvements in production, resource efficiency and animal welfare simultaneous. For example, there have been improvements in milk yield, udder health, feed efficiency and general health in dairy cows (e.g. NAV, 2018), leg health, feed efficiency and growth rate in pigs (e.g. Aasmundstad *et al.*, 2014) and foot pad dermatitis, feed efficiency and growth in broilers (e.g. Kapell *et al.*, 2012).

Sustainable development of animal production, which includes the animals in the system, needs to consider economic, environmental and social aspects (e.g. Dockes *et al.*, 2011). Increased emphasis on achieving the UN SDGs and increasing consumer and societal interest in animal welfare require the development of long-term sustainable breeding strategies. Sustainable breeding strategies need to consider the requirements of all stakeholders in the food chain, e.g. farmers, retailers and consumers (Gamborg *et al.*, 2005). An important step in the development of breeding goals is to identify traits that are important and which should be included in the breeding evaluation (Boelling *et al.*, 2003).

Decisions on the traits included in modern breeding goals are generally based on expert opinion and what is currently practically and economically feasible. They are more seldom based on farmers' preferences. Although all stakeholder groups are important in the process of developing sustainable breeding strategies, farmers have essential knowledge about how animals are functioning on their farms and they ultimately decide which animals they will use and what genetic material they will purchase. In a study investigating the breeding traits perceived by dairy and pig farmers to affect profitability, environmental load and animal welfare, both pig and dairy farmers considered that all suggested breeding traits have impacts on profitability (Wallenbeck *et al.*, 2013). There were also similarities between pig and dairy farmers regarding traits considered to be related to environmental load and traits related to animal welfare. In the same study, genetic simulations indicated that a breeding goal aiming at decreased environmental impact would lead to genetic improvement in environmental and production traits, while a breeding goal aiming for improved animal welfare would lead to genetic improvement in animal welfare traits without affecting productivity traits (Wallenbeck *et al.*, 2015). Overall, that study showed that it is theoretically possible to design breeding goals which simultaneously improve animal welfare, decrease environmental load and improve production efficiency

Modern breeding is continually developing new approaches, e.g. genomic selection by improved statistical modelling, with the aim of increasing the rate of genetic improvement. Increased genetic change over time can improve the genetic potential for improvements in animal welfare traits as long as these traits are included in the breeding goals or have favourable genetic correlations to traits included in the breeding goals.

Future challenges and ways in which future breeding for improved animal welfare can meet societal expectations and challenges

Breeding for future production environments. Societal expectations and challenges will influence the development of future animal production systems, and thus the context for animal breeding. If evaluation and selection of parents is performed in a production environment diverging from that of commercial production animals, genotype by environment interactions may occur, with a particular phenotypic trait being partly governed by different genes in different environments. One way to take variation in production environments into account is to breed for genetic placidity. A major challenge for animal breeding is to cater for future production environments influenced by e.g. climate change and animal welfare demands.

Including new animal welfare traits. In recent years, housing and management practices aimed at improving animal welfare and methods for automatic surveillance of animal welfare indicators, e.g. methods related to precision livestock farming, have been developed. Examples include processing information derived from video recordings, automatic feeding systems, milking robots and activity monitors, while physiological parameters such as biomarkers are also relevant animal welfare indicators. In parallel, breeding technologies have been developed and improved. There are great possibilities to include animal welfare traits derived from automatic surveillance in both genetic and genomic selection (Brito *et al.*, 2020). Genomic selection, with the opportunity to include a larger number of traits in the selection, has particular potential, as the concept of animal welfare is broad (covering animal health, physiology and behaviour). Thus further development requires a transdisciplinary approach, including e.g. genetics, immunology, physiology, computer science, ethology and animal welfare science.

Another area with great opportunities to improve the genetic potential for animal welfare-related traits is associative breeding focusing on social (or associated) genetic effects, i.e. selecting animals based on their genetic potential to affect their group members (Muir, 2005; Bijma, 2007). Associative breeding is of particular interest in attempts to select against injurious social behaviour. The method is promising for improving production and welfare simultaneously, and captures the entire behavioural profile of the animals as it considers the consequences of their behaviour, not their behaviour as such, in the evaluation (Turner *et al.*, 2018).

References

Aasmundstad T., Gjerlaug-Enger E., Grindflek E., Vangen O. 2014. Genetic trends of conformation traits and genetic correlations to osteochondrosis in boars. Animal, 8, 1045–1052.

Bijma, P., Muir, W., Van Arendonk, J.. 2007. Multilevel selection 1: Quantitative genetics of inheritance and response to selection. Genetics. 175(1), p 277-288.

Boelling, D., Groen, A.F., Sorensen, P., Madsen, P. & Jensen, J. (2003). Genetic improvement of livestock for organic farming systems. Livestock Production Science 80(1-2), 79-88.

Brito, L.F., Oliveira, H.R., McConn, B.R., Schinckel, A.P., Arrazola, A., Marchant-Forde, J.N., Johnson, J.S. 2020. Large-Scale Phenotyping of Livestock Welfare in Commercial Production Systems: A New Frontier in Animal Breeding. Front Genet. 2020 Jul 31;11:793. https://doi.org/10.3389/fgene.2020.00793. PMID: 32849798; PMCID: PMC7411239.

Dockes, A.C., Magdelaine, P., Daridan, D., Guillaumin, A., Remondet, M., Selmi, A., Gilbert, H., Mignon-Grasteau, S., Phocas, F., 2011, Farming and breeding goals for sustainable animal breeding - views and expectations of stakeholders from the production chains and the general public. Inra Prod. Anim. 24, 285-296.

EconWelfare. 2011. Findings, conclusions and recommendations concerning effective policy instruments in the route towards higher animal welfare in EU. D.05.

Eurobarometer. 2016. Attitudes of Europeans towards Animal Welfare. Special Eurobarometer 442.

Fraser, D., Weary, D.M., Pajor, E.A., Milligan, B.N. 1997. A scientific conception of animal welfare that reflects ethical concerns. Animal Welfare. 6, 187-205.

Gamborg, C., Sandøe, P. 2005. Sustainability in farm animal breeding: A review. Livest. Prod. Sci. 92, 221–231.

Kapell, D. N. R. G., W. G. Hill, A. M. Neeteson, J. McAdam, A. N. M. Koerhuis, and S. Avendano. 2012. Genetic parameters of foot-pad dermatitis and body weight in purebred broiler lines in 2 contrasting environments. Poult. Sci. 91:565–574.

Keeling, L., Tunón, H., Olmos Antillón, G., Berg, C., Jone,s M., Stuardo, L., Swanson, J., Wallenbeck, A., Winckler, C. and Blokhuis, H. 2019 Animal Welfare and the United Nations Sustainable Development Goals. Front. Vet. Sci. 6:336.

Leonardsson. 2011. Trends and future developments in animal welfare. ENCAW 6.

Muir, W. 2005. Incorporation of competitive effects in forest tree or animal breeding programs. Genetics. 170(3), p 1247-1259.

NAV. 2018. Nordic cattle genetic evaluation. http://www.sweebv.info/ba52nycknav.aspx

Olmos Antillón, G., Tunón, H., de Oliveira, D., Jones, M., Wallenbeck, A., Swanson, J., Blokhuis, H., Keeling, L. 2021. Animal Welfare and the United Nations' Sustainable Development Goals – Broadening Students' Perspectives. Sustainability 2021, 13, 3328.

Turner, S. P.; Camerlink, I.; Baxter, E. M.; D'Eath, R. B.; Desire, S.; Roehe, R. 2018. Breeding for pig welfare: opportunities and challenges. In: Advances in pig welfare: a volume in herd and flock welfare. Editor: Spinka, M. Elsevier; Oxford; UK

United Nations. 2015. Transforming our World: The 2030 Agenda for Sustainable Development. New York, NY.

Van Der Steen, H. A. M.; Prall, G. F. W.; Plastow, G. S. 2005. Application of genomics to the pork industry. Journal of Animal Science. 83(E), p E1-E8.

Wallenbeck, A., Ahlman, T, Rydhmer, L. and Röcklinsberg, H. 2013. Farmers' views on the impact of breeding traits on animal welfare, productivity and environment. In: Röcklinsberg, H & Sandin, P. (eds). The ethics of consumption – the citizen, the market and the law. Wageningen: Wageningen Academic Publishers. 215-220.

Wallenbeck, A., Mirkena, T., Ahlman, T. and Rydhmer, L. 2015. Dairy breeding programs focusing on animal welfare and environment – a simulation study. Proceedings of the 66th Congress of the EAAP, Warszawa, Poland.

83. Swine inflammation and necrosis syndrome (SINS) and its association with biting behavior after weaning

N.L. Leite[1*], E.F. Knol[2], S. Nuphaus[2], S. Tsuruta[1], R. Vogelzang[2] and D. Lourenco[1]

[1]University of Georgia, 425 River Rd, 30602, Athens, USA; [2]Topigs Norsvin, Schoenaker 6, 6641 SZ Beuningen, the Netherlands; nataliagaloro@uga.edu

Abstract

In this study, we estimated genetic (co)variance components of the swine inflammation and necrosis syndrome (SINS), postweaning biting damage (BD), and several production traits in a three-way crossbred population. Additionally, we evaluated the benefits of modeling BD using social interaction models in an F1 gilt population. Our results showed that SINS and BD have substantial genetic variances and are moderately and positively correlated. That suggests that piglets genetically more prone to be born with SINS signs are also more likely to suffer from biting after weaning. Unfavorable correlations between SINS and BD with important production traits, such as backfat thickness and loin depth, require close attention. Modeling BD using social interaction models increased the prediction accuracy of breeding values, reduced bias, and increased the total genetic variation by 2.5 folds, thus highlighting better possibilities for breeding for overall pig welfare.

Introduction

The swine inflammation and necrosis syndrome (SINS) results from a blood flow constraint of capillaries, leading to loss of skin function at the far ends of body parts, such as the tail, ear, and teats. This syndrome can affect up to two-thirds of all newborn piglets, but it can also be observed later in a pig's life. Under such conditions, pigs might suffer from a chronic reduction of their welfare and productivity throughout the growing process (Reiner *et al.*, 2021). Besides the direct effects of SINS, affected animals may be potentially more susceptible to bites after weaning because affected parts of the body, such as the tail and ears, can trigger the suckling and biting behavior of pen mates.

Biting behavior challenges the pig industry in coping with the increasing animal welfare standards. Besides the clear welfare issue, damaged animals might have decreased productivity and be more susceptible to infectious diseases due to open wounds, undoubtedly affecting individual and herd health status. Decreasing biting behavior is needed and might be achieved through improvements in the environment and genetic selection (Schrøder-Petersen *et al.*, 2001).

Biting behavior can be phenotypically expressed as a combination of, for instance, tail, ear, and vulva biting, which are known to result from animal social interaction. Although social interaction is usually taken as a complete environmental effect, animals might behave differently towards their pen mates due to their different genetic backgrounds. The genetic effect of an animal over the phenotype of others in the same group is called the indirect genetic effect, and social interaction models can account for it (Bijma *et al.*, 2007).

Accounting for indirect genetic effects using social interaction models might increase the total genetic variance of a trait and, consequently, improve the rate of genetic gain. Moreover, it could help manage negatively correlated responses in situations where the direct and indirect effects are negatively correlated (Bijma *et al.*, 2007).

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_83, © N.L. Leite *et al.* 2022

Our objectives were to: (1) estimate genetic (co)variance components of the swine inflammation and necrosis syndrome (SINS), postweaning biting damage (BD), and several production traits; (2) evaluate the benefits of modeling BD using social interaction models.

Materials & methods

Genetic relationship between SINS, BD, and production traits. A total of 5,960 three-way crossbred piglets born from 80 sires and 240 dams were evaluated. Piglets presenting any signs of teat, claw, ear, or ear SINS at birth were scored as 2, and piglets with no SINS signs were scored as 1. Information on birth weight (BW), weaning weight (WW), BD (1 or 2), backfat carcass thickness (BF), and loin depth (LOD) were also recorded. Biting damage was recorded on the transferring day from the nursery to the finishing phase (9 weeks of age). The binary codification for BD was chosen because of the low incidence of the trait and the inference problems associated with the evaluation of traits with multiple categories with small variance. Four multiple-trait animal models with combinations of SINS, BD, and an alternative production trait were used to assess trait heritabilities and genetic correlations. Alternative traits were BW (Model A), WW (Model B), BF (Model C), and LOD (Model D). The maternal effect was included only in the BW, WW, and SINS models. Normality was assumed for all traits, and the BLUF90 family of programs was used for further analyses (Misztal *et al.*, 2014).

Modeling BD using social interaction models. A total of 46,340 F1 gilts were scored for BD (1 or 2) at the end of the growing phase (21 weeks of age). Gilts were housed in similar commercial conditions and housed in groups of 10.7±0.7 individuals. Two animal models were used for variance component estimation: an animal model only accounting for direct genetic effects (Model 1) and a second model extended with indirect effects (Model 2), as described as follows:

$$y=X\beta+Z_Da_D+Wc+Fg+e \quad (1)$$

$$y=X\beta+Z_Da_D+Z_Sa_S+Wc+Fg+e \quad (2)$$

where y is the vector of phenotypes, β is the vector of fixed effects of the scorer, line, and age covariate, a_D is the vector of direct genetic effects, a_S is the vector of indirect genetic effects, c is the vector of the common environment of litter, g is the vector of group effects, and e is the vector of residuals; X, Z_D, Z_S, W, and F are incidence matrices for the effects in the vectors β, a_D, a_S, c, and g, respectively.

A forward validation was performed to evaluate the prediction accuracy, bias, and dispersion of both models. The validation set was composed of 14,710 young gilts born from 2020 to 2021 that had classical breeding values (EBV; Model 1) and total EBV (TBV; Model 2) estimated with a complete dataset (including all phenotypic information available) and with a partial dataset (excluding phenotypes of the validation set). TBV were calculated as:

$$TBV = a_D + a_S(N-1) \quad (3)$$

where N is the average number of pen mates in the population.

Both EBV and TBV from partial and complete datasets were later used to calculate prediction accuracy, dispersion, and bias of both models following the Linear Regression (LR) method (Legarra *et al.*, 2018). Phenotypes were assumed to follow a normal distribution, and analyses were performed using the BLUF90 family of programs (Misztal *et al.*, 2014).

Results

Genetic relationship between SINS, BD, and production traits. Estimates of direct heritabilities and genetic correlations in the four multiple-trait models are presented in Table 1. The direct heritability of SINS ranged from 0.15 to 0.24, whereas heritabilities for BD were lower, ranging from 0.05 to 0.06. Direct genetic correlations between SINS and BD were moderate and positive (from 0.19 to 0.50). The direct effect of SINS was negatively correlated with BW (-0.30), WW (-0.40), and LOD (-0.16) but not significantly correlated with BF (0.05). BD was strongly correlated with WW (-0.61), BF (-0.84), and LD (0.76) but not significantly correlated with BW (-0.11).

Modeling BD using social interaction models. BD heritability in Model 1 was low at 0.04, whereas the total heritable variance in Model 2 was about 2.5-fold greater (0.10). The genetic correlation of direct and indirect effects was not significantly different from zero (-0.06). Validation results are shown in Table 2. Model 2 showed a substantial advantage in prediction accuracy (0.29) over Model 1 (0.20), which did not account for indirect genetic effects. Bias was increased for Model 2 compared to Model 1, whereas dispersion was greatly improved by the addition of indirect effects in Model 2.

Discussion

Genetic relationship between SINS, BD, and production traits. The low but significant heritability estimates for SINS and BD indicate that overall welfare improvement through selection is possible in the long term. The intermediate direct genetic correlation between SINS and BD indicates that piglets genetically more prone to be born with SINS signs are also more likely to suffer from biting after weaning. Such correlations can be associated with the triggering aspect of the necrosed skin on the suckling and biting behavior of pen mates after weaning.

Table 1. Direct heritabilities (diagonal) and genetic correlations (off-diagonal) between SINS, BD, and production traits.

Model 1				Model 2			
	SINS	**BD**	**BW**		**SINS**	**BD**	**WW**
SINS	0.21 (0.02)[1]	0.19 (0.08)	-0.30 (0.08)	SINS	0.23 (0.03)	0.23 (0.09)	-0.40 (0.08)
BD		0.05 (0.02)	-0.11 (0.09)	BD		0.05 (0.01)	-0.61 (0.06)
BW			0.07 (0.01)	WW			0.11 (0.02)
Model 3				**Model 4**			
	SINS	**BD**	**BF**		**SINS**	**BD**	**LOD**
SINS	0.24 (0.03)	0.43 (0.07)	0.05 (0.05)	SINS	0.15 (0.02)	0.50 (0.09)	-0.16 (0.07)
BD		0.06 (0.01)	-0.84 (0.04)	BD		0.05 (0.01)	0.76 (0.06)
BF			0.63 (0.07)	LOD			0.42 (0.06)

[1] Estimate standard error.

Table 2. Breeding value prediction accuracy, bias, and dispersion in Models 1 and 2.

	Model 1	**Model 2**
Prediction accuracy	0.20	0.29
Bias	-0.27	-0.55
Dispersion	0.46	0.91

The negative genetic relationship between SINS and early growth traits (i.e. BW and WW) indicates that, at the piglet genetic level, a high incidence of SINS may be associated with low birth and weaning weights. However, there might be a low or no significant association between SINS signs at birth and carcass traits (i.e. BF and LOD), possibly due to the long separation in time scale between such traits.

BD was highly and unfavorably correlated with economic production traits, such as BF and LOD, making efforts for selecting against BD even more critical in pig breeding programs. On the other hand, BD was favorably correlated with WW, indicating that piglets with better genetic potential for growth during the lactation stage might be less genetically prone to be injured by biters during hierarchical rearrangements after weaning.

Modeling BD using social interaction models. Our results underline important genetic components on BD behavior in growing F1 gilts. Modeling indirect genetic effects substantially increased the trait total genetic variance, indicating that the genetic component on the behavior of animals over their pen mates could substantially increase opportunities for genetic improvement against biting behavior. The nonsignificant correlation between direct and indirect effects suggests that neither antagonist nor collaborative behavior among animals exists for BD. Therefore, a selection index for overall BD improvement should consider both direct and indirect effects.

The greater genetic variance obtained through modeling indirect effects with social interaction models combined with the associated higher prediction accuracy and smaller dispersion of breeding values highlights an opportunity for more accurate BD predictions and, therefore, faster genetic improvement.

Overall, enough genetic variation exists for SINS and BD; thus, improving overall welfare through genetic selection is feasible. Selecting animals less genetically prone to developing SINS is also expected to positively affect piglet genetics for heavier weaning weight and reduced BD. The strong and unfavorable genetic correlation between BD and important production traits under selection deserves close attention. Accounting for indirect effects of BD using social interaction models might greatly increase possibilities for faster and more accurate genetic improvement.

References

Bijma P., Muir W.M., and Van Arendonk J.A.M. (2007). Genetics. 175(1), 277-288. https://doi.org/10.1534/genetics.106.062711

Legarra A., and Reverter A. (2018). Genetics Selection Evolution. 50(1), 1-18. https://doi.org/10.1186/s12711-018-0426-6

Misztal I., Tsuruta S., Lourenco D., Aguilar I., Legarra A *et al.* Manual for BLUPF90 family of programs – Available at: http://nce.ads.uga.edu/wiki/lib/exe/fetch.php?media=blupf90_all2.pdf

Reiner G., Kuehling J., Loewenstein F., Lechner M., and Becker S. (2021). Animals 11(6):1670. https://doi.org/10.3390/ani11061670

Schrøder-Petersen D.L., and Simonsen H.B. (2001). The Veterinary Journal. 162(3), 196-210. https://doi.org/10.1053/tvjl.2001.0605

84. Breeding for future social license for beef production

S.P. Miller[1,2*] and K.J. Retallick[2]

[1]AGBU, a joint venture of NSW Department of Primary Industries and University of New England, 2351, Armidale, Australia; [2]Angus Genetics Inc., 3201 Frederick Ave., Saint Joseph 64506, USA; steve.miller@une.edu.au

Abstract

Examples from Angus cattle provide evidence of positive genetic trends for production traits and, with increasing frequency, for traits more directly related to welfare outcomes. Genetic improvement programs are optimized when clear market signals linked to the commercial sector are in place. When welfare related traits like foot structure were added to profit driven selection indexes, recording levels increased and the accuracy of the evaluation for these traits improved. It is postulated that effective recording and genetic selection for improved welfare should be of interest to consumers. When faced with an informed choice, consumers should be more attracted to beef products which are sourced from production systems where through selection, welfare issues such as dystocia, lameness and reduced longevity are actively minimized. Beef breeders should be making efforts now to ensure a future 'social license' for their products by preparing for such economic feedback in the future.

Introduction

Beef cattle have been improved considerably through selection, resulting in cattle which better meet consumer demands, generate greater profits for breeders and are more competitive as a protein source. However, a focus on existing breeding objectives may not ensure future success for beef production. Breeding priorities for the future will need to address changing consumer demands.

Consumers are becoming more informed and more critical of their purchases across a range of products, with ethical considerations and environmental impacts being increasingly important. It is expected other papers will address reducing cattle's impact on the environment at this congress. This paper will focus on genetic solutions to improve welfare, enabling what is described as a future 'social license'. The term social license refers to the consumer acceptance of the practises underpinning beef production. Although the concept of a social license to produce beef is a relatively new one, this paper highlights genetic improvement initiatives that are already well established and address aspects of animal welfare. Thoughts are put forward as to how direct consumer pressure in the marketplace could accelerate the adoption, development, and genetic improvement of welfare related traits.

Materials & methods

To illustrate genetic improvement and changes to recording practices over time genetic trends and counts of performance records were acquired from the American Angus Association's as used in the National Cattle Evaluation (American Angus Association 2022).

Results

Recent genetic trends for traits influencing animal welfare, along with selected production traits are presented in Figure 1 as are the changes in trait recording levels.

Discussion

The genetic trends in Figure 1 demonstrate improvement across a range of traits, which has been supported by the levels of phenotype recording, also presented. How these trends relate to improving welfare will be discussed for each trait separately.

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_84, © S.P. Miller and K.J. Retallick 2022

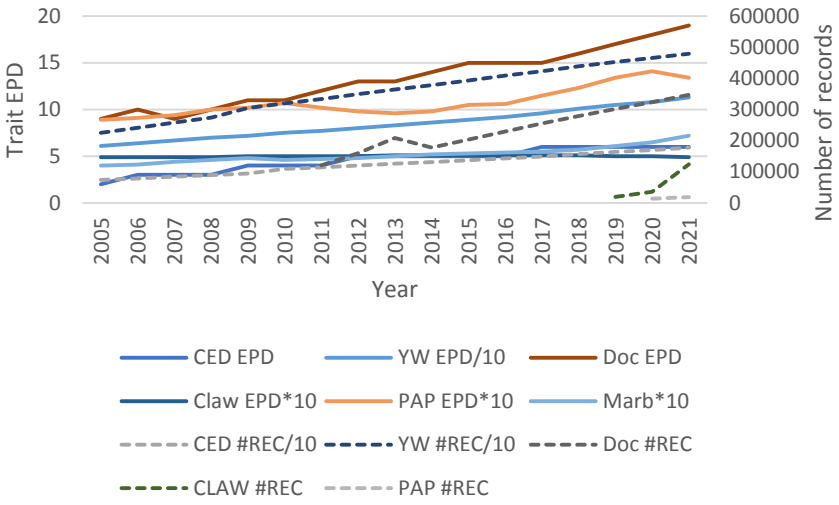


Figure 1. Genetic trends and records available for genetic evaluation for selected traits in American Angus cattle. Traits illustrated are: Calving Ease Direct (CED), Yearling Weight (YW), Docility (Doc), Foot Claw Set (Claw), Pulmonary Arterial Pressure (PAP), Carcass Marbling (Marb). The number of records used in the evaluation from all sources including Canada and Australia, where applicable (#REC).

Calving difficulty. A remarkable thing about the American Angus genetic trends is the continued increase in growth while improving calving ease (reducing dystocia), given the positive genetic correlation between post-weaning gain and birth weight (0.29). Through effective multi-trait selection, aided by the application of economic selection indexes, the Angus breed is now the heaviest breed at a year of age, while also being one of the leading breeds for calving ease as evaluated by the USDA (Kuehn and Thallman 2020). Commercial producers in the USA are conscious of selecting sires that will present very low dystocia rates. There is no doubt dystocia is an animal welfare concern, and it should be recognized that selection for improved calving ease plays a significant role in improving animal well-being.

Docility. The first Docility EPD were presented in 2011 based on 120,770 records and there has been steady improvement in both genetic merit and recording of the trait since. Docility also ranked high in the American Angus survey from 2019 (Santos *et al.*, 2019), but this was also noticeable in interviews involving large processors and feedlots, which indicated that cattle with better temperament were better for their businesses. The docility EPD was added to the maternal component of Angus economic selection indexes ($Values) in 2019. The emphasis has been on improving temperament for the betterment of the farmer, but it should be realized that cattle with better temperament also have better welfare outcomes with *inter alia* less stress and bruising.

Pulmonary arterial pressure (PAP). A disease which affects cattle at higher altitudes in the USA (typically over 1,800 m) is caused by right-heart failure and severe cases lead to death. It is a heritable condition that can be measured with the related trait of Pulmonary Arterial Pressure and significant research has been done on this trait by scientists at Colorado State University (Crawford *et al.*, 2016). Despite the complexity of recording the trait, the PAP measure has been adopted in the region. In 2019, the first EPD were released for PAP, including genomic prediction through the weekly single step evaluation, enabling selection for this difficult trait for breeders across the country. Speaking directly with farmers with affected stock, this disease puts stress on the cattle in their care but also the mental wellbeing of the farmers who need to deal with these losses.

Foot Score. Improving foot structure remains a high priority for breeders as demonstrated by recent survey work (Santos *et al.*, 2019). For this reason, there was much interest in the launch of EPDs for claw set and foot angle in 2018. The evaluation has seen a significant increase in foot score recording since the foot score EPD were incorporated into the economic multiple trait selection indexes ($Values) in 2019. The records contributing to EPDs spiked in 2020 due to a new joint evaluation incorporating records from Angus Australia as described by Alvarenga *et al.* (2022) in a paper submitted to this congress. Cattlemen are searching for 'trouble free' cattle and those with better foot structure will suffer less from lameness and are less likely to be culled. It is clear that improving foot structure is a 'step' towards better animal welfare.

Hair Shedding. Based on work from the University of Missouri (Durbin *et al.*, 2020) a hair shedding research EPD was first released in 2020, based on analyses of 14,396 records and a 2021 re-evaluation involved 22,564 hair shedding phenotypes. The phenotype is based on a simple score (1-5) describing the proportion of winter coat an animal has shed, typically recorded in late spring. This EPD is of most interest to breeders in hotter states, and those in areas that suffer from endophyte infected fescue. With a warming environment, and Angus cattle pushing into hotter regions, such efforts to improve heat tolerance will remain important.

Longevity. Recent research has investigated the potential of a random regression model to estimate breeding values for longevity in Canadian and American Angus but more work is required to further develop a research EPD for release to the membership (Oliveira *et al.*, 2021). In an effort to further improve the data structure to determine longevity, changes have been made to encourage more whole-herd inventory-based reporting from members. Improving longevity will contribute to improved welfare and enhance social license as cows will be less likely to be culled early due to health and structural problems.

Fostering selection for improved animal welfare

Reasons behind breeders' choice to record and select for the traits impacting welfare is likely two-fold. Breeders are also farmers; they naturally want to maintain and improve the health of animals in their care. An example of this not mentioned above is the extensive genotyping Angus breeders have undertaken to reduce the incidence of recessive genetic conditions that result in dead or deformed calves at birth. The other reason for this selection pressure, also pointed out above, is the relationship between these EPDs, $Values and prices for the bulls they sell. By including welfare related traits in the economic selection indexes, which contribute significantly to bull prices, selection of welfare related traits naturally follows.

Proposed here is an additional avenue to put greater economic incentive on animal welfare traits. Afterall, breeding is inevitably about trade-offs. The genetic progress for growth and carcass above is obvious. Any pressure on welfare traits will reduce the progress for these big economic drivers. Welfare traits can become more important economic drivers if more incentives are available in the commercial sector. Work on profit-based selection indexes makes it clear that traits related directly to carcass value, such as marbling in Angus cattle in America, naturally receive significant weighting. If welfare related traits were more directly related to profit in the commercial industry, this would influence selection pressure in the seedstock sector.

To date consumer pressure on production related practices has been evident in other industries, and the objection to caged birds in the laying industry is a good example. Now consumers in developed countries are typically faced with a wide array of choices for eggs, with cage-free, free-range, etc. being predominant options at a premium to cage-laid eggs. This is a typical example of how the marketplace is impacted by welfare concerns, in a straightforward categorical way. The result here being pressure to reduce cage-based production systems.

In beef cattle we could see welfare related traits being addressed in a similar manner. Horns vs polled is a good example, with de-horning becoming an unacceptable practice when there is a genetic alternative available. A breeding program to increase the frequency of polled in the commercial tier is described by Johnston *et al.* (2022) in a paper submitted to this congress. With this economic driver looming, as well as obvious husbandry benefits, the prices of 'polled genetics' in breeds that are introgressing this allele show a significant premium.

The genetic incentive for better welfare should not be restricted to such categorical traits. Consider calving difficulty, where a surgical calving is a greater compromise to welfare than horns disbudded at birth. Like horns, the calving problem could have been avoided in most cases with selection. If consumer pressure can promote differential prices for things like 'cage-free' in hens and perhaps polledness in the future, does it not make sense that this could be extended to all traits impacting welfare? Consider commercial producers purchasing bulls facing the choice between Bull A, which has no genetic description of welfare related traits (as is common in the marketplace now) and Bull B, which is from a seedstock breeder who actively records, and makes EPDs for welfare related traits available to buyers. Should the producer purchasing from supplier B not be rewarded for the due diligence demonstrated to use the tools available to improve animal welfare at the genetic level?

It is argued here that this should be given similar emphasis to moving hens from cages or eliminating horns. If the economic incentive was in the commercial marketplace to reward such selection practices, the increase in recording, selection pressure, and development of new traits could be significant. It has been demonstrated time and again that the greatest incentives for change in cattle breeding always come from the commercial marketplace. Cattle breeders are making good progress in improving welfare related traits now, but if greater direct pull from the commercial sector was present, through demand from consumers, this would improve much faster. It seems such pressure will come and the seedstock sector should be proactive in getting ahead of this demand to ensure a future 'social license' for their products.

References

American Angus Association (2022) Angus National Cattle Evaluation. Available at: https://www.angus.org/Nce/Heritabilities

Crawford N.F.,Thomas M.G.,Holt T.N.,Speidel S.E. *et al.* (2016) J.Anim.Sci. 94(11):4483–4490. https://doi.org/10.2527/jas.2016-0703

Durbin, H.J., Lu D., Yampara-Iquise H., Miller S.P. and Decker J.E. (2020) Genet. Sel. Evol. 52:1 https://doi.org/10.1186/s12711-020-00584-0

Kuehn, L and Thallman M. (2020) https://beefimprovement.org/wp-content/uploads/2021/01/21_ABEPDfactsheet.pdf

Oliveira H.R., Brito L.F., Miller S.P. and Schenkel F.S. (2020) Animals 10(12):2410 https://doi.org/10.3390/ani10122410

Santos B., Archer J.A., Martin-Collado D., Quinton C.D., Crowley J. *et al.* (2019) Proc. of the 23rd AAABG, Armidale, Australia.

85. Selection for piglet mortality and litter size in outdoor organic pig production systems

T.T. Chu[1,2*], R.M. Zaalberg[1], J. Jensen[1] and T.M. Villumsen[1]

[1]Aarhus University, Center for Quantitative Genetics and Genomics, 8830 Tjele, Denmark; [2]Vietnam National University of Agriculture, Faculty of Animal Science, Vietnam; chu.thinh@qgg.au.dk

Abstract

This study evaluated breeding goals for increasing litter size of sows 11 days after farrowing (LS11) in organic production systems while aiming to reduce piglet mortality (Mort11). Investigated breeding goals included LS11 only, or Mort11 and total number of piglets born (TNB) combined. Response to selection, expressed as LS11, was derived using selection index theory. We found that selection based on breeding goal of LS11 only, or using equal economic weights for TNB and Mort11 would increase LS11, but at the costs of increasing piglet mortality. By defining LS11 as two traits of TNB and Mort11, selection could reduce piglet mortality through possibility of altering economic weights for the traits in breeding goals, and through exploiting heritability differences of TNB and Mort11 in selection. By increasing economic weight for Mort11, breeding program for organic pig production could achieve zero genetic change in Mort11 while still gaining more TNB.

Introduction

In non-organic Danish pig production, litter size of sows has increased almost linearly from 2003 to 2019 (Nielsen *et al.*, 2013; Hansen, 2020). In the same period, mortality rate of piglets decreased steadily until 2016, but showed an increasing trend after 2016 (Hansen, 2020). This increase in piglet mortality could be due to a non-linear effect of litter size on mortality traits. However, trait definition in the breeding goal used in some breeding programs could be a problem as well. For example, using litter size at day 5 after farrowing, commonly known as LS5, for selection (Nielsen *et al.*, 2013) might be suboptimal for reducing the mortality of piglets, because this approach assumes the trait total number born per litter (TNB) and mortality traits have the same genetic background. We hypothesized that by decomposing litter size into TNB and a mortality trait, the number of piglets alive can be improved.

Mortality of piglets reduces animal welfare and generates economic losses. Both are major concerns in organic as well as non-organic production systems. However, the welfare concern of piglet mortality is particularly important for the organic system, because consumers and society have high animal welfare requirements and expectations for organic systems (Font-i-Furnols *et al.*, 2019). The use of a trait such as LS5 in selection might increase the number of piglets alive at day 5 after farrowing, but at the cost of an overall increased number of dead piglets. From a welfare point of view, this is not desirable. The trait LS5 is commonly used in the Danish conventional system. In the organic system, number of piglets alive at 11 days after farrowing (LS11) may be more suitable, because of management practices.

The aim of this study was to evaluate breeding goals for increasing number of piglets alive from sows in outdoor organic pig production systems. The specific objectives were to: (1) estimate genetic parameters for sow traits of TNB, number of piglets dead at 0-11 days of piglet age (Mort11) and LS11; (2) evaluate effect of different economic weights for TNB and Mort11 in breeding goals; (3) assess the use of LS11 *versus* TNB and Mort11 in the breeding goals; (4) develop a breeding goal for increasing LS11 without an increase in Mort11.

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_85, © T.T. Chu *et al.* 2022

Materials & methods

Data were collected from the largest organic nucleus and multiplier pig farm in Denmark. Mort11 and TNB were recorded from 3,977 litters of 1,219 sows. The phenotypes of TNB were total number of piglets born per litter of sows. The phenotypes of Mort11 were number of dead piglets per litter that included stillborn piglets and piglets that died from farrowing to 11 days of age. The phenotypes of LS11 were: LS11 = TNB – Mort11. This study considered records of litter size after 11 days of farrowing only, because piglet mortality during this period accounted for the majority (80%) of pre-weaning piglet mortality in the outdoor organic pig production (Chu *et al.*, 2021). In addition, recording of Mort11 was easier and more accurate compared to mortality records at a later period in the outdoor system. Phenotyping of piglet mortality in the Danish outdoor piglet production system is challenging after removing hut fenders, allowing piglets to move freely throughout the field.

For estimation of variance components, pedigree-based linear BLUP models were used. The statistical model for TNB and Mort11 included three fixed effects (breed of the litter, parity of the sow and herd-year-season), two random effects (direct genetic effects of the sow and permanent environment effects caused by sow), and residual term. Bivariate models for TNB and Mort11 were run using the average information restricted maximum likelihood estimation method (DMUAI) in the DMU package (Madsen and Jensen, 2013). Genetic parameters of LS11 were not estimated, but derived from TNB and Mort11. This is because the genetic parameters derived and estimated are mathematically equivalent. We did not want to introduce estimation problem to the comparison of breeding goals for LS11 *versus* TNB and Mort11. In addition, in our preliminary analysis, the use of genetic parameters of LS11 that were either estimated or derived led to the same conclusions for this study.

Phenotypic variance (σ^2_{pLS11}) of LS11 was derived from phenotypic variance estimates of TNB σ^2_{pTNB} and Mort11 $\sigma^2_{pMort11}$:

$$\sigma^2_{p_{LS11}} = \sigma^2_{p_{TNB}} + \sigma^2_{p_{Mort11}} - 2\sigma_{p_{TNB},p_{Mort11}} \quad (1)$$

Other variances and covariances were derived in similar ways.

Responses to selection expressed as LS11 were derived using selection index theory. Investigated breeding goals (H_1 and H_2) were:

$$H_1 = v_{TNB} \times \text{TNB} + v_{Mort11} \times \text{Mort11} \quad (2)$$

$$H_2 = v_{LS11} \times \text{LS11} \quad (3)$$

where v_{TNB}, v_{Mort11} and v_{LS11} are economic weights of trait TNB, Mort11, and LS11, respectively; v_{LS11} =1.

The selection indexes (I_1, I_2) corresponding to H1 and H2 were:

$$I_1 = b_{TNB}\overline{TNB} + b_{Mort11}\overline{Mort11} \quad (4)$$

$$I_2 = b_{LS11}\overline{LS11} \quad (5)$$

where $\overline{TNB}$, $\overline{Mort11}$, and $\overline{LS11}$ are the mean of TNB, Mort11 and LS, respectively with assumption that sows had on average 2 litters (2 repeated records) at the time of selection; b_{TNB}, b_{Mort11} and b_{LS11} are selection indexes.

Expected response to selection and additive genetic change (S_{g_i}) for trait j in the breeding goal were calculated using formula as described in Gibson and Dekkers, (2003). For breeding goal H_2, (S_{g_i} of TNB and Mort11 were derived from correlated response of LS11.

Desired gain indexes were derived as described by Brascamp (1984), so that that the genetic change of Mort11 was zero. After that, economic weight for H_1 in the case of desired gain were derived.

Results & discussion

The mean of TNB per litter was 15.2 with the standard deviation of 4.0. The mean of Mort11 per litter was 2.3 with the standard deviation of 2.4. Overall piglet mortality rate from pre-farrowing to 11 days of piglet age was 15%. This value was lower than the mortality rate reported in Denmark for organic herds (~27% (Rangstrup-Christensen *et al.*, 2018)) and non-organic herds (~20% (Hansen, 2020)). The lower observed mortality level was likely due to the high level of management on the study farm.

Figure 1 shows the data on dead piglets collected for this study, of which about 62% were stillborn piglets and piglets that died before 1 day of age. It was difficult to distinguish between stillborn and dead at 0 or 1 day of age as herdsman might not be present at the day of birth delivery. About 10 to 11 days of piglet age, hut fenders were removed. At this time, uncounted dead piglets were recorded, thus we could see a slightly higher piglet death compared to day 5 to 8.

Heritability and genetic variances of TNB were higher than those of Mort11 although these differences were not statistically significant (Table 1). Genetic correlation between TNB and Mort11 was positive. In other words, the larger litter size of sows, the more piglets died. Parameters of LS11 were derived from estimates of bivariate models of TNB and Mort11. Genetic correlation between LS11 and TNB were highly positive. The genetic correlation between LS11 and Mort11 was also positive, but at low level.

Response to selection in units of LS11 was highest when the breeding goal of selection included TNB and Mort11 traits with an equal economic weight (Table 2). However, the potential gain in LS11 for this breeding goal came from increasing additive genetics of TNB, while at the same time piglet mortality increased. These differences in potential genetic changes for TNB and Mort11 could be explained by the difference in genetic variation, and the positive correlation between the two traits (Table 1).

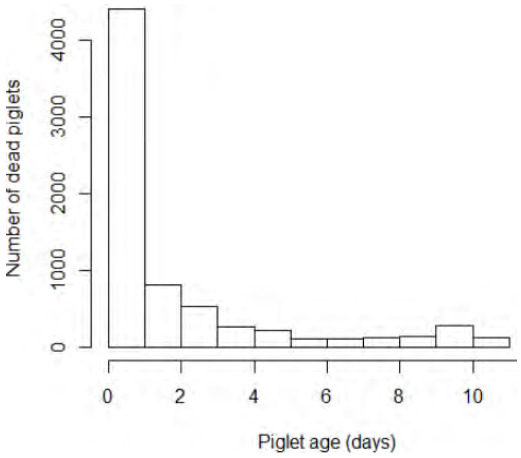


Figure 1. Histogram of piglet mortality from 0 to 11 days of piglet age.

Table 1. Genetic parameters for TNB, Mort11 and LS11: heritability (diagonal), genetic correlations (below diagonal) and phenotypic correlation (above diagonal).[1]

Trait	TNB	Mort11	LS11	Phenotypic variance
TNB	0.09	0.45	0.80	13.45
Mort11	0.66	0.06	-0.18	4.99
LS11	0.87	0.20	0.06	11.08

[1] Standard errors of heritability estimates were ~0.02.

By altering the breeding goal and focusing on Mort11 only, piglet mortality could decrease (Table 2). However, TNB and as well as LS11 decreased, which might not be acceptable from an economic perspective. Therefore, we used a desired gain breeding goal to achieve zero genetic change in Mort11, which might be more acceptable to society and organic consumers (Font-i-Furnols *et al.*, 2019). With the desired gain indexes, more economic weight was put on Mort11. LS11 increased, although the increase in response to selection was lower compared to the scenario using equal economic weights for Mort11 and TNB.

A breeding goal that only consisted of LS11 led to lower response to selection compared to the scenario that included TNB and Mort11 with equal economic weight in the breeding goal. The two breeding goals were equivalent in economic weights for component traits. However, the higher response to selection of the breeding goal using TNB and Mort11 was due to the possibility of having different selection indexes (b_{TNB} and b_{Mort11}) for the two traits. In other words, differences in heritability of the two traits were taken into account. In contrast, the use of LS11 only would assume equal index for TNB and Mort11 (=b_{LS11}).

Conclusions

Selection based on breeding goal of LS11 only, or using equal economic index weights for TNB and Mort11 would increase number of piglets alive at 11 days after farrowing for sows, but at the costs of increasing the number of dead piglets slightly.

By defining number of piglets alive for sows as two traits of TNB and Mort11, selection could reduce piglet mortality through the flexibility of altering economic index weights for the traits in breeding goals, and through utilizing differences in heritability of TNB and Mort11.

By increasing the economic index weight for Mort11, breeding program for organic pig production could achieve zero genetic change in piglet mortality until day 11 while still gaining more TNB.

Table 2. Response to selection expressed in LS11 based on different breeding goals.

Breeding goal[1]	v_{TNB}	v_{Mort11}	v_{LS11}	S_{gTNB}	$S_{gMort11}$	LS11
H_1 [a]	1	-1	-	0.572	0.149	0.424
H_1 [b]	1	0	-	0.624	0.234	0.389
H_1 [b]	0	-1	-	-0.507	-0.288	-0.218
H_1 [c]	1	-1.76	-	0.363	0.000	0.363
H_2 [d]	-	-	1	0.450	0.053	0.397

[1] [a] equal economic indexes for TNB and Mort11; [b] alternative economic weights; [c] desired gain indexes; [d] optimum index using derived parameters from bivariate model of TNB and Mort11.

References

Brascamp, E. W. (1984) Anim. Breed. Abstr. 52:645–654.
Chu T.T., Zaalberg R.M., Villumsen T.M., Jensen J. (2021) 72nd EEAP, Davos, Switzerland.
Font-i-Furnols M., Skrlep M., Aluwe M. (2019) IOP Conference Series: Earth and Environmental Science.
Gibson J., Dekkers J. (2003) Design and economics of animal breeding strategies.
Hansen C. (2020) Seges Svineproduktion NOTAT NR. 2014.
Madsen P., Jensen J. (2013) DMU: A Package for Analysing Multivariate Mixed Models.
Nielsen B., Su G., Lund M.S., Madsen P. (2013) J. of Anim. Sci. 91, 2575-2582.
Rangstrup-Christensen L., Krogh M.A., Pedersen L.J., Sørensen J.T. (2018). animal 12, 810-818.

86. Heritability of daily activity over time in broilers

E.D. Ellen[1*], M. van der Sluis[1], B. de Klerk[2], J. Henshall[2], Y. de Haas[1], T.B. Rodenburg[1,3] and P. Bijma[1]

[1]Wageningen University & Research, Droevendaalsesteeg 1, 6708 PB Wageningen, the Netherlands; [2]Cobb Europe, Koorstraat 2, 5831 GH Boxmeer, the Netherlands; [3]Utrecht University, Yalelaan 2, 3584 CM Utrecht, the Netherlands; esther.ellen@wur.nl

Abstract

Individual activity is related to the health, welfare and performance of broilers. In previous work using a radio frequency identification system, we were able to collect individual activity data over time, and activity was found to be heritable. The aim of the current study was to estimate genetic parameters for activity while taking into account the effect of age. Therefore, two repeated measurement models were fitted. The model with an intercept and slope for genetic and permanent environmental effect, and a heterogeneous residual variance, yielded the best goodness-of-fit. First results showed that daily heritability varied between 7 and 30%. The heritability was lowest at the start and at the end of the production life. This study shows potential for selection on activity in broilers. Furthermore, this study shows that activity patterns (in time) can be changed, as both the intercept and slope of activity are heritable.

Introduction

Activity levels of broilers are related to their health, welfare and performance (Bokkers *et al.*, 2007; Van Hertem *et al.*, 2018; van der Sluis *et al.*, 2019, 2021). Therefore, individual activity levels appear to have potential to serve as a proxy for other traits in broiler breeding programmes. However, before implementation of activity as a proxy in breeding programmes, it is important to gain insight into the heritability of activity. Several studies have examined this topic. For example, it has been reported that the heritability of activity (recorded as the number of observations during which an animal was moving) at 16 days old, in a cross between two broiler lines that were divergently selected on digestive efficiency, was low (0.09±0.07) and not significantly different from zero (Mignon-Grasteau *et al.*, 2017). In another study, the average activity from hatching until approximately 35 days old was found to be heritable (0.31±0.11; Ellen *et al.*, submitted).

Studies on the heritability of activity have, however, been limited due to the labour required for personal observations, complicating for example the study of changes in heritability with age. To genetically improve activity in broilers, it is, however, important to take into account changing activity levels and variation due to age. In addition, previous research showed that activity decreases with age, suggesting that heritability might change over time. Therefore, research is needed to investigate the heritability pattern of activity. In this study, we examine the heritability of activity in broilers based on individual records of daily activity levels, recorded using a radio frequency identification (RFID) tracking system, over time. We compare two repeated measurements models that take into account heterogeneous variance and an effect of time to estimate genetic parameters for activity. The results of this study contribute to an improved understanding of the genetic variation of daily activity in broilers.

Materials & methods

Animals, housing and pedigree. Data on a total of 387 purebred male broilers were used for the analysis, collected across five consecutive production batches. The broilers were housed on a farm in the Netherlands, in a pen with a size of 4.7 m^2. Food and water were provided *ad libitum*, and wood shavings were provided as bedding.

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_86, © E.D. Ellen *et al.* 2022

Nearly the same sires (31 in total) and dams (96 in total) were used in each batch, but with different numbers of offspring. Each sire was mated with approximately 3 dams, and each dam contributed on average 3.7 male offspring. Approximately six generations of pedigree were included in the calculation of the relationship matrix.

Activity recording. All broilers were tracked continuously with an RFID tracking system, using leg tags on the broilers and an RFID antenna grid underneath the pen floor (see van der Sluis *et al.* (2020) for details on the activity tracking). Approximate distances moved during the main light period from 07.00-23.00 were calculated for each day and bird, using the recorded tag positions during this period. The recorded distance was then divided by the recording duration between 07.00 and 23.00 for that specific day, to obtain an average distance moved per hour for each day. This allowed for comparisons between days and batches, even when data were missing for part of a day due to for example weighing of the birds.

Data analysis. For each day and bird, activity was recorded, resulting in 12,232 activity records. R software version 4.1.0 (R Core Team, 2021) was used for an initial analysis and to decide which fixed effects to include in the full model to estimate genetic parameters. The initial model included a fixed effect for batch, day, and the interaction between batch and day. The ASReml software package (Gilmour *et al.*, 2015) was used to estimate genetic parameters for daily activity. Two models were compared using the Likelihood-ratio test (Wald, 1943):

$$y = Xb + Za + Ze_p + e \quad (1)$$

$$y = Xb + Za_0 + Za1.xt + Ze_{p0} + Zep1.xt + e \quad (2)$$

where y is a vector of daily activities, b is a vector of fixed effects, with incidence matrix X linking observations to the fixed effects, a is a vector of breeding values, with incidence matrix Z linking records to individuals, e_p is a vector of permanent environmental effects of individuals to account for repeated measurements, and e is a vector of random residuals, allowing for a different variance for each day. For Model 2, both a genetic and a permanent environmental effect are fitted as an intercept, and as a function of time, where $x_t = 36–Day$ (Day ranging from 1 to 35). We used $x_t = 36–Day$, rather than $x_t = Day$, because an initial analysis showed that the variance decreased over the growing period.

The genetic variance as a function of time is, $\sigma^2_{A_t} = \sigma^2_{A_0} + x_t^2\,\sigma^2_{A_1}$, where t is Day (1-35), $\sigma^2_{A_0}$ is the genetic variance in the intercept, and $\sigma^2_{A_1}$ is the genetic variance in the slope. The variance due to the permanent environmental effect, as a function of time, is $\sigma^2_{P_{e_t}} = \sigma^2_{P_{e_0}} + x_t^2\,\sigma^2_{P_{e_1}}$. We assumed that effects on intercept and slope were independent, to avoid overfitting the data, which contained only 387 individuals. The phenotypic variance as a function of time is $\sigma^2_{P_t} = \sigma^2_{A_t} + \sigma^2_{P_{e_t}} + \sigma^2_{e_t}$, where $\sigma^2_{e_t}$ is the residual variance for a specific day. The heritability for each day is $h^2_t = \sigma^2_{A_t}/\sigma^2_{P_t}$.

Results

Figure 1 shows the average daily activity per batch. As expected, activity decreased with age. Furthermore, differences in activity levels were found between the different batches. Table 1 shows the comparison of the different genetic models. Statistical comparison showed that the model with an intercept and slope for both the genetic and permanent environmental effect yielded the best goodness-of-fit. Figure 2 shows the heritability for each day, which ranged from 0.07 to 0.30.

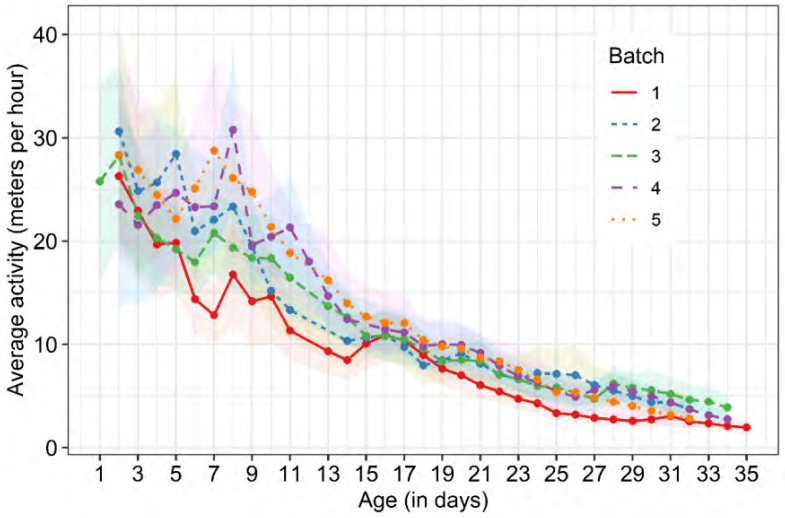


Figure 1. Average daily activity per production batch.

Table 1. Model comparison using the Likelihood ratio-test.

Model	Log Likelihood	LR[1]	P	σ^2_{A0}	σ^2_{A1}
1	-14514.35				
2	-13520.87	1333.34	<0.001	0.25±0.16	0.005±0.002

[1] LR is two times the difference in log likelihood between the complex and less-complex model.

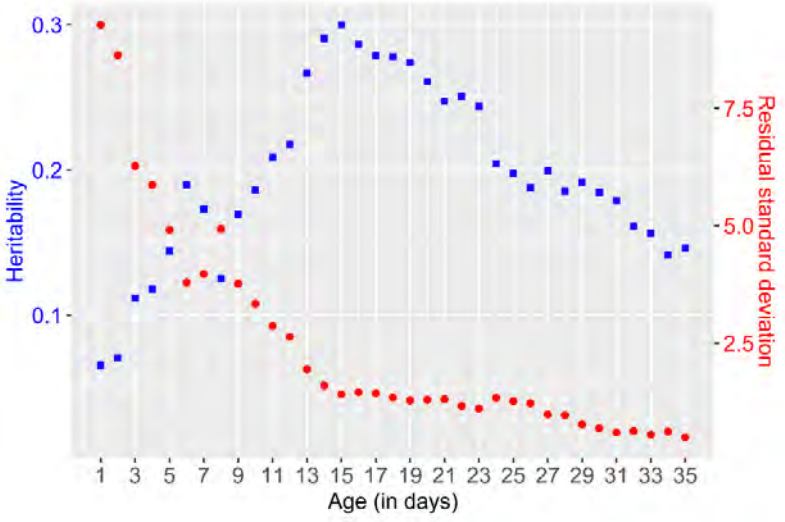


Figure 2. Heritability and residual standard deviation (2nd y-axis) over time.

Discussion

In this study, we found that a repeated measurements model with an intercept and slope for both the genetic and permanent environmental effects, and a heterogeneous residual variance, yielded the best fit. Daily activity is heritable, and heritability changes over time. Despite the limited number of individuals in our data and the presence of up to four random effects in the model, all analyses showed good convergence, and all estimated genetic parameters were significantly different from zero. This is, at least in part, due to the large number of records per individual (35 here), illustrating the power of longitudinal sensor data for genetic analysis.

In our previous study (Ellen *et al.*, *submitted*), we estimated genetic parameters for overall broiler activity and showed that this trait is heritable. However, the activity level changes considerably over time, and so does the variation in activity per day (Figure 1). Furthermore, when repeated observations are available (such as daily activity records), it could be of added value to consider a model that can handle these kind of longitudinal data, to better pick up on changes over time, instead of summarizing across periods and running separate analyses. In this study, a repeated measurements model with a genetic and a permanent environmental effect as a linear function of time was therefore implemented. However, activity is not necessarily fully linear (see Figure 1). In other studies dealing with longitudinal data, random regression models with non-linear terms have also been implemented. Future research could focus on finding the most appropriate model to take into account the effect of time, potentially also as a non-linear function, for daily activity levels in broilers.

In this study, the heritability of daily activity varied between 0.07 and 0.30. Overall, the estimated heritability was lowest at the start and at the end of the production life. At 15 days of age the highest heritability was found. Moreover, differences between animals in their activity patterns over time were observed. This highlights the value of repeated activity recordings, as single-age records might miss important nuances to the (heritability of) activity in broilers. Therefore, future research could focus on investigating the relation between the intercept and slope and the possibilities of changing the activity pattern of a broiler in relation to their health and performance. Furthermore, detailed interpretation of the results and how to implement these results in a broiler breeding program is needed.

Funding

This study was financially supported by the Dutch Ministry of Economic Affairs (TKI Agri and Food project 16022) and the Breed4Food partners Cobb Europe, CRV, Hendrix Genetics and Topigs Norsvin.

References

Bokkers E.A.M., Zimmerman P.H., Rodenburg T.B., and Koene P. (2007) Appl. Anim. Behav. Sci. 108(1-2):129-142. https://doi.org/10.1016/j.applanim.2006.10.011

Gilmour A.R., Gogel B.J., Cullis B.R., Welham S.J. and Thompson R. (2015). ASReml User Guide Release 4.1 Functional Specification, VSN International Ltd, Hemel Hempstead, UK

Mignon-Grasteau S., Chantry-Darmon C., Boscher M.Y., Sellier N., Le Bihan-Duval E. *et al.* (2017) Behav. Genet. 47(1):114-124. https://doi.org/10.1007/s10519-016-9807-1

R Core Team (2021) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.

Van der Sluis M., de Haas Y., de Klerk B., Rodenburg T.B., and Ellen E.D. (2020) Sensors 20(13):3612. https://doi.org/10.3390/s20133612

Van der Sluis M., de Klerk B., Ellen E.D., de Haas Y., Hijink T. *et al.* (2019) Animals 9(8):580. https://doi.org/10.3390/ani9080580

Van der Sluis M., Ellen E.D., de Klerk B., Rodenburg T.B., and de Haas Y. (2021) Poult. Sci. 100(9):101300. https://doi.org/10.1016/j.psj.2021.101300

Van Hertem T., Norton T., Berckmans D., and Vranken E. (2018) Biosyst. Eng. 173:93-102. https://doi.org/10.1016/j.biosystemseng.2018.07.002

Wald, A. (1943) Trans. Am. Math. Soc. 54:426-482. https://doi.org/10.2307/1990256

Part 8

Novel traits: health and well-being

87. Resilience indicators based on individual feed intake from a high-health swine nucleus

C.D. Bierman[1*], R.A. Kemp[1] and J.C.M. Dekkers[2]

[1]Genesus Inc, P.O. Box 278, ROHOYO, Oakville, MB, Canada; [2]Iowa State University, Ames, IA 50011, USA; cbierman@genesus.com

Abstract

Resilience to disease holds economic value in swine production, yet placing selection emphasis directly on selection candidates is difficult due to the requirement of high health nucleus herds. New traits derived from individual feeding behavior have been previously found to correlate favorably with disease resilience, which we parallel with stress resilience here. The foundation data for these novel traits are commonly recorded within nucleus herds' individual feed intake (FI) recording systems. Parameter estimates within a higher stress and lower stress dataset identified differences in heritability between the two scenarios. Resilience FI traits are heritable inside a high health nucleus environment and tend to have higher heritability under more stressful conditions. Genetic correlations between FI resilience and production traits range in magnitude and favorability dependent on stress condition. Optimal selection should emphasize relationships between resilience and production traits within expected stress scenarios for desired trait gains.

Introduction

Improving profitability for swine producers is the main breeding goal focus. Mortality and morbidity negatively impact profitability (Boyd 2012) but are difficult to directly include in breeding programs. Higher resilience to stressors implies reduced mortality and morbidity and resilience traits could contribute to genetic improvement in morbidity and survival, leading to increased profitability.

Feed intake has received significant focus in terminal sire programs. Appetite drives growth rate and is linked to robustness (Ocepek *et al.* 2019). Stressors such as disease negatively impact feed intake and subsequent growth (Nguyen-Ba *et al.* 2020). Feed intake behavior under a disease challenge has been investigated, resulting in novel FI resilience phenotypes that are heritable and genetically correlated with production and mortality outcomes under such stress (Cheng *et al.* 2020, Putz *et al.* 2019), suggesting they are plausible indicator traits of disease resilience.

Genetic selection for disease resilience in nucleus farms poses challenges due to their high health status. However, non-disease stressors such as temperature and humidity, movement, and social interactions can also affect feeding behaviors (Pearce *et al.* 2014). As individual FI is commonly measured, data to derive the FI resilience phenotypes of Putz *et al.* (2019) and Cheng *et al.* (2020) are available. Harlizius *et al.* (2020) reported FI resilience phenotypes from nucleus boars correlated favorably with survival of their progeny in a multi-pathogen disease challenge. Under the premise that any stress elicits variation in FI and feeding behavior, a parallel interpretation to the results for FI resilience traits reported under disease challenge (stress) can be made for non-disease stress-induced scenarios, which would allow selection of nucleus candidates. Therefore, the objectives of this study were to determine if heritable variation in FI resilience traits exist within a high health setting and to identify whether stress characterized by climatic variation impacts genetic parameter estimates.

Materials & methods

Data. Automatic recording of daily FI was acquired on Duroc animals using Insentec IVOG® feeders (Marknesse, Netherlands). Animals were grouped by sex in pens of 24 pigs and 0.80 m^2 per pig. Pigs started

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_87, © C.D. Bierman *et al.* 2022

test at an average of 82.9 kg and concluded at 117.4 kg (sex-averaged weights). Quality control of daily FI standardized the intake timeline to 42 days prior to test conclusion. Missing daily records from the filter process were allowed to remain missing. Traditional phenotypes included average daily feed intake (ADFI) and growth rate adjusted to 120 kg (Age120). Novel FI resilience traits from Putz *et al.* (2019) included variation in daily FI (VAR_{FI}), variation in daily FI duration (VAR_{Dur}), and the proportion of off-feed days (OFF_{FI}) and duration of off-feed (OFF_{Dur}) events. The month the animal concluded test was combined into higher and lower stress categories based on the previous 5-year average monthly temperature for the farm's location in southwest Manitoba, Canada. This partitioning provided a seasonality effect designed to separate phenotypes into stress categories based on climatic conditions. Although barn temperature and ventilation are electronically controlled, humidity and hot/cold microenvironments are difficult to overcome during climatic extremes, resulting in different stress levels.

Analysis. After quality control, resilience FI phenotypes were computed from 3,580,517 daily FI records from 7,498 animals tested from 2016 to 2021. Variance components were estimated using the Echidna software (Gilmour, 2021) based on the following animal model:

$$y_{ijklm} = age_{ijklm} + Sex_i + cg_j + animal_{ijklm} + litter_{ijkl} + e_{ijklm} \qquad (1)$$

where y_{ijklm} is the trait phenotype; age_{ijklm} is a covariate of age at test conclusion; Sex_i and CG_j are fixed effects of sex and contemporary group, respectively; $litter_{ijkl}$ is the random litter effect ~ $N(0, I\sigma^2_l)$, where σ^2_l is the litter variance; $animal_{ijklm}$ is the random animal genetic effect ~ $N(0,A\sigma^2_k)$, where A is the numerator relationship matrix, and σ^2_k is the additive genetic (co)-variance; e_{ijklm} is the residual effect ~ $N(0, I\sigma^2_e)$, where σ^2_e is the residual variance. Mean heritability estimates (minimum of 6 per trait) were obtained from eight multivariate models. Bivariate models including FI resilience traits with either Age120 or ADFI were used to estimate genetic correlations.

Results

Definition of the stress categories resulted in 7 months of higher stress and 5 months of lower stress (Table 1). The range in average high and low temps for the higher stress category was 25 to 26 and -22 to -9 °C, respectively. The range in average high and low temperatures for the lower stress category was 9 to 24 and -4 to 12 °C, respectively. Assessment of mean FI resilience traits between the two stress categories suggests that the data were successfully partitioned in terms of level of stress (Table 1). Higher means for the FI resilience traits for both sexes for the higher stress category suggest less resilience during the stressful timeframe.

Table 1. Trait means (standard deviations) by gender under different stress categories.

	High stress		Low stress	
	Female	Male	Female	Male
n	1559	2858	1149	1932
ADFI (kg)	3.11 (0.60)	2.91 (0.69)	3.01 (0.65)	2.97 (0.56)
Age120 (days)	169.4 (11.9)	159.0 (9.9)	168.2 (11.9)	158.1 (9.7)
OFF_{dur}	0.05 (0.07)	0.07 (0.10)	0.04 (0.09)	0.05 (0.08)
OFF_{FI}	0.05 (0.07)	0.06 (0.08)	0.05 (0.09)	0.04 (0.07)
VAR_{FI} (kg)	0.93 (0.27)	0.82 (0.28)	0.83 (0.28)	0.77 (0.25)
VAR_{dur} (min)	16.5 (6.1)	14.2 (5.1)	14.3 (5.7)	13.4 (4.9)

[1] OFF trait unit = proportion.

Genetic variances were higher for all traits under the high stress category (13 to 225%) while changes in residual variances were variable (Table 2). Estimates of heritability for OFF_{dur} and OFF_{FI} were low in the low stress category and moderate in the high stress category whereas VAR_{FI} and VAR_{dur} were moderate and similar in both stress categories.

The estimate of genetic correlation (r_g) between ADFI and Age120 was high and negative as expected and unaffected by stress category (Table 3). In addition, estimates of r_g between OFF traits and ADFI were moderate to high, and stronger under higher stress than under lower stress. Although not as high, estimates of r_g between OFF traits and Age120 were also higher under high stress. Estimates of r_g of the VAR traits with ADFI and Age120 were less consistent when comparing stress categories within traits; with a high SE, $r_g(VAR_{FI} : ADFI)$ was zero under high stress, while $r_g(VAR_{FI} : Age120)$ was twice as large in the low stress than the high stress category. Again with a high SE, estimates of $r_g(VAR_{dur}$ with ADFI and Age120 were stronger under high stress.

Discussion

Individuals with high values for the FI resilience traits are expected to be less resilient, as their daily intakes fluctuate more and they have more off-feed days. These traits have been shown to be genetic indicator traits for resilience through their genetic correlations with mortality, treatment rates, and growth under a severe health challenge (Cheng *et al.*, 2020; Putz *et al.*, 2019). Here we report estimates of h^2 for these same traits in a high-health environment for two stress categories defined by the expected level of climatic stress. Estimates of h^2 for FI resilience traits in the low stress category were higher in this population than the h^2 estimates under disease stress reported by Cheng *et al.* (2020). This is encouraging, as it suggests genetic

Table 2. Estimates of heritability and SE for performance and feed intake resilience traits under different stress categories.[1]

	High stress				Low stress			
	σ^2_g	σ^2_e	Heritability	SE	σ^2_g	σ^2_e	Heritability	SE
ADFI	0.179	0.106	0.46	0.04	0.104	0.156	0.35	0.05
Age120	52.3	50.5	0.42	0.04	46.0	49.8	0.40	0.05
OFF_{dur}	0.002	0.005	0.25	0.04	0.001	0.004	0.18	0.04
OFF_{FI}	0.002	0.002	0.31	0.04	0.001	0.003	0.22	0.04
VAR_{FI}	0.025	0.030	0.33	0.04	0.018	0.027	0.31	0.04
VAR_{dur}	13.0	6.6	0.50	0.04	11.5	7.2	0.50	0.04

[1] σ^2_g = genetic variation; σ^2_e = residual variation.

Table 3. Estimates of genetic correlations (SE) between feed intake resilience traits and performance under different stress categories.

	ADFI		Age120	
	Higher stress	Lower stress	Higher stress	Lower stress
ADFI	-	-	-0.80 (0.05)	-0.82 (0.06)
OFF_{dur}	-0.87 (0.05)	-0.66 (0.09)	0.19 (0.16)	-0.07 (0.18)
OFF_{FI}	-0.95 (0.02)	-0.82 (0.07)	0.70 (0.12)	0.28 (0.18)
VAR_{FI}	0.00 (0.15)	0.49 (0.16)	-0.33 (0.13)	-0.65 (0.14)
VAR_{dur}	-0.22 (0.11)	-0.12 (0.16)	-0.39 (0.12)	-0.36 (0.16)

variation in these traits is expressed even without the severe stress of disease challenges. Furthermore, h^2 estimates were generally higher under the high stress category, with the largest differences for OFF phenotypes, where h^2 increased by 54-64%. The dataset spanned 5 years, limiting potential noise from short term farm management influence outside of any contemporary group effect.

High negative r_g between OFF traits and ADFI corroborate results from Cheng *et al.* (2020). Although r_g(OFF : ADFI) was 16-30% more negative in the high stress category, inclusion of OFF traits collected during low stress could also benefit genetic improvement. Putz *et al.* (2019) reported low correlations between ADFI and the VAR traits, which is also duplicated here under the high stress category. The specific estimate of 0.49 for $r_g(VAR_{FI} : ADFI)$ in the low stress category suggests that when pigs are not undergoing climatic stress, targeting lower ADFI genetically reflects selection for lower variation in daily feed intake, which suggests higher stress resilience. This estimate should be investigated further, as this correlation was zero in the high stress category. Estimates of r_g(OFF : Age120) were generally low and favorable, while r_g(VAR : Age120) were moderately unfavorable. The latter estimates require further study, as they disagree with previous reports (Putz *et al.* 2019) and would reduce selection response for growth rate.

In conclusion, FI resilience traits are heritable in a high health setting during times of both high and low stress. Placing selection emphasis on data from different stress characterizations should be driven by genetic correlations, as resilience may already be emphasized indirectly. Selection on OFF traits may not be as beneficial when collected during the high compared to low stress category. The opposite may be true for VAR traits, for which r_g with ADFI and Age120 were neither high nor consistently favorable. Assuming the r_g with resilience holds true, and that stress levels on pig farms may not be consistent over time, genetic improvement of resilience may benefit from continual inclusion of FI resilience traits in the breeding goal.

References

Boyd D. (2012) Proc. of the Am. Assoc. of Swine Veterinarians, Perry City, IA United States.

Cheng J., Putz A.M., Harding J.C.S., Dyck M.K., Fortin F. *et al.* (2020) J Anim Sci. 98:8:1-14. https://doi.org/10.1093/jas/skaa244

Gilmour, A.R., (2019) Echidna Mixed Model Software www.EchidnaMMS.org

Harlizius B., Mathur P., Knol E.F. (2020) J Anim Sci. 18:98(Suppl 1):S150-S154. https://doi.org/10.1093/jas/skaa141

Nguyen-Ba H., van Milgen J., Taghipoor M. (2020) Animal 14:2:253-260. https://doi.org/10.1017/S1751731119001976

Ocepek M., Andersen-Ranberg I., Edwards S. A., Fredriksen B., Framstad T. *et al.* (2016), J Anim Sci. 94:8:3550–3560. https://doi.org/10.2527/jas.2016-0386

Pearce S.C., Sanz-Fernandez M.V., Hollis J.H., Baumgard L.H., Gabler N.K. (2014) J Anim Sci. 92(12):5444-54. https://doi.org/10.2527/jas.2014-8407

Putz A.M., Harding J.C.S., Dyck M.K., Fortin F., Plastow G.S., Dekkers J.C.M. *et al.* (2019)

Front. Genet. 9:660:1-14. https://doi.org/10.3389/fgene.2018.00660

88. Comparison of milk yield based resilience indicators across dairy cattle breeds

G. Bonekamp[1*], M. Poppe[1], J. ten Napel[1], C. Kamphuis[1], Y. de Haas[1] and I. Adriaens[1,2,3]

[1]Wageningen University & Research, P.O. Box 338, 6700 AH Wageningen, the Netherlands; [2]KU Leuven, Kleinhoefstraat 4, 2440 Geel, Belgium; [3]RAFT Solutions Ltd., Sunley Raynes Farm, Galphay Ln, Ripon HG4 3AJ, United Kingdom; gerbrich.bonekamp@wur.nl

Abstract

Resilience is increasingly recognized as an important trait for dairy cattle to improve their functioning and welfare. The log transformed variance (LnVar) and autocorrelation (r_{auto}) of daily milk yield deviations from an expected lactation curve have been studied as potential resilience indicators, so far only using data of Holstein Friesian cattle. The aim of this research was to compare the resilience indicators between different breeds present at Dutch dairy farms and to estimate the effect of crossbreeding on these indicators. Significant differences in LnVar and r_{auto} were found across twelve breeds, with the breed effects on LnVar different from the effects on r_{auto}. We estimated negligible heterosis effects for r_{auto} and LnVar. This study suggests that different breeds respond differently to environmental disturbances and that different breeds might have different levels or types of resilience.

Introduction

Recent local and global developments of the dairy farming industry, such as up-scaling of farm size, have created a demand for healthy and easy-to-manage cows (Elgersma *et al.*, 2018). At the same time, future aspirations towards more circular and agroecological dairy farming, might lead to more environmental challenges for animals due to variable feed quality or exposure to weather extremes. To improve animal functioning and welfare, a more resilient cow is desired (Berghof *et al.*, 2019). A resilient cow is minimally affected by a disturbance or rapidly returns to its former state upon exposure (Colditz and Hine, 2016; Berghof *et al.*, 2019).

Increasing cow resilience through genetic selection requires large-scale quantification of this trait (Berghof *et al.*, 2019). Scheffer *et al.* (2018) proposed to quantify resilience using automatically recorded high-frequency data of traits that are affected by disturbances. For dairy cows, such trait could be milk yield. Milk production drops when cows perceive a disturbance or when they have a disease (Elgersma *et al.*, 2018). In the past, fluctuations in daily milk yield (DMY) have been proposed to create indicators for resilience of individual cows (Colditz and Hine, 2016; Elgersma *et al.*, 2018; Poppe *et al.*, 2020; Adriaens *et al.*, 2020). Poppe *et al.* (2020) calculated resilience indicators from the deviations of DMY records from an expected lactation curve. In the data set of Poppe *et al.* (2020), consisting of purebred Holstein Friesian (HF) cows only, a low variance and autocorrelation were shown to indicate a high level of resilience.

Health, fertility, longevity and production level differ between breeds. For example, Bieber *et al.* (2019) showed that local breeds from Sweden, Austria, Switzerland and Poland tended to live longer, have a better health and fertility, but produced less than the commercial breeds HF or Brown Swiss. They described the local breeds to be more robust and therefore more suitable for locally adapted organic or agroecological dairy production systems. However, large scale analyses of resilience traits based on DMY of non-HF breeds have not been performed. In this study, we compared the resilience indicators proposed by Poppe *et al.* (2020) across breeds and estimated the effect of crossbreeding on these indicators.

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_88, © G. Bonekamp *et al.* 2022

Materials & methods

The dataset consisted of 2,763,357,043 milk yield records from farms with an automatic or conventional milking system provided by breeding company CRV (Arnhem, the Netherlands). These records included animal ID, milk yield and the date and start and end time of the milking of in total 1,528,030 dairy cows. Additionally, pedigree data of each animal were available. Data editing steps and calculation of resilience indicators were similar to Poppe *et al.* (2020). In the current study, purebred animals (at least 87.5% of one breed) and crossbred animals (50% HF and 50% of another breed) were selected. The resilience indicators were calculated only for first parities of cows that were milked by an automatic milking system and whose age at first calving was more than 640 days. For these animals, DMY was calculated. All DMY records made after 350 days in milk were removed from the dataset. Animals that had their first DMY record later than 14 days after calving, that had less than 50 DMY records or that had more than 5% missing records per lactation were removed from the dataset.

Calculation of resilience indicators. The resilience indicators used in this study were based on deviations of DMY records from an expected lactation curve that corrects for the lactation shape in absence of disturbances. For the calculation of this expected lactation curve, a polynomial quantile regression (Equation 1) was fitted on the DMY records of each cow:

$$DMY_t = \beta_0 + \beta_1 t + \beta_2 t^2 + \beta_3 t^3 + \beta_4 t^4 + \varepsilon \quad (1)$$

where DMY_T was the observed DMY on lactation day t and ε was a random error term.

Next, variance and autocorrelation were calculated from the daily deviations of DMY from the expected lactation curve. The variance of those deviations had a right-skewed distribution. To create a normally distributed variable a log-transformation was applied to the variance trait (LnVar). The autocorrelation was calculated with a time lag of 1 day (r_{auto}).

Statistical analysis. Breed and heterosis effects on average daily milk yield (ADMY) and on the resilience indicators LnVar and r_{auto} were estimated using linear regression models. For ADMY and autocorrelation the following model was used:

$$y = \mu + FLV + MON + BS + ZRB + NRB + MRY + BBL + JER + G + FH + AYS + cross + \varepsilon \quad (2)$$

Where y was the observed ADMY or r_{auto}, the HF breed was taken as intercept, the other fixed effects represent the individual breed proportions, with respectively FLV as Fleckvieh, *MON* as Montbéliarde, *BS* as Brown Swiss, *ZRB* as Swedish Red, *NRB* as Norwegian Red, *MRY* as Meuse-Rhine-Yssel, *BBL* as Belgium Blue, *JER* as Jersey, *G* as Groninger Whiteheaded, *FH* as Dutch Friesian and *AYS* the proportion of Ayrshire. *cross* was the fixed effect of an animal being a cross (1) or purebred (0) cow to estimate the heterosis effect, and ε was a random error term. Additional to the fixed effects of Equation 2, the model for LnVar included the ADMY as a covariate to correct for the scale effect of milk yield level across breeds. This was done based on the knowledge that an increase in the mean of a trait typically leads to a (proportional) increase of its variance (Falconer and Mackey, 1996).

Results

The final dataset included 450,642 pure and crossbred animals of twelve different breeds (Table 1). The breeds Meuse-Rhine-Yssel, Jersey and Dutch Friesian were mostly present as purebreds, while other breeds were more frequently used for crossbreeding. The individual fixed breed effects showed that ADMY of

all breeds was significantly lower than that of the Holstein Friesian. The heterosis effect on ADMY was significant and was estimated at +0.9 kg.

The r_{auto} differed slightly between breeds (Figure 1). The r_{auto} of Fleckvieh, Montbéliarde, Brown Swiss, Belgium Blue and Jersey was significantly lower than HF, suggesting that their recovery rates are higher. In contrast, Swedish and Norwegian Red and Dutch Friesian had a significantly higher r_{auto} than HF, suggesting that their recovery rate was lower. The heterosis effect on r_{auto} was estimated at -0.007, which is negligible compared to the breed effects.

The LnVar increased with 0.021 per kg ADMY. Breed effects of Montbéliarde and Swedish Red showed a significantly higher LnVar than HF (Figure 2), which is associated with a more variable milk yield. Fleckvieh, Norwegian Red, Jersey, Dutch Friesian and especially the Groninger Whiteheaded and Meuse-Rhine-Yssel breeds showed a significant lower LnVar than HF, suggesting that those breeds have a more stable milk yield in presence of common disturbances. The heterosis effect was estimated at +0.028, which is negligible.

Discussion

The aim of this study was to compare breeds present at Dutch dairy farms for the resilience indicators LnVar and r_{auto}. Differences in LnVar and r_{auto} were found across breeds, demonstrating cattle breeds differ in milk yield variability and recovery rate. This suggests that the level and type of resilience varies between breeds. Some breeds score better for LnVar and others score better for r_{auto}, this suggests that different breeds respond differently to environmental disturbances.

Table 1. Number of pure- and crossbreds and the breed effects on the average daily milk yield (ADMY) for the breeds in the final dataset.

Breed[1]	HF	MRY	FLE	JER	MON	FH	BS	G	ZRB	NRB	BB	AYS
Purebreds	426,335	1,818	1,010	757	663	282	137	33				
Crossbreds		879	7,338	411	3,456	62	3,133	149	2,196	1,470	472	41
ADMY (kg)	27.2	20.0	22.1	17.9	22.8	20.9	21.5	13.6	20.0	23.3	20.6	19.4

[1] HF = Holstein Friesian, MRY = Meuse-Rhine-Yssel, FLE = Fleckvieh, JER = Jersey, MON = Montbéliarde, FH = Dutch Friesian, BS = Brown Swiss, G = Groninger Whiteheaded, ZRB = Swedish Red, NRB = Norwegian Red, BB = Belgium Blue, AYS = Ayrshire.

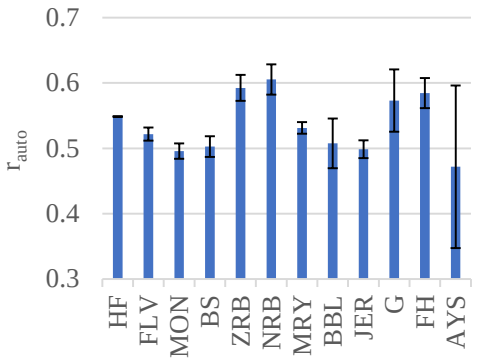


Figure 1. Breed effects on the r_{auto}. Error bars represent the 95% confidence interval.

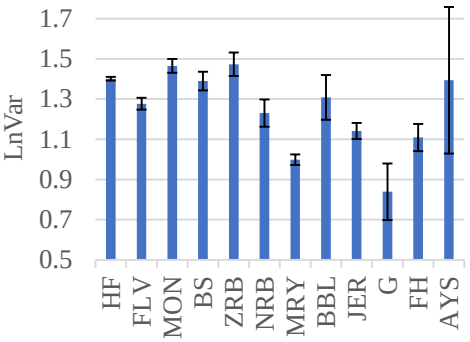


Figure 2. Breed effects on the LnVar, using daily milk yield as covariate. Error bars represent the 95% confidence interval.

In this study, heterosis effects for the resilience indicators were negligible. That was against our expectations because heterosis effects are strongest for health and functional traits (Sørensen *et al.*, 2008). Still, crossbreeding could improve LnVar and r_{auto} of the offspring because some breeds score better than HF. The effect of ADMY as a covariate in the model of LnVar was of considerable strength, taking into account the difference in ADMY between breeds. This is largely explained by a statistical effect, but there might be underlying biological reasons too (Poppe *et al.* 2021). More research is needed to understand that relationship better.

There is no proof in this study that breeds with more favourable LnVar and r_{auto} are indeed more resilient because we have no information on the environmental challenges or on an animal's vulnerabilities that are causing drops in milk yield. The type of production system that the animals are kept in, and thus the type or level of environmental challenges, might be different for the different breeds. In addition, different breeds might be specifically more susceptible to different health issues. So, drops in milk yield might be caused more by mastitis in one breed but more by claw problems in the other. Further research is needed to explore this.

This study suggests that different breeds respond differently to environmental challenges and that different breeds might have different levels or types of resilience (Friggens *et al.*, 2021). This study, however, involved only two resilience indicators, that were both based on milk yield. For a deeper understanding of the resilience of different cattle breeds, more indicators derived from fluctuations in DMY and/or more sources of information, such as animal activity, could be explored (Adriaens *et al.* 2020).

Acknowledgements

This study was financially supported by the Dutch Ministry of Economic Affairs (TKI Agri and Food project 16022) and the Breed4Food partners Cobb Europe, CRV, Hendrix Genetics and Topigs Norsvin, as well as by the European Union's Horizon 2020 research and innovation program (GenTORE) under grant agreement No. 727213.

References

Adriaens, I., Friggens, N.C., Ouweltjes, W., Scott, H., Aernouts, B. *et al.* (2020). J. Dairy Sci. 103:7155-7171. https://doi.org/10.3168/jds.2019-17826

Berghof, T.V.L., Poppe, M. and Mulder, H.A. (2019). Front. Genet. 9:692. https://doi.org/10.3389/fgene.2018.00692

Bieber, A., Wallenbeck, A., Leiber, F., Fuerst-Waltl, B., Winckler, C. *et al.* (2019). J. Dairy Sci. 102(6):5330-5341. https://doi.org/10.3168/jds.2018-16147

Colditz, I.G., and Hine, B.C. (2016). Anim. Prod. Sci. 56(12):1961–1983. https://doi.org/10.1071/AN15297

Elgersma, G.G., de Jong, G., van der Linde, R., and Mulder, H.A. (2018). J. Dairy Sci. 101(2):1240–1250. https://doi.org/10.3168/jds.2017-13270

Falconer, D.S. and Mackay, T.F.C. (1996). Introduction to Quantitative Genetics. 4th ed, Pearson Education Limited, Essex, UK.

Friggens, N.C., Adriaens, I., Boré, R., Cozzi, G., Jurquet, J. *et al.* (2021). https://doi.org/10.5281/zenodo.5473794

Poppe, M., Veerkamp, R.F., van Pelt, M.L., and Mulder, H.A. (2020). J. Dairy Sci. 103(2):1667–1684. https://doi.org/10.3168/jds.2019-17290.

Poppe, M., Mulder, H.A., Kamphuis, C., and Veerkamp, R.F. (2021). J. Dairy Sci. 104:616-627. https://doi.org/10.3168/jds.2020-18525

Scheffer, M., Bolhuis, J.E., Borsboom, D., Buchman, T.G., Gijzel, S.M.W. *et al.*. (2018). Proc. Natl. Acad. Sci. U. S. A. 115(47):11883–11890. https://doi.org/10.1073/pnas.1810630115

Sørensen, M.K., Norberg, E., Pedersen, J., and Christensen, L.G. (2008). J. Dairy Sci. 91(11), 4116-4128. https://doi.org/10.3168/jds.2008-1273

89. Genotype-by-environment interactions for health traits across productivity level or health status of herds in Normande

S. Aguerre[1,2*], A. Vinet[2] and S. Mattalia[1,2]

[1]Institut de l'Elevage, UMT eBIS, Domaine de Vilvert, 78350 Jouy-en-Josas, France; [2]Université Paris Saclay, INRAE, AgroParisTech, GABI, Domaine de Vilvert, 78350 Jouy-en-Josas, France; sophie.aguerre@idele.fr

Abstract

Genotype-by-environment (G×E) interactions for somatic cell score (SCS) and clinical mastitis (CM) were investigated in first lactation Normande cows. Milk and SCS herd-year effects estimated in the French national evaluation were used as environment descriptors of productivity level and health environment of the herds, respectively. Variance and covariance components were estimated using bivariate and reaction norm models. For SCS, heritability estimates ranged from 0.17 to 0.23 according to productivity level. Genetic correlations higher than 0.9 were estimated between groups, showing no G×E interaction for this trait with both models. For CM, the estimates of heritability ranged from to 0.01 to 0.07. Genetic correlations between and 0.3 and 0.99 were estimated across productivity levels and health environments, indicating G×E interactions for this trait.

Introduction

G×E interactions are currently ignored in the national genetic evaluation in dairy cattle in France. But the question arises as to whether these interactions should be taken into account due to the diversity of environments and herd management practices. In a previous study, Huquet *et al.* (2012) investigated G×E interactions for production traits in French dairy cattle. No re-ranking was observed but they found a significant scaling effect for milk yield, with a higher genetic variance in the herds with more intensive management practices.

The aim of this paper was to investigate the magnitude of G×E interactions on SCS and CM in Normande cattle breed. Very diverse definitions of the herd environment have been used in the literature. In this study, herd environment description was based on herd-year effects estimated in the French genetic evaluations of milk production and SCS. The former is assumed to be a descriptor of the productivity level of a given herd, while the latter describes its sanitary status, regardless to its genetic level or the age of the animals.

Materials & methods

Description of the dataset. Data were obtained from the French national database. First parity phenotypes of Normande cows recorded between 2015 and 2020 were included in the analyses. Phenotypes were lactation somatic cell scores (SCS) and clinical mastitis events (CM). Lactation SCS was computed as a weighted average of the test-day SCS ($SCS=\log_2(\text{Somatic Cell Count}/100{,}000)+3$) adjusted for days in milk. CM phenotype was a binary variable set to 1 when at least one clinical mastitis event was recorded during the first 150 days of lactation, and 0 otherwise. Milk and SCS herd-year effects, reflecting productivity level and sanitary status of the herds, respectively, were used as descriptors of the environment. These estimates were obtained as a by-product of the French national genetic evaluation of milk yield and SCS performed in October 2020.

Data analysis. Only cows having completed their entire lactation within the same herd have been included in the analyses to ensure the phenotype was expressed in the right environment. Herds with less than 15 cows, thus with limited accuracy of herd-year effects, were discarded.

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_89, © S. Aguerre *et al.* 2022

The environment description was based on the deciles of the herd-year effect distribution. Two methods were investigated to estimate variance and covariance components for SCS and CM. First, a bivariate analysis was performed, in which the herds with extreme levels of productivity or sanitary status were selected. Then, performances were analysed with a reaction norm model, using the herd-year effects to describe the environment. SCS performances were analysed in relation to the productivity level only. For CM, two separate analyses were performed for the productivity level and the sanitary status, respectively.

For the bivariate analysis, herds randomly selected in the deciles 2-3 and deciles 7-8 defined two groups with contrasted situations. Bivariate models included the fixed effects of herd-year, age at first calving, month of calving, the breeding value as a random effect, and the residual. Performances from 43,111 cows (100,111 animals in the pedigree file) for SCS and 43,182 cows (99,883 animals in the pedigree file) and 48,690 cows (112,810 animals in the pedigree file) for CM with milk and SCS herd-year effect respectively were included.

For reaction norm models, herds were randomly selected in each decile. The average herd-year effects for each decile were used as regression variable. Reaction norm models included Legendre polynomials of order 2 for both traits:

$$y = X\beta + \Sigma_{r=0}^{2} Z_r(t)\, g_{ir} + e \quad (1)$$

Performances from 44,504 cows (85,541 animals in pedigree file) for SCS and 39,161 cows (75,681 animals in pedigree file) and 38,748 cows (74,896 animals in pedigree file) were included for CM with milk and SCS herd-year effect respectively.

For both methods, pedigree information was traced back on three generations. The analyses were performed using Wombat software (Meyer, 2007).

Results

With bivariate models, heritabilities of 0.17 and 0.20 (±0.02) were estimated for SCS in low and high levels of productivity groups respectively (Table 1). No G×E interaction was found for SCS, with a genetic correlation of 0.99 (±0.02) between the two groups. For CM, heritabilities between 0.01 (±0.006) and 0.03 (±0.008) were estimated (Table 1). Genetic correlations between low and high productivity level groups as well as between good and poor health environment groups were 0.7 (±0.2) indicating the existence of G×E interactions.

Table 1. Descriptive statistics, heritability and genetic correlation estimates for SCS and CM in herds with low and high productivity level and in herds with good and poor health environments with bivariate models.

Trait	Herd-Year effect	Group	Mean (std)	Frequency (%)	H^2	r_g
SCS	Milk	Low	3.15 (1.32)		0.17±0.02	0.99±0.02
		High	2.99 (1.33)		0.20±0.02	
CM	Milk	Low		8.94	0.014±0.006	0.70±0.20
		High		10.38	0.025±0.008	
	SCS	Good		10.04	0.020±0.007	0.70±0.20
		Poor		9.46	0.025±0.007	

With reaction norm models, heritabilities of SCS along the milk herd-year effects ranged from 0.18 to 0.23 (Table 2). Genetic correlations between SCS at different milk herd-year effects were all higher than 0.9 (s.e. ranging from 0.003 between close groups to 0.52 between the most distant groups), indicating no G×E interaction for this trait. Heritabilities of CM along milk herd-year effects were low, ranging from 0.02 to 0.06 (Table 2). Similar values were obtained for the heritabilities of CM along SCS herd-year effects (0.01 to 0.07).

Genetic correlations between CM at different milk herd-year effects ranged from 0.44 between groups with the most distant herd-year effects to 0.99 between groups with the closest herd-year effects (Figure 1a). The same trend could be observed for genetic correlations between CM along the SCS herd-year effects with values between 0.3 and 0.99 (Figure 1b). However, the sampling errors were quite high and the estimates were not significantly different from 0 when using SCS herd-year effects as environment descriptor.

Discussion

Our estimates for SCS and CM heritabilities were consistent with the literature (Boichard and Rupp, 1999; Govignon-Gion *et al*., 2016).

Regarding G×E interactions, the analyses with bivariate and reaction norm models gave similar results. Both models supported the existence of G×E interactions for CM, but not for SCS. Our results are in

Table 2. Descriptive statistics and heritability estimates for SCS and CM according to productivity level and health environment with reaction norm model.

Trait	Herd-Year effect		Deciles									
			1	2	3	4	5	6	7	8	9	10
SCS	Milk	Mean	3.42	3.20	3.24	3.16	3.11	3.06	3.04	3.01	3.00	2.95
		std	1.33	1.32	1.34	1.34	1.32	1.36	1.31	1.32	1.32	1.36
		h^2	*0.20*	*0.19*	*0.19*	*0.19*	*0.20*	*0.20*	*0.21*	*0.21*	*0.22*	*0.23*
CM	Milk	Frequency		9.81	8.84	10.80	9.12	8.42	11.38	10.56	9.81	
		h^2		*0.02*	*0.03*	*0.03*	*0.03*	*0.03*	*0.04*	*0.04*	*0.06*	
	SCS	Frequency		9.90	9.87	9.72	8.32	8.02	8.75	10.07	9.13	
		h^2		*0.07*	*0.03*	*0.02*	*0.01*	*0.01*	*0.01*	*0.01*	*0.01*	

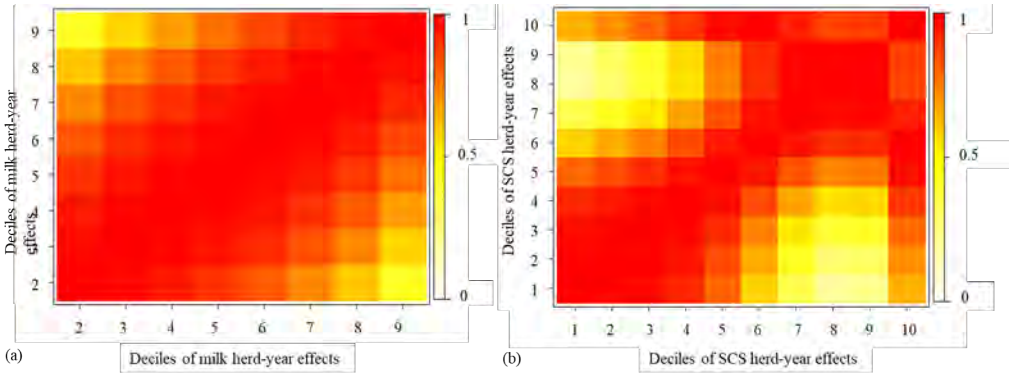


Figure 1. Heatmaps of the genetic correlations between clinical mastitis at different milk herd-year effects (a) and SCS herd-year effects (b).

agreement with previous studies reporting genetic correlations between 0.8 and 1 for SCS at the lactation level across environments (Carlén *et al.*, 2005). However, studies on test-day level are more informative and tend to show the existence of G×E for SCS in early lactation related to a higher incidence of mastitis in early stages of lactation (Calus *et al.*, 2006). Similarly to Carlén *et al.* (2005), our bivariate analysis suggested the existence of G×E interactions for mastitis in different SCS environments. Differences in heritability could not be explained by phenotypic differences between deciles as the average frequencies of CM were similar for all deciles (Table 2).

The use of milk and SCS herd-year effects is an original approach to describe the environment in G×E interactions studies. They represent good estimates of the overall productivity level and health status of the herds and they are easily accessible since they are computed as part of the national genetic evaluations. The use of SCS herd-year effects was particularly interesting since few indicators on the health status of herds are available, these data being difficult to collect routinely. Proxies could be used as shown by Calus *et al.* (2006) who combined bulk milk SCC as an indicator of hygienic conditions to days in milk in their G×E analyses; and Carlén *et al.* (2005) who used herd-year average of somatic cell count and mastitis. Our results indicate that SCS herd-year effects are another useful proxy for health environment.

Different breeding strategies could be considered to take into account G×E interactions in the breeding programs. A guideline was provided by Mulder *et al.* (2006) who showed that it was optimal to have environment-specific breeding programs if the genetic correlation between environments was lower than 0.61.

Conclusions

This study showed that milk and SCS herd-year effects are useful proxies for the productivity level and health environment of the herds. We report the existence of G×E interaction for clinical mastitis in first lactation Normande cows. Taking these interactions into account in breeding programs would allow to adapt breeding strategies to various environments.

Acknowledgements

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 727213 (GENTORE).

References

Calus M.P.L., Janss L.L.G., and Veerkamp R.F. (2006) J. Dairy Sci. 89(12):4846-4857. https://doi.org/10.3168/jds.S0022-0302(06)72533-4

Carlén E., Jansson K., and Strandberg E. (2005) Proc of the 56th Ann. Mtg. EAAP, Uppsala, Sweden

Govignon-Gion A., Dassonneville R., Baloche G. and Ducrocq V. (2016) Animal 10(4):558-565. https://doi.org/10.1017/S1751731115002529

Huquet B., Leclerc H., and Ducrocq V. (2012) Genet. Sel. Evol. 44(35). https://doi.org/10.1186/1297-9686-44-35

Meyer K. (2007). WOMBAT – A tool for mixed model analyses in quantitative genetics by REML, J. Zhejiang Uni. Sci. B 8:815–821. https://doi.org/10.1631/jzus.2007.B0815

Mulder H.A., Veerkamp R.F., Ducro B.J., van Arendonk J.A.M., and Bijma P. (2006) J. Dairy Sci. 89(5):1740-1752. https://doi.org/10.3168/jds.S0022-0302(06)72242-1

Rupp R., Boichard D. (1999). J. Dairy Sci., 82, 2198-2204. https://doi.org/10.3168/jds.S0022-0302(99)75465-2

90. In search of a new tool for phenotyping tick resistance in cattle

O. Matika[1,3*], S. Foster[2], N. Githaka[4], C. Mwendia[5], H. Brown[1], K. Watson[3], A. Djikeng[3] and M. Birkett[2]

[1]The Roslin Institute and R(D)SVS, University of Edinburgh, EH25 9RG, United Kingdom; [2]Biointeractions and Crop Protection Department, Rothamsted Research, Harpenden, Herts, AL5 2JQ, United Kingdom; [3]Centre for Tropical Livestock Genetics and Health, Roslin Institute, University of Edinburgh, EH25 9RG, United Kingdom; [4]Animal and Human Health Programme, [4]International Livestock Research Institute, Nairobi, Kenya; [5]Biochemistry & Molecular Biology, Egerton University, Kenya; oswald.matika@roslin.ed.ac.uk

Abstract

Ticks and tick borne diseases cause significant loss in livestock production with about 80% world's cattle at risk. The cost of chemical control is high and there is an ever increasing tick resistance to chemicals. Genetic selection as alternative long term control strategy is constrained by laborious phenotyping using tick counts or scores. This study explores the use of host volatile semiochemicals that may be attractants or repellents to ticks as a novel phenotype with potential to be used as a proxy in selection programmes. About a 100 young cattle composed of *Bos indicus* and *Bos taurus* were artificially infested with 2,500 *R. decoloratus* larvae with daily female ticks (4.5 mm) counts taken from day 20 post infection. Volatile compounds we sampled before and after tick infestation. We identified three pre-infestation volatile compounds that were associated with tick resistance (P-value<0.01) and one post infestation (P-value<0.05) using 6 day repeated measure analysis. The high correlation coefficients (r=0.66) between repeated records with all volatile compounds support the potential predictive value for volatile compounds in selective breeding programmes for tick resistance in cattle.

Introduction

Ticks and tick borne diseases cause significant loss of production with at risk about 80% of the world's cattle. Economic losses from ticks and tick-borne diseases have been estimated to range from US$20-30 billion annually (Lew-Tabor and Rodriguez Valle, 2016). Host resistance to ticks has been reported as moderately to highly heritable (Burrow and Henshall, 2014) which translates into a long term, low-cost intervention and requires no extra labour or resources (Frisch, 1999). Although host resistance to ticks can be considered as one of the most economical tick control strategies, improvements in host resistance have been largely neglected (Frisch, 1999). The main reason is that the gold standard tick phenotyping methodology involves tick counts which are laborious, time-consuming and require skilled animal technicians as well as expensive infrastructure to constrain animals. A simpler, cost-effective method of identifying individual animal variation in resistance to ticks (phenotyping) under research, commercial and smallholder production systems is urgently needed to enable improvements in host resistance to ticks. In the past, knowledge of different tick pheromones, allomones and kairomones has been used to develop novel tick control products by incorporating tick pheromones and small amounts of pesticide to attract and kill ticks on their hosts or in vegetation (Latha, 2012). Ticks like other haematophagous arthropods use host olfactory and visual cues known as volatile semiochemicals, which generally regarded as attractants, but also can be repellents in some cases (Logan and Birkett, 2007). In order to use the semiochemicals as proxies for tick resistance, we need to quantify within animal variation of different semiochemicals and how they relate to tick loads. In this study, we investigate within animal variation in tick burdens between *Bos taurus* and *Bos indicus* cattle using artificial tick infestation and quantifying differences in volatile semiochemical production. Evidence for a correlated response between volatile semiochemical and tick loads could underpin a development of a new low-cost and rapid tool for phenotyping cattle for resistance to ticks, with potential use for selection and animal breeding in both commercial and smallholder cattle production systems.

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_90, © O. Matika *et al.* 2022

Materials & methods

Animals, ticks and tick infestation. The study animals composed of 24 males and 28 females East African Zebu cattle (*B. indicus*) aged 12 to 18 months purchased from local farmers in Western Kenya were transported to the study site at ILRI, Nairobi, Kenya and male *B. taurus* (45 Holstein-Friesian and 9 Ayrshire) steers from ILRI farm. Cattle were artificial infested with 2,500 Rhipicephalus (Boophilus) Decoloratus (*R. decoloratus*) larvae, with daily counts of female ticks (4.5 mm) conducted over a 6-day period from day 20 post infestation.

Volatile compounds sampling and gas chromatography (GC) analysis. Each animal was held in a crush and volatile organic compounds (VOCs) were collected using air entrainment using the anterior groin of an animal by a hand-held air sampling pump. The funnel was held gently against the skin to ensure a seal around the funnel perimeter. Collections were carried out for 5 minutes. Two samples were collected per animal, i.e. prior to, and 21-23 days after, tick infestation.

VOCs were analysed by high resolution gas chromatography-flame ionization detector (GC-FID) analysis. Nine VOCs (isoamyl acetate, (*E*)-2-heptenal, benzaldehyde, 6-methyl-5-hepten-2-one, hexanal, decane, (*RS*)-limonene, undecane and tridecane) were selected as candidates for investigation based on previous reports as non-host semiochemicals for ticks.

Statistical analysis. Volatile organic compound data were analyzed separately as nine pre- infection values (BI), nine post infection (AI) values and nine differences (diff) between the pre and post infection values resulting in 27 volatile compound traits. The tick counts were treated as six daily counts modelled as repeated measures or each daily count as separate trait and total or average. The distribution of raw tick counts and volatile compounds were mostly skewed; hence they were transformed to approximate normality using log10 (x + 1) with a constant added to allow for the zero counts in our data. Data were subsequently analysed in SAS software (SAS, 2012) for repeated measure analysis modelling the six days tick count effects in the model. The final model fitted for the repeat measure analyses included effects of day, group, body weight and volatile compounds. The fixed effect of 'group' was created by combining the effects of sex (Male or Female) and breed (Zebu or Taurus) to give a group effect for male Zebu, female Zebu and male Taurus animals.

Single trait analysis fitted individual daily tick counts, tick totals or averages analysed using GLM (SAS, 2020) using general linear models but removing the effects of day in the models. The models explored both breed and sex as fixed effects or group, plus the covariates already mentioned above.

Trait ~ breed +sex (or group) + body weight + volatile compounds

Where trait was individual daily log transformed tick counts, log transformed tick totals or log transformed tick averages.

The correlation between transformed tick counts and volatile organic compounds was explored using Pearson correlation coefficients using SAS (2012).

Results

There was wide variation between pre- infection values (BI, CV% ranging from 18.4 to 59.4) and post infection (AI, CV% ranging from 8.2 to 65.2) across sexes and breeds. The CV% was generally higher for BI in female Zebu than Males, except for volatiles BI-(*E*)-2-heptenal/benzaldehyde) and BI-(*E*)-2-heptenal. The CV% of log transformed tick counts ranged from 31.1 to 84.0.

The repeat measure analysis before infestation identified three volatile compounds (BI-(*E*)-2-heptenal, BI-6-methyl-5-hepten-2-one and BI-hexyl acetate) that were significantly (*P*-value<0.05) associated with low tick counts suggesting that they could be used as predictive indicators of tick resistance. All fixed and covariate effects of day, group, body weight and volatile compounds significantly ($P<0.05$) accounted for other variations in the model. Results for after infection tick counts for repeat analysis identified a single volatile compound (AItrans-2-heptenal/benzaldehyde) that was significant ($P<0.05$) on volatile compounds.

There were three volatile compound differences (Diff(AI-BI): (*E*)-2-heptenal/benzaldehyde, 6-methyl-5-hepten-2-one, hexyl acetate)) that were significantly ($P<0.05$) associated with tick counts whilst two of the compounds (6-methyl-5-hepten-2-one and hexyl acetate) were previously identified as significantly associated with tick counts in the BI results.

Five single day tick count and volatile compound combinations were identified in the GLM analysis as significant ($P<0.05$) where isoamyl acetate (TickD25 by AI-isoamyl acetate) and limonene (TickD25 by AI-limonene) had not been previously observed in the repeat measure analysis. Two other volatile compounds ((*E*)-2-heptenal/benzaldehyde and 6-methyl-5-hepten-2-one) were observed in the previous analyses with *P*-values ranging from <0.015 to 0.045. These volatile compounds represent potential proxies for use in predicting tick resistance without counting ticks on the animals. The volatile compound (lnBI6-methyl-5-hepten-2-one) was significantly ($P<0.05$) associated with tick counts during the first 3 days of sampling and also for tick count totals and tick average.

The high correlation coefficients, r=0.66 and r=0.73 respectively for models including repeated records or mean tick counts with all volatile compounds support the potential predictive value for volatile compounds in selective breeding programmes for tick resistance in cattle.

Discussion

The within breed natural differential attractiveness of individual animals to disease-transmitting cattle flies was found to be partly due to differences in volatile semiochemicals emitted by the host (Birkett *et al.*, 2004). Other studies (Jensen *et al.*, 2004) reported a genetic basis to host resistance to cattle flies. High correlation between volatile semiochemical production and tick resistance in cattle could underpin the development of a new low cost tool for phenotyping cattle for selection and breeding for tick resistance in livestock breeding programmes.

The work to develop new anti-tick vaccines are ongoing (Guerrero *et al.*, 2014, Lew-Tabor and Rodriguez Valle, 2016), but new vaccines are highly unlikely to confer total protection against ticks. It has been postulated (Frisch, 1999), that the best method of reducing the impacts of these stressors to improve productivity and animal welfare is to breed cattle that are well-adapted to the different types of stress, and work in concert with other management interventions.

We identified volatile compounds with some naturally found in naïve animals, if this is can be replicated in other studies, it would allow sampling of animals irrespective of the tick infestation status. Results for post infestation compound can be taken to show host responses to an ongoing tick challenge. In sub-Saharan Africa, ticks and tick borne diseases are a major constraint, example of livestock tick-borne diseases reported in Kenya include theileriosis, babesiosis, anaplasmosis and ehrlichiosis, which reflect the diversity of the transmitting tick vectors (Mwamuye *et al.*, 2017).

This search for potential new volatile-based resistance phenotype has the potential to simultaneously improve cattle resistance to both ticks and biting flies. The new phenotype would enhance predictive power across breeds and countries to enhance existing efforts already being undertaken in this regards (Cardoso *et al.* 2021). Our current method will need further refinement to satisfy the requirements of a simple, cost-effective phenotype available for use in large cattle herds as a replacement for the laborious tick counts.

References

Birkett, M.A., Agelopoulos, N., Jensen, K-M.V., Jespersen, J.B., Pickett, J.A., *et al.* (2004). Medical and Veterinary Entomology, 18, 313-322.

Burrow, HM and Henshall, J.M. (2014). Proc 10th World Congress Genetics Applied to Livestock Production, 17–22 August 2014, Vancouver, Canada.

Cardoso, F.F, Matika, O., Djikeng, A., Mapholi, N., Burrow, H.M., *et al.* (2021). Frontiers in Immunology. https://www.frontiersin.org/articles/10.3389/fimmu.2021.620847/full

Jensen, K-M.V., Jespersen, J.B., Birkett, M.A., Pickett, J.A., Thomas, G., *et al.* (2004). Medical and Veterinary Entomology, 18, 275-280.

Frisch, J.E. (1999). Int. J. Parasit. 29, 57-71.

Guerrero, F.E., Andreotti, R., Bendele, K.G., Cunha, R.G., Miller, R.J. *et al.*,(2014). Parasit. Vect. (2014), 7, 475. https://doi.org/10.1186/s13071-014-0475-9

Latha, B.R. (2012) J. Vet. & Anim. Sci. (Lahore) 43, 1–10.

Lew-Tabor & Rodriguez Valle (2016). Ticks Tick Borne Dis 7, 573

Mwamuye, M.M., Kariuki, E., Omondi, D., Kabii, J., Odongo, D., *et al.* (2017). Ticks Tick Borne Dis. 2, 208-218. https://doi.org/10.1016/j.ttbdis.2016.09.002. Epub 2016 Sep 5.

SAS, Base SAS Procedure's Guide: Statistical Procedures, 2002-2012, SAS® 9.4, SAS Institute Inc., Cary, NC, USA.

91. Lameness in genomic evaluation for foot and claw disorders in Czech Holstein cows

L. Zavadilová*, E. Kašná and Z. Krupová

Institute of Animal Science, Přátelství 815, 104 00 Praha-Uhříněves, Czech Republic;
zavadilova.ludmila@vuzv.cz

Abstract

Genetic parameters and genomic breeding values for three groups of foot and claw disorders (CD): infectious diseases (ID), claw horn lesions (CHL) and overall claw disorder (OCD) were estimated. Lameness (L) represented the indicator trait for CD. L figured in the two-trait or multi-trait model with ID, CHL or OCD to increase the accuracy of genomic breeding values. Data from 189,893 lactations of 57,711 Holstein cows were recorded on 74 farms in the Czech Republic from 2017 to 2021. The linear animal model and single-step genomic prediction were employed. Genetic correlations between L and ID were 0.27; CHL 0.60; OCD 0.67. An increase in the accuracy of the breeding value appeared in the two-trait model compared to the single trait model. Due to the multi-trait model, the increase in accuracy was no longer as significant as the two-trait model. Paper was supported by the projects MZERO0718 and QK1910320.

Introduction

Foot and claw diseases/disorders are among the main reasons for the deterioration of dairy cows' welfare, production, and reproductive system (Heringstad *et al.* 2018). Welfare problems arise because of claw lesions that are usually long-term and painful (Van der Waaij *et al.* 2005). Other fundamental consequences manifest in the production and reproductive system of the cow (Charfeddine and Pérez-Cabal 2017). The heritability of foot and claw disorders has been verified in several dairy cattle breeds (Chapinal *et al.* 2013). The painful disorder often manifests as lameness (L), e.g. impaired locomotion of variable severity and deviation from normal gait or posture (Greenough 2007). Linear mixed models are often employed to estimate genetic parameters for claw disorders, e.g. especially when the genetic correlation with other traits is estimated (Chapinal *et al.* 2013). The objectives of this study were to use lameness as the second trait in the genomic evaluation of claw health traits in Czech Holstein cattle by a genomic two-trait model for increasing the accuracy of the genomic prediction.

Materials & methods

Cases of foot and claw diseases/disorders (CD) from 189,893 lactations of 57,711 Holstein cows were recorded on 74 farms in the Czech Republic from 2017 to 2021. Records of CD were gathered by farmers and registered voluntarily in the national cattle health monitoring system 'The Diary of Diseases and Medication' (Kasna *et al.* 2017). This recording system was implemented in August 2018 after a one-year trial period in the Czech Republic. It consists of farmers' online health recording form and a simplified key of diagnoses based on ICAR recommendations. Holstein Cattle Breeders Association of the Czech Republic and the Czech and Moravian Breeding Corporation provided health and genomic data. Three groups of CD were defined: infectious diseases (ID), including digital and interdigital dermatitis; interdigital phlegmon and heel horn erosion; claw horn lesions (CHL), including ulcers, white line disease, horn fissures, and double sole; and overall claw diseases (OCD) comprising all the recorded diseases/disorders. Monitoring also included lameness (L). For analyses, foot and claw diseases/disorders and lameness were defined as 0/1 occurrence per lactation.

Description of model equation. The following linear animal model was used to estimate genetic parameters and genomic breeding values for foot and claw disease/disorder traits and L traits:

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_91, © L. Zavadilová *et al.* 2022

$$y_{ijklm} = parity_agegroup_i + herd_j + year_season_k + PE_l + A_m + e_{ijklm}, \quad (1)$$

where y_{ijklm} is the analysed trait: ID, CHL, OCD, L 0/1 occurrence per lactation; $parity_agegroup_i$ is the effect of parity combined with age at calving class (15 levels: first, second, third, fourth, and five and higher parity; 3 classes of age at calving per parity); $herd_j$ is the effect of herd j (up to 74 levels); $year_season_k$ is the combined effect of calving year k (4 levels) and $season_l$ is the effect of calving season (4 levels: January–March, April–June, July–September, and October–December); PE_l is the random permanent environmental effect on cow traits across parity; A_m is the random additive genetic effect of $animal_m$, and e_{ijklm} is the random residual effect. For the prediction of genomic breeding values, a single-step procedure was applied (Aguilar *et al.* 2010, Christensen and Lund, 2010) with 7,337 genomic animals: 5,215 bulls and 2,122 cows with CD phenotype; a total number of effective SNPs used in the calculation of G matrix was 38,084; effective animals 7,105. The relative breeding values in % (RBV) were calculated using a reference level derived from sires born in 2010 (mean=100%) and SD=12. Higher relative breeding values mean a favourable value for foot and claw disorders, i.e. higher resistance to them. The predictions of breeding values by single-trait model for ID, CHL and OCD; by two-trait model for ID and L, CHL and L, OCD and L, and by multi-trait model for ID, CHL, OCD and L. Lameness was considered as an indicator trait included in the two-trait or multi-trait model prediction for an increase of the accuracy. The basic editing and preparation of datasets, processing of results and basic statistical evaluation were carried out by the SAS 9.4 programme (SAS. 2016). Program package BLUPF90 (Misztal *et al.* 2018) was used to estimate genetic parameters and genomic breeding values. The number of animals in particular models and lactation incidence rate (LIR) in % are presented in Table 1. The datasets for two-trait and multi-trait models were combined according to cows and parity from datasets used for single-trait analyses.

Results

As shown in Table 2, there was a sound increase in the accuracy of the breeding values for all the listed animal categories for the two-trait model compared to the single trait model, which included the particular CD trait and L. When the multi-trait model was employed, the increase in accuracy was no longer as high as for the two-trait model. Notably, for CHL, the accuracy of the breeding values decreased. According to particular traits, the genetic trends for bulls and cows were very similar when compared between single-trait and two-trait models. However, genetic trends based on the multi-trait model differ considerably from previous models, especially for CHL, as is shown in Figure 1. The differences between genetic trends by different models for other traits were less noticeable for ID and OCD than for CHL. The proportion of individual diseases to ID (in lactations) was 52% dermatitis digitalis, 8% dermatitis interdigitalis, 35%

Table 1. Number of animals for different traits models.

Trait[1]	Model	LIR in %	No. of farms	No. of animals in the pedigree
ID	Single	15.12	39	84,644
CHL	Single	15.10	25	56,510
OCD	Single	19.42	51	102,874
L	Single	8.38	43	84,597
ID, L	Two	See single	68	126,176
CHL, L	Two	See single	60	70,730
CHL, L	Two	See single	60	109,277
OCD, L	Two	See single	74	137,202
ID,CHL,OCD,L	Multi	See single	74	137,448

[1] ID = infectious diseases; CHL = claw horn lesions; OCD = overall claw diseases; L = lameness.

interdigital phlegmon and 4% heel horn erosion; around 10% of all ID lactations combined two diseases. CHL was constituted mainly by ulcers (92%). For OCD, the proportion of lactation with both infectious disease and horn lesion was 29%. The heritabilities of traits were ID 0.067; CHL 0.055; OCD 0.056; L 0.026. Based on the two-trait models, the genetic correlation between ID and L was 0.27; CHL and L 0.60 and between OCD and L 0.67.

Discussion

The genetic evaluation is employed for improving foot and claw health of Holstein cows in the Czech Republic using three groups of CD traits: ID, CHL, and OCD. This procedure appears appropriate because foot and claw health traits are usually grouped according to aetiology and pathogenesis (Greenough, 2007) to represent the most frequent disorders and achieve adequate records for genetic evaluation. The other monitored trait is lameness, but it does not enter the genomic evaluation. Because the amount of data for CD data is not optimal, we also try to use L data in the genetic evaluation of CD traits. We proved true what Bicalho *et al.* (2007), and Weber *et al.* (2013) found that L is highly correlated with CD, as we also found especially for CHL and OCD. Heritabilities of all analysed traits were low as usual in the literature (Van

Table 2. Accuracy of genomic breeding values for different models.

Trait[1]		Genomic cows; no. 2,212		Genomic bulls; no. 5,215		Cows with phenotype[1]	
	Model	Mean±SD	Range	Mean±SD	Range	Mean±SD	Range
ID	Single	0.28±0.05	0.09-0.45	0.23±0.10	0.00-0.86	0.14±0.06	0.00-0.45
	Two	0.41±0.05	0.23-0.53	0.34±0.13	0.00-0.93	0.20±0.07	0.01-0.53
	Multi	0.45±0.06	0.21-0.60	0.38±0.15	0.00-0.92	0.26±0.09	0.01-0.58
CHL	Single	0.22±0.05	0.05-0.38	0.18±0.09	0.00-0.84	0.11±0.06	0.00-0.36
	Two	0.41±0.05	0.23-0.53	0.34±0.13	0.00-0.95	0.22±0.08	0.01-0.49
	Multi	0.36±0.06	0.14-0.52	0.30±0.14	0.00-0.92	0.21±0.08	0.01-0.48
OCD	Single	0.31±0.05	0.13-0.49	0.27±0.11	0.00-0.88	0.14±0.06	0.00-0.45
	Two	0.43±0.05	0.24-0.56	0.35±0.13	0.00-0.93	0.20±0.07	0.01-0.53
	Multi	0.44±0.06	0.19-0.59	0.36±0.15	0.00-0.91	0.26±0.09	0.01-0.58

[1] no. of Cows with phenotype: ID infectious diseases – 30,159; CHL claw horn lesions – 17,418; OCD overall claw diseases – 39,592.

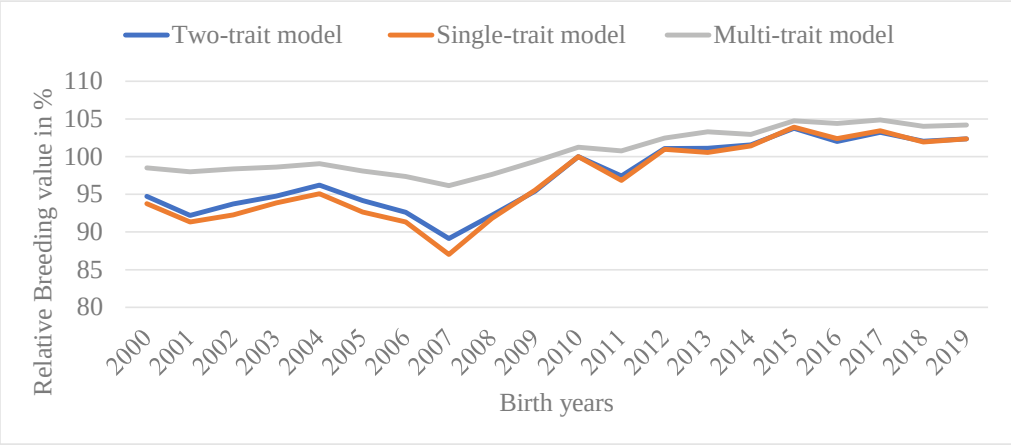


Figure 1. Genetic trends in claw horn lesions for bulls.

der Waaij *et al.* 2005; Weber *et al.* 2013; Heringstadt *et al.* 2018). The presented analysis displayed the very positive impact of the two or multi-trait model on GEBV accuracy. Increasing the amount of information due to the employment of lameness in two-trait model analysis led to an increase in accuracy compared to single-trait genomic prediction. However, we prefer using a two-trait model over a multi-trait model because it has led to a higher increase in accuracy while not expressively altering genetic trends for the particular CD trait.

References

Aguilar I., Misztal I., Johnson D.L., Legarra A., Tsuruta S. *et al.* (2010) J Dairy Sci 93(2): 743–752. https://doi.org/10.3168/jds.2009-2730

Bicalho R.C., Vokey F., Erb H.N. and Guard C. L. (2007) J Dairy Sci 90(10): 4586–4591. https://doi.org/10.3168/jds.2007-0297007-0297

Chapinal N., Koeck A., Sewalem A., Kelton D. F., Mason, S. *et al.* (2013) J Dairy Sci 96(4): 2596–2604. https://doi.org/10.3168/jds.2012-6071

Charfeddine N. and Pérez-Cabal M.A. (2017) J Dairy Sci 100(1): 1–13. https://doi.org/10.3168/jds.2016-11434.

Christensen O.F. and Lund M.S. (2010) Genet Sel Evol. 42: 2. https://doi.org/10.1186/1297-9686-42-2

Heringstad B., Egger-Danner C., Charfeddine N., Pryce J. E., Stock, K. F. *et al.* (2018) J Dairy Sci 101(6): 4801–4821. https://doi.org/10.3168/jds.2017-13531.

Greenough P.R. (2007) Bovine laminitis and lameness. Elsevier Limited Philadelphia PA USA

Kasna E., Zavadilova L., Krupova Z., Slosarkova S., Fleischer P. *et al.* (2017) Proc. of EAAP – 68th Annual Meeting, Tallinn.

Misztal I., Tsuruta S., Lourenco D.A.L., Masuda Y., Aguilar I., *et al.* (2018) Manual for BLUPF90 family programs. University of Georgia. Available at: http://nce.ads.uga.edu/wiki/doku.php?id=documentation

SAS (2016): Release 9.4 (TS1M1) of the SAS® System for Microsoft® Windows®. SAS Institute, Inc., Cary, USA.

Van der Waaij E.H., Holzhauer M., Ellen E., Kamphuis C. and De Jong G. (2005) J Dairy Sci 88(10): 3672–3678. https://doi.org/10.3168/jds.S0022-0302(05)73053-8

Weber A., Stamer E., Junge W. and Thaller G. (2013) J Dairy Sci 96(5): 3310–3318. https://doi.org/10.3168/jds.2012-6261

92. The lasting effect of udder inflammation on milk lactose content in dairy cows

A. Costa[1*], H. Bovenhuis[2], C. Egger-Danner[3], B. Fuerst-Waltl[4] and M. Penasa[1]

[1]Dept. Agronomy, Food, Natural resources, Animals and Environment, University of Padova, Viale dell'Università 16, 35020 Legnaro (PD), Italy; [2]Animal Breeding and Genomics Group, Wageningen University, P.O. Box 338, 6700 AH Wageningen, the Netherlands; [3]ZuchtData EDV-Dienstleistungen GmbH, Dresdner Str. 89/19, A-1200 Vienna, Austria; [4]Dept. Sustainable Agricultural Systems, University of Natural Resources and Life Sciences, Vienna, Gregor Mendel-Str. 33, 1180 Vienna, Austria; angela.costa@unipd.it

Abstract

Milk lactose content (LC) reduces with parity in dairy cattle and shows an inverse relation with somatic cell count. The aim of this study was to investigate if the lifetime decrease in LC is due to cumulative number of experienced udder inflammations. For this purpose, milk test-day data of Austrian Fleckvieh cows were available together with records of acute (AM) and chronic mastitis (CM) cases. In addition, inflammations were also defined by using high test-day somatic cell count: ≥200,000 (SCC_{200}) or ≥400,000 cells/ml (SCC_{400}). Inflammation events were calculated for each test-day using data collected across the cows' lifespan, i.e. across all available lactations. The effect of cumulative inflammations on LC was highly significant, with LC being the highest when 0 mastitis events were experienced and significantly different from LC estimated for test-day records linked to ≥1 event. Findings demonstrate that the lifetime LC decrease can be seen as a mammary memory indicator and represents the combined result of physiological aging and past inflammations.

Introduction

A negative correlation between milk lactose content (LC) and somatic cell count (SCC) has been observed in several studies, but only in recent years some attempts have been made to explore why and how LC changes when mastitis occurs. Both the genetic and the phenotypic correlations between LC and SCC and between LC and mastitis confirm the opposite trend of these features. A progressive drop in average milk LC is observed during productive life of cows, i.e. across parities. Costa *et al.* (2020) suggested that LC decreases with parity for two main reasons: physiological aging and cumulative effect of inflammation events. In both cases, the alveolar permeability is subjected to damage and impairment of its functionality, which lead to a loss of lactose from the alveolar lumen (milk) to blood and finally to urine. The reduction in alveolar LC causes an alteration of the osmotic equilibrium, which is compensated by sodium and potassium. In this study, we explored the role of cumulative mastitis events on milk LC using cows' test-day (TD) records and validated health data.

Materials & methods

Data. Data used in the present study included milk and health records of cows born between 1997 and 2017 reared in Austrian farms under validated health data recording system (Egger-Danner *et al.* 2012) located in Styria and Lower Austria. Only Fleckvieh cows with information available from parity 1 were retained and the maximum parity was set at 10. In addition, only TD records from 5 to 500 days in milk were considered and ≥3 cows/farm and ≥3 TD/lactation were required. After editing, date of first calving was within a window from 1 January 2011 to 29 February 2020. This guaranteed the presence of reliable health records and allowed the completion of those lactations started in spring 2020. Finally, 432,839 TD from 53,099 lactations and 18,404 cows located on 566 farms were left. Both acute (AM) and chronic mastitis (CM) events were available. The cumulative number of mastitis events were assigned to each TD of a cow starting from the first available TD in parity 1 onwards, i.e. across lifespan, regardless of the distance

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_92, © A. Costa *et al.* 2022

between the diagnosis date and the date of TD. Mastitis events were defined using diagnosed AM, AM+CM or based on high SCC. In particular, TD SCC ≥200,000 (SCC_{200}) or TD SCC ≥400,000 cells/ml (SCC_{400}) were used as threshold. The TD records were classified as reported in Table 1 according to cumulative AM events, i.e. 0, 1, and 2, with the latter including TD records with 2 or more AM. For the sum of cumulative AM+CM, 4 classes were defined, i.e. from 0 (no diagnoses) to 3 (≥3 diagnoses). Furthermore, 5 classes of SCC_{200} and 5 classes of SCC_{400} were created based on the number of TD with 'high' SCC. Specifically, the classes 0, 1, 2, 3, and 4 included 0, 1, 2, 3, and ≥4 'high' TD SCC, respectively. Counting of mastitis events did not stop at each lactation, so that for each TD all events recorded previously in life were considered (Figure 1). Seasons of calving were defined as: Mar-May, Jun-Aug, Sep-Nov, and Dec-Feb.

Effect of mastitis. The following mixed model was used in SAS software v. 9.4 (PROC HPMIXED) to estimate LC with increasing number of experienced inflammations, i.e. including the fixed effect of AM, AM+CM, SCC_{200}, or SCC_{400} separately in four analyses:

$$y_{ilklmn} = u + PO_i + LS_j + ME_k + (PO \times LS)_{ij} + (LS \times ME)_{jk} + YS_l + FC_m + DF_n + e_{ilklmn} \quad (1)$$

where y_{ijklmn} is the milk LC at the given TD, u is the overall intercept of the model, PO_i is the fixed effect of the i^{th} parity (i = 1 to 6, with parities from 6 to 10 grouped together), LS_j is the fixed effect of the j^{th} lactation stage (l = 1 to 7, each class being 50 days wide, except for the last class which included TD from 305 to 500 days in milk), ME_k is the fixed effect of the k^{th} cumulative mastitis event (Table 1), YS_l is the fixed temporal effect of the l^{th} year-season of calving (l = 1 to 43), and FC_m, DF_n, and e_{ilklmn} are the random effects of the m^{th} cow, the n^{th} herd, and the residual, respectively. Random effects were assumed to be independent and normally distributed.

Table 1. Distribution of test-day records across classes of fixed effects.

Fixed effect	Cumulated traits	Frequency (%)				
		0	1	2	3	4
AM	n. past acute mastitis	88.01	8.01	3.98	-	-
AM+CM	n. past acute plus chronic mastitis	84.68	9.41	4.07	1.84	-
SCC_{200}	n. past test-day somatic cell count ≥200,000	51.39	13.74	10.2	5.79	18.88
SCC_{400}	n. past test-day somatic cell count ≥400,000	69.42	12.86	7.51	3.49	6.72

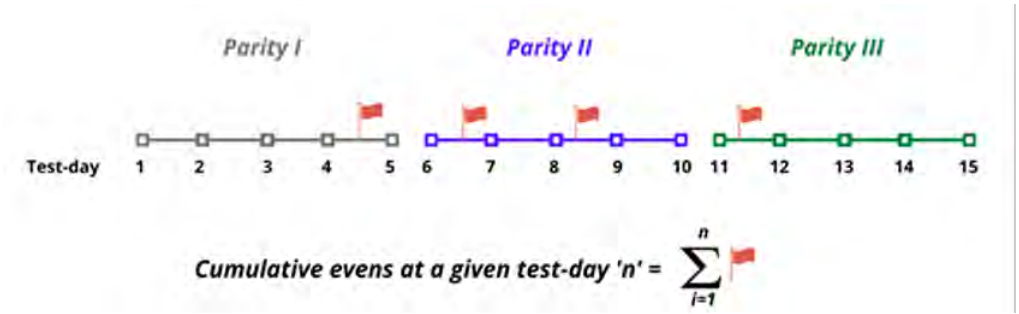


Figure 1. Example of mastitis events calculation for a given test-day record along a cow's lifespan.

Results

Milk LC averaged 4.79% and varied from 3.00 to 5.63%. Each TD was assigned to the cumulative number of mastitis events defined as previously described. The binary events AM and CM were mainly recorded in early and late lactation, respectively, while high SCC values were observed throughout the whole lactation (Figure 2).

All fixed effects in model (1), including the number of mastitis events, were highly significant in all analyses ($P<0.001$). The highest level of LC was observed in class 0 and this estimate was significantly different than those of other classes (Table 2). In addition, the gap between class 0 and class 1 was wider when fixed effect of mastitis was based on SCC_{200} and SCC_{400}.

The lifetime decline in LC is thus due to both parity and past inflammations, and this is particularly evident when looking at the estimates of SCC_{200} or SCC_{400}. In fact, the LC decrease was less noticeable if the effect was either AM or AM+CM. The F-value of fixed effect of parity ranged from 3,104.09 to 5,438.63, while the F-value of mastitis event was 29.87, 18.46, 1,392.03, and 998.99 for AM, AM+CM, SCC_{200} and SCC_{400}, respectively.

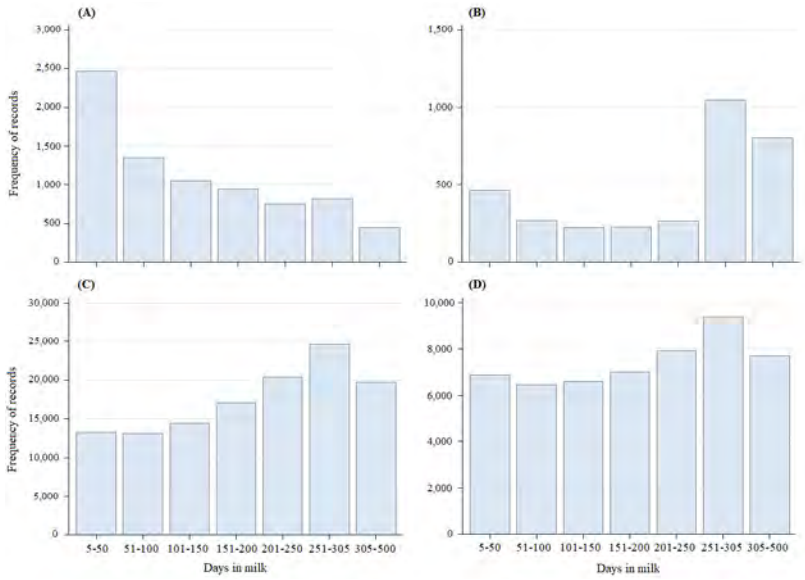


Figure 2. Distribution of mastitis events identified using acute (A) and chronic (B) diagnoses and milk TD SCC ≥200,000 (C) and ≥400,000 cells/ml (D) across days in milk.

Table 2. Least squares means[1] of lactose (%) for cumulative number of mastitis events.

Fixed effect[2]	Class 0	Class 1	Class 2	Class 3	Class 4
AM	4.725 [a]	4.717 [b]	4.714 [b]	-	-
AM+CM	4.724 [a]	4.719 [b]	4.719 [b]	4.712 [c]	-
SCC_{200}	4.763 [a]	4.730 [b]	4.715 [c]	4.705 [d]	4.686 [e]
SCC_{400}	4.746 [a]	4.710 [b]	4.699 [c]	4.687 [d]	4.678 [e]

[1] Superscript letters within row indicate significantly different estimates ($P<0.05$).

[2] Cumulative number of past acute mastitis (AM), acute plus chronic mastitis (AM+CM), and test-day somatic cell count ≥200,000 (SCC_{200}) and ≥400,000 cells/ml (SCC_{400}).

Discussion

Costa *et al.* (2020) opened the debate on the LC reduction across parities in dairy cows and suggested that the drop in LC may be used as a marker of mammary memory. A larger drop in LC within the productive life indicates that the cow faced up more infections, i.e. inflammations. Findings suggest that there is a lasting effect of udder inflammation on the alveolar structure and epithelium, that is detectable through changes in LC. The effect of parity (e.g. tissue aging plus milking mechanical stress) was separated from the effect of cumulative inflammations and, particularly when considering the results of SCC_{200} or SCC_{400}, the hypothesis that the lifetime LC reduction is due to both aging and cumulative udder inflammations was confirmed. Accurate identification of mastitis is challenging, particularly in presence of binary records. On the other hand, identification of mastitis based on milk SCC may led to false positive cases. Thus, in the present study the use of AM and AM+CM may plausibly have led to less glaring findings compared to continuous SCC data. It frequently happens that cows may not be sampled for official TD milk analyses if one or more diagnosis of mastitis is recorded during the lactation. For instance, farmers could opt for culling or for an earlier drying off. In the Austrian system specifically, when AM or CM is recorded it means that a treatment was administered to the lactating cow, speeding up the recovery and likely reducing the potential damage due to inflammation process. Conversely, using milk SCC for identification of inflammations seems a less restrictive approach, allowing detection of LC change after the event(s) in those cows with: (1) considerable number of cumulated high SCC TD; and (2) likely no or few official recorded mastitis diagnoses. The distribution of the different mastitis events within lactation (Figure 2) supports the idea that binary health traits have a different connotation compared to SCC quantitative data. Understanding if there is interaction between parity and cumulative event on LC sounds interesting, even if a small number of TD is expected to be associated with multiple inflammation events in parity 1, inevitably leading to distorted results. Vice versa, it is plausible that very old cows, e.g. in parity >5, have a limited number of TD linked to 0 past inflammations. In conclusion, this study demonstrates that the lifetime change in LC can be considered a mammary memory indicator, representing the combined result of aging and mastitis events in Fleckvieh cows.

Acknowledgements

Egger-Danner C. and Fuerst-Waltl B. were supported by the Project D4Dairy (Digitalisation, Data integration, Detection and Decision support in Dairying) founded by Austrian Federal Ministry for Climate Action, Environment, Energy, Mobility, Innovation and Technology, Federal Ministry for Digital and Economic Affairs and provinces of Lower Austria and Vienna in the framework of COMET program (Austrian Research Promotion Agency, grant number 872039).

References

Costa A., Bovenhuis H., Penasa M. (2020) J Dairy Sci 103:11574-11584. https://doi.org/10.3168/jds.2020-18615

Egger-Danner C., Fuerst-Waltl B., Obritzhauser W., Fuerst, C., Schwarzenbacher H., *et al.* (2012) J Dairy Sci 95:2765-2777. https://doi.org/10.3168/jds.2011-4876

93. Estimation of genetic parameters for milk-related traits derived from automatic milking systems in Holstein cattle

L.S. Gloria[1*], J.P. Boerman[1], V.B. Pedrosa[1], S.Y. Chen[1,2], M.E. Montes Gonzalez[1], J.S. Doucette[3] and L.F. Brito[1]

[1]Department of Animal Sciences, Purdue University, West Lafayette, IN 47907, USA; [2]Farm Animal Genetic Resources Exploration and Innovation Key Laboratory of Sichuan Province, Sichuan Agricultural University, Chengdu 611130, China; [3]Agriculture Information Technology (AgIT), Purdue University, West Lafayette, IN 47907, USA; lgloria@purdue.edu

Abstract

We estimated genomic-based genetic parameters for daily milk yield (MY), milk electrical conductivity (EC), and somatic cell score (SCS) from automatic milking systems (AMS) in North American Holstein cattle. Random regression models (RRM) enable the estimation of genetic parameters for each lactation day as well as the genetic correlation coefficients among MY, SCS, and EC. The data contained records of 4,118 cows genotyped with 72,820 SNP. The heritabilities and genetic correlations for MY, SCS, and EC and were estimated fitting third-order Legendre orthogonal polynomial. The daily heritability estimates ranged from 0.07-0.28 for MY, 0.03-0.08 for SCS, and 0.39-0.50 for EC. The genetic correlation among MY and other traits had high variation, whereas the genetic correlations between SCS and EC were moderate and positive. These results contribute to a better understanding of the genetic variability of longitudinal traits derived from AMS and can be used when refining dairy cattle selection indexes.

Introduction

Automatic milking systems (AMS; also known as milking robots) are becoming more common in dairy farms around the world due to the improvements in cow welfare, increased productivity, and reduced labor (Samoré *et al.*, 2011). AMS also generates a large amount of data, including phenotypes for production, behavior, health, and milk quality traits (Barkema *et al.*, 2015). AMS provide longitudinal measurements of indicators such as milk electrical conductivity. In addition, the availability of repeated records enable the use of random regression models (RRM) to estimate genetic parameters for each lactation day. RRM provide robust genetic modelling of time-specific effects for longitudinal traits, with key advantages compared to other statistical models, such as: (1) prediction of genetic parameters and genetic correlation among traits for the times-series within the range of data collection (Oliveira *et al.*, 2019); (2) less computational demand; and (3) not assuming that all records represent the same trait.

Electrical conductivity (EC) measures the ability of a solution to conduct an electric current between two electrodes, and it is measured in milliSiemens (mS). AMS are equipped with sensors for measuring EC, which is significantly correlated with somatic cell score (SCS) – an indirect measure of mastitis (Martin *et al.*, 2018). Data from AMS have been used to estimate genetic parameters for milk yield traits (e.g. Dechow *et al.*, 2020), which are highly genetic correlated with traits derived from conventional milking systems (Mulder *et al.*, 2004). Due to their already existing importance for breeding programs, AMS-based data has been collected for evaluating animal health traits such as EC (Romero *et al.*, 2017). In this context, the present study aimed to estimate genetic parameters for daily milk yield, EC, and SCS derived from AMS in North American Holstein cattle.

Materials & methods

Dataset. The data was recorded using 36 Lely AMS between 2017 to 2021 and contained 1,752,713 milking records of 4,118 genotyped Holstein cows from a commercial dairy farm located in Plymouth, IN, USA.

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_93, © L.S. Gloria *et al.* 2022

The traits evaluated are total milk yield per milking (MY, kg/day); EC (mS/cm), which was averaged across quarters; and somatic cell count (SCC, cells/ml) measured from the average across quarters. SCC was log-transformed to SCS using the formula (Piwczyński *et al.*, 2021):

$$SCS = log_2\left(\frac{SCC}{10{,}000}\right) + 3 \quad (1)$$

Genetic analyses. The descriptive statistics of the data was done using the R software (R Core Team, v 4.1.2), after eliminating records deviating 3.5 SD from the mean as potential outliers. Genotype imputation was performed using the FImpute software (Sargolzaei *et al.*, 2014) to converge all SNP panels available into a common set of 72,820 SNPs. Additionally, we performed the quality control (QC) analysis to remove the individuals and SNPs with call rate lower than 0.90, minor allele frequency (MAF) lower than 0.05, and extreme deviation from Hardy-Weinberg equilibrium ($P<10^{-8}$) using the BLUPF90 family programs (Misztal *et al.*, 2018). After QC 57,600 SNPs and 4,118 animals remained for further analyses.

Variance components and genomic-based genetic parameters for each trait were estimated using the AIREMLF90 software (Misztal *et al.*, 2018). RRM was used to describe the covariance structure of the data as a function of days in milk (DIM) (de Boor, 1980; Kirkpatrick *et al.*, 1990; Meyer, 2005). In matrix notation, the model equation can be defined as:

$$y_{ijkl} = YS_k + LN_l + \sum_{m=1}^{m} PM_m \emptyset_m(t_{ij}) + \sum_{m=1}^{m} a_{im} \emptyset_m(t_{ij}) + e_{ijkl}\,, \quad (2)$$

where y_{ijkl} is the predicted AMS traits of i^{th} animal on DIM j within year-season and lactation number, YS and LN are the fixed effects of year-season and lactation number, respectively; PM_m and a_{im} are the m random regression coefficients that describes the permanent environment and the additive genetic effect for the i^{th} animal; t_{ij} is the time of data collection (DIM j); $\emptyset_m(t_{ij})$ is a regression function according to DIM j using third-order Legendre orthogonal polynomial (LP). The residual effect (e_{ijkl}) was assumed to be homogeneous throughout lactation and the random effects were assumed to be normally distributed with zero mean.

Results

Heritabilities for MY, SCS, and EC are shown in Figure 1, in which the standard errors ranged from 0.02 to 0.07. The heritability for daily MY was lower (approximately 0.06) up to DIM 60 and then increased up to 0.28 at 305 DIM. The heritability for SCS increased from 0.02 to 0.08. The heritability for EC was moderate to high (0.39 to 0.50) with a low standard deviation (≈0.02) across DIM estimates. This indicates that EC is highly heritable throughout the entire lactation.

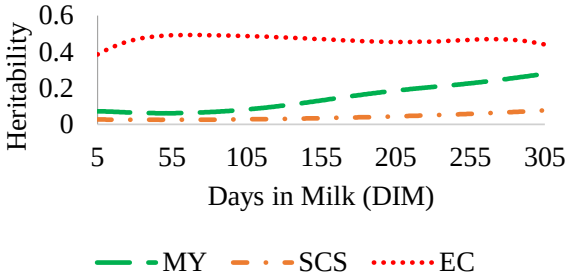


Figure 1. Daily heritability estimates of milk yield, somatic cell score, and milk electrical conductivity in North American Holstein cows. MY = Milk yield, kg; SCS = Somatic cell score, log2; EC = Milk electrical conductivity in mS/cm

Genetic correlations. The genomic-based genetic correlations among MY, EC, and SCS are presented in Figure 2, showing high variation among MY and the other studied traits. These correlation curves presented similar shape compared to a phenotypic lactation curve, increasing at the beginning until achieving a peak [MY and EC (0.42), MY and SCS (0.39)], and then decreasing until a minimum value [MY and EC (-0.32), MY and SCS (-0.42)]. On the other hand, the genetic correlations between SCS and EC are moderate and positive ranging from 0.23 to 0.42.

Discussion

The use of RRM enabled the estimation of time-dependent variance components, in which all three traits were shown to be heritable throughout lactation. However, a differential pattern on their genetic relationship across time was observed. The heritability for MY had a nonlinear relationship along the lactation, which is in agreement with previous reports such as for French Holstein cows milked by AMS (Tribout *et al.*, 2018). These authors using different methods and polynomials to model the random effects obtained heritability values ranging from 0.02 to 0.26 for MY, which are similar to those found in our study on North American Holstein cattle. In addition, a MY heritability of 0.11 was reported in US Holsteins milked using AMS (Dechow et. al., 2020). The heritability for SCS is in the lower range of the literature for Holstein cattle, but it is still heritable throughout the lactation. SCS has been used as an indirect measure of mastitis (Norberg *et al.*, 2004) and commonly used in dairy cattle breeding programs. EC has also been used for evaluating mastitis in dairy herds over the last decade. The use of EC as a selection trait is not new, Schukken *et al.* (1997) was the first to suggest EC as an indicator trait in a breeding program, and with increased use of AMS during this period there is a unique opportunity to closely monitor the mastitis status of dairy cows and incrementing the existing selection indexes for improved udder health. Previous studies have also observed that EC is moderately to highly heritable in other cattle populations (e.g. Piwczyński *et al.*, 2021).

Genetic improvement towards improving mammary gland health in dairy cattle is very important from both economical and animal welfare point of view. Previous studies have reported high variation and unfavorable genetic correlations between MY and SCS (Piwczyński *et al.*, 2021; Samoré *et al.*, 2011). According to Norberg *et al.* (2004), there are significantly differences for healthy, sub-clinically infected, and clinically-infected cows, as demonstrated by the means of EC (mS/cm) obtained by Norberg *et al.* (2004): 5.3 (±0.03), 5.75 (±0.04) and 6.73 (±0.06), respectively. The large amount of information generated through AMS combined with genomic information provide a great opportunity for selecting animals towards improved udder health, which will consequently contribute from both an economical and animal welfare perspective to the long-term sustainability of dairy farms.

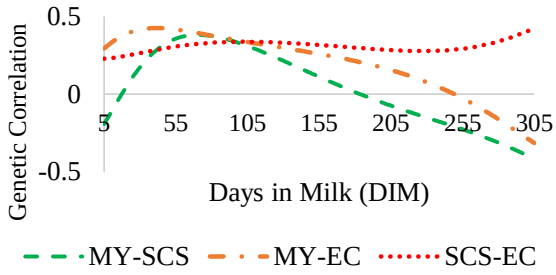


Figure 2. MY = Milk yield, kg; SCS = Somatic cell score, log2; EC = Electrical conductivity in milk, mS/cm

Acknowledgements

The authors thank Homestead Dairy, LLC and especially Brian and Jill Houin for inviting us on their farm and for allowing our research group to utilize their farm records. We also thank Lely for giving us access to their datasets. This activity was funded by Purdue University as part of AgSEED Crossroads funding to support Indiana's Agriculture and Rural Development.

References

Barkema H. W., von Keyserlingk M. A., Kastelic J. P., Lam T. J. G. M., Luby C. *et al.* (2015). Journal of dairy science, 98(11): 7426-7445. https://doi.org/10.3168/jds.2015-9377

Dechow C. D., Sondericker K. S., Enab A. A., and Hardie L. C. (2020). Journal of dairy science, 103(12): 11503-11514. https://doi.org/10.3168/jds.2020-18786

Martin P., Barkema H. W., Brito L. F., Narayana S. G., and Miglior F. (2018). Journal of dairy science, 101(3), 2724-2736. https://doi.org/10.3168/jds.2017-13554

Misztal I., Legarra A., and Aguilar I. (2009). Journal of dairy science, 92(9): 4648-4655. https://doi.org/10.3168/jds.2009-2064

Mulder H. A., Groen A. F., De Jong G., and Bijma P. (2004). Journal of dairy science, 87(5): 1487-1495. https://doi.org/10.3168/jds.S0022-0302(04)73300-7

Nixon M., Bohmanova J., Jamrozik J., Schaeffer L. R., Hand K. *et al.* (2009). Journal of Dairy Science, 92(7): 3422-3430. https://doi.org/10.3168/jds.2008-1689

Norberg E., Hogeveen H., Korsgaard I. R., Friggens N. C., Sloth K. H. M. N. *et al.* (2004). Journal of dairy science, 87(4): 1099-1107. https://doi.org/10.3168/jds.S00220302(04)732567

Oliveira H. R., Brito L. F., Lourenco D. A. L., Silva F. F., Jamrozik J., *et al.* (2019). Journal of dairy science, 102(9): 7664-7683.https://doi.org/10.3168/jds.2019-16265

Piwczyński D., Sitkowska B., and Ptak E. (2021). Animal, 15(2): 100094. https://doi.org/10.1016/j.animal.2020.100094

Romero G., Roca A., Alejandro M., Muelas, R., and Díaz J. R. (2017). Journal of dairy science, 100(2): 1555-1567. https://doi.org/10.3168/jds.2016-11544

Samoré A. B., Román-Ponce S. I., Vacirca F., Frigo E., Canavesi F. *et al.* (2011). Journal of dairy science, 94(8): 4081-4089. https://doi.org/10.3168/jds.2010-3611

Sargolzaei, M., Chesnais, J.P., and Schenkel, F.S. (2014). BMC Genomics, 15:478. https://doi.org/10.1186/1471-2164-15-478

Schukken Y.H., Lam T.J.G.M., Barkema H.M. (1997) Proc., International workshop on genetic improvement of functional traits in cattle health. Uppsala, Sweden.

Tribout, T., Minery, S., Dassé, B., Saunier, D., Ducrocq, V. *et al.* (2018). Proc. of the 11th WCGALP, Auckland, New Zealand.

94. Identification of genome regions and promising candidate genes linked to innate immune capacity on young Holstein calves

P. Martin[1], S. Fresco[1,2*], L. Jouneau[3], S. Taussat[2], A. Vinet[1], R. Lefebvre[1], D. Descamps[3], C. Ferret[3], C. Chottin[3], V. Pietralunga[3], T. Pezier[4], S. Lacroix-Lamandé[4], M. Boussaha[1], S. Barbey[5], V. Denizot[6], D. Boichard[1], S. Riffault[3] and F. Laurent[4]

[1]Université Paris-Saclay, INRAE, AgroParisTech, GABI, 78350 Jouy-en-Josas, France; [2]Allice, 149 rue de Bercy, 75012 Paris, France; [3]Université Paris-Saclay, INRAE, UVSQ, VIM, 78350 Jouy-en-Josas, France; [4]INRAE, UMR1282 Infectiologie et Santé Publique, 37380 Nouzilly, France; [5]INRAE, UE326 Domaine Expérimental du Pin, 61310 Gouffern en Auge, France; [6]Innoval, GDS Bretagne, 35538 Noyal-sur-Vilaine, France; solene.fresco@inrae.fr

Abstract

In order to investigate the genomic control of innate immune response, the concentration levels of 14 cytokines produced naturally or in response to whole blood stimulation with three pattern recognition receptors agonists were determined in 623 Holstein calves enrolled in the Healthycalf project. The effect of the immunostimulants and the relationship among the cytokines produced were observed through a principal components analysis, and 50K GWAS were performed on all cytokine × condition traits. Results showed that most cytokines are oriented together following the first component that explained 70% of the variability and partially discriminated the agonistic conditions. Additionally, 40 genome regions were found in association with at least one trait with some regions being largely shared between conditions and cytokines. Seven positional candidate genes were considered as especially interesting due to their functional link to innate immune response reported from the literature.

Introduction

In most national breeding programs, the included health traits relate to adult animal health. However, calf health is also a major economical and welfare burden. In France, depending on the breed, between 10 and 20% of the dairy calves die before six months of age (Leclerc *et al.*, 2016). If a large part of deaths occurs in the first 48h and can be related to calving, a considerable and variable part happens between 48h and 2 months of age with a detrimental influence of diarrheas (Jegou *et al.*, 2006). A better knowledge of the mechanisms and genetic determinism of calves' innate immune response could help reducing the prevalence of diarrheas and respiratory symptoms and possible subsequent deaths. In this study, we aim to identify genome regions and candidate genes associated with innate immune response through the concentration level of various cytokines.

Materials & methods

Between 2018 and 2021, a large immune monitoring study was performed on Holstein calves from various commercial farms and one experimental farm (https://doi.org/10.15454/1.5483257052131956E12) all located in the Brittany and Normandy regions of France. A total of 623 calves were sampled at 1-2 weeks of age by veterinarians or qualified technicians. The authors would like to especially acknowledge H. Chapuy, C. Cazals, K. Toukmidine and C. Mechin for this work. Blood samples from the jugular vein were collected in four monovette syringes (Sarstedt, Marnay, France) containing heparin and pre-filled with pattern recognition receptors-agonists (LPS, NOD2L, R848) or a non-stimulated condition with vehicle only. Samples were incubated at 38.5 °C for 24 h and plasma recovered stored at -20 °C for further cytokine determination. An additional sample of blood or cartilage (when twin births) was taken for genotyping. This study is part of the Healthycalf project that was funded by Apis-Gene (https://www.apis-gene.com/).

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_94, © P. Martin *et al.* 2022

Determination of cytokine concentrations. Concentrations of 15 cytokines/chemokines (IFN-γ, IL-1α et IL-1β, IL-1RA, IL-2, IL-4, IL-6, IL-8, IL-10, IL-17a, IP-10, CCL2, CCL3, CCL4 and TNFα) were measured in plasma samples with a custom bovine MILLIPLEX® MAP assay (SPRCUS706, Merck Millipore). Data were recorded on a MAGPIX® using xPONENT® software (Luminex). This system allows the quantification of a large number of samples and of 15 cytokines concomitantly. The wide assay range generally allows for accurate measurement of the 15 cytokines at a given plasma dilution (1/10). When a concentration was out of range (below or above), we assigned the minimum/maximum threshold for this cytokine. For IL-8, more than half the samples were out of range and therefore this cytokine was discarded from the analyses. For the other cytokines, the proportion of assigned samples varied from 1 to 22%.

Genotyping and quality controls. All animals were genotyped using the EuroGMD SNPchip designed for EuroGenomics (Illumina Inc., San Diego, CA, USA). DNA extraction and genotyping were performed at LABOGENA (www.labogena.fr, Jouy-en-Josas, France). The following quality controls were applied: an individual call rate higher than 95%, a SNP call rate higher than 90%, a minor allele frequency higher than 2% and genotype frequencies in Hardy-Weinberg equilibrium with $P>10^{-4}$. A total of 573 animals and 46, 910 SNP from the 29 autosomes passed the quality controls.

Statistical analysis. In order to visualize the effect of the condition (LPS, NOD2L, R848 or non-stimulated) and the relationships among the 14 cytokines, a principal component analysis (PCA) was performed using the miXOmics package of the R Software (Rohart *et al.*, 2017).

Genome-wide association study. Single-trait association analyses were performed on 64 traits, corresponding to the 14 cytokines × 4 conditions, plus 8 traits corresponding to the coordinates of each sample under each condition on the two first components of the PCA analysis. The following mixed linear model was used with the mlma option of GCTA software:

$$y= 1\mu + xb + u + e \quad (1)$$

where y is the vector of phenotypes; μ is the overall mean; b is the additive fixed effect of the SNP to be tested for association; x is the vector of predicted allele assays ranging from 0 to 2; $u \sim N(0,G\sigma_u^2)$ is the vector of random polygenic effects (with G being the genomic relationship matrix and σ_u^2 being the estimated polygenic variance based on the null model) and $e \sim N(0,I\sigma_e^2)$ is the vector of random residual effects (with I being the identity matrix and σ_e^2 the residual variance). In a first intend, SNP with –logP>5 were selected, this threshold being on the range of the chromosome-wide threshold of significance under a Bonferroni correction of α=5%.

Results

The results for the PCA are reported in Figure 1. The first axis accounts for more than 70% of the variability while the second axis explains only 10%. Most of the cytokine concentrations are correlated following the first component, which is also able to discriminate the different conditions, reflecting the overall intensity of immune reaction depending on the agonist. The second axis discriminates some specific cytokines (CCL2 and CCL4 versus IP10 and IFNγ) but to a lower extent. No structure is observed depending on the farm, the year or the calf sex (results not shown).

Considering the GWAS results for the 64 traits, a total of 145 SNP reached the threshold, corresponding to 40 different chromosomic regions. From these regions, 18 are detected associated with one trait only while six regions are shared between two traits, four between three traits, two between four traits and 10 regions, presented in Table 1, are shared between at least six traits. The same region was frequently

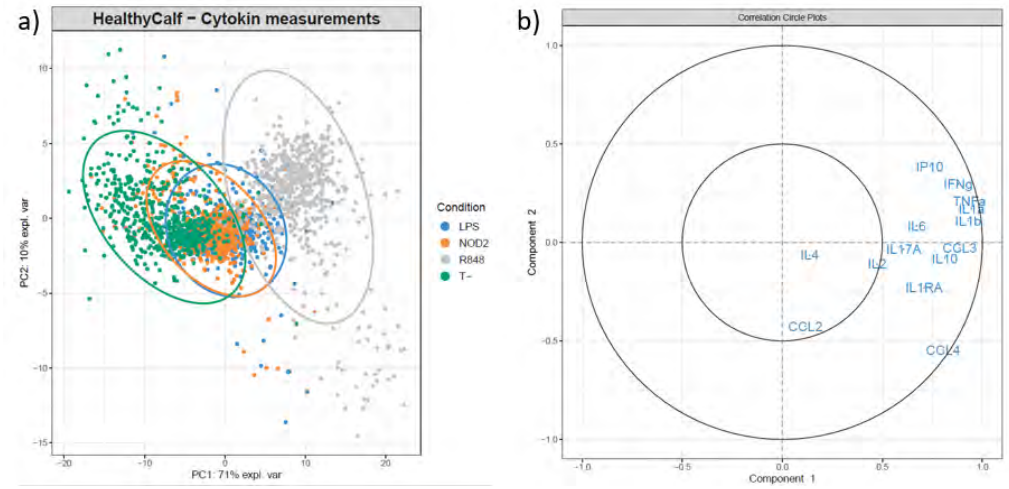


Figure 1. Principal component analysis results with (a) the plot of all samples colorized by condition and (b) the correlation circle plot among the cytokines.

Table 1. Details of the 10 regions identified by the GWAS associated with at least 6 different traits.

Chromosome	SNPmax Position (mb)	Minor allele frequency	Associated cytokines (with number of associated conditions)
2	36.0	0.13	IL1RA (4), IL6 (2)
2	72.7	0.12	IFNγ (1), IL1α(1), IL1RA (3), IL6 (3), TNFα (1)
3	70.4	0.09	IL1RA (4), IL6 (3), TNFα (1)
7	85.4	0.16	IL1α (1), IL1RA (4), IL6 (2), TNFα (1)
8	82.2	0.11	IL1RA (4), IL6 (2)
12	51.9	0.14	IL1RA (3), IL6 (3)
17	66.1	0.08	IFNγ (1), IL10 (1), IL1RA (4), IL6 (3), TNFα (1)
19	31.8	0.03	IL1RA (2), IL2 (2), IL6 (2), TNFα (1)
19	54.5	0.32	CCL4 (2), IFNγ (2), IL10 (2), IL1α (3), IL1β (3), IL1RA (2), IL2 (1), IL4 (4), PC1_LPS, TNFα (3)
24	12.5	0.24	IL17A (1), IL1α (1), IL1b (2), IL4 (1), PC1_LPS, TNFα (1)

associated with more than one condition of the same cytokine but different cytokines also shared some regions. The cytokines sharing the highest number of regions were the ones highly correlated with the first axis of the PCA, while cytokines like CCL2 were mostly associated with regions specific to them. It is however interesting to note that the number of regions associated with the PC traits was small (only five regions overall the eight traits).

Discussion

These results show that the mechanisms of innate immune response are complex with a genomic control that is partially shared between some cytokines but mostly specific for others and also with possible variations depending on the agonist. Among the numerous regions found associated with the traits here, some contain genes that have already been linked with immunity processes and are especially interesting candidate genes: LYAR (Yang *et al.*, 2019) and OTOP1 (Wang *et al.*, 2014), both located in a region of chromosome 6 associated with CCL4; SLC46A2 (Paik *et al.*, 2017), located in a region associated with

IL6 on chromosome 8; SEC14L (Li *et al.*, 2013) found in the largely shared region of chromosome 19; CXCL17 (Lee *et al.*, 2013) on chromosome 18 in a region associated with IL6; OTULIN (Fiil *et al.*, 2013) on chromosome 20 associated with IL4 and finally TREX1, on chromosome 22 associated with IL1β (Yan *et al.*, 2010). The next step will be to confront these cytokine profiles with the clinical status of the animals to determine if there is a link between the cytokine profile observed and the health status of the animals allowing the identification of predictive markers of sensitivity/resistance to diseases in young calves.

References

Fiil B.K., Damgaard R.B., Wagner S.A., Keusekotten K., Fritsch M., *et al.* (2013). Mol Cell 50:818–830. https://doi.org/10.1016/j.molcel.2013.06.004.

Jegou V., Porhiel J.Y., Brunschwig P., and Jouanne D. (2006). Proc. of 13eme Renc. Rech. Ruminants, Paris, France.

Leclerc H., Lefebvre R., Douguet M., Phocas F., and Mattalia S. (2016). Proc. of 23eme Renc. Rech. Ruminants, Paris, France.

Lee W.-Y., Wang C.-J., Lin T.-Y., Hsiao C.-L., and Luo C.-W. (2013). Am J Physiol Endocrinol Metab 304:E32-40. https://doi.org/10.1152/ajpendo.00083.2012.

Li M.-T., Di W., Xu H., Yang Y.-K., Chen H.-W., *et al.* (2013). J Virol 87:10037–10046. https://doi.org/10.1128/JVI.01073-13.

Paik D., Monahan A., Caffrey D.R., Elling R., Goldman W.E., *et al.* (2017). J Immunol 199:263–270. https://doi.org/10.4049/jimmunol.1600409.

Rohart F., Gautier B., Singh A., and Le Cao K.-A. (2017). PLOS Computational Biology 13:e1005752. https://doi.org/10.1371/journal.pcbi.1005752.

Wang G.-X., Cho K.W., Uhm M., Hu C.-R., Li S., *et al.* (2014). Diabetes 63:1340–1352. https://doi.org/10.2337/db13-1139.

Yan N., Regalado-Magdos A.D., Stiggelbout B., Lee-Kirsch M.A., and Lieberman J. (2010). Nat Immunol 11:1005–1013. https://doi.org/10.1038/ni.1941.

Yang C., Liu X., Cheng T., Xiao R., Gao Q., *et al.* (2019). J Virol 93:e00769-19. https://doi.org/10.1128/JVI.00769-19.

95. Assessment of miRNAs in plasma and leucocytes as potential biomarkers for rumen health in cows fed forage and high-grain diets

O.E. Ojo[1], C. Pacífico[1], S. Johanns[1], S. Ricci[1], R. Rivera-Chacon[1], A. Sener-Aydemir[1], E. Castillo-Lopez[1], N. Reisinger[2], Q. Zebeli[1] and S. Kreuzer-Redmer[1*]

[1]CD Lab Innovative Gut Health Concepts of Livestock, Institute of Animal Nutrition and Functional Plant Compounds, University of Veterinary Medicine Vienna, Austria; [2]BIOMIN Research Center, Tulln an der Donau, Austria; susanne.kreuzer-redmer@vetmeduni.ac.at

Abstract

MicroRNAs are a class of small non-coding RNAs, which regulate post-transcriptional gene expression, and could serve as biomarkers for several biological conditions. Sub-acute ruminal acidosis (SARA), a metabolic disorder characterized by low ruminal pH and the accumulation of volatile fatty acids, is linked to feeding of starch-rich diets. To evaluate the suitability of miRNAs as potential biomarkers in cattle for the diagnosis of SARA, a comprehensive identification of miRNA profiles in plasma and leucocytes was conducted by next generation sequencing technologies. Four dry Holstein cows were fed a control diet of 100% forage and subsequently transitioned to a 65% high-grain diet. In plasma and leucocytes, 520 and 730 miRNAs were found, respectively. Bta-miR-2285u, bta-miR-30b-3p and bta-miR-12034 were differentially expressed between diets, and therefore chosen as potential SARA biomarker candidates and further analysed by RT-qPCR in more cows. Bta-miR-2285 and bta-miR-30-3p could be confirmed as biomarker candidates for rumen health.

Introduction and background

MicroRNAs (miRNAs) are small (≈22 nucleotides long) non-coding RNAs which are crucial regulators of gene expression. These short transcripts limit protein translation through a process referred to as 'gene silencing', where miRNA binds to mRNA molecules and either transiently blocks translation of proteins, or degrades the bound mRNA (Duchaine and Fabian, 2019). Biofluids and other biological matrices, such as plasma, saliva, or faeces are common targets for biomarker profiling as they are obtained through minimally invasive techniques, and contain biologically important molecules which may be predictive of disease outcomes (Tesfaye *et al.*, 2016). Shaughnessy and colleagues reported that specific miRNAs had shown to be differentially expressed in faecal samples of colorectal cancer cases relative to those from healthy human controls (Shaughnessy *et al.*, 2020). MiRNAs are conserved between different species and target most of the protein-coding genes that are involved in various developmental and biological processes (Sun *et al.*, 2019). MicroRNAs in bovine are suggested to play important roles in diverse molecular functions and biological processes which includes rumen development, adipogenesis, mammary gland and ovary development, immunity, and infection. They can also serve as biomarker for production traits (e.g. milk yield and quality), and could be used for early diagnosis of pregnancy (Sun *et al.*, 2019). MicroRNA biomarkers are referred to as the especially produced or enriched miRNAs in a particular tissue under a specific biological condition. Ioannidis and Donadeu reported and identified the miR-802 in cattle, a circulating miRNA shown to regulate insulin sensitivity and lipid metabolism, that could also potentially serve as a specific biomarker of liver function, a key parameter in the context of post-partum negative energy balance in dairy cows (Ioannidis and Donadeu, 2018). The discovery of novel biomarkers in association with specific production traits and disease conditions in cattle is of significant agricultural importance.

The current definition and ruminal pH threshold for SARA varies among studies. However, SARA usually occurs when ruminal pH remains in the range of 5.2 and 6 for a longer period. In this study, SARA was defined as pH disturbances below a threshold of ruminal pH<5.8 for 320 min/d. Despite being reportedly

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_95, © O.E. Ojo *et al.* 2022

associated with a systemic inflammation of different organs and tissues in dairy cows, there is no typical clinical sign of illness in SARA-affected cows (Abdela, 2016). Thus, biomarkers are pressingly needed to identify cows that are at a greater risk of developing SARA. Therefore, the main aim of this study was to comprehensively identify miRNA profiles in plasma and leucocytes to evaluate miRNAs as biomarkers in cattle exposed to SARA due to high-grain feeding during a prolonged period.

Materials & methods

Blood samples were collected in EDTA-coated tubes (Becton Dickinson, USA) from the jugular vein of four ruminal-cannulated Holstein cows, average age of 11.3±2.29 years, with a weight of 909.75±98.17 kg. Cows were fed a diet of 100% forage and were transitioned after one week of adaptation to a 65% high-grain diet, which was fed for a total of 4 weeks. The animals were held at the research and training farm of the University of Veterinary Medicine Vienna in Kremesberg, Austria. The trial was approved by the ethical committee, with the protocol number BMBWF-68.205/0033-V/3b/2019. Cell-free plasma and leucocytes were obtained using a two-step centrifugation protocol. Total RNA was isolated using the NucleoSpin miRNA Kit (Machery-Nagel) following the manufacturers protocol with few modifications.

Small RNA libraries were prepared from blood plasma and blood leucocytes using the NEXTflex Small RNA-Seq Kit (Bioo Scientific). A 50 bp single end sequencing approach was performed on an Illumina NovaSeq 6000 platform by the company CeGat (Tübingen, Germany). Read processing and pre-filtering was conducted in sRNAbench and R software 4.0.2. Normalised reads count (reads per million mapped, RPMM) obtained from sRNAbench were filtered (>10 reads) in R using the *tidyr* (version 1.1.4) package. Reads with an average PHRED score below 20 were discarded and all miRNAs that had less than 10 total read counts across the studies samples within each cow were filtered out. Bioinformatics and differential expression analysis was carried out using DESeq2 (version 4.1) (Love *et al.*, 2014). Differentially expressed (DE) miRNAs were considered if false discovery rate (FDR)≤0.05 and a |log2 (fold change) | ≥1. TargetScan (McGeary *et al.*, 2019) and miRBase (Griffiths-Jones, 2006) were used to predict DE miRNA target genes. The filtered and sorted data were further analysed using Venny 2.1 (Oliveros, J.C, 2007-15) for identification of unique and shared miRNAs across animals and diets. The four cows plus one additional cow was tested for RT-qPCR to confirm the accuracy of the sequencing results. Furthermore, a later time point of the experiment, week 4, was included. Quantification of miRNA was done by a poly-A-technique according to the manufacturer's protocol (miRNA 1st-Strand cDNA Synthesis Kit, Agilent Technologies). Therefore, 1 µg of total RNA was polyadenylated for 30 min at 37 °C terminated by 95 °C for 5 min in a total volume of 10 µl. In a second step, 10 µl of the elongated miRNA was reverse transcribed using the AffinityScript RT and the Universal Reverse Primer (3.125 µM) in 20 µl volume. MiRNA-specific primers were designed and validated. RT-qPCR assays were performed using the miRNA QPCR Master Mix (Agilent) on a ViiA7 instrument (Life technologies). MiR-107 and miR-103 served as reference miRNAs to normalize for RNA content. Relative expression was calculated using the ddCT-method. Relative values were calculated relative to the mean expression of the forage-based diet.

Results & discussion

In total, 520 miRNAs were discovered in plasma, with 360 miRNAs (Figure 1a) shared between forage (week1) and high grain (week2), while 730 miRNAs were identified in leucocytes, with 650 miRNAs (Figure 1b) shared between weeks 1 and 2 in all cows. When the cows were fed the high-grain diet, 63 miRNAs were identified in plasma, and they are solely expressed when cows are fed a high grain diet, resulting in higher circulating miRNAs. A total of 730 miRNAs were found in leucocytes, with 650 miRNAs shared between weeks 1 and 2. 17 miRNAs were only expressed when the forage-based diet (week 1) was fed, while 34 miRNAs were only expressed during high-grain feeding (week 2).

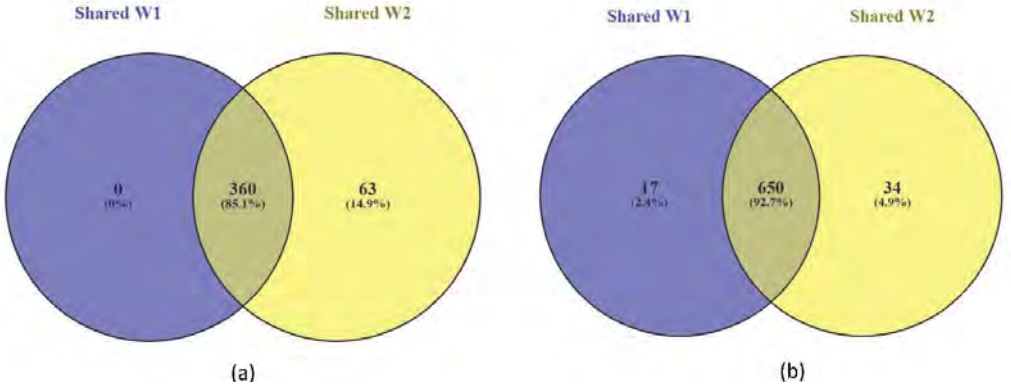


Figure 1. A schematic diagram of shared miRNAs in (a) plasma and (b) leucocytes between forage and high-grain diet.

Twelve miRNAs were differentially expressed in plasma samples (forage vs high-grain diet), from which 10 miRNAs were found to be up-regulated and 2 miRNAs were down-regulated. A total of 25 miRNAs were differentially expressed in leucocytes (forage vs high-grain diet). From these, 5 miRNAs were upregulated, and 20 miRNAs were found to be downregulated. The miRNAs bta-miR-2285u, bta-miR-30b-3p and bta-miR-12034 were chosen for further analyses. RT-qPCR validation included a further sampling time point at week 4, where the cows were already fed high grain for three weeks. These candidate miRNAs obtained from plasma were selected with respect to the total read counts of the filtered and sorted miRNAs in all the cows fed on a high grain diet (n=63), as well as the differentially expressed miRNAs (n=12) in terms of their read count in increasing order and log2 fold change. A Kruskal-Walli's test showed a strong influence (P=0.035) of the high-grain diet on the expression of miR-2285u (Figure 2A). Bta-miR-2285, a huge class of miRNAs, is believed to be rumen specific (Sun *et al.*, 2019). Hence miR-2285-u could be a promising circulating-miRNA candidate biomarker for rumen health. Furthermore, we found a higher expression of miR-30b-3p on week 2 (P=0.098), compared to the forage feeding at week 1 (Figure 2B). The same direction of effect was detected within the next generation sequencing data. On week four, after three weeks of high-grain feeding, the expression level was again at the level of forage feeding. Nevertheless, bta-miR-30b-3p has the potential to be a good candidate to detect short-term imbalances due to high-grain diet. Our result goes in line with the study of Muroya and colleagues, who identified bta-miR-30b-3p to be highly expressed in plasma of cattle fed a high-grain diet compared to grazing cattle (Muroya *et al.*, 2016). In the case of bta-miR-12034 we could not confirm a regulation by feeding high-grain diets through qPCR (Figure 2C).

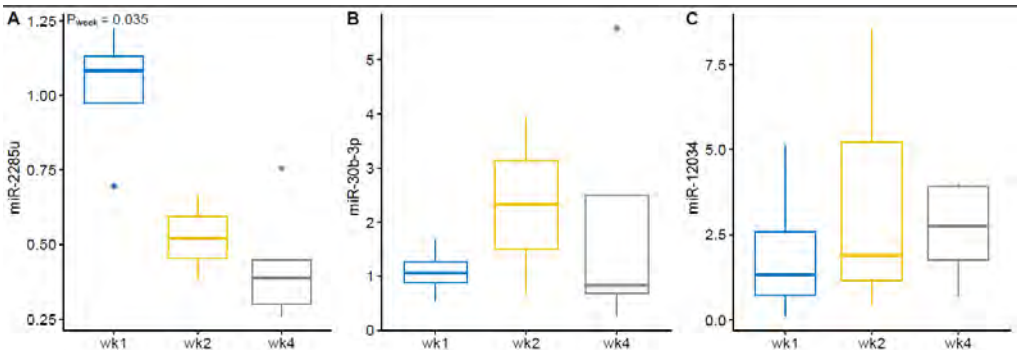


Figure 2. Relative miRNA expression of selected bovine miRNAs (A: miR-2285u, B: miR-30b-3p, C: miR-12034) by RT-qPCR in blood plasma of cows (n=5) at sampling time point week 1, when cows were feed a forage based diet and at sampling time points week 2 and week 4, when cows were fed a high-grain diet with 65% concentrate.

Conclusions

Our findings show that dietary modifications affect the release and expression of miRNAs in systemic circulation, potentially modulating post-transcriptional gene expression in SARA-affected cattle. MicroRNAs bta-miR-2285 and bta-miR-30-3p appear to be good candidates to serve as biomarker for rumen health and will be further analyzed.

References

Abdela, N. (2016) Achievements in the Life Sciences 10(2): 187–196. https://doi.org/10.1016/j.als.2016.11.006

Duchaine, T. F., and Fabian, M. R. (2019) Cold Spring Harbor Perspectives in Biology 11(3). https://doi.org/10.1101/cshperspect.a032771

Griffiths-Jones, S. (2006) Nucleic Acids Research 34(90001): D140–D144. https://doi.org/10.1093/nar/gkj112

Ioannidis, J., and Donadeu, F. X. (2018) BMC Genomics 19. https://doi.org/10.1186/s12864-018-4646-5

Love, M. I., Huber, W., and Anders, S. (2014) Genome Biology 15(12): 550. https://doi.org/10.1186/s13059-014-0550-8

McGeary, S. E., Lin, K. S., Shi, C. Y., Pham, T. M., Bisaria, N., *et al.* (2019) Science. https://doi.org/10.1126/science.aav1741

Muroya, S., Shibata, M., Hayashi, M., Oe, M., and Ojima, K. (2016) PloS One 11(9): e0162496. https://doi.org/10.1371/journal.pone.0162496

Shaughnessy, R. G., Farrell, D., Stojkovic, B., Browne, J. A., Kenny, K. *et al.* (2020) Scientific Reports 10(1): 5908. https://doi.org/10.1038/s41598-020-62843-w

Sun, H.-Z., Chen, Y., and Guan, L. L. (2019) Scientific Data 6(1): 190013. https://doi.org/10.1038/sdata.2019.13

Tesfaye, D., Salilew-Wondim, D., Gebremedhn, S., Sohel, M. M. H., Pandey, H. O., *et al.* (2016) Reproduction, Fertility, and Development 29(1): 8–23. https://doi.org/10.1071/RD16266

96. Variance component estimation of longitudinal alpha diversity metrics of the ocular microbiome in preweaned beef cattle

A.D. Lakamp*, A.C. Neujahr, M.M. Hille, J.D. Loy, S.C. Fernando and M.L. Spangler

University of Nebraska-Lincoln, 3940 Fair St, Lincoln, NE, USA; alakamp2@unl.edu

Abstract

The objective of this study was to examine the relationship between host genetics and four alpha diversity measures of the ocular microbiome in beef cattle. Ocular microbiome samples were taken from 223 preweaned calves at four time points (T1-T4). The average age at T1 and T4 were 65 and 204 days, respectively. The alpha diversity metrics used were observed diversity, the Chao1 index, the Simpson index, and the Shannon index. Variance components were estimated for each diversity metric at each time point using univariate ssGBLUP animal models. Heritability estimates for observed and Chao1 were moderate at T1 (0.35 and 0.36, respectively), high at T2 (0.58 and 0.68, respectively), zero for T3, and increased at T4 (0.45 and 0.55, respectively). Heritability estimates for the Simpson and Shannon indexes were moderate at T1 (0.42 and 0.50, respectively), low to moderate at T2 (0.14 and 0.30, respectively), and zero at T3 and T4.

Introduction

In beef cattle, the composition and diversity of the ocular microbiome are of interest due to their possible relationship with infectious bovine keratoconjunctivitis (IBK), commonly known as bovine pinkeye. Previous research has focused on the effect host genetics on the incidence of IBK infection (Snowder *et al.*, 2005); however, to our knowledge, the relationship between host genetics and ocular microbiome composition and diversity has not been previously investigated. Nonetheless, several studies have demonstrated the abundance of several ruminal taxa to be heritable (Li *et al.*, 2019; Sasson *et al.*, 2017; Wallace *et al.*, 2019). Moreover, the alpha diversity of ruminal bacteria and archaea is also heritable (Li *et al.*, 2019). Therefore, it is possible host genetics also affect the diversity and composition of the ocular microbiome.

The objective of this study was to estimate the degree to which different alpha diversity metrics of the ocular microbiome were heritable in preweaned beef calves.

Materials & methods

Animals, study design, and phenotypes. Data used for the current study was previously described by Bartenslager *et al.* (2021). Briefly, a group preweaned beef calves (n=223) comprised of varying percentages of Angus, Simmental, Red Angus, Hereford, and Gelbvieh from a single herd-year-season of birth were used. Animals were randomly administered one of three IBK vaccine treatments. The corneal and conjunctival surfaces were sampled using an eye swab in each calf at four time points (T1-T4). The first swab was prior to vaccination (day 0) when the average calf age was 65±15 days. Swabs at T2 and T3 corresponded to day 21 (vaccination booster) and 41, respectively, post initial swab collection. The final swab was taken on day 139 when the average calf age was 204±15 days.

Each ocular microbiome sample was subjected to DNA extraction and amplification of the V4 region of the 16S rRNA gene using barcoded primers. This was followed by sequencing of the V4 region on the Illumina MiSeq platform using 250 bp paired-end sequencing strategy (Kozich *et al.*, 2013). Data analysis was performed after quality filtering using the DADA2 pipeline (Callahan *et al.*, 2016) and the phyloseq package v1.26.1 (McMurdie and Holmes, 2013) in R software. This method resulted in groups of microbes collectively known as amplicon sequence variants (ASVs). Finally, ASV taxonomy was assigned using the SILVA reference alignment v138.

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_96, © A.D. Lakamp *et al.* 2022

The alpha diversity metrics selected as phenotypes for the current study were observed diversity, the Chao1 diversity index (Chao and Chiu, 2016), the Simpson index of diversity (Simpson, 1949), and the Shannon diversity index (Shannon, 1948). The diversity indexes can be interpreted as follows; observed diversity is the number of different ASVs identified in the sample, Chao1 estimate is the diversity estimate after accounting for rare species in the sample. For both observed and Chao1 a higher value represents greater diversity. In the case of Shannon and Simpson indexes, both richness (number of species present) and abundance (number of individuals per species) are considered. Given Shannon diversity is based on measuring uncertainty, a low diversity will result in a low uncertainty value, and therefore, as diversity increases the Shannon index would increase. However, the Simpson index uses the mean proportional abundance information of each species. Therefore, as the number of species increases the proportional abundance for a given species decreases. As such, as diversity increases the Simpson index would decrease. Thus, 1-Simpson index was used as the diversity measurement which would increase with increased diversity, and in this sense, would be similar in interpretation as the other indexes used.

Data were edited to remove animals with different vaccination treatments on day 0 and day 21. Additionally, not all animals had observations for all time points (n=211, 211, 187, and 189, for T1-T4, respectively). In total, 223 animals had observations for at least one time point.

Statistical analysis. A pairwise Wilcoxon rank sum test was performed to determine the statistical significance differences between time points within each metric using R software (R Core Team, 2021).

A four-generation pedigree was constructed for all 223 animals. This resulted in a pedigree of 1158 unique animals. Genotypes were imputed to a common 50K platform by the American Simmental Association and were available for 434 of the animals in the pedigree, including 170 of the animals with phenotypes. Genotypic filtering removed any SNP with a call rate less than 0.90 and a minor allele frequency less than 0.05. In total, 42,897 SNP remained for analysis.

Variance components were estimated using univariate animal models in the BLUPF90 suite of programs (Mitszal *et al.*, 2002). A single-step GBLUP framework (Aguilar *et al.*, 2010) was used to enable using phenotypes from animals with and without genotypes available. Fixed effects included sex, IBK vaccination treatment, and linear covariates of calf age at the time of sampling, individual and dam breed fraction (derived from pedigree), and individual and maternal heterosis. Random effects included a direct additive genetic component and a residual.

Results

The average value of each metric at each time point is detailed in Table 1. In general, alpha diversity decreased from T1 to T2 and then increased such that alpha diversity at T4 was similar to that at T1.

Heritability estimates (Table 2) varied depending on metric and time point. The observed diversity and the Chao1 index followed a similar pattern. At T1, estimates for both traits were moderate. At T2 heritability estimates for both traits increased and were high. As alpha diversity decreased at T3, so did the additive genetic. Thus, the heritability estimates were zero. However, at T4 heritability estimates increased to levels similar to T1 and T2.

Heritability estimates for the Simpson and Shannon indexes at T1 were moderate to high. Both traits exhibited a decrease in additive genetic variance at T2 where heritability estimates were low to moderate. Similar to the observed and Chao1, the heritability estimates at T3 for the Simpson and Shannon indexes were zero. In contrast to the observed and Chao1, heritability estimates at T4 for the Simpson and Shannon indexes remained near zero.

Table 1. Average measurements for the alpha diversity metrics of observed diversity, Chao1 index, Simpson index, Shannon index for the ocular microbiome of preweaned calves at different ages.

	Average measurement[1] (standard deviation)			
Avg. calf age (in days)	**Observed diversity**	**Chao1 index**	**Simpson index**	**Shannon index**
65	158.0 (124)[a]	143.7 (97)[a]	0.5898 (0.22)[a]	1.917 (1.2)[a]
86	27.01 (31)[b]	28.29 (31)[b]	0.3276 (0.15)[b]	0.7415 (0.40)[b]
106	65.50 (48)[c]	63.99 (43)[c]	0.4502 (0.16)[c]	1.140 (0.58)[c]
204	216.9 (134)[d]	193.2 (104)[d]	0.6047 (0.22)[a]	2.309 (1.1)[d]

[1] The greater the measurement, the greater the alpha diversity.

[2] Different superscript letters within a column indicate significantly different means determined by Wilcoxon rank sum test.

Table 2. Heritability estimates (standard error) for the alpha diversity metrics of observed diversity, Chao1 index, Simpson index, Shannon index for the ocular microbiome of preweaned calves at different ages.

	Heritability estimate (standard error)			
No. animals	**Observed diversity**	**Chao1 index**	**Simpson index**	**Shannon index**
211	0.36 (0.23)	0.35 (0.22)	0.50 (0.21)	0.42 (0.22)
211	0.58 (0.25)	0.68 (0.25)	0.14 (0.19)	0.30 (0.21)
187	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)
189	0.45 (0.27)	0.55 (0.28)	0.0 (0.0)	0.05 (0.20)

Discussion

In the current study, each ocular sampling can be considered as a perturbation that altered the microbiome of the eye. Bartenslager *et al.* (2021) suggested that the decrease in alpha diversity from T1 to T3 was related to the perturbation caused by the ocular swabbing conducted during sampling. Further, the authors suggested that the increase in diversity from T2 to pre-perturbation levels at T4 was evidence that the ocular microbiome was slowly re-establishing when allowed to recover without perturbation.

Heritability estimates of observed diversity and the Chao1 index followed a very similar pattern across time points. This is because Chao1 is a nonparametric estimation of species richness and was intended to achieve the same goal as observed diversity but be more sensitive to rare species (Chao and Chiu, 2016). However, as described above, both the Simpson index of diversity and the Shannon index include both richness and abundance. This resulted in the two measurements, and their corresponding heritability estimates, to behave similarly.

In general, all metrics were moderately to highly heritable at T1. This indicates that at a young age, the alpha diversity of the ocular microbiome may be partially influenced by host genetics. At T2, heritability estimates of the metrics which focus on richness (observed and Chao1) increased, while those that included relative abundance (Simpson and Shannon) began to decrease. This increased heritability with decreased diversity suggests that the dominant bacterial species present during this time period may be under host genetic control. Heritability estimates at T3, post-vaccination booster, were zero across all alpha diversity metrics suggesting that the perturbation may have resulted in allochthonous microbes from the environment to colonize the eye with overdominance of some bacterial species. Despite alpha diversity increasing at T4 to levels similar to that of T1, only observed diversity and the Chao1 index showed increased heritability. This may be because over time autochthonous microbes may have increased in abundance and outcompeted the allochthonous microbial species present in the ocular microbiome establishing microbial species that

were more under host genetic control. However, the reason for the absence of increased heritability in the Simpson and Shannon indexes at T4 may be because although similar microbes were identified at T4, the abundance of these microbes may have been different compared to T1.

Host genetics does appear to influence the alpha diversity of the ocular microbiome in young beef calves. Additional cohorts are being sampled and will add to the current study. Further investigations will include estimating genetic correlations between alpha diversity metrics within a time point and between time points within a diversity metric and estimating the heritability of different microbial taxa (genus, family phylum) within the bovine ocular microbiome.

References

Aguilar I., Misztal I., Johnson D.J., Legarra A., *et al.* (2010) J Dairy Sci 93:743–752. https://doi.org/10.3168/jds.2009-2730

Bartenslager A.C., Althuge N.D., Loy J.D., Hille M.M., *et al.* (2021) Anim Microbiome 3. https://doi.org/10.1186/s42523-021-00079-3

Callahan B.J., McMurdie P.J., Rosen M.J., *et al.* (2016) Nat Methods 13:581–583. https://doi.org/10.1038/nmeth.3869

Chao A., and Chiu C.H. (2016) Nonparametric estimation and comparison of species richness. In: eLS. John Wiley & Sons Ltd, Chichester, UK, pp. 1-11. https://doi.org/10.1002/9780470015902.a0026329

Kozich J.J., Westcott S.L., Baxter N.T., Highlander S.K., and Schloss P.D. (2013) Appl Environ Microb 79(17):5112-5120. https://doi.org/10.1128/AEM.01043-13

Li F., Li C., Chen Y., Liu J., *et al.* (2019) Microbiome 7. https://doi.org/10.1186/s40168-019-0699-1

McMurdie P.J., and Holmes S. (2013) PLOS ONE. 8(4). https://doi.org/10.1371/journal.pone.0061217

Misztal I., Tsuruta S., Strabel T., Auvray B., *et al.* (2002) Proc. of the 7th WCGALP, Montpellier, France.

R Core Team, R: A language and environment for statistical computing. Available at: https://www.R-project.org/.

Sasson G., Ben-Shabat S.K., Seroussi E., Doron-Faigenboim A., *et al.* (2017) mBio 8(4). https://doi.org/10.1128/mBio.00703-17.

Shannon C.E. (1948) Bell Syst Tech J 27:379–423.

Simpson E.H. (1949) Nature 163:688.

Snowder G.D., Van Vleck L.D., Cundiff L.V., and Bennett G.L. (2005) J Anim Sci 83:507–518. https://doi.org/10.2527/2005.833507x.

Wallace R.J., Sasson G., Garnsworthy P.C., Tapio I., *et al.* (2019) Sci Adv 5. https://doi.org/10.1126/sciadv.aav8391

97. Gene expression patterns observed from the skin of selected cattle breeds infested with *Amblyomma hebraeum* ticks

M. Letuka[1], D. Wragg[2,3], A. Bester-van der Merwe[1], K. Dzama[1], L.T. Nesengani[4], O. Matika[2,3], B. Dube[5] and N.O. Mapholi[4*]

[1]Department of Animal Sciences, University of Stellenbosch, South Africa; [2]Centre for Tropical Livestock Genetics and Health (CTLGH), Roslin Institute, University of Edinburgh, Easter Bush Campus, EH25 9RG, United Kingdom; [3]The Roslin Institute and R(D)SVS, University of Edinburgh, Easter Bush Campus, EH25 9RG, United Kingdom; [4]Department of Agriculture and Animal Health, University of South Africa, South Africa; [5]Agricultural Research Council, South Africa, Private X2, Irene, 0062, Pretoria, South Africa; maphon@unisa.ac.za

Abstract

This study aimed to identify genes that mediate immune responses in tick-tolerant Nguni, tick-susceptible Angus, and F2 Nguni X Angus population, infested with *Amblyomma hebraeum* ticks. Skin biopsy samples were collected prior and at post-infestation. RNA was isolated from the skin biopsies and sequenced on the Illumina HiSeq2500 Illumina sequencer. Overall, 139 significantly differentially expressed genes were identified in the current study. Calcium signalling pathway was the most enriched pathway linked to Five genes (PHKG1, PHKA1, CAMK21, NOS2, HRC). NOS2 gene was highly (significantly, $P>0.05$) expressed in the susceptible group before tick infestation and highly expressed in resistant group after tick infestation. Similar pattern was observed with HRC gene. HRC gene has been previously reported to play a role in regulating immune response against infections. This research reports differential gene expression following tick bites, thus providing preliminary data concerning the underlying mechanisms associated with the immune response.

Introduction

Ticks and tick-borne diseases (TTBDs) are a global problem and considered as one of the major challenges to livestock health and performance (Jongejan and Uilenberg 2004). The infestation of cattle by ticks can lead to weight loss, infertility, damage to hides and even death (Mapholi *et al.* 2016). *Amblyomma hebraeum* are native and common in Southern Africa, commonly known as the bont tick (Mapholi *et al.* 2016). The tick diseases are a major impediment to improvement of livestock industry particularly in Africa, due to a large number of tick species present and the variety of associated diseases (Walker 2003). Numerous tick control strategies have been employed including the application of acaricides. However, development of resistance to chemicals by ticks limit the effectiveness of this control strategy. This has created rising demands for different approaches that can alleviate the negative effects of ticks on livestock production thereby enhancing the livestock industries' contribution to the world's economy (Mapholi *et al.* 2014) (Porto Neto *et al.* 2011). Understanding the pattern of gene expression associated with tick resistance and susceptibility will create an opportunity to develop genomic selection programs for tick resistance as a complementary approach that is sustainable in the long term (Piper *et al.* 2008). Furthermore, genomic selection for host resistance may be an alternative for tick control. The current study identified genes that mediate immune responses to cattle infested with *Amblyomma hebraeum* ticks. This approach provides tools that will allow selection towards host resistance.

Materials & methods

Study design. This study was carried out at the Agricultural research council (ARC), South Africa. Ethical approval for the present study was obtained from the ARC- Institute's ethics committee. A total of 18 cattle (Nguni, Angus and F2 progeny from Nguni-Angus cross) were selected for the current study. Pathogen

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_97, © M. Letuka *et al.* 2022

free *Amblyomma hebraeum* larvae, were used in the artificial infestation of the animals for a period of 10 days. Cotton-stoppered vials containing tick larvae (100 larvae per animal) were placed into the calico bags secured to the shaven area on the animal's back. On the 11th day, the number of engorged ticks and the number of ticks that did not survive were counted. The animals were categorized as: high (1-15 ticks), moderate (16-45) and low (46-100) according to the yield of engorged ticks at the end of the artificial infestation period (Mapholi *et al.* 2016).

Skin biopsy and RNA sample collection and sequencing. Skin biopsy samples were collected before infestation and 72 hours post-infestation. RNA was isolated from the skin biopsies following Trizol reagent (Invitrogen) manufacturer's instructions and sequenced on the Illumina HiSeq2500 Illumina sequencer. Analyses were conducted in R using DeSeq2 to identify differentially expressed genes between resistant and susceptible animals. FASTQC v11.9 was used to assess the quality of the raw reads. The reads were aligned to the reference genome *Bos taurus* ARS-UCD1.2 using STAR v2.7.3a. Functional Analysis was performed using DAVID v6.8 (Database for Annotation Visualisation and Integrated Discovery).

A total of 576,676,968 reads were observed from 32 samples after trimming. A principal components analysis (PCA) plot generated using the qqplot package of DESeq2 was used to visualize similarities and variations between pre and post infestation samples (Figure 1). DESeq2 was employed to identify differentially expressed genes (DEGs) by comparing the skin biopsy samples at different time points.

Results & discussion

The overall tick count indicated a considerable variation in the number of ticks fed in animal within a period of 10 days after tick infestation. The overall mean tick count observed amongst all the animals was 33. The Angus group was a low tick resistance group and had an average tick count of 118. The high resistant group had the lowest number of engorged ticks with average tick count of 2. There are several factors that can affect the likelihood of a tick feeding successfully on a host. The breed effect, age of the host and the age of the ticks, a study by Jones and the team (Jones *et al.* 2015) reported that feeding success increases with the host age and decreases with tick age. The study further highlighted that prior exposure of hosts to ticks stimulates immunological responses that deter tick feeding (Jones *et al.* 2015), and that a non-defensive physiological change such as anaemia in the host can reduce blood feeding.

A total of 22,147 differentially expressed genes were detected. At the corrected *P*-value (P_{adj}<0.05), the number of genes that were significantly differentially expressed was 159. Functional enrichment analysis was performed for the identified 159 genes using DAVID Bioinformatics Resources 6.8. The analysis successfully assigned 138 genes to *B. taurus*. The remaining 21 genes were unassigned and therefore excluded from further analysis.

A total of 62 out of 138 genes (45%) were assigned to 23 KEGG pathways. The most enriched pathway was the Calcium signalling pathway.

This pathway revealed 10 genes (NCX, ROC, MCU, SERCA, HRC, TnC, MLCK, PHK, CAMK and NOS2). The 10 genes were compared to the list of 138 differentially expressed genes and only 5 genes namely, PHKG1, PHKA1, CAMK2A, NOS2 and HRC could be linked to the calcium signalling pathway. The five genes were used to create boxplots to compare the pattern across the study groups (Figure 2).

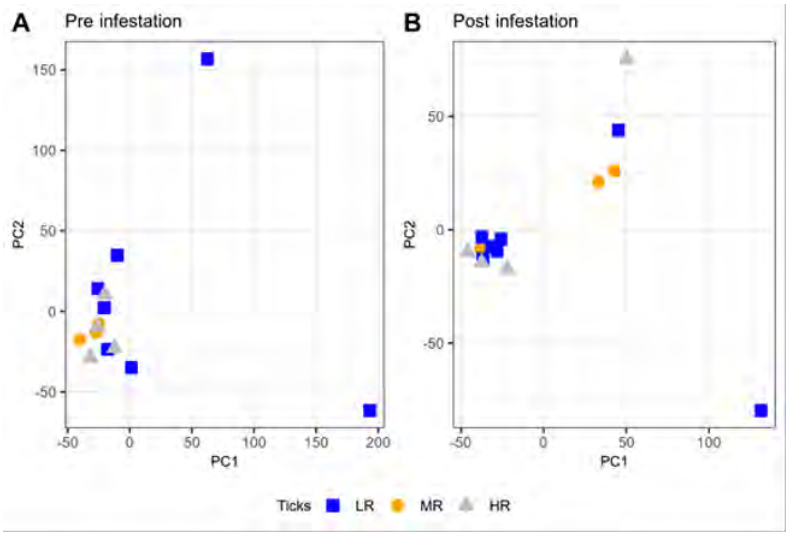


Figure 1. PCA plot depicting the effect of time on gene expression variation. The blue squares (LR=Low Resistant) on this plot refer to the tick susceptible groups, which is the Angus group and the LR group. The yellow circles refer to the moderate resistant group (MR=Moderate Resistant). The grey triangles (HR= High Resistant) refer to the tick tolerant groups, which is the Nguni and HR groups.

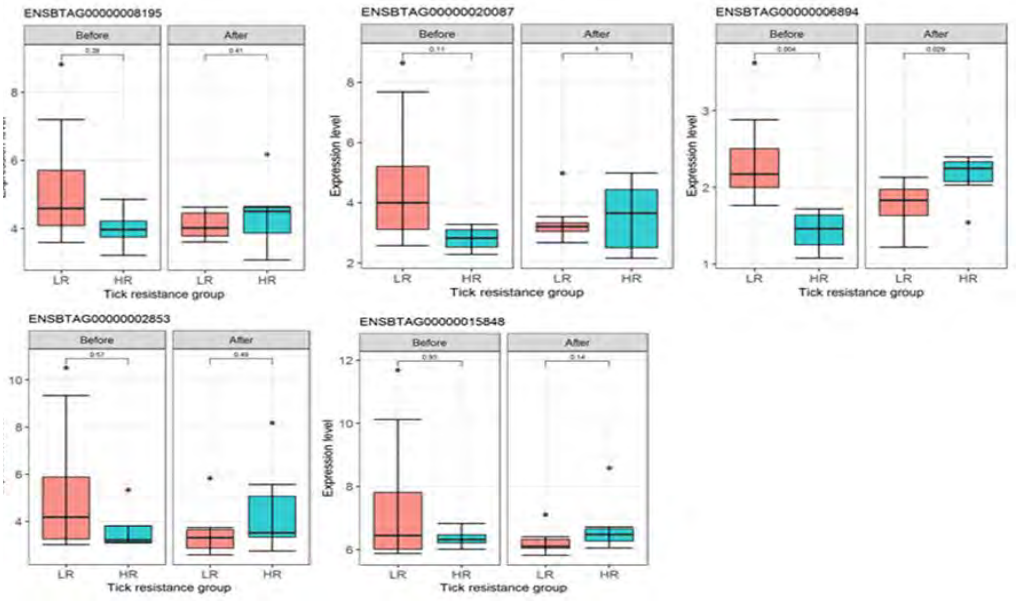


Figure 2. Boxplots showing before and after infestation gene expression comparisons of two groups (LR= Low Resistant, and HR= High Resistant) for the following genes: PHKG1 (ENSBTAG00000008195), PHKA1 (ENSBTAG00000015848), CAMK2A (ENSBTAG00000020087), NOS2 (ENSBTAG00000006894), and HRC (ENSBTAG00000002853). Expression level indicates variance stabilized read counts on a log2 scale. The numbers at the top indicate the significance levels. The *P*-values were obtained using the two sample Wilcoxon test. The bar represents the mean ± SD.

To overcome the challenge of sample size, the tick susceptible groups (LR and Angus) were combined into one group (LR group) and the tick resistant groups (HR and Nguni) were combined into HR group. Interestingly, the genes that were identified by the current study and their involvement in the Calcium (Ca^{2+}) pathway, were not previously reported in previous studies of a similar nature. For instance, (Moré *et al.* 2019) and (Otto *et al.* 2019) reported similar genes that are involved in biological processes that are relevant to various immune system functions.

Conclusions

The current study offered evidence that the Ca^{2+} signalling pathway maybe implicated in host resistance during the first 10 days of tick attachment. Most importantly, this research reports differential gene expression in the skin following tick bites, thus providing preliminary data concerning the underlying mechanisms associated with the immune response.

References

Jones C.R., Brunner J.L., Scoles G.A. and Owen J.P. (2015) Parasites & Vectors 8(1)**:** 340. https://doi.org/10.1186/s13071-015-0955-6

Jongejan F., and Uilenberg G. (2004) Parasitology 129(S1)**:** S3-S14. https://doi.org/10.1017/S0031182004005967

Mapholi N.O., Maiwashe A., Matika O., Riggio V., Bishop S. *et al.* (2016) Ticks and Tick-Borne Diseases 7(3)**:** 487-497.

Mapholi N.O., Marufu M.C., Maiwashe A., Banga C.B., Muchenje V. *et al.* (2014) Ticks and Tick-borne Diseases 5(5)**:** 475-483. https://doi.org/10.1016/j.ttbdis.2014.04.006

Moré D.D., Cardoso F.F., Mudadu M.A., Malagó-Jr W., Gulias-Gomes C.C. *et al.* (2019) BMC Genomics 20(1)**:** 998. https://doi.org/10.1186/s12864-019-6360-3

Otto P.I., Guimarães S.E.F., Verardo L.L., Azevedo A.L.S., Vandenplas J. *et al.* (2019) Journal of Dairy Science 102(9)**:** 8148-8158. https://doi.org/10.3168/jds.2018-15305

Piper E.K., Jackson L.A., Bagnall N.H., Kongsuwan K.K., Lew A.E. *et al.* (2008) Veterinary Immunology and Immunopathology 126(1)**:** 110-119. https://doi.org/10.1016/j.vetimm.2008.06.011

Porto Neto L.R., Jonsson N.N., D'occhio M.J. and Barendse W. (2011) Veterinary Parasitology 180(3)**:** 165-172. https://doi.org/10.1016/j.vetpar.2011.05.048

Walker A.R. (2003) *Ticks of domestic animals in Africa: a guide to identification of species*. Bioscience Reports Edinburgh.

98. Combined effect of genetics and gut microbiota on variations in vaccine response in hens

A. Lecoeur[1*], F. Blanc[1], D. Gourichon[2], N. Bruneau[1], T. Burlot[3], F. Calenge[1] and M.H. Pinard-van der Laan[1]

[1]Université Paris-Saclay, INRAE, AgroParisTech, UMR1313 GABI, 78350 Jouy-en-Josas, France; [2]INRAE, UE PEAT, 37380 Nouzilly, France; [3]NOVOGEN, 5 rue des compagnons, Secteur du Vau Ballier, 22960 Plédran, France; alexandre.lecoeur@inrae.fr

Abstract

Genetic variabilities cause variations in the efficiency of most vaccines. Variations in gut microbiota composition may be also implied. We studied 4 experimental or commercial lines of laying hens. A strong disruption of the gut microbiota was induced by a cocktail of antibiotics given at two different times. Vaccine response was monitored by hemagglutination inhibition assays and the composition of the cecal microbiota was assessed by the analysis of amplified 16S rRNA gene sequences. The vaccine response level differed according to the line after vaccination. The effect of antibiotics on the vaccine response varied according to the line, with different dynamics over time. Finally, antibiotics intake impacted the composition of the caecal microbiota at the end of the experiment for each line. This experiment confirms the effect of both gut microbiota and genetic lineage on the level of vaccine response.

Introduction

Vaccination is the most effective strategy for preventing infectious diseases. However, most vaccines vary in their efficacy. Host genetic variabilities and variations in gut microbiota composition are two of the likely drivers of this variation in efficacy (Ciabattini *et al.*, 2019; Zimmerman *et al.*, 2018). We have demonstrated the role of host genetics in vaccine response variation by successfully performing genetic selection of White Leghorn laying hens for a stronger vaccine response against Newcastle disease virus (NDV) (Pinard-van der Laan, 2002): a 'ND3 line' has been selected for high response for more than 25 years now, while keeping an unselected control 'TEM' line. Furthermore, the gut microbiota, in hens as in all farm animals, interacts closely with host immunity (Broom *et al.*, 2018), so it is hypothesized to have an effect on the vaccine response as well. This role is less documented, but a few studies demonstrate its influence (Valdez *et al.*, 2014; Borey *et al.*, 2021).

The objective of the present study was to investigate the effect of both genetic background and caecal microbiota on vaccine response. We used 4 laying hen lines: the 2 experimental lines ND3 and TEM and 2 commercial lines from the Novogen breeding company, one Rhode Island line (RIR) and one White Leghorn line (LEG). To study the impact of the microbiota, we modified its composition via the ingestion of an antibiotic cocktail.

Materials & methods

Animals. A total of 204 animals were used (ND3=48, TEM=48, RIR=48 and LEG=60). All animals hatched together at the PEAT unit (Nouzilly) and were then distributed in 3 groups for each line: 1. Group WA: 24 individuals vaccinated and not treated with antibiotics, 2. Group A1: 12 animals vaccinated and treated with antibiotics from D1 to D3, 3. Group A19: 12 animals vaccinated and treated with antibiotics the three days before NDV vaccination. In addition, a control group of 12 animals of the TEM line was neither vaccinated nor treated (Group CTR). The animals were raised in cages of 12 hens and killed at D42 by electronarcosis and cervical dislocation. The animals were weighed before their vaccination (D22) and before their sacrifice (D42). Blood samples were taken at D36, D40 and D42 to measure the vaccination

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_98, © A. Lecoeur *et al.* 2022

response. Caecal contents and spleen were collected post-mortem from each animal. Animals were raised and sacrificed according to French regulations (APAFIS#29749-2021021018217492 authorization).

Caecal microbiota. The gut microbiota was disrupted by a cocktail of 3 antibiotics (Amoxicillin, Streptomycin and Metronidazole) of different families with different methods of action. Caecal contents collected at D42 were frozen immediately in liquid nitrogen. DNAs from these samples were extracted by the @bridge platform (INRAE Jouy-en-Josas). Sequencing was performed by an Illumina Miseq 2×250 bp sequencer, using amplicons of the V3-V4 region of the 16S rRNA gene obtained with the primers Forward: ACGGRAGGCAGCAG and Reverse: TACCAGGGTATCTAATCCT. Raw sequences were analysed using the R package Dada2. The ASV table obtained was analysed with the phyloseq and Deseq2 packages to assess the global diversity of the caecal microbiotas collected and individual taxonomic compositions.

Vaccination and immunity. Animals were vaccinated at 5 weeks (D22) against NDV by ocular injection and vaccine response was measured by hemagglutination inhibition test.

Results

Weight gain. Antibiotics treatment caused neither excess mortality nor health problems. All lines differed significantly from each other for their weight (data not shown). The RIR line showed a faster growth than the three other lines (times 2.6 vs 2.3). The effect of antibiotics ingestion differed according to the line: For the RIR line, the A1 group tended to grow more than A19 (P_{adj}<0.1). For the TEM line, the A19 group had a faster growth than A1, WA and CTR.

Vaccine response. ND3 showed the highest response (Figure 1). The TEM line showed an intermediate level of response, while the two commercial lines had lower levels of response than the two experimental lines, and similar to each other. This ranking of lines was similar for the three time points considered (D36, D40 and D42). We observed an effect of antibiotics intake on the vaccine response, varying according to the line: For the RIR line, no significant difference between the 3 groups was observed, probably due to a high variability of the A1 group. For the LEG line, group A19 showed a weaker vaccine response than the other two groups. This difference between groups, which was slight at D36, increased over time with a peak in vaccine response 3 weeks after injection. The response levels of the ND3 and TEM lines were stable between groups. The variability of the response was very low and decreased with time.

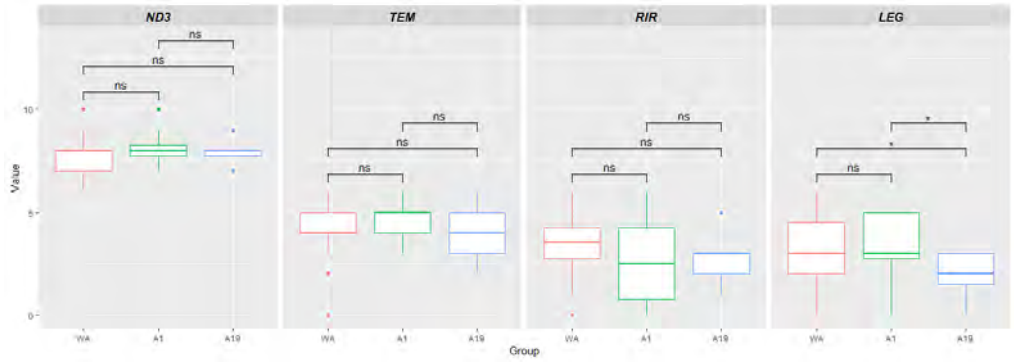


Figure 1. vaccine response levels of the four genetic lines studied (RIR, LEG, ND3 and TEMP) in three different animal groups: WA (without antibiotics), A1 (antibiotics received from day 1), A19 (antibiotics received from day 19, just before vaccination).

Caecal microbiota. We obtained 12.37 million raw sequences, with an average of 49,700 sequences per sample. At the output of Dada2, the taxonomy table of the 204 individuals contained 5,854,649 sequences, with an average of 28,283 sequences per sample, and 9,609 ASV (Amplicon Sequence Variant - significance) in total. Six individuals with too few sequences were eliminated from the analyses (<8,500 sequences). The dataset was normalized to 15,928 sequences per individual. The caecal microbiota of the 4 lines was largely dominated by the Phylum *Firmicutes* (between 95 and 99% of the sequences), followed by *Bacteroidota* (0.1-3.3%), *Actinobacteriota* (0.19-0.38%), *Proteobacteria* (0.1-0.43%). We also observed some sequences of the phylum *Desulfobacterota*. The most represented family was *Ruminococcaceae* (46-66%), followed by *Lachnospiraceae* (8-14%), *Oscillospiraceae* (4.5-7%), *Rikenellaceae* (0.1-3.3%). The analysis of different alpha diversity criteria (Richness, Chao1, Shannon and Simpson) describing the microbiota of the four lineages revealed the existence of differences in microbiota structure between lines. The Shannon and Simpson indices showed significant differences between lines (data not shown). We also compared lines and groups for 4 indices describing the beta-diversity of microbiotas: Bray-Curtis, Jaccard, Unifrac and Weighted Unifrac. They were higher in the RIR line compared to the ND3 and LEG lines, and lower in the ND3 line compared to the RIR and TEM lines. We observed for the 4 indices a line effect, a group effect and a significant line × group interaction.

Discussion

We compared the vaccine response levels of four experimental or commercial laying hen lines treated or not with a cocktail of antibiotics. The animals studied were raised at the same time under identical conditions in the same room after hatching in the same hatchery. Therefore, phenotypic differences between animal groups cannot be attributed to differences in the breeding environment.

By comparing groups of animals treated vs non–treated with antibiotics, we first observed that the antibiotics ingestion led to a microbiota disruption in the hen lines (Figure 2). This effect was different according to the time at which the antibiotics were given (Figure 2). When comparing the different lines for their caecal microbiota composition, we observed global differences in two alpha diversity indices and

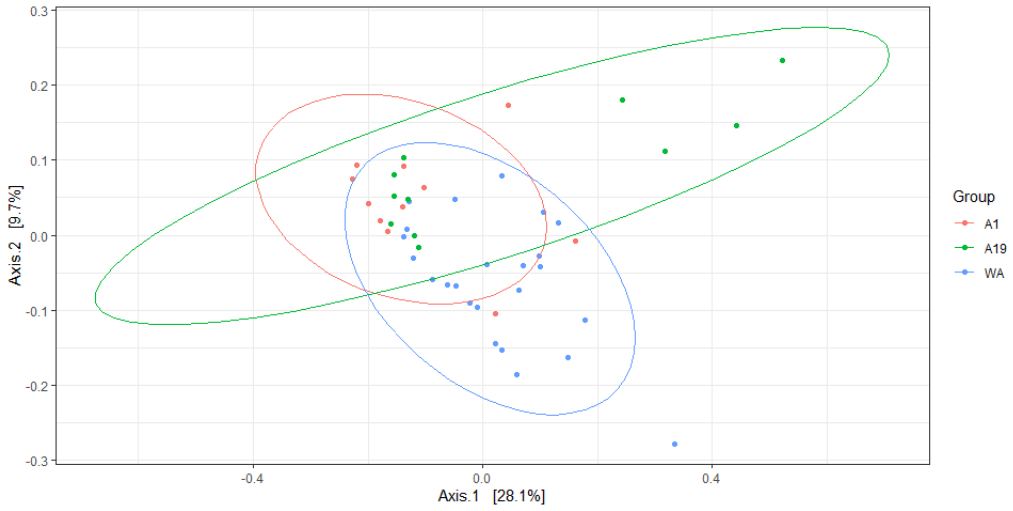


Figure 2. NMDS representation of the bray-curtis distances of the individual caecal microbiota composition in the three treatment groups of the RIR line: WA (without antibiotics), A1 (antibiotics from day 1), A19 (antibiotics from day 19, before vaccination).

in beta-diversity indices, which means that microbiota compositions vary according to the line. The role of host genetics in shaping the caecal microbiota in the chicken has already been demonstrated (i.e. Borey *et al.*, 2020). However, microbiotas of the 4 lines are altered in a similar way by antibiotics ingestion (data not shown). We thus showed that our model of microbiota perturbation is effective for all the lines studied and can be used for further studies.

In parallel, we observed that the 4 genetic lines displayed contrasted vaccine response levels and dynamics over time, with the ND3 line logically displaying the highest vaccine response, and ND3 and TEM displaying the more stable vaccine response (Figure 1). The latter lines have been selected for the stability of their vaccine response, which could explain that a disruption of this characteristic is more difficult. After antibiotics ingestion, the vaccine response was affected differently according to the line (Figure 1). The vaccine response of the LEG line was the most affected by the antibiotics ingestion.

In conclusion, this study allowed us to validate our model of microbiota disruption by antibiotic treatment. The latter induced a relatively long-lasting disturbance of the microbiota without generating any health problem, at least during the 6 weeks of our experiment. This first application of our model allowed us to confirm that host genetic variability partly controls the level and dynamics of the vaccine response. Antibiotics ingestion impacted the vaccine response and the caecal microbiota composition. However, the role of the caecal microbiota in the vaccine response perturbation induced by the antibiotics ingestion, although likely, remains uncertain at this stage and needs further exploration.

References

Borey, M., Estellé, J., Caidi, A., Bruneau, N., Coville, J.-L, *et al.* (2020). PloS One 15(5), e0232418. https://doi.org/10.1371/journal.pone.0232418

Broom, L. J., & Kogut, M. H. (2018). Vet Immunol Immunopathol 204: 44–51. https://doi.org/10.1016/j.vetimm.2018.10.002

Ciabattini, A., Olivieri, R., Lazzeri, E., & Medaglini, D. (2019). Frontiers in Microbiology, 10, 1305. https://doi.org/10.3389/fmicb.2019.01305

Pinard-van der Laan, M.-H. (2002). Vet Immunol Immunopathol, 87(3–4), 199–205. https://doi.org/10.1016/s0165-2427(02)00075-2

Valdez, Y., Brown, E. M., & Finlay, B. B. (2014). Trends Immunol. 25113637. https://pubmed.ncbi.nlm.nih.gov/25113637

Zimmermann, P., & Curtis, N. (2018). Vaccine 36(30): 4433–4439. https://doi.org/10.1016/j.vaccine.2018.04.066

99. Resilience indicators based on daily milk yield data for genetic selection in dairy cattle

M. Poppe[1,2*], H.A. Mulder[1], H. Hogeveen[1], C. Kamphuis[1], G. Bonekamp[1], M.L. van Pelt[2], E. Mullaart[2] and R.F. Veerkamp[1]

[1]Wageningen University & Research, P.O. Box 338, 6700 AH Wageningen, the Netherlands; [2]CRV, Wassenaarweg 20, 6843 NW Arnhem, the Netherlands; marieke.poppe@crv4all.com

Abstract

Dairy cows face many kinds of environmental challenges throughout their life, such as pathogens, sudden changes in feed quality, or heat waves. Resilient cows are cows that can cope well with such challenges – they are minimally affected and recover quickly. Having resilient cows is beneficial for both the cows themselves and farmers. We investigated if newly developed traits, based on daily milk yield can be used to genetically select for improved resilience. Cows change their milk yield in response to challenges. The hypothesis was, therefore, that daily patterns in milk yield can indicate resilience. Breeding for low variability in milk yield and for low autocorrelation, i.e. dependency between day-to-day deviations in milk yield, can help to improve resilience. Genetic improvement of resilience will likely improve welfare and profitability of cows, and improve job satisfaction of farmers.

Introduction

Resilience is the ability of animals to be minimally affected by, and to quickly recover from environmental disturbances, such as pathogens, extreme weather, or changes in feed quality (Colditz and Hine, 2016; Berghof *et al.*, 2019). Resilience is important for the welfare of animals and work pleasure of farmers, especially given the expected increase of disturbances in the future (Boichard and Brochard, 2012; Elgersma *et al.*, 2018).

There are several ways to improve resilience of animals, and one of them is through genetic selection (Colditz and Hine, 2016; Berghof *et al.*, 2019). However, as resilience is a broad concept, it is difficult to quantify and thus difficult to identify which animals are genetically most resilient. Therefore, indicator traits are needed to improve resilience.

Nowadays, many sensor and other automated data are available on animals (Rutten *et al.*, 2013), and these data may contain information about resilience. Many traits recorded by sensors and automated systems are sensitive to disturbances. Hence, the fluctuation patterns in such longitudinal traits are expected to contain information about resilience (Scheffer *et al.*, 2018; Berghof *et al.*, 2019). Two traits based on longitudinal data have been proposed as indicators of resilience: the variance and lag-1 autocorrelation (Scheffer *et al.*, 2018). Variance indicates the variability of a trait, and thus the sensitivity of the animal to disturbances. Autocorrelation indicates the dependency of subsequent records on each other, and thus at what rate the animal recovers from disturbances.

In this paper we summarize results on the resilience indicators variance and autocorrelation of daily milk yield. We focus on the heritability and genetic correlations with other resilience-related traits, the associations with economic performance and we discuss its usefulness for selection.

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_99, © M. Poppe *et al.* 2022

Materials & methods

Data. The original data set contained 1,782,373,113 milk yield records on 1,120,550 cows, obtained during single milk visits to automatic milking systems (AMS) and conventional milking systems. Only records from AMS were selected and single milk visit records were converted to daily records. Only cows were selected that were registered in the herd-book and that were more than 87.5% Holstein Friesian. The milk yield data were corrected for lactation curve shape by fitting a polynomial quantile regression curve for each cow. The residuals from the fitted curves contained information about short-term fluctuations that are informative about resilience. These residuals were used to calculate the resilience indicators for each cow in lactation 1: variance (LnVar; 198,725 cows) and autocorrelation (r_{auto}, 198,580 cows).

Analyses. We estimated genetic parameters for the resilience indicators using mixed animal models in ASReml (Gilmour *et al.*, 2015). We estimated genetic correlations with traits that are currently in routine genetic evaluation using the multiple across country evaluation (MACE) procedure (Schaeffer, 1994; Klei and Weigel, 1998; Interbull, 2017). We estimated genetic correlations with several traits describing milk yield response to real-life disturbances at herd level using sire-maternal grandsire models in ASReml (Gilmour *et al.*, 2015). Real-life disturbances at herd level were selected based on days where mean milk yield of all cows in a herd suddenly declined, probably due to a disturbance. The traits describing response to the disturbances at herd level were the total yield loss during the temporary decline in milk yield (kg) and the time that the temporary decline lasted (days).

In addition to genetic analyses, we analysed the phenotypic association between the resilience indicators and lifetime gross margin. We performed an analysis of covariance based on records of 1,337 cows from 21 herds, with mean milk yield in lactation 1 as a covariate. Lifetime gross margin of each cow was calculated as the difference between all lifetime revenues (milk, calves, slaughter) and lifetime costs (feed, calving, rearing, insemination, treatment, destruction).

Results

All results are presented in Table 1. Both LnVar and r_{auto} were heritable, with LnVar having a higher heritability (0.21) than r_{auto} (0.09). Both had a positive genetic correlation with mean milk yield and negative genetic correlations with udder health, ketosis, and longevity, with the strongest correlation always for LnVar. These correlations indicate that cows with low LnVar or r_{auto} (good resilience) had low milk yield, good udder health and few ketosis cases, and high longevity, genetically. LnVar had a strong genetic correlation with total yield loss upon real-life herd disturbances (0.90), which means that cows with low LnVar had only a small decline in milk yield upon disturbances. The r_{auto} had no genetic correlation with yield loss upon herd disturbances, but was strongly genetically correlated with the time until milk yield recovered again. This means that cows with low r_{auto} had quick recovery upon real-life disturbances at herd level. The r_{auto} did not have a significant association with lifetime gross margin. However, LnVar did have a significant association with lifetime gross margin, which means that cows with low LnVar in first lactation had generated more profit at the end of life than cows with the same milk yield but with high LnVar.

Discussion

Low LnVar and r_{auto} were expected to be indicators of good resilience based on the assumption that resilient cows have milk yield that does not fluctuate much around its expected value, and if it deviates, recovers quickly (Scheffer *et al.*, 2018; Berghof *et al.*, 2019). Both resilience indicators were heritable, which means that they can be changed using genetic selection. The favourable genetic correlations with health traits, longevity (Poppe *et al.*, 2020; 2021a), and response in milk yield to real-life disturbances at herd level (Poppe *et al.*, 2021b) confirm that selection for these traits will result in improved resilience.

Table 1. Genetic variance (VarA) and heritability (h^2) of resilience indicators, genetic correlations (r_g) with other traits (SE between brackets), and regression coefficients explaining relation with lifetime gross margin.

		Resilience indicator[1]	
		LnVar	**r_{auto}**
Genetic parameters	VarA	0.056 (0.002)	0.003 (0.000)
	h^2	0.21 (0.01)	0.09 (0.01)
r_g with other traits[2]	AMY (kg)	0.79 (0.02)	0.16 (0.04)
	UH[3]	-0.32	-0.21
	KET[3]	-0.48	-0.15
	LON[3]	-0.16	-0.02
	Yield loss (kg)	0.90 (0.05)	-0.01 (0.13)
	Yield loss length (d)	-0.00 (0.29)	0.97 (0.35)
Association with lifetime gross margin	Regression coefficient	-427.98	404.77
	P-value	0.01	0.41

[1] LnVar = natural log-transformed variance of deviations from expected milk yield; r_{auto} = lag-1 autocorrelation of deviations from expected milk yield.
[2] AMY = average daily milk yield; UH = udder health; KET = ketosis; LON = longevity; yield loss = total yield loss upon real-life disturbances at herd level; yield loss length = time until yield is recovered upon real-life disturbances at herd level.
[3] Genetic correlations are estimated with Multiple Across Country Evaluation procedure, which does not give standard errors.

The strong genetic correlations with traits quantifying response to real-life disturbances at herd level confirm that the resilience indicators contain information about the severity of response to disturbances and recovery rate. The current breeding goal does not include response and recovery to disturbances yet, but only incidence of disease for example (CRV, 2020). Therefore, the resilience indicators contain new and relevant information, and resilience could be considered as an additional trait in the breeding goal.

The favourable association between LnVar and lifetime gross margin also suggests that LnVar has economic importance. However, some of the traits used to calculate lifetime gross margin are also included in the current breeding goal, such as milk yield, longevity, and fertility. To avoid double counting in the breeding goal, the marginal economic value, or effect of LnVar partial on these traits should be established first. Moreover, resilience is expected to have an additional non-economic value through its association with animal welfare. Hence, alternative methods are needed to determine the weight in the breeding goal, such as desired gains (Nielsen *et al.*, 2005), choice experiments with farmers (Nielsen *et al.*, 2011), or more research into associations with labour costs (Berghof *et al.*, 2019). Once the weight of resilience in the breeding goal is determined, an index of LnVar and r_{auto} can be used to genetically improve resilience.

Funding

This study was financially supported by the Dutch Ministry of Economic Affairs (TKI Agri and Food project 16022) and the Breed4Food partners Cobb Europe, CRV, Hendrix Genetics and Topigs Norsvin. Additional financial support was provided by the European Union's Horizon 2020 research and innovation program (GenTORE) under grant agreement No. 727213.

References

Berghof, T.V.L., Poppe, M. and Mulder, H.A. (2019). Front. Genet. 9:692. https://doi.org/10.3389/fgene.2018.00692
Boichard, D., and Brochard, M. (2012). Animal 6(4):544–550. https://doi.org/10.1017/S1751731112000018.
Colditz, I.G., and Hine, B.C. (2016). Anim. Prod. Sci. 56(12):1961–1983. https://doi.org/10.1071/AN15297

CRV. (2020). E-20 NVI. Available at: https://cooperatiecrv-be6.kxcdn.com/wp-content/uploads/2020/11/E_20-NVI_aug2020_en.pdf

Elgersma, G.G., de Jong, G., van der Linde, R., and Mulder, H.A. (2018). J. Dairy Sci. 101(2):1240–1250. https://doi.org/10.3168/jds.2017-13270

Gilmour, A.R., Gogel, B.J., Cullis, B.R., Welham, S.J., and Thompson, R. (2015). ASReml User Guide Release 4.1 Functional Specification. VSN International Ltd., Hemel Hempstead, UK

Interbull. (2017). Interbull Code of Practice - Methods of International Evaluation. Available at: http://www.interbull.org/ib/cop_chap5

Klei, L., and Weigel, K.A.. (1998). Interbull Bull. 17:8–14. https://journal.interbull.org/index.php/ib/article/.

Nielsen, H.M., Christensen, L.G., and Groen, A.F. (2005). J. Dairy Sci. 88(5):1882–1890. https://doi.org/10.3168/jds.S0022-0302(05)72864-2

Nielsen, H.M., Olesen, I., Navrud, S., Kolstad, K., and Amer, P. (2011). J. Agric. Environ. Ethics 24:309–330. https://doi.org/10.1007/s10806-010-9264-4

Poppe, M., Veerkamp, R.F., van Pelt, M.L., and Mulder, H.A. (2020). J.Dairy Sci. 103(2):1667–1684. https://doi.org/10.3168/jds.2019-17290.

Poppe, M., Bonekamp, G., van Pelt, M.L., and Mulder, H.A. (2021a). J. Dairy Sci. 104(2):1967-1981. https://doi.org/10.3168/jds.2020-19245.

Poppe, M., Mulder, H.A., and Veerkamp, R.F. Veerkamp. (2021b). J. Dairy Sci. 104(7): 8094-8106. https://doi.org/10.3168/jds.2020-19817.

Rutten, C.J., Velthuis, A.G.J., Steeneveld, W., and Hogeveen, H.. (2013). J. Dairy Sci. 96(4):1928–1952. https://doi.org/10.3168/jds.2012-6107

Schaeffer, L.R. (1994). J. Dairy Sci. 77(9):2671–2678. https://doi.org/10.3168/jds.S0022-0302(94)77209-X

Scheffer, M., Bolhuis, J.E., Borsboom, D., Buchman, T.G., Gijzel, S.M.W. *et al.*. (2018). Proc. Natl. Acad. Sci. U. S. A. 115(47):11883–11890. https://doi.org/10.1073/pnas.1810630115

100. Genetic evaluation of heat tolerance in Italian Holstein breed

R. Finocchiaro[1*], F. Galluzzo[1], J.B.C.H.M van Kaam[1], M. Marusi[1] and M. Cassandro[1,2]

[1]Associazione Nazionale Allevatori della Razza Frisona, Bruna e Jersey Italiana (ANAFIBJ), Via Bergamo 192, 26100 Cremona (CR), Italy; [2]Department of Agronomy, Food, Natural resources, Animals and Environment (DAFNAE), University of Padova, Viale dell'Università 16, 35020 Legnaro (PD), Italy; raffaellafinocchiaro@anafi.it

Abstract

Heat stress (HS) on dairy cows is triggered by a combination of high ambient temperature and high relative humidity: this phenomenon leads to physiological disorders that have a negative impact on the animal's wellbeing and productive and reproductive performances. The objective of this study was to identify the starting point of HS on milk production, measured on the temperature-humidity index (THI) scale, to define the heritability of the tolerance to HS based on milk production and to estimate breeding values for this trait in the Italian Holstein population. A single-trait repeatability linear animal model was used and the heritability was found to be 0.16. In conclusion, the possibility to select for heat tolerance (HT) based on milk production was confirmed. Further analyses will be performed to assess the possibility of selection for HT based on other productive and functional traits and to define its optimum weight in a selection index.

Introduction

High ambient temperature combined with high relative humidity negatively affects both productive (Bernabucci *et al.*, 2014) and reproductive (Ravagnolo and Misztal, 2002) performance of dairy cows, resulting in a negative impact on herd profitability: this phenomenon is regarded as HS. Not all animals are affected the same way by HS. In Holstein cattle, tolerance to HS was found to be moderately heritable (Aguilar *et al.*, 2009; Bernabucci *et al.*, 2014), meaning that a selection for this trait is feasible. As the majority of the Italian Holstein population lives in areas characterized by hot and humid summer, a genetic evaluation for HT may be beneficial. The aim of our study was to investigate if heat stress results in different effects on Italian Holstein performance due to genetic differences in heat tolerance.

Materials & methods

Data and data editing. Data came from the Italian National Breeders Association (AIA) and consisted, after editing, of 41,063,377 test-day milk production records of 2,591,002 Holstein cows. Heat tolerance was modelled by using data from weather stations. Daily maximum temperature and relative humidity were collected from 1994 till now. Weather measurements (777,980 daily records) were collected from 137 stations distributed across the national territory. Geographical coordinates and altitudes from all weather stations and from the communities of the farms were available. Distances between weather stations and farms were computed. Milk test day records were linked with the nearest weather stations within 80 km. Weather stations more than 500 m above or below the farm were omitted. Weather estimates for the farm were computed for the geometric center of the chosen weather stations as a weighted average accounting for the distance from the farm. The average distance between the geometric center of the chosen nearby weather stations and the farms was 13.5 km.

Holstein test-day records for milk production from the first three parities were included in this study. Records were from cows registered between 2004 and 2021. Cows were required to have, within parity, at least 3 test-days and the first one within 60 days from calving and test-days between 5 and 305 DIM. Cows were also required to have the first parity recorded. First parity cows were required to have an age at parity of minimum 22 months, second parities greater or equal than 34 months and later parity ones between

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_100, © R. Finocchiaro *et al.* 2022

48 and 76 months. Moreover, in order to effectively compare the contemporaries for herd-year-season of recording (HYSR), groups with less than 30 observations were excluded from the analysis. The pedigree went back to four generations.

Because longer periods of heat stress might have a more severe effect than shorter periods, the average weather measurements of 2, 4, 5, 7, 10 and 14 days before test-day recording were considered.

The temperature-humidity index (THI) has been calculated as proposed by Kelly and Bond (1971) by combining maximum temperature, in degrees Celsius, and average relative humidity (as percentage) with the following expression:

$$THI = T_{max} - [0.55\ x\ (1 - RH)]\ x\ (T_{max} - 14.4) \quad (1)$$

where T_{max} is the maximum daily temperature in °C and RH is the average daily relative humidity.

The THI threshold value (THI_{thr}), above which milk yield declines, was identified visually after the computation of least square means of the fixed effect THI class. A function of THI, f(THI), was then created as in Equation 2:

$$f(THI) = \begin{cases} 0 & if\ THI \leq THI_{thr} \\ THI - THI_{thr} & if\ THI > THI_{thr} \end{cases} \quad (2)$$

The visual inspection showed that the threshold value of THI from which milk production starts to decline is ≥24 (Celsius like scale), and this value was defined as THI_{thr}. In addition, the average of 7 days before test-day recording has been found to be the most effective period for modelling the decline in milk production.

Statistical model. The selected model was a single trait repeatability linear animal model. Fixed effects were identified with SAS PROC GLM (SAS Institute Inc., Cary, NC, 2017): only the ones which resulted statistically significant (P<0.05) were retained. The effect of HS on production is modeled as a random regression.

The defined model is described in Equation 3:

$$Y_{ijklmno} = HYSR_i + YS_j + STPY_k + a_{general_l} + a_{heat_l}[f(THI_m)] + PE_{general_n} + PE_{heat_n}[f(THI_m)] + e_{ijklmno} \quad (3)$$

with Y_{ijklmn} is the measurement of daily milk production (kg/d), $HYSR_i$ is the fixed effect herd-year-season of recording i; YS_j is the fixed effect of year-season of calving j, $STPY_k$ is the fixed effect of stage of lactation-parity-year of calving class k. Regarding random effects, $a_{general_l}$ is the general additive genetic effect of cow l, a_{heat_l} is the additive genetic effect of HT for cow l, $PE_{general_n}$ as the general permanent environmental effect of cow l based on repeated measures both within and across lactations, PE_{heat_n} as the permanent environmental effect of HT for cow l and $e_{ijklmno}$ as the residuals. Parities were defined as first, second and later parities while seasons as December to February, March to May, June to August and September to November.

Estimation of genetic parameters. Variance components estimation was performed with the Gibbs sampler THRGIBBS1F90 (Misztal *et al.*, 2002). Three analyses were performed on three samples of 200 herds each and a minimum of 30,000 cows each (average number of test-day records per sample: 817,481) randomly extracted from the full dataset, with 500,000 iterations, a burn-in of 50,000 and a thinning rate of 10. Post-Gibbs analysis was performed using the software POSTGIBBSF90 (Misztal *et al.*, 2002) using

the remaining 45,000 samples. Convergence was assessed by visual inspection. Heat tolerance heritability based on milk production was calculated as described in Equation 4:

$$h^2 = \frac{\sigma^2_{a_{general}} + \sigma^2_{a_{heat}} + 2\sigma_{a_{general}}\sigma_{a_{heat}}}{\sigma^2_{a_{general}} + \sigma^2_{a_{heat}} + 2\sigma_{a_{general}}\sigma_{a_{heat}} + \sigma^2_{pe_{general}} + \sigma^2_{pe_{heat}} + 2\sigma_{pe_{general}}\sigma_{pe_{heat}} + \sigma^2_e} \quad (4)$$

with $\sigma^2_{a_{general}}$ as the general additive genetic effect variance, $\sigma^2_{a_{heat}}$ as the additive genetic effect variance for HT, $\sigma_{a_{general}}\sigma_{a_{heat}}$ as the covariance between the two, $\sigma^2_{pe_{general}}$ as the general permanent environmental effect variance, $\sigma^2_{pe_{heat}}$ as the permanent environmental effect variance for HT, $\sigma_{pe_{general}}\sigma_{pe_{heat}}$ as the covariance between the two and σ^2_e as the residual variance.

Breeding value estimation. Heat tolerance breeding values (BVs) for daily milk production were estimated with MiX99 (Lidauer *et al.*, 2019) and expressed on a scale with mean of 100 and standard deviation of 5: values higher than the mean refer to individuals that transmit a higher tolerance to HS.

Results

Estimation of genetic parameters. Posterior means and standard deviations for variance components and heritability were computed for the three samples. The mean heritability of the samples' posterior means was 0.16 (standard deviation=0.006), confirming the feasibility of a selection for HT. As expected, the genetic correlation between the general additive genetic variance and the HT genetic variance was negative, -0.45 (0.048), meaning that high producing cows tend to be more susceptible to HS.

Genetic trend. The genetic trend for HT by year of birth of bulls is represented in Figure 1. The figure shows a strong decreasing trend in the early '90s, a plateau and finally an important increase in the last 20 years. This trend reflects the introduction of new breeding objectives in the Italian selection index, which progressively shifted some weight from production to functionality, therefore the positive trend can be explained by the negative genetic correlation found between milk production and HT.

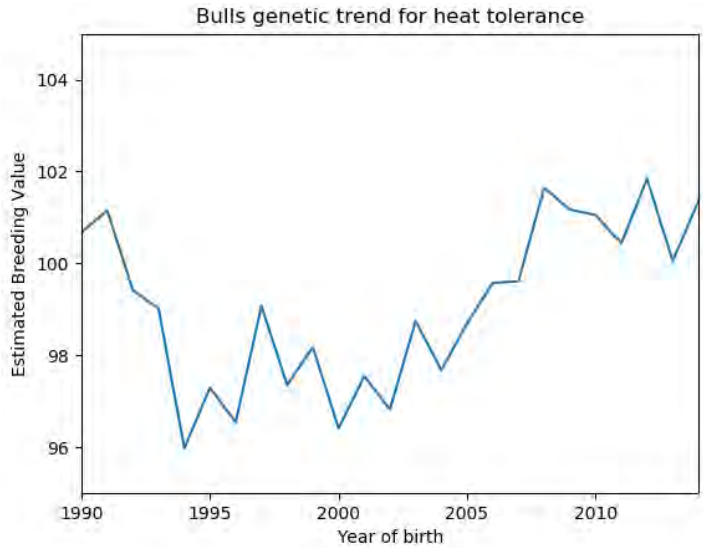


Figure 1. Genetic trend of heat tolerance by year of birth.

Conclusions

The feasibility of a selection for HT based on milk production was confirmed by the moderate heritability found for this trait, 0.16 (0.006). The negative genetic correlation between HT and milk production was also confirmed (-0.45), meaning that HT should not be used as a selection criterion by itself: economic analyses will be performed to define the optimum weight for this trait in a selection index. The use of the Italian selection index (PFT) has already selected, indirectly, for more heat tolerant bulls, it is clear that the availability of a direct index would allow a better selection for more heat tolerant individuals.

References

Aguilar, I., I. Misztal, and S. Tsuruta. 2009. Genetic components of heat stress for dairy cattle with multiple lactations. J. Dairy Sci. 92:5702–5711. https://doi.org/10.3168/jds.2008-1928.

Bernabucci, U., S. Biffani, L. Buggiotti, A. Vitali, N. Lacetera, and A. Nardone. 2014. The effects of heat stress in Italian Holstein dairy cattle. J. Dairy Sci. 97:471–486. https://doi.org/10.3168/jds.2013-6611.

Kelly, C.F., and T.E. Bond. 1971. Bioclimatic factors and their measurement. National Academy of Sciences, Washington, DC.

Lidauer, M., K. Matilainen, I. Strandèn, E. Mäntysaari, T. Pitkänen, and M. Taskinen. 2019. MiX99 : Technical reference guide for MiX99 pre-processor.

Misztal, I., S. Tsuruta, T. Strabel, B. Auvray, T. Druet, and D.H. Lee. 2002. BLUPF90 and related programs (BGF90). Commun. No. 28-07. Page in 7th World Congress on Genetics Applied to Livestick Production, Montpellier, France.

Ravagnolo, O., and I. Misztal. 2002. Effect of heat stress on nonreturn rate in holsteins: Fixed-model analyses. J. Dairy Sci. 85:3101–3106. https://doi.org/10.3168/jds.S0022-0302(02)74397-X.

SAS Institute Inc., Cary, NC, U. 2017. Sas® 9.4. SAS Inst. Inc., Cary, NC, USA 1–8.

101. Inferring genetic parameters for prenatal heat stress effects on calf diseases and cow productivity

T. Yin*, K. Halli, S.F. Vanvanhossou, M. Bohlouli and S. König

Institute of Animal Breeding and Genetics, Justus-Liebig-University Gießen, Ludwigstr. 21B, 35390 Gießen, Germany; tong.yin@agrar.uni-giessen.de

Abstract

The aim of this study was to infer genetic and genomic effects of heat stress (HS) during late gestation of dams on disease occurrences and productivity of ~20,000 genotyped calves and cows, respectively, from an across-generation perspective. The climatic descriptor was the temperature-humidity index (THI) during the last week of pregnancy. For diseases, climatic sensitivity was studied via random regression coefficients nested within maternal genetic effects on prenatal THI alterations. For production traits, a genomic interaction model with THI≥60 indicating prenatal HS, was applied. Maternal heritablities for pneumonia increased from 0.03 to 0.10 with increasing THI from 60 to 74, but was largest for diarrhea at THI 60 (0.07). Genotype by HS interaction variances for production traits explained maximal 3.79% of the total variances. The novel modelling strategies enabled the assessment of maternal climatic sensitivity and the estimation of variances caused by HS.

Introduction

As a heat sensitive breed, Holstein cattle showed variations in phenotypic performances and breeding values under heat stress (HS) conditions. For example, protein yield significantly decreased at THI 68 (Gernand *et al.*, 2019). Genetically, random regression models were applied to infer genetic (co)variance components along a continuous climatic descriptor, i.e. the temperature-humidity index (THI; Bohlouli *et al.*, 2019). Such modelling approach was also used to detect possible genotype × environment interactions (G×E; Bohlouli *et al.*, 2019). Most of the genetic studies focused on immediate responses to HS, i.e. considering THI at the measuring day, or the average THI shortly before the measuring day. Phenotypically, HS during late gestation in dams, termed as prenatal HS, significantly impaired growth, immune competence and functional traits in offspring (Tao and Dahl, 2013). Monteiro *et al.* (2016) studied long-term effects prenatal HS on female fertility and performance traits recorded in offspring and identified most detrimental impact of HS from the last 6 weeks of gestation. Genetically, Halli *et al.* (2021) and Kipp *et al.* (2021) identified alterations of genetic variances for primary traits and G×E in young dual-purpose cattle and in first lactating dairy cows in response to prenatal THI.

In this regard, the aim of this study was to apply alternative modelling strategies and additionally considering genomic data to infer the effects of prenatal HS on genetic parameters for calf diseases and for test-day and lactation production traits in German Holstein cattle. Solutions for genetic effects were used in ongoing genome wide association studies (GWAS). Suggestively associated SNPs from GWAS were considered in gene annotations.

Materials & methods

Calf diseases and cow production traits were recorded in 53 large-scale co-operator herds located in two federal states located from the north-eastern part of Germany.

Dataset for calf diseases. The calf diseases pneumonia (PNEU) and diarrhea (DIAR) were diagnosed by veterinarians or herd producers based on the official hierarchical health diagnosis key from the International Committee for Animal Recording (Stock *et al.*, 2013). For PNEU, we focused on the period

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_101, © T. Yin *et al.* 2022

from 0 to 180 d after birth for female calves. DIAR mostly occurs in a short period after birth (0 to 30 days of age). Hence, both sexes were considered for DIAR. For each of the calf diseases, at least one entry within the respective period implied a score = 1 (diseased calve); otherwise, a score = 0 (healthy calve) was assigned. The pedigree file included 338,158 animals. Among them, 19,247 Holstein cattle were genotyped with the *Illumina BovineSNP50 v2 BeadChip* or the *Illumina Bovine Eurogenomics 10K low-density chip*. The number of genotyped calves with phenotypes was 12,728 for PNEU and 10,170 for DIAR. After quality control, 41,135 SNPs were considered for ongoing analyses.

Dataset for cow production traits. Test-day records from the first official test-day after calving (5 to 35 days in milk) included milk yield (TMY), fat percentage (TFP), fat yield (TFY), protein percentage (TPP), protein yield (TPY) and somatic cell score (TSCS). First lactation records were considered for milk yield (LMY), fat percentage (LFP), fat yield (LFY), protein percentage (LPP) and protein yield (LPY). The lactation lengths for lactation records varied from 275 to 305 days in milk. In total, 41,139 SNPs from 14,188 genotyped cows were used for the ongoing genomic analyses. All genotyped cows had milk production records.

Meteorological data. Pairwise distances (in km) were calculated between weather stations and cowherds. According to the minimal distances, we allocated 33 weather stations to the 53 herds. Hourly THI were calculated considering hourly temperature (T) and relative humidity (RH) as follows (NRC, 1971):

$$THI = [1.8 \times T + 32] - [0.55 - 0.0055 \times RH] \times [1.8 \times T - 26].$$

Afterwards, we calculated the prenatal THI for the week before birth (0-7 d) by averaging hourly THI. For calf diseases, all THI values <60 were set to 60 to eliminate the impact of possible cold load. The average THI for the week before calving was scaled using -1+2((THI-60)/(74-60)), where 60 = minimum THI value and 74 = maximum THI value. The scaled THI were used in the analyses for calf diseases. For cow production traits, a dummy variable with values of either 0 (THI<60) or 1 (THI≥60) was used to indicate prenatal HS. We focused on different strategies to consider THI, because data structures and applied models were not the same for the two trait categories.

Statistical models. For calf diseases, an animal model with maternal effects considering prenatal THI in the week before birth, was applied. The statistical in matrix notation was:

$$y = Xb + Qd + Wm + Sp_m + e \quad (1)$$

where y = vector of records for PNEU or DIAR; b = vector of fixed effects including herd, birth year, birth month, gestation length classes, lactation-calving-age classes for dams, sex-birth type and regressions on scaled THI nested within birth year; d = vector of direct additive-genetic effects; m = vector of maternal genetic effects for random intercepts and slopes on scaled THI; p_m = vector of maternal permanent environmental effects for random intercepts and slopes on scaled THI; e = vector of random residual effects.

For cow production traits, a genomic interaction model (Yao *et al.*, 2017) considering interactions between genotype and prenatal HS was defined:

$$y = Xb + Zg + Wg_{hs} + e \quad (2)$$

where y = vector of observations for test-day or lactation records; b = vector of fixed effects including herd, calving year, calving month, age at first calving classes, lactation-calving-age classes for dams, classes for days in milk (for the test-day traits), lactation length classes (for lactation traits) and a dummy variable

indicating the presence of HS during the week before calving; g = vector of additive genomic effects; g_{hs} = vector of genotype by HS interaction effects for cows with prenatal HS; e = vector of random residuals. The variances of g and g_{hs} were defined as common genomic and interaction variance, respectively. The common genomic variance consistently existed for all cows regardless of prenatal HS. The interaction variance represented the variance due to prenatal HS. Proportions of the common genomic and the interaction variance in relation to the phenotypic variance were defined as common heritability (h_c^2) and interaction heritability (h_i^2), respectively. Additionally, a genomic animal model without interaction effects was applied to identify changes of the heritabilities due to genotype by HS interactions. Gibbs sampling and AIREML algorithms as implemented in the software packages THRGIBBS1f90 and AIREMLf90 (Misztal *et al.*, 2002) were used to estimate variance components for calf diseases and cow production traits, respectively.

Results

Direct heritabilities for PNEU were quite stable with 0.10 at THI 60 to 0.08 at THI 74. For DIAR, the direct heritabilities ranged from 0.16 to 0.17. Maternal heritabilities for PNEU increased from 0.03 to 0.10 along the THI gradient. Oppositely, the maternal heritability for DIAR was slightly larger at THI 60 (0.07) than under prenatal HS conditions. The genetic correlations between maternal effects at THI 60 with the remaining THI values for both calf diseases decreased with increasing THI distances (Figure 1). The smallest correlations of 0.42 for DIAR and 0.26 for PNEU were estimated between the most distant prenatal THI values.

From model 2, h_c^2 were 0.13 for TMY, 0.19 for TFP, 0.15 for TFY, 0.24 for TPP, 0.10 for TPY, and 0.08 for TSCS (Figure 2). For lactation milk production traits, h_c^2 were 0.37 for LMY, 0.72 for LFP, 0.35 for LFY, 0.70 for LPP, and 0.28 for LPY. In general, common genomic heritabilities for milk production traits were identical to genomic heritabilities estimated from the alternative model without G×E interaction effects (results not shown). Among all traits, TPP displayed the largest heritability due to genotype by HS interactions (3.79%). For TFP, TFY, TPY, TSCS, LFP, and LPY, h_i^2 were <1%.

Discussion

Increases in direct heritabilities for cow diseases (Kipp *et al.*, 2021) and milk yield (Bohlouli *et al.*, 2019) along prenatal and prompt THI, respectively, indicate pronounced genetic differentiation in challenging environments. Compared to cow traits, genetic effects of calf diseases comprise direct and maternal genetic effects. Consequently, we focused on the effects of prenatal THI on maternal genetic effects. For PNEU, larger maternal genetic variations were identified in dams exposed to extreme prenatal HS. Oppositely,

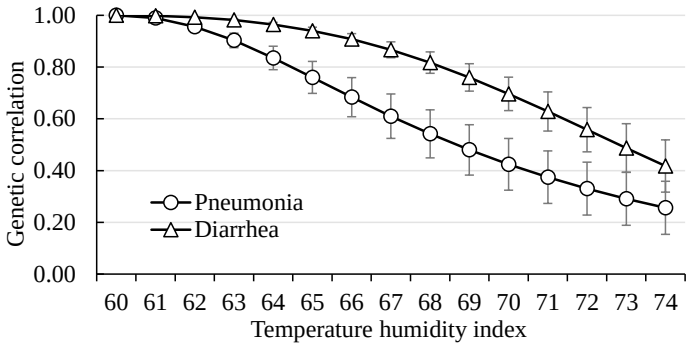


Figure 1. Posterior means for genetic correlations with respective posterior standard deviations between maternal effects at THI 60 and at the remaining THI.

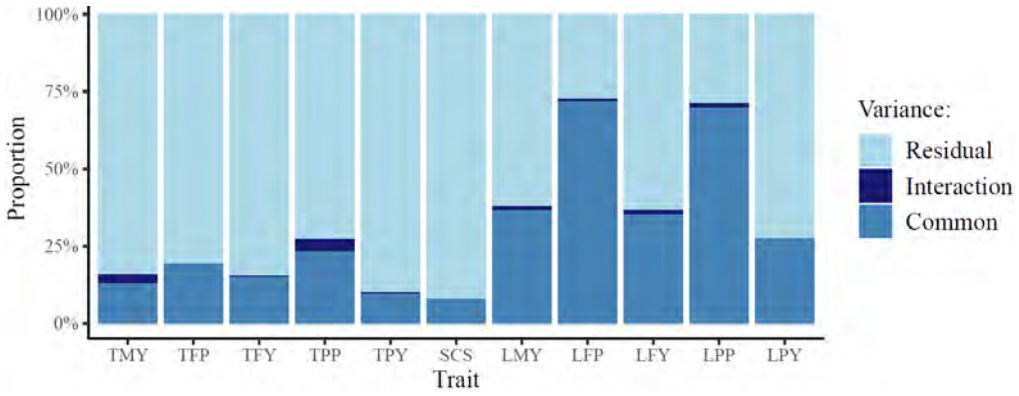


Figure 2. Proportions of common genomic, interaction and residual variances in relation to the total variance for test-day and lactation production traits.

DIAR displayed larger maternal genetic variances under comfort conditions. Hence, responses of maternal genetic effects on prenatal THI alterations are trait-specific. Genetic correlations smaller than 0.80 for calf diseases recorded at THI 60 with corresponding diseases at THI 74 suggest obvious time-lagged G×E for maternal genetic effects. Time-lagged G×E for direct genetic effects due to HS in late pregnancy were reported by Kipp *et al.* (2021) for primary and functional traits in lactating cows. For cow production traits, the ranking of sires according to estimated breeding values under comfort and prenatal HS conditions was almost the same. In analogy, identical heritabilities were estimated for production traits from the model with or without genotype by HS interactions. Additionally, for all traits, the interaction variances only explained small proportions of the total phenotypic variances (up to 3.79%). Similarly, Kipp *et al.* (2021) confirmed the absence of time-lagged G×E for milk production traits, using a random regression sire model. In contrast, ratios of environment-specific genomic variances to overall variances for feed efficiency traits (Yao *et al.*, 2017) were larger than h_i^2 from this study. Consequently, the effect of prenatal HS during late pregnancy is smaller than the prompt herd effect due to geographical impact across latitudes. Nevertheless, effects of HS on direct and maternal genetic (co)variance components of traits in offspring are first indications for epigenetic mechanisms, suggestion ongoing molecular genetic studies in this regard.

References

Bohlouli, M., Alijani, S., Naderi, S., Yin, T., and König, S. (2019). J. Dairy Sci. 102(1):488–502.

Gernand, E., König, S., and Kipp, C. (2019). J. Dairy Sci. 102(7):6660–6671.

Halli, K., Brügemann, K., Bohlouli, M., Yin, T., and König, S. (2021). J. Anim. Sci. 99(5).

Kipp, C., Brügemann, K., Yin, T., Halli, K., and König, S. (2021). J. Dairy Sci. 104(9):10029–10039.

Misztal, I., Tsuruta, S., Strabel, T., Auvray, B., Druet, T., *et al.* (2002). Proc. of the 7th WCGALP, Montpellier, France.

Monteiro, A.P.A., Tao, S., Thompson, I.M.T., and Dahl, G.E. (2016). J. Dairy Sci. 99(10):8443–8450.

NRC (1971). A guide to environmental research on animals. National Academy Science, Washington, D.C., United States.

Stock, K.F., Cole, J., Pryce, J., Gengler, N., Bradley, A., *et al.* (2013). ICAR Tech. Ser. 17:75–81.

Tao, S., and Dahl, G.E. (2013). J. Dairy Sci. 96(7):4079–4093.

Yao, C., de los Campos, G., VandeHaar, M.J., Spurlock, D.M., Armentano, L.E., *et al.* (2017). J. Dairy Sci. 100(3):2007–2016.

102. A genetic evaluation for calf health: preliminary analysis

C. Lynch[1*], F. S. Schenkel[1], K. Houlahan[1], N. van Staaveren[1], F. Miglior[1,2] and C.F. Baes[1,3]

[1]Centre for Genetic Improvement of Livestock, Department of Animal Biosciences, University of Guelph, Canada; [2]Lactanet Canada, Guelph, Ontario, Canada; [3]Institute of Genetics, Vetsuisse Faculty, University of Bern, Bern, Switzerland; clynch@uoguelph.ca

Abstract

Growing consumer awareness means that the dairy industry needs to take proactive steps to improve animal health. Limited work has been conducted for improving Canadian calf health through genetics, and therefore, this study aimed to understand the current impact of calf health diseases on Canadian farms and to estimate genetic parameters for calf health traits. Calf health data comprised of 76,405 Canadian Holstein calf disease records for respiratory problems and diarrhea, collected on 1,890 Canadian dairy herds from 2009 to 2020. Incidence rates for respiratory problems were similar to values previously reported in literature, whereas diarrhea incidence rates were lower. Heritabilities of 0.01 and 0.04 were estimated for diarrhea and respiratory problems, respectively, which were lower than estimates reported in the literature. To maximise estimates stricter herd inclusion protocols, incorporating genomic information and splitting the traits by time of occurrence need to be investigated.

Introduction

Emerging issues in dairy, related to animal welfare and social acceptability mean that the dairy industry must take proactive steps to improve animal health. Calf-hood diseases have a major impact on the economic viability of cattle operations, due to costs of treatment and long-term effects on performance (Svensson and Hultgren 2008; Mohd Nor *et al.* 2013). In the literature incidence rates for diarrhea range from 23 to 44%, while producers reported treating roughly 12% of pre-weaned calves for respiratory diseases (USDA 2010, 2018; Urie *et al.* 2018; Haagan *et al.* 2021). Though incidence reports are not available in Canada, diarrhea and respiratory diseases account for 53 and 21% of pre-weaned calf mortality in Canada, respectively (Murray 2011). Therefore, efforts to improve dairy calf health are vital.

From a genetic perspective, efforts aimed at breeding for increased disease resilience offer opportunities. Reported heritability estimates for respiratory disease and diarrhea range between 4 to 10% (Henderson *et al.* 2011; McCorquodale *et al.* 2013; Haagan *et al.* 2021), indicating the potential for genetic selection. In Canada, research has looked at improving overall cow health through initiatives such as Immunity+ provided by Semex, but this is proprietary, and no strategy for including calf health traits within national evaluations has been proposed. Therefore, the objective of this study was to understand the current impact of calf health diseases on Canadian farms and to estimate genetic parameters for calf health traits using existing data available. This work will provide the foundations for the inclusion of calf health traits within national Canadian genetic evaluations. Specifically, these results can be used to identify gaps in data collection methods which should be addressed in future developments for successful implementation of a national calf health selection strategy.

Materials & methods

Data. Calf health data recorded by dairy producers through management software was provided by Lactanet Canada (Guelph, Ontario). It comprised 76,405 Holstein calf disease records for respiratory problems and diarrhea, collected on 1,890 Canadian dairy herds from 2009 to 2020. All other calves present on the farm in a given year without a disease record were assumed to be healthy. To be considered in the analysis, records must have occurred within the first six months of the calf's life and each herd required a minimum

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_102, © C. Lynch *et al.* 2022

annual disease incidence rate of at least 1%. This was calculated using the number of disease events in a herd per year over the total number of heifer calves raised on that farm in a given year. Additionally, herds had to have consecutive calf health records for years leading to and including 2020, or a minimum of three consecutive years of data available if there were no 2020 records. These metrics provided insight to which herds were consistently collecting calf disease information. A total of 62,168 records were used in the analysis after data validation and cleaning (Table 1).

Models. Univariate threshold animal models with a probit link function were used for both traits. A Bayesian approach based on the Gibbs sampling algorithm using the Markov chain Monte Carlo method with 1,000,000 iterations, 100,000 burn-in and 50 thinning interval was used for the analysis. Convergence of all chains was achieved, based on the Heidelberger and Welch, and Geweke convergence diagnostic tests. The general form of the model used was defined as:

$$l = Xb + Za + e \quad (1)$$

where l is a vector of underlying liabilities corresponding to the binary observation (0= healthy, 1= diseased), b is a vector of systematic fixed effects of year born-month born and herd, a is a vector of random additive genetic effects, e is a vector of random residuals and, X and Z are corresponding incidence matrices.

Results

Incidence rates for each disease by year are presented in Table 2. The median incidence rate for each year was always lower than the mean for both traits. This was due to a number of herds with high incidence rates inflating the mean value. Given this, the median value should be seen as a better representation of the true average. For diarrhea, median incidence rate in herds ranged from 5 to 10% with the highest incidence rate

Table 1. Final datasets of calf health records.

Trait	Diseased	Healthy	Total records (herds)
Diarrhea	18,887	101,857	120,744 (425)
Respiratory problems	43,281	212,502	255,783 (664)

Table 2. Descriptive statistics of diseases incidence rates in herds across years.

Year	Diarrhea			Respiratory problems		
	Mean %	Median %	Herds	Mean %	Median%	Herds
2009	10	5	54	16	8	133
2010	12	7	68	13	8	154
2011	12	8	86	15	10	187
2012	12	7	93	16	9	205
2013	11	7	98	16	10	230
2014	13	8	93	15	10	245
2015	13	8	107	16	10	253
2016	11	7	85	13	9	228
2017	11	6	98	14	10	260
2018	12	8	128	16	11	327
2019	13	8	166	17	12	355
2020	16	10	145	16	11	323

seen in 2020. There was also a considerable increase in the number of herds collecting calf health data with this number rising from 54 herds in 2009 to 145 in 2020. For respiratory problems, median incidence rate in herds ranged from 8 to 12%, with the highest incidence rate seen in 2019. Again, there was a considerable increase in the number of herds collecting calf health data with this number rising from 133 herds in 2009 to 323 in 2020.

Variance components for each trait are presented in Table 3. Heritabilities were originally estimated on the liability scale and converted to the observed scale. Heritabilities of 0.01 and 0.04 were estimated for diarrhea and respiratory problems, respectively.

Discussion

Reported incidence rates for respiratory problems were similar to the mean incidence rates previously reported in the literature, while for diarrhea values were lower (USDA 2010, 2018; Urie *et al.* 2018; Haagan *et al.* 2021). These differences may be due to several reasons, including quality of data collection and management practices. However, it is evident that consistent levels of calf disease were observed across years, indicating that there is an opportunity to reduce incidence rates and improve farm efficiency.

Heritabilities estimates for diarrhea and respiratory problems within this study were lower than estimates reported in the literature (Henderson *et al.* 2011; Mahmoud *et al.* 2017; Haagan *et al.* 2021). These differences could be attributed to several reasons including the lack of detailed information on individual calves, such as, birth weights and colostrum intake, which are directly related to the occurrence of respiratory problems and diarrhea (Henderson *et al.* 2011; Mahmoud *et al.* 2017). Another challenge is the uncertainty surrounding recording practices on farms. Farms without strict recording practices may have had sick animals that were assumed healthy in the current study, impacting model performance. Haagen *et al.* (2021) reported heritability estimates of 0.10 and 0.08 for respiratory problems and scours, respectively, with a similar lack of direct calf information as found in this study. The main difference was the use of genomic data for the estimation of genetic parameters. Furthermore, Vukasinovic *et al.* (2018) found higher estimates for respiratory problems by splitting the trait by age of occurrence, with heritabilities of 0.10 between 0 to 3d of age, 0.06 at 4 to 14d of age, and 0.05 at 15 to 50d of age, all above the estimate of 0.04 between day 0-365 of age. Given this, methods such as stricter herd inclusion protocols, incorporating genomic information, and looking at traits at specific times needs to be investigated with the current data to try to maximise model performance and improve parameter estimates.

Conclusions

Growing consumer awareness means that the dairy industry needs to take proactive steps to improve animal health. In Canada, research has looked at improving overall cow health through initiatives such as Immunity+ provided by Semex, but no strategy for including calf health traits within national evaluations

Table 3. Variance components of disease traits in Canadian dairy calves.

Parameters[1]	Diarrhea	Respiratory problems
Residual variance	1.00	1.00
Genetic variance (HPD)	0.02 (0.01-0.02)	0.05 (0.04-0.05)
Liability scale h^2 (HPD)	0.02 (0.01-0.02)	0.05 (0.04-0.05)
Observed scale h^2	0.01	0.04

[1] HPD = highest probability density.

has been proposed. This study determined the current impact of calf health diseases on Canadian farms and provided initial estimates of genetic parameters for respiratory problems and diarrhea in Canadian dairy calves and, although the estimates were lower than those seen in other studies, they provide the foundations for future work. To improve estimates stricter herd inclusion protocols, incorporation of genomic information and splitting the traits by time of occurrence need to be investigated.

Acknowledgements

The authors acknowledge funders, partners and collaborators of the Resilient Dairy Genome Project (http://www.resilientdairy.ca).

References

Haagen I.W., Hardie L.C., Heins B.J., and C.D. Dechow. (2021) J. Dairy Sci. 104:11770–11778. https://doi.org/10.3168/jds.2021-20432

Henderson L., Miglior F., Sewalem A., Wormuth J., Kelton D., *et al.*. (2011) J. Dairy Sci. 94:6181–6187. https://doi.org/10.3168/jds.2011-4347

Mahmoud M., Yin T., Brügemann K., and König S. (2017) J. Dairy Sci. 100:2017–2031. https://doi.org/10.3168/jds.2016-11767

McCorquodale C.E., Sewalem A., Miglior F., Robinson A. Koeck A., *et al.*. (2013) J. Dairy Sci. 96:1880–1885 https://doi.org/10.3168/jds.2012-5735

Mohd Nor N., Steeneveld W., van Werven T., Mourits M.C.M., and Hogeveen H. (2013) J. Dairy Sci. 96:981–992. http://dx.doi.org/10.3168/jds.2012-5741

Murray B. (2011) Optimizing calf survival at birth. Ontario Ministry of Agriculture, Food, and Rural Affairs. Available at: http://www.omafra.gov.on.ca/english/livestock/dairy/facts/optbirth.htm

Svensson C., and Hultgren J. (2008) J. Dairy Sci. 91:1510–1518. https://doi.org/10.3168/jds.2007-0235

Urie N. J., Lombard J. E., Shivley C. B., Kopral C. A., Adams A. E., *et al.*. (2018) J. Dairy Sci. 101:9229–9244. https://doi.org/10.3168/jds.2017-14019

USDA. (2010) Dairy 2007: Heifer calf health and management practices on U.S. dairy operations, 2007. USDA-APHIS-VS-CEAH. Available at: https://www.aphis.usda.gov/animal_health/nahms/dairy/downloads/dairy07/Dairy07_allpubs.pdf

USDA. (2018) Dairy 2014: Health and Management Practices on U.S. Dairy Operations, 2014. USDA-APHIS-VS-CEAH-NAHMS. Available at: https://www.aphis.usda.gov/animal_health/nahms/dairy/downloads/dairy14/Dairy14_dr_PartIII.pdf

Vukasinovic N., Gonzalez-Peña D., Brooker J.J., Desai P., and Denise S.K. (2018) Proc. of the 11th WCGALP, Auckland, New Zealand.

103. Development of a single step genomic evaluation of health traits and relationships with dairy traits in French Holsteins

H. Leclerc[1,2*], A. Barbat[2], R. Vallée[3,2], F. Guillaume[4], L. Manciaux[5], J.-B. Davière[6] and M. Philippe[4]

[1]ALLICE, 149 rue de Bercy, 75012 Paris, France; [2]Univ. Paris Saclay, INRAE, AgroParisTech, GABI, 78350 Jouy-en-Josas, France; [3]IDELE, 149 rue Bercy, 75012, France; [4]EVOLUTION, CS 10040, 35538 Noyal sur Vilaine, France; [5]INNOVAL, CS 80038, 35538 Noyal sur Vilaine, France; [6]SEENERGI, CS 84215, 53942 St Berthevin, France; helene.leclerc@inrae.fr

Abstract

Selecting for better health has become of primary importance for the dairy industry. In France, where no national database exists for health traits (except mastitis), a collaborative project has been implemented to explore and evaluate farmer health records. Around 10% of herds record health events accurately. Genetic parameters of 4 disorders have been studied here: Displaced Abomasum (DA), Milk Fever (MF), Metritis (MET) and Retained Placenta (RP). Heritabili-ties are low (1-2%), but health traits have moderate genetic correlations with some dairy traits (0.15-0.45). Based on EBV obtained with ssGBLUP for the four health traits, the daughters' incidence of best EBV bulls versus worst bulls is reduced by a factor of 2 minimum.

Introduction

Health traits are of increasing importance for dairy farmers, both for economic reasons (treat-ments' cost/production loss), for working time and for animal well-being. Since 2000, French laws require breeders to record all treatments in a diary. The pathology causing a treatment may also be recorded. Rapidly, a growing number of breeders started recording using one of the available IT application (software, web, etc.), thus making disease data available to their support organisations. In 2015, a collaborative regional project named 'GénoSanté' was initiated by the *Evolution* Breeding Company, together with milk recording and herd support organisations (DHI) of Western France, breeding companies and R&D organisations. The aim of this project was to improve animal health through new tools both for management and selection (Leclerc *et al.*, 2019), adding extra-value to information recorded by farmers in health management diaries. In 2015, a study on a subset of IT application showed that 15% of herds recorded health events exhaustively (Bourrigan *et al.*, 2015) and suggested that these data could help to develop references for management tools and genetic evaluations. The objectives were: (1) to validate the feasibility of using breeders recorded health data for gene-tic purpose; (2) to estimate genetic relationships between some available health traits of interest and other dairy traits; (3) to develop a Single Step Genetic Evaluation of those health traits.

Materials & methods

The health database. 742,973 multi-breed health events (mastitis discarded), corresponding to 178 treatments reasons, collected by farmers from 2008 to 2021 in 16,619 herds were available. Due to lack of supervision in the data acquisition system, the consistency of data was checked by cross-referencing all the data available for those animals with the national genetic databases. Six different categories of animals were established: calves less than 1 month old, calves from 1 to 6 months old, heifers over 6 months old, pregnant heifers, cows, and males. A list of possible diseases was defined for each category. For instance, calving pathologies such as uterine prolapse are only possible for the 'cow' category. Inconsistent data (2.1%) and uncategorized data (3.9%) were discarded. Data from herds with less than 20 lactations per year were excluded (5.6%), as well as data from collective treatments that did not allow to assess individual sensitivity (22.9%). Among health disorders, four traits were selected: Displaced Abomasum (DA), Milk Fever (MF), Metritis (MET) and Retained Placenta (RP). Lameness and ketosis

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_103, © H. Leclerc *et al.* 2022

disorders recorded in herd diaries were not studied here, genetic evaluations for ketosis (through milk MIR spectra of Acetone and Beta-Hydroxybuty-rate (BHB)) and lameness (from claw lesions) being already available (Leclerc *et al*., 2019).

Completeness at herd level. To ensure that data available are from reliable herds, a minimum of 3 different health disorders and health events recorded for at least 15% of the lactating cows per year were required to consider the health status of the herd as complete for that year, regardless of breeds. A minimum of 5 consecutive years of completeness was required to select a herd, to ensure this period covers the cow productive lifespan and to have a dataset quality suitable for the estimation of genetic parameters.

Traits' definition. Health disorders were defined as binary traits (0=no case / 1=at least one case) based on whether or not the cow had at least one disease case recorded during the lactation period (0-60 days after calving for DA, 0-7 d for MF, 0-150 for MET, 0-14 d for RP). MET and RP periods are similar to those used in Canada (Koeck *et al.*, 2012). Breeders are more or less sensitive to certain diseases, so only herds with at least one case recorded in the considered year, the previous or the following one have contemporaries' cows considered as 'no case' (617 herds considered for DA, 756 for MF, 767 for MET and 845 for RP).

Information from the national database was added for other traits on 409,031 lactations (parity 1 to 5) from 2006 to 2021 (Table 1): Milk production traits, recorded between 7 and 40 days in milk (DIM) (kg of milk, Fat%, Pro%, SCS, acetone and BHB estimated from MIR spectra on milk sample ($100\times\log_2([\text{acetone or BHB}]+1)$); number of insemination (#INS), interval between calving and first insemination (ICFINS), interval between calving and last insemination (ICLINS), and number of mastitis cases in the lactation (#MAS).

Analysis. This study, focused on Holsteins' data from the exhaustive herds, was split into 3 parts: (1) variance components estimation within the 4 health traits (herd diaries database); (2) estimation of genetic correlations between the 4 health traits, and acetone/BHB (Milk database), udder health, fertility and production traits (national database); (3) genomic evaluation of the 4 health traits by SSGBLUP.

Pedigree and genotyped animals. Animal pedigree was built by tracing the pedigree of the 167,498 cows with data on 5 generations (339,406 animals). A total 23,132 animals genotyped (or imputed) on the Illumina Bovine SNP50 BeadChip were used for ssGBLUP.

Model. Bivariate and Multivariate linear animal models, with repeated observations across lactations, were fitted by a REML using Wombat software (Meyer, 2007). Multivariate analysis were performed among the 4 health traits, and bivariate for other traits combinations.

Table 1. Summary statistics of analyzed traits.

Trait	N°	#Lact	Mean	Std	Trait	N°	#Lact	Mean	Std
DA (0/1)	a	193,650	0.011	0.102	#MAS	c	409,031	0.44	0.85
MF (0/1)	a	257,482	0.018	0.134	Milk (kg)	b	351,492	32.71	8.26
MET (0/1)	a	215,837	0.063	0.243	Fat%	b	351,492	4.08	0.71
ND (0/1)	a	285,252	0.062	0.241	Pro%	b	351,491	3.03	0.30
#INS	c	351,565	2.21	1.55	SCS	b	351,491	2.42	1.88
ICFINS (days)	c	351,565	91.64	38.15	Acetone	b	239,402	20.89	13.57
ICLINS (days)	c	289,916	134.22	73.13	BHB	b	223,954	15.15	12.17

$$y= X\beta+Za+Zp+e \quad (1)$$

where y is the vector of observations for the traits; β the vector of fixed effects consisting of a herd×year (a,b,c), calving age×parity (a,b,c), calving month×year (a,b), length of dry period×parity (a,b), parity×DIM (a) [a,b,c as described in Table 1]; a ($\sim N(0,A\sigma^2_g)$) and p ($\sim N(0,I\sigma^2_{pe})$) the vectors of additive genetic and permanent environment, X & Z the incidence matrices.

A SingleStep genetic evaluation using HSSGBLUP software (Tribout *et al.*, 2020) using uni-variate model similar to variance component estimation was performed on the 4 health traits.

Results

Around 2/3 of health data remained after data edits. It was used to qualify the completeness of 10,019 herds and 8.6% of herds (867) have fulfilled the 5 consecutive years of completeness.

The disease frequencies of the 4 health traits were low, from 1.1% for DA to 6.3% for MET (Table 1). MF frequency were highly dependent of parity (close to 0 in 1st and 2nd up to 9% in 5th). Heritability estimates of the 4 health traits in linear model were between 1 and 2% (Table 2). MET genetic correlations were high with RP (0.64) to moderate with DA (0.31).

Genetic correlations between health traits and other dairy traits are shown in Table 3 (for production traits (dark grey cells), a negative correlation means a favourable relationship). DA and MF have favourable correlations with ketosis (Acetone & BHB) and fertility traits. MET, and RP to a lesser extent, have positive genetic correlations with #INS and ICLINS. Acetone and BHB have moderate favourable correlations with SCS and Pro% and unfavourable with Fat%.

Table 2. Heritability (SE in parenthesis), repeatability, genetic correlations (SE) (above the diagonal) and phenotypic correlations (below the diagonal) within four health traits.

			Correlations			
Trait	**h^2**	**r_p**	**DA**	**MF**	**MET**	**RP**
DA	0.020 (0.002)	0.020		0.09 (0.08)	**0.31** (0.08)	0.14 (0.07)
MF	0.015 (0.002)	0.029	0.013 (0.003)		-0.05 (0.09)	0.04 (0.07)
MET	0.010 (0.001)	0.027	0.036 (0.003)	0.012 (0.003)		**0.64** (0.06)
RP	0.018 (0.002)	0.034	0.033 (0.003)	0.012 (0.002)	0.122 (0.002)	

Table 3. Genetic parameters between health traits, dairy and fertility-mastitis traits.

			Genetic correlations					
Trait	**h^2**	**r_p**	**DA**	**MF**	**MET**	**RP**	**Acetone**	**BHB**
Milk	0.122 (0.004)	0.218	-0.13	-0.02	-0.12	-0.11	0.06	0.03
Fat%	0.214 (0.004)	0.262	0.03	0.07	-0.05	-0.11	0.13	**0.25**
Pro%	0.274 (0.004)	0.311	**-0.16**	-0.08	-0.08	-0.11	**-0.25**	**-0.28**
SCS	0.079 (0.003)	0.141	**0.15**	0.10	**0.17**	0.01	**0.18**	**0.20**
Acetone	0.112 (0.004)	0.150	**0.28**	**0.27**	0.01	-0.02	-	**0.95**
BHB	0.101 (0.004)	0.152	**0.18**	**0.27**	-0.10	-0.06	**0.95**	-
#INS	0.044 (0.003)	0.107	**0.18**	-0.03	**0.45**	0.13	-	-
ICFINS	0.081 (0.003)	0.162	**0.22**	**0.21**	0.10	0.09	-	-
ICLINS	0.060 (0.003)	0.120	**0.22**	**0.20**	**0.36**	**0.18**	-	-
#MAS	0.076 (0.003)	0.152	-0.04	**0.15**	0.14	0.06	-	-

Genomic EBV for 4 health traits have been estimated in an univariate ssGBLUP. Based on 636-779 bulls with at least 20 daughters, in comparison with the 25% worst bulls, the 25% best ones have on average an incidence divided by 2 (MET=8.6% vs 4.2% / RP=8.9% vs 3.9% / MF=3.1% vs 1.3%) to more than 10 (DA=2.2% vs 0.2). The female genetic trend (Figure 1) show a favourable evolution of the genetic level for DA, MET and RP since 2008.

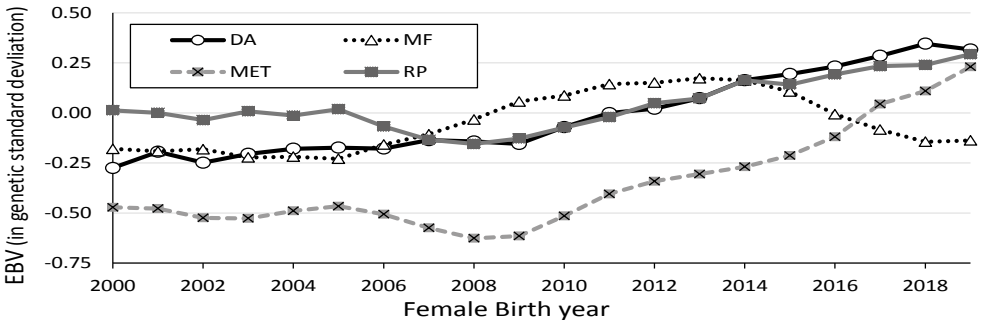


Figure 1. Female genetic trends for DA, MF, MET and RP.

Discussion

Farmer-observed health events recorded in diaries can be used to develop genetic evaluation, even if the number of farms is very limited. At the end, 8.6% of herds value 41% of consistent data (a 3 years completeness period would value 18% of herds – 56% of data). This could be improved by focusing DHI effort to increase participation. Highlighting the benefits of disease recording for herd management and genetic evaluations is essential.

Linear model was preferred to threshold model due to practical consideration on multivariate model and ssGBLUP software although the latter is generally recommended with binary traits, such as health traits. Heritability estimates for the 4 health traits (1-2%) were, slightly lower than in Koeck *et al.* (2012), but consistent with the review of Pryce *et al.* (2016). High genetic correlations between some of the 4 health traits (MET-RP and DA-MET) are consistent to the ones on Canadian study (Koeck *et al.*, 2012) with lower standard errors. Moderate but favour-able correlations between the 4 health traits and functional/other health traits give a promising situation that could, if the latter group of traits are used as predictors, increase the reliabilities of EBV for health traits. The favourable genetic trends observed for DA, MET and RP since 2008 could be an indirect effect of the inclusion of functional traits in Total Merit Index.

Acknowledgements

GénoSanté was labelled by the Valorial competitiveness cluster. It has received funded from the Fonds Unique Interministériel (BPI France, regions of Brittany, Alsace, Pays de Loire).

References

Bourrigan X., Blériot G., Bouissel C., *et al.* (2015) Renc. Rech. Ruminants 22:41-44.

Koeck A., Miglior F., Kelton D.F. and Schenkel F.S. (2012) J. Dairy Sci. 95:4099-4108. https://doi.org/10.3168/jds.2011-5127

Leclerc H., Croué I., Vallée R., Baur A., Barbat A., *et al.* (2019) ICAR Techn. 24:373-381.

Meyer K. (2007) J ZHEJIANG UNIV-SC B 8(11):815-821 https://doi.org/10.1631/jzus.2007.B0815

Pryce J.E., Parker Gaddis K.L., Koeck A., Bastin C., Abdelsayed M. *et al.* (2016) J. Dairy Sci. 99:6855-6873. 4099-4108. https://doi.org/10.3168/jds.2016-10854

Tribout T., Ducrocq V. and Boichard D. (2020) Proc. of the 6th ICQG, Brisbane, Australia.

104. Genetic selection for reduced susceptibility to subclinical ketosis using mid-infrared predicted traits

A. Köck[1*], L.M. Dale[2], A. Werner[2], M. Mayerhofer[1], F.J. Auer[3] and C. Egger-Danner[1]

[1]ZuchtData EDV-Dienstleistungen GmbH, Dresdner Str. 89, 1200 Vienna, Austria; [2]Regional State Association for Performance and Quality Inspection in Animal Breeding of Baden Württemberg, Heinrich-Baumann-Str. 1-3, 70190 Stuttgart, Germany; [3]LKV-Austria, Dresdner Str. 89, 1200 Vienna, Austria; koeck@zuchtdata.at

Abstract

The objective of this study was to evaluate the value of milk mid-infrared (MIR) predicted traits in breeding programs in reducing the incidence of subclinical ketosis. Data from 670 cows on 49 farms were collected from September 2020 to March 2021. A handheld electronic device (WellionVet BELUA, MED TRUST Handels GmbH, Marz, Austria) for the detection of subclinical ketosis (β-hydroxybutyrate-concentration >1.2 mmol/l) was used on all cows at milk recording in early lactation. In addition to fat-to-protein ratio, data from newly developed MIR traits (KetoMIR, β-hydroxybutyrate, acetone) were available at each milk recording. Heritabilities for all traits were high, ranging from 0.16 (acetone) to 0.30 (KetoMIR). Genetic correlations between β-hydroxybutyrate results gained by the handheld device and traits predicted by MIR ranged from 0.60 to 0.73. The moderate to high genetic correlations between MIR-predicted traits and subclinical ketosis suggest that consideration of these traits would reduce subclinical ketosis.

Introduction

In Austria, the frequency of clinical ketosis is very low (1.2%) (ZuchtData 2020). Subclinical ketosis occurs much more frequently. Data from a previous project showed a frequency of subclinical ketosis of around 14% (ZAR 2016). Since affected cows usually show no signs of disease at all, detection is very difficult. For the diagnosis of subclinical ketosis, various tests for the measurement of ketone bodies in blood, milk or urine are available to the farmer. Khol *et al.* (2019) investigated a handheld electronic device for the detection of subclinical ketosis (β-hydroxybutyrate-concentration (BHB)>1.2 mmol/l), in capillary blood and venous whole blood in cows (WellionVet BELUA, MED TRUST Handels GmbH, Marz, Austria). Altogether, 98% of all the samples were correctly classified as subclinical ketosis or non-subclinical ketosis by the handheld device in capillary blood (sensitivity 0.96, specificity 0.98) and 97.4% in venous whole blood (sensitivity 0.889, specificity 0.991), respectively. This rapid test has high accuracy but means additional work. Therefore, when detecting subclinical ketosis, the milk components from the milk performance recording provide valuable practical information that does not require any additional work. Milk analysis using mid-infrared spectroscopy (MIR) is a fast and inexpensive way of examining milk samples on a large scale for fat, protein, lactose and urea. The classical detection method of subclinical ketosis within routine performance recording is a fat-protein ratio>1.5 at the beginning of lactation. Recently, MIR equations for milk components with low concentrations such as BHB and acetone were developed. Detection of these components is done by indirect links with global milk composition, as MIR technology cannot directly detect components below 100 ppm (<0.1 g/l) (Dardenne *et al.* 2015). A new indicator trait for ketosis, KetoMIR, was developed by LKV Baden-Württemberg (Dale *et al.* 2018, Werner *et al.* 2019). The new concept of KetoMIR is to offer a ketosis risk index via milk samples that is more accurate than the fat-protein ratio. The objective of this study was to evaluate the value of fat-protein-ratio and these newly developed MIR traits (KetoMIR, BHB-MIR, Acetone-MIR) for genetic selection for reduced susceptibility to subclinical ketosis. Cows with subclinical ketosis were identified by using the handheld electronic device. Heritabilities and genetic correlations were estimated for all traits.

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_104, © A. Köck *et al.* 2022

Materials & methods

Data. This project involved 49 farms with metabolic problems from September 2020 to March 2021. Data collection was carried out by the regional milk recording organisations. As part of the study, capillary blood was analyzed for BHB concentration in all cows during milk recording in early lactation (1st and 2nd test day after calving) using a handheld device (WellionVet BELUA, MED TRUST Handels GmbH, Marz, Austria). In addition, feeding behaviour (1=does not eat, 2 = eats reduced, 3 = eats normal) and metabolism (ketosis prevention, ketosis treatment, milk fever prevention, milk fever treatment) were recorded.

Traits. For the analyses, BHB measurements from the handheld device were divided into 3 classes (1 = BHB≤1.2 mmol/l (healthy), 2 = BHB=1.3-2.9 mmol/l (subclinical ketosis), 3 = BHB≥3.0 mmol/l (clinical ketosis)). KetoMIR was defined as a three-class index (1 = low ketosis risk, 2 = medium ketosis risk, 3 = high ketosis risk). BHB-MIR, Aceton-MIR and fat-protein-ratio were considered as continuous traits. All traits were available on each test-day.

Data edits. Analyses were carried out for Fleckvieh cows from all parities. In total 504 records from 323 Fleckvieh cows were available for analyses. A summary of statistics of the analyzed data set is given in Table 1. An animal pedigree file was generated by tracing the pedigree of cows 5 generations back, ending up with 3,523 animals.

Model. Data were analyzed with univariate and bivariate linear animal models using the average information-restricted maximum likelihood (AI-REML) procedure in the DMU package (Madsen and Jensen, 2008). The following model was applied to all traits:

$$y = X\beta + Z_h h + Z_{pe} pe + Z_a a + e \quad (1)$$

where y is a vector of observations; β is a vector of systematic effects, including fixed effects of month of recording, parity and lactation stage; h is a vector of random herd effects; pe is a vector of random permanent environmental effects; a is a vector of random animal additive genetic effects; e is a vector of random residuals; and X, Z_h, Z_{pe}, and Z_a are the corresponding incidence matrices. Parity had five classes with 1, 2, 3, 4 and 5+. Lactation stage was defined in classes (1 = ≤20 DIM, 2 = 21-40 DIM, 3 = 41-60 DIM, 4 = 61-80 DIM, 5 = >80 DIM).

Results

Figure 1 shows the percentage of healthy cows (BHB≤1.2 mmol/l), cows with subclinical ketosis (BHB=1.3-2.9 mmol/l) and clinical ketosis (BHB≥3.0 mmol/l) detected by the handheld device.

Table 1. Summary statistics of the analysed traits.

	Number of records	Mean	Minimum	Maximum
BHB – handheld device	504	1.298	1	3
KetoMIR	474	1.227	1	3
BHB-MIR	474	1.025	0.370	3.630
Aceton-MIR	474	0.097	0.020	0.420
Fat-protein-ratio	474	1.352	0.629	2.407

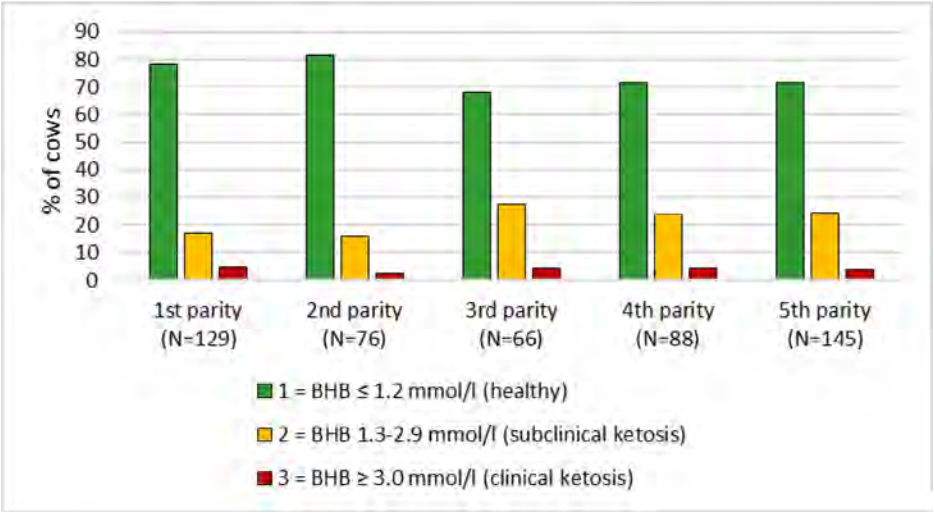


Figure 1. Percentage of healthy cows (BHB≤1.2 mmol/l), cows with subclinical ketosis (BHB=1.3-2.9 mmol/l) and clinical ketosis (BHB≥3.0 mmol/l) detected by the handheld device.

Heritabilities and genetic correlations for the analysed traits are shown in Table 2. Heritabilities for all traits were high, 0.26 for BHB results gained by the handheld device and between 0.16 and 0.30 for MIR predicted traits. The genetic correlations between BHB concentration from the handheld device and the traits predicted by MIR ranged from 0.60 to 0.73.

Discussion

Overall, the heritability estimates found in our study were consistent with previous literature results. Hamann *et al.* (2017) found heritabilities in the range of 0.22 to 0.39 for KetoMIR. Belay *et al.* (2017) reported heritability estimates for predicted blood BHB at different stages of lactation, ranging from 0.250 to 0.365. Benedet *et al.* (2020) found a heritability of 0.21 for BHB-MIR from 5 to 35 DIM. Genetic correlations between BHB results gained by a handheld device and MIR predicted traits could not be found in the literature. However genetic correlations with clinical ketosis from veterinary diagnoses were available. A moderate genetic correlation of 0.47 between predicted blood BHB at first test day and clinical ketosis was reported by Belay *et al.* (2017). Because the standard errors of the estimates in our study were high, the results should be taken with caution. However, the moderate to high genetic correlations between MIR-predicted traits and subclinical ketosis suggest that consideration of these traits would reduce subclinical ketosis.

Table 2. Heritabilities (on diagonal) and genetic correlations (above diagonal) with standard errors in brackets.

	BHB – handheld device	KetoMIR	BHB-MIR	Aceton-MIR	Fat-protein-ratio
BHB – handheld device	0.26	0.61 (0.35)	0.60 (0.32)	0.73 (0.33)	0.60 (0.37)
KetoMIR		0.30	0.48 (0.37)	0.35 (0.39)	0.20 (0.45)
BHB-MIR			0.19	n.c.[1]	0.72 (0.32)
Aceton-MIR				0.16	0.61 (0.38)
Fat-protein-ratio					0.20

[1] n.c. = no convergence.

Acknowledgements

This work was conducted within the COMET-Project D4Dairy (Digitalisation, Data integration, Detection and Decision support in Dairying, Project number: 872039) that is supported by BMK, BMDW and the provinces of Lower Austria and Vienna in the framework of COMET-Compentence Centers for Excellent Technologies. The COMET program is handled by the FFG, grant number 872039.

References

Belay T.K., Svendsen M., Kowalski Z.M., and Ådnøy T. (2017) J. Dairy Sci. 100: 6298-6311. https://doi.org/10.3168/jds.2016-12458.

Benedet A., Costa A., De Marchi M., and Penasa M. (2020) J. Dairy Sci. 103: 6354-6363. https://doi.org/10.3168/jds.2019-17916.

Dale L. M., and Werner A. (2018) Personal communication.

Dardenne P., Fernandez Pierna J. A., Abbas O., Baeten V., and Lecler, B. (2015) Proc. Final Optimir Scientific and Expert Meeting: From Milk Analysis to Advisory Tools, Namur, Belgium.

Hamann H., Werner A., Dale L., and Herold P. (2017) Proc. of the 41st ICAR Conference, Edinburgh, United Kingdom.

Khol J. L., Freigassner K., Stanitznig A., Tichy A., and Wittek T. (2019) Polish Journal of Veterinary Sciences 22: 557-564. https://doi.org/10.24425/pjvs.2019.129964

Werner A., Gollé-Leidreiter F., Droessler K., Auer F. J., Mayerhofer M., *et al.* (2019) Proc. of the 43rd ICAR Conference, Prague, Czech Republic.

ZuchtData (2020) ZuchtData Annual Report, 2019.

ZAR (2016) Efficient Cow – Final report on research project no. 100861. Available at: https://dafne.at/projekte/efficient-cow.

105. Heritability of scurs in Wagyu and tropically adapted composite beef cattle

G. Moser* and M.J. Kelly

Australian Agricultural Company, 76 Skyring Terrace, QLD, 4006, Australia; gmoser@aaco.com.au

Abstract

Breeding genetically hornless cattle provides an alternative to the invasive process of dehorning. Accurate genomic tests for poll status are available but the understanding of scurs genetics remains limited. Scurs status of two breed types was recorded in a large breeding operation in northern Australia (n=1,599). Estimates of SNP-based heritability ranged from 0.26-0.77 depending on breed, sex, and scurs phenotype. Several genome regions were significantly associated with scurs; the largest effects mapped to BTA1 and BTA12. Our findings confirm a complex, highly polygenic genetic architecture underlying scurs, modified by age and sex.

Introduction

Increasing frequency of the polled trait has positive implications for animal welfare, productivity, labour requirements and safety of station staff. The poll genetic test is a highly effective tool to increase the frequency of polled animals, with 99% accuracy in most cattle breeds (Randhawa *et al.*, 2020).

A major limitation of the poll genetic test is that it is not predictive of scurs, which can arise when only one copy of the poll gene is present. Scurs are loosely attached hornlike formations, which are usually smaller than true horns and develop slightly later in life (Grobler *et al.*, 2021). Scurs are not necessarily associated with the same issues as horns, so their removal often constitutes an unnecessary dehorning procedure.

Several studies, albeit with relatively small sample sizes, suggest that scurs are heritable, but genetically complex and affected by poll status, sex and breed (Aldersey *et al.*, 2020; Gehrke *et al.*, 2020; Grobler *et al.*, 2021). In the current study, we have utilised a greater sample size to further our understanding of the genetic architecture of scurs.

Materials & methods

SNPs and genotypes. DNA samples were collected at branding and genotyped with either GeneSeek® Genomic Profiler™ 100K or 50K arrays by Neogen Australasia. A total of 675,476 SNPs remained after imputation to high-density using BEAGLE (Browning and Browning, 2007) and QC. Poll test results are reported as P_cP_c, P_cP_f, P_fP_f for homozygous poll; HP_c, HP_f for heterozygous poll and *HH* for horned, where the specific poll allele is included as a subscript (Celtic (P_c), Friesian (P_f)).

Samples and phenotypes. Phenotypes were recorded for two groups of beef cattle verified by DNA testing as heterozygous polled (total n=1,599), with 10% being Wagyu and 90% Mitchell. Mitchell is a tropically adapted composite breed developed from Brahman, Shorthorn, Charolais, Senepol, Belmont Red and Santa Gertrudis in the 1980s. All cattle were assessed as horned, polled or scurred by station staff during routine handling at branding (2-6 months of age).

Males were excluded from routine dehorning during branding and later reassessed and scored for scurs type between 12 to 16 months of age. Observed phenotypes were scored into five categories using a modified version of the scoring system proposed by Gehrke *et al.* (2020). Polled animals are assigned to category 1 and scurred animals to category 2 to 5 depending on the type and size of scurs (Figure 1). For analysis, the

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_105, © G. Moser and M.J. Kelly 2022

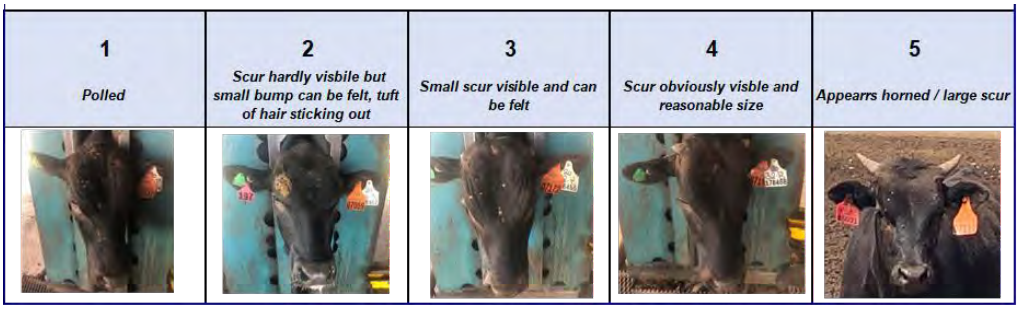


Figure 1. Scoring system for poll phenotype.

coded horn phenotypes were converted into two binary traits. In coding model BC1 only poll animals are counted as cases, in BC2 animals in categories 1 and 2 are combined and categories 3, 4, and 5 make up the controls (Table 1).

Females were assessed once at branding and phenotype recorded as either 'polled' or not Using this routinely collected data, a binary poll/scurs phenotype was constructed for heterozygous animals, with females not recorded polled being classified scurs. Heterozygous animals can only be polled or scurred given the high accuracy of the DNA test. Given manual palpation of the head is generally not performed by station staff at branding, small scurs (Figure 1, score 2) would mostly be recorded as polled, corresponding to coding BC2 for males. This is supported by an observation made in Mitchell males, where 213 bulls recorded as polled at branding presented with scurs as yearlings.

Statistical methods. We estimated genomic heritability and performed association analyses between SNPs and scurs phenotypes. Estimation of SNP based heritability of scurs traits used genome-based restricted maximum likelihood (GREML) implemented in the software WOMBAT (Meyer, 2007). To scan the genome for genetic variants associated with scurs we performed a genome-wide association study (GWAS) in PLINK (Chang *et al.*, 2015), treating BC1 and BC2 as binary phenotypes. To account for the genetic diversity within the Mitchell breed and to correct for population stratification effects in GWAS we fitted genomic breed composition (Boerner and Wittenburg, 2018) as covariates.

Results

The phenotypic variance explained by all SNPs using animals carrying the P_c allele are shown in Table 2. Estimates were moderate to high, and higher for BC2 compared to BC1. All estimates had low precision as shown by the large standard error. In Mitchell, SNPs explained more of the phenotypic variance in males than in females. Using a bivariate GREML we estimated a genetic correlation of 0.64 (0.18) between male and female BC2 phenotypes. Including animals carrying the P_f allele gave very similar results to those shown in Tab 2.

Table 1. Number of HP_c (HP_f in parenthesis) animals for coding BC1 and BC2.

		BC1		BC2	
Breed	**Sex**	**Cases**	**Controls**	**Cases**	**Controls**
Mitchell	Female	-	-	890 (70)	150 (4)
	Male	60 (8)	247 (17)	212 (22)	95 (3)
Wagyu	Male	15 (1)	123 (14)	54 (5)	84 (10)

Table 2. SNP based heritability (standard errors in parenthesis) for coding BC1 and BC2 estimated using HP_c animals.

Breed	Trait	Female	Male	Female + Male
Mitchell	BC1		0.26 (0.17)	
Wagyu			0.33 (0.24)	
Mitchell+Wagyu			0.42 (0.15)	
Mitchell	BC2	0.41 (0.07)	0.77 (0.19)	0.36 (0.06)
Wagyu			0.60 (0.26)	
Mitchell+Wagyu			0.56 (0.14)	0.35 (0.06)

For convenience, we use *P*-value thresholds commonly applied in human studies of 10^{-5} and 5×10^{-8} for suggestive significant and genome-wide significant associations (Visscher *et al.*, 2017). Across analyses, suggestive associations were found on many chromosomes and genome-wide significant peaks were mapped to seven different chromosomes (Figure 2).

Discussion

The presence of scurs was found to be moderately to highly heritable whether coded BC1 or BC2. In contrast to the findings of Gehrke *et al.* (2020), heritability was higher for BC2 compared to BC1. Nevertheless, both studies together strongly suggest that scurs are heritable and can be changed through selection. Gehrke *et al.* (2020) found that the P_f allele suppressed scurs more than the P_c allele. Despite only 10% of animals carrying the P_f allele, we only presented results for P_c animals to avoid potential problems caused

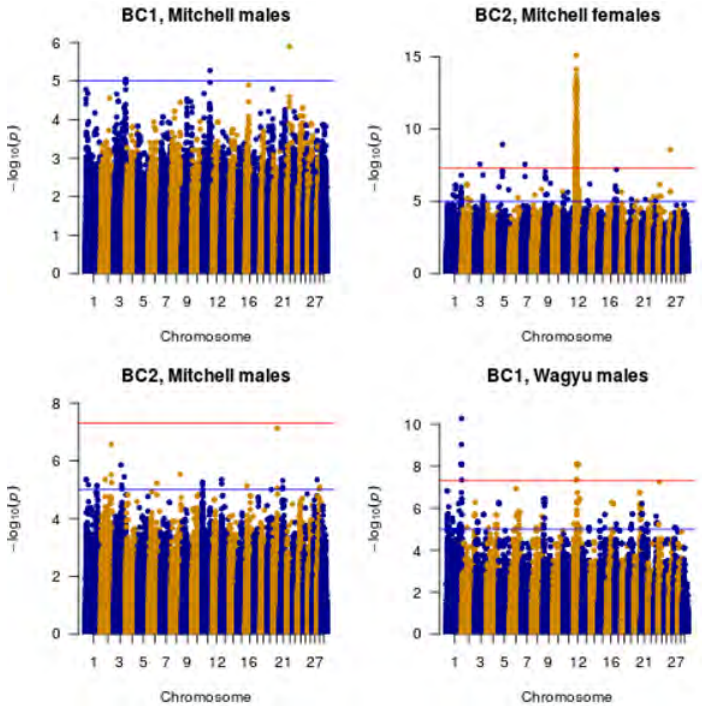


Figure 2. GWAS results for scurs with different phenotype codings. The horizontal lines indicate suggestive ($\alpha=10^{-5}$, blue) and genome-wide ($\alpha=5\times10^{-8}$, red) significance thresholds.

by allelic heterogeneity. Furthermore, confounding between founder breed and poll allele might be present, particularly in Wagyu where a very limited number of poll carriers were used for introgression into the naturally horned Wagyu breed.

The results of the current study do not support regulation of scurs development by a single or a few genes. However, there are several studies that reported putative QTL regions involved in the development of scurs (reviewed in Aldersey *et al.*, 2020; Grobler *et al.*, 2021). A recent study of Holstein-Friesian cattle described three QTL on BTA12 (Gehrke *et al.* 2020). We also found associations in Mitchell females and Wagyu males on BTA12, however, the QTL were mapped to different locations. In addition, two regions distal to the poll locus on BTA1 were detected for both phenotype codings in both breeds and sexes. Although evidence exists supporting a small number of QTL regions, together these QTL regions only explain a small component of the genetic variation in scur presence.

In this study we investigated scurs phenotypes in polled Wagyu and tropical adapted composite males and females. Heritability estimates indicate that reducing the frequency of scurs through selection is feasible. Our findings confirm a complex, highly polygenic genetic architecture underlying poll phenotypes, presumably modified by age, sex and other environmental factors. Several QTLs have been identified, however, our ability to further investigate putative regions is limited by the small sample size of our study.

References

Aldersey, J. E., Sonstegard, T. S., Williams, J. L., & Bottema, C. D. K. (2020). Anim Genet 51:166–176. https://doi.org/10.1111/age.12915

Boerner V., and Wittenburg D (2018) Frontiers in Genetics 9:185. https://doi.org/10.3389/fgene.2018.00185.

Browning S.R., and Browning B.L. (2007) Am J Hum Genet 81:1084-1097. https://doi.org/10.1086/5219

Chang C.C, Chow C.C., Tellier L.C.A.M., Vattikuti S., Purcell S.M., and Lee J.J. (2015) GigaScience 4. https://10.1186/s13742-015-0047-8

Gehrke, L.J., Capitan A., Scheper C., König S., Upadhyay M., *et al.* (2020) Genet Sel Evol 52(6):1–13. https://doi.org/10.1186/s12711-020-0525-z

Grobler, R., van Marle-Köster, E., and Visser, C. (2021) Livestock Science, 247, 1-7. https://doi.org/10.1016/j.livsci.2021.104479

Meyer K. (2007) Science B 8: 815–821. https://doi.org/10.1631/jzus.2007.B0815

Randhawa I.A.S., Burn B.M., McGowan M.R., Porto-Neto L.R., Hayes B.J., *et al.* (2020) G3 Genes|Genomes|Genetics 10(2):539–544.https://doi.org/10.1534/g3.119.400866

Visscher P.M., Wray N.R., Zhang Q., Slar P.M., McCarty M.C. et al (2017) Am J Hum Genet 101(1):5-22. https://doi.org/10.1016/j.ajhg.2017.06.005

106. Genetic parameters for social hierarchy of Nellore cattle using data from electronic feeding system

J.P.S. Valente[1*], M. Deniz[1], K.T. de Sousa[1], R.C. Canesin[2], S.F.M. Bonilha[2], M.E.Z. Mercadante[2] and L.T. Dias[1]

[1]*Sector of Agricultural Sciences, Federal University of Paraná, 80035-050, Curitiba, Brazil;*
[2]*Beef Cattle Research Center, Institute of Animal Science, 14174-000, Sertãozinho-SP, Brazil;*
juliadepaulasoaresvalente@gmail.com

Abstract

Genetic parameters for traits related to social hierarchy of 970 Nellore young bulls (282±39 days old), from feed event data collected by electronic feeding system, were estimated. Social hierarchy was evaluated from three traits: dominance value (continuous), social rank (binomial), and social hierarchy (three categories). The heritability estimates were moderate (0.23 to 0.29). Genetic correlation (r_g) between social hierarchy, and dry matter intake per visit and time at the feeder per visit were negative (-0.28 to -0.47), while r_g with the number of visits at the feeder were positive (0.21 to 0.34). In general, the r_g between social hierarchy and growth traits were close to zero, suggesting that selection for growth does not affect the social hierarchy. In conclusion, there are genetic effects on the expression of social hierarchy traits. Furthermore, the social hierarchy is genetically related to feeding behaviour, but not to growth traits.

Introduction

The social relationships developed by cattle result in the establishment of social hierarchy within the group (Phillips and Rind, 2002). One of the main functions of social relationships is to regulate access to resources such as food (Bica *et al.*, 2019; de Sousa *et al.*, 2021), water and shade (Deniz *et al.*, 2021).

Understanding the social dynamics of animals is important to establish effective management at the farms and to comprehend how productive traits are affected by animals' behaviour. However, in the genetic field, studies about the social organization of cattle are scarce due to the difficulty to obtain the social behaviour data. Furthermore, most studies use direct visual observation to determine the social hierarchy (Bica *et al.*, 2019; de Sousa *et al.*, 2021; Deniz *et al.*, 2021).

With technology advancement, studies have shown the feasibility of determining the social hierarchy from feed events registered by electronic feeding system (Huzzey *et al.*, 2014; McDonald *et al.*, 2019). A short time interval between the visits of two different animals at the same feeder can be used to establish dominance relationship among animals. This study aimed to estimate genetic parameters for dominance value, social rank, and social hierarchy of Nellore cattle from feed event data collected by electronic feeding system.

Materials & methods

Animals. Data of 970 Nellore young bulls, born from 2012 to 2020, were obtained in 20 feed efficiency tests (282±39 days old at the beginning of the test), conducted at the Beef Cattle Research Center, Institute of Animal Science (IZ), Sertãozinho-SP, Brazil.

Housing. The feed efficiency test lasted 84±12 days. The groups of animals were housed in one of the three collective pens available at the IZ: Pen 1 (1,844 m^2): one common water feeder and five GrowSafe System (Vytelle – Kansas City, Missouri, United States of America) electronic feeder; Pen 2 (3,687 m^2): one common water feeder and 10 GrowSafe System electronic feeder; and Pen 3 (3,020 m^2): four Intergado®

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_106, © J.P.S. Valente *et al.* 2022

(Contagem, Minas Gerais, Brazil) water electronic feeders and 12 Intergado® electronic feeder. The feeders had physical barriers that allowed access to only one animal at a time. The animals received an electronic ear tag for individual recognition at the feeder. The diet was offered twice a day (8:00 am and 3:00 pm), and the animals had *ad libitum* access to food, water, and mineral salt.

Social hierarchy determination. The method to determine the social hierarchy of the animals in each feed efficiency test group (48±19 animals per group) was based on the replacements that occurred at the electronic feeders (Foris *et al.*, 2019). For the analyses, the feeding event data collected by the electronic troughs during the feed efficiency test were used to estimate one dominance value for each animal. The interval of up to 10 seconds between the exit of one animal and the entry of another animal in the same feeder was considered to determine a replacement since longer intervals can increase the probability of false-positive results (Foris *et al.*, 2019). For each group, two interspersed days per week were selected until the end of the test in order to minimize the number of inconsistencies (De Vries, 1998). On average, 22 test days were used for each group. In addition, only data collected during the periods after feeding were used: three hours in the morning (07:00:00 am to 09:59:59 am) and three hours in the afternoon (02:00:00 pm to 4:59:59 pm). Such periods were selected because the greatest number of replacements between animals occurs following the fresh food delivery (DeVries and Von Keyserlingk, 2005; Val-Laillet *et al.*, 2008). Weeks without information due to technical failures in the electronic feeder were discarded.

The social hierarchy was determined based on the equations proposed by Kondo and Hurnik (1990) from 'socialh' R package (Valente *et al.*, 2021). The dominance value (DV) for each animal was calculated as the sum of all relationships of this animal with each other animal within the group throughout the days. The social rank (SR) [high (1) and low (2)] and the social hierarchy (SH) [subordinate (1), intermediate (2), and dominant (3)] were obtained from the dominance value, as described by Coimbra *et al.* (2012).

Statistical analysis. The heritability estimates of DV, SR and SH were obtained by one-trait analysis, and the genetic correlations estimates were obtained from two-traits analyses with the following traits: average dry matter intake per feeder visit (DMI_v; kg/visit, n=1,003), average time spent at the feeder per visit (TF_v; min/visit, n=1,003), average frequency of visits to the feeder (FF, visits/day, n=1,003), weight at the beginning of the test (BW, kg, n=1,781), average weight throughout the test (AW, kg, n=1,781), weight at the end of test (FW, kg, n=1,781), postweaning weight (PW, kg, n=9,936), and average daily gain (ADG, kg/day, n=1,781). The model can be written in matrix form as follows: $y = X\beta + Z\alpha + e$, where y is the vector of the traits observed; β is the vector of fixed effects [test group (or contemporary group for PW) and the linear effect of animal age at the beginning of the test and the linear and quadratic effects of the dam age as covariates]; α is the vector of direct additive genetic effects; X and Z are incidence matrices for each effect, respectively, and e is the vector of residual effects. It was assumed that $E(y) = X\beta$, $E(a) = 0$, $E(e) = 0$ and $Var(\alpha) = A \otimes S_a$ and $Var(e) = I \otimes S_e$, where S_a is the matrix of genetic (co)variance between social hierarchy, feeding behaviour and growth traits; S_e is the residual (co)variance matrix, A is the pedigree-based relationship matrix (12,228 animals, five generations back), and I is the identity matrix.

The (co)variance components for DV trait were estimated fitting linear models by the AIREML software (Misztal *et al.*, 2021), while the (co)variance components for categorical traits (SR, SH) were estimated fitting threshold models by the THRGIBBS and POSTGIBBS software (Misztal *et al.*, 2021) (500,000 interactions; 50,000 cycles of burn-in; sample every 20 interactions; resulting in 22,500 effective samples).

Results & discussion

Heritability estimates for DV, SR and SH were moderate (Table 1). The values indicate that these traits would respond to direct selection. Despite the relevant presence of the environmental effect on the traits, it is interesting to note the existence of the direct genetic effects on DV, SR and SH which should be considered in the evaluation of animals. Previous studies in the literature presented heritability estimates for dominance value of dairy cows: 0.40 for heifers (Beilharz *et al.*, 1966) and 0.15 for lactating cows (Dickson *et al.*, 1970). These results differ from the present study probably due to the breed and the huge difference between beef and dairy cattle.

Genetic correlation estimates (r_g) between social hierarchy traits (DV, SR and SH), and feeding behaviour traits (DMI_v, TF_v and FF), and growth traits (BW, AW, FW, PW, and ADG) are presented in Table 2. The r_g of DV, SR, SH with DMIv and TFv was negative and moderate, while the r_g with FF was positive. These results suggest that animals at high social positions get access to the feeder more frequently, spend less time, and eat less by feed event when compared to animals at low social positions. On the other hand, selection for growth seems to have no effect on the animal social hierarchy. The r_g between DV, SR, and BW, AW, FW, PW, and ADG indicated a low genetic association among social hierarchy and growth traits. However, the r_g estimates presented high standard deviations or standard errors as a consequence of small number of animals with social hierarchy information, so more studies are necessary to confirm these genetic associations.

Table 1. Estimates of the additive genetic variance (σ_a^2), residual variance ($\sigma_e^2$and heritability (h^2) for dominance value (DV), social rank (SR) and social hierarchy (SH) in Nellore cattle.

Trait	Mean ± SD	σ_a^2	σ_e^2	H
DV	-0.06±9.98	23.08 (7.59)	78.60 (6.97)	0.23 (0.07)
SR[1]	1.48±0.50	0.45 (0.22)	1.02 (0.66)	0.29 (0.10)
SH[1]	1.99±0.68	0.11 (0.04)	0.34 (0.04)	0.25 (0.08)

[1] For SR and SH, σ_a^2, σ_e^2, and h^2 are posterior mean (posterior standard deviation).

Table 2. Genetic correlations between dominance value (DV), social rank (SR) and social hierarchy (SH), with feeding behaviour and growth traits in Nellore cattle.[1]

Trait	Mean ± SD	Genetic correlation		
		DV	SR[2]	SH[2]
DMI_v	0.11±0.04	-0.37 (0.17)	-0.36 (0.17)	-0.28 (0.16)
TF_v	00:01:17±00:00:37	-0.47 (0.16)	-0.44 (0.16)	-0.46 (0.15)
FF	88±24	0.34 (0.18)	0.27 (0.19)	0.21 (0.19)
BW	245±49	0.09 (0.16)	0.03 (0.13)	0.31 (0.13)
AW	290±53	0.07 (0.15)	0.03 (0.13)	0.25 (0.12)
FW	332±59	0.06 (0.15)	0.04 (0.12)	0.20 (0.12)
PW	303±53	-0.08 (0.14)	-0.03 (0.12)	-0.02 (0.12)
ADG	0.00±0.62	-0.03 (0.17)	-0.06 (0.14)	-0.09 (0.14)

[1] DMI_v (kg/visit): dry matter intake per feeder visit; TF_v (min/visit): time spent at the feeder per visit; FF (visits/day): frequency of visits to the feeder; BW (kg): weight at the beginning of the test; AW (kg): average weight throughout the test; FW (kg): weight at the end of the test; and PW (kg): postweaning weight; ADG (kg/day): average daily gain.

[2] For SR and SH, genetic correlations are posterior mean (posterior standard deviation).

Conclusions

As a preliminary study conducted with Nellore cattle, the results indicated the presence of genetic effects on the expression of social hierarchy traits. Social hierarchy is genetically associated with feeding behaviour traits but it is not genetically associated with growth traits. The results found for social hierarchy will help to understand how this trait interferes in the expression of productive traits. In the future, information regarding social hierarchy can probably be included in beef cattle breeding programs.

Acknowledgements

JPSV, MD, and KTS acknowledge CAPES (Coordination for the Improvement of High Education Personnel, Brazil- Finance Code 001) for the scholarship. Also, we thank the São Paulo Research Foundation (FAPESP, grant #2017/10630-2 and grant #2017/50339-5).

References

Beilharz, R.G., Butcher, D.F., Freeman, A.E. (1966) J. Dairy Sci 49:887–892. https://doi.org/10.3168/jds.S0022-0302(66)87964-X

Bica, G.S., Teixeira, D.L., Hötzel, M.J., Machado Filho, L.C.P. (2019) Appl. Anim. Behav. Sci – 220. https://doi.org/10.1016/j.applanim.2019.104852

Coimbra, P.A.D., Machado Filho, L.C.P., Hötzel, M.J. (2012) Appl. Anim. Behav. Sci. 139:175–182. https://doi.org/10.1016/j.applanim.2012.04.009

De Vries, H. (1998) Anim. Behav. 55:827–843. https://doi.org/10.1006/anbe.1997.0708

Deniz, M., de Sousa, K.T., Moro, M.F., do Vale, M.M., Dittrich, J.R. *et al.* (2021) Appl. Anim. Behav. Sci. 244, 105467. https://doi.org/10.1016/j.applanim.2021.105467

De Sousa, K.T., Machado Filho, L.C.P., Bica, G.S., Deniz, M., Hötzel, M.J. (2021) Appl. Anim. Behav. Sci. 234, 105172. https://doi.org/10.1016/j.applanim.2020.105172

DeVries, T.J., Von Keyserlingk, M.A.G. (2005) J. Dairy Sci 88:625–631. https://doi.org/10.3168/jds.S0022-0302(05)72726-0

Dickson, D.P., Barr, G.R., Johnson, L.P., Wieckert, D.A. (1970) J. Dairy Sci 53:904–907. https://doi.org/10.3168/jds.S0022-0302(70)86316-0

Foris, B., Thompson, A.J., von Keyserlingk, M.A.G., Melzer, N., Weary, D.M. (2019) J. Dairy Sci. 102: 9176–9186. https://doi.org/10.3168/jds.2019-16697

Huzzey, J.M., Weary, D.M., Tiau, B.Y.F., Von Keyserlingk, M.A.G. (2014) J. Dairy Sci. 97:2953–2958. https://doi.org/10.3168/jds.2013-7434

Kondo, S., Hurnik, J.F. (1990) Appl. Anim. Behav. Sci. 27:287–297. https://doi.org/10.1016/0168-1591(90)90125-W

McDonald, P. V., von Keyserlingk, M.A.G., Weary, D.M. (2019) J. Dairy Sci. 102:3495–3500. https://doi.org/10.3168/jds.2018-15585

Misztal, I., Tsuruta, S., Lourenco, D.A.L., Aguilar, I., Legarra, A. *et al.* (2014) Manual for BLUPF90 family of programs. Available at: http://nce.ads.uga.edu/wiki/lib/exe/fetch.php?media=blupf90_all2.pdf

Phillips, C.J.C., Rind, M.I. (2002) J. Dairy Sci. 85:51–59. https://doi.org/10.3168/jds.S0022-0302(02)74052-6

Valente, J.P.S, Deniz, M. and de Sousa, K.T. (2021) socialh: Rank and Social Hierarchy for Gregarious Animals. Available at: https://CRAN.R-project.org/package=socialh

Val-Laillet, D., Passillé, A.M. de, Rushen, J., von Keyserlingk, M.A.G. (2008) Appl. Anim. Behav. Sci. 111:158–172. https://doi.org/10.1016/j.applanim.2007.06.001

107. Pigs' weighing order stability is favourably linked with tail biting and production: a new proxy for social structure?

W. Gorssen*, R. Meyermans, S. Janssens and N. Buys

KU Leuven, Department of Biosystems, Center for Animal Breeding and Genetics, Kasteelpark Arenberg 30 - bus 2472, 3001 Leuven, Belgium; wim.gorssen@kuleuven.be

Abstract

In this study, we weighed fattening pigs pen-by-pen every two weeks to study their weighing order as a proxy for social group structure. Hereby, pigs were allowed to choose the order of weighing themselves. We investigated the genetics of the stability of the weighing order and the (genetic) relationship with production and welfare traits. In total, 15,435 recordings from 1,864 crossbred pigs (149 pens) with known pedigree were analysed. Weighing order stability increased over consecutive weight recordings, with correlations for consecutive recordings increasing from r=0.11 to r=0.44. Weighing order stability was lowly heritable (h^2=1.6-5.2%). However, a low weighing order stability was (genetically) correlated with lower average daily gain and lean meat percentage and with more tail/ear biting, lameness and mortality. These results indicate that social structure in fattening pigs can possibly be estimated by weighing order stability, and that a more stable weighing order is favourably correlated with production and welfare traits.

Introduction

Pigs are highly social animals, and hence they form complex relationships. In nature, pigs live in groups based on matriarchal hierarchies, whereas in modern pig farming, pigs from different families are often grouped in pens (Rydhmer and Canario, 2014). Creating new groups of pigs often results in aggression to assess a new social hierarchy, possibly leading to injuries such as ear-biting and tail-biting (Rydhmer and Canario, 2014). Moreover, the emerging fields of social and behavioural genetics have shown that social effects in pigs are heritable (Rydhmer and Canario, 2014; Cheng *et al.*, 2009), and can be exploited in breeding programs to improve overall feed efficiency (Bergsma *et al.*, 2008) or to reduce aggression and biting behaviour (Camerlink *et al.*, 2013). However, we still have little understanding about the importance of social stability and the relation to social behaviour and production traits (Camerlink *et al.*, 2021).

In this study, individual fattening pigs were weighed pen-by-pen and scored every two weeks on a commercial farm. We hypothesize that the weighing order is related to pigs' social behavior, and/or that the stability of these weighing orders is (partly) under genetic control. Moreover, we explored the (genetic) relationship between pigs' weighing order and important traits for pig production.

Materials & methods

Study population and data collection. Data were recorded between July 2020 and July 2021 at a progeny testing farm for the breeding organization 'Vlaamse Piétrain Fokkerij' (Belgium). Pigs were kept in commercial conditions and where housed in seventeen identical compartments of eight pens (4.0×2.5 m) with on average thirteen pigs per pen. Pens were of mixed sex (surgically castrated males and females) and penmates descended from the same Piétrain sire. Pigs were identified with radio frequency identification (RFID) ear tags.

Pigs were weighed individually for the first time nine days after arrival at the fattening farm (average age 83.1±5.6 days; average weight 29.9±6.5 kg). Hereafter, pigs were weighed every two weeks. At the last (eight) recording, pigs had an average age of 172.8 days (SD=7.6) and weight of 104.9 kg (SD=13.4). Pen-by-pen, all individuals were driven to the stable's central hallway, after which pigs were weighed one after

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_107, © W. Gorssen *et al.* 2022

the other. Per pen, the pigs' weighing order was determined based on time of weighing: the first weighed pig of a pen was scored with '1', the second with '2', etc. Moreover, ear swellings (hematomas) and ear biting victims were scored (0=none, 1=left or right, 2=left and right) as well as tail biting victims (0=none, 1=small scratches, 3=bloody and/or infected tail). Mortality (0=alive, 1=dead) and lameness (0=normal, 1=lame) were recorded as well. All recordings were collected by the first author. Lean meat percentage was individually recorded at the slaughterhouse of the Belgian Pork Group using AutoFom III™ (Meer, Belgium). The final dataset consisted of 15,435 recordings from 1,864 pigs (149 pens) originating from 135 Piétrain sires and 266 crossbred dams with known pedigree.

Animal samplings were conducted following the procedures approved by the ethical committee of KU Leuven (P004/2020).

Analysis. Pigs' weighing orders were first rescaled on a pen level from '0' (first) to '100' (last) to account for the number of pigs in each pen (varying from 8 to 15). Hereafter, weighing order stability was estimated by the standard deviation of weighing orders per pig over all eight recordings (SD-order$^{1\text{-}8}$) and over the last five recordings (SD-order$^{4\text{-}8}$), since weighing order seemed to stabilize in general after the first three weighings (Table 1). These standard deviations were used as new phenotypes. For mortality, lameness, hematomas and ear and tail biting victims the maximum observed value over all recordings was retained.

Heritabilities (h^2) and genetic correlations (r_g) were estimated using single or bivariate animal models with average information REML, implemented in airemlf90 and invoked with the R-package breedR (Muñoz and Sanchez, 2019). All models were of the form:

$$y= Xb + Za + Wc + e \quad (1)$$

Where *y* is the vector with phenotypes for the studied trait(s); *b* is the vector containing the fixed effects (sex, 2 levels; parity of dam, 4 levels) and covariates (age last recording); *a* is the vector of additive genetic effects (4,989 levels); *c* is the vector of random environmental effects (17 levels); *e* is the vector of residual effects; *X*, *Z* and *W* are incidence matrices for fixed effects, random animal effects and random permanent environmental effects. The random environmental effect *c* is a combination of farm and date of entrance at farm.

Results

In general, the stability of pigs' weighing order increased over time (Table 1). Spearman correlations of consecutive recordings ranged from r=0.11 to r=0.44. The average correlation of all pairwise combinations for all pens was r=0.21, although large differences were observed between pens from r=-0.05 (low stability) to r=0.50 (high stability). After rescaling, standard deviations of weighing orders were approximately normally distributed (μ=27.0±7.0), ranging between 0.0 and 50.0.

Heritability estimates for weighing order stability were very low: h^2 of SD-order$^{1\text{-}8}$ was estimated at 1.6% (SE=2.4), whereas SD-order$^{4\text{-}8}$ had an estimated h^2 of 5.2% (SE=3.0). However, low to moderate genetic correlations were found (Table 2) between weighing order stability and ear biting scores (r_g=0.22-0.27) and tail biting scores (r_g=0.37-0.40), although standard errors were high (Table 2). Average daily gain and lean meat percentage had low to moderate negative r_g estimates (r=-0.14 to r=-0.41). For mortality and lameness, there was a phenotypic relationship with SD-order$^{4\text{-}8}$ (r=0.15 and r =0.23), but standard errors were too high for these traits to estimate r_g.

Table 1. Pairwise Spearman correlations of pigs' weighing orders over eight recordings. For consecutive weight recordings, the Spearman correlation increases from r=0.11 to r=0.44, indicating the weighing orders of pigs within a pen stabilizes over time.

Order at weighing	2	3	4	5	6	7	8
1	0.11	0.06	0.05	0.04	0.01	0.02	0.06
2		0.31	0.24	0.17	0.15	0.14	0.10
3			0.40	0.28	0.21	0.18	0.12
4				0.39	0.30	0.24	0.20
5					0.40	0.33	0.28
6						0.40	0.35
7							0.44

Table 2. Phenotypic and genetic correlations between weighing order stability and other traits. Phenotypic (Pearson) correlations were estimated at a pen level: mean standard deviation of weighing order vs mean trait level. For lameness and mortality, standard errors of genetic correlations were >1 and hence results are not shown. ADG_{Youth}: average daily gain from birth until first recording. ADG_{Life}: average daily gain from birth until last recording. °: $P<0.10$,*: $P<0.05$.

	Phenotypic correlation		Genetic correlation	
Trait	SD-order$^{1-8}$	SD-order$^{4-8}$	SD-order$^{1-8}$ (se)	SD-order$^{4-8}$ (se)
ADG_{Youth}	0.06	-0.02	-0.41 (0.61)	-0.20 (0.36)
ADG_{Life}	0.02	-0.04	-0.32 (0.52)	-0.35 (0.46)
Meat percentage	-0.09	-0.07	-0.34 (0.83)	-0.14 (0.33)
Ear biting	0.12	0.15°	0.27 (0.53)	0.22 (0.16)
Tail biting	0.09	0.15°	0.40 (0.55)	0.37 (0.21)
Ear swelling	-0.02	-0.06	-0.22 (0.75)	-0.21 (0.34)
Lameness	0.23*	0.23*	NA	NA
Mortality	0.09	0.15°	NA	NA

Discussion

This study has shown that the weighing order of a pen of fattening pigs generally stabilizes over time, and that the stability of this weighing order is lowly heritable (h^2=1.6-5.2%). Moreover, the weighing order stability was (genetically) favourably correlated with biting behaviour, lameness, mortality and production traits, although standard errors of genetic correlations were high. These findings suggest that the weighing order stability might have potential to assess social stability and aggression in the pen. Different studies have found an association between social genetics and production, aggression, skin lesions and/or biting behaviour (Rydhmer and Canario, 2014; Bergsma *et al.*, 2008; Camerlink *et al.*, 2013, Angarita *et al.*, 2019). However, where most studies assume that pigs' social hierarchy stabilizes a few weeks after mixing groups, we found considerable differences between pens with regard to the stability of weighing order over time.

We would like to note that the specific weighing setup might be important to interpret (or replicate) these findings. In this study, pigs were weighed individually pen-by-pen and during the weighing process pigs had plenty of space in the central hallway, making it possible for them to (partly) determine the order of weighing, although pigs also interacted with the person performing the weight recordings. This new 'environment' might result in different social rankings than would be observed in the pen or possibly with a different weighing setup. Moreover, it is unclear whether pigs ranked first for weighing order would be

more dominant, as going first could also be an expression of fleeing the group. For future research, it would be interesting to relate the stability of weighing orders with pigs' observed behaviour and social rankings in the pen.

To conclude, our results show a relation between weighing order stability in fattening pigs and increased production and welfare. Since social behaviour is hard (or costly) to phenotype in pigs, the stability of weighing orders might be interesting for future research as a proxy for general social stability.

Acknowledgements

This study was partially funded by an FR PhD fellowship (1104320N) and an SB PhD fellowship (1S37119N) of the Research Foundation Flanders (FWO). We would like to acknowledge the Vlaamse Piétrain Fokkerij, Steven Broekx, Nathalie Cenens, Simon Cardinaels and Hannelore Vermoesen for their help.

References

Angarita, B. K., Cantet, R. J., Wurtz, K. E., O'Malley, C. I., Siegford, J. M., *et al.* (2019). Estimation of indirect social genetic effects for skin lesion count in group-housed pigs by quantifying behavioral interactions. Journal of animal science, 97(9):3658-3668. https://doi.org/10.1093/jas/skz322

Bergsma, R., Kanis, E., Knol, E. F., & Bijma, P. (2008). The contribution of social effects to heritable variation in finishing traits of domestic pigs (Sus scrofa). Genetics, 178(3):1559-1570. https://doi.org/10.1534/genetics.107.084236

Camerlink, I., Turner, S. P., Bijma, P., & Bolhuis, J. E. (2013). Indirect genetic effects and housing conditions in relation to aggressive behaviour in pigs. PloS one, 8(6):e65136. https://doi.org/10.1371/journal.pone.0065136

Camerlink, I., Proβegger, C., Kubala, D., Galunder, K., & Rault, J. L. (2021). Keeping littermates together instead of social mixing benefits pig social behaviour and growth post-weaning. Applied Animal Behaviour Science, 235:105230. https://doi.org/10.1016/j.applanim.2021.105230

Cheng, J., Janssens, S., & Buys, N. (2009). Full sib pens of pigs are not suitable to identify variance component of associative effect: a simulation study using Gibbs sampling. BMC genetics, 10(1), 1-9. https://doi.org/10.1186/1471-2156-10-9

Rodriguez, L. S., & Munoz, F. (2016). breedR: Statistical methods for forest genetic resources analysts. In IUFRO Genomics and Forest Tree Genetics. https://hal.inrae.fr/hal-02741136

Rydhmer, L., & Canario, L. (2014). Behavioral genetics in pigs and relations to welfare. In Genetics and the behavior of domestic animals (pp. 397-434). Academic Press. https://doi.org/10.1016/B978-0-12-394586-0.00011-1

108. Application of machine learning algorithms to develop behavioural and stress phenotypes for sheep

H. Keshavarzi*, D.L.M. Campbell and C. Lee

Agriculture and Foo, Commonwealth Scientific and Industrial Research Organization (CSIRO), Armidale, NSW, 2350, Australia; hamideh.keshavarzi@csiro.au

Abstract

This study aimed to evaluate the performance of Machine Learning (ML) in relating behavioural phenotypes to stress susceptibility in sheep. Hierarchical clustering was used to categorize 335 sheep for behavioural scores and relate these to cortisol levels in response to an acute stressor (shearing). The individuals were categorized as bold, calm, and fearful based on their behaviour in two behavioural tests, namely the arena and corridor tests. Venn diagrams were applied to plot the common animals between stress and behaviour clusters. Correlation between cortisol and behavioural traits was low (-0.16-0.20); and the Venn diagram did not show a high proportion of common individuals between behaviour and stress clusters. Application of ML is useful to characterise behavioural phenotypes but these were not related to variation shown in the more difficult to measure trait of stress susceptibility in sheep.

Introduction

The development of new technological devices has enabled researchers to collect large datasets on the movement (Keshavarzi *et al.*, 2020), social interactions (Krause *et al.*, 2013), and other behavioural traits such as vocalisations (Blumstein *et al.*, 2011) of individual animals. Challenges such as unknown nonlinear dependencies between variables and failure to meet the assumptions of many classical statistical models have made these datasets complex to analyse and describe (Valletta *et al.*, 2017). However, Machine Learning (ML) methods have been demonstrated to successfully classify samples based on unknown relationships and handle large datasets with missing values (Keshavarzi *et al.*, 2020) which overcomes some of these challenges.

A potential use of ML in animal science is clustering animals based on their behaviour, in which the method puts animals into groups that share common characteristics. This would enable the development of new behavioural phenotypes and these could be related to temperament, personality or resilience, e.g. a response to stress. As variation in behaviour is easier to measure than variation in response to stress, it may be possible to identify proxy behavioural measures for stress. Clustering methods have been widely used in animal behaviour studies to implement various tasks such as classifying species, vocalisations or behaviours or detecting emotional states in animals (Valletta *et al.*, 2017). The relationship between such behaviours with welfare indicators may be detectable with ML.

Previous studies have reported the association between cortisol as a physiological indicator of stress with some behavioural traits such as vocalisation, restraint (Grandin and Shivley, 2015), and fearfulness (Lee *et al.*, 2014). We, therefore, aimed to utilise the ability of ML to reveal the possible relationship between cortisol and behavioural traits in this study using sheep. Further understanding of how behavioural phenotypes relate to a physiological stress response using ML may enable more simple and improved welfare-related phenotypes to be developed.

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_108, © H. Keshavarzi *et al.* 2022

Materials & methods

Ethical approval of animal experimentation. The protocol and conduct of the study were approved by the Chiswick Armidale Animal Ethics Committee, under the New South Wales Animal Research Act 1985.

Animals and data. A total of 409 Merino × Romney sheep (ewes and wethers) were used in the study. Two weeks after weaning (3 months of age), fit and healthy lambs were individually subjected to behavioural tests (arena, and corridor) as described below.

Arena test. A motivational choice test that measures the conflict between social attraction and fear of a human. Briefly, individual sheep were placed in a 7×3 m outdoor arena for three minutes. The arena was divided into virtual zones. A human stood in front of a pen of four familiar flockmates at one end of the arena. Recorded behaviours included: zone score (quantified the total time spent near to the human and other sheep), the number of zones crossed, and the number of vocalisations.

Corridor test. The corridor test evaluates sheep's reaction to social isolation combined with repeated approaches by a human. The corridor consisted of a 0.6 m wide rectangular circuit (4.2×7.2×1.2 m) with an inner lining of rubber mesh. The circuit was virtually divided into six equal zones. The test consisted of three phases: (1) isolation- sheep left alone in the corridor for 30s; (2) pursuit- human entered the corridor and walked two circuits for 1 min; and (3) confinement- sheep placed into one arm of the corridor with a stationary human standing at one end for 30s. Video recorded behaviours included: the number of zones crossed (phase 1), the zones of the human and the animal, and the time the human could see the animal (phase 2), and zones entered and crossed (phase 3).

Cortisol. At 9 months of age, cortisol response to shearing was extracted by radioimmunoassay from blood samples collected 15 min post-shearing.

The recorded data were limited to animals with all information for all behavioural tests. After editing the data, the final dataset included 335 observations/individuals.

Clustering. Clustering of individuals was performed based on important traits identified from the behavioural tests. The hierarchical clustering algorithm was used to cluster animals into sub-classes. The analysis was implemented in R (R Core Team, 2018) using the function '*hclust*' with the Ward D method. The number of clusters was considered to be k=3. Based on the obtained results from arena and corridor tests (Figure 1), animals were classified as: fearful, bold or calm. Animals were also categorised into high (HC), medium (MC), and low cortisol (LC) based on plasma cortisol concentrations. The common animals between behavioural and stress clusters were then plotted using the package of 'ggVennDiagram' in R. Correlation between variables was also calculated in R (R Core Team, 2018).

Results

Figure 2 presents the number of shared animals between behavioural (fearful, bold, and calm) clusters based on the arena (first row) and corridor (second row) tests and their physiological response to stress (cortisol). All clusters of behavioural tests show overlap with different concentrations of plasma cortisol which indicates that there was no relationship between behaviour and cortisol of sheep in the current study.

Discussion

This study demonstrates the performance of machine learning to understand relationships between stress susceptibility and behaviour in sheep. Individuals were categorised into three groups based on behavioural tests, including fearful, bold, and calm animals. Animals were also grouped into three clusters based on

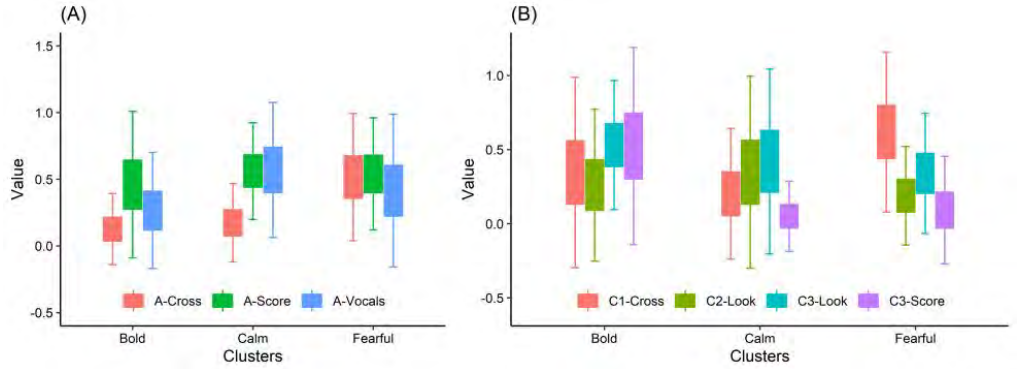


Figure 1. Results of sheep classification using Hierarchical clustering based on selected variables from arena (A) and corridor (B) behavioural tests. Individuals with the same colour belong to the same cluster.

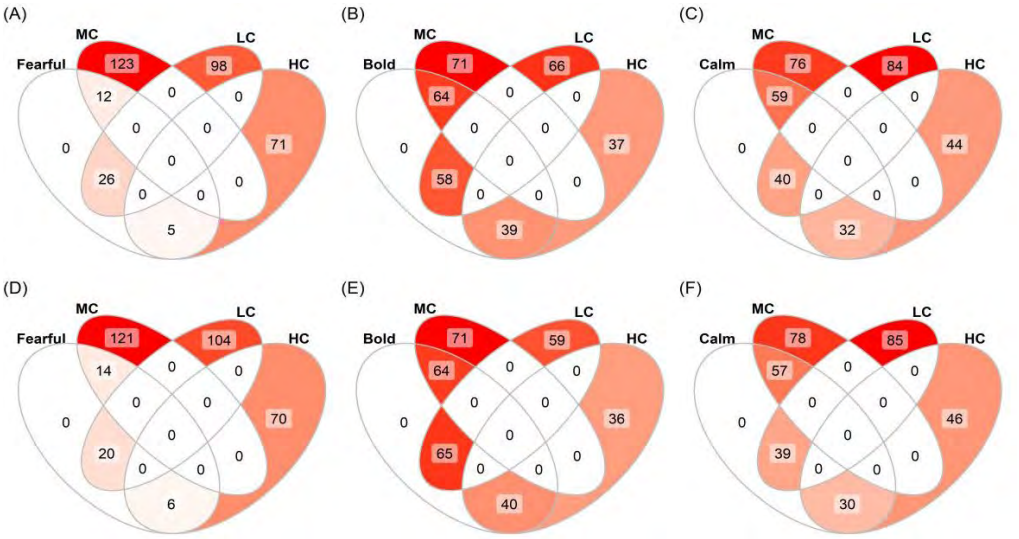


Figure 2. Venn diagram depicting the number of animals in each cluster based on behavioural tests (Arena (first row) and corridor (second row)) vs plasma cortisol, and their overlap. HC, MC, and LC represent high, medium, and low cortisol, respectively.

plasma cortisol responses to stress: high (HC), medium (MC), and low cortisol (LC) concentrations. Venn diagrams were then applied to plot the common animals between stress and behaviour clusters. It was predicted that the behavioural clusters would strongly relate to the stress responses of sheep with a high proportion of shared animals between desired sets (i.e. fearful with HC and bold with LC). This would enable a potential proxy non-invasive behavioural indicator for a more invasive trait (cortisol). However, this was not supported as there was little overlap between the behavioural and stress response clusters. This aligns with the low correlation (-0.16-0.20) between cortisol and behavioural parameters found in the current study. Previous studies have found relationships between physiological measures of stress and behaviour in sheep. For example, high cortisol ewes exhibited fewer bleats (Caroprese *et al.*, 2010) and vocalizations (Diess *et al.*, 2009) and low cortisol responses were associated with reduced fearfulness (Lee

et al., 2014). A possible explanation for the differences in findings may be the different tests and stress measures used. Further studies are needed to determine if other behavioural tests may relate more strongly to physiological stress responses.

Nevertheless, this study demonstrates how the application of ML can be used to gain new insights into the relationship between traits in addition to their potential use to predict (e.g. abortion incidence in dairy cattle (Keshavarzi *et al.*, 2020b)) and cluster individuals into sub-groups based on their behaviour (Teimouri *et al.*, 2018)). This may lead to more accurate behavioural phenotypes for improved welfare in the future.

References

Blumstein D.T., Mennill D.J., Clemins P., Girod L., Yao K., *et al.* (2011) J. Appl. Ecol. 48: 758-767. https://doi.org/10.1111/j.1365-2664.2011.01993.x

Caroprese M., Albenzio M., Marzano A., Schena L., Annicchiarico G., *et al.* (2010) J. Dairy Sci. 93:2395–2403. https://doi.org/10.3168/jds.2009-2604

Diess V., Temple D., Liyour S., Racine C., Boux J., *et al.* (2009) Appl. Anim. Behav. Sci. 119:193–202. https://doi.org/10.1016/j.applanim.2009.03.018

Grandin T., Shivley C. (2015) Animals 5: 1233-1251. https://doi.org/10.3390/ani5040409

Keshavarzi H., Lee C., Johnson M., Abbott D, Ni W., Campbell D. L. M. (2021a) Sensors 21:924. https://doi.org/10.3390/s21030924

Keshavarzi H., Sadeghi-Sefidmazgi A., Mirzaei A., Ravanifard R. (2020b) Prev. Vet Med. 175: 104869. https://doi.org/10.1016/j.prevetmed.2019.104869

Krause J., Krause S., Arlinghaus R., Psorakis I., Roberts S., *et al.* (2013) Trends Ecol. Evol. 28 (9): 541-551. https://doi.org/j.tree.2013.06.002

Lee T.K., Lee C., Bischof R., Lambert G.W., Clarke I.J., *et al.* (2014) Psych neuroendocrinology 47: 166-177. http://dx.doi.org/10.1016/j.psyneuen.2014.05.015

R Core Team (2018) R Foundation for Statistical Computing. Vienna, Austria.

Valletta J.J., Torney C., Kings M., Thornton A., Maddence J. (2017) Anim. Behav. 124: 203-220. https://doi.org/10.1016/j.anbehav.2016.12.005

Teimouri M., Indah Ulf G., Sickel H., Tveite H. (2018) Int. 7: 78. https://doi.org/10.3390/ijgi7020078

109. Heritability of social interactions in dairy cattle

L. Rönnegård[1,2*], I. Hansson[1] and W.F. Fikse[3]

[1]Swedish University of Agricultural Sciences, Uppsala, Sweden; [2]Dalarna University, Falun, Sweden; [3]Växa, Uppsala Sweden; lars.ronnegard@slu.se

Abstract

Social interactions in dairy cattle are important to understand since they influence both welfare and disease transmission. We investigated the number of contacts a cow has with other group members closer than 2.5 m. The aim was to quantify how heritable such a trait is and whether cows more related to other cows in the same group have more contacts. Using location data of cows, we quantified the number of contacts between cows and documented in which areas in the barn these interactions occurred. An animal model was fitted using pedigree information from 163 cows in two different milking groups of around 100 cows in each group (group size varied between days). The positions of all cows were recorded once every second to compute the number of contacts. The estimated heritability varied between 0 and 0.20 and was largest in the resting area.

Introduction

Resource-based aggression and sociability are important traits affecting animal welfare for dairy cows. In resource-based aggression, cows approach each other in their competition for resources, such as water or feed (Gibbons *et al.*, 2010; Haskell *et al.*, 2014). Sociability is the willingness to be close to other animals of the same group. Both sociability and resource-based aggression causes cows to be in close proximity to other cows and may be treated as two types of social interactions.

Heritability of social interactions for dairy cows have not been investigated thoroughly. Temperament is a factor that might affect social interactions and is known to be heritable in both beef and dairy cattle (Costilla *et al.*, 2020; Haskell *et al.*, 2014). Nevertheless, there are few, if any, studies on the heritability of social interactions *per se*.

The aim of this study was to estimate the heritability of social interactions in dairy cattle in an indoor environment. To this end, indoor locational data collected every second on all cows over a two-week period were used.

Materials & methods

Data description. Data was collected from a commercial dairy farm in Sweden that housed around 210 lactating dairy cows in a non-insulated free-stall barn. The barn was divided into two milking groups, G1 and G2, where G1 consisted of cows in mid and late lactation and G2 consisted predominately of cows in early and mid-lactation. The pen area was 14×56 m for each group and there were 103 cubicles in G1 and 102 cubicles in G2, with rubber mattresses and sawdust as bedding material. The cows were fed a total mixed ration 12 times a day and had approximately 0.6 m of feed space per cow. The cows were milked twice a day in a milking parlor, and they spent around 1.5 h in the milking parlor during each milking event. Each lactating cow was equipped with a tag mounted on the top of the neck collar connected to a real-time location system (CowView, GEA Farm Technologies, Bönen, Germany) automatically collecting individual positioning data of the cows with one-second fix rate. Raw positioning data were extracted from 16 to 29 October 2020. Individual attribute data such as parity, calving date, and tag-ID were provided by the farm and information about breed, insemination records, pregnancy diagnoses, claw trimming records, and pedigree were extracted from the Swedish official milk recording scheme.

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_109, © L. Rönnegård *et al.* 2022

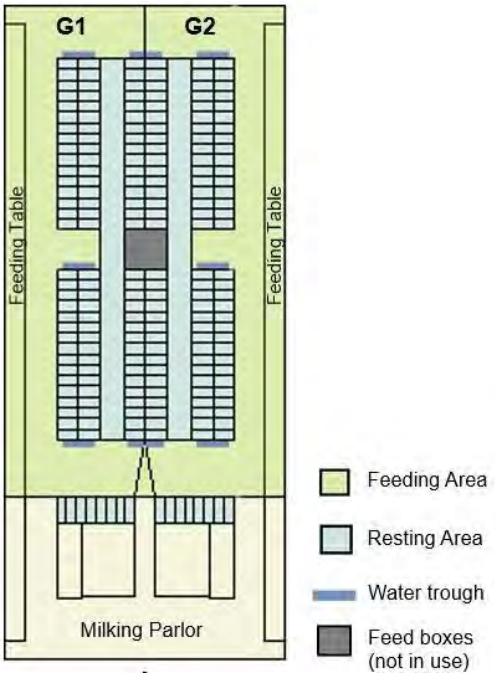


Figure 1. Layout of the barn. Data from two groups were used in the analysis. The late lactating cows (G1) were kept in the half of the barn area depicted by G1 (i.e. the left-hand side), whereas the early lactating cows (G2) were kept in the other half of the barn.

Statistical models. The total duration of contacts for each cow and day was calculated separately for two functional areas in the barn, the feeding area and resting area (Figure 1). The time spent in the two areas were accounted for, as there might be individual variation in time budgets. The response variable was computed as the ratio between total duration of contacts and the time spent in the specific area, and is referred to as the 'average number of contacts'. Cows closer than 2.5 m from each other were defined to be in contact. Contacts that lasted in total less than 10 minutes per day were removed to reduce noise.

Data from the two groups, G1 and G2, were analyzed separately. The response variable was computed for all individuals present on all 14 days using the connections with all individuals present on each day. Both groups were not constant during the study period, and only responses for animals present in the group during the entire period were consider for the subsequent analysis. The number of individuals present in group G1 on all 14 days was 80, and 83 in group G2.

Two linear mixed models with the same explanatory variables were fitted, one for each functional area. Fixed effects were date, parity, lactation stage, breed, pregnancy status, estrus, udder health and claw health, as described in Table 1. Furthermore, level of relatedness was included as a fixed effect nested within breed. The level of relatedness measures how related each cow is with the rest of all the cows in the same group and was computed as the row sums of the additive relationship matrix minus the diagonal element. Random effects were permanent environmental effect and animal effect. The additive relationship matrix was computed, and the model was fitted using the pedigreemm (Vazquez *et al.*, 2010) package in R. Box-Cox transformation was applied to the response variable average number of contacts.

Table 1. Data description of the cows included in the analysis.

Trait	No. of cows in group G1 (n=80)	No. of cows in group G2 (n=83)
Parity		
1	21	23
2	28	22
3	31	38
Lactation stage		
Early (7-49 DIM)	0	17
Mid (50-179 DIM)	16	60
Late (>180 DIM)	64	6
Breed		
Holstein	21	27
Red Dairy Cattle	22	19
Crossbred	37	37
Estrus		
In estrus	2	28
Not in estrus	65	54
No information	13	
Pregnancy status		
Pregnant	62	39
Open	18	47
Udder health		
Low (0-130,000 SCC/ml)	50	58
Mid (130,000-300,000 SCC/ml)	8	11
High (>300,000 SCC/ml)	22	14
Claw Health		
Remark	24	24
No remark	54	58
No information	2	1

Results

The estimated heritabilities were larger in the resting area than the feeding area (Table 2), and the repeatabilities were consistent between groups for each of the two defined areas.

For group G1, in the resting area, the following fixed effects were significant (with $P<0.05$), date and claw health. For group G1, in the feeding area, the following fixed effects were significant: date, breed, pregnancy and lactation stage. For group G2, in the resting area, the following fixed effects were significant (with $P<0.05$), date, pregnancy, estrus and udder health. For group G2, in the feeding area, the following fixed effects were significant, date, parity, lactation stage, and estrus. The fixed effect of level of relatedness was found to be non-significant ($P>0.05$) for both groups and both investigated areas.

Table 2. Estimated heritabilities (h^2) and repeatabilities (r^2) for the two groups in the two different areas. Twice the change in REML likelihood for a model with and without a random animal effect is given as $-2\log LR_{REML}$.

	Group G1			Group G2		
	h^2	r^2	$-2\log LR_{REML}$	h^2	r^2	$-2\log LR_{REML}$
Resting area	0.20	0.47	0.14	0.13	0.50	0.17
Feeding area	0.14	0.35	0.24	0	0.36	0

Discussion

Using a real-time location system, we have computed the number of contacts a cow has with other group members closer than 2.5 m. The estimated heritability for number of contacts varied between 0 to 20% and was found to be largest in the resting area of the investigated commercial free-stall farm. However, for all four models the difference in REML likelihoods (for a model with and without animal effect) were rather low and the estimated heritabilities could not be concluded to be significantly different from 0. Furthermore, we expected cows closely related to other cows in the same group to spend more time together than unrelated individuals. However, the level of relatedness did not have a significant effect on the number of contacts a cow had.

In the current study we constructed social networks to compute the number of contacts each cow had. However, all the information in the social network was not included in the analysis. In terms of social network attributes, we estimated the heritability of social contacts using node attributes (i.e. information on individual cows) as the response variable, whereas an alternative approach is to analyse edge weights (i.e. information on pairs of cows). For instance, suppose there might be a tendency for cows with similar attributes to stay close to each other more often than others, then this is an effect possible to capture in a model with edge weights as response, but not in the model applied in our analysis. In our analysis we focused on whether cows with similar attributes stay close to any cow more often than cows with different attributes. For further information on estimating heritability of social phenotypes in human and wild populations using edge weights, see Radersma (2021).

In our analysis, the social networks were constructed using information on whether a cow was in close proximity to other cows, which does not necessarily capture true social interactions. Also, it would be useful to be able to distinguish between positive and negative social interactions, which is not an obvious task having only positional data. Further investigations are needed on how to define social interactions based on location data.

We do anticipate future studies within this field since social interactions are important to understand for further development of optimal management, welfare, and strategies to counteract disease transmission.

References

Costilla, R., Kemper, K. E., Byrne, E. M., Porto-Neto, L. R., Carvalheiro, R., Purfield, D. C., Doyle, J. L., Berry, D. P., Moore, S. S., Wray, N. R., & Hayes, B. J. (2020). Genetic control of temperament traits across species: Association of autism spectrum disorder risk genes with cattle temperament. Genetics Selection Evolution, 52(1), 51.

Gibbons, J. M., Lawrence, A. B., & Haskell, M. J. (2010). Measuring sociability in dairy cows. Applied Animal Behaviour Science, 122(2), 84–91.

Haskell, M. J., Simm, G., & Turner, S. P. (2014). Genetic selection for temperament traits in dairy and beef cattle. Frontiers in Genetics, 5, 368.

Radersma, R. (2021). Estimating heritability of social phenotypes from social networks. Methods in Ecology and Evolution, 12(1), 42–53.

Vazquez, A. I., Bates, D. M., Rosa, G. J. M., Gianola, D., & Weigel, K. A. (2010). Technical note: An R package for fitting generalized linear mixed models in animal breeding1. Journal of Animal Science, 88(2), 497–504.

110. Uncertainty about the social network among dairy cows impairs estimation of indirect genetic effects

W.F. Fikse[1*], I. Hansson[2] and L. Rönnegård[2,3]

[1]Växa, Uppsala, Sweden;[2]Swedish University of Agricultural Sciences, Uppsala, Sweden; [3]Dalarna University, Falun, Sweden; freddy.fikse@vxa.se

Abstract

Social interaction among animals living in groups mean that an individual's phenotype can be influenced by conspecifics. One approach for modelling of social genetic effects utilizes intensity of interactions measuring the strength of the interaction. Quantifying the intensity of interactions is not always straightforward, and the aim was to investigate consequences of uncertainty about intensity of interactions. Social networks among 10,000 cows in 100 herds were simulated, and a cow's phenotype was affected by on average five conspecifics. Variance components for social genetic effects were estimated with three different scenarios. Using correct intensities for all conspecifics yielded unbiased estimates. Estimates were still unbiased, but had lower precision, when only one conspecific was included. When intensities were subject to random noise, the variance of the social genetic effect was downwards biased. Hence, estimation of intensity of interactions is of higher priority than identifying all conspecifics for elucidating social genetic effects.

Introduction

Phenotypic performance of animals living in groups is not only affected by their own genetic potential, but also by group mates and their genetic potential. These social effects have been reported to exhibit considerable genetic variation in livestock production species like pigs and poultry (Ellen *et al.*, 2014). Social interactions occur even in dairy cattle (Bouissou *et al.*, 2001), but studies on the genetics of social interactions in dairy cattle are thus far scarce.

The model predominantly used for uncovering social genetic effects assumes that an animals' social genetic effect on its conspecifics is uniform. This model is often applied in the context of animals housed in groups of fixed size and composition. For forest trees, Cappa and Cantet (2008) proposed to consider intensity of competition, in their study quantified as the distance between the focal tree and a neighboring tree. That is, the total additive genetic competition that is exerted over plant i from trees $j_1, j_2, \ldots j_m$ equals $\sum_{j=1}^{m_i} f_{ij}\boldsymbol{a}_{c,j}$, with $a_{c,j}$ being the social genetic effect of tree j, and f_{ij} the intensity of competition (or more neutrally, intensity of interaction) between tree j and i. The intensity can be interpreted as a weighting factor that expresses how intense pairs of individuals interact in relation to all other individuals in the group. Angarita *et al.* (2019) applied this model to group-housed pigs and defined the intensity as the total time of aggressive interactions that took place between the pair of animals over a 9-h period after groups were created.

Social networks in dairy cattle are dynamic (Bouissou *et al.*, 2001), as cows enter and leave the barn in relation to calving and drying off and because the behavior may change over time, for example when time spent feeding alters due to different energy requirements across the lactation (Løvendahl and Munksgaard, 2016). Real-time location systems can provide detailed information about the position of cows in the barn and can be used to determine how long a cow is in the proximity of other cows. Such data can be used to construct network graphs (Chopra *et al.*, 2020), but it remains uncertain whether these accurately reflect the true social network since proximity can occur for reasons other than social interactions.

Data structure has been shown to be of importance for estimating social genetic effects (Cappa and Cantet, 2008). The aim of the present study was to investigate, by simulation, how uncertainty about social networks affects estimation of variance components for social genetic effects in dairy cattle scenario.

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_110, © W.F. Fikse *et al.* 2022

Methods

Phenotypes for a conceptional trait were simulated for 10,000 females in 100 herds. Within each herd, half of the females were descendants of 5,000 base population females and were offspring to the other half of the females in the herd. All females were randomly assigned a sire out of 200 unrelated sires, each having 50 female progenies.

For each animal, two genetic traits were simulated: a direct genetic effect and a social genetic effect, assumed to follow a normal distribution, ~ N(0, A $\otimes$ G), where A is the numerator relationship among individuals and G the variance-covariance matrix among traits. The variance of the direct and social genetic effect was 2 and 1, respectively, and the genetic correlation between both traits was set to 0.18.

A social network was generated within each herd. First, the number of conspecifics an animal had interactions with (degree) was generated from a normal distribution (~ N(5,4)). Next, a random social network was generated with the sample_deqseq() function of the igraph package (Sardi and Nepusz, 2006), using the degrees as input. Lastly, for each contact the intensity was sampled from a gamma distribution with parameters 1 (shape) and 2 (rate).

The phenotype for each female was generated as the sum of the direct genetic effect, the product of the intensity and the social genetic effect of all conspecifics the animal had contact with, and a random residual (with variance 2.5). The total heritable variation was about 2.5.

Variance components were estimated using DMU (Madsen *et al.*, 2014) for three different scenarios:

1. Using true intensities for all conspecifics the animal interacted with.
2. Using true intensity for just one (randomly chosen) conspecific the animal interacted with.
3. Using imprecise intensities for all conspecifics the animal interacted with. The imprecise intensities were generated by adding random noise (~ N(0,0.04)) to the true intensities.

The simulation was repeated 25 times and the mean and SD of estimated variance components are presented.

Results

Estimates of variance components were unbiased when true intensity for all conspecifics were included in the analysis (Table 1). When only one conspecific was assumed known, and with the correct intensity, estimates of genetic (co)variances were still unbiased but had lower precision. The social interactions unaccounted for when most conspecifics were presumed unknown caused an overestimation of the residual variance. Using imprecise intensities resulted in an underestimation of the variance of social genetic effects and an overestimation of the residual variance.

Discussion

Constructing a social network of dairy cows in a barn is not an easy task, even with modern technology continuously monitoring dairy cows in a barn. Processing data from real-time locations systems on the position of a cow in the barn and extracting information about the social network requires separating real social interactions from proximity because of other reasons (Chopra *et al.*, 2020). Similar challenges exist if cows are recorded with video. Our results indicate that imprecise information on intensity of interaction results in more problems in variance component estimation than missing information. Studies like these can serve as guidance what to prioritize in constructing social network for elucidating genetics of social interactions.

Table 1. Mean and SD of estimated (ratio of) variance components (vDGE = variance of direct genetic effect, vSGE = variance of social genetic effect, rG = genetic correlation between direct and social genetic effect, vE = residual variance).

	vDGE		vSGE		rG		Ve	
	Mean	**SD**	**Mean**	**SD**	**Mean**	**SD**	**Mean**	**SD**
Simulated	*2.00*		*1.00*		*0.18*		*2.50*	
Scenario 1	1.99	0.12	1.00	0.04	0.18	0.03	2.51	0.10
Scenario 2	2.00	0.15	0.99	0.10	0.18	0.07	4.43	0.16
Scenario 3	1.99	0.12	0.88	0.04	0.18	0.03	2.63	0.10

Scenario 1 = using true intensity for all conspecifics, Scenario 2 = using true intensity for just one conspecific, Scenario 3 = using imprecise intensity for all conspecifics.

The simulated network was a simple, random, symmetric network. Social networks of dairy cows are more complex in practice. Social interactions between cows can be positive or negative, are influenced by familial relationships and familiarity of the cows and can depend on cow attributes like stage of lactation or parity number (Bouissou *et al.*, 2001). Also, the present simulation assumed the degree (number of contacts) to be identically and independent distributed, whereas one can envision that it has a heritable component.

A social network can mathematically be represented as a distance matrix, where each element describes the intensity of the contact between a pair of individuals. This representation relates very well to the model for social genetic effects based on intensity of contacts. This model allows for different number of conspecifics per individual and for non-symmetric networks (where the social interaction effect of individual i on individual j can be different from the effect of individual j on individual i). Extension to consider repeated observations for the trait, for example multiple test-day yields, is easily accomplished. By specifying different intensities for different observations on the same individual, the dynamics of social networks over time can be accommodated for.

Conclusions

Uncertainty in the social network will affect estimation of variance components for a social genetic effects model. Not being able to identify all conspecifics resulted in unbiased estimates, albeit with lower precision. Hence, precise estimation of intensity of interactions is of higher priority than identifying all conspecifics an animal interacts with in constructing social networks for genetic analyses of social effects in dairy cattle.

References

Angarita B.K., Cantet R.J.C., Wurtz K.E., O'Malley C.I., Siegford J.M, *et al.* (2019) J. Anim Sci. 98, 3658-3668. https://doi.org/10.1093/jas/skz244

Bouissou M.F., Boissy A., Neindre P. le, and Veissier I. (2001) The social behaviour of cattle. In: Keeling L.J. and Gonyou, H.W. (eds.), Social Behaviour in Farm Animals. CABI, Wallingford, pp. 113–145. https://doi.org/10.1079/9780851993973.0113

Cappa E.P., and Cantet R.J.C. (2008) Silvae Genet. 57(2), 45-56. https://doi.org/10.1515/sg-2008-0008

Chopra K., Hodges H.R., Barker Z.E., Vázquez Diosdado J.A., Amory J.R. *et al.* (2020) Front. Vet. Sci. 7., 583715. https://doi.org/10.3389/fvets.2020.583715

Csardi G., and Nepusz T. (2006) InterJournal Compl. Syst., 1695. https://igraph.org

Ellen E.D., Rodenburg T.B., Aalbers G.A.A., Bolhuis J.E., Camerlink I., *et al.* (2014) Front. Genet. 5, 377. https://doi.org/10.3389/fgene.2014.00377

Løvendahl P., and Munksgaard L. (2016) J. Dairy Sci. 99, 408–417. https://doi.org/10.3168/jds.2015-9838

Madsen P., Su G., Labouriau R., Christensen O. (2014). Proc. of the 10th WCGALP, Vancouver, Canada.

111. Genetic and phenotypic variation and consistency of cow's preference and circadian use of robotic milking units

P. Løvendahl and A.J. Buitenhuis*

Center for Quantitative Genetics and Genomics, Aarhus University, AU-Foulum, 8830 Tjele, Denmark; bart.buitenhuis@qgg.au.dk

Abstract

Many dairy herds use automatic milking stations (AMS) and in large herds, cows usually have access to two or more AMS, and consequently must make a choice between them whenever they go for milking. For this study, we assume that the degree of using the same milking box is an expression of preference. We used a total of 4,665 cows from a Holstein and a Jersey cohort covering 12 commercial dairy herds in Denmark to estimate heritability, repeatability and individual level correlations between preference choice score (PCS), Time_profile (preference for specific clock hours) and associated milking and behaviour traits. The average PCS was 0.43 (Holstein) and 0.41 (Jersey), respectively. Both PCS and Time_profile had low heritability (0.07/0.11 Holstein, 0.13/0.04 Jersey) and moderate repeatability (0.47/0.40 Holstein and 0.50/0.42 Jersey) and were weakly correlated to each other (Holstein 0.18 and Jersey 0.17) and weakly correlated to milk yield (r_i range 0.0 to -0.10).

Introduction

Dairy cows have been subject to intensive genetic selection for production traits for decades, and many cows are living in housed environments interacting with the confinements of housing, and with various automated installations, including robotic milking machines. In these modern dairy herds cows are making a number of choices every day. For example, at milking cows form an order of entry to milking parlours, and through their use show preference to their favourite side of the milking parlour (Hopster *et al.* 1998). When accessing automated milking stations (AMS, milking robots), they may acquire a routine with preference for a specific one, or be more flexible. If such routines are long lasting, it indicates a degree of consistency, which may have a genetically determined component (Koolhaas and Van Reenen, 1999; Løvendahl *et al.*, 2016). Similarly, cows may develop routines or preferences for being milked at specific hours of the day. The consistency of these routines are sometimes very strong, and the strength of it was shown to be both an acquired feature, but also to include some genetic variation (Løvendahl *et al.*, 2016). However, little evidence is available on how different aspects of behaviour at robotic milking are related, and how these relate to key production traits at the phenotypic and the genetic level. The aim of the present study was to investigate preference consistency, in Holstein and in Jersey cows having access to two, three or four milking boxes, and their use of time of day to be milked, for repeatability over time and for genetic variation, and eventually for correlation to milk yield and time spent in the milking box. The quantitative genetic study was based on records from commercial dairy herds in Denmark.

Materials & methods

Overall design, herds and cows. This investigation was conducted as a cohort study using robotic milking data from 12 Danish dairy herds, having either Jersey or Holstein cows of mixed parities (1-5) between October 2014 and September 2019. There were 2,407 Jersey cows with 28,952 records and 2,258 Holstein cows with 31,290 records. Cows were kept in groups having access to either 2, 3 or 4 milking boxes with free cow traffic, so that cows had a choice at every milking visit. The genetic relationships between cows within each breed was known and traced back 5 generations through the NAV database (NAV, SEGES, Skejby, Denmark). The Hostein pedigree consisted of 20,229 and the Jersey pedigree consisted of 14,543 animals. This study used non-invasive production data and was therefore exempt from animal experimental ethical approval.

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_111, © P. Løvendahl and A.J. Buitenhuis 2022

Milking data recording. At each milking the time of entry and exit from robot was recorded together with milk yield. From these data the time spent in the milking box (Boxtime, minutes) was calculated. The interval between milkings were expressed as milking frequency (i.e. milkings per 24 h, MilkFreq), because cows have voluntary access to milking boxes and thereby will have variable milking intervals. Milk composition data to calculate the energy corrected milk (ECM) (Sjaunja *et al.* 1990) was measured through the national recording organization RYK (RYK, Skejby, Denmark) at the Eurofins laboratory (Eurofins, Vejen, Denmark).

Milking box preference consistency score calculations. Preference consistency score (PCS) was calculated based on frequency of access's to each of the milking units in a given time period, here chosen at 15 days ('segment'). Cows accessed milking boxes voluntarily so it is assumed that the box being visited most frequently is the preferred one. If all boxes were equally used the frequency of first choice will be equal to 1/number of choices. The PCS is then calculated using:

$$PCS = (Actual_fraction - (1/number\ of\ choices)) / (1 - (1/number\ of\ choices)) \quad (1)$$

Diurnal milking time profile. The time profile of being milked at specific times of the day was here estimated based on frequency of milkings at each hour of the day, condensed in a single number indicating the variation. It was assumed that in case a cow has no hour-preference, the frequency in each hour-bin is equal to all other hour-bins. In contrast, if all or most milkings will fall into one or a few hour-bins, the deviation from equal use of hour-bins will be large. In between these extremes will be a continuum of intermediate cases. The deviation from equal frequency in all hour-bins was quantified using a Chi-square goodness-of-fit approach, and log-transformed to obtain an approximate normal distribution, better suited for analysis of variance.

Statistical models and parameter estimates. Variance-covariance components were estimated within breed across lactation using uni- or bi-variate models in DMU software package (DMU version 6 Release 4.7; Madsen and Jensen, 2008):

$$y = Xb + Z_{pe}a_{pe} + Z_g a_g + Ie \quad (3)$$

where y is the vector of observed traits, b is a vector of fixed effects of herd-year-season, number of AMS units (2, 3 or 4), parity-group (first, second or older=3-5), segment (1-20) by parity interaction. The matrix X contain the design of the fixed effect factors. The random effects are the additive genetic effect of cow (a_g), the permanent (cow) environmental effect across segments (a_{pe}) and the random error effect (e).These have incidence matrices Z_g being the relationship A-matrix and Z_{pe}, being incidence cow within parity. I being the identity matrix and e the random error term. Random (co)variance components were assumed independent following a normal distribution $N(0,\sigma_a^2)$, $N(0,\sigma_{pe}^2)$, and $N(0,\sigma_e^2)$. Genetic parameters were calculated from (co)variance estimates with associated asymptotic standard errors.

Results

The use of available milking boxes differed among cows, because some cows almost entirely used just one box while other cows spread their milkings over all available boxes. This was effectively quantified by the PCS, with values ranging between zero and 1.0, and an overall mean of 0.43 (sd=0.28) and 0.41 (sd=0.27) in Holstein and Jersey cows. This value is clearly larger than zero, where a zero value would indicate even use of all boxes. First parity Jersey cows had the highest PCS followed by first parity Holstein, and lowest values were found in the group of older Jersey cows. The PCS decreased with age in Jersey cows, but remained at similar level for all ages in Holsteins. The Time_profile varied among cows, and high values show that

some cows make most of their milkings during few clock hours, whereas the variation show that other cows spread their milkings over several clock hours. The first parity Jersey cows, and to a lesser degree, the first parity Holstein cows condensed their milking events into fewer clock hours than older cows of both breeds. There was a general trend for the time profile to increase gradually during progression of lactation, and Jersey cows had always a slightly higher Time_profile score than Holsteins. The milking frequency increased during early lactation for young cows of both breeds, and then decreased gradually from about 100 DIM. In contrast, milking frequency decreased throughout lactation in all groups of older cows. The time spent in the milking box increased in early lactation for all breed and age groups, and then declined after reaching a peak at around 50 DIM. Heritability estimates for the behavioural traits and ECM are presented in Table 1.

The genetic and the individual level correlations between PCS and Time_profile (r_i ~ 0.10 to 0.18) showed that cows who more consistently use a given milking box, also tend to have a stronger Time_profile, i.e. be milked more often at the same time of the day. Those cows that are more consistent also have somewhat lower milking frequency (r_i ~ -0.18 to -0.16). However, consistency seem almost unrelated to total daily time spent in the milking box and neither to milk yield. Individual cows who show a stronger Time_profile were also spending less time in the milking box, and may have slightly lower milk yield (r_i ~ -0.10). High milk yield was weakly related to Time_profile and even less to PCS, but positively with higher milking frequency and with more time per day spent in the milking box. Overall, the phenotypic correlations closely resembled the individual level correlations and correlations in the Jersey cows were almost similar to those obtained in Holsteins.

Discussion

Dairy cows entered the milking box on around 3 times per day in this study, and each time they had to wait for a vacant box, either being their favourite one, or another one of at least 2, and often 3 or 4 boxes. There were no obvious reasons for cows to prefer a specific box in general, as the available boxes within a group were used almost equally much, as also found in a previous study (Løvendahl *et al.*, 2016). Assuming a specific cow had a preferred box she would have to wait until it was available, or to choose one of the other boxes. In the waiting area, she could also chase other cows away from the queue to get access to her preferred box. In this context, we interpret higher use of a specific box as preference and measure it over a short period of 15 d. In case only two possible choices are available this is reasonably simple, and a

Table 1. Estimates of heritability and repeatability for milking behaviour consistency traits in Holstein and Jersey cows, across whole lactation.

Trait[1]	Breed	Heritability (h^2±SE)	Repeatability (t±SE)
PCS	Holstein	0.07±0.02	0.47±0.01
	Jersey	0.13±0.03	0.50±0.01
Time_profile	Holstein	0.11±0.02	0.40±0.01
	Jersey	0.04±0.02	0.42±0.01
MilkFreq (no/d)	Holstein	0.22±0.04	0.72±0.01
	Jersey	0.17±0.04	0.64±0.01
Boxtime (min/d)	Holstein	0.25±0.05	0.77±0.01
	Jersey	0.27±0.05	0.75±0.01
ECM_Yield (kg/d)	Holstein	0.08±0.03	0.65±0.01
	Jersey	0.10±0.03	0.68±0.01

[1] PCS: preference consistency score; Time_profile: use of hours at specific time of the day; ECM: Energy Corrected Milk.

'consistency score' can be used (Hopster *et al.* 1998, Løvendahl *et al.* 2016). However, for this study some cows had 2 boxes, others had 3 or 4 boxes to choose between, so in order to adjust for number of choices an adjustment of the consistency score was implemented. This adjusted score, now called a preference consistency score PCS gave values between zero and 1.00 based on how many of the milkings were on the most frequently used box. One should note that the PCS itself resembles a repeatability coefficient, but here used as a response variable.

This study showed that differences between cows in use of specific milking boxes are repeatable and to some degree also heritable. This extends earlier findings from a preliminary study only in Holsteins and only in herds with two milking boxes (Løvendahl *et al.*, 2016), to include many more cows and two breeds, and situations also comprising 3 or 4 milking boxes.

This study was primarily designed to quantify genetic and phenotypic co-variation in milking behavior traits as possible correlated response to selection for production, but not to include these traits in selection programs. Although more than 4,600 cows provided data, genetic correlations had large standard errors, whereas individual level correlations were used as proxies and gave similar results for the two breeds. Obtaining even larger cohorts or using genomic information may help resolve this problem in some future studies.

References

Hopster H., van der Werf J.T.N., and Blokhuis H. (1998) Appl. Anim. Behav. Sci. 55:213-229. https://doi.org/10.1016/S0168-1591(97)00064-6

Koolhaas J. M., and Van Reenen C.G. (2016) J. Anim. Sci. 94: 2284–2296. https://doi.org/10.2527/jas.2015-0125

Løvendahl P., Sørensen L.P., Bjerring M., and Lassen J. (2016) J. Dairy Sci. 99:9857–9863. https://doi.org/10.3168/jds.2016-11287

Madsen P., and Jensen J. (2008) A user's guide to DMU: A package for analysing multivariate mixed models. Version 6 release 4.7. Available at: http://dmu.agrsci.dk.

Sjaunja L.O., Baevre L., Junkkarinen L., Pedersen J. and Setälä J. (1990) Proc. of the 27th ICEPMA, Paris, France.

112. Genetic and environmental influences on maternal calving difficulty in South African Holstein cattle

N.E. Mkhize[1,2*], N.O. Mapholi[1,3], B. Dube[2], M.A. Madilindi[2], O. Matika[4] and C.B. Banga[2,3,5]

[1]Discipline of Life and Consumer Sciences, College of Agriculture, Environmental Sciences, University of South Africa, PO Box 392, Pioneer Avenue, Florida,1709, South Africa; [2]ARC-Animal Production, Private Bag X2, Irene, 0062, South Africa; [3]Department of Agriculture and Animal Health, University of South Africa, Private Bag X6, Florida 1710, South Africa; [4]Centre for Tropical Livestock Genetics and Health (CTLGH), Roslin Institute, University of Edinburgh, Easter Bush Campus, EH25 9RG, Edinburgh, United Kingdom; [5]Department of Animal Sciences, Faculty of Science, Tshwane University of Technology, Private Bag X680, Pretoria, 0001, South Africa; mkhizene88@gmail.com

Abstract

Calving difficulty is an important trait in dairy production, as it affects profitability and animal welfare. Environmental factors and genetic parameters on maternal calving difficulty (MCD) were estimated for South African Holstein cattle using data from the National Milk Recording Scheme. The original data set comprised of calving records of 10,495,603 of cows 192,052, in 14 herds that calved between 1949 and 2018. Environmental effects were tested by the General Linear Models (GLM) procedure of the Statistical Analysis System. Maternal heritability for calving difficulty in the first three lactations was estimated by the ASReml software, using linear animal and repeatability animal models. Sex of calf, parity, herd-year-season and age of dam at calving had significant effects ($P<0.05$) on MCD. Heritability estimates for MCD were low, ranging from 0.04±0.09 in second parity to 0.12±0.10 in third parity. Both estimates of heritability and repeatability of MCD across parities were the same in magnitude (0.04±0.04). Despite the low heritability estimates, these results provide the basis for developing genetic prediction models for MCD in South African Holstein cattle.

Introduction

Calving difficulty (dystocia) is a reproductive trait of outstanding importance in dairy cattle (Abdela and Ahmed, 2016). Along with calf and cow mortality, it results in economic losses due to increased labour and veterinary costs, and other management expenses (Kaya *et al.*, 2015). Difficulties during parturition also affect long-term animal performance through health issues, fertility problems, reduced production, and involuntary culling (Abdela and Ahmed, 2016; Bujko *et al.*, 2018). Dystocia is also a welfare problem, as it causes pain and injury to the cow, which directly leads to poor cow welfare (Abdela and Ahmed, 2016). Furthermore, calving-related infections indirectly affect human health, as they require increased use of antibiotics, which contaminate the cow's milk, leading to possible microbial resistance (Silvestre *et al.*, 2019).

Proper herd management practices may alleviate the effects of dystocia (Newman, 2008); however, several studies have demonstrated that genetic selection for better calving difficulty, is a better option in the long term (Alam *et al.*, 2017). Calving difficulty, along with other functional traits, is becoming increasingly important in dairy cattle breeding objectives worldwide (Egger-Danner *et al.*, 2015).

In South Africa, calving performance traits are currently not included in the national genetic evaluation programme for dairy cattle, and genetic parameter estimates for these traits are not available. Therefore, the primary aim of this study was to estimate environmental and genetic influences on maternal calving difficulty in South African Holstein cattle, in order to develop models for genetic prediction of the trait.

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_112, © N.E. Mkhize *et al.* 2022

Materials & methods

Data. Data comprised of 14,250 calving records of 8,832 cows from 14 herds participating in the National Milk Recording Scheme, that calved between 2009 and 2018, after editing. Calving difficulty was measured as a binary trait, with animals experiencing a difficult calving being scored 1, and those having a normal calving being scored 0. The pedigree file consisted of 27,091 animals, and was built based on consideration of animals with calving records and going three generations back.

Statistical analysis. Non-genetic effects on MCD were tested by a fixed effects model, using the General Linear Models (GLM) procedure of SAS (SAS Institute, 2012).

The heritability of MCD for each parity was estimated by applying the following model Equation 1, using the ASReml 3.0 software (Gilmour *et al.,* 2015). Only maternal effects of MCD were included in the model:

$$y = Xb + Za + e \quad (1)$$

Where, y is the vector of observations for calving difficulty (0=normal, 1=difficult); b is the vector of fixed effects (herd-year-season and calf sex); a is the vector of random additive genetic effects of the cow; X is the incidence matrix relating observations to fixed effects; Z is the incidence matrix relating observations to random additive genetic effects; e is the vector of random residual effects.

Heritability and repeatability across parities were estimated by the following model Equation 2, using the ASReml 3.0 software (Gilmour *et al.,* 2015):

$$y = Xb + Za + Wpe + e \quad (2)$$

Where y, X, Z, a and e are the same as in Equation 1, b is the vector of fixed effects (herd-year-season, calf sex and parity), pe is a vector of permanent environmental effects and W is an incidence matrix relating observations to permanent environmental effects.

Results & discussion

Incidence of calving difficulty was significantly higher ($P<0.05$) for male (14%) than female (12%) calves (Figure 1A). This is a widely observed fact (Atashi *et al.*, 2012; McHugh *et al.*, 2014) and is attributable to the fact that male calves are generally bigger in size than females, which may put more strain on the dam's birth canal during parturition. First-parity cows had a significantly higher ($P<0.05$) incidence of calving difficulty (16%) than second-parity (11%) and third-parity (11%) cows (Figure 1B). Calving assistance has been mostly found to be higher in first-parity than in later-parity calvings, and calving complications occur more in heifers than in cows (Eaglen *et al.*, 2012; Ghafariania *et al.*, 2014). This may be attributable to poor pelvic development in 2-year-old heifers which often is not fully compensated by a smaller calf (Kebede *et al.*, 2017). Herd-year-season and age of dam at calving also had significant effects ($P<0.05$) on maternal calving difficulty, in agreement with other studies (e.g. Vallée *et al.*, 2013).

Heritability estimates for MCD in the first-three parities are presented in Table 1. Estimates were low, ranging from 0.04±0.09 in second parity to 0.12±0.10 in third parity, with an overall estimate of 0.04±0.04 across the parities. Comparable estimates have been reported in previous studies (Jamrozik and Miller, 2014; Alam *et al.*, 2017) and suggest low accuracy of selection for MCD. However, a study by Vostrý *et al.* (2014) reported higher heritability estimates, ranging from 0.23 to 0.43 in Czech Charolais cattle. Discrepancies in heritability estimates for calving difficulty may be attributable to the manner in which the trait is measured and the statistical models used. In some cases, calving difficulty is scored on a scale of up to 5, unlike the

Figure 1. Incidence of calving difficulty by calf sex (A) and parity (B) in Holstein cows (means (%) of calving difficulty with different superscripts differ significantly at $P<0.05$).

Table 1. Heritability estimates of maternal calving difficulty within and across first-three parities in Holstein cows.

Parity	Heritability ± standard error
1	0.10±0.04
2	0.04±0.09
3	0.12±0.10
All	0.04±0.04

binary measurement used in the current study. Analysis of calving ease with linear models has been found to yield additive genetic variance estimates that are consistently smaller than those obtained with threshold models (Silvestre *et al.*, 2019). A low repeatability estimate of 0.04±0.04 was observed for MCD across the first three parities, which is comparable to those reported in Iranian Holstein cows (Ghiasi *et al.*, 2014). Thus, first parity records may not be used reliably to predict calving performance in later parities.

Conclusions

Sex of calf, parity, herd-year-season and age of dam at calving are significant sources of variation in maternal calving difficulty and, therefore, should be included in models for genetic analysis of MCD. Low heritability estimates suggest poor accuracy of selection on MCD in South African Holstein cows. Further research needs to be conducted, to explore methods of improving accuracy of selection. This may include marker assisted selection or promoting the large-scale recording of calving performance on South African herds, using clear guidelines and standards.

References

Abdela N., and Ahmed W.M. (2016) J. of Repro. and Infert., 7**:** 63-74.

Alam M., Dang C., Choi G., Choy T.J., Lee Y.H. *et al.* (2017) Asian-Aust. J. of Anim. Sci. 30:1225.

Atashi H., Abdolmohammadi A., Dadpasand M., and Asaadi A. (2012) Asian-Aust. J. of Anim. Sci. 25:447.

Bujko J., Candrák J., Strapák P., Žitný J., Hrnčár C. (2018) Albanian j. agric. sci. 17 (1):31-36.

Eaglen S.A., Coffey M.P., Woolliams J.A., and Wall, E. (2012) Genet. Sel. Evol. 44:1-13.

Egger-Danner C., Cole J., Pryce J., Gengler N., Heringstad B., *et al.* (2015) Animal 9:191-207. https://doi.org/s1751731114002614

Ghafariania Y., Babaiee M., and Vatankhah M. (2014) I. J. of Fund.and Appl. Life Sci. 4:1148-1155.

Ghiasi H., Khaldari M. and Taherkhani R. (2014) Live. Sci. 165:22-26.

Gilmour A.R., Gogel B.J., Cullis B.R., and Thompson, R. (2015) ASREML User's Guide Release 3.0. VSN International Ltd., Hemel Hempstead, U.K.

Jamrozik J., and Miller, S. (2014) Live. Sci. 162:42-49.

Kaya I., Uzmay C., and Ayyilmaz, T. (2015) Turk. J. of Vete. and Anim. Sci. 39:87-95.

Kebede H., Jimma A., Getiso A., and Zelke B. (2017) J. Fish. Livest Prod. 5: 237. https://doi.org/10.4172/2332-2608.1000237

SAS. (2012). Base SAS® 9.4 Procedures Guide: Statistical Procedures, Second Edition, Cary, NC: SAS Institute Inc.

Silvestre A., Martins Â., Santos V., and Colaço J. (2019) Ital. J. Anim. Sci. 18:80-87. https://doi.org/10.1080/1828051x.2018.1482801

Vallée A.,Van Arendonk J.A.M., and Bovenhuis H. (2013) J. Anim. Sci. 91 (12):5582-5588.

Vostrý L., Veselá Z., Svitáková A., and Vydrová H.V. (2014) Czech J. of Anim. Sci. 5:302-309.

113. Genetic evaluation of gestation length in Italian Holstein breed

F. Galluzzo[1*], J.B.C.H.M. van Kaam[1], R. Finocchiaro[1], V. Ferrari[1,2], M. Marusi[1] and M. Cassandro[1,2]

[1]Associazione Nazionale Allevatori della Razza Frisona, Bruna e Jersey Italiana (ANAFIBJ), Via Bergamo 192, 26100 Cremona (CR), Italy; [2]Department of Agronomy, Food, Natural resources, Animals and Environment (DAFNAE), University of Padova, Viale dell'Università 16, 35020 Legnaro (PD), Italy; ferdinandogalluzzo@anafi.it

Abstract

Gestation length (GL) affects both the cow and the newborn calf. Genetic components of GL are both direct and maternal, meaning that both the calf (direct effect) and the cow (maternal effect) genotypes contribute to determine it. The objective of this study was to define how much of the direct and maternal components is heritable in the Italian Holstein population and to estimate breeding values. A multiple-trait repeatability linear animal model was defined, considering first and later parities as different traits. The heritability of the direct effect was 0.44 (posterior standard deviation: 0.004) for first parities and 0.36 (0.003) for later. The heritability of maternal effect was 0.09 (0.001) for first parities and 0.06 (0.001) for later. In conclusion, the possibility of selecting for GL was confirmed. Further analyses will be performed in order to quantify the economic importance of GL and to identify its relationship with other traits.

Introduction

GL affects both the cow and the newborn calf and the effects come both from the calf's (direct effect) and the cow's (maternal effect) genotypes (Norman *et al.*, 2009, 2011; Eaglen *et al.*, 2013). As nowadays the importance of reducing costs has become crucial in farm management, a deep investigation on the possibility to select for GL and its effects on other economically important traits is needed.

Materials &methods

Data and data editing. Data came from the Italian National Breeders Association (AIA) and consisted of 24,705,045 calvings of Holstein cows. From the original dataset, only calvings from 2004 onwards and with a GL between 260 and 305 days were kept. Furthermore, only animals with both parents known were used. Records from embryo-transfer or multiple births were deleted. Regarding parity, only observations coming from cows with up to 10 calvings were analyzed. Age at parity had to be between 18 and 45 months for primiparous and between 33 and 240 months for multiparous. In order to have an effective comparison between contemporaries for herd-year-season of calving (HYS), groups with less than 20 observations were excluded. The pedigree went up to four generations. Below, in Table 1, the description of the final dataset.

Table 1. Description of the dataset.

Parameter	Value	Parameter	Value
Number of observations (calves)	6,620,306	Observations/HYS[1]	41.90
Number of dams	3,388,361	Male calves – frequency	0.49
Observations/dam	1.95	Female calves – frequency	0.51
Number of herds	6,061	Mean GL[2] male calves – days	279.03 (6.12)
Observations/herd	1,092.28	Mean GL[2] female calves – days	278.05 (5.85)
Frequency parity 1	0.36	Mean GL[2] parity 1 – days	277.71 (5.97)
Frequency parity 2+	0.64	Mean GL[2] parity 2+ – days	278.991 (5.97)
Number of HYS[1]	157,998	Number of records in pedigree	9,389,954

[1] HYS: Herd/Year/Season class.
[2] GL: gestation length.

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_113, © F. Galluzzo *et al.* 2022

Statistical model. The genetic evaluation was based on a multiple trait repeatability linear animal model, considering first parities and later parities as different traits. Fixed effects were selected with a stepwise approach using SAS PROC GLM (SAS Institute Inc., Cary, NC, 2017).

The defined model for first parity cows is described in Equation 1:

$$Y_{ijklmn} = HYS_i + SEX_{jk} + AGEC_l + a_m + d_n + e_{ijklmno} \quad (1)$$

with Y_{ijklmn} as the phenotypic observation of GL in days, HYS_i as the fixed effect of HYS i, SEX_{jk} as the fixed effect of sex j of the newborn calf within year of birth k, $AGEC_l$ as the fixed effect of age at calving l, a_m as the random direct additive genetic effect for calf m, d_n as the random maternal additive genetic effect of dam n and $e_{ijklmno}$ as the residuals. Regarding first parity cows, the fixed effect of sex was considered within year of birth because of the increasing use of sexed semen in the last years.

The defined model for later parity cows is described in Equation 2:

$$Y_{ijklmn} = HYS_i + SEX_j + P_k + mpe_l + a_m + d_l + e_{ijklmn} \quad (2)$$

with SEX_j as the fixed effect of SEX j of the newborn calf, P_k as the fixed effect of Parity k (2,3,4,5+), mpe_l as the random maternal permanent environment effect of dam l, a_m as the random direct additive genetic effect for calf m, d_l as the random maternal additive genetic effect of dam l and e_{ijklmn} as the residuals. For later parities, maternal permanent environment was considered because of the presence of across-lactations repeated records.

Estimation of genetic parameters. Variance components estimation was performed with the Gibbs sampler THRGIBBS1F90 (Misztal *et al.*, 2002), which handles linear traits as well as categorical. The analysis was performed on the entire dataset, with 90,000 iterations, a burn-in of 45,000 and a thinning rate of 10. Post-Gibbs analysis was performed using the software POSTGIBBSF90 (Misztal *et al.*, 2002) using the remaining 4,500 samples. Convergence was assessed visually.

For primiparous, the direct and maternal heritabilities were calculated as described in Equation 3 and Equation 4 respectively:

$$h_d^2 = \frac{\sigma_a^2}{\sigma_a^2+\sigma_d^2+\sigma_e^2} \quad (3)$$

$$h_m^2 = \frac{\sigma_d^2}{\sigma_a^2+\sigma_d^2+\sigma_e^2} \quad (4)$$

with h_d^2 as the direct heritability, h_m^2 as the maternal heritability, σ_a^2 as the direct additive genetic effect variance, with σ_d^2 as the maternal additive genetic effect variance and σ_e^2 as the residual variance.

As an animal model with repeated measures, the heritabilities for multiparous were calculated as described in Equation 5 and Equation 6 respectively:

$$h_d^2 = \frac{\sigma_a^2}{\sigma_a^2+\sigma_d^2+\sigma_{mpe}^2+\sigma_e^2} \quad (5)$$

$$h_m^2 = \frac{\sigma_d^2}{\sigma_a^2+\sigma_d^2+\sigma_{mpe}^2+\sigma_e^2} \quad (6)$$

with σ_{mpe}^2 as the maternal permanent environment effect variance and σ_e^2 as the residual variance.

Breeding value estimation. Breeding values (BVs) were estimated with MiX99 (Lidauer *et al.*, 2019) and expressed on a scale with mean of 100 and standard deviation of 5: values higher than the mean refer to individuals that transmit a longer gestation period.

Results

Estimation of genetic parameters. For first parities, post-Gibbs analysis gave a posterior mean (PM) of 0.44 (0.004) for direct heritability and 0.09 (0.001) for the maternal one. For later parities, the PM for direct heritability was equal to 0.36 (0.003) and 0.06 (0.001) for the maternal. The genetic correlation between the direct and maternal GL was equal to -0.50 for first parities and -0.22 for later parities. The genetic correlation between parities was equal to 0.92 for the direct component and 0.85 for the maternal one.

Genetic trend. The genetic trend by bulls' year of birth for direct and maternal effects of GL is represented in Figure 1 and reflects their negative genetic correlation. While direct GL shows a decreasing trend, the maternal one shows a slight increase, especially for first parities over the last 20 years. The decreasing trend of the direct effect may be explained by the negative correlation of direct GL BVs with production traits' BVs (unpublished data): selection for production may have led to such trend.

Conclusions

Heritability estimates, mainly the direct components, confirmed the possibility to select for GL. The genetic correlations between parities showed that first and later parities should be treated as different traits in the genetic evaluation for GL. The genetic correlation between the direct and the maternal components showed, for both traits, the antagonistic relationship between the effect of calf and dam genotypes. The estimation of the genomic breeding values for GL is in progress. Further studies will be led to define how GL relates with other economically important traits, to identify its optimum value and to investigate the possibility to include GL in the genetic evaluation for calving traits.

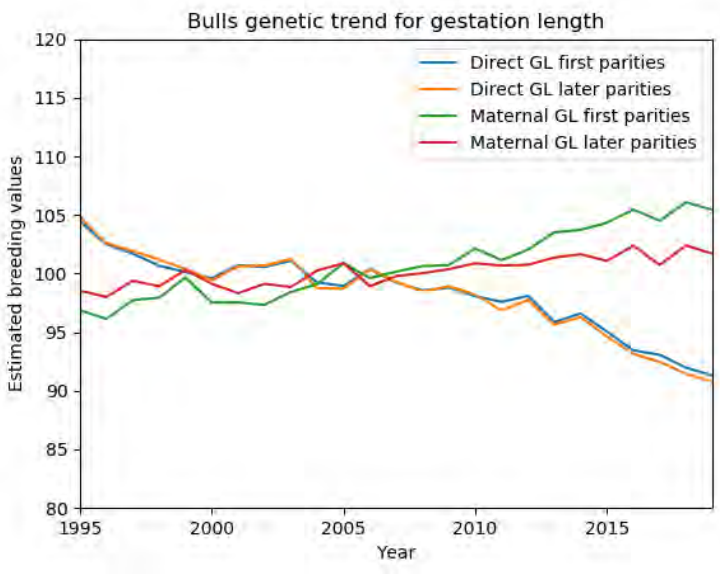


Figure 1. Direct and maternal gestation length (GL) genetic trend by year of birth.

References

Eaglen, S.A.E., M.P. Coffey, J.A. Woolliams, and E. Wall. 2013. Direct and maternal genetic relationships between calving ease, gestation length, milk production, fertility, type, and lifespan of Holstein-Friesian primiparous cows. J. Dairy Sci. 96:4015–4025. https://doi.org/10.3168/jds.2012-6229.

Lidauer, M., K. Matilainen, I. Strandèn, E. Mäntysaari, T. Pitkänen, and M. Taskinen. 2019. MiX99 : Technical reference guide for MiX99 pre-processor.

Misztal, I., S. Tsuruta, T. Strabel, B. Auvray, T. Druet, and D.H. Lee. 2002. BLUPF90 and related programs (BGF90). Commun. No. 28-07. Page in 7th World Congress on Genetics Applied to Livestick Production, Montpellier, France.

Norman, H.D., J.R. Wright, M.T. Kuhn, S.M. Hubbard, J.B. Cole, and P.M. VanRaden. 2009. Genetic and environmental factors that affect gestation length in dairy cattle. J. Dairy Sci. 92:2259–2269. https://doi.org/10.3168/jds.2007-0982.

Norman, H.D., J.R. Wright, and R.H. Miller. 2011. Potential consequences of selection to change gestation length on performance of Holstein cows. J. Dairy Sci. 94:1005–1010. https://doi.org/10.3168/jds.2010-3732.

SAS Institute Inc., Cary, NC, U. 2017. Sas® 9.4. SAS Inst. Inc., Cary, NC, USA 1–8.

114. Genetic evaluation of cystic ovarian disease in Czech Holstein cattle

E. Kašná*, L. Zavadilová and Z. Krupová

Institute of Animal Science, Přátelství 815, 10400 Prague, the Czech Republic; kasna.eva@vuzv.cz

Abstract

We estimated the genetic parameters and genomic breeding values for cystic ovarian disease (COD) resistance. We used data of 30,729 Holstein cows recorded by Czech farmers between 2017-2021. We conducted pedigree- and genomic-based evaluation. Estimated heritability was small (h^2=0.02). Breeding values were normally distributed. Their reliability increased with added genomic information evaluated with single-step genomic method. This increase was the highest in genotyped cows, where the mean reliability increased threefold. Genetic trends showed gradual improvement of resistance to cystic ovarian disease in Czech Holstein cows. The study was supported by the Ministry of Agriculture of the Czech Republic, project NAZV QK1810253 and institutional support MZE-RO0718.

Introduction

Cystic ovarian disease is economically significant ovarian dysfunction as it increases days open, treating costs and culling due to infertility (Kelton *et al.*, 1998). Its incidence has been reported to vary between 1 and 30% in dairy herds (Kelton *et al.* 1998, Vanholder *et al.* 2006). COD occurs most often during the early postpartum period when there is a transition from the non-cyclic condition (pregnancy) to the establishment of regular cyclicity. It develops when one or more follicles fail to ovulate, they do not regress but maintain growth and steroidogenesis (Vanholder *et al.*, 2006). The primary cause of ovarian cyst development is not clearly established, but the most widely accepted hypothesis is that COD results from a neuroendocrine imbalance involving the hypothalamic-hypophyseal-gonadal axis (Hooijer *et al.* 2001). The possible causes of this imbalance may include stress, herd management, nutritional status, body condition and infectious disease (Silvia *et al.* 2002).

COD is usually defined as the presence of one smooth, rounded structure of more than 20-25 mm in diameter on one or both ovaries persisted for up to 10 days without luteal tissue (Kelton *et al.* 1998). Estimates of COD heritability are usually lower than 0.10 and vary according to the data set and applied method (Jamrozik *et al.* 2021, Koeck *et al.* 2014, Heringstad 2010, Hooijer *et al.* 2001). Studies of genetic correlation showed an unfavourable association between COD and milk production traits (Hooijer *et al.*, 2001, Koeck *et al.* 2014) which indicates that selection on better milk performance would lead to increased COD frequency.

In the Czech Republic, the data on dairy cows' health traits are recorded voluntarily by farmers since July 1, 2017. Data are recorded to the national cattle health database via web application with implemented key of diagnoses based on ICAR recommendations. In 2021, a national genomic evaluation of clinical mastitis was introduced for Czech Holstein cattle (Zavadilová *et al.* 2021). The implementation of genetic-based selection of another health traits, such as reproductive disorders, is actively investigated. The objective of this study was to perform pedigree and genomic-based analysis of COD in Czech Holstein cattle using traditional BLUP and ssGBLUP (single-step genomic BLUP) approaches.

Materials & methods

Data. Data on COD were retrieved from national database of health traits (Czech Moravian Breeding Corporation). COD was scored as 0 (no case) or 1 (at least 1 case) in the period from 30 to 305 days after

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_114, © E. Kašná *et al.* 2022

calving. We included only lactations started between July 1, 2017, and October 31, 2021. Herds with lower than 1% COD incidence were excluded. The data set comprised 47,407 lactations from 30,729 Holstein cows producing in 80 herds.

Pedigree information was retrieved from the national database (Plemdat Ltd.). Genomic data were provided by the Holstein Cattle Breeders Association of the Czech Republic. Animals were genotyped with the Illumina Bovine SNP50 BeadChip (50K, Illumina Inc., San Diego, CA). The genotyped population for the single-step method included 5,394 bulls and 30,095 cows born between 1973 and 2021.

Genetic evaluation model. Single trait animal model was employed for data analysis. The model equation was:

$$y = X\beta + Z_{pe}pe + Z_a a + e, \quad (1)$$

where y is a vector of COD observations (0 – no case, 1 – at least one case in the period from 30 to 305 days after calving); β is a vector of systematic effects (including HYS = fixed effect of herd-year-season with 235 classes; Par = fixed effect of parity of cow with three classes, 1-first, 2-second, 3-third and later); pe is a vector of random cow's permanent environment effect where pe ~ $N(0,I\sigma_{pe}^2)$ with σ_{pe}^2 the permanent environment variance, a = is a vector of random additive genetic effects where a ~ $N(0,H\sigma_a^2)$ with σ_a^2 the additive genetic variance; and e is a vector of random residual errors, where e ~ $N(0,I\sigma_e^2)$, with σ_e^2 the error variance. X, Z_{pe} and Z_a are the corresponding incidence matrices. H is a modified genetic relationship matrix that combines pedigree-based A and genomic-based G relationship matrices. The significance of fixed effects was examined in the preliminary least square analysis (procedure GLM; SAS 9.4.). The variance components estimation was conducted using an AIREML approach; conventional breeding values (EBV) were estimated using the BLUP approach; genomic breeding values (GEBV) were predicted using the ssGBLUP approach, and reliabilities calculated with ACCF90GS program, as implemented in the BLUPF90 family of programs (Misztal *et al.* 2018).

Results

The incidence of COD in Czech Holstein cows was 8.1%. The herd incidence varied between 1.0 and 32.0%. The incidence was lower in primiparous (6.6%) compare to older (8.9%) cows ($P<0.001$). The incidence was higher

Variance components estimation gave phenotype variance $\sigma_p^2=0.069$, and additive genetic variance $\sigma_a^2=0.001$, which led to the heritability $h^2=0.02$. The variance due to permanent environmental effect was twice as high as the additive genetic variance. EBV distribution was close to normal with mean 0.008 and SD 0.012 for EBV and 0.202 for GEBV.

Mean reliabilities of breeding values (Table 1) were higher with the inclusion of genomic information. The highest increase was achieved in genotyped cows whose mean reliability of genomic breeding values increased threefold compare to conventional breeding values. For genomic sires, genomic reliability was 4 p.p. higher. For cows with phenotype, the increase in reliability was only minor.

GEBV were expressed as relative breeding values (RBV) with mean 100 and SD 12, for base bulls born 2010. Genetic trends (Figure 1) based on RBV with reverse sign (higher values mean better resistance to COD) showed gradual improvement in cows since 2012. For bulls, RBV fluctuated and did not show a clear trend.

Table 1. Reliability of breeding values.

		Conventional breeding values			Genomic breeding values		
Sample	**N**	**Mean**	**Min.**	**Max.**	**Mean**	**Min.**	**Max.**
Genomic cows	30,095	0.04	0.00	0.25	0.13	0.00	0.35
Genomic bulls	5,394	0.05	0.00	0.91	0.09	0.00	0.91
Cows with phenotype	30,729	0.08	0.00	0.26	0.08	0.00	0.31

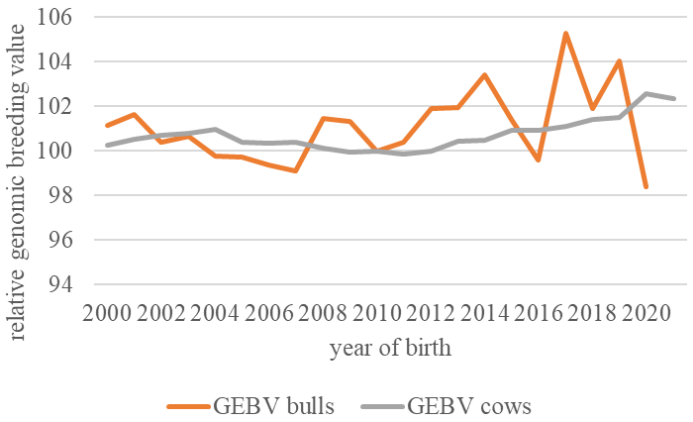


Figure 1. Genetic trend for COD in Czech Holstein bulls and cows.

Discussion

Previous studies of COD resistance (Hooijer *et al.* 2001, Heringstad 2010, Koeck *et al.* 2014, Jamrozik *et al.* 2021) showed, that it has a genetic background and thus could be enhanced by selection. A common problem is the lack of reliable data on COD occurrence. In the Czech Republic, the data are recorded by farmers on voluntarily basis for five consecutive years. Our study conducted on this data showed, that the incidence of COD, its variance components and heritability were comparable with other population studies of Holstein cattle (Vanholder *et al.* 2006, Koeck *et al.* 2014, Jamrozik *et al.* 2021). However, the reliability of EBV estimated with conventional BLUP method was not satisfying because of low heritability and low number of records. One way to increase the reliability of EBV is to incorporate genomic data (Parker Gaddis *et al.* 2014). For that we applied the single-step genomic method, that permits simultaneous utilization of phenotypic, full pedigree and genomic information (Aguilar *et al.* 2010). Genomic information helped to increase the mean reliability of GEBV mainly in the group of genotyped cows that missed phenotype records. In groups with phenotype records (own record, daughters' record) the improvement was lower.

Genetic trends for COD are the result of indirect selection response. Hoojier *et al.* (2001) reported an unfavourable genetic correlation between COD and milk production traits, which implies higher occurrence of COD in cows selected for higher milk production. Jamrozik *et al.* (2021) found unfavourable correlation of COD with length of productive life, protein yield, and several conformation traits (mammary system, rear attachment width, fore attachment), but favourable correlation with fertility and calving performance traits. Some of these traits are also incorporated in the Czech Holstein selection index with weights 28% (protein yield), 15% (fertility), 13% (udder conformation) and 5% (longevity).

Our study showed, that farm-collected health data are useful for routine genomic evaluation and selection of Czech Holstein cattle for better COD resistance.

References

Aguilar I., Misztal I., Johnson D.L., Legarra A., Tsuruta S. *et al.* (2010) J. Dairy Sci. 93(2):743-752. https://doi.org/10.3168/jds.2009-2730

Heringstad B. (2010) J. Dairy Sci. 93(6):2751–2756. https://doi.org/10.3168/jds.2009-2879

Hooijer G.A., Lubbers R.B., Ducro B.J., van Arendonk J.A., Kaal-Lansbergen L.M., *et al.* (2001) J. Dairy Sci. 84(1):286-91. https://doi.org/10.3168/jds.S0022-0302(01)74478-5

Jamrozik J., Kistemaker G.J., Van Doormaal B.J., Baes C.F., and Miglior F. (2021) Proc. of the 2021 Interbull Meeting, Leeuwarden, the Netherlands.

Kelton D.F., Lissemore K.D., and Martin R.E. (1998) J. Dairy Sci. 81(9):2502-2509. https://doi.org/10.3168/jds.S0022-0302(98)70142-0

Koeck A., Loker S., Miglior F., Kelton D.F., Jamrozik J., *et al.* (2014) J. Dairy Sci. 97(9):5806-5813. https://doi.org/10.3168/jds.2013-7785

Misztal I., Tsuruta S., Lourenco D.A.L., Masuda Y., Aguilar I., *et al.* (2018) Manual for BLUPF90 family programs. University of Georgia. Available at: http://nce.ads.uga.edu/wiki/doku.php?id=documentation

Parker Gaddis K.L., Cole J.B., Clay J.S. and Maltecca C. (2014) J. Dairy Sci. 97(5):3190-3199. https://doi.org/10.3168/jds.2013-7543

Silvia W.J., Hatler T.B., Nugent A.M., and Laranja da Fonseca L.F. (2002) Domestic Animal Endocrinology 23(1-2):167–177. https://doi.org/10.1016/S0739-7240(02)00154-6

Vanholder T., Opsomer G., and De Kruif A. (2006) Reprod. Nutr. Dev. 46:105-119. https://doi.org/10.1051/rnd:2006003

Zavadilová L., Kašná E., and Krupová Z. (2021) Odhad genomických plemenných hodnot na zvýšení odolnosti vůči klinické mastitidě u holštýnského skotu. Institute Of Animal Science, Prague, Czech Republic.

115. Entropy of broiler activity: individual variation and consistency

M. van der Sluis[1,2*], L. Asher[3], T.B. Rodenburg[1,2], Y. de Haas[1], B. de Klerk[4] and E.D. Ellen[1]

[1]Wageningen University & Research, P.O. Box 338, 6700 AH Wageningen, the Netherlands; [2]Utrecht University, 3508 TD Utrecht, the Netherlands; [3]Newcastle University, NE1 7RU, Newcastle upon Tyne, United Kingdom; [4]Cobb Europe, 5831 GH Boxmeer, the Netherlands; malou.vandersluis@wur.nl

Abstract

Animal behaviour is complex and comparing average levels of behavioural activities is sometimes insufficient to pick up on behavioural differences between (groups of) animals. Entropy, a measure of the randomness or regularity of time series, might help us to describe aspects of behaviour better. In this study, we determined daily entropy in individual broiler activity levels over time, based on a time series of observed activity per 15 minutes across a day, to assess individual variation and consistency in entropy of activity. Activity data for calculation of entropy were available for 79 broilers for a maximum of 21 days. We observed individual variation in entropy between broilers, but the level of entropy was not very consistent across days. The individual variation in entropy indicates potential for future research to look into whether and how entropy differences relate to other traits in broilers.

Introduction

Animal behaviour is complex and multi-dimensional (Asher *et al.*, 2009) and studies have indicated that mean levels of behaviours are not always sensitive to differences between groups of animals. For example, Dawkins *et al.* (2012) studied broiler flocks using optical flow patterns and observed no correlation between the mean gait score (i.e. walking ability) of the flock and the mean optical flow, but did observe correlations between the mean gait score and the skew and kurtosis of optical flow. This indicates that descriptors of behaviours that capture more than mean behavioural activity levels have potential to provide more insight into (individual) differences in behaviour. One measure that has potential is entropy. Entropy is a measure of the randomness or regularity of time series based on the existence of patterns (Delgado-Bonal and Marshak, 2019) and it has been indicated that entropy can be informative for, or predictive of, human behaviour, health and well-being (e.g. Glenn *et al.* (2006)). Although to a smaller extent, entropy has also been studied, and found to be informative, in non-human animals (e.g. Stamps *et al.* (2013)). In this study, we aimed to describe individual broiler locomotor activity patterns over time and assess individual differences in these patterns, using entropy. Individual differences in entropy have potential to be linked to other traits in broilers and may in the future aid in individual broiler phenotyping.

Materials & methods

Animals and housing. Individual-level broiler activity data were collected for 24 hours per day, from 1 to 33 days old, in a group of 80 birds housed in a pen with a size of 1.8×2.6 m.

Activity tracking. To record the individual broilers' activity levels, we fitted the pen with an earlier-validated RFID activity tracking system (Van der Sluis *et al.*, 2020). This system consisted of tags, that could be fitted to the birds' legs, and a grid of antennas that was placed underneath the floor of the pen. These antennas could register the presence of tags and a log file was stored with information on the tag ID (i.e. bird ID), the time of registration and the antenna (i.e. location) at which the tag was registered. To calculate activity levels, we first determined for every minute whether an antenna switch (i.e. position change) was registered within this minute for this animal. We then calculated the number of minutes in which an antenna switch was registered per 15-minute bin. Subsequently, these values were categorized into four classes based on the quartiles observed in the data. Class one consisted of the observations up to

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_115, © M. van der Sluis *et al.* 2022

and including the largest value observed in the lowest 25% of the data. Class two consisted of observations up to and including the largest value observed in the lowest 50% of the data, that were not already in class one. The upper limit for the third class was the 75% quartile, and the upper limit for the fourth class was the 100% quartile. In this process, we excluded values of zero (those were added as a fifth class), as a large part of the data consisted of zeroes due to low activity levels during the night. As activity is known to decrease with age in broilers (e.g. Weeks *et al.,* 2000), quartiles were determined within each day instead of across the full tracking period (Figure 1). Days on which there were overlapping numbers of antenna switches across the quartiles (e.g. one antenna switch was identified as both the first and second quartile limit) were excluded. This was the case for 28-days-old and older.

Entropy calculation. The resulting classes for the 15-minute bins of activity across each day were used as the underlying time series for the sample entropy (SampEn) calculation. The SampEn values can therefore be interpreted as a measure of regularity in activity within days, with lower values indicating more regularity and higher values indicating more randomness. However, SampEn can only be accurately calculated for days without missing data and therefore days on which the birds were weighed or their leg tags were checked, were excluded. This resulted in 21 days of data available for SampEn calculation, for which the pracma package (Borchers, 2021) was used in R (R Core Team, 2021). A total of 1455 entropy values were available, for a total of 79 individuals.

Statistics. All statistics were performed in R. Descriptive statistics were used to obtain a first overview of individual entropy values over time. To determine whether the individual entropy levels were consistent over time (i.e. the same individuals had relatively high or low entropy values on different days), we calculated Spearman rank correlations between the entropy values on different days, using the Hmisc package (Harrell Jr *et al.*, 2021). Visualisations were made using the ggplot2 (Wickham, 2016) and corrplot (Wei and Simko, 2021) packages. The level of statistical significance was set at 0.05.

Results

The overall mean entropy was 0.96 (SD 0.25), with a range from 0.07 to 1.65. Over time, the median entropy remained similar, but there did appear to be individual differences in entropy, as shown in Figure 2, where quite a large range of entropy values can be seen. The ranking of the entropy values of individual animals over days was not very consistent (Figure 3).

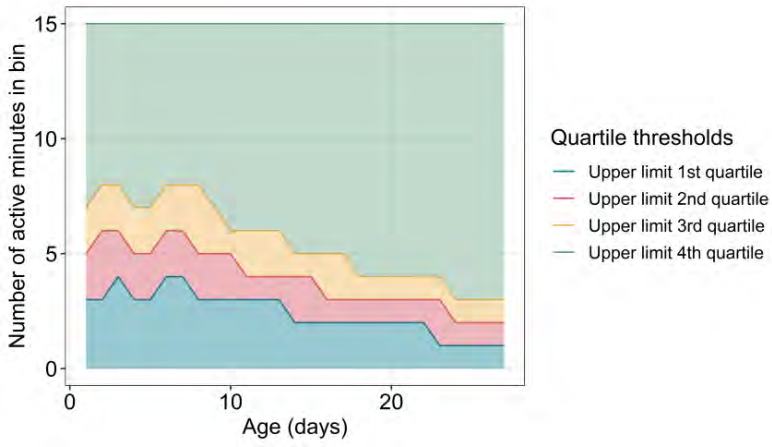


Figure 1. Quartile thresholds over time for number of active minutes per 15-minute bin.

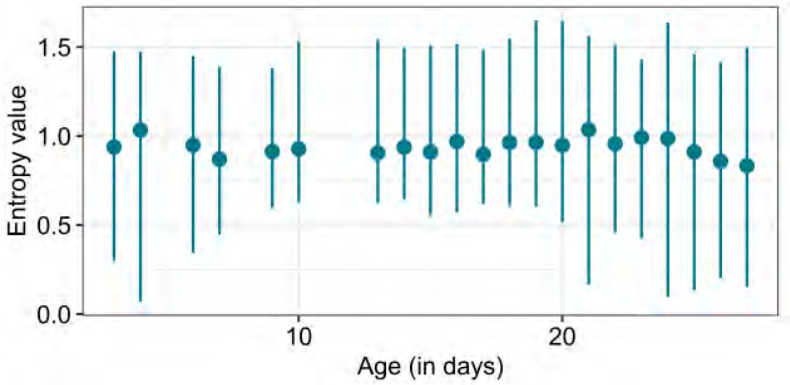


Figure 2. Median entropy values (circles) over time. Bars show the range from the minimum to the maximum observed entropy value for that specific day.

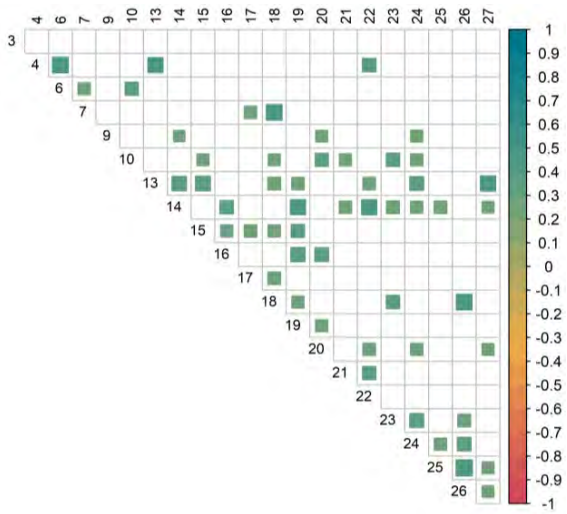


Figure 3. Spearman rank correlations between entropy values of individual animals on different days. On the axes, the day number is shown. Colours and sizes of squares indicate the strength of the correlation. Only correlations with $P<0.05$ are shown, squares associated with non-significant correlations are left empty.

Discussion

In this study, we examined entropy of activity of individual broilers over time and assessed whether there are consistent individual differences in entropy. We observed that the median entropy in activity from 3 to 27 days old was similar, but that there was quite a large range of observed entropy values within a day, suggesting that there are individual differences in entropy. This indicates potential for future research to look into whether and how such entropy differences are related to other traits. Other studies have indicated that differences in entropy can be linked to, for example, animal health. Fasmer and Johansen (2016) observed that spontaneously hypertensive rats displayed a higher complexity of movement time series (that is, a higher entropy) than control rats. However, the lack of strong correlations in entropy values over time observed here suggests that these differences are not consistent. Studies in laying hens have observed quite some consistency in movement patterns for some individuals, when looking at transitions

between five zones in a commercial aviary (Rufener *et al.*, 2018). In this study, however, the activity patterns were simplified considerably by looking at 15-minute bins and five activity categories, and no locations or directions of movement were taken into account. Possibly, if a more detailed overview of activity would have been used as the underlying time series, more consistency in activity or movement patterns would have been observed. On the other hand, the lack of consistency in entropy might be interesting too: perhaps changes in entropy are related to specific traits or environmental conditions of the broilers, but this requires further investigation. Overall, it would be worth exploring different ways of applying entropy to animal behaviour data in the future, to understand how to best capture the behaviour feature of interest and to examine possible relationships with other traits.

Funding

This study was financially supported by the Dutch Ministry of Economic Affairs (TKI Agri and Food project 16022) and the Breed4Food partners Cobb Europe, CRV, Hendrix Genetics and Topigs Norsvin.

References

Asher L., Collins L.M., Ortiz-Pelaez A., Drewe J.A., Nicol C.J. *et al.* (2009) J. R. Soc. Interface 6(4):1103-1119. https://doi.org/10.1098/rsif.2009.0221

Borchers H.W. (2021) pracma: Practical Numerical Math Functions. R package version 2.3.3. https://CRAN.R-project.org/package=pracma

Dawkins M.S., Cain R., and Roberts S.J. (2012) Anim. Behav. 84(1):219–223. https://doi.org/10.1016/j.anbehav.2012.04.036

Delgado-Bonal A., and Marshak A. (2019) Entropy 21(6):541. https://doi.org/10.3390/e21060541

Fasmer O.B., and Johansen E.B. (2016) Behav. Brain Funct. 12:32. https://doi.org/10.1186/s12993-016-0117-9

Glenn T., Whybrow P.C., Rasgon N., Grof P., Alda M., *et al.* (2006) Bipolar Disord. 8(5):424-429. https://doi.org/10.1111/j.1399-5618.2006.00373.x

Harrell Jr F.E., contributions from Dupont C. and many others. (2021) Hmisc: Harrell Miscellaneous. R package version 4.5-0. https://CRAN.R-project.org/package=Hmisc

R Core Team (2021) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.

Rufener C., Berezowski J., Maximiano Sousa F., Abreu Y., Asher L. *et al.* (2018) Sci. Rep. 8:12303. https://doi.org/10.1038/s41598-018-29962-x

Stamps J.A., Saltz J.B., and Krishnan V.V. (2013) Anim. Behav. 86(3):641-649. https://doi.org/10.1016/j.anbehav.2013.07.012

van der Sluis M., de Haas Y., de Klerk B., Rodenburg T.B., and Ellen E.D. (2020) Sensors 20(13):3612. https://doi.org/10.3390/s20133612

Weeks C.A., Danbury T.D., Davies H.C., Hunt P., and Kestin S.C. (2000) Appl. Anim. Behav. Sci. 67(1-2):111-125. https://doi.org/10.1016/S0168-1591(99)00102-1

Wei T., and Simko V. (2021) R package 'corrplot': Visualization of a Correlation Matrix (Version 0.90). https://github.com/taiyun/corrplot

Wickham H. (2016) Ggplot2: elegant graphics for data analysis. Springer. https://doi.org/10.1007/978-3-319-24277-4

116. Laying hens tweet: they also lead and follow – implications for genetic selection against gregarious nesting

L. Bécot[1,2*], N. Bédère[2], T. Burlot[1] and P. Le Roy[2]

[1]NOVOGEN, 5 rue des Compagnons 22960 Plédran, France; [2]PEGASE, INRAE, Institut Agro, 16 Le Clos 35590 Saint-Gilles, France; lorry.becot@inrae.fr

Abstract

Gregarious nesting occurs when a hen chooses a nest already occupied while other nests are vacant, increasing welfare and production problems. This study aims to propose two predictors of gregarious nesting: weighted indegree (WID) and weighted outdegree (WOD), which are the number of times a hen entered an individual nest next to a vacant nest or an occupied nest, respectively. Nesting behaviour was recorded using individual electronic nests for 210 days. Genetic parameters for WID, WOD and nesting behaviour traits were estimated for 1,268 and 1,080 Rhode Island Red and White Leghorn respectively. For both lines, heritability estimates were high for WID (0.40/0.53) and low for WOD (0.12/0.15). Moderate genetic correlations were estimated between WID and WOD (+0.26/-0.35) and strong genetic correlations with some nesting behaviour traits. WID and WOD could help to select against gregarious nesting but the effects for hens using collective nests remain to be evaluated.

Introduction

In the EU, the majority of laying hens are now raised in cage-free systems, such as barn, free-range, and organic (52% in 2020; European Commission, 2021). In these systems, hens lay their eggs in collective nests, allowing them to express natural nesting behaviours (Cronin *et al.*, 2012). Gregarious nesting can occur when a hen chooses a nest already occupied while other nests are vacant (Riber, 2010). This behaviour can lead to mortality with suffocation in the nest and the increase of broken eggs and floor eggs (van den Oever *et al.*, 2021). Selecting hens against gregarious nesting could help to reduce welfare problems and to increase egg production in cage-free systems. However, this behaviour is difficult to record on large scale. Currently, little is known about its genetic background. The development of electronic nests continuously recording the nesting behaviour of hens raised in groups, could make it possible to identify predictors of gregarious nesting (Bécot *et al.*, 2021a). This study aims to propose predictors of gregarious nesting, estimate their heritability and their genetic correlations with nesting behaviour traits.

Materials & methods

Birds. Two laying hens' pure lines – Rhode Island Red (RIR) and White Leghorn (WL) – belonging to the NOVOGEN company (Plédran, France) were mobilised. For RIR, the phenotypes were recorded between 2018 and 2021 on 1,268 hens (three batches) issued from 144 sires and 506 dams. For WL, the phenotypes were recorded between 2019 and 2021 on 1,080 hens (two batches) issued from 100 sires and 372 dams. For both lines, more than 5 generations were used in the pedigree to estimate genetic parameters. For each line and batch, birds were raised in a barn on the floor until 17 weeks of age. Birds were then transferred and raised in another barn until 64 weeks of age. This barn was composed of floor pens (one by line) and was equipped with individual electronic nests overlaid on two rows (about 1 nest for 5 hens). Hens were raised with cocks (about 1 cock for 10 hens).

Recording system. Nesting behaviour was recorded continuously by radio-frequency identification and using individual electronic nests developed by NOVOGEN. Hens were individually recognised when they visited a nest by a transponder tagged on their leg. Thus, the hen's identification, the localisation of the nest visited, the time of entry in the nest, and the time of exit from the nest were recorded for each visit.

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_116, © L. Bécot *et al.* 2022

Oviposition time was recorded with an egg sensor located behind each nest, allowing us to know if the hen had laid an egg. Data were recorded between 200 and 409 days of age. Thus, 233,127 and 198,760 visits with oviposition were collected for RIR and WL, respectively.

Traits. The recording period was divided into seven 30-days periods. The pairs of hens using a nest for laying at the same time, in nests separated by a horizontal distance of 0 or 1 nest (0 for those at the same horizontal position, on top and bottom row) were selected to calculate predictors of gregarious nesting. This represented on average 2,617 and 3,486 pairs by day for RIR and WL respectively. For each hen of the pairs, the first who was entered in the nests was considered as a leader and the second as a follower. Using this information, two social network traits were calculated by hen as predictors of gregarious nesting: weighted indegree (WID) and weighted outdegree (WOD; Bondy and Murty, 1976). WID was the number of times the hen was the leader. WOD was the number of times the hen was the follower. Finally, hens with at least 15 eggs laid in the nests for a given 30-days period were kept for analyses, as hens with fewer eggs had fewer opportunities to express gregarious nesting. Thus, 8,137 and 7,124 phenotypes respectively associated with 1,268 RIR and 1,080 WL were kept for analyses (Table 1).

Five nesting behaviour traits were then calculated for each of these hens and each 30-days periods. The laying rate in the nests (LRN) was calculated to represent the egg production in the nests. For a given 30-days period, LRN was the egg number laid in the nests divided by 30 and then multiplied by 100. The mean oviposition time (MOT) and the logarithm of the variance of oviposition time (LVOT) were calculated to characterize the laying rhythm, using a procedure described in Bécot *et al.* (2021b). To consider the pre-laying behaviour referring to the nest choice for laying, the mean distance between nests of laying (MDN) was calculated. The distance between nests of laying corresponded to the horizontal distance, in the number of nests, which separated the nests of laying for two consecutive days. To avoid measuring different behaviours, this distance was not calculated for the laying separated by one or more days without laying in the nests. The mean laying duration (MLD) was the mean of the time spent in the nest for visits with oviposition. The time spent in the nest was calculated by the difference between the time of exit from the nest and the time of entry in the nest. The summary statistics of the 7 traits are given in Table 1. All traits were standardized to facilitate the convergence of models.

Model. Genetic parameters and phenotypic correlations were estimated using the restricted maximum likelihood method applied to a multi-trait animal model. The environmental effects of the model included the fixed effect of the hen's hatch date and the random effect of the common environment of the hen.

Table 1. Summary statistics of phenotypic data.[1]

	Rhode Island Red				White Leghorn			
Trait	**Mean**	**Std**	**Min**	**Max**	**Mean**	**Std**	**Min**	**Max**
WID	64.13	40.63	1.00	317.00	92.45	47.73	5.00	425.00
WOD	64.36	24.53	0.00	134.00	92.60	24.76	2.00	137.00
LRN (%)	94.57	7.56	50.00	100.00	92.11	8.33	50.00	100.00
MOT (hh:mm)	02:47	01:23	-00:23[a]	08:12	03:36	01:05	-00:12[a]	07:31
LVOT (sq. h)	-0.62	1.42	-6.28	3.29	-0.15	1.25	-4.14	3.54
MDN (nests)	7.70	4.09	0.12	31.50	10.55	4.64	1.48	33.17
MLD (min)	41.23	18.99	6.54	123.80	58.75	27.99	9.17	193

[1] WID = weighted indegree; WOD = weighted outdegree; LRN = laying rate in the nests; MOT = mean oviposition time; LVOT = logarithm of the variance of oviposition time; MDN = mean distance between nests of laying; MLD = mean laying duration. [a] MOT before the lights were turned on.

Variance and covariance components were estimated for the 7 traits together with REMLF90 (Misztal *et al.*, 2002). The estimated values were used as starting points to estimate standards errors with AIREMLF90 (Misztal *et al.*, 2002).

Results & discussion

Heritability estimates. Heritability estimates for WID and WOD were high (0.40) and low (0.12) for RIR respectively (Table 2). Similar results were found for WL with heritability estimates from 0.53 and 0.15 for WID and WOD respectively (Table 3). To our knowledge, no heritability estimates for these traits were reported in the literature. However, van den Oever *et al.* (2021) observed a different expression of gregarious nesting (considering the number of eggs per nest) between five broiler breeders lines and suggested the influence of genetic background.

Genetic correlations. Moderate genetic correlations, from +0.26 and -0.35, were estimated between WID and WOD for RIR and WL respectively (Table 2 and 3). Counterintuitively, WID and WOD are not genetically opposite and would provide different information linked to gregarious nesting.

For RIR, strong genetic correlations were estimated between WID and the traits MOT (-0.61), LVOT (-0.51) and MLD (+0.78). For WL, strong genetic correlations were estimated between WID and MLD (+0.87). For both lines, these results suggest selecting the leader hens, who have a high breeding value for WID, would increase MLD. This is not desirable as high MLD increases the occupation rate of the nests and therefore the risks for some hens to lay floor eggs (Bécot *et al.*, 2021a).

Table 2. Genetic parameters and phenotypic correlations[a] for Rhode Island Red.[1]

Trait	h^2	r_g / r_p WID	r_g / r_p WOD
WID	0.40 (0.05)	1 / 1	
WOD	0.12 (0.03)	0.26 (0.15) / 0.33	1 / 1
LRN	0.07 (0.02)	0.02 (0.28) / 0.21	0.69 (0.42) / 0.37
MOT	0.48 (0.06)	-0.61 (0.07) / -0.55	-0.13 (0.15) / -0.22
LVOT	0.19 (0.03)	-0.51 (0.10) / -0.34	-0.36 (0.15) / -0.30
MDN	0.19 (0.04)	0.14 (0.14) / 0.15	0.20 (0.19) / 0.11
MLD	0.46 (0.06)	0.78 (0.05) / 0.66	0.02 (0.16) / 0.00

[1] Traits abbreviations are described in Table 1. [a] Heritability estimates (h^2) for all traits. Genetic (r_g) and phenotypic (r_p) correlations between predictors of gregarious nesting (WID and WOD) and nesting behaviour traits. Standard errors are in parenthesis.

Table 3. Genetic parameters and phenotypic correlations[a] for White Leghorn.[1]

Trait	h^2	r_g / r_p WID	r_g / r_p WOD
WID	0.53 (0.06)	1 / 1	
WOD	0.15 (0.04)	-0.35 (0.15) / 0.06	1 / 1
LRN	0.09 (0.03)	-0.18 (0.25) / 0.22	0.36 (0.23) / 0.47
MOT	0.39 (0.06)	-0.33 (0.11) / -0.38	0.43 (0.14) / 0.37
LVOT	0.21 (0.04)	0.02 (0.13) / -0.18	-0.22 (0.16) / -0.31
MDN	0.24 (0.05)	0.14 (0.14) / 0.04	-0.08 (0.18) / -0.05
MLD	0.70 (0.07)	0.87 (0.03) / 0.73	-0.38 (0.14) / -0.11

[1] Traits abbreviations are described in Table 1. [a] Heritability estimates (h^2) for all traits. Genetic (r_g) and phenotypic (r_p) correlations between predictors of gregarious nesting (WID and WOD) and nesting behaviour traits. Standard errors are in parenthesis.

Strong genetic correlations were estimated between WOD and LRN (+0.69; but with a high standard error) for RIR or between WOD and MOT (+0.43) for WL. This suggests selecting the not follower hens, who have a low breeding value for WOD, would decrease LRN and increase MOT for RIR and WL respectively.

For both lines, low genetic correlations, ranging from -0.08 and +0.14, were estimated between MDN and the traits WID and WOD, except between WOD and MDN for RIR (+0.20). Some studies suggest that hens laying eggs in the same nests, i.e. with a low MDN, have a gregarious nesting (Riber, 2010; van den Oever *et al.*, 2021). MDN would be complementary to WID and WOD, each of them capturing a different part of the genetic variance for gregarious nesting.

Selecting the leader and the not follower hens, as well as hens laying their eggs in nests far from each other (high breeding value for MDN), could help to reduce gregarious nesting. However, this behaviour appears complex because it is influenced by multiple factors currently poorly known (Riber, 2012). The strong and unfavourable genetic correlations for WID and WOD with some nesting behaviour traits suggest to be careful before use such traits in breeding programmes. The real response to selection of WID and WOD to reduce gregarious nesting must be evaluated.

Conclusions

This study highlighted that WID and WOD are heritable traits in a RIR and a WL lines. Such traits could be interesting to reduce gregarious nesting by selection. However, strong and unfavourable genetic correlations were estimated with MLD and LRN, according to the trait and the line. Some studies are needed to better understand the potential effects for selection of WID and WOD, especially consequences on gregarious nesting of commercial hens raised with collective nests.

References

Bécot L., Bédère N., and Le Roy P. (2021a) INRAE Prod. Anim. 34(1):1-14. https://doi.org/10.20870/productions-animales.2021.34.1.4680

Bécot L., Bédère N., Burlot T., Coton J., and Le Roy P. (2021b) PLOS ONE 16:e0251037. https://doi.org/10.1371/journal.pone.0251037

Bondy A., and Murty U. S. R. (1976) Graph theory with applications. Macmillan, London, UK.

Cronin G.M., Barnett the late J.L., Hemsworth P.H. (2012) Anim. Prod. Sci. 52:398-405. https://doi.org/10.1071/AN11258

European Commission (2021) Market situation for eggs.

Misztal I., Tsuruta S., Strabel T., Auvray B., Druet T *et al.* (2002) Proc. of the 7th WCGALP, Montpellier, France.

Riber, A. B. (2010) App. Anim. Behav. Sci. 123(1-2):24-31. https://doi.org/10.1016/j.applanim.2009.12.016

Riber, A. B. (2012) App. Anim. Behav. Sci. 138(1-2):70-78. https://doi.org/10.1016/j.applanim.2012.01.009

Van den Oever A. C. M., Kemp B., Rodenburg T. B., van de Ven L. J. F., and Bolhuis J. E. (2021) Anim. 15(1):1-7 https://doi.org/10.1016/j.animal.2020.100030

117. Analysis of social behaviors in large groups: simulation and genetic evaluation

Z. Wang*, H.P. Doekes and P. Bijma

Wageningen University & Research, P.O. Box 338, 6700 AH Wageningen, the Netherlands; zhuoshi.wang@wur.nl

Abstract

Harmful social interactions, such as injurious feather pecking in poultry and tail biting in swine, reduce both animal welfare and efficiency. While these traits are heritable, application of breeding is still limited due to the lack of proper genetic models and precise phenotyping methods for large groups. In the near future, large scale longitudinal data on social interactions will become available thanks to developments in computer vision and artificial intelligence. Here we present models to simulate and analyze such data, which are an extension of the classic social genetic model. Latent traits were defined to represent the tendency of individuals to be engaged in behavioral interactions, distinguishing performer and recipient effects. Binary interaction records were simulated and subsequently analyzed using generalized linear mixed models. Results show that high accuracies of estimated breeding values can be obtained (0.4-0.7), despite the low observed-scale heritability of the binomial trait (0.05-0.2). We conclude that our model can be promising for breeding value estimation for social traits in large groups.

Introduction

Livestock regularly engage in behavioral interactions, and modern group-housing systems may increase the occurrence of harmful social behaviors, such as injurious feather pecking in poultry and tail biting in swine. To reduce the occurrence of harmful social behaviors, management strategies have been proposed (Van Krimpen *et al.*, 2016; De Vries *et al.*, 2004). However, the application of breeding strategies to behavioral traits is still limited.

One challenge of applying breeding strategies is that the social behavior of an individual may depend not only on the effect of the genes of the individual itself (known as the direct genetic effect, DGE), but also on that of its group mates (known as the indirect genetic effect, IGE). In the last two decades, genetic models for social traits have been developed to estimate DGE and IGE (Muir, 2005; Bijma *et al.*, 2007; Ellen *et al.*, 2014). These models assume that each individual has the opportunity to affect the phenotype of each of its pen mates equally. Therefore, these models are particularly useful in small groups, in which animals have ample interactions with all pen mates and the strength of social interaction can be assumed similar between all individuals.

For large groups the assumption of equal interaction does not hold anymore. In recent years, social network analysis has been successfully applied to large animal populations (Krause *et al.*, 2021; Radersma 2021). Social network analysis requires detailed records of the social interactions, including time, physical position, duration, and performer and recipient identity. Recording every interaction event between individuals in a large population used to be time- and labour-demanding, but is gradually becoming feasible due to developments in animal detection and tracking technologies. With these technologies, distinguishing performer and recipient effect is possible. Therefore, the terms performer and recipient effect will be used in the rest of this paper instead of DGE and IGE. It is expected that in the near future these detailed records of social interaction will become available. However, there is still a lack of methods to translate such data into estimates of individual performer effect and recipient effect.

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_117, © Z. Wang *et al.* 2022

In this study, we: (1) extended the social effect model for genetic analysis of large-scale longitudinal data on animal social behaviors; and (2) evaluated the effect of number of individuals and number of social interactions on accuracy of breeding values (BV) estimated with this model. First, latent traits were defined to represent the tendency of individuals to be engaged in behavioral interactions, distinguishing performer and recipient effects. Second, social interactions with known performer and recipient were simulated under various population settings based on an assumed genetic structure. Last, statistical models were applied to estimate BV of social traits from the simulated records, and to evaluate the accuracies of the EBV.

Materials & methods

Trait definition. For each individual, two latent traits were defined to represent its liability to be engaged in a social interaction as a performer (trait α) or a recipient (trait β). Each trait consisted of two parts: a heritable effect (the BV) and a permanent environmental effect. The latter was used to account for similarity between repeated records on the same individual.

$$P_{\alpha,i} = \mu + A_{\alpha,i} + Ep_{\alpha,i} \quad (1)$$

$$P_{\beta,i} = \mu + A_{\beta,i} + Ep_{\beta,i} \quad (2)$$

where $P_{\alpha,i}$ and $P_{\beta,i}$ were the individual's normally-distributed liability for each trait, $A_{\alpha,i}$ and $A_{\beta,i}$ were the individual's BVs, and $Ep_{\alpha,i}$ and $Ep_{\beta,i}$ were the permanent environmental effects. For the results presented here, we assumed that both traits were independent (genetic correlation of 0) and had the same genetic variance ($\sigma_{A\alpha}^2 = \sigma_{A\beta}^2$), and that the permanent environmental variances were equal to the corresponding genetic variances.

For interaction between individuals i and j (see section 'Simulations' for conditions related to their proximity), the probability of i performing the social interaction towards j is given by:

$$p_{ij} = \text{logistic}(P_{\alpha,i} + P_{\beta,i}) = 1/\ [1+\exp(-P_{\alpha,i} - P_{\beta,i})] \quad (3)$$

Thus, the logistic function was used to rescale $(P_{\alpha,i} + P_{\beta,i})$ from the real number domain of the liability to the probability domain of 0 to 1. Binary interaction records were then generated by sampling a random number from a Bernoulli distribution, where 1 meant the interaction took place, while 0 meant it did not take place.

Simulations. For social interactions to occur, animals need to be in each other's proximity. In this study, we simulated the dynamic physical position of animals through agent-based modelling, using three behaviors that were alternated: eating (at feeders in the pen), resting and walking. Only when animals encountered each other, a 0/1 interaction record was generated as described above. We generated 2000 individuals from 100 sires and 400 dams, and each dam produced 5 offspring. Simulations were run for 20 pens, each with 100 random individuals, and until a total of 10,000 interactions took place in a pen. Given a simulated mean interaction probability of 1% (E(p) = 0.01), this implied a total of ~1,000,000 encounters per pen.

Finding realistic input values. Observed-scale heritabilities for binomial traits related to animal behavior are typically in the range of 0.05 to 0.20. Kjaer and Sørensen (1997), for example, found heritabilities in this range for number of pecks performed, based on an average of ~25 interactions per individual. To find realistic input values for our simulations, we simulated a dataset where individuals performed on average 25 pecks, then estimated the observed-scale heritability with ordinary linear mixed models, and tuned the input value for the genetic variance until observed-scale heritability was either 0.05, 0.1 or 0.2 (Table 1).

Breeding value estimation. To estimate BV for the performer effect (α) and the recipient effect (β), a generalized linear mixed model with a logit link-function and a binomial distribution was fitted in ASReml 4.1 (Gilmour *et al.*, 2015):

$$\text{logit}(p_{ij}) = \mu + Pen_k + A_{\alpha,i} + A_{\beta,i} + Ep_{\alpha,i} + Ep_{\beta,i} \quad (4)$$

where p_{ij} is the probability that i performs a social interaction towards j, Pen_k is a fixed pen effect, A is the BV and Ep is the permanent environment effect.

Results

Table 1 shows the genetic variances (σ_A^2) that correspond to observed scale heritabilities of total number of interactions performed (h_o^2) of 0.02, 0.05 and 0.1. These genetic variances correspond to interaction probabilities of top and bottom ranking individuals ($\pm 2\sigma_A$) that differed by a factor of 1.5 for $h_o^2 = 0.05$, 2.5 for $h_o^2 = 0.10$, and 5.5 for $h_o^2 = 0.20$. Hence, despite the low observed-scale heritability, the genetic differences between individuals in their tendency to engage in social interactions were very large.

To examine the relationship between the number of individuals and the accuracy of EBV, we included 5, 6, …, up to 20 pens in the genetic analysis (thus 500, 600, …, 2,000 individuals). When 500 individuals were included, the GLMM yielded an accuracy of EBV of 0.30, 0.36 and 0.44, given heritabilities of 0.05, 0.10 and 0.2, respectively (Figure 1a). When all the 2,000 individuals were included in the analysis, the accuracies were 0.47, 0.59 and 0.71.

To investigate the effect of the number of interactions on accuracy of EBVs, the GLMM was fitted by including the first 5, 10, 20, 50 or 100 thousand interaction records. Accuracy increased with an increasing number of records because more observations of sibs were included, but benefit was small above 20 thousand records (Figure 1b).

Table 1. Simulated genetic variances (σ_A^2) required to obtain a certain observed-scale heritability (h_o^2), the corresponding m ± $2\sigma_A$ range of expected interaction probabilities, and accuracy of EBV yielded by the ordinary linear mixed model, using data from all 20 pens.

h_0^2	σ_A^2	$\bar{p}(A_a = \mu - 2\sigma_A)$	$\bar{p}(A_a = \mu + 2\sigma_A)$	Accuracy LMM
0.05	0.012	0.008	0.012	0.173
0.1	0.038	0.006	0.015	0.195
0.2	0.170	0.004	0.022	0.241

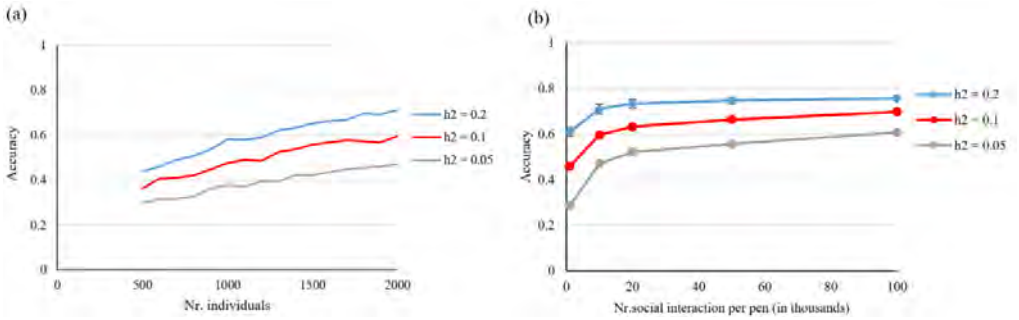


Figure 1. Accuracy of EBV from the GLMM for (a) different numbers of individuals involved and (b) different numbers of observed interactions.

Discussion

Here we developed an individual-based simulation method to generate social interaction events, and a GLMM to estimated breeding values for performer and recipient effect from the resulting data. Results showed that our method yielded a high accuracy of EBV (of 0.47, 0.59 and 0.71 for 2,000 individuals with h_o^2 of 0.05, 0.1 and 0.2, respectively) in comparison to an ordinary LMM (accuracies of 0.17, 0.20 and 0.24), and that genetic differences in the tendency to express a behavior can be large despite a low observed-scale heritability (Table 1).

This simulation shows that, even though the genetic variance (σ_A^2) was the same as the permanent environment variance (σ_{Ep}^2) on the liability scale, the observed-scale heritability can still be very low (e.g. 0.05). This low heritability is due to the random sampling of the event, which introduces extra variance in the observed records.

Our results indicate that the breeding values can be accurately estimated for social interaction traits even if the observed level heritability is very low. Similar results were found when both traits were positively or negatively correlated (results not shown). Note that in these simulations, the maximum information used for BV estimation was own performance, 4 full sibs and 15 half sibs. The accuracy was quite high with this amount of information, and we expect that higher accuracy can be achieved with genomic prediction. This high accuracy can partly be explained by the repeated measures of the trait. In our simulations, one individual could interact with 99 pen mates and was on average engaged in 200 interactions and 20,000 encounters. This means that the breeding value of each individual was expressed 20,000 times. While this may seem a large amount of data, one can easily imagine that an animal performs 10 tail bites or feather pecks a day in total. In that case, our data would correspond to one month of recording, which is shorter than the typical duration of, e.g. the fattening period in pigs or the laying period in laying hens.

In summary, we presented methods to simulate and analyse behavioral data, which distinguishes between performer and recipient effects, and is applicable to large-scale longitudinal data on animals kept in large groups, which will become available in the near future.

Acknowledgements

This work was financially supported by the Dutch Research Council (NWO-TTW).

References

Bijma P., Muir W.M., & Van Arendonk, J.A. (2007). Genetics, 175(1), 277-288. https://doi.org/10.1534/genetics.106.062711

Brinker T., Bijma P., Visscher J., Rodenburg T. B., and Ellen E. D. (2014). Genetics Selection Evolution, 46(1), 1-10. https://doi.org/10.1186/1297-9686-46-33

DeVries T. J., Von Keyserlingk M. A. G., & Weary D. M. (2004). Journal of dairy science, 87(5), 1432-1438. https://doi.org/10.3168/jds.S0022-0302(04)73293-2

Ellen E.D., Rodenburg T. B., Albers G. A., *et al.* (2014). Frontiers in genetics, 5, 377. https://doi.org/10.3389/fgene.2014.00377

Gilmour A. R., Gogel B. J., Cullis B. R., Welham S., and Thompson R. (2015). ASReml user guide release 4.1 structural specification. Hemel hempstead: VSN international ltd.

Kjaer J. B., and Sørensen P. (1997). British Poultry Science, 38(4), 333-341. https://doi.org/10.1080/00071669708417999

Krause H. V., Baum K., Baumann A., and Krasnova H. (2021). Media Psychology, 24(1), 10-47. https://doi.org/10.1080/15213269.2019.1656646

Muir, W. M. (2005). Genetics 170(3):1247-1259.

Radersma R. (2021). Methods in Ecology and Evolution, 12(1), 42-53. https://doi.org/10.1111/2041-210X.13499

Van Krimpen M. M., Leenstra F., Maurer V., and Bestman M. (2016). Journal of Applied Poultry Research, 25(1), 129-138. https://doi.org/10.3382/japr/pfv048

118. Divergent genetic selections for social attractiveness or tolerance toward humans in sheep

D. Hazard[1*], E. Delval[2], S. Douls[3], C. Durand[3], G. Bonnafe[3], D. Foulquié[3], D. Marcon[4], C. Allain[3], S. Parisot[3] and A. Boissy[2]

[1]GenPhySE (Génétique, Physiologie et Systèmes d'Elevage), Université de Toulouse, INRAE, ENVT, 31326 Castanet-Tolosan, France; [2]UMR Herbivores, Université de Clermont, INRAE, VetAgro Sup, 63122 Saint-Genès Champanelle, France; [3]UE Domaine de la Fage, INRAE, 12250 Saint-Jean Saint-Paul, France; [4]UE P3R, INRAE, 18390 Osmoy, France; dominique.hazard@inrae.fr

Abstract

Including behavioural traits in genetic selection could be an advantageous strategy for improving adaptation and welfare of farm animals to extensive farming systems in an agroecological perspective. The aim of the present study was to investigate in sheep the efficiency of an early selection for social attractiveness or tolerance to humans. Four divergent lines were created according to the reactivity of lambs to social separation or to human presence. Lambs were individually exposed just after weaning to two behavioural tests. The arena test aims at assessing the motivation of the lambs towards conspecifics and the corridor test aims at assessing the reactivity of the lambs to an approaching shepherd. Several behaviours were measured including vocalizations and flight distance from a human used as selection criteria. After two generations of selection, the divergent selection led to highly significant line differences for social attractiveness or tolerance to humans, respectively 1.8 and 1.0 genetic standard deviation between high and low lines. Such divergent selections also affected additional behaviours expressed in the tests.

Introduction

Genetic selection including behavioural traits could be an advantageous strategy for improving robustness and welfare of farm animals in various farming conditions by minimizing unsuitable responses to changes in their social and physical environment, limiting an excessive fear of humans and improving sociability (Mignon-Grasteau *et al.*, 2005). Farm animals are social and gregarious, and relational behaviours are essential for ensuring social cohesion, social facilitation, offspring survival and docility toward humans. Breed differences and genetic variation within breed have been reported in sheep for early social behaviours and found to be heritable, and associated with some QTL, suggesting such behaviours could be selected (Boissy *et al.*, 2005; Beausoleil *et al.*, 2012; Hazard *et al.*, 2014; Cloete *et al.*, 2020). In addition, such early social reactivity of lambs was identified as a robust trait (Hazard *et al.*, 2016). However, up to now, any study has demonstrated the feasibility, efficiency and effects of genetic selection for early reactivity to social separation or towards a human in domestic sheep. Thus, the aim of the present study was to investigate in sheep the efficiency of an early genetic selection for social attractiveness or tolerance to humans using experimental divergent lines.

Materials & methods

Animals and management. The experimental animals were Romane meat sheep, reared at the INRAE experimental farm of La Fage (Saint-Jean Saint-Paul, Causse du Larzac, France) exclusively outdoors under harsh environmental conditions. The flock comprised about 250 reproductive females reared on 280 ha of rangelands. The farming system and management characteristics have been described previously by Gonzalez *et al.* (Gonzalez-Garcia *et al.*, 2014). Ewes lambed for the first time at 2 years of age and all the lambs were born outdoor in the spring.

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_118, © D. Hazard *et al.* 2022

Behavioural tests and responses. Each year during 7 years, in average 300 female and male lambs were individually exposed just after weaning to two behavioural tests. The arena test (AT) consisted of two successive phases evaluating: (1) reactivity to social isolation (AT1); (2) the motivation of the lamb towards conspecifics in presence of a motionless human (AT2). The arena test pen consisted in an unfamiliar enclosure virtually divided into 7 zones as described in detail by Ligout *et al.* (Ligout *et al.*, 2011). The corridor test (CT) consisted of two successive phases evaluating: (1) reactivity to social isolation (CT1) and; (2) reactivity to an approaching human (CT2). The test pen consisted in a closed, wide rectangular circuit and has been described in detail by Boissy *et al.* (Boissy *et al.*, 2005). Several behaviours were measured: vocalizations (i.e. frequency of high-pitched bleats), locomotion (i.e. number of virtual zones crossed), vigilance posture (i.e. time spent by the animal motionless), the proximity score (i.e. weighting of time spent in virtual zones, a high score indicated a high duration spent close to conspecifics and a human). The mean flight distance (DIST) separating the human and the lamb and the time during which the human saw the lamb (SEEN) were measured in CT2.

Statistical handling. Two synthetic variables were constructed using the first component of each principal component analysis (PCA) (PRINCOMP procedure in the SAS® software). A PCA was performed on vocalizations across the different phases of behavioural tests (AT1, AT2 and CT1) and the resulting synthetic variable (HBLEAT) was used as indicator of the reactivity to social isolation. A PCA was performed on DIST and SEEN variables and the resulting synthetic variable (HUMAPPRO) was used as indicator of the reactivity to an approaching human. For divergent selection, individual estimated breeding values (EBV) were calculated for HBLEAT and HUMAPPRO with univariate analyses under an animal mixed model using the REML methodology implemented in the ASREML software (Gilmour *et al.*, 2009). In order to constitute the divergent lines, extreme animals were chosen as sires (in average 10 sires per generation and line) or dams (in average 70 dams per generation and line) at each generation according to their high or low EBV for HBLEAT or HUMAPPRO while intermediary for HUMAPPRO or HBLEAT, respectively. Selection intensities were in average 10 and 45% for sires and dams, respectively. Two generations of selected animals were produced (in average 270 female and male lambs per generation and line). Linear models (proc GLM, SAS Institute Inc.) were applied to test whether the divergent selections on behavioural traits had a significant effect on the recorded traits considering the adequate fixed effects in the analyses. The linear model used was $y_{ijklm} = \mu + line_i + sex_j + lsr_k + age_l + year_m + e_{ijklm}$; where y_{ijklm} was the behavioural observations of lambs; μ was the overall mean of the population; line, sex, lsr, age and year were the fixed effects of the behavioural line *i*, the sex *j* of the lambs, the litter size born and reared *k*, age *l* of the dam, and year *m* of the experiment, respectively.

Results & discussion

For gregarious species such as sheep, vocalization and locomotion behaviours in response to social separation are interpreted as an active way to re-establish social links (Boissy *et al.*, 2005; Beausoleil *et al.*, 2012). The divergence between low (S-) and high (S+) lines selected on social attractiveness reached 1.8 genetic standard deviation after two generations of selection (Figure 1). The divergence between low (H-) and high (H+) lines selected on tolerance toward humans reached 1.0 genetic standard deviation after two generations of selection. Responses to selection obtained were consistent with those expected considering the heritabilities previously estimated for both traits (h^2=0.5±0.06 for social attractiveness and 0.15±0.04 for tolerance toward humans) and the intensity of selection applied. As expected, the divergent lines for tolerance toward humans were intermediary for social attractiveness compared to S- and S+ lines even if a slight difference was observed between H+ and H- lines for the social criterion (Figure 1). A high difference was observed between S- and S+ lines for tolerance toward humans while no difference was expected between these lines for this criterion. Such difference could be linked to the low positive genetic correlation previously reported between criteria for social attractiveness and tolerance toward humans

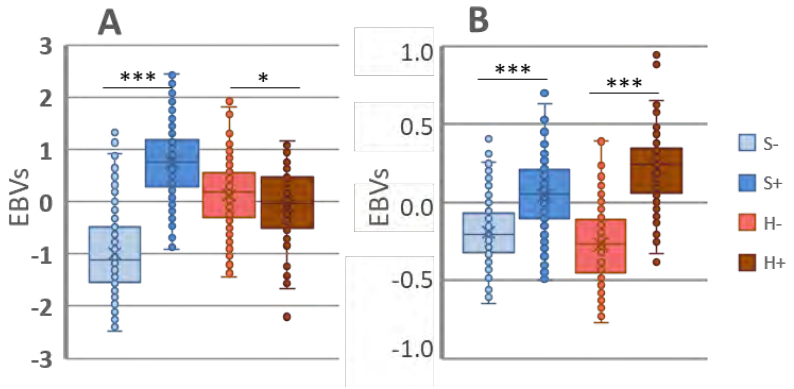


Figure 1. Estimated breeding values for social attractiveness (A) and tolerance toward humans (B) in the four divergent lines after two generations of selection. S- (n=279) and S+ (n=314) lines, low and high social attractiveness, respectively; H- (n=233) and H+ (n=149) lines, low and high tolerance toward humans, respectively. Pairwise comparison between Least Squares Means for line effect: *, and *** for *P*-value<0.05 and 0.001, respectively.

(r_g=0.25±0.11) (Hazard *et al.*, 2016). This difference may also be due to the selection criterion that involved measurement of reactivity to social separation in presence of a human (i.e. vocalizations in AT2). Thus, conflict behaviour between the social motivation toward conspecifics and avoidance of a motionless human included in the selection criterion may thus have resulted in differences between S- and S+ lines for both social attractiveness and tolerance toward humans.

Effects of divergent selections on the different behaviours expressed in response to social separation with or without human presence in arena and corridor tests were also investigated (Table 1). Selection for social attractiveness resulted in a two-fold higher expression of vocalizations in response to social isolation without or with human presence. This selection for social attractiveness also resulted in a higher locomotor activity in response to visual isolation from conspecifics (i.e. AT1 phase) as well as a higher proximity score with conspecifics in presence of a human (i.e. AT2 phase). Thus, selection for social attractiveness based on vocalizations not only affected vocalization behaviour but also locomotor activity in a context

Table 1. Behaviours in the four divergent lines after two generations of selection.

Behaviours	Test	S-[1]	S+	Line[2]	H-	H+	Line
Vocalizations	AT1	8.0	16.2	***	13.6	11.9	***
Vocalisations	AT2	3.0	7.6	***	5.7	5.0	NS
Locomotion	AT1	23.3	27.2	***	26.5	24.0	*
Locomotion	AT2	6.6	7.2	NS	7.2	4.9	**
Vigilance	AT1	18.6	18.2	NS	17.7	19.3	NS
Proximity	AT2	23.9	27.9	*	25.0	29.1	*
Flight distance	CT2	5.5	5.3	*	5.6	5.1	***

[1] S- and S+ lines, low and high reactivity to social separation, respectively; H- and H+ lines, low and high tolerance toward a human, respectively. AT1, arena test 1 (visual social isolation); AT2, arena test 2 (social isolation in presence of a human); CT2, corridor test 2 (reactivity to an approaching human); Vocalizations: frequency; Locomotion: number; Vigilance: duration in sec; Proximity: duration in sec; Flight distance: meter.

[2] Line effect: *, ** and *** for *P*-value<0.05, 0.01 and 0.001, respectively.

specific way. In one hand, more social lambs (S+) increased vocalizations and locomotion behaviours in response to visual separation from conspecifics (i.e. AT1 phase) probably to search actively for them and re-establish social links. On the other hand, when lambs were separated from conspecifics by a grid barrier and a human (i.e. AT2 phase), S+ lambs showed higher levels of vocalizations and proximity while they expressed quite similar level of locomotion compared with S- lambs. This suggested that locomotor activity of S+ lambs was more frequently observed in virtual zones close to conspecifics and the human in the second phase of arena test. As expected, the flight distance of lambs from an approaching human decreased with the selection for tolerance toward humans (Table 1). This selection also resulted in a slight decrease in vocalization and locomotion behaviours in response to social isolation as well as an increase in the proximity score toward conspecifics in presence of a human. We hypothesized that genetic selection for lambs more tolerant toward humans also reduced the overall reactivity to social separation and exposure to humans, may be by reducing overall stress responses.

Funding

Part of this project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 772787 (SMARTER project) and INRAE (COMPAGNE project).

References

Beausoleil N. J., Blache D., Stafford K. J., Mellor D. J., and Noble A. D. L. (2012). Appl. Anim. Behav. Sci. 139: 74-85. https://doi.org/10.1016/j.applanim.2012.02.020

Boissy A., Bouix J., Orgeur P., Poindron P., Bibe B. *et al.* (2005). Genet. Sel. Evol. 37: 381-401. https://doi.org/10.1051/gse:2005007

Cloete S. W. P., Burger M., Scholtz A. J., Cloete J. J. E., Kruger A. C. M. *et al.* (2020). Appl. Anim. Behav. Sci. 233: 105152. https://doi.org/10.1016/j.applanim.2020.105152

Gilmour A. R., Gogel B. J., Cullis B. R., and Thompson R. (2009). www.vsni.co.uk.

Gonzalez-Garcia E., de Figuereido V. G., Foulquie D., Jousserand E., Autran P. *et al.* (2014). Domest. Anim. Endocrinol. 46: 37-48. https://doi.org/10.1016/j.domaniend.2013.10.002

Hazard D., Bouix J., Chassier M., Delval E., Foulquie D. *et al.* (2016). J. Anim. Sci. 94: 1459-1471. https://doi.org/10.2527/jas2015-0277

Hazard D., Moreno C., Foulquié D., Delval E., François D. *et al.* (2014). BMC Genom. 15: 778. https://doi.org/10.1186/1471-2164-15-778

Ligout S., Foulquie D., Sebe F., Bouix J., and Boissy A. (2011). Appl. Anim. Behav. Sci. 135: 57-62. https://doi.org/10.1016/j.applanim.2011.09.004

Mignon-Grasteau S., Boissy A., Bouix J., Faure J.-M., Fisher A. D. *et al.* (2005). Livest. Prod. Sci. 93: 3-14. https://doi.org/10.1016/j.livprodsci.2004.11.001

119. Genetic parameters of novel behaviour traits derived from social network analysis in pigs

S. Agha[1*], S. Foister[2], R. Roehe[2], S.P. Turner[2] and A. Doeschl-Wilson[1]

[1]The Roslin Institute, University of Edinburgh, Easter Bush, Midlothian EH25 9RG, Scotland, United Kingdom; [2]Animal and Veterinary Sciences Department, Scotland's Rural College, West Mains Road, Edinburgh EH9 3JG, Scotland, United Kingdom; sagha@exseed.ed.ac.uk

Abstract

Social network analysis (SNA) has provided novel traits that describe the role of individual pigs in aggression. The objectives were to estimate the genetic parameters for these SNA traits and quantify their genetic association with skin lesions. Pigs were video recorded for 24h post-mixing. The observed fight and bullying behaviour of each animal was used as input for the SNA. Skin lesions were counted on different body parts at 24h (SL24h) and 3wk (SL3wk) post-mixing. A Bayesian approach estimated the genetic parameters of SNA traits and their association with skin lesions. SNA traits were heritable (h^2=0.18 to 0.26) and strongly genetically correlated (rg>0.88). Positive genetic correlations were observed between 3 of the 4 studied SNA centrality traits with anterior SL24h, although they showed negative genetic correlations with anterior SL3wk. This study provides a first step towards potential integration of SNA traits into a multi-trait selection index for improving pigs' welfare.

Introduction

Grouping of unacquainted pigs is a common procedure in commercial farms, but can lead to elevated aggression and injuries (Boumans *et al.* 2018). Genetic selection could help in providing a long-term solution for controlling aggression in pigs (Turner 2011). Skin lesions have been considered as a practical indicator traits for aggression (Wurtz *et al.* 2017). However, they are inadequate for identifying individuals that contribute most to the aggression in a pen. On the other hand, Social Network Analysis (SNA) provides quantitative 'centrality traits' that describe the position of each animal in the social structure (Agha *et al.* 2020). Previous studies have shown strong phenotypic association between SNA centrality traits and skin lesions at the pen level (Foister *et al.* 2018). However, the existence of a genetic component contributing to SNA traits is still unknown. Our objectives were to (1) estimate the genetic parameters for SNA centrality traits, and (2) quantify the genetic association between SNA and skin lesion traits.

Material & methods

The dataset consisted of 1,146 commercial pigs, 698 of them were purebred Yorkshire and 448 were Yorkshire × Landrace. The animals were the progeny of 82 sires and 217 dams. The total number of animals in the two-generation pedigree was 2,427. Animals were distributed into 78 pens, i.e. social groups, each containing 15 animals with the same sex and breed which were formed using 3 animals from each of 5 litters. Fresh skin lesions were counted for each body region 24h (SL24h) and 3 weeks (SL3wk) post-mixing. The body regions were identified as anterior (head, neck, forelegs, and shoulders), central (flanks and back) and posterior (hind legs and rump). The skin lesion records were log transformed ($y= \log_e (x + 1)$) to approach the normal distribution (Full details in Desire *et al.* 2016).

Animals were video recorded for 24h post-mixing. Each animal that initiated or received an aggressive interaction was registered. Aggressive interactions consisted either of fighting or bullying behaviour. Fighting was defined as aggression that lasted at least one second where both pigs engaged in biting, pushing or head knocking the opponent, whereas bullying was defined as when one pig received or delivered aggression with no observable retaliation occurring (Turner *et al.* 2009). All of these aggression

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_119, © S. Agha *et al.* 2022

interactions were entered in the SNA as undirected, allowing for bidirectional initiation of interactions, and unweighted, i.e. the frequency or duration of interactions between a given dyad were not considered. SNA was applied to transform the numerical data of aggressive interactions between animals into graphs, where the animals are displayed in terms of nodes (Farine and Whitehead 2015). The SNA traits for each animal describe its position in the network, i.e. the pen (Table 1). The SNA traits considered in this study were derived from Foister *et al.* (2018) using the 'igraph' package in R (Csárdi *et al.* 2016). All the listed centrality traits showed considerably skewed distributions; therefore, a square root transformation was applied to reduce skewness.

A series of univariate and bivariate analyses were used to estimate the genetic variance components of all transformed SNA and skin lesion traits using the following linear animal model

$$y=Xb + Za + Wc + e$$

where y is the vector of records for the traits, and X, Z and W the incidence matrices of systematic effects, genetic effects and environmental (pen) effects, respectively. Vectors b, a, c and e represent systematic effects (genetic line, sex, batch), additive direct genetic effects, common environmental pen effects and residual error, respectively, whereas body weight at time of mixing was fitted as a covariate. Bayesian analyses were performed using BLUPF90 software (Misztal *et al.* 2018). Heritability and genetic correlations between traits were calculated.

Results

The posterior means of heritability for SNA traits ranged from 0.18 to 0.26. SNA traits were strongly genetically correlated (Table 2).

Positive genetic correlations were observed between all SNA traits and anterior SL24h, except for clustering coefficient, which was weakly negatively correlated with anterior SL24h with HPD95% covering zero (Table 3).

Table 1. The definition of the social network analysis traits considered in this study.

Measures	Interpretation
Degree centrality	The number of animals that a particular animal directly engaged with.
Eigenvector centrality	Takes into consideration both the number of aggressive interactions of the focal individual and the number of aggressive interactions that its social partners have.
Betweenness centrality	Measures the relative importance of the animal in connecting different subgroups directly engaging in aggression.
Clustering coefficient	Quantifies what proportion of an individual's social partners are also connected, and divides this by the possible number of theoretical connections that could have occurred between the social partners.

Table 2. Posterior means of heritability (diagonal) and genetic correlations (off-diagonal) and the 95% highest posterior density intervals (HPD95%) of the SNA traits.

Trait	Degree centrality	Eigenvector centrality	Betweenness centrality	Clustering coefficient
Degree centrality	0.26 (0.11, 0.41)	0.98 (0.96, 1)	0.99 (0.95, 1)	-0.88(-0.99, -0.72)
Eigenvector centrality		0.22 (0.10, 0.36)	0.94 (0.84, 0.99)	-0.95 (-0.99, -0.85)
Betweenness centrality			0.26 (0.11, 0.40)	-0.95 (-0.99, -0.90)
Clustering coefficient				0.18 (0.06, 0.29)

Conversely, estimated genetic correlations between the majority of SNA traits and anterior SL3wk were found to be negative with the HPD95% not including zero, except for clustering coefficient, which was found to be moderately to strongly positively correlated with both anterior and central SL3wk post-mixing (Table 4).

Discussion

To the best of our knowledge, the genetic parameters for SNA traits for aggressive behaviour in pigs and genetic associations to skin lesions have not been previously estimated. The magnitude of the heritability of SNA traits indicates that these traits are partially under genetic regulation and could be utilized for selective breeding. Remarkably, heritability estimates for SNA traits, which partly capture social interactions between group members other than the individual in question, were at similar magnitude as those of dyadic behavioural traits previously estimated from the same dataset (Desire *et al.* 2016) and fall within the range of heritability estimates for skin lesions at 24h post-mixing (Turner *et al.* 2009; Wurtz *et al.* 2017).

Positive genetic correlations between 3 of the 4 SNA centrality traits with anterior SL24h, and their negative genetic correlations with anterior SL3wk suggests that individuals with genetic predisposition for high centrality immediately after mixing would tend to suffer high injuries on the anterior part of the body at this stage but would tend to have lower injuries in the stable group. Except for anterior lesions 24h post-mixing, clustering coefficient was the only trait that was found to be positively genetically correlated with skin lesions at both time points suggesting that using this trait at mixing as selection criterion may not inflict a trade-off for reducing lesions at different time points or different body parts. However, this study only provides a first step towards potential integration of SNA traits into a multi-trait selection index for improving animal welfare by reducing both aggressive interactions as well as resulting injuries. The recent advances in automated capture of animal behaviour would facilitate the application of SNA to derive novel behavioural phenotypes that describe an animal's relative contribution to pen level aggression in commercial pig farms.

Table 3. Posterior means of the genetic correlations and the 95% highest posterior density intervals (HPD95%) of SNA and skin lesions recorded 24h post-mixing.

Trait	Anterior skin lesions 24h (HPD95%)	Central skin lesions 24h (HPD95%)	Posterior skin lesions 24h (HPD95%)
Degree centrality	0.62 (0.34, 0.88)	0.01 (-0.46, 0.44)	-0.02 (-0.60, 0.56)
Eigenvector centrality	0.54 (0.15, 1.00)	-0.19 (-1.00, 0.39)	-0.13 (-1.00, 0.59)
Betweenness centrality	0.46 (-0.01, 0.87)	-0.10 (-0.71, 0.40)	-0.27 (-1.00, 0.41)
Clustering coefficient	-0.14 (-0.56, 0.34)	0.63 (0.08, 1.00)	0.73 (0.24, 1.00)

Table 4. Posterior means of the genetic correlations and the 95% highest posterior density intervals (HPD95%) of SNA skin lesions recorded 3 weeks post-mixing.

Trait	Anterior skin lesions 3wks (HPD95%)	Central skin lesions 3wks (HPD95%)	Posterior skin lesions 3wks (HPD95%)
Degree centrality	-0.37 (-0.66, -0.06)	-0.18 (-0.56, 0.20)	0.06(-0.53, 0.80)
Eigenvector centrality	-0.47 (-0.84, -0.10)	-0.23 (-0.67, 0.19)	-0.17(-0.76, 0.40)
Betweenness centrality	-0.57 (-0.99, -0.24)	-0.49 (-0.87, -0.09)	-0.78 (-1.00, 0.18)
Clustering coefficient	0.68 (0.28, 1.00)	0.62 (0.26, 0.97)	0.40 (-0.22, 0.98)

Conclusions

SNA provides promising new phenotypes to reduce aggression and resulting injuries in pigs through selective breeding. Future studies are needed to determine the response to selection for these traits on pen level aggression and other economically important traits in order to simultaneously improve the performance and welfare of pigs.

References

Agha S., Fàbrega E., Quintanilla R., and Sánchez J.P. (2020). Animals 10. https://doi.org/10.3390/ani10112123.

Boumans I. J. M., Imke J., de Boer M., Hofstede G. J., and Bokkers E. A. M. (2018). Physiology and Behavior 194: 23–40. https://doi.org/10.1016/j.physbeh.2018.04.032.

Csárdi G., Nepusz T., and Airoldi E.M. (2016). InterJournal, Complex Systems, 1695. Available at: https://igraph.org.

Desire S., Turner S. P., D'Eath R. B., Lewis C. R. G., and Roehe R. (2016). Animal 10: 1243–1253. https://doi.org/10.1017/S1751731116000112.

Farine D. R. and Whitehead H. (2015). Journal of Animal Ecology 84 (5):1144–63. https://doi.org/10.1111/1365-2656.12418.

Foister S., Doeschl-Wilson A., Roehe R., Arnott G., Boyle L., and Turner S. P. (2018). PLoS ONE 13(10). https://doi.org/10.1371/journal.pone.0205122.

Misztal I., Tsuruta S., Lourenco D. A. L, Masuda Y., Aguilar I., Legarra A., *et al.* (2018). BLUPF90 software. Available at: http://nce.ads.uga.edu/wiki/doku.php?id=documentation.

Turner S. P., Roehe R., D'Eath R. B., Ison S. H., Farish M., *et al.* (2009). Journal of Animal Science 87(10):3076–82. https://doi.org/10.1007/s10519-007-9171-2.

Turner S. P. (2011). Applied Animal Behaviour Science 134(1):1–9. https://doi.org/10.1016/j.applanim.2011.06.001.

Wurtz K. E., Siegford J. M., Bates R. O., Ernst C. W., and Steibel J. P. (2017). Journal of Animal Science 95(10):4310–4317. https://doi.org/10.2527/jas2017.1757.

120. Genetic parameter estimation of various body temperature and respiration rate indicators in maternal-line pigs under heat stress conditions

H. Wen[1], J.S. Johnson[2], P.H.F. Freitas[1], J.M. Maskal[1], M.K. Byrd[1], F. Tiezzi[3], Y. Huang[4], A.P. Schinckel[1] and L.F. Brito[1*]

[1]Department of Animal Sciences, Purdue University, West Lafayette, IN 47906, USA; [2]USDA-ARS Livestock Behavior Research Unit, West Lafayette, IN 47907, USA; [3]Department of Agriculture, Food, Environment and Forestry (DAGRI), University of Florence, Piazzale delle Cascine 18, 50144 Firenze, Italy; [4]Smithfield Premium Genetic, Rose Hill, NC 28458, USA; britol@purdue.edu

Abstract

This study aimed at investigating genetic parameters of six physiological indicators of heat stress (HS) in maternal-line pigs, including automatically-measured vaginal temperature (T_V); ear (T_E), shoulder (T_S), rump (T_R), and tail (T_T) temperature; and respiration rate (RR) under heat stress conditions. T_V had the highest heritability (0.34±0.03) and repeatability (0.57) estimates, whereas the heritability of skin surface temperature (SST) and RR ranged from 0.04±0.01 to 0.08±0.01. The genetic correlations among skin surface temperature (SST) indicators ranged from 0.76 to 0.87, indicating that SST can be measured at any of these locations for genetic selection purposes. T_V is positively correlated with SST and RR. Furthermore, SST and RR are also heritable. Although collecting more repeated records are recommended, they are highly correlated at the genetic level, except for T_E. Thus, T_V, SST, and RR are key indicator traits to be included in swine breeding programs for mitigating the negative effects of heat stress.

Introduction

Heat stress (HS) causes considerable welfare issues and economic losses in worldwide pig production due to a reduction in animal productivity, reproductive performance, and animal health. Currently, heat tolerance (HT) is not included in most swine selection indexes. Continued genetic selection for production and reproduction traits has significantly increased metabolic heat production in sows (Stinn and Xin, 2014; Cabezon *et al.*, 2017), and has likely resulted in greater HS sensitivity due to increased heat gain (McConn *et al.*, 2021). Therefore, there is an urgent need to identify feasible-to-measure and reliable HS indicator traits for including in selection schemes. Genetic selection could enable animals to better adapt to challenging climatic conditions and decrease the negative effects of HS on profitability, animal welfare, and productive efficiency. Physiological indicators, including body temperature, respiration rate, panting score, and pulse rate, might be key indicators of HS in pigs (Gourdine *et al.*, 2017). In this context, the main objectives of this study were to: (1) describe the phenotypic variability of six HS indicators [T_V, SST (ear, shoulder, rump, and tail), RR]; and, (2) estimate variance components and genetic parameters (including heritability and genetic correlations) for these six traits in lactating sows under heat stress conditions.

Materials & methods

Animals and dataset. All the datasets analyzed in this study were collected at a commercial farm in Maple Hill, NC, USA (34.70738, -77.73653). Phenotypic records (T_V, T_E, T_S, T_R, T_T, RR) were collected in 1,645 multiparous (parities 2 to 7) lactating sows (Landrace × Large White cross) from June 5th, 2021 to July 30th, 2021. The T_V of 1,381 lactating sows was measured automatically every 10 minutes using a digital thermometer (iButton model DS1921H, calibrated accuracy ±0.01 °C; resolution=0.125 °C; Dallas Semiconductor, Maxim, Irving, TX, USA). The SST of all sows was measured manually four times a day (8:00, 12:00, 16:00, 20:00) using an infrared temperature gun (Raytek model ST61; accuracy = ±1%; emissivity = 0.98; resolution = 0.10 °C; Raytek Corporation, Fluke Process Instruments, Everett, WA, USA). RR was

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_120, © H. Wen *et al.* 2022

collected by counting flank movements for 15 s and multiplying by 4 to calculate breaths per minute (bpm) four times a day (as described in Johnson *et al.*, 2016). The ambient temperature (T_a) and relative humidity (RH) of each room was recorded every five minutes using within-room data loggers (HOBO; Onset ®; Bourne, MA, USA). Dew point (T_{dp}) was calculated by T_a and RH using the Magnus-Tetens equation (Buck, 1981). Outliers were discarded if they deviated by more than 3.5 SD from the mean. A total of 1,639 sows within the studied population were genotyped using the PorcineSNP50K Bead Chip (50,703 SNPs; Illumina, San Diego, CA, USA). SNP markers were kept with a minor allele frequency (MAF) >0.01, SNPs and animals with call rate above 0.90 using the BLUPF90 family software (Misztal *et al.*, 2002). After the quality control, 49,547 SNPs were used for further analyses.

Data analysis. Multi-trait (T_E, T_S, T_R, T_T) and bivariate (T_V and RR with the other five traits) repeatability model analyses were performed using the BLUPF90 family programs (Misztal *et al.*, 2002). The statistical model used can be described as:

$$y = X\beta + Z_a a = Z_{pe} pe + e, \quad (1)$$

where y is the vector of phenotypic records for T_V, SST, and RR; β is a vector of fixed effects, including parity, location, time of measurement, body size, scorer of body size, hair density, scorer of hair density, and T_{dp}; a is the random genetic animal effect and assumed to follow $\mathbf{a} \sim N(0, \mathbf{G}\sigma_a^2)$. The *G* matrix was created based on the method of VanRaden (2008); *pe* is the vector of random permanent environmental effect; X, Z_a, Z_{pe} are the incidence matrices; and e is the residual effect assumed to follow $\mathbf{e} \sim N(0, \mathbf{I}\sigma_e^2)$. The (Co)variances for all traits were estimated using the AIREMLF90 program (Misztal *et al.*, 2002). The heritability, repeatability, and genetic correlations were defined as follows:

$h^2 = \frac{\sigma_a^2}{\sigma_a^2+\sigma_{pe}^2+\sigma_e^2}$, $r_e = \frac{\sigma_a^2+\sigma_{pe}^2}{\sigma_a^2+\sigma_{pe}^2+\sigma_e^2}$, $r_{g_{ij}} = \frac{Cov_{ij}}{\sqrt{Var_i \times Var_j}}$, respectively.

Results

Descriptive statistics from the data available are shown in Table 1. The average ± SD T_V, T_E, T_S, T_R, T_T, and RR were 39.74±0.77, 36.74±1.07, 36.46±1.08, 37.23±0.92, 36.90±0.95, 72.98±28.28, respectively. The average values of SST were remarkably lower than that of T_V. Besides, the high average and maximum T_V and RR indicated that sows are under heat stress conditions. Variance components, heritability, and repeatability estimates for all the traits evaluated are listed in Table 2. The heritability estimates for T_V, T_E, T_S, T_R, T_T, and RR are 0.34±0.03, 0.04±0.01, 0.06±0.01, 0.08±0.01, 0.06±0.01, and 0.17±0.01, respectively. The repeatability estimates for T_V, T_E, T_S, T_R, T_T, and RR are 0.58, 0.09, 0.19, 0.21, 0.18, and 0.19, respectively. T_V had the highest heritability and repeatability estimates. The genetic correlations among all the six traits are shown in Table 3. T_V is moderately and positively correlated with SST (from 0.32 to 0.46) and RR (0.22±0.11). The correlations among SST traits were positive and ranged from 0.75 (T_R with T_T) to 0.84 (T_E with T_S), indicating that SST are genetically similar. Apart from T_E (0.16±0.08), the remaining SST are highly correlated with RR. The genetic relationships between RR and T_S (0.87±0.02) and T_R (0.81±0.01) were positive, while RR had a surprisingly negative correlation with T_T (-0.71±0.01).

Discussion

The presence of genetic variation in the traits evaluated in this study indicates that there are great opportunities for genetic improvement in swine populations. The moderate-to-high heritability estimate of T_V (0.34±0.03) in our study was in agreement with the estimate of 0.32±0.18 under low THI conditions reported by Sarlo *et al.* (2019), who measured T_V every 5 minutes in Brahman–Angus heifers. AI-Kanaan (2016) reported similar heritability estimates (0.018 to 0.035 on different body areas) for SST and high positive genetic correlations among SST (0.688 to 0.937) in cattle. The heritability estimate of RR in our

Table 1. Descriptive statistics for six physiological indicators of heat stress in lactating sows.[1]

Trait	No. of records	No. of sows	Mean	SD	Minimum	Maximum
T_V / ℃	932,926	1,381	39.74	0.77	37.08	42.72
T_E / ℃	25,568	1,645	36.74	1.07	32.50	40.70
T_S / ℃	25,572	1,645	36.46	1.08	32.30	39.80
T_R / ℃	25,571	1,643	37.23	0.92	33.60	39.90
T_T / ℃	25,570	1,643	36.90	0.95	33.20	40.00
RR	25,815	1,643	72.98	28.28	12.00	172.00

[1] T_V=Vaginal temperature; T_E=Ear temperature; T_S=Shoulder temperature; T_R=Rump temperature; T_T=Tail temperature; RR= Respiration rate.

Table 2. Genetic components and parameters for six physiological indicators of heat stress in lactating sows.[1]

Trait	Variance components			Heritability	Repeatability
	σ^2_a	σ^2_{pe}	σ^2_e		
T_V	0.07	0.11	0.13	0.34±0.03	0.58
T_E	0.03	0.04	0.70	0.04±0.01	0.09
T_S	0.04	0.10	0.59	0.06±0.01	0.19
T_R	0.03	0.06	0.34	0.08±0.01	0.21
T_T	0.03	0.06	0.43	0.06±0.01	0.18
RR	34.50	70.21	445.53	0.07±0.01	0.19

[1] σ^2_a, σ^2_{pe} and σ^2_e are estimated additive variance, permanent variance and residual variance, respectively; T_V=Vaginal temperature; T_E=Ear temperature; T_S=Shoulder temperature; T_R=Rump temperature; T_T=Tail temperature; RR= Respiration rate.

study is higher than in Holstein (0.04±0.01; Luo *et al.*, 2021) and in a Holstein and German black pied cattle mixed population (0.05±0.04; AI- Kanaan, 2016). However, RR was fitted as a categorical trait in their studies. Low to moderate repeatability estimates of SST and RR implied that their records changed largely within 24 hours. Luo *et al.* (2021) reported similar repeatability of RR (0.14) during heat stress in Holstein cattle. In cattle, genetic correlations between core body temperature and RR were 0.25±0.12 (Gourdine *et al.*, 2016) and 0.26±0.11 (Luo *et al.*, 2021). In general, T_V had the highest heritability and repeatability estimates, and was moderately related to SST and RR. Furthermore, SST and RR are less heritable and repeatable, but highly correlated at the genetic level, except for T_E. The physiological indicators of HS evaluated here can be included in swine breeding programs for reducing the negative effects of HS. Although there is a value in recording all traits, based on the genetic parameters obtained, the automatically-measured T_V appears to be the most recommended indicator trait for selecting for HT in maternal-line pigs.

Table 3. Genetic correlations estimations among six physiological indicators of heat stress in lactating sows.[1]

Trait	T_E	T_S	T_R	T_T	RR
T_V	0.38±0.04	0.32±0.05	0.46±0.07	0.40±0.05	0.22±0.11
T_E		0.84±0.01	0.83±0.02	0.76±0.08	0.16±0.08
T_S			0.83±0.01	0.83±0.01	0.87±0.02
T_R				0.75±0.02	0.81±0.01
T_T					-0.71±0.01

[1] T_V=Vaginal temperature; T_E=Ear temperature; T_S=Shoulder temperature; T_R=Rump temperature; T_T=Tail temperature; RR= Respiration rate.

Acknowledgements

This research was supported by the Agriculture and Food Research Initiative (AFRI) Competitive Grant number 2020-67015-31575 from the USDA National Institute of Food and Agriculture.

Reference

Al-Kanaan A. (2016) Ph.D. thesis, University of Kassel, Germany.

Baumgard L. H., and Rhoads R. P. Jr. (2013) Annu. Rev. Anim. Biosci. 1: 311–337. https://doi.org/10.1146/annurev-animal-031412-103644.

Buck A. L. (1981) J. Appl. Meteorol 20:1527-1532. https://doi.org/10.1175/1520-0450

Cabezón F. A., Schinckel A. P., Richert B. T., Peralta W. A., and Gandarillas M. (2017) Prof. Anim. Sci. 33:357–362. https://doi.org/10.15232/pas.2016-0158.

Gourdine J. L., Rauw W. M., Gilbert H., and Poullet N. (2021) Front. Vet. Sci. 13(8):770480. https://doi.org/10.3389/fvets.2021.770480.

Johnson J. S., Sapkota A., and Lay D. C. (2016). J. Appl. Physiol. 120:1249–1259. https://doi.org/10.1152/japplphysiol.00685.2015.

Luo H., Li X., Hu L., Xu W., and Chu Q. et al (2021) J. Dairy Sci. 104(4):4441-4451. https://doi.org/10.3168/jds.2020-18725.

McConn B.R., Gaskill B.N., Schinckel A.P., Green-Miller A.R., and Lay D.C. Jr., *et al.* (2021) J. Anim. Sci. 99:1-8. https://doi.org/10.1093/jas/skab181.

Misztal I., Tsuruta S., Strabel T., Auvray B., and Druet T. (2002). Proc. of the 7th WCGALP, Montpellier, France.

Sarlo D. K. M., Hamblen H., Hansen P. J., Dikmen S., Oltenacu P. A. *et al.* (2019) J. Anim. Sci. 97(8):3246-3252. https://doi.org/10.1093/jas/skz188.

Stinn J.P., and Xin. H. (2014) Trans. ASABE. 57: 1517-1528. https://doi.org/10.13031/trans.57.10711.

VanRaden P. M. (2008) J. Dairy Sci. 91(11):4414-4423. https://doi.org/10.3168/jds.2007-0980.

121. Genetics of body weight and length of Large White piglets and relationship to maternal performance

L. Canario[1*], Y. Billon[2], W. Hébrard[2] and J. Demars[1]

[1]UMR1388 GenPhySE, INRAE/INPT ENSAT/INPT ENVT, 31326, Castanet, Tolosan, France; [2]UE GenESI, INRAE, 17700, Surgères, France; laurianne.canario@inrae.fr

Abstract

Body weight and body metrics of newborn piglets reflect their level of development. We studied a Large White population raised in a single experimental herd. Survival and growth performance of 10,101 piglets originating from 689 litters were analysed. In 440 of these litters, 1,320 piglets contrasting for birth weight were recorded for body length and circumference. At piglet level, the estimated direct and maternal heritability values were h^2_d=0.02 and h^2_m=0.13 for birth weight, h^2_d=0.65 and h^2_m=0.21 for the ponderal index and h^2_d=0.23 and h^2_m=0.01 for the ratio of circumference to body length. The genetic correlation between the ponderal index and litter size and gestation length tended to be negative (r_g=-0.38 and -0.35, respectively with SE=0.33).

Introduction

The developmental status of piglets at birth largely determines their survival and growth until weaning. Intrauterine growth retardation results in piglets that are weaker at birth due to the brain-sparing effect, evidenced by a higher brain to organ ratio. Morphometric measurements can be used to measure maturity in addition to body weight, which is only a partial indicator of maturity at birth. A study of head characteristics of newborn pigs showed that developmental delay is influenced by genetics (Matheson *et al.*, 2018). Developmental differences can surface in measurements that combine weight and several lengths on the piglet's body. We worked on objective morphometric measurements, body length and thoracic circumference, obtained by intra-litter sampling. First, analyses at the piglet level allowed to disentangle direct genetic effects from maternal genetic effects. Second, although the metrics data are at piglet level, growth retardation could be considered as a property of the mother, partly due to her uterine capacity and nutritional status at the end of gestation. Genetic relationships of body weight and length, with litter size and gestation length of the sow, and litter growth during lactation were estimated.

Materials & methods

Animals and data. Data were collected from a purebred Large White population raised at the INRAE GenESI experimental farm (https://doi.org/10.15454/1.5572415481185847E12). The project received financial support from the French National Agency for Research (PIPETTE project, ANR-18-CE20-0018). Sows were managed under a batch farrowing system, with three weeks interval. Farrowing was not induced, and interventions including cross-fostering were limited. Piglet death was recorded daily. Each individual piglet, regardless of whether it was stillborn or alive, was weighed within 24h after birth. At the same time, three live-born piglets per litter were measured for body length (crown to rump) and circumference (thorax): the lightest, the one closest to the mean litter weight, and the heaviest of the litter. All piglets were weighed again at 3 wks of age. In this period, their growth depended solely on milk intake. Sows were inseminated with semen from AI Large White boars from the French national scheme, for which variability was maximised after 2018, over the period when morphometric measurements were recorded. Piglets were weaned at 4 weeks of age. Two data sets were constructed. The first data set included individual birth weight (IBW) and weight at weaning (IW3) from which was calculated the individual daily weight gain from birth to 3 weeks, as IWG0-3 = (IW3-IBW)/age at 3 wks (20±3 days). The density index, the body mass index, the ponderal index (PI), and the ratio of circumference to body length (ratioL) were calculated

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_121, © L. Canario *et al.* 2022

from weight and metrics. Litter growth traits, i.e. litter mean birth weight (MBW) and daily weight gain (IWG03) were calculated from individual records. Gestation length was recorded, in full days, as farrowing date minus successful mating date. Prolificacy included the number of piglets born in total (TNB) and that survived until weaning (SURV). The data set contained data on 385 sows with 10,101 piglets born in 689 litters, of which 7,019 were weighed twice and measured at birth. Among sows, 67% had their own weight performance at birth and 3 weeks of age. The data set was augmented with litter records (all traits but morphometrics traits) collected from older and contemporary sows, from end of 2015 to end of 2021. All sows considered were raised in the same housing system, i.e. crate; and with the same management practices. The pedigree file, based on sow pedigree and sire pedigree included 20,666 animals. The second data set included piglet and litter records.

Statistical analysis. All analyses were carried out with the REML procedure of the ASReml software (Gilmour *et al*., 2015). The goodness of fit was used to compare nested models following a chi-square statistics. Models for analyses included the fixed effect of parity (5 levels) and year-month (54 levels). The gender effect (2 levels at birth, 3 levels at 3 wks) was included in all piglet analyses. For piglet traits including IWB, IWG03 and PI, the complete model (possible simplified to reach convergence) was:

$$\mathbf{y} = \mathbf{Xb} + \mathbf{Z}_D\mathbf{a}_D + \mathbf{Z}_M\mathbf{a}_M + \mathbf{Vc} + \mathbf{Wl} + \mathbf{e}$$

where y is the vector of observations; X, V and W are known incidence matrices; b is a vector for fixed effects; a_D is a vector of direct additive genetic effects, and a_M is a vector of maternal genetic effects. The model accounted for a covariance between direct and maternal genetic effects, using the variance structure

$$\begin{bmatrix}\mathbf{a}_D\\ \mathbf{a}_M\end{bmatrix} \sim MVN\left(\begin{matrix}\mathbf{0}\\ \mathbf{0}\end{matrix}, \begin{bmatrix}\sigma^2_{A_D} & \sigma_{A_{DM}}\\ \sigma_{A_{DM}} & \sigma^2_{A_M}\end{bmatrix} \otimes \mathbf{A}\right)$$

where A denotes the matrix of additive genetic relationships with $\sigma^2_{A_D}$ and $\sigma^2_{A_M}$ respectively the direct and maternal additive genetic variances and $\sigma_{A_{DM}}$ the co-variance between direct and maternal genetic effects. Next, c is a vector of random permanent environmental effects of the dam, $\mathbf{c} \sim N(0, \mathbf{I}\sigma^2_c)$; l is a vector of random litter effects, $\mathbf{l} \sim N(0, \mathbf{I}\sigma^2_l)$; and e is a vector of residuals, $\mathbf{e} \sim N(0, \mathbf{I}\sigma^2_e)$. Z_D and Z_M are known incidence matrices for the direct and maternal genetic effects. The phenotypic variance was calculated as $\sigma^2_P = \sigma^2_{A_D} + \sigma^2_c + \sigma^2_l + \sigma^2_{A_M} + \sigma^2_e$.

Analyses with a direct-maternal model were applied. Genetic and phenotypic correlations between piglet traits and between a piglet trait and a litter trait (NBT or GEST) were obtained from sets of bivariate analyses. To overcome convergence difficulties, the model could be simplified to a sow model, considering only the random effects of the dam, i.e. genetics and permanent environment.

Results

The distribution of piglet IBW and PI according to IBW of sampled piglets is given in Figure 1. IBW ($\bar{x}$=1.37 kg SD 0.37) and PI ($\bar{x}$=96.35 SD 13.9) were negatively correlated (y=-0.33x+96; R^2=0.0001; n=1,320). Mean TNB was 15.7 (SD 3.4).

The estimates of heritability and variance components for piglet traits are given in Table 1, considering IBW, IWG03, PI and ratio L as traits of both the piglet and the sow. The size of the direct effects and maternal effects on IWG03 was similar. The maternal heritability on IBW was higher than the direct heritability. We found large direct genetic effects on PI that refer to the piglet merit for achieving proper development at birth. Moreover, the maternal heritability was substantial. Applying a two-trait animal model, with piglet as animal and only the permanent environment of the sow, the direct (piglet) heritability was inflated for IBW

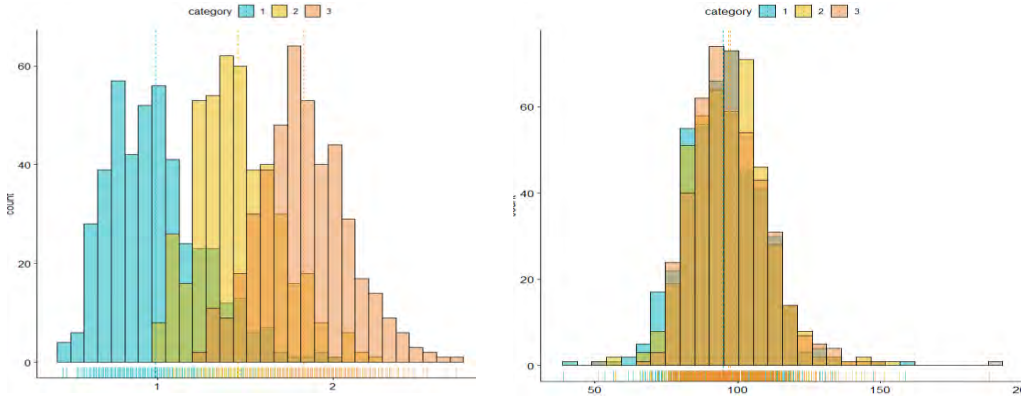


Figure 1. Distribution of piglet IBW (left graph) and PI (right graph) according to within-litter category (1: lightest, 2: mean, 3: heaviest in the litter).

(h^2_d=0.42 SE=0.09) and deflated for PI (h^2_d=0.42 SE=0.09), and the direct genetic correlation between IBW and PI was positive but low (r_g=0.21 SE=0.13; r_p=0.11 SE=0.03).

Since PI depends on sow nutritional state in the end of the gestation phase, next we analysed the piglet measurements as sow traits with an animal model including only direct (sow) effects and sow permanent environment. The estimates of genetic and phenotypic correlations between PI and performance traits are given in Table 2. Similar estimates were obtained with PI adjusted for litter size. At the sow scale, although estimates were not precise, genetic correlations indicate a trend for PI to be correlated negatively with litter size at birth and with gestation length. Sows with higher genetic merit for piglet PI produce smaller litters and have shorter gestation. Since the standard errors of the (co)variances are high, the results need to be confirmed with a larger data set.

Table 1. REML estimates of direct (h^2_d) and maternal (h^2_m) heritabilities and direct-maternal correlation (with SE) for piglet traits.

Estimates[1]	h^2_d	h^2_m	$r_{d,m}$	l^2 (%)	c^2 (%)
IBW	$0.02_{0.09}$	$0.23_{0.05}$	$-0.22_{0.61}$	11.2	1.4
IWG03	$0.12_{0.06}$	$0.12_{0.05}$	$-0.35_{0.28}$	6.7	6.4
PI	$0.65_{0.16}$	$0.21_{0.11}$	$-0.76_{0.01}$	ni[2]	7.8
ratioL	$0.23_{0.17}$	$0.01_{0.11}$	-0.01	13.9	0.5

[1] l^2 and c^2 are respectively the part of phenotypic variance (%) that is explained by litter effects and sow permanent environment effects.
[2] Effect not included in the model.

Table 2. REML estimates of genetic (r_g) and phenotypic correlations (r_p) (with SE) between piglet PI (y_1 h^2_d =0.09 SE 0.05) and litter traits (y_2) from a sow model[1].

PI with	TNB	DGEST	MBW	SURV	$IBW_{(sow)}$
$r_{g\ y1,y2}$	$-0.38_{0.33}$	$-0.33_{0.33}$	$-0.19_{0.33}$	$0.17_{0.41}$	$-0.18_{0.30}$
$r_{p\ y1,y2}$	$-0.05_{0.04}$	$-0.04_{0.03}$	$0.04_{0.04}$	$-0.04_{0.03}$	$0.37_{0.13}$
h^2 of y_2	$0.26_{0.09}$	$0.28_{0.08}$	$0.30_{0.09}$	$0.17_{0.08}$	$0.14_{0.04}$

[1] Models for both traits included the fixed effects of parity and batch, and the random effects of the sow genetics and permanent environment.

Discussion

The heritability for litter size was slightly higher than described in the literature. Heritability estimates for other litter traits were in agreement with previous results (Banville *et al*., 2014; Matheson *et al*., 2018). We investigated new body measures with purposive intra-litter sampling. Although known to be moderately heritable in adult pigs (Johnson and Nugent, 2003), we found a heritability value close to zero for the newborn piglet density index (h^2<0.01) and that of the BMI was not estimable. The sampling limited our ability to estimate the litter effect. Genetic estimates were often of low precision. We obtained the first estimates for PI, a trait of interest that describes whether the newborn piglet was malnourished in utero. This metric is also used in humans for diagnosing fetal growth disorders. The heritabilities for this trait were estimated to be high and a strong genetic antagonism between direct and maternal effects was evidenced. PI is influenced by the genotype of the piglet as well as that of the sow. The genetic correlation with IBW at sow scale was low, which indicate that different genetic determinisms underlie the two traits. Tabuaciri *et al*. (2011) used several lines to create a dataset and averaged the individual records of all piglets in the litter. Their heritability of mean PI was h^2=0.07 (SE=0.08) and the antagonism with MBW was stronger than in the current study (r_g=-0.56; SE=0.61). Moreover, our results tend to indicate that selection for litter size favours the production of piglets with lower PI at birth, in line with the delayed physiological maturity observed in piglets born to modern Large White sows (Canario *et al*., 2007).

References

Banville M., Riquet J., Bahon D., Sourdioux M., Canario L. (2014) J. Anim. Breed. Genet. 132(4):328-37. https://doi.org/10.1111/jbg.12122.

Canario L., Pere M.C., Tribout T., Thomas F., David C., *et al*. (2007) Anim. 1:1409. https://doi.org/10.1017/S1751731107000766.

Gilmour, A.R., Gogel B.J., Cullis B.R., Welham S.J., Thompson R. (2015). ASReml User Guide Release 4.1.

Johnson Z.B., Nugent R.A. (2003) J Anim Sci. 81(8):1943-9. https://doi.org/10.2527/2003.8181943x.

Matheson S.M., Walling G.A., Edwards, S.A. (2018) Genet Sel Evol 50:46. https://doi.org/10.1186/s12711-018-0417-7

Tabuaciri P., Bunter K.L., Graser, H.U. (2011) Proc. Assoc. Advmt. Anim. Breed. Genet. 19: 115-118.

Part 9

Novel phenotyping tools: machine learning

122. Automated individual walking distance of group-housed broilers: a comparison between ground-truth, RFID, and video

J.E. Doornweerd[1*], I. Fodor[1], G. Kootstra[1], R.F. Veerkamp[1], B. de Klerk[2], M. van der Sluis[1], A.C. Bouwman[1] and E.D. Ellen[1]

[1]*Wageningen University & Research, P.O. Box 338, 6700 AH, Wageningen, the Netherlands;* [2]*Cobb Europe, Koorstraat 2, 5831 GH, Boxmeer, the Netherlands; janerik.doornweerd@wur.nl*

Abstract

Individual broiler behaviour could serve as a proxy for an individual's health, welfare and performance. However, large-scale individual behaviour recording is difficult. Wearable and remote sensors have both been investigated to record individual broiler behaviour. In this study, we validated the use of automated tracking using video or radio frequency identification (RFID) by comparing the total estimated walking distance to ground-truth annotated from video. By using colour-marked birds we simplified the computer vision task of YOLOv5. The estimated walking distances by video pixel-based closely matched ground-truth pixel-based (≤2 cm difference). Overall, the RFID-system also closely matched ground-truth pixel-based, except for one movement event (75 cm underestimation). Over the entire duration of the trial, RFID and video gave similar animal rankings, but RFID seems more error prone due to missing reads. RFID is suitable to track broilers over time, but video could be more accurate and practical for on-farm application.

Introduction

There is a growing interest in the individual behaviour of broilers (Van der Sluis *et al.*, 2019), as behaviour could serve as a proxy for the individual's health (Hart, 1988), welfare (Van Hertem *et al.*, 2018), and performance (Reiter and Bessei, 2001). However, broilers are often kept in large groups (i.e. thousands of birds), which makes recording individual broiler behaviour difficult, but breeding programs require automated large-scale individual behaviour records.

Several sensors have been investigated to record individual broiler behaviour. Wearable sensors, such as ultra-wideband tracking (UWB; Van der Sluis *et al.*, 2019), radio frequency identification (RFID; Van der Sluis *et al.*, 2020), and accelerometers (Yang *et al.*, 2021) can record individual broiler behaviour. However, these wearable sensors are impractical and expensive, as these sensors either need batteries (UWB, accelerometer) or an extensive network of antennas (RFID). Alternatively, a remote sensor, such as a video camera, could be used. A video camera can observe multiple animals in a group, but it is challenging to automatically identify individual birds. Hence these cameras have been mainly used for flock level observations (Dawkins *et al.*, 2021; Li *et al.*, 2020; Van Hertem *et al.*, 2018). Nevertheless, individual broiler behaviour could be recorded with video cameras with the use of external markers such as colour-marks or RFID.

The aim of this study was to validate the use of automated tracking using video or RFID by comparing the total estimated walking distance to the ground-truth walking distance acquired from manual annotation on the video data. Furthermore, video and RFID were compared over the entire duration of the trial to assess tracking performance over a longer duration. By using colour-marked birds we simplified the computer vision task.

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_122, © J.E. Doornweerd *et al.* 2022

Materials & methods

Animals and housing. The data used in this study were collected in the research of Van der Sluis *et al.* (2020) in collaboration with Cobb Europe. Van der Sluis *et al.* (2020) followed broilers from hatch until 34 days of age. Forty male broilers from two genetic crosses were housed in one pen (1.8×.6 m). At hatch, all birds were fitted with a RFID-tag. Four birds, two from each cross, were marked in four different colours. From 15 days of age, 34 out of 40 birds were equipped with an UWB-tag, because the number of available UWB-tags was limited. In the present study, only data at 18 days of age were used. The birds were kept under standard lighting conditions. Feed and water were provided *ad libitum* (Van der Sluis *et al.*, 2020).

Radio frequency identification system. A passive RFID-system from Dorset Identification (Dorset Identification B.V., Aalten, the Netherlands) was used to track individual broilers. Each broiler was fitted with a high-frequency RFID-tag (13.56 MHz, 15×3.7 mm, <1.0 gram) around its leg. In total, 30 sub-floor high-frequency RFID-antennas (32×41 cm each) covered the entire surface area of the pen. The RFID-system continuously recorded time, RFID-tag, and the antenna that registered the tag (Van der Sluis *et al.*, 2020).

Video. An RGB-camera placed above the pen recorded the birds from 10:44 AM until 12:59 PM at 18 days of age. The two-hour period was recorded in multiple videos of approximately 7 minutes. The video recordings were made with a Zavio B6210 2MP camera (Zavio Inc., Hsinchu City, Taiwan) at 25 frames per second in full-HD.

Frame extraction and annotation. Five-hundred frames were randomly selected from all extracted frames for annotation. The selected frames were randomly divided between two annotators (Annotator A and B). The Computer Vision Annotation Tool (CVAT) (Sekachev *et al.*, 2019) was used for annotation. Three out of the four colour-marked birds were annotated, since the colour of one bird was inconsistent over days. The birds were labelled by bounding boxes, each bird having its own colour-based class (black, lightblue, pink). Inter- and intra-annotator reliability were calculated based on 25 randomly selected frames from the 500 frames, which were annotated twice by both annotators. Reliability was assessed with the intersection over union (IoU), calculated by dividing the area of overlap with the area of the union of the respective bounding boxes. Expressed in median IoU, the inter-annotator reliability was 0.92 (interquartile range [IQR] = 0.08), and the intra-annotator reliability was 0.95 (IQR=0.04) and 0.95 (IQR=0.07) for Annotator A and B, respectively.

Ground-truth. Five one-minute-long snippets were randomly sampled from 5 different seven-minute-long videos at 3 randomly sampled frames per second. The frames were randomly sampled to increase frame differences. From the five one-minute-long snippets, three movement events (5 to 10 seconds), one for each annotated bird, were annotated with bounding boxes for ground-truth.

You Only Look Once (YOLO). YOLOv5 was used to automatically detect the colour-marked birds in the videos (Jocher *et al.*, 2021). A pre-trained YOLOv5s model was trained on 350 randomly selected frames and tested on the remaining 150 frames, with mostly default hyperparameters, except for batch size (lowered to 6). Model performance, reported as mean average precision (mAP@[.5:.95]; value between 0 and 1 in which 1 is perfect), was 0.798 on the train set and 0.812 on test set.

Distance calculations. Animal walking distance was calculated for RFID, video, and ground-truth. The distance moved by the animal was based on the Euclidean distance between the centre points of the recorded antennas in the grid (RFID; Van der Sluis *et al.*, 2020), the centre points coordinates of the detected bounding boxes of YOLOv5s (video) or the centre points coordinates of the annotated bounding

boxes (ground-truth). The Euclidean distance was calibrated using one side of the pen. The animal was assumed to have remained at the last recorded RFID-antenna if no new read was received. The centre point coordinates of the detected and annotated bounding boxes were also translated to RFID-antennas and distances were calculated with the same method as described for the RFID-system (i.e. as RFID).

Results

Walking distance validation. In all movement events, the walking distance recorded by ground-truth pixel-based and video pixel-based agreed within 2 cm (Table 1). Overall, the RFID-system approximated the ground-truth pixel-based and video pixel-based, except for the black-marked broiler where it underestimated the movement with 75 cm. Ground-truth as RFID and video as RFID agreed, RFID reported an antenna switch for the pink-marked broiler and missed switches for the black-marked broiler.

Total walking distance. Overall, the recorded total walking distance substantially differed between RFID and video (Table 2). Compared to video, the RFID-system underreported the total walking distance of the animals. However, both RFID and video gave similar animal rankings, although not when the total walking distance on video was calculated as RFID.

Discussion

The aim of this study was to validate the automated tracking of broilers using video or RFID by comparing the total walking distance to ground-truth from video. Furthermore, video and RFID were compared to assess tracking performance over the entire duration of the trial. This study was the first time we were able to track broilers over time on video with computer vision, and in a setting that allowed us to visually confirm the location of individual broilers

Overall, the estimated walking distances by RFID and video pixel-based closely matched ground-truth pixel-based. However, for the black-marked broiler, RFID deviated with 75 cm compared to ground-truth pixel-based. The RFID-system seems to underestimate the total walking distance of the birds compared to video (Table 2). The RFID-system does not register movement within an antenna and can only register switches between antennas. However, not all antenna switches were registered, something that could be related to the hardware, the software or both. Similar observations were made in Van der Sluis *et al.* (2020).

Table 1. Walking distance (m) per movement event per broiler as recorded by ground-truth, RFID and video.

	Ground-truth		RFID	Video	
Broiler	as RFID	Pixel-based		as RFID	Pixel-based
Black	0.83	1.11	0.36	0.83	1.13
Light blue	0.36	0.34	0.36	0.36	0.35
Pink	0.0	0.48	0.45	0.0	0.48

Table 2. Total walking distance (m) per broiler as recorded by RFID and video.

	RFID	Video	
Broiler		as RFID	Pixel-based
Black	24.10	74.99	59.22
Light blue	24.79	63.68	68.68
Pink	27.85	94.42	80.58

Video pixel-based had a close-match with ground-truth pixel-based (≤2 cm difference, Table 1). Video could potentially overestimate the total walking distance over longer periods of time due to cumulative detection noise, but it could also be due to posture changes while lying and pecking behaviour (Table 2). This potential overestimation could be solved with a minimum movement threshold.

In conclusion, the estimated walking distances by video pixel-based closely matched ground-truth pixel-based (≤2 cm difference). Overall, the RFID-system also closely matched ground-truth pixel-based, except for one movement event (75 cm underestimation). Over the entire duration of the trial, both RFID and video gave similar animal rankings, but RFID seems more error prone due to missing reads. Nevertheless, RFID is suitable to track broilers over time, but video could be more accurate and practical for on-farm application. The study improved the understanding of both sensor systems and can help to develop a system for automated large-scale individual broiler behaviour recording.

Ethical statement

Data were collected under control of Cobb Europe. Cobb Europe complies with Dutch legislation on animal welfare. The Animal Welfare Body of Wageningen Research confirmed that this study was not an animal experiment under the Law on Animal Experiments.

Funding

This study was financially supported by the Dutch Ministry of Economic Affairs (TKI Agri and Food project 16022) and the Breed4Food partners Cobb Europe, CRV, Hendrix Genetics and Topigs Norsvin.

References

Dawkins, M. S., Wang, L., Ellwood, S. A., Roberts, S. J., & Gebhardt-Henrich, S. G. (2021). Appl. Anim. Behav. Sci., 234. https://doi.org/10.1016/j.applanim.2020.105180

Hart, B. L. (1988). Neurosci. Biobehav. Rev., 12(2), 123–137. https://doi.org/10.1016/S0195-5616(91)50028-0

Jocher, G., Stoken, A., Chaurasia, A., Borovec, J., *et al.* (2021). ultralytics/yolov5: v6.0. Available at https://doi.org/10.5281/ZENODO.5563715

Li, G., Zhao, Y., Purswell, J. L., Du, Q., *et al.* (2020). Comput. Electron. Agric., 175(May). https://doi.org/10.1016/j.compag.2020.105596

Reiter, K., and Bessei, W. (2001). Proc. of the 6th European Symposium on Poultry Welfare, Zollikofen, Switzerland.

Sekachev, B., Manovich, N., & Zhavoronkov, A. (2019). Computer Vision Annotation Tool. Available at https://doi.org/10.5281/zenodo.3497106

Van der Sluis, M., De Haas, Y., De Klerk, B., Rodenburg, B. T., & Ellen, E. D. (2020). Sensors, 20(13). https://doi.org/10.3390/s20133612

Van der Sluis, M., De Klerk, B., Ellen, E. D., De Haas, Y., *et al.* (2019). Animals, 9(8). https://doi.org/10.3390/ani9080580.

Van Hertem, T., Norton, T., Berckmans, D., & Vranken, E. (2018). Biosyst. Eng., 173, 93–102. https://doi.org/10.1016/j.biosystemseng.2018.07.002

Yang, X., Zhao, Y., Street, G. M., Huang, Y., *et al.* (2021). Animal, 15(7), https://doi.org/10.1016/j.animal.2021.100269

123. Automated processing and phenotype extraction of ovine medical images using deep learning and computer vision

J.F. Robson*, S.J. Denholm and M. Coffey

Scotland's Rural College, Peter Wilson Building, King's Buildings, Edinburgh, Scotland, EH9, United Kingdom; james.robson@sruc.ac.uk

Abstract

In this study, we applied deep neural networks, in the form of a generative adversarial network, to perform image-to-image processing steps needed for ovine phenotype analysis from CT scans of sheep. Key phenotypes such as gigot geometry and tissue distribution were determined using a computer vision pipeline. The results of the image processing using a trained network are strikingly similar (a similarity index of 98%) when used on unseen test images. The combined pipeline was able to process and determine the phenotypes at a speed of 0.11 s per medical image compared to approximately 30 min for manual processing. We hope this pipeline represents the first step towards automated phenotype extraction for ovine genetic breeding programmes.

Introduction

Increase in global food demand has led to livestock breeders seeking to produce breeding lines more able to match economic demand which have genetic advantages to primary traits such as growth speed and reduced feed intake. Recent improvements in genomic technologies such as detection of single nucleotide polymorphisms and whole genome sequencing have allowed unparalleled insight into the driving factors which guide animal phenotypes and successful genomic breeding selection has been able to identify traits which are desirable economically or welfare indicators. As the number and biological complexity of known phenotypes is increasing, there is a call to innovate new ways to detect phenotypes faster and more accurately in addition to detecting and preserving those of potential future relevance.

Machine learning, artificial intelligence and deep learning have been successfully implemented to increase phenotype detection speed and accuracy within many different medical areas including brain cancer and covid detection (Lundervold and Lundervold, 2019). Recently, the same technology has been applied to areas of agricultural science such as detection of bovine tuberculosis status based upon milk spectral data (Denholm *et al.*, 2020).

Similar networks called generative adversarial networks (GANs) can also be trained to translate images, rather than classify or categorise, and have traditionally been used for a variety of image translational tasks including noise removal from low-dose CT, motion artefact correction and produce CT images from MR images (Goodfellow *et al.*, 2020).

By combining GANs with another machine learning technique, computer vision (CV), any images generated by the GAN can then be analysed to extract data of interest in a fully automated way. Computer vison is a research field which aims to extract understanding or context from images and can use both traditional mathematical regression techniques as well as deep learning classification networks. Application of CV can range from simple inspection of food quality and ripeness by counting the number of pixels within images of fruit and vegetables which fall within certain colour hue ranges up to more complex tasks such as identifying road signage or pedestrians to guide automated driving systems (Deva Koresh, 2019). The aim of the present study was to develop a completely automated system for the extraction of phenotypes relevant for breeding programmes.

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_123, © J.F. Robson *et al.* 2022

Materials & methods

Development of the deep learning framework involved an initial manual data collection phase followed by network training, examination of network performance and development of computer vision code to explore phenotypes.

Ovine CT scan collection and tissue calibration. A single cross-sectional 2D images was taken through the top of the leg at the point of the ischium for each lamb using a Somatom Scope (Siemens, located at SRUC CT Unit, Edinburgh) with a slice thickness of 10 mm for a variety of breeds including Beltex, Blue Texel, Charollais, Hampshire Down, Meatlinc, Shropshire, Southdown, Suffolk and Texel as performed previously (Bunger *et al.*, 2011). The images at this stage are referred to as 'raw' images and contain non-subject objects such as scanning cradles. Tissue density windows corresponding to fat, muscle and bone were calculated based upon tissue dissection previously performed (Scholz *et al.*, 2015).

Manual image processing and phenotype analysis. Image objects superfluous for downstream phenotype analysis such as scanning cradles and testes were removed manually using STAR image processing software. Images at this point were then considered as 'ground truth'. From the ground truth images, tissue phenotypes could then be extracted by calculating tissue distribution and geometric phenotypes were measured manually in a 'click and drag' fashion. The total processing time from raw to phenotyped image took approximately 30 min.

Deep neural network (GAN) training and testing. A dataset containing 126 raw and ground truth image pairs of mixed breed ovine CT scans taken between 2019-2020 were used for GAN training. Raw and ground truth image pairs were first split into training (n=101) and validation (n=25) datasets (80 and 20% respectively). Thirty-two previously unseen raw CT scans taken from 2018-2019 and belonging to the breed Charollais were passed through the trained GAN model to produce 'predicted' images.

Calculation of image similarity. Mean squared error (MSE) and structural similarity index (SSI) metrics were used to compare the raw and ground truth images with the resulting predicted images using functions contained within the SciKit Image Python image processing library (Gandhi and Kulkarni, 2013).

Computer vision powered phenotype analysis. Automated phenotype extraction from ground truth and predicted images was performed using a pipeline which incorporated known pixel intensity value thresholding for each component of the carcass. Geometric phenotypes were computed by implementation of area, contour and perimeter functions within the SciKit Image library as part of a custom script (Van der Walt *et al.*, 2014).

Computing hardware and software. All computation was performed on a NVIDIA DGX Station using one TESLA V100 GPU running on CUDA 10.1. The GAN trained in this study is an implementation of AUTOMAP and Pix2Pix which has been optimised for use with paired image datasets (Kaji and Kida, 2019).

Results

The trained model was able to perform the major structural alterations within the image dataset needed to transform the raw CT scans into something which, by eye, strongly resembled the ground truth images as shown below in Figure 1a. Likeness of the raw, ground truth and predicted image sets (n=32) were compared pairwise using MSE and SSI. Both raw vs ground truth and raw vs predicted showed the lowest image similarity values with an average MSE of (58,674±17,766 and 58,008±17,319) and with average SSIs of (0.49±0.025 and 0.48±0.024) respectively indicating a high degree of image dissimilarity. On the other

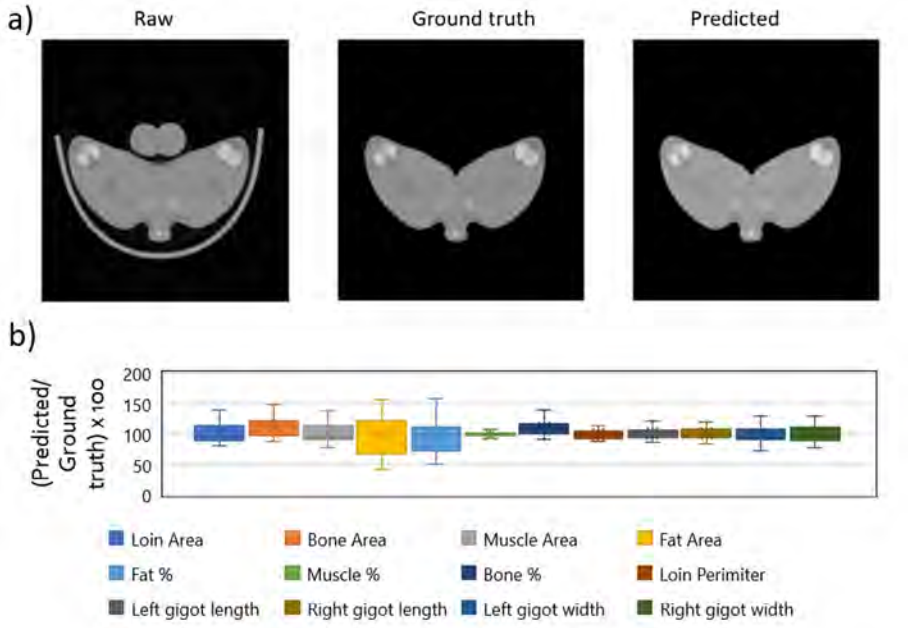


Figure 1. Performance of neural network and computer vision pipeline.

hand, comparing images in the ground truth and predicted datasets showed a much lower average MSE (1,028±1201) and a far higher average SSI of (0.98±0.0035) indicating a far greater similarity and indicated high accuracy of the trained model in mimicking the manual processing of CT scan images.

Phenotypes from both predicted and ground truth datasets were extracted using the computer vision pipeline and compared to determine the suitability of predicted images for phenotype (Figure 1b). Across all phenotypes, the average values were on average 101.44% that of the ground truth value with a standard deviation of 12.90% (n=32). Muscle % was the best performing predicted phenotype with estimated values between 93.67 and 106.65%. On the other hand, calculating fat area was the worst performing predicted phenotype with estimated values between 42.50 and 156.18%.

Discussion

In summary, we believe this research represents the first case of using an automated phenotype detection pipeline on agricultural animal medical images. This was achieved by using a combined GAN-CV pipeline to analyse agricultural medical images in a fully automated fashion. By feeding a paired image dataset into a GAN, we were able to perform the various image processing steps needed to produce a predicted image, containing only the relevant tissues, with accuracies of 98% (SSI) which rivalled that of manual processing and at a fraction of the cost. Phenotypes were then extracted or calculated from these predicted images by applying CV techniques as part of an automated pipeline.

We hope to immediately expand this highly accurate GAN-CV pipeline to process and extract phenotypes from other key CT scan sections such as the 8th thoracic vertebrae and 5th lumbar vertebrae positions. Further on, we hope to develop a pipeline to process a complete set of layered CT images to produce an accurate 3D model from which a multitude of phenotypes can then be extracted, such as spine length and vertebrate number, and detect phenotypes which are best explored in 3D space such as organ

morphology. By continuing this research, we will further expand the automated extraction of phenotypes from agricultural medical imaging data and use the findings to guide genetic and genomic breeding programmes.

References

AGandhi, S. and C.V.Kulkarni. (2013) *Int. J. Sci. Eng. Res.* 4(7): 930–934.

Bunger, L., Macfarlane, J.M., R., N., Conington, J., A., K., Moore, K., Glasbey, C.A., *et al.* (2011) *CT Scanning - Tech. Appl.* 329–348. https://doi.org/10.5772/23798

Denholm, S.J., Brand, W., Mitchell, A.P., Wells, A.T., Krzyzelewski, T., *et al.* (2020) *J. Dairy Sci.* 103(10): 9355–9367. Elsevier Inc. https://doi.org/10.3168/jds.2020-18328

Deva Koresh, J. (2019) *J. Innov. Image Process.* 1(01): 11–19. https://doi.org/10.36548/jiip.2019.1.002

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., *et al.* (2020) *Commun. ACM* 63(11): 139–144. https://doi.org/10.1145/3422622

Kaji, S. and Kida, S. (2019) *Radiol. Phys. Technol.* 12(3): 235–248. https://doi.org/10.1007/s12194-019-00520-y

Lundervold, A.S. and Lundervold, A. (2019) *Z. Med. Phys.* 29(2): 102–127. Elsevier B.V. https://doi.org/10.1016/j.zemedi.2018.11.002

Scholz, A.M., Bünger, L., Kongsro, J., Baulain, U. and Mitchell, A.D. (2015) *Animal* 9(7): 1250–1264. https://doi.org/10.1017/S1751731115000336

Walt, S. Van Der, Schönberger, J.L., Nunez-Iglesias, J., Boulogne, F., Warner, J.D., *et al.* (2014) *PeerJ* 2014(1): 1–18. https://doi.org/10.7717/peerj.453

124. VTag: automatic pipeline to annotate video data for pig phenomics studies

C.J. Chen[1,2], G. Morota[2] and H. Cheng[1*]

[1]Department of Animal Science, University of California, Davis, CA, USA; [2]Department of Animal and Poultry Sciences, Virginia Polytechnic Institute and State University, Blacksburg, VA, USA; qtlcheng@ucdavis.edu

Abstract

Due to the rising demand of animal products, computer vision (CV) has become a successful practice to automate monitoring important behaviors, including tail biting and lameness, in the pig industry. However, most CV systems are driven by supervised learning algorithms, the robustness of which heavily relies on high-quality hand-annotated data and laborious efforts. Therefore, we derived a software, Virtual Tag (VTag), to facilitate annotating video data for animal phenomics studies. By taking a top-view RGB video as input data, VTag can automatically derive annotations of pig position coordinates by the built-in algorithm. Users are allowed to further fine-tune the annotations through the graphical user interface in VTag and export the results for the training of other CV systems. Along with the coordinate annotations, the individual-level of pig activity is also generated as farming guidance. In sum, VTag can serve as a platform to accelerate pig behavioral studies.

Introduction

Due to the rising demand of animal products, precision agriculture, which aims to massively collect detailed measurements of animal traits, has been an essential practice to maximize the resource efficiency in the pig industry. The poor efficiency can be reflected in pig behaviors or the appearance. For example, pig lameness is an indicator of fractures, skin lesions or trauma, which have negative impact on the body condition and result in severe economic losses. Harmful animal behaviors such as tail-biting (Larsen *et al.*, 2019), aggression (Turner *et al.*, 2006), and excessive mounting (Teixeira and Boyle, 2014) also diminish animal welfare and reduce the production efficiency. Hence, to effectively monitor animal behaviors has become an important topic in the pig industry.

The technology of computer vision (CV), which is inspired by human vision that can intuitively focus on objects of interest and exclude noisy signals, has been used to automate monitoring the pig activity and shown promising results solely based on video information. For example, coupling with unmanned aerial vehicles (UAV), CV is possible to automate cattle counting in real-time with labor costs substantially reduced (Xu *et al.*, 2020), whereas it used to be costly to manage cattle in large-scale pasture lands. In smaller-scale indoor farms, CV systems were also used to detect body cleanness (Li *et al.*, 2019), structure (Liu *et al.*, 2020), positions (Yang *et al.*, 2018; Zhang *et al.*, 2019; Huang *et al.*, 2020), and behaviors (Okura *et al.*, 2019; Fuentes *et al.*, 2020; Ren *et al.*, 2021) for animal productions. However, the robustness of CV systems relies on laborious efforts to obtain a large amount of hand-annotated data, which, for example, store the positions (x-y coordinates) of studied objects from images by manual efforts. In addition, as the uniform annotation format for phenomics studies is not settled, the annotated data are difficult to be shared for community members to validate and extend. Outside precision agriculture, the availability of uniformly formatted and annotated image data has derived revolutionized applications; e.g. in the deep learning domain, annotated image databases, such as ImageNet (Deng *et al.*, 2009) and Cityscapes (Cordts *et al.*, 2016), have been heavily used in benchmarking newly developed models and derived applications for self-driving cars and cancer diagnostics. Current phenomics studies must solely rely on researchers' own resources due to the absence of standard pipeline or software that can generate uniformed annotated images. This limitation has become a significant difficulty for wider deployment of precision farming because of the costly work in generating annotated data.

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_124, © C.J. Chen *et al.* 2022

Therefore, in this paper, we derived a phenomics software, Virtual Tag, to automatically annotate pig positions using images captured by a top-view RGB video. The complete codes and content can be found on github.com/poissonfish/vtag. VTag is developed to meet three major goals:

- Automatically tracking pig positions without requiring training data set.
- Allowing users to annotate video data via a friendly graphical interface.
- Generating uniformed output format that is compatible in other analytical platforms.

Materials & methods

Automatic annotation pipeline. VTag enables tracking pig individuals automatically in two steps: semantic segmentation and instance segmentation. The step of semantic segmentation, which separates objects of interest (i.e. pigs in this study) from the background, uses pig motions to infer their positions. The space area with higher pixel variation across certain time frames indicates stronger observed motion. As the top-view video is expected to only capture group-housed pigs, pig positions should be highly overlapped with the area associated with strong observed motion. After segmenting the object of interest, the next step is to recognize pig identities. This step is also called instance segmentation. VTag extracts imagery or temporal describers, such as skin textures or moving trajectories, from the input video by multi-dimensional convolution operations to identify individual-level features. Then, the describers are used as clustering features in an unsupervised clustering algorithm for VTag to label different pig identities. After the two tracking steps, the annotations of pig position coordinates are generated automatically for the input video data. Through the VTag interface, users have access to the annotation results and is able to make necessary adjustments through the graphical user interface (GUI) of VTag by simple clicking and dragging. Therefore, annotating video data can be both efficient and accurate.

Properties and design of the annotated data. The annotations will be saved in the format of HDF5. This format has multiple merits for the research community: (1) multi-interface support: As much software for phenomics studies were released in R, Python, or GIS plugins, the format is readable in different interfaces for versatile analyses; (2) heterogeneous data integration: phenome annotations are a mixture of images and tabular data; the proposed format is able to store heterogeneous data and provide self-describing metadata for users to manage the collection of data sets; (3) partial data access: Imagery data usually is large in file sizes which imposes a critical burden to computing resources. With the support of partial data access, computers with limited memory are also capable of analyzing annotated phenome data. The output HDF5 has a data hierarchy listed in Table 1.

Results

The graphical interface has three components: the video preview, the playback controller, and the pipeline configuration. The preview (Figure 1a) will show the results of the annotated video for both pig contours and position centroids with a distinct color for each individual. Through the controller panel (Figure 1b), the video will play, pause, go forward or backward one frame by mouse-clicking the corresponding symbols, or pressing the arrows keys and space key on the keyboard. It is also possible to traverse the video to any frame by hovering the cursor on the playback progress bar. Each frame in the progress bar is colored in a gradient scale from yellow to blue, showing the levels of observed interaction between individuals. The

Table 1. Stored attributes of the annotated data.

Attributes	Data type	Description
Phenotypes	Tabular	Three-column data frame: object identity, trait name, phenotype.
Trial date	Text	Datetime format in MM-DD-YYYY or MM-DD-YYYY hh:mm.
Trial location	Text	Geo-Information in CRS, or general description of locations.
Trial scope	Text	Individual, house-grouped, field plots, etc.
DOI	Text	DOI number if it is published data.
Published	Text	Datetime format in MM-DD-YYYY.
Object positions (contour)	Numerical matrix	Four-dimension matrix: time frame by width by height by spectral channel.
Object positions (coordinate)	Tabular data	Three-column data frame: object identity, position coordinate, time frame.
Measurement	Text	RGB colors, near infra-red, depth, thermal, etc.
Measurement (details)	Numerial	One-dimension vector: scanned wavelengths
Studied species	Text	General name or Binomial nomenclature for studied species.

yellow color indicates the strong interaction and short distance between individuals are observed. On the right side of the interface is a panel for configuring the annotated results and outputs (Figure 1c). Several preview items can be adjusted including play speed (frames per second), annotation opacity, and displayed information. The output file paths and annotated attributes are also configurable in this section. In addition to the position coordinates, the outputs will cover several derived information, which includes the activity at specific time frame (Figure 1d), individual level of activity (Figure 1e), cumulated activity at specific time frame (Figure 1f), and spatial hot spots of pig movements (Figure 1g). In sum, VTag provides a friendly platform to annotate video data and generate informative farming guidance. These resources can pace an efficient way for further sophisticated work in behavioral research and animal phenomics studies.

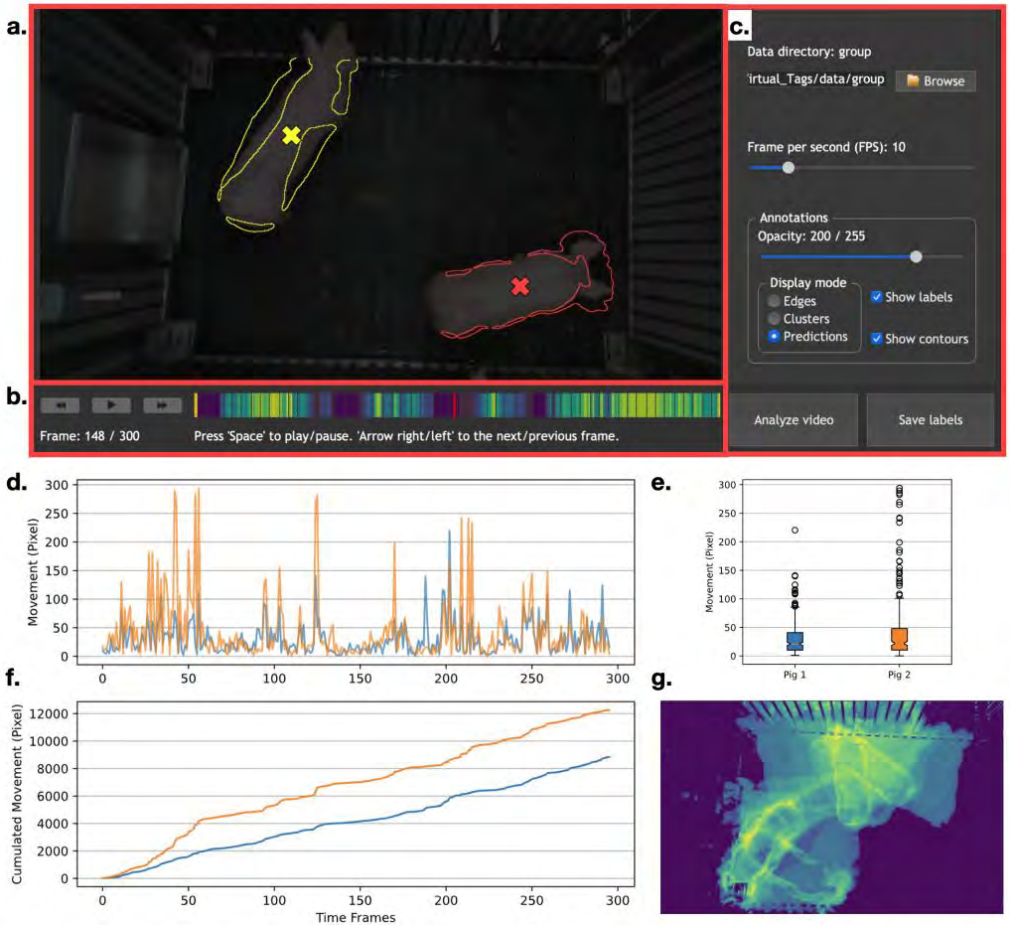


Figure 1. Overview of VTag. 1a-1c: Graphical user interface of VTag. 1d-1f: Individual-level of monitored unit time activity. 1g: Spatial hot spots of pig movement.

References

Cordts, M., M. Omran, S. Ramos, T. Rehfeld, *et al.* (2016). The Cityscapes Dataset for Semantic Urban Scene Understanding. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, Las Vegas, NV, USA. p. 3213–3223. Available from: http://ieeexplore.ieee.org/document/7780719/

Deng, J., W. Dong, R. Socher, L.-J. Li, *et al.* (2009). ImageNet: A large-scale hierarchical image database. In: 2009 IEEE Conference on Computer Vision and Pattern Recognition. p. 248–255.

Fuentes, A., S. Yoon, J. Park, and D. S. Park. (2020). Deep learning-based hierarchical cattle behavior recognition with spatio-temporal information. Computers and Electronics in Agriculture. 177:105627. https://doi.org/10.1016/j.compag.2020.105627.

Huang, W., W. Zhu, C. Ma, and Y. Guo. (2020). Weber Texture Local Descriptor for Identification of Group-Housed Pigs. Sensors. 20:4649. https://doi.org/10.3390/s20164649.

Larsen, M. L. V., H. M.-L. Andersen, and L. J. Pedersen. (2019). Changes in activity and object manipulation before tail damage in finisher pigs as an early detector of tail biting. Animal. 13:1037–1044. https://doi.org/10.1017/S1751731118002689.

Li, X., C. Cai, R. Zhang, L. Ju, and J. He. (2019). Deep cascaded convolutional models for cattle pose estimation. Computers and Electronics in Agriculture. 164:104885. https://doi.org/10.1016/j.compag.2019.104885.

Liu, H., A. R. Reibman, and J. P. Boerman. (2020). Video analytic system for detecting cow structure. Computers and Electronics in Agriculture. 178:105761. https://doi.org/10.1016/j.compag.2020.105761.

Okura, F., S. Ikuma, Y. Makihara, D. Muramatsu, *et al.* (2019). RGB-D video-based individual identification of dairy cows using gait and texture analyses. Computers and Electronics in Agriculture. 165:104944. https://doi.org/10.1016/j.compag.2019.104944.

Ren, K., G. Bernes, M. Hetta, and J. Karlsson. (2021). Tracking and analysing social interactions in dairy cattle with real-time locating system and machine learning. Journal of Systems Architecture. 116:102139. https://doi.org/10.1016/j.sysarc.2021.102139.

Teixeira, D. L., and L. A. Boyle. (2014). A comparison of the impact of behaviours performed by entire male and female pigs prior to slaughter on skin lesion scores of the carcass. Livestock Science. 170:142–149. https://doi.org/10.1016/j.livsci.2014.09.026.

Turner, S. P., I. White, S. Brotherstone, M. Farnworth, *et al.* (2006). Heritability of post-mixing aggressiveness in grower-stage pigs and its relationship with production traits. Animal Science. 82:615–620. https://doi.org/10.1079/ASC200678.

Xu, B., W. Wang, G. Falzon, P. Kwan, *et al.* (2020). Automated cattle counting using Mask R-CNN in quadcopter vision system. Computers and Electronics in Agriculture. 171:105300. https://doi.org/10.1016/j.compag.2020.105300.

Yang, A., H. Huang, C. Zheng, X. Zhu, *et al.* (2018). High-accuracy image segmentation for lactating sows using a fully convolutional network. Biosystems Engineering. 176:36–47. https://doi.org/10.1016/j.biosystemseng.2018.10.005.

Zhang, L., H. Gray, X. Ye, L. Collins, and N. Allinson. (2019). Automatic Individual Pig Detection and Tracking in Pig Farms. Sensors. 19:1188. https://doi.org/10.3390/s19051188.

125. Milk spectra selection for equations development to predict milk technological traits

M. Frizzarin[1,2], I.C. Gormley[2], A. Casa[2] and S. McParland[1*]

[1]Teagasc, AGRIC, Moorepark, Fermoy P61 P302, Co. Cork, Ireland; [2]School of Mathematics and Statistics, UCD, Ireland; sinead.mcparland@teagasc.ie

Abstract

Using all observations in a dataset to predict a phenotype may be sub-optimal for poorly represented predictands. Here we investigate the use of only observations similar to the predictand to predict milk technological traits using mid infrared spectrometry (MIRS) data. Similarity between observations and the predictand were judged according to: (1) the correlation between the MIR spectrum of the predictand and the other available spectra (the LOCAL approach); (2) the Mahalanobis distance between the predictand spectrum's principal components (PCs) and the other spectra PCs (the local changepoint approach; LCPA). Results are compared to the prediction performance of the traditional approach whereby the entire dataset is used to inform prediction (the global approach). The global approach provided the lowest root mean square error of validation (RMSEV) for 5 traits, while the LCPA had the lowest RMSEV for the remaining trait. When only observations with spectra poorly represented in the dataset were considered, the LCPA had the lowest RMSEV for 2 traits.

Introduction

The use of mid-infrared spectrometry (MIRS) data to predict milk and animal traits is well established (De Marchi *et al.*, 2014). The benchmark statistical method used to relate the spectra to a phenotype is partial least squares regression (PLSR; Wold 2001) whereby, in general, all available data are used in the development of the PLSR prediction equation. However, if the dataset analysed contains observations grouped in clusters or non-linear associations are present between the spectra and the trait of interest, the utilization of all the available observations may present some limits (Lesnoff *et al.*, 2020). These limits can be overcome using statistical approaches such as cluster methods (Franzoi *et al.*, 2018) or local methods (Shenk *et al.*, 1997), both of which use only part of the available observations in a dataset. Cluster methods identify a specific number of clusters in the dataset, and then develop a specific equation for each of these clusters. Local methods group observations in a dataset according to similarity between the predictand (neighbours) and other observations. The same number of neighbours are generally selected for all the observations in a dataset.

The objective of the present paper was to test a new local approach, the local changepoint approach (LCPA), to predict milk technological traits using MIRS data. The novelty of the tested approach is that the selected neighbours were identified in an objective, data driven manner, and the number of neighbours changed for each observation according to its characteristics. The prediction performance of the novel approach was then compared to global PLSR and the local method developed by Shenk *et al.* (1997; LOCAL).

Materials & methods

Data and data editing. Milk samples (n=713) were collected from 622 cows differing in breed, parity, and stage of lactation. Samples were analysed using a MilkoScan FT6000 and the resulting spectra were stored; samples were subsequently assessed to quantify rennet coagulation time (RCT), curd firming time (k20), curd firmness at 30 and 60 minutes (a30 and a60), casein micelle size (CMS), and pH.

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_125, © M. Frizzarin *et al.* 2022

A single observation for each cow was retained, identified as that with the greatest Mahalanobis distance between the principal component scores (PCs) of the observation and the average PCs scores of the entire dataset. Spectra were transformed from transmittance to absorbance, and noise regions were deleted. Outliers for the milk technological traits were identified as those greater than 3 standard deviations from the trait mean. Only observations with values for all the 6 traits were used in the subsequent analyses. The final dataset comprised 348 samples.

Statistical methods. Three approaches to perform prediction were tested; PLSR equations were developed using the statistical software R (R core team, 2020) for all approaches. Neighbours were identified according to: (1) the correlation between the predictand spectrum and the other spectra in the dataset (the LOCAL approach); and (2) the Mahalanobis distance between the predictand spectrum PCs and the other spectra PCs (LCPA). Results were compared to the global approach whereby all samples in the dataset were considered as neighbours.

The LOCAL approach tests different numbers of neighbours and number of PLSR factors, and selects for the final analyses the combination of neighbours and factors which produces the lowest root mean square error (RMSE). These tuning parameters are then used for all the observations in the dataset.

During the LCPA analyses, the PCs of the spectra was calculated with both standardized and unstandardized spectra, and the Mahalanobis distance between the first 4 PCs of the spectra was used to measure the similarity. Then, to identify the optimal set of neighbours, changepoint analysis, which identifies a single or multiple changes in the mean of a vector, was used (Killick *et al.*, 2012). In the context of neighbours' identification, the vector of Mahalanobis distances between the predictand and the other spectra was used after being sorted from the smallest to the largest values. Changepoint analysis identified multiple points where the mean in the similarity vector changed and the first changepoint, which selected at least 20 observations, was used to select neighbours. During the changepoint analysis, the PELT method associated with the SIC and MBIC penalties were tested.

Analyses. During the global PLSR analyses, all samples in the dataset, with the exception of the predictand, were used to identify the optimal number of PLSR factors using leave-one-out (LOO) cross validation (CV). Then, all dataset observations (except the predictand) and the optimal number of factors were used to predict the predictand trait value.

Nine different LOCAL equations were developed. In each equation a different combination of number of neighbours and number of PLSR factors were considered. The number of neighbours used was 25, 50, and 100, and the number of PLSR factors used was 5, 10, and 20.

For the LCPA, neighbours were identified using the changepoint analysis, and were used to define the optimal number of PLSR factors using LOO CV. Then the neighbours and the optimal number of PLSR factors were used to predict the predictand trait values.

Measuring prediction performance. The prediction performance of the different approaches tested was quantified using the RMSEV and the correlation between the true and the predicted trait values (r).

Results

Prediction performance. Table 1 summarizes the RMSEV and r for the different approaches tested. Regardless of the approach employed, pH consistently had the greatest r, while CMS consistently had the lowest r. When comparing the predictions across approaches, the global approach had the lowest RMSEV

for 5 traits (i.e. RCT, a30, a60, CMS, pH), while the LCPA using the MBIC penalty without standardizing the spectra before the PCs calculation had the lowest RMSEV for k20. If only the local approaches are compared, the LCPA had the lowest RMSEV for 4 traits (i.e. RCT, k20, a30, CMS), while the LOCAL approach had the lowest RMSEV compared to LCPA for a60 and pH.

Prediction for spectra at the edge of the principal component space. Prediction performance for the 5% spectra with the greatest Mahalanobis distance between the PCs of the predictand and the average PC scores of the entire dataset was considered. For such outlying samples, the global approach had the lowest RMSEV compared to other approaches for 4 traits (i.e. RCT, a30, a60, and pH), while LCPA had the lowest RMSEV compared to the other approaches for k20 and CMS. The LCPA outperformed the LOCAL approach for all the traits with the exception of pH.

Table 1. Root mean square error of validation (RMSEV) and correlation (r) between the true and predicted trait values obtained from global, LOCAL, and local changepoint approaches to Partial Least Squares Regression Analysis.

	Global		Local		Local changepoint	
Trait	RMSEV	r	RMSEV	r	RMSEV	r
RCT, min	**5.74**	**0.51**	6.29	0.35	6.23	0.44
k20, min	2.55	0.47	2.71	0.42	**2.54**	**0.50**
a30, mm	**12.04**	**0.64**	13.25	0.54	13.01	0.56
a60, mm	**11.12**	**0.46**	11.41	0.38	12.38	0.33
CMS, mm	**24.17**	**0.32**	25.22	0.21	25.16	0.22
pH	**0.060**	**0.77**	0.065	0.73	0.068	0.72

[1] Bold font identifies the lowest RMSEV and highest correlation value for each trait.

[2] Results for the LOCAL and local changepoint approach are those produced from the optimal combination of tuning parameters.

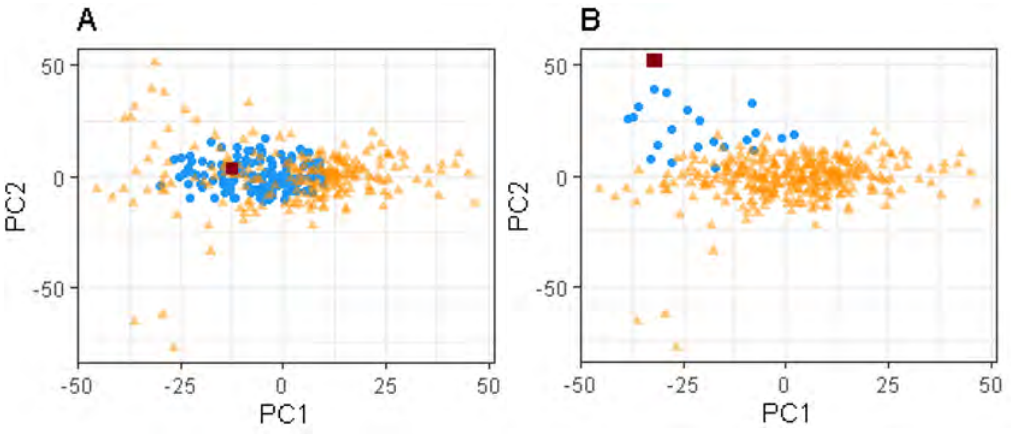


Figure 1. Neighbours selected according to the MBIC penalty after standardising the spectra for a target spectrum (red square) that lies in the centre ((A) n=133, blue dots) and at the edge ((B) n=21, blue dots) of the four-dimensional principal component space. The yellow triangles represent the other spectra not selected as neighbours.

Discussion

The local approaches presented in the literature generally select a fixed number of neighbours for use in prediction, regardless of the characteristics of the predictand. However, the number of samples with similar characteristics to the predictand may differ, particularly when we compare samples at the centre and at the edge of the dataset PCs space. In this study, samples at the centre of the PCs space had a greater number of neighbours identified when compared to samples at the edge of the PCs space (Figure 1). The sample with the greatest number of neighbours (133) was the 25th most central point (Figure 1A), while the sample with the lowest number of neighbours (21) was the fourth farthest point from the centre (Figure 1B). The benefit of varying the number of neighbours according to the characteristics of the spectra can be partly visible looking at the RMSEV when the 5% of observations with the greatest Mahalanobis distance from the centre were analysed.

Local methods have been successfully used to improve prediction analyses when multiple datasets were considered (Lesnoff *et al.*, 2021). However, those datasets comprised samples which originated from different sources and when these datasets were plotted in their first 2 PCs space, clusters were generally evident. The dataset considered in the present paper is composed only of milk samples with no obvious clusters in the PC space. This milk dataset can be thus considered an homogeneous dataset.

The application of local approaches in homogeneous datasets presents some limits, as reported also by Rabenarivo *et al.* (2013) who used local methods to predict soil traits. They reported better prediction results using the global approach compared to the local approach, with few exceptions. The limits of the local approaches when applied to homogeneous datasets can be related to the fact that all the observations in the dataset are similar to each other, and thus are informative during the prediction. In more heterogeneous datasets, the utilization of the local approaches can identify sub-groups and exclude uninformative samples during the predictions.

Conclusions

The local changepoint approach is a data driven and objective method for neighbours' selection, and in the present dataset improved the performance compared to the LOCAL approach. However, compared to the global approach, the local approaches had poorer predictions, probably due to the homogeneity of the dataset. The local approach can be useful for some specific traits, and more in general with heterogeneous datasets.

References

De Marchi M., D., Toffanin V., Cassandro M. and Penasa M (2014) JDS 97, 1–16. https://doi.org/10.3168/jds.2013-6799

Franzoi M, C. L. Manuelian, L. Rovigatti, E. Donati, and M. De Marchi (2018) JDS, Volume 101, Issue 8, Pages 6835-6841, ISSN 0022-0302, https://doi.org/10.3168/jds.2017-14343.

Killick, R.; Fearnhead, P.; Eckley, I.A. (2012) J. Am. Stat. Assoc., 107, 1590–1598.

Lesnoff, M., M. Metz, J.M. Roger. (2020) J. Chemom. 34 (5) e3209. https://doi.org/10.1002/cem.3209

R Core Team (2020). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/

Rabenarivo, M.; Chapuis-Lardy, *et al.* (2013) J. Near Infrared Spectrosc. 2013, 21, 495–509.

Shenk J. S., M. O. Westerhaus, and P. Berzaghi (1997) J. Near Infrared Spectrosc. 5, 223-232.

Wold S, Sjöström M, Eriksson L. (2001) Chemom Intel Lab Syst. 58(2):109-130. https://doi.org/10.1016/S0169-7439(01)00155-1.

126. Using supervised machine learning for honey harvest prediction

J. Ramirez-Diaz[1,2*], A. Manunza[1], T.A. De Oliveira[3], T. Bobbo[1,4], V. Silva[3], P. Cozzi[1], S. Biffani[1], A. Stella[1] and G. Minozzi[5]

[1]Institute of Agricultural Biology and Biotechnology – Italian National Research Council – CNR, Via Edoardo Bassini 15, 20133 Milan, Italy; [2]Department of Animal Science, Food and Technology – DIANA, Università Cattolica del Sacro Cuore, Via Emilia Parmense 84, 29122 Piacenza, Italy; [3]Statistics Department, Paraíba State University, Rua Baraunas 351, 58429-500 Paraíba, Brazil; [4]Department of Agricultural and Environmental Sciences (DISAA), University of Milan, Via Celoria 2, 20133 Milan, Italy; [5]Department of Veterinary Medicine DIMEVET, University of Milan, Via dell'Università 6, 26900 Lodi, Italy; johanna.ramirezdiaz@ibba.cnr.it

Abstract

Honeybees are a critical pollinator with a high economic and social impact on farming and natural ecosystems. Suitable climatic and ecological conditions play a crucial role on honey production. Machine learning has created an opportunity for predicted patterns in complex data structures (e.g. non-linear associations). We used 349 honey harvest records (kg) and 1,047 records of temperature, precipitation and atmospheric pressure from 6 apiaries collected between 2016 and 2019. Linear regression, Random Forests and eXtreme Gradient Boosting – XGBoost algorithms were employed to predict total honey harvest (THH). The lowest RMSE in testing set was estimated by XGBoost algorithms. Although the features were likely not enough to adequately capture the total variation of honey harvest the model performed well within its constraints. Using supervised machine learning was possible to predict THH with the error of 8.24 kg. More accurate phenotypes may better serve estimates of breeding values and improve selection outcomes.

Introduction

The *Apis mellifera* species are widely distributed in diverse climates around the world as a result of being potential for honey production. Climatic conditions have a direct influence on inside/outside colony activities like thermoregulation (Stabentheiner *et al.*, 2010), egg laying rate (Abou-Shaara *et al.*, 2017), food storing (Seeley and Visscher, 1985), flight and foraging activity (Haftom Gebremedhn *et al.*, 2014) and honey production (Neupane and Thapa, 2005). One of the biggest challenge's beekeepers face is estimating the effects of extremely low temperatures and high humidity during the winter season, which can lead to a significant reduction in honey harvest. Machine Learning Methods (MLM) have been useful to predicted complex biological data in agriculture (Liakos *et al.*, 2018), in meteorology (Choi *et al.*, 2018) and ecology (Peters *et al.*, 2014). In honeybees, MLM have been used for predicting health traits (Bjerge *et al.*, 2019), behaviour (Blut *et al.*, 2017), genetics (Fathian *et al.*, 2007), queen body mass (Prešern and Smodiš Škerl, 2019) and honey production (Karadas and Kadırhanogullari, 2017). Supervised regression models (SRM) are a group of learning algorithms that allow predicting targets (dependent variables) from a set of variables or features (predictors). The learning process is divided into training and testing steps. In training step, a data sample is taken as input in which features are learned by algorithm and build the learning model, then this model is used to make the prediction at new data during testing step (Nasteski, 2018). The next step is to choose the right metrics to assess its effectiveness. Our objective was to develop a predictive model for honey harvesting that incorporates the climatic conditions of winter prior to harvest. This approach can help beekeepers to identify potential limitations and establish new strategies to adaptively manage beehives in the coming seasons.

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_126, © J. Ramirez-Diaz *et al.* 2022

Materials & methods

Data sources. We used 349 honey harvest records (kg) from hives distributed in 6 apiaries collected between 2016 and 2019 and located in Lombardy region (Italy). The total honey harvest (THH) is the total weight of honey produced per hives per apiary between spring (March) to summer season (June). The number of hives with honey harvest records by apiary(year) were: Apiary A: 41(2019), Apiary B: 27(2016) and 41(2017), Apiary C: 31(2016), Apiary D: 32(2016) and 16(2017), Apiary E: 100(2018) and 30 (2019) and Apiary E: 31(2016). Geographic coordinates (latitude and longitude) were used to retrieve climatic data from meteorological stations for each apiary. Daily records of the atmospheric pressure (Hpa), temperature (°C) and rainfall (mm) between 2015-2019 for the winter prior to harvest were recovered from of the Copernicus Earth Observation program (https://www.copernicus.eu/it). The variation of temperature, rainfall and atmospheric pressure in the winter prior to honey harvest were 3.87 to 6.53 (°C); 66.10 to 100 mm and 5.23 to 7.00 Hpa (Table 1).

Preprocessing – exploratory analysis and standardization. For exploratory analysis, descriptive statistics (mean, standard deviation, interquartile range) were calculated (Table 2). Distributions, scatter diagrams and Pearson correlations were also estimated to identify outliers and to determine the relationship between climatic variables and THH. Subsequently, the hives with a THH lower than 5 kg were removed. A total of 1,047 observations from winter climatic conditions were included in the models. In addition, we rescaled the quantitative features using the z-score because the use of original scale may put more weight on the variables with a large range (Ahsan *et al.*, 2021).

Supervised machine learning analyses. For predictive analytics, three different SRM were used: Linear regression, Random Forest and eXtreme Gradient Boosting – XGBoost. Consequently, the database was divided into the training and testing set, with a chosen ratio of 70/30. The training set was used to create the initial models that they learned for each algorithm and to adjust the hyperparameters for cross-validation using 4 folds and 10 repetitions using the GridSearch technique that allow selected the best

Table 1. Climatic conditions in the winter prior to the honey harvest.

Apiary	Winter (December to March)	Climatic conditions		
		Temperature (°C)	Rainfall (mm/month)	Pressure (Hpa/month)
A	2018-2019	5.23	66.10	6.43
B	2015-2016	3.87	100.00	5.23
	2017-2018	3.97	85.30	5.23
C – D – F	2015-2016	6.40	90.00	6.93
D	2017-2018	6.53	69.20	7.00
E	2017-2018	4.37	80.90	6.13
	2018-2019	5.23	66.10	6.43

Table 2. Descriptive statistics of quantitative features and THH.

	Mean ± SD	Median	Q1	Q3
THH (kg)[1]	20.83±9.87	19.10	13.70	27.20
Temperature	5.10±3.06	5.09	3.80	6.80
Atmospheric pressure (Hpa)	6.27±0.99	6.30	5.40	7.00
Precipitation (mm)	81.83±23.88	84.30	70.50	103.72

[1] THH = total honey harvest (kg).

combination of hyperparameters. The root mean square error (RMSE) and the coefficient determination (R^2) were employed to measure model performance. RMSE is the most common metric for evaluating linear regression and represent the average magnitude of the error. All analyzes were performed using tidymodels package (Kuhn *et al.* (2020) in R v.4.0.3 (R core team 2020).

Results

Distribution of THH was asymmetric (mean: 20.83; median: 19.10). Correlations close to zero were found between the features and THH, indicating that there is no linear relationship between them. RMSE was used to evaluate the predictive performance of the models. The RMSE mean estimated was ±8.54 kg of THH (Table 3). The RMSE mean variation between training and testing set was close to 7%, this suggests that the algorithms used for the analysis of THH were consistent and that the use of winter season climatic variables are important on the prediction of THH.

In addition, a XGBoost had an optimal hyperparameters configuration that allowed a higher prediction performance. The best model was the eXtreme Gradient Boosting (XGBoost) with the lowest RMSE (8.24 on testing set) and R2 (coefficient determination) value of 0.29.

Discussion

In this study, the lowest RMSE (8.24 kg) was estimated fitting the XGboost algorithm that included apiary, year of harvest, and monthly observations of temperature, precipitation, and atmospheric pressure, which was similar to the value of MAE (mean absolute error; 10.3 kg) found by Campbell *et al.* (2020) using 115 climatic and remotely sensed features, and greater than the RMSE (1.85 kg) obtained by Karadas and Kadırhanogullari (2017) in models that included also the number of full beehives, working period in apiculture during year and control frequency of beehives in summer the results from 85 apiaries. The best algorithm should have the lowest values of RMSE (Ali *et al.*, 2015) and the highest estimation of adjusted R^2 (adj.R^2). In our results the estimation for adj.R^2 was 0.29 on XGboost algorithm, which was considerably smaller than the adj.R^2 reported by Karadas and Kadırhanogullari (2017) using MARS algorithm, nevertheless, our model considered less features and apiaries, and by Sakanovic *et al.* (2019) who included climatic variations such as carbon dioxide, temperature, oxygen levels and humidity present in apiaries using IoT (internet of things) technology. Although the features were likely not enough to adequately capture the total variation of honey harvest, the model performed well within its constraints and suggested an important effect of winter season climatic condition on the honey harvest. Finally, it is feasible to create a more informative algorithm that includes other climatic conditions, human management systems, and ecological and botanical information to find effective independent variables in beekeeping. A more informative algorithm could offer opportunities for the THH prediction and the development of better management and genetic/genomic selection strategies for the apiaries.

Table 3. Root mean square error (RMSE) for the models studied in training and testing.

Supervised regression models	Training set	Testing set
Linear regression	9.66	9.15
Random Forest	8.90	8.25
XGBoost[1]	8.87	8.24

[1] XGBoost = eXtreme gradient boosting algorithm.

Acknowledgements

This research was funded by the BEENOMIX 2.0 project funded by the Lombardy Region (FEASR program), PSR (grant number 201801057971 – G44I19001910002)

References

Abou-Shaara, H.F., Owayss A.A., Ibrahim Y.Y., and Basuny N.K. (2017) Insectes Soc. 64:455-463 https://doi.org/10.1007/s00040-017-0573-8

Ahsan, M.M., Mahmud M.A.P., Saha P.K., Gupta K.D., and Siddique Z. (2021) Technologies 9(3):52. https://doi.org/10.3390/technologies9030052

Ali M., Eyduran E., Tariq M.M., Tirink C., Abbas F *et al.*, (2015) Pakistan J. Zool. 47:1579–1585.

Bjerge K., Frigaard C.E., Mikkelsen P.H., Nielsen T.H., Misbih M *et al.*, (2019) Comput. Electron. Agric. 164:104898. https://doi.org/10.1016/j.compag.2019.104898

Blut C., Crespi A, Mersch D, Keller L., Zhao L *et al.*, (2017) Sci. Rep. 7(1):17663. https://doi.org/10.1038/s41598-017-17863-4

Campbell T., Dixon K.W., Dods K., Fearns P., and Handcock R. (2020) Agriculture 10(4):118. https://doi.org/10.3390/agriculture10040118

Copernicus, European Earth Observation program, Available at https://www.copernicus.eu/it

Choi C., Kim J., Kim J., Kim D., Bae Y *et al.*, (2018) Adv. Meteorol. 2018:1–11. https://doi.org/10.1155/2018/5024930

Fathian M., Amiri B., and Maroosi A. (2007) Appl. Math. Comput. 190(2):1502–1513. https://doi.org/10.1016/j.amc.2007.02.029

Haftom G., Alemayehu T., and Tesfay B. (2014) Livest. Res. Rural Dev. 26 (11). http://www.lrrd.org/lrrd26/1/haft26011.html

Karadas K., and Kadırhanogullari I.H. (2017) Pak. J. Zool. 49:1611–1619. https://doi.org/10.17582/journal.pjz/2017.49.5.1611.1619.

Kuhn M., and Wickham H. (2020) available at https://tidymodels.org

Liakos K., Busato P., Moshou D., Pearson S., and Bochtis D. (2018) Sensors 18(8):2674. https://doi.org/10.3390/s18082674

Nasteski V. (2017). Horizont 4:51-62. https://doi.org/10.20544/HORIZONS.B.04.1.17.P05.

Neupane K., and Thapa R. (2005) J. Inst. Agric. Anim. Sci. 26:143–148. https://doi.org/10.3126/jiaas.v26i0.667

Peters D.P.C., Havstad K.M., Cushing J., Tweedie C, Fuentes O *et al.*, (2014) Ecosphere 5(67)
https://doi.org/10.1890/ES13-00359.1

Prešern J., and Smodiš M.I. (2019) Apidologie 50:745–757. https://doi.org/10.1007/s13592-019-00683-y

Sakanovic S., Dogru N., Keco D., and Kevric J. (2019) 8th Mediterranean Conference on Embedded Computing (MECO). Budva, Montenegro

Seeley T.D., and Visscher P.K. (1985) Ecol. Entomol. 10:81–88. https://doi.org/10.1111/j.1365-2311.1985.tb00537.x.

Stabentheiner A., Kovac H., and Brodschneider R. (2010) PLoS One 5. https://doi.org/10.1371/journal.pone.0008967

127. Validation of computer vision algorithms for classifying video segments applied to behavioural phenotyping of pigs

J.P. Steibel[1*], J. Han[1], C. Chen[3], J. Siegford[1], T. Norton[2] and D. Colbry[1]

[1]Michigan State University, USA; [2]Katholieke Universiteit of Leuven, Belgium; [3]Jiangsu University, China; steibelj@msu.edu

Abstract

We compared validations for classification of behavioural interactions between pigs using computer vision and deep learning. We recorded interactions between pigs at a single-space feeder for 6 groups of 10 finishing pigs/group over the course of 6 weeks then compared different strategies to validate findings. Random cross validation is the most common procedure used in similar studies and consists of randomly partitioning video episodes into training and testing. This resulted in the highest classification accuracy (95%). Validation blocked-by-time used videos from early recordings for training and videos recorded three weeks later (same pigs) for testing (accuracy = 86%). Validation blocked-by-feeder used videos from one feeder for training and videos from another feeder (different pigs) for validation. This approach resulted in accuracies of 86 and 77% depending on which feeder was used for training and which feeder was used for validation. These results are worth considering when designing large scale behavioral.

Introduction

Computer vision is being used for automating behavioural phenotyping of pigs, including the detection of feeding and drinking events, interaction of pigs with enrichment objects and aggression between pigs (Chen, Zhu, Steibel, *et al.* 2020; Chen, Zhu, Oczak, *et al.* 2020). In most studies deep learning is used for automatically learning features from images and videos and for classifying them into different behavioural classes. Using this approach, good classifier properties under random cross validation are usually observed. One limitation for the applications of those approaches to large-scale phenotyping is the availability of training data, which typically requires extensive manual annotation of images and video segments. An important question is how to validate the resulting classifiers for large scale application in genetic evaluations involving large number of animals, groups, farms and timepoints. We investigated alternative cross validation schemes for classifying behavioural interactions between pigs at the feeder. The chosen problem has multiple potential application in animal breeding. For instance, detecting social interactions at the feeder can be used to parameterize competition effects in genetic evaluation models (Angarita *et al.* 2021).

Materials & methods

Experimental design. Two pens equipped with single-space automatic feed intake recording systems and digital video cameras were available for this study. Six groups of pigs were rotated through these pens every week for six consecutive weeks. Each group was recorded twice at different stages of the production cycle following the schedule in Table 1.

Table 1. Rotation schedule of social groups (SG) through feeders/pens.

	Week 1	Week 2	Week 3	Week 4	Week 5	Week 6
Feeder 1	SG 1	SG 2	SG 3	SG 1	SG 2	SG 3
Feeder 2	SG 4	SG 5	SG 6	SG 4	SG 5	SG 6

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_127, © J.P. Steibel *et al.* 2022

Within each week, the first 5 days were for (re) acclimation of pigs to the new pen and the last two days were used for video recording of pig behavior at the feeder during the daylight hours (approximately from 8 AM to 4 PM).

Video recording and labelling. Video was recorded using the RGB module of the Intel® RealSense™ D435 camera. Videos of the feeder region were cropped at a resolution of 982×238 pixels with a target frame rate of 30 frames per second. Video segments containing two pigs at the feeder were segmented into 30 frame units and manually annotated by a trained observer according to the observed interaction at the feeder. After an initial assessment of a limited number of videos, an ethogram was built to classify video segments into five types of interactions between pigs at the feeder: (1) direct head to body contact; (2) indirect/ unintended head to body contact; (3) mounting; (4) levering; and (5) no contact. After further checking and validation, we decided to merge the two head to body categories into a single category of head to body.

Classification algorithm. We used deep learning (DL) and followed previously published work on classification of video episodes for behavioural phenotyping (Li *et al.* 2021). In brief, our algorithm was comprised of two stages (Figure 1). First, a pre-trained convolutional neural network was used for feature extraction. Second, a long short-term memory DL algorithm was trained from scratch for classification of video segments using the features extracted in the previous step.

When training DL models, it is extremely important to carefully chose hyperparamter values including network architecture, optimizer options, etc. (Han *et al.* 2021). For this study we used a combination of previously suggested options and grid search using internal cross validation.

A particular challenge in this case was the imbalance of the behavioural classes. Levering and mounting were clearly less frequent interactions, however, they are particularly important ones as they often lead to immediate replacement of pigs competing for feeding space and are forcible physical interactions between the pigs. Yet, when training DL classifiers, it is important to keep classes balanced. Up sampling or down

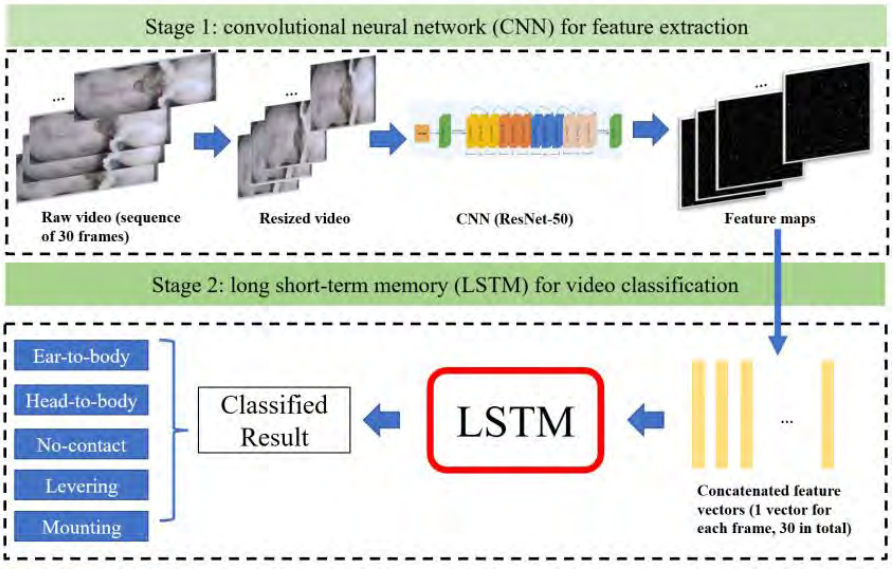


Figure 1. Deep learning pipeline for pig's aggressive behavior detection based on videos.

sampling are commonly used for this purpose. We were able to partially alleviate this problem through up sampling of the least frequent classes through data augmentation by selecting partially overlapping video segments of mounting and levering events. However, the imbalance persisted. Thus, we used a multi-class focal loss function that is specifically recommended for class-imbalanced data (Equation 1).

$$loss_{FL} = -\sum_{i=1}^{C} \alpha(1-y_i)^{\gamma} t_i log(y_i) \quad (1)$$

where C is the number of behavior classes, i is the class index, t_i is an indicator variable of correct/incorrect classification, y_i denotes the probability of behavioural class i (obtained from the last softmax activation function in the last layer), α is the focusing parameter, and γ is the balancing parameter.

The performance of the trained classifiers was assessed under cross validation (next section) using two metrics: class-specific recall (Equation 2) and class-specific precision (Equation 3).

$$recall_i = \frac{Number\ of\ correctly\ classified\ episodes\ for\ class\ i}{Total\ number\ of\ annotated\ episodes\ for\ class\ i} \quad (2)$$

$$precision_i = \frac{Number\ of\ correctly\ classified\ episodes\ for\ class\ i}{Total\ number\ of\ episodes\ predicted\ for\ class\ i} \quad (3)$$

Validation strategies. The structure of our data allowed assessing the properties of classifiers using different validation partitions. Under random cross-validation, five different random training and validation sets were selected. For validation blocking by time, video segments from the first three weeks were used for training and video segments from the last three weeks were used for validation. For the validation blocking by feeder, video segments recorded at one feeder were used for training and video segments recorded at another feeder were used for validation. To facilitate comparison of results, the size of the training set was kept fixed at 7,500 video episodes for all validation schemes and the validation was repeated five times by randomly selecting different test and validation video episodes, subject to the constraints of each validation scheme.

Results

Random cross validation. The performance of the multi-class classification algorithm of video segments under random cross validation was on par with previously published work using binary classification of video episodes for aggression, drinking and feeding. Overall classification accuracy was above 95%, precision ranged between 93% for levering and 97% for head-to-body interactions and recall varied between 85% for levering and 97% for mounting (Figure 2). This is remarkably good performance considering the complexity of multi-class classification with heavy imbalance among classes, as is the case in this example.

Validation blocked-by-time. Overall, using early episodes for training the algorithm to classify video episodes recorded later from the same social groups resulted in a significant loss in performance. Overall classification accuracy was 86±3%. Class-specific precision ranged between 35% (levering) and 97% (head-to-body contact) and recall ranged between 82% (no contact) and 96% (levering). The low precision in detecting levering was due to many head-to-body contact events that were classified as levering.

Validation blocked-by-feeder. When using video episodes from one feeder for training to classify video episodes recorded on another feeder, the performance dropped even further. But most remarkably, the performance differed between the two partitions in this validation strategy. Specifically, using video episodes recorded at feeder 1 (social groups 1, 2, 3) for training to classify video episodes recorded at feeder 2 (social groups 4, 5, 6) resulted in better accuracy (86±1%) compared to using episodes recorded at feeder 2 to classify video episodes recorded at feeder 1 (77±3%). Moreover, the specific pattern of mis-

classifications also differed between the two implementations of this validation. A principal component analysis of features extracted from frames and input into the LSTM classifier revealed that the first two principal components clearly separated video frames according to feeder at which they were taken. However, higher order principal components indicated that variation between frames taken at feeder 2 was less than variation between frames taken at feeder one. This hints at the possibility that training on the most variable set and predicting on the less variable set produces better performance compared to the opposite scenario.

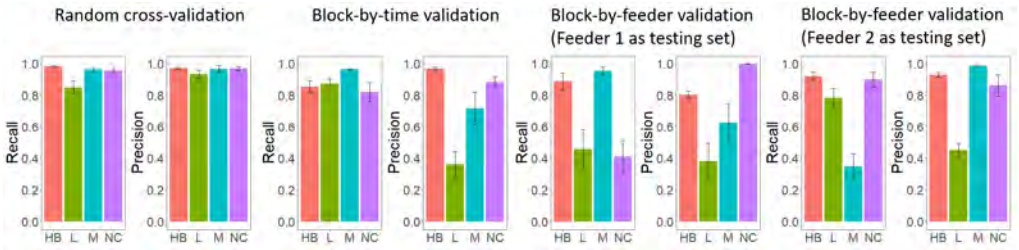


Figure 2. Recall and Precision of video classifiers of behavioral interactions under alternative validation strategies. HB: head to body, L: levering, M: mounting, NC: no contact.

Discussion

Deep learning algorithms combining transfer learning for feature extraction and di-novo training for classification are a very promising tool for video classification using computer vision for animal behavioural phenotyping. However, as these algorithms are developed and tested under realistic production conditions it is important to conduct validations that are commensurate with the intended applications. For instance, for research applications where a large amount of video is collected over time and across social groups for later analysis, random cross validation seems to produce the best results. On the other side, if 'on-the-flight' training is necessary, where videos segments recorded early are annotated and used to train algorithms to classify video segments recorded in later production stages, validation over time gives more realistic results than random cross validation. Validation blocking by time is also worth considering when training computer vision for animal identification algorithms that will be used through the production cycle of each individual. Finally, training classifiers on video recordings of certain groups of animals to be used in videos recorded on different groups will likely require larger training sets and more extensive validation as demonstrated in our study through the blocking-by-feeder scenario.

References

Angarita, B. K., J. Han, R. J. C. C. Cantet, S. K. Chewning, K. E. Wurtz, *et al.* 2021. J. Anim. Sci. 99(5): skab042.

Chen, C., W. Zhu, M. Oczak, K. Maschat, J. Baumgartner, *et al.* 2020a. Comput. Electron. Agric. 175(2020):105580.

Chen, C., W. Zhu, J. Steibel, J. Siegford, J. Han, *et al.* 2020b. Biosyst. Eng. 196(2020):1-14

Han, J., C. Gondro, K. Reid, and J. P. Steibel. 2021.. G3 Genes|Genomes|Genetics. 11(7): jkab032.

Li, G., Y. Huang, Z. Chen, G. D. Chesser, J. L. Purswell, *et al.* 2021. Sensors. 21(4):1–42.

128. Automated estimation of pose features in broilers using computer vision

I. Fodor[1*], J.E. Doornweerd[1], B. de Klerk[2] and A.C. Bouwman[1]

[1]*Wageningen University & Research, P.O. Box 338, 6700 AH Wageningen, the Netherlands;*
[2]*Cobb Europe, 5831 GH Boxmeer, the Netherlands; istvan.fodor@wur.nl*

Abstract

The importance of sensor-based phenotyping is increasing in the broiler industry, aiming at improved animal health and welfare, and lower economic losses. We analysed 11 pose features of 87 individual chickens at three ages (day 14, 21, and 33) using video recordings filmed from behind while the broilers walked through a corridor one by one. A pre-trained deep learning model was trained on a limited number of frames (n=181) to adapt it to a new environment for accurate keypoint detection. Extraction of the three poses of interest (double support, left and right steps) was fully automated. Significant (P<0.05) differences were observed among ages in hock joint angles, shank versus horizontal line angles, normalized drumstick and shank lengths, and hock/knee distance ratio. Recording walking broilers from behind proved to be useful in obtaining relevant pose features in an automated way.

Introduction

Poor leg health is a significant welfare and economic issue in the broiler industry (Knowles *et al.*, 2008). Sensor-based phenotypes collected at large scale can support the improvement of genetic gain of locomotion if incorporated in the breeding strategy. In a study of 51,000 broilers from 176 flocks, 27.6% of the birds showed poor locomotion (defined by gait score ≥3 on a 0-5 scale), and 3.3% were almost unable to walk (gait score ≥4; Knowles *et al.*, 2008). Several approaches have been proposed to monitor locomotion in poultry, including inertial measurement units (Bouwman *et al.*, 2020) and cameras (Aydin, 2017; Caplen *et al.*, 2012). In a study, broilers were filmed from above while walking through a corridor, and broilers with higher (≥3 on a 0-4 scale) gait scores had shorter steps, lower speed, and more pronounced lateral body oscillation (Aydin, 2017). In another study, the pose of standing broilers was compared to their gait using photogrammetry from 21 to 42 days of age (Mendes *et al.*, 2016). The aim of our study was to analyse the pose of individual broilers across ages using video images under semi-practical circumstances.

Materials & methods

Experimental setup and video recording. Individually tagged male broilers (n=109) from the same cross were filmed three times (at 14, 21, and 33 days of age) at a Cobb test facility in the Netherlands. During testing, birds had to walk one by one through a corridor created within their pen with dimensions 3 m (length) × 0.4 m (width). An Intel® RealSense™ Depth Camera D415 (Intel Corp., Santa Clara, CA, USA) camera was placed at the start of the corridor, which recorded the chickens from behind during walking. Some birds were removed from the experimental groups for other purposes or due to mortality, therefore, 87 birds remained for video recording by 33 days of age. The RGB video recordings of the camera were converted to .mp4 format with 12 FPS and 1,280×720 pixels resolution for further analysis using a custom Python code.

Keypoint detection. Eight keypoints (head, neck, left and right knees, hocks, and feet) were detected using a pre-trained broiler pose estimation deep learning model developed in DeepLabCut (Mathis *et al.*, 2018). Details of the pre-trained broiler pose estimation model are described in detail elsewhere (Doornweerd *et al.*, 2021). However, this pre-trained model struggled with accurately detecting the keypoints of the chickens in the new environment (e.g. different illumination, red drinkers above the head of the chickens,

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_128, © I. Fodor *et al.* 2022

other chickens at the end of the corridor). Therefore, video frames of some birds that had been removed from the current setup between day 14 and 33 were used for retraining of this model. Altogether, 181 frames were used for retraining (day 14: five birds, day 21: eight birds), whereas 40 frames (day 14: two birds, day 21: two birds) were used for testing the retrained model, which resulted in a 82/18 train-test split. This model was then used for the analysis of the newly acquired videos, along with a spline filter and dynamic cropping (the latter in day 14 videos only) to reduce noise in the keypoint estimates. Coordinates and likelihoods of the estimated keypoints were used for pose extraction and calculating the features.

Automated pose extraction. Birds that did not walk during the test (n_{D21}=1, n_{D33}=3) were excluded from further analyses. Briefly, the first and last 10% of the frames of each video were discarded. Then, maximum leg lift and feet being on the ground were detected, based on the local maxima and minima in the longitudinal data of keypoint coordinates. Based on these two distinct feet positions, pose proposals were generated for the three specific poses of interest: double support stance, left and right steps with maximum leg lift. Pose proposals were then sorted based on the normalized vertical difference between the left and right feet, using the vertical difference between the highest knee and lowest foot in the given frame for normalization. Frames with double support phase were selected by minimizing the vertical difference between the left and right foot. On the contrary, steps were selected based on maximizing the vertical difference between feet. For each bird-age-pose combination, the three best frames (based on the vertical differences as sorting criteria) were chosen to calculate pose features. The quality of automatically detected poses was checked visually by extracting the respective frames of 10 randomly selected birds from each age.

Pose features. Altogether, 11 features were defined, where the left and right sides of the same attribute counted as two separate features: normalized step height (%), hock joint lateral angles (degrees), medial angle of the shank and the horizontal line (degrees), normalized drumstick length, normalized shank length, and the ratio between the horizontal distance of hocks and that of the knees (Figure 1). The same normalization was performed as with pose proposals described above. Pose features of the three extracted best frames were averaged to represent the given feature of the bird at that specific pose at a given age.

Statistical analysis. Birds alive on day 33 were included in the analyses. Extracted features were analysed in R version 4.0.2 (R Core Team, 2020). Differences by age were tested using a linear mixed model with age as a categorical variable and chicken as a random factor, using the lme4 package in R (Bates *et al.*, 2015). Pairwise post hoc comparison of age groups was performed using Tukey contrasts. The level of significance was set to 0.05.

Results

Based on the visual check of the labelled keypoints using the original pre-trained model, label jumps between the bird of interest and birds in the background were more common at younger ages. However, label jumps between the drinker and the head occurred quite frequently irrespective of age. Further training

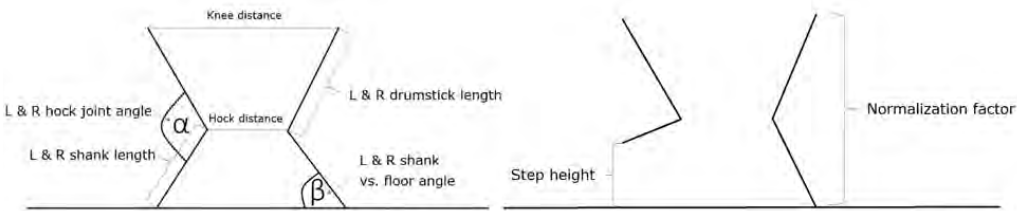


Figure 1. Schematic representation of the analysed features at double support (left) and steps (right).

of the pre-trained broiler pose estimation model substantially reduced the occurrence of these errors. The training and test errors of the final model were 2.11 pixels and 4.02 pixels, respectively, at 100,000 iterations.

The automated pose extraction algorithm correctly identified and selected the relevant poses. Examples of automatically extracted double support stance, left and right steps from day 14 are shown in Figure 2.

Sharper hock joint angles ($P<0.01$) and shank vs horizontal line angles ($P<0.05$) were observed on day 33 than on day 21 (Table 1). Normalized drumstick lengths were shorter ($P<0.0001$), whereas normalized shank lengths were longer ($P<0.05$) on day 21 and 33 than on day 14. The hock/knee distance ratio was larger ($P<0.01$) on day 21 and day 33 compared to day 14. No clear trends were observed regarding step height by age.

Figure 2. Automatically extracted double support stance, left and right steps on day 14 (images cropped and enlarged).

Table 1. Descriptive statistics of pose features by age.

Feature	D14		D21		D33	
	Mean	SD	Mean	SD	Mean	SD
Left hock angle (°)	153.7	7.1	153.6	6.6	150.1	8.2
Right hock angle (°)	154.5	5.8	157.2	5.5	154.0	6.6
Left shank vs horizontal line angle (°)	84.24	4.76	85.38	6.47	82.93	6.65
Right shank vs horizontal line angle (°)	83.06	5.94	83.58	4.55	81.18	6.33
Left drumstick rel. length	0.583	0.048	0.542	0.047	0.535	0.046
Right drumstick rel. length	0.601	0.047	0.564	0.046	0.572	0.037
Left shank rel. length	0.423	0.033	0.443	0.061	0.454	0.051
Right shank rel. length	0.411	0.037	0.449	0.047	0.439	0.040
Hock/knee distance ratio	0.586	0.094	0.625	0.058	0.644	0.066
Left step height (%)	36.47	6.53	39.08	6.79	39.20	8.71
Right step height (%)	39.20	7.08	36.00	7.36	32.08	9.59

Discussion

We demonstrated the usability of recording walking broilers from behind for monitoring certain locomotion features. Further training of a pre-trained pose estimation model with a limited number of frames improved its performance in the new environment. The automated pose extraction algorithm proved to be useful for high-throughput pose extraction. Although, further improvements could be made to this algorithm, because in some instances the double support stance was somewhat asymmetric, and not the highest leg lift was captured during some steps.

In contrast with the photogrammetry method used by Mendes *et al.* (2016), video recordings have the potential to be upscaled to practical circumstances. Using a top view of the chickens, Aydin (2017) calculated features such as speed, step frequency and length, and lateral body oscillation. In our setup, angles, hock-knee distance ratio, and step height were the most informative features, whereas relative drumstick and shank lengths were more heavily influenced by perspective. Further work could focus on the relationship of these video-derived features with gait.

Ethical statement

Data were collected under control of Cobb Europe. Cobb Europe complies with Dutch legislation on animal welfare.

Funding

This study was financially supported by the Dutch Ministry of Economic Affairs (TKI Agri and Food project 16022) and the Breed4Food partners Cobb Europe, CRV, Hendrix Genetics and Topigs Norsvin.

References

Aydin A. (2017) Comput. Electron. Agr. 136:140-146. https://doi.org/10.1016/j.compag.2017.02.019

Bates D., Maechler M., Bolker B., Walker S. (2015) J. Stat. Softw. 67(1):1-48. https://doi.org/10.18637/jss.v067.i01

Bouwman A., Savchuk A., Abbaspourghomi A. and Visser B. (2020) Front. Genet. 11:207. https://doi.org/10.3389/fgene.2020.00207

Caplen G., Hothersall B., Murrell J.C., Nicol C.J., Waterman-Pearson A.E., *et al.* (2012) PLoS ONE 7(7): e40800. https://doi.org/10.1371/journal.pone.0040800

Doornweerd J.E., Kootstra G., Veerkamp, R. F., Ellen E.D., van der Eijk J.A.J., *et al.* (2021) Front. Anim. Sci. 2:791290. https://doi.org/10.3389/fanim.2021.791290

Knowles T.G., Kestin S.C., Haslam S.M., Brown S.N., Green L.E., *et al.* (2008) PLoS ONE 3(2): e1545. https://doi.org/10.1371/journal.pone.0001545

Mathis A., Mamidanna, P., Cury K.M., Abe T., Murthy V.N., *et al.* (2018) Nat. Neurosci. 21, 1281–1289. https://doi.org/10.1038/s41593-018-0209-y

Mendes A.S., Paixão S.J., Sikorski R.R., Bonamigo D.V., Morello M.G., *et al.* (2016) Braz. J. Poult. Sci. 18(4):573-580. https://doi.org/10.1590/1806-9061-2015-0079

R Core Team (2020) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

129. Predicting carcass cut yields in cattle from digital images using artificial intelligence

D. Matthews[1*], T. Pabiou[1], R.D. Evans[1], C. Beder[2] and A. Daly[2]

[1]Irish Cattle Breeding Federation, Ballincollig, Co. Cork, Ireland; [2]Munster Technological University, Bishopstown, Cork, Ireland; dmatthews@icbf.com

Abstract

Deep Learning (DL) is a successful tool for many image classification problems but has yet to be applied to carcass images. The aim of this study was to train DL models to predict carcass cut yields and compare predictions to more standard machine learning (ML) methods. Three approaches were undertaken to predict the grouped carcass cut yields of Grilling cuts and Roasting cuts from a large dataset of 54,598 and 69,246 animals respectively. The approaches taken were: (1) animal phenotypic data used as features for a range of ML algorithms; (2) carcass images used to train Convolutional Neural Networks; and (3) carcass dimensions measured directly from the images, combined with associated phenotypic data and used as feature data for ML algorithms. Results showed that DL models can be trained to predict carcass cuts yields but an approach that uses carcass dimensions in ML algorithms performs slightly better in absolute terms.

Introduction

Currently, beef carcasses in the European Union are graded for payment and processing using the EUROP classification grid for conformation and fat. The degree of association between EUROP predicted scores and meat yield of a carcass, both on a total meat yield level or on a carcass cut level, is highly variable (Conroy *et al.*, 2009; Conroy *et al.*, 2010; Drennan *et al.*, 2008; Pabiou *et al.*, 2011) yet meat yield to the processor is the primary driver of a carcass' subsequent economic value. Nonetheless, payments to primary producers, price reporting, and trade to wholesalers are all still based on the EUROP system. It is therefore desirable at an industry level to have mechanisms to predict the actual meat yield of a carcass accurately.

From an animal breeding perspective, EUROP predictions are often used as genetic traits notwithstanding the variable association between EUROP predicted scores and a carcass' economic value. Connolly *et al.* (2019) established the benefit of including breeding values for meat yields in the Irish breeding program. However, meat yields have yet to be included as genetic traits in the Irish breeding program, mainly due to the difficulty in recording meat yields during carcass processing. If accurate meat yield predictions could be calculated these would offer a potential approach for including meat yields in breeding programmes and thus, allow producers to breed for higher yields of more valuable cuts of meat. In this study we have investigated the performance of Deep Learning (DL) to predict grouped carcass cut yields from images and compared the models to machine learning (ML) models using animal data and also using carcass dimensions measured directly from the carcass images.

Materials & methods

Data collection. Data from Irish cattle slaughtered for meat between 2011 to 2018 was available from a range of sources: carcass images and phenotypic data captured at different stages throughout the animal's lifetime were collated along with target variables of meat yields for different processed cuts of the carcass. The yield of specific meat cuts in kilograms were recorded during routine carcass processing in an Irish commercial abattoir. As such, the dataset contained steers, heifers, cows and young bulls. Also, the vast majority of animals were commercial crossbred animals with varying proportions of beef and dairy breeds. The meat cuts recorded and used in this study were the Silverside, Topside, Knuckle, Rump, Striploin, Fillet, and Cube-roll. Only cut data from animals that: (1) had a yield recorded for each of the 7 cuts

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_129, © D. Matthews *et al.* 2022

examined; (2) had a weight within 3 standard deviations of their respective cut mean; (3) matched to an animal record present in the ICBF database; and (4) also matched to a carcass image were included in the study. Similar to Pabiou *et al.* (2011), the meat cuts representing the highest value cuts of the carcass were grouped together into Grilling cuts (combined yield of striploin, fillet and cube-roll) and Roasting cuts (combined yield of silverside, topside, knuckle and rump). The final clean dataset contained 54,598 Grilling yields and 69,246 Roasting yields. Meat yields recorded in the abattoir were matched to animal records in the ICBF database and features deemed informative for this study were extracted. Categorical data was one-hot encoded into multiple features and joined to continuous features which were standardised. Two photographs of the carcasses were taken immediately after slaughter: a standard two-dimensional image (2D) and a structured light image (3D). The video image analysis (VIA) system used across Irish abattoirs is the VBS2000 (E+V GmbH, Germany) and prediction of carcass conformation and carcass fat on the EUROP scale uses proprietary VIA software. In this study, the 2D and 3D images collected in abattoirs are used to train novel models, but the propriety software was not.

Models. To investigate the predictive performance of models created in this study, two distinct datasets were created: a training set and a hold-out test set. The test set was approximately 10% of the total dataset and was chosen randomly. Three different approaches were examined for predicting Grilling and Roasting cuts largely determined by the suitability of the data type to a specific ML algorithm. The 3 approaches were: (1) phenotypic data as input features for a range of ML regression algorithms, (2) carcass images fed into a DL neural network adapted to predict a regression value rather than classification, and (3) measuring 346 carcass dimensions from the carcass image and concatenating these measurements with the phenotypic data available to use as features in a range of ML algorithms.

The Root Mean Squared Error (RMSE) was used as the main determinant of a model's predictive ability. Mean Absolute Error (MAE) and coefficient of determination (R^2) were also calculated as a final measure of accuracy.

Results

Of all ML algorithms tested in Approaches 1 and 3, Gradient Boosting performed the best and therefore those results are presented here. Table 1 presents the predictive performance of optimised models on the hold-out test set. As this test set has never been included in any training or validation set, it allows for a more direct comparison between the different predictive models. For Grilling yield, all models performed very similarly with small RMSE and MAE metrics relative to the Grilling yield mean of 35.1 kg. The difference between the best performing and worst performing models was 0.16 kg RMSE. Approach 1, utilising the phenotypic animal data, had the lowest predictive performance on the test set with a RMSE of 2.94 kg. Predicting Grilling yield directly from digital images used to train a DL model reduced RMSE to 2.84 kg for both '2D' and '3D' images. Including carcass measurements, taken from the digital image, marginally reduced the RMSE further to 2.78 kg giving the highest predictive performance for the Grilling yield overall. Removing the conformation and fat score predicted previously from the digital image by a different process increased the RMSE fractionally to 2.81 kg.

The performance of the models trained to predict Roasting yield were also very similar across approaches with a 0.16 kg RMSE differential between the best performing and worst performing models (Table 1). Approach 1, utilising the phenotypic animal data to predict the Roasting yield resulted in an RMSE of 3.23 kg on the test set which in turn had a mean yield of 78.2 kg. The DL models used in Approach 2 had fractionally worse performance with RMSE values of 3.26 kg for both '2D' and '3D' images respectively. Models produced from Approach 3 had the best predictive performance at an RMSE of 3.19 kg without including conformation and fat scores, and 3.07 kg when including these scores. Overall, from a production

Table 1. Model performance for prediction of Grilling and Roasting yields on hold-out test data. RMSE and MAE are in kilograms.

			Grilling			Roasting		
Approach	**Data**	**Algorithm**	**RMSE**	**MAE**	**R^2**	**RMSE**	**MAE**	**R^2**
1	Phenotypic	Gradient Boosting	2.94	2.10	0.70	3.23	2.22	0.93
2	2D Images	Deep Learning	2.84	2.07	0.72	3.26	2.22	0.93
2	3D Images	Deep Learning	2.84	2.05	0.72	3.26	2.17	0.93
3	Phenotypic & Carcass measurements[1]	Gradient Boosting	2.81	2.00	0.72	3.19	2.13	0.93
3	Phenotypic & Carcass measurements[2]	Gradient Boosting	2.78	1.98	0.73	3.07	2.02	0.94

[1] No fat and conformation score.

[2] Including fat and conformation score.

viewpoint there are small differences in performance between models. The significance of this is discussed further below.

Discussion

One of the key results of this study is that there is little difference in the performance of the DL and ML models. From a practical point of view, once the DL models are trained, they are easier to deploy than the ML models. This is because the DL models require the carcass images only, captured in the abattoir within minutes of slaughter. In contrast the ML models require significant more data collection. In Approach 3, carcass measurements were combined with the associated phenotypic data. Matching carcass images to animal phenotypic data adds a layer of complexity and likely removes the possibility of an immediate prediction as nearly always the associated phenotypic data is stored in databases not instantaneously available. Consequently, it may be preferred at a processing level to obtain an immediate cut yield prediction with a slightly reduced predictive performance. In this scenario, our DL models predicting grouped carcass cuts from digital images could offer a mechanism to achieve just that.

A marked improvement in all approaches for predicting Roasting yields compared to predicting Grilling yields was observed in this study. Pabiou *et al.* (2011) also found variation in predictive performance for different areas of the carcass. The EUROP conformation classification places large emphasis on the hindquarters of a carcass. The Roasting cuts are exclusively in this area of the carcass, and it can therefore be postulated that predicted conformation score is more highly correlated to Roasting yield than Grilling yield. However, DL models also had much improved predictive capability for Roasting yield over Grilling yield when trained on the images alone without any predicted conformation score included in the model. The Roasting yield has a larger mean weight than the Grilling yield and makes up a larger proportion of the total carcass. Therefore, many more pixels from the digital image contain direct information on the Roasting yield than the Grilling yield. Conversely, the Grilling cuts have a smaller proportion of image pixels conveying direct information on these cuts. Furthermore, the fillet cut which is one of three Grilling cuts, is located beneath the rump and striploin cuts and therefore is not directly represented in any image pixel. Many successful applications of DL applied to images match the performance of a trained human (Miotto *et al.*, 2018). Fundamentally, these DL applications do not observe visual information beyond the capability of the human eye. Given that little information on Grilling yield is directly available from the image it is our opinion that the DL models presented here are approaching the predictive ability of a trained human grader.

In this study we have used 43,628 and 55,362 training records (in addition to 5,466 and 6,935 validation records) for Grilling and Roasting cuts respectively. Compared to the training set sizes observed in many successful DL models such as those trained on the ImageNet database with ~20 million images (Deng *et al.*

2009), the size here is limited especially considering the similarity between carcass images. We have applied transfer learning to overcome this issue of a smaller dataset, however, a much larger dataset would be an optimum solution. Nonetheless, given the impracticality of obtaining a carcass image training dataset of millions of records with associated meat yields, a more suitable option may be to collect more information from the digital images. This is commonly applied with the current VIA processes for EUROP score prediction and we have also investigated this approach by including carcass measurements in ML models. Including these measurements, as carried out in Approach 3, gave the best overall performance, albeit by a very small margin. However, predicted conformation and fat scores are still necessary features to include in the model to maximise performance. The fact that a reduction in performance does occur, indicates that the current VIA processes effectively capture available information from the carcass images, and therefore demonstrates the potential for the current VIA processes and models to be re-trained on meat yield rather than EUROP classification.

We conclude that DL models applied to meat yield data hold potential and may be more practical to deploy at an abattoir level, but our results suggest that the training datasets need to be significantly larger. In the scenario of more limited datasets, the standard approach of VIA and further image processing to capture important carcass dimensions and contours which flow into predictive ML models, has better predictive ability for carcass meat yields.

References

Connolly, S. M., Cromie, A. R., Sleator, R. D., and Berry, D. P. (2019). Translational Animal Science, 3(1), 1-13. https://doi.org/10.1093/tas/txy129

Conroy, S. B., Drennan, M. J., Kenny, D. A., and McGee, M. (2009). Animal, 3(11), 1613-1624. https://doi.org/10.1017/S1751731109990395

Conroy, S. B., Drennan, M. J., McGee, M., Keane, M. G., Kenny, D. A., *et al.* (2010). Animal, 4(2), 234-241. https://doi.org/10.1017/S1751731109991121

Deng, J., Dong, W., Socher, R., Li, L.J., Li, K. *et al.* (2009). Proc. of the 2009 IEEE conference on computer vision and pattern recognition (pp. 248-255).

Drennan, M. J., McGee, M., and Keane, M. G. (2008). Animal, 2(5), 752-760. https://doi.org/10.1017/S1751731108001754

Miotto, R., Wang, F., Wang, S., Jiang, X., and Dudley, J. T. (2018). Briefings in bioinformatics, 19(6), 1236-1246. https://doi.org/10.1093/bib/bbx044

Pabiou, T., Fikse, W. F., Cromie, A. R., Keane, M. G., Näsholm, A., *et al.* (2011). Livestock Science, 137(1-3), 130-140. https://doi.org/10.1016/j.livsci.2010.10.012

130. Digital phenotyping to improve Atlantic salmon feed efficiency and welfare

G.F. Difford[1,2*], S.K. Lindberg[1], K. Heia[1], B. Hatlen[1], E. Durland[1], B. Gjerde[1], R. Alvestad[1], V. Mota[1], C. Noble[1], A. Norris[3], S. Boison[3], N.K. Afseth[1], J.P. Wold[1] and A.K. Sonesson[1]

[1]Norwegian Institute of Food, Fisheries and Aquaculture (Nofima), P.O. Box 210, 1432 Ås, Norway; [2]Norwegian University of Life Sciences, Norway; [3]MOWI AS, Sandviksboder 77AB, Bergen, Norway; gareth.difford@nofima.no

Abstract

Recent developments in rapid digital phenotyping methods have resulted in new phenotypes in finfish aquaculture like feed intake, growth, and body fat towards improving the feed efficiency complex. Similarly, digital operational welfare indicators have been newly developed for Atlantic salmon including: the number of lice, the number and severity of wounds, condition of fins and smoltification status. The methods employed across these diverse phenotypes vary in approach but show some similarities. In general, they focus on being high throughput and non-invasive, reducing the fish time out of water by using semi-automated image or vibrational spectroscopy. All phenotypes tested were significantly heritable (0.18-0.61) and correlated to their respective gold standards (0.70-0.98). The added benefit of being either higher throughput, cheaper or less invasive than their gold standard counter parts; means digital phenotyping methods will likely continue to develop to meet the needs of future genetic selection programmes.

Introduction

The Atlantic salmon industry faces challenges to maintain good health and animal welfare, reduce environmental impacts, whilst remaining profitable. The sheer scale of aquaculture operations, as well as the operational and biological limitations of recording fish out of water, has spurred new research into rapid and non-invasive digital phenotyping methods. Genetic selection for improved feed efficiency is one approach that can lead to reductions in feed costs and environmental impact, as more than 50% of operational costs and 70% of the carbon foot print are related to feed (Iversen *et al.*, 2020; Winther *et al.*, 2020). Ensuring individual animal welfare is a moral and legal imperative, however genetic research into animal welfare is still in its infancy due to difficulties obtaining phenotypes.

Both feed efficiency and animal welfare traits are complex phenotypes comprised of numerous subtraits. Feed efficiency requires recording of growth and lipid deposition in different body tissues and, most importantly, feed intake. A standardised set of traits entitled 'operational welfare indicators' was recently established (Noble *et al.*, 2018) including smoltification status (determining readiness for seawater), number of ectoparasites, presence and severity of skin wounds and fin condition. Trained observers score these operational welfare indicators, which entails a rapid assessment of as many as 20 different traits. A conundrum while recording this data is that individual welfare and feed intake should not be compromised by the act of measuring traits themselves. The longer a fish remains out of water while being handled, the higher the chance of harming its welfare and stress induced loss of appetite. We present a number of potential rapid, non-invasive, digital phenotypes in development to address these challenges.

Materials & methods

Different digital phenotyping methods used in the feed efficiency complex or operational welfare indicator suites are described below with an emphasis on the basic principles, number of salmon recorded and their life stages.

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_130, © G.F. Difford *et al.* 2022

The feed efficiency complex. The feed efficiency complex is a group of traits including feed intake, growth and lipid deposition, all of which are now being digitalised to increase the frequency and quality of recordings. Growth of Atlantic salmon can now be continuously and non-invasively measured underwater at sea using diode frames. As fish swim through the frame, they obstruct the infrared light providing the length and depth record of the fish. Weight is then estimated using allometric predictions. Our innovation combines the underwater passive integrated transponder (PIT) readers into the diode frame thereby linking the radio frequency identification of the individual salmon with their growth phenotypes (Difford *et al.*, 2020). To date, more than 5,000 Atlantic salmon have been continuously phenotyped until 2 kg.

The addition of radio opaque markers to pelleted feeds makes it possible to determine the mass of feed consumed by an individual fish using a calibration equation for the relationship between marker numbers and feed mass. Individual fish are then X-rayed and the number of markers (which can be as high as 30-500) must be counted. Since manual counting is time consuming and tedious, we developed semi-automated image analysis together with automated deep learning tools to count the number of markers. Hitherto, more than 1,200 Atlantic salmon have been recorded for feed intake, ranging from 50-1000 gram.

The third component of feed efficiency is body lipid content in whole fish or in the muscle and other tissues at slaughter, resulting in a wide variety of spectroscopic methods. Currently, two non-invasive methods are in use for whole fish, a handheld dielectric spectroscopic method (Distel Fat Meter™) and an online rapid and non-invasive NIR sensor (SmartSensor). Both methods have been recorded simultaneously on 680 post-smolt Atlantic salmon. Three different methods using Raman and NIR sensors on salmon fillets or ground meat have been simultaneously tested on more than 500 Atlantic salmon broodstock (Difford *et al.*, 2021).

Operational welfare indicators. To reduce excessive handling and time during welfare scoring, we developed an imaging system with rapid delivery of fish via conveyor belts. This system makes it possible to image both sides of an anesthetised fish in under 30 seconds. By handling the fish with care and monitoring their recovery post-anaesthesia, none of the measured fish showed impaired welfare. The images undergo image analysis and deep learning methods to determine numerous traits including fish smoltification status, fin morphology, presence of wounds, as well as salmon lice numbers. For genetic evaluations, standardised *in vivo* lice challenge models are used, where large numbers of fish (2,000-3,000) must be counted for lice within a 2-3 day period with between 0-180 lice per fish, before the chalimus lice (<1 mm in length) become motile. To this end, we have digitally recorded more than 1,200 genotyped smolt infected with over 18,000 salmon lice. We have also successfully scored 450 Atlantic salmon during their transformation from fresh- to saltwater, where observable changes in their morphology, including skin and fin colorations were captured. We have successfully recorded more than 2,500 Atlantic salmon parr and post smolts for presence and severity of skin lesions, scale loss, damage to pectoral, dorsal, pelvic and caudal fins for indicators to welfare.

Results & discussion

We present the digital phenotypes, methods, number of recorded salmon, heritability and validation results for each method in Table 1. A variety of digital methods are employed each with different forms of multivariate data captured either as images, vibrational spectra, or sensors in time series. To unlock the phenotyping potential of these diverse data streams, we applied a wide variety of skills in image analysis, machine learning and deep learning. As the data types continue to diversify and the volumes increase, a large proportion of phenomics will be devoted towards data processing, compression and handling.

Table 1. Digital phenotypes, methods, number of fish, heritability estimates and validation in Atlantic salmon.

Phenotypes	Methods	N[1]	h^2	Validation[3]
Feed efficiency complex				
Feed intake	Xray	1,200	0.22	rp=0.85
Length	Infrared Diode frames	5,162	0.22	rg=0.90
Weight	Infrared Diode frames	5,161	0.18	rg=0.84
Lipid (fillet)	NIR spectroscopy	523	0.60	rg=0.93
Lipid (fillet)	Raman spectroscopy	523	0.52	rg=0.98
Lipid (whole)	Dielectric spectroscopy (Distell meter™)	680	0.45	rp=0.87
Lipid (whole)	NIR spectroscopy (SmartSensor)	680	0.61	rp=0.80
Operational welfare indicators				
Skin & fin	Spectral imaging	2,500		rp=0.94
Smoltification	Spectral imaging	450		rp=0.90
Lice count	Spectral imaging	1,200		rp=0.70

[1] N = number of fish.

[2] h^2 = heritability.

[3] Validation using either phenotypic correlations (rp) or genetic correlations (rg).

All new digital phenotyping methods must be compared to the appropriate gold standard methods in order to benchmark the level of their agreement. In a few instances it is possible to have enough phenotypes with both methods to estimate genetic correlations (Table 1). However, in most instances the gold standard is too expensive or invasive and only a small number of fish can be recorded. In such cases, we estimated the phenotypic correlation. In all cases the correlations were above 0.7 and mostly over 0.84. Assessing the costs of recording, the heritability, the gain in throughput over the gold standard method and proportion of variation in the gold standard captured, all contribute to assessing the merit of including digital phenotypes in the breeding goal.

Importantly, this work shows that digital methods can be adapted to meet the needs of fish biology. This is the first instance where individual growth of Atlantic salmon was measured underwater at sea and found to be heritable. The ability to record salmon in their commercial environments opens up the possibility of improving growth, yield, and, ultimately, feed efficiency. Furthermore, the shift from relying on human-scored operational welfare traits to rapid, image-based methods has resulted in larger sample sizes in far shorter time (Figure 1). This not only reduces the welfare burden on the fish being measured but also

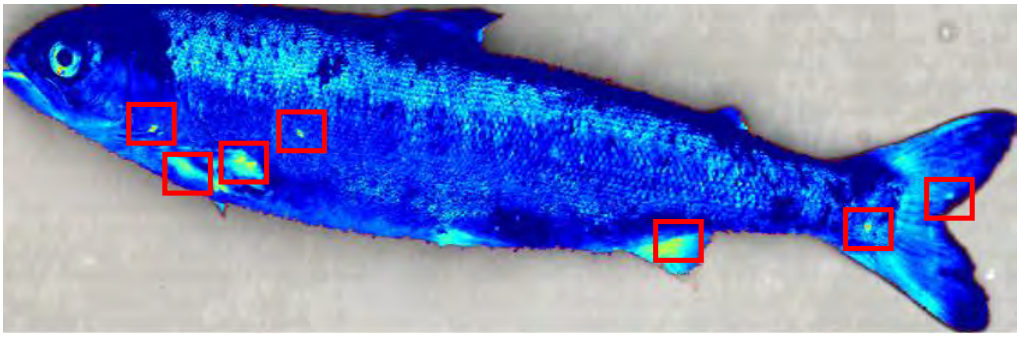


Figure 1. Enhanced image of Atlantic salmon with copepod salmon lice attached (red).

the financial burden of requiring human resources, a costly exercise in Norway. Currently, no studies have reported heritability estimates for welfare traits in Atlantic salmon, and more research is needed to address this.

Method development is not static; rather, improvements are continuous as new technology and digital phenotypes are created, which must then be rigorously validated. For example, significant genomic heritability of feed intake in Atlantic salmon is reported for the first time using digital Xray radiography (Figure 2). Whilst these estimates are from land-based research facilities, the latest developments in portable Xray technology make cage-side imaging of larger fish feasible. Atlantic salmon breeding programmes of the future will likely contain many digital phenotypes in their breeding goals.

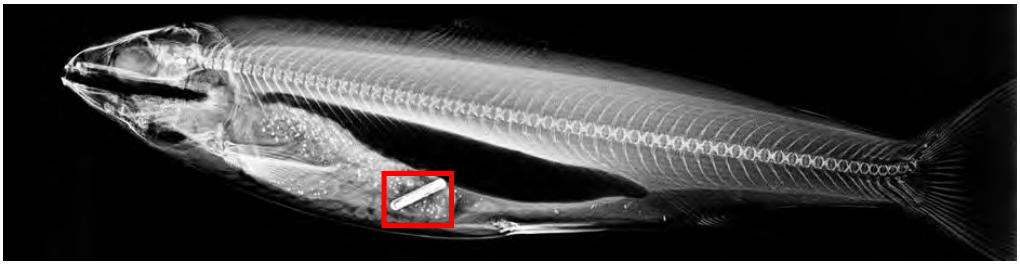


Figure 2. Digital radiograph of Atlantic salmon with PIT tag (red) and opaque bead markers in digestive tract.

Acknowledgements

The authors acknowledge project funding from the following EU Horizon 2020 projects grant projects 'AquaIMPACT' and 'NewTechAqua' under grant no. 818367. As well as the Research council of Norway projects 'SmoltFieldGen' 267650 and 'Genomics of Omega 3' 244200 and internal Nofima projects PrecisionVision, Vision2Precision and MOWI Ramme.

References

Difford, G.F., Boison, S.A., Khaw, H.L., Gjerde, B., 2020. Comput. Electron. Agric. 173, 105411. https://doi.org/10.1016/j.compag.2020.105411

Difford, G.F., Horn, S.S., Dankel, K.R., Ruyter, B., Dagnachew, *et al.*, 2021.. Genet. Sel. Evol. 53. https://doi.org/10.1186/s12711-021-00605-6

Iversen, A., Asche, F., Hermansen, Ø., Nystøyl, R., 2020. Aquaculture 522, 735089. https://doi.org/10.1016/j.aquaculture.2020.735089

Noble, C., Gismervik, K., Iversen, M.H., Kolarevic, J., Nilsson, *et al.*, 2018..

Winther, U., Hognes, E.S., Jafarzadeh, S., Ziegler, F., 2020. Greenhouse gas emissions of Norwegian seafood products in 2017.

131. Prediction of roughage intake of dairy cows combining milk mid-infrared spectra and cow variables by deep learning

G.T. Gebregiwergis[1*], N.K. Afseth[2], A. Kidane[1], E. Prestløkken[1] and T.H.E Meuwissen[1]

[1]Department of Animal and Aquacultural Sciences, Faculty of Biosciences, Norwegian University of Life Sciences, P.O. Box 5003, 1430, Ås, Norway; [2]Nofima AS, Norwegian Institute of Food, Fisheries and Aquaculture Research, Ås, Norway; gebreyohans.tesfaye.gebregiwergis@nmbu.no

Abstract

Combining milk Mid-infrared (MIR) spectra with other cow descriptive and performance variables improves the prediction accuracy of total dry matter intake (TDMI) in dairy cows above using the milk MIR spectra only. However, the improvement in accuracy of prediction depends also on data fusion and prediction methods. Our objective was to evaluate low- and high-level data fusion strategies with partial least square (PLS) and artificial neural networks (ANN). Prediction equations were developed using a total of 384 observations and 4 different models. Predictions were performed using ANN, PLS and multiple linear regression (MLR). MLR was implemented using the variables A (milk yield, concentrate dry matter intake, week of lactation, parity and metabolic body weight). ANN was implemented using variables A, using only milk MIR spectra data or combining milk MIR spectra and variables A. PLS was implemented using variables A or using the milk MIR spectra and variables A. The high-level data fusion method using ANN had the highest out of sample coefficient of determination (R^2) of 0.62. Direct merging of milk MIR spectra data with variables A explained better the TDMI than the milk MIR spectra data using both ANN and PLS prediction methods. Generally, the ANN methods perform better than PLS at low as well as at high level data fusion. Merging both milk MIR spectra data and animal variables at high level fusion enables to exploit better the information in the data. This method has thus a great potential to predict TDMI in commercial dairy farms. However, our finding needs additional evidence using large dataset and external validation.

Introduction

Milk MIR spectra are useful for predicting TDMI in dairy cows (Dórea *et al.* 2018; Lahart *et al.* 2019; Grelet *et al.* 2020). Based on most studies, combining milk MIR spectra data with other cow descriptive and performance variables, such as milk yield, lactation stage, parity, body weight, and feeding behavior, gave better predictions of TDMI than predictions based on milk MIR spectra data alone (Dórea *et al.* 2018; Wallén *et al.* 2018; Lahart *et al.* 2019). Moreover, combining individual cow concentrate intake data with milk MIR spectra improved prediction accuracy of DMI (Wallén *et al.* 2018; Rachah *et al.* 2020). However, the improvement in accuracy of prediction of TDMI depends also on the prediction method. Dórea *et al.* (2018) reported a better performance using artificial neural network (ANN) than partial least squares regression (PLS) to predict TDMI using milk MIR spectra and animal variables in dairy cows. Generally, direct combination of descriptive and performance variables into milk MIR spectral data improves TDMI prediction accuracy, but how well the method would be able to fully exploit this data in both ANN and PLS prediction methods is not known.

To fully utilize the benefits of integrating the milk MIR spectra data and other descriptive and performance variables of cows, different data fusion methods were used in this study in both prediction methods (PLS and ANN). A data fusion method can be used to combine the outputs and complementary data from multiple instruments to generate better inferences than a single method (Biancolillo *et al.* 2014). There are three levels of data fusion: low level, mid-level, and high level (Biancolillo *et al.* 2014; Li *et al.* 2019), and we will investigate the low- and high-level here.

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_131, © G.T. Gebregiwergis *et al.* 2022

We hypothesized that the high-level data fusion with advanced analytical methods, such as with machine learning gives better prediction accuracy of TDMI than the low-level data fusion with PLS prediction. Therefore, the objective was to evaluate the low (direct merging of milk MIR spectra data with other descriptive and performance variables)- and high-level data fusion strategies. Both PLS and ANN prediction methods were assessed using the low and high data fusion strategies to predict TDMI in dairy cows.

Materials & methods

Data. Data were collected from a feeding experiment of dairy cows at the dairy research farm of the Norwegian University of Life Sciences (Ås, Norway). The experiment lasted for 54 days and used 48 early- and mid-lactation Norwegian red dairy cows. A detailed description of the experiment objective, design, and results are available at Kidane *et al.* (2018). In the experiment period, concentrate dry matter intake (CDMI), Total dry matter intake (TDMI), and milk yield (MY) data were available daily per cow. The daily measurements of MY, TDMI and CDMI of a specific date was merged into weekly averages for each experimental week.

Milk MIR spectra. Each cow was milked twice daily (between 06:15 to 08:15 h=a.m. and 05:00 to 17:00 h=p.m.) using milking machines. During the first two weeks of the experiment, only Monday milk samples were collected. During weeks three to eight, milk samples were collected Monday and Thursday. Morning and evening samples were kept separately and analyzed using a FOSS instrument (MilkoScan 6000; Foss Analytical, Hilleroed, Denmark).

In the Foss MIR spectrum, there are 1,060 data points corresponding to the absorption of infrared light in the 900 to 5,000 cm^{-1} region. In order to predict TDMI, wavelength regions recommended by Wallén *et al.* (2018) were used: between 926 and 1,593, between 1,745 and 3,061, and between 3,781 and 5,149 cm^{-1}.

For each MIR wavelength, the morning and evening measurement averages for week 1 and week 2 were calculated but from weeks 3 to 8, Mondays and Thursdays morning and evening measurements were averaged. In the end, there were 384 records available, that is, one MIR spectrum for each cow each week.

TDMI model development and validation. Prediction equations were developed using 10-fold cross-validation. Three alternative models were considered to predict TDMI of dairy cows. These were, partial least squares regression (PLS), artificial neural network (ANN) or MLR (Multiple linear regression). The models were as follows: Model 1 included MY, CDMI, WL, Parity and $BW^{0.75}$; model 2 included milk MIR spectra data; model 3 included all variables in model 1 and 2. Model 1 was implemented using both MLR and ANN as statistical approaches to predict TDMI. Model 2 and 3 were implemented with both PLS and ANN.

The cross-validation coefficient of determination (R^2) between true and predicted TDMI values and the root mean squared error (RMSE) were used to assess the accuracy and robustness of the models.

Models fitted using ANN were performed using H2O's AutoML (automatic machine learning) in R (LeDell and Poirier 2020). All parameters were left at their default values, except for max_runtime_secs=36,000, nfolds=10 and include_algos=deeplearning. The training data comprised 80% (288) and the test data 20% (96) of the data set. As in the PLS and MLR approaches, R^2 and RMSE were used to assess the accuracy and robustness of the models in the 10-fold cross-validation, full training data set and test data set.

High-level data fusion. In high-level data fusion, the results of model 1 and model 2 were integrated to obtain an ensemble decision and binary linear regression was used to solve the advanced data fusion problem as described by Li *et al.* (2019).

Results

Our results supported our hypothesis that high-level data fusion using ANN had the highest coefficient of determination (R^2=0,62) which means 62% of the variation in the total dry matter intake (TDMI) was explained using the model (Table 1).

Direct merging (low-level data fusion or model 3) of milk MIR spectra with other variables such as MY, CDMI, WL, Parity and $BW^{0.75}$ explained TDMI better than the milk MIR spectra using PLS prediction (Table 2). Multiple linear regression (MLR) using other variables (MY, CDMI, WL, Parity and $BW^{0.75}$) (model 1) explained TDMI better than the milk MIR spectra (model 1) (R^2=0.39 vs R^2= 0.38 in the calibration and R^2=0.37 vs R^2=0.33 in the validation) (Table 2). In case of ANN prediction methods, the low-level data fusion (model 3) had the highest R^2 next to the high-level data fusion method (Table 1).

Discussion

Our results showed that the ANN method is better for prediction of TDMI than the PLS method. This implies that the relationship between the milk MIR spectra and TDMI is not only a linear relationship since the former method capture both the linear and non-linear relationship and other complex relationships (Dórea *et al.* 2018). Our results also showed that adding animal variables to the milk MIR spectra improves

Table 1. Root mean square error (RMSE) and coefficient of determination (R^2) for low -and high-level data fusion obtained from predicting total dry matter intake using milk MIR spectra data and other variables (MY, CDMI, parity and $BW^{0.75}$) and artificial neural networks (ANN).

Model	Predictor	10- fold cross- validation on training data		Full training data set		Test data set
		R^2	RMSE	R^2	RMSE	R^2
1	[1]A	0.52	0.14	0.53	0.14	0.40
2	[2]MIR	0.46	0.15	0.48	0.15	0.53
3	MIR+A	0.56		0.58		0.605
[3]H	-	-	-	0.69	0.11	0.62

[1] MY, CDMI, WL, Parity and $BW^{0.75}$.
[2] Milk MIR spectra data.
[3] High level data fusion.

Table 2. Summary of partial least squares (PLS) regression, multiple linear regression (MLR) and high-level data fusion prediction models for total dry matter intake in dairy cows using milk MIR spectra data, MY, CDMI, parity and $BW^{0.75}$.

Trait	Models	Predictors	Factor	Calibration		Cross validation	
				R^2	RMSEP	R^2CV	RMSECV
TDMI	1	[1]A	-	0.39	2.43	0.37	2.44
	2	[2]MIR	7	0.38	2.43	0.33	2.53
	3	MIR+A	3	0.38	2.42	0.37	2.44
	[3]H	-		0.42	2.35	0.38	2.41

[1] MY, CDMI, WL, Parity and $BW^{0.75}$.
[2] Milk MIR spectra data, [3]high level data fusion.

the prediction accuracy of TDMI especially using ANN. However, it matters how we combine the milk MIR spectra with other cow variables in order to fully exploit the information in the data and improve accuracy of prediction. Instead of directly adding the animal variables to the milk MIR spectra as addition explanatory variables and use it for the prediction of TDMI, merging both milk MIR spectra and animal variables at high-level enables to exploit better the information in the data. Moreover, the animal variables are easy to record and may be already in the dairy farm management system. Thus, merging milk MIR spectra data with animal variables at high-level have great potential to predict TDMI in commercial dairy farm industries and this tool can be practically implemented without additional infrastructure and without extra costs. However, this finding should be verified using a large dataset and external validation.

Acknowledgements

To the Norwegian Research Council (project no. 268124 /E500) for financial support.

References

Biancolillo A., Bucci R., Magrì A.L., Magrì A.D. and Marini F. (2014) Analytica chimica acta 820(23-31.

Dórea J., Rosa G., Weld K. and Armentano L. (2018) Journal of dairy science 101(7)**:** 5878-5889.

Grelet C., Froidmont E., Foldager L., Salavati M., Hostens M. *et al.* (2020) Journal of dairy science 103(5)**:** 4435-4445.

Kidane A., Øverland M., Mydland L.T. and Prestløkken E. (2018) Livestock Science 214(42-50.

Lahart B., Mcparland S., Kennedy E., Boland T., Condon T. *et al.* (2019) Journal of dairy science 102(10)**:** 8907-8918.

Ledell E., and Poirier S. (2020) H2o automl: Scalable automatic machine learning, pp. in *Proceedings of the AutoML Workshop at ICML*.

Li Y., Xiong Y. and Min S. (2019) Vibrational Spectroscopy 101(20-27.

Rachah A., Reksen O., Afseth N.K., Tafintseva V., Ferneborg S. *et al.* (2020) Journal of Dairy Research 87(4)**:** 436-443.

Wallén S., Prestløkken E., Meuwissen T., Mcparland S. and Berry D. (2018) Journal of dairy science 101(7)**:** 6232-6243.

132. CFIT – Cattle Feed InTake – a 3D camera based system to measure individual feed intake and predict body weight in commercial farms

J. Lassen[1,2*], J.R. Thomasen[1] and S. Borchersen[1]

[1]Viking Genetics, Ebeltoftvej 16, Assentoft, 8960 Randers, Denmark; [2]Center for Quantitative Genetics and Genomics, Aarhus University, Blichers Alle 20, 8830 Tjele, Denmark; jalas@vikinggenetics.com

Abstract

A 3D camera technology to measure feed intake and body weight on individual cows in commercial farms has been developed. The camera based method for feed intake measurements is validated using scale measures in a commercial farm. In this study the squared correlation between camera data and scale measures on daily intake was 0.90. Body weight is predicted based on images of back couvertures and corresponding scale measures in three dairy breeds: Jersey (301 cows), Holstein (398 cows) and Red dairy cows (103 cows). Root mean square error between observed and predicted weight was 23.5, 31.0 and 22.7 kg and the repeatabilites were 0.86, 0.83 and 0.90 for Jersey, Red dairy cattle and Holstein respectively. Several management tools continue to be developed based on the data for the farmers that have the equipment installed and more traits are expected to be developed based on the 3D methodology in the future.

Introduction

Feed intake (FI) is the highest variable cost in the dairy farm. Therefore even marginal improvements in the amount of feed that is used to produce an amount of milk and meat will have a huge effect on the turnover for a farmer. Milk and meat production has been core traits in most breeding systems for several decades, but not until recently feed efficiency has been a trait that has been evaluated. This is not due to lack of value of the trait, but more due to lack of data on the trait at an individual level. Individual FI measures has only been available on cows from research farms, and often only for limited parts of lactation and in few lactations per cow. FI is a main driver for cow health and welfare. Most diseases occur in early lactation so if data is not available in this period it is very difficult to provide information that possibly could prevent diseases from developing. The system is currently installed in 20 dairy herds in Denmark where more than 10,000 cows are providing data.

FI measures historically has been done using scale based measures. These systems are relatively expensive and labour intensive. Therefore they are also impractical for measurements in commercial herds. Over the recent years both 3D cameras, hardware such as graphic cards and algorithms around image analysis has been under drastic improvements. This includes among others artificial intelligence (AI) algorithms such as the convoluted neural network method called MASK-CNN (He *et al.*, 2018). Combining all these developments into a working system is very difficult.

Since 2016 a 3D camera system has been developed to measure individual FI records and make body weight (BW) prediction at an individual cow level. In this paper the methodology as well as the system will be described and discussed.

Materials & methods

System and data description. In order to make individual FI and BW measures 3D cameras was used in a top down position (Microsoft Kinect). Each camera is placed 2.5 meter apart and 4.5 meter from the empty feeding table in order to cover the entire feeding table. Also, in the exit corridor of the milking system an ear tag reader and a 3D camera was placed. Three type of images were recorded from the 3D camera:

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_132, © J. Lassen *et al.* 2022

normal pictures, infrared (IR) pictures and depth pictures. These images were indicating the distance from the camera to the object that can be seen by the camera. When cows left the milking system their electronic ear tag was read and at the same time a 3D picture was taken of the back of the cow. For BW prediction as well as a reference to predict the same cow these pictures were stored and used. The identification was done based on the contours, colour and patterns of the back of the cow, when eating at the feeding table. The MASK-CNN algorithm based on instance segmentation prediction was used for this (He *et al.*, 2018) and this approach was validated in an earlier study (Thomasen *et al.*, 2018; WO 2017/001538). When a cow entered the head to the feeding table the last image of the feeding pile before she put in her head and began eating was stored from virtual boxes defined in the feeding table. When the cow had finalized the meal and took the head out again the first new image of the feeding pile was stored. In the next step the height in each pixel from two stored images were now subtracted from each other and the removed feed was quantified on pixel level. Five variables were stored from each specific visit: the ID of the cow, the placement in the barn, time when the meal was initiated and finalized and the amount of feed eaten. During a visit feed would be allocated to a specific cow from a total of 5 virtual boxes – the virtual box right below the cow as well as the two virtual boxes to the left and the right respectively. If a virtual box was shared during a feeding visit cows would also share the feed taken from this box during the two cow specific visits.

Feed intake trial. FI was measured both with cameras and on a scale in an area of the herd where only one cow could attend the feed. Daily intake was estimated with the camera based FI algorithm as well as with the scale data as the difference between provided feed and residual feed for a day. Six different cows was used in the trial and each cow was measured for 4 days. Correlations between these measures where estimated.

Body weight prediction. Weight data of the cows has been recorded in 6 weeks periods in 4 herds: 2 Jersey (JER), 1 Red Dairy Cow (RDC) and 1 Holstein (HOL) after every milking. Using the partial least square approach (PLS) where the 12 variables that described the most of the variance where chosen to predict BW (SAS software version 9.4). In a 10-fold crossvalidation the root mean square error (RSME) was estimated. Variance components for BW was analysed using the following model:

Weight = mean + herd + week of year + lactation number + week of lactation + animal + residual

Where herd, week of year, lactation number and week of lactation was fixed effects and animal and residual was random effects. The repeatability was estimated as the random animal variance divided by the total variance.

Results

Mean daily FI based on scales in the feeding trial was 55.4 kg of feed and the squared correlation between observed FI from the scales and the predicted FI from the cameras was 0.90. Descriptive statistics from scales as well as RMSE, variance components and repeatabilities for camera based BW in JER, HOL and RDC cows are presented in Table 1.

The mean BW was 678 kg for Holstein, 508 kg for Jersey and 624 kg for RDC. The RMSE were 23.5 kg (4.6% of mean), 31.0 kg (5.38% of mean) and 22.7 kg (3.60% of mean) for JER, HOL and RDC respectively. Repeatabilities were 0.86, 0.83 and 0.90 for JER, HOL and RDC respectively.

Table 1. Breed, mean observed weight, RMSE in kg and in % of mean, animal variance, residual variance, repeatabilities and number of animals used in the study.

Breed	Mean obs weight	STD	RMSE kg/%	Var animal	Var Res	Repeatability	Number of animals
JER	508	48.8	23.5/4.60	950	153	0.86	300
HOL	678	84.2	31/5,38	2,527	512	0.83	398
RDC	624	59.6	22.7/3.60	2,216	233	0.90	101

Discussion

The validation trial of the FI algorithm showed a squared correlation between FI measured from scales and cameras of 0.90. This indicates a strong relationship between the two methods. We are currently not aware of any other studies that have made such a comparison. BW based on 3D image analysis was predicted based on scale measures with RMSE between 22.7 and 31 kg. This is lower or in the same range as other studies (Song *et al.*, 2018). Several parameters can influence these results including camera and scales, information used from images, software and statistical model.

A major priority for developing this system was to obtain data throughout lactation. In relation to genetic analysis several studies have shown that genetic correlation change during lactation between FI, milk yield and BW (Manzanilla-Pech *et al.*, 2014; Li *et al.*, 2017). Estimates for the genetic correlation between FI and milk yield varies from -0.80 in early lactation up to 0.8 in mid and late lactation (Manzanilla-Pech *et al.*, 2014). Selection for improved efficiency in early lactation where FI and BW changes specifically needs to be improved based on records obtained in mid lactation then might lead to even lower FI in early lactation. From a management perspective also FI records and BW are interesting throughout lactation. In early lactation the absolute majority of health problems occurs both related to mastitis, BW changes, reproduction and nutrition. In mid lactation the farmer wants to know which cow are most efficient to optimize culling strategy and in late lactation the FI is interesting in order to optimize strategies in relation to drying off cows. With this system all information will be available for the farmer. Heritability estimates for FI and BW is presented in another abstract in this conference (Manzanilla-Pech *et al.*, 2022). Also the data is used for national prediction of saved feed which consist of a maintenance part and a physiological part described as residual feed intake (NAV 2020; Stephansen *et al.*, 2021). The identification algorithm was validated in an earlier study where more than 6,000 cow visits were annotated and in over 99% of the visits the right ID was predicted (Thomasen *et al.*, 2018; WO 2017/001538). This system opens up for working with phenotypes such as body condition score and lameness.

Conclusions

FI and BW can be predicted based on image analysis using a 3D camera system. The squared correlation between FI measured with scales and cameras was 0.90. Repeatability was between 0.86 and 0.90 for BW.

Acknowledgements

The research leading to these results has received funding from the Innovation Fund Denmark. Project Cattle Feed Intake – CFIT number 9090-00083B.

References

He, K., G. Gkioxari, P. Dollar and R. Girshick (2018) Mask R-CNN. arXiv:1703.06870v3

Lassen, J., J. R. Thomasen, R. H. Hansen, G. G. B. Nielsen, E. Olsen, P. R. B. Stentebjerg, *et al.* (2018). Proc. of the 11th WCGALP. Auckland, New Zeland.

Li, B., B.Berglund, W.F.Fikse, J.Lassen, M.H.Lidauer, P.Mäntysaari, P.Løvendahl. (2017). J. Dairy Sci. 100(11): 9076-9084. https://doi.org/10.3168/jds.2017-12775

Manzanilla-Pech, C.I.V., R.B. Stephansen, T. Andersen and J. Lassen (2022) Proc. of the 12th WCGALP. Rotterdam, the Netherlands.

Manzanilla-Pech, C.I.V., R. Veerkamp, Y. de Haas, M. Calus, and J. tenNapel. (2017). J. Dairy Sci. 100(11):9103-91014. https://doi.org/10.3168/jds.2017-12741

NAV. (2020). https://www.nordicebv.info/wp-content/uploads/2020/11/Metabolic-efficiency-included-in-the-Saved-feed-index.pdf

Song, X., E. A. M. Bokkers, P. P. J. van der Tol, P. W. G. Groot Koerkamp, *et al.* (2018) J. Dairy Sci. 101:4448–4459https://doi.org/10.3168/jds.2017-13094

Stephansen, R.V., M. H. Lidauer, U.S. Nielsen, J. Pösö, F. Fikse, *et al.* Proc. of the 2021 Interbull meeting.

Thomasen, J. R., J. Lassen, G. G. B. Nielsen, C. Borggard, P. R. B. Stentebjerg, *et al.* (2018) Proc. of the 11th WCGALP. Auckland, New Zealand.

WO 2017/001538 System and method for identification of individual animals based on images of the back.

WO 2014/166498 System for determining feed consumption of at least one animal.

133. Video-based analysis of dairy cow behaviour: detection of lying down and standing up

I. Adriaens*, W. Ouweltjes, I. Hulsegge and C. Kamphuis

Wageningen University & Research, P.O. Box 338, 6700 AH Wageningen, the Netherlands;
ines.adriaens@wur.nl

Abstract

Digital agriculture offers opportunities for improved monitoring and precision phenotyping of farm animals, crucial to achieving a more sustainable livestock production sector. Video-based analysis enables the quantification of animal behaviour in a non-invasive, automated way with few sensors. To unlock its full potential, appropriate computer vision techniques are needed. In this study, we propose an algorithm to detect lying-down and standing-up behaviour in dairy cows based on changes in bounding box properties detected via YOLOv5 and tracked with DeepSORT. We analysed 86 videos with a standing-up or lying-down event. With different criteria applied to the bounding box time series, we could detect up to respectively 92.3 and 80% for standing-up and lying-down events, respectively, with an accuracy of less than 2 seconds. Using bounding box properties as proxy for body shape and location, a general cow detection algorithm can serve multiple behavioural analyses simultaneously, whilst interpretability of the algorithms is maintained.

Introduction

Today's society faces a plethora of challenges to achieve a sustainable, socially accepted, safe and sufficient food production. Digital agriculture and related technological solutions contribute to addressing these challenges, as they allow to monitor and phenotype production entities in an automated and standardized way (Basso and Antle, 2020). In precision livestock farming, both individual animals and groups are monitored via automated sensor technologies, providing farmers, veterinarians, nutritionists and other stakeholders with information on health, fertility and welfare. Besides individual on-animal sensors, more and more also off-animal technologies that allow simultaneous and non-invasive monitoring of multiple animals emerge.

A main challenge in precision livestock farming is the translation of raw and sometimes unstructured sensor data into valuable animal or herd parameters that can be used for decision support. Often, this translation needs dedicated tools for data interpretation, including animal identification, feature extraction, time-series modelling and detection of abnormalities. The development of robust and targeted data-processing pipelines is therefore a crucial step in the digital agricultural transition (Basso and Antle, 2020).

Automated analysis of animal behavior provides new tools for welfare and health monitoring and (precision) phenotyping, as many links between health disturbances and behavior in livestock are known. Traditionally, behavior is observed visually, rendering it a subjective and tedious task with only limited value in commercial farm settings. Automation of behavior assessment allows to use behavioral traits in monitoring and decision support tools for the entire herd with limited observation costs, and is therefore seen as an important focus in current research and development (Rushen *et al.*, 2012).

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_133, © I. Adriaens *et al.* 2022

In dairy cows, lying behavior changes upon disease such as lameness, mastitis and metabolic problems, and can therefore be used to assess health and welfare status (Dittrich *et al.*, 2019). In the past, on-animal sensors (e.g. 3D accelerometers) have been used for detection of health problems, with varying success. Off-animal sensors such as video cameras have the advantage that a single sensor can be used for: (1) multiple animals simultaneously; and (2) many different behavioral traits, based on spatial as well as temporal features of the individual cows. However, to extract valuable information from the unstructured video footage, targeted data processing is crucial, for example via the use of cow detection, key point identification or pose estimation (Porto *et al.*, 2013).

In this study, we aimed at developing a tool for the detection of 'lying down' (LD) and getting up (GU) events from video footage collected in a freestall barn as leverage for lying behaviour analysis in dairy cows. We started from state-of-the-art cow detection and tracking models, and propose the use of bounding box (bb) properties to predict the LD and GU events.

Materials & methods

Data. Video footage of a freestall barn with synthetic flooring was collected at the research facility 'Dairy Campus' in 2019. Cows were fitted with a 3D accelerometer (IceQube, IceRobotics Inc. Edinburgh, Schotland), from which the timing of GU and LD events was derived. Based on the location of the animals, 1 out of 4 cameras were selected, and the footage surrounding the GU or LD event was isolated to obtain videos of approximately 460 frames (frame rate 15f/s). To collect ground truth information on timing of the GU and LD events, the cows' poses were manually observed and the exact start and end frame were determined and recorded together with the animals' coordinates when lying down. When the target cow was occluded by other cows, videos were removed from the analysis.

Cow detection and tracking. A cow detection model (YOLOv5x) combined with a tracking algorithm (DeepSORT) trained and validated on a separate dataset on the high performance computing infrastructure of WUR (the Netherlands) were applied on each video to obtain bounding box coordinates (x,y), their width (w) and their height (h) for each of the cows (Kamphuis *et al.*, 2022). All further data processing was done in Matlab R2020b (The Mathworks Inc., Natick, MA, USA), and corresponding code can be found on https://git.wur.nl/iadriaens/kb_ddht_ai. Bounding box data of the target cows was gathered using the Euclidian distance of the centre of each bb with the ground truth location of the target cow during lying down (i.e. using either the first (for GU events) or last (for LD events) 50 frames of each clip). When no cow detection had taken place, or when the target cow was not detected in more than 10% of the frames and could not be reidentified, the clip was removed from the analysis. Examples of the bb are shown in Figure 1.

Figure 1. Example of bounding box changes of cow during a getting-up (left panel) and lying-down (right panel) event.

Bounding box properties. The bb coordinates, width and height were smoothed with a third-order Savitsky-Golay filter with window size of 25 frames to correct for noise and instability caused by the YOLOv5 detection algorithm. Next, we calculated the Euclidian distance from the center of the frame (*cntr*), the circumference of the bb (*circ*), and the ratio between width and height (*rwh*) as follows:

$$cntr = \sqrt{(x_{bb} - x_c)^2 + (y_{bb} - y_c)^2} \quad (1)$$

$$circ = 2 * h_{bb} + 2 * w_{bb} \quad (2)$$

$$rwh = \frac{w_{bb}}{h_{bb}} \quad (3)$$

These bb properties (bbp) were standardized with a min-max standardisation as follows:

$$\text{bbp}_{st} = \frac{bbp - \min(bbp)}{\max(bbp) - \min(bbp)} \quad (4)$$

Finally, the ratios between each of the bbp_{st} (e.g. *centr*/*circ*, *circ*/*rwh*) to represent simultaneous change across bbp were computed.

Event detection and statistics. Unique statistical change points, and local and absolute minima and maxima of the bbp_{st}, of the ratios, and of their first and second derivatives were calculated as proxies for sudden change and maximal slopes in the bbp throughout frames. Next, the location of first and last local peaks within the frame time series was compared with the annotated location of the start and end of the GU/LD events and the frame and time distance was calculated. Appropriate bbp for the detection of GU and LD are those consistently close to the target event throughout all videos, independently of the location and orientation of the cow. The detection was considered sufficiently accurate when the detection was within 2 to 5 seconds (i.e. 30-75 frames) of the actual event.

Results

Eighty-four clips were retained for the analysis, from which 45 with a LD event and 39 with a GU event. The results are summarized in Table 1 with the best performing bbp printed in bolt. When both events were considered, the maximal difference in the *circ* and *cntr* showed the best detection accuracy, with an average distance of resp. 1.3 and 2 seconds from the event, and 82.1% of the videos having a distance of less than 2 seconds from the event. Considering only the GU behaviour, the best detection criteria were the maxima of the second derivatives of *cntr* and *circ*, with average distances of respectively 0.7 and 1.7 seconds and 92.3 to 74.4% of the events being detected within 2 seconds of the ground truth. For LD detection, additional to the maximal differences in *cntr* and *circ*, also the ratios between *cntr* and *circ* and the changepoints of these ratios in all three bbp performed well, with up to 80% of the LD events detected within 2 seconds and 86.7% within 5 seconds of the ground truth.

Discussion

This study investigated the possibility to use a cow detection and tracking algorithm to identify lying down and getting up events via changes in bbp time series over frames. The idea relied on the expectation that during lying down, body shape and location remain constant, while during the GU/LD event, they suddenly change. When the detection algorithm works well and draws its bb tightly and accurately around the animal, the defined bbp were expected to link with lying behaviour. In this study, we showed the potential of using bbp for detection of lying down and getting up events with video footage, and demonstrated that the criteria most suitable for detecting LD or GU separately might differ.

Table 1. Time distance and % events detected within 2 to 5 seconds with the different bounding box properties for all (GU+LD), get-up (GU) and lying-down (LD) events.

	Time distances			% within 2-5 seconds		
	GU+LD	**GU**	**LD**	**GU+LD**	**GU**	**LD**
max_dif_cntr[1]	**1.3±3.0**	1.4±3.5	**1.3±2.5**	**82.1-85.7**	84.6-87.2	**80.0-84.4**
max_dif_circ[2]	**2.0±3.7**	2.9±4.6	**1.3±2.5**	**71.4-77.4**	61.5-74.4	**80.0-80.0**
ratiopeak_cntr_circ[3]	4.3±5.2	7.0±6.4	**2.0±2.0**	45.2-65.5	30.8-43.6	**57.8-84.4**
cpts_rat[4]	3.2±4.6	4.6±5.4	**2.1±3.6**	64.3-73.8	48.7-59.0	**77.8-86.7**
der2_cntr[5]	3.4±3.3	**0.7±1.1**	5.8±2.6	46.4-59.5	**92.3-97.4**	6.7-26.7
der2_circ[6]	2.8±3.1	**1.7±2.2**	3.7±3.5	54.8-67.9	**74.4-82.1**	37.8-55.6

[1] maximal difference in centerdistance.

[2] maximal difference in circumference.

[3] peak in ratio between centerdistance and circumference.

[4] statistical change point in ratios across the 3 bbp.

[5] second derivative of center distance.

[6] second derivative of circumference.

In the past, accelerometer sensors have been used to automatically detect lying down or getting up (Ledgerwood *et al.*, 2010), and also in this study, this served as ground truth benchmark. As these sensors are individually attached to the cows, a separate sensor has to be purchased for each animal, which increases costs. Also maintenance (e.g. battery, broken sensors replacements) and the limited number of potential behaviours monitored are drawbacks for this type of technology that are overcome by using cameras. While accelerometers do not require a separate identification step needed for machine vision, these sensors have the disadvantage to not provide spatial information regarding the behaviours monitored.

The advantage of using bbp over e.g. key point based methods, or even over methods that specifically annotate videos for lying animals (Porto *et al.*, 2013) is that it allows a general cow-detection algorithm (for which only bb annotations are needed) to be used as a basis for multiple location and body-shape based behavioural analyses. This limits annotation time and contributes to interpretability and understandability of the methodology, which is an advantage when aiming for e.g. monitoring and decision support. The same bb and tracking could be used for e.g. analysis of feeding behaviour (based on location and orientation of the bb at the feeding rack) or social behaviour (by calculating proximity). The alternative of including the behaviour as input for the (deep learning) video analysis would require many more time consuming and tedious annotation work, with a lower number of potential applications.

Future work can focus on further automatizing the algorithm, testing it on more and longer videos and multiple cows simultaneously. Additionally, a regression model in which several of the bbp time series can be combined potentially further improves detection accuracy.

Video-derived behavioural analysis allows for automated monitoring of multiple animals and multiple traits simultaneously in a cost-effective way using few sensors. In the future, this can contribute to improved monitoring and precision phenotyping, and as such, ultimately leverage a more sustainable dairy sector.

Acknowledgements

This study was part of Knowledge Base Data Driven High Tech (KB-38-008-002), funded by the Dutch Ministry of Agriculture, Nature and Food Quality.

References

Basso B., and Antle, J. (2020) Nature Sustainability 3:254-256. https://doi.org/10.1038/s41893-020-0510-0.

Dittrich I., Gertz M., and Krieter J. (2019) Heliyon 5(11):e02902. https://doi.org/10.1016/j.heliyon.2019.e02902.

Ledgerwood DN., Winckler C., and Tucker CB. (2010) J Dairy Sci 93(11):5129-5139. https://doi.org/10.3168/jds.2009-2945.

Kamphuis C., Adriaens I., Ouweltjes W., and Hulsegge I. (2022) Proc. of the 12th WCGALP, Rotterdam, the Netherlands.

Porto S., Arcidiancono C., Anguzza U., and Cascone G. (2013) Biosyst Eng 115(2):184-194. https://doi.org/10.1016/j.biosystemseng.2013.03.002.

Rushen J., Chapinal N., and de Passillé AM. (2012) Anim Welfare 21:339-350. https://doi.org/10.7120/09627286.21.3.339.

134. Image-based cattle conformation prediction using deep learning methods

D. Segelke*, H. Alkhoder and J. Wabbersen

Vereinigte Informationssysteme Tierhaltung w.V. (vit), Heinrich-Schröder-Weg 1, 27283 Verden, Germany; dierck.segelke@vit.de

Abstract

Automated image-based phenotyping has become more feasible due to the improvement in deep learning methods aiming at identifying regions of interests within an image. To investigate the possibility of using standard deep learning tools to predict the conformation score out of an image, a dataset of 8,950 images obtained from 3,561 different cows was used. Each animal was also scored by a classifier. The conformation traits analysed included: udder depth, rear leg set and rump angle. These showed that automated extraction of relevant regions within an image are possible and the defined regions can be used to predict a classification score for an animal. Image-based (deep learning) heritability scores were 0.35, 0.20 and 0.25 for udder depth, rear leg set and rump angle, respectively. Genetic correlations between both methods were highly positive for all traits.

Introduction

In commercial dairy herds reduced costs of genotyping result in a growing genotyping interest. However, to receive reliable breeding values, accurate phenotyping in large scale has become increasingly important. Due to the high labour costs, the total cost of phenotyping for breeding programs is high. Furthermore, differences between and within classifier scores is another drawback. Improved computer-based image analyses might be the solution to reduce phenotyping costs while obtaining reliable phenotype scores for breeding value estimations. Nye *et al.* (2020) showed that estimating conformational traits in dairy cattle is feasible using a two-step deep learning automated phenotyping and segmentation approach. High heritabilities suggest that more meaningful biological information could be extracted from image data. Manzanilla Pech *et al.* (2021) indicated that reliable genetic parameters for feed efficiency can be derived from 3D camera-based systems. Poppe *et al.* (2019) reported another approach for affordable extraction of large-scale automated conformation traits. They applied their method on cartesian teat coordinates measured by automatic milking systems to record udder conformation traits. The aims of this study were to investigate whether classification scores can be predicted from image data and to estimate genetic parameters from this novel data.

Materials & methods

Data. To collect the images, the vit mobile app for reporting the classification results was extended to support both taking and uploading images of defined body regions. To receive comparable images, the app shows a silhouette of the body region of interest. The classifiers were asked to take up to seven different images per cow, covering all important routine conformational traits. Out of these seven different images per cow, a single image was used for the prediction of one trait. The traits in this study include udder depth (UDE), rump angle (RAN), and rear leg set (RLS). The exact number of images for the studied traits is shown in Table 1. Images were randomly split into four distinct subsets. The splitting process took place at different times for the three traits while new images were still coming in. Hence, there were different numbers and ratios of images in each of the four subsets. Images that were received after the splitting were moved to the cross-validation and test sets. The prediction for each trait relies on two models connected in series, which were trained and validated: One model to find the regions of interest (ROIs) in the image, as depicted in Figure 1 and Figure 2, and a second model to predict the score by using the post-processed

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_134, © D. Segelke *et al.* 2022

regions of interest found based on the first model. In Table 1, the subset used for training the ROI detection model is called training set and the subset used for validating this first model is called validation set. The subset called cross-validation set was used to select and tune the second model of the pipeline (i.e. the model that predicts the trait's score). The test set was finally used to evaluate the presented system on images that were never touched before in order to get an unbiased performance estimate. It should be noted that the present system does not estimate a score for all images available in the test set. This is due to several reasons such as: images without cows, images with multiple cows in the centre of the image, images that do not include all necessary ROIs, or failed detection of all necessary ROIs. Besides the image data set, a data set with classification scores on all linear conformation traits on the same cows were available. Classification scores ranged from 1 to 9 and were scored by 15 professional classifiers.

Two-step system. All ROI detection models were trained with the TensorFlow Object Detection API. For UDE, a pre-trained EfficientDet D1 was fine-tuned with the labelled training set in order to detect different morphological regions within an image. The outcome is a list of bounding boxes of the ROIs searched for, as depicted in Figure 1. The list of bounding boxes was then cleaned to only include the ones from the cow in the centre of the image. The second model, which predicts the UDE score, is just a simple formula dividing the distance between udder floor and hock by the distance between hock and buttock, followed by an affine transformation to receive scores from 1 to 9. The divisor of the formula approximately corrects different distances between the camera and the cow that leads to different sizes of the cow in the image.

Table 1. Nb. of cows in different subsets for udder depth, rump angle and rear leg set.

Dataset	Udder depth	Rump angle	Rear leg set
Total	3,302	2,734	2,914
Training	326	282	182
Validation	81	73	30
Cross-validation	1,060	937	1,226
Test	1,835	1,442	1,476
Heritability estimation	1,551	1,170	1,430

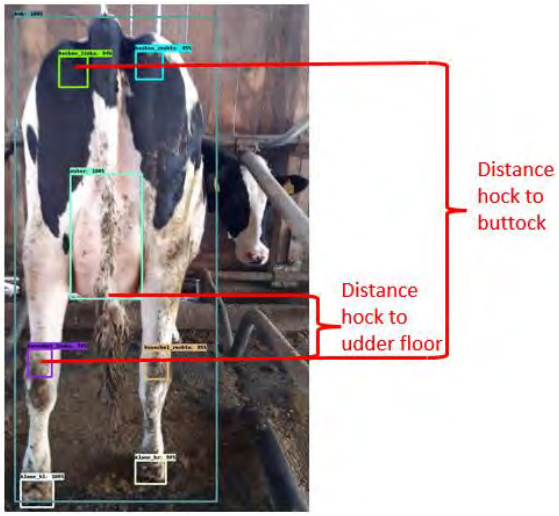


Figure 1. Schematic example of the processing steps for udder depth.

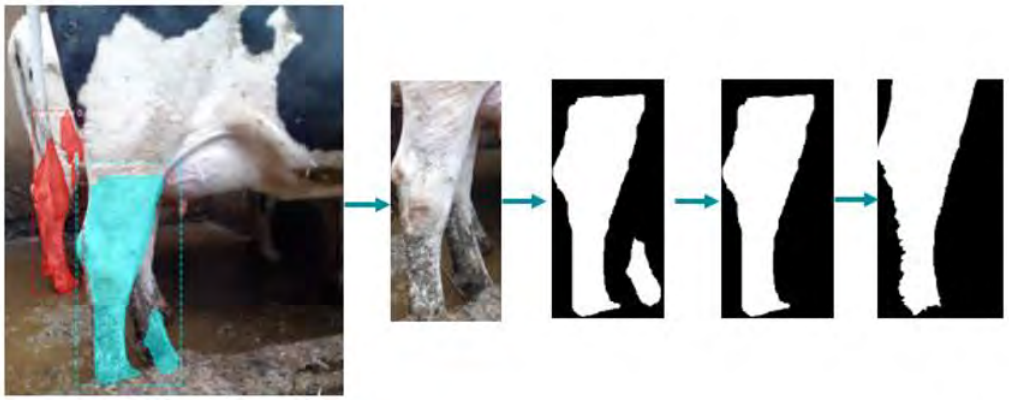


Figure 2. Schematic example of first processing steps for rear leg set.

For RAN, the same pre-trained EfficientDet D1 was fine-tuned to detect pins, tailheads, hooks, and withers. A similar post-processing as for UDE was used. In order to estimate scores, the bounding box information was used to create features such as distances between boxes, and angles between line segments connecting the centres of bounding boxes. These features were used as input for a feedforward neural network (FFNN) to estimate the conformation score. The training of the FFNN was performed by using the scores of the classifiers.

For RLS, a pre-trained Mask R-CNN was fine-tuned to detect hind legs. The outcome of this model is visualized in Figure 2. Mask R-CNN is a model for instance segmentation, meaning that it predicts, in addition to bounding boxes, a pixel mask for every hind leg it detects in the image. This way, the shape of the hind leg of interest is extracted by searching for a centred and large enough hind leg detection in the image, removing small blobs not belonging to the hind leg, and correcting the boundary of the hind leg with a method called GrabCut. The outcome of this post-processing is fed into a convolutional neural network to predict the trait's score.

Statistical models. To investigate whether derived traits from the images have a biological meaning, variance, and covariance components for both scored and image-based traits were estimated by fitting a bivariate animal model based on a restricted maximum likelihood (REML) approach using the software ASReml 4.1. The general model can be described in matrix notation as follows:

$$Y = X\beta + Z\alpha + e \quad (1)$$

where Y was the vector of observations (scored conformation trait, imaged conformation trait); β were the vectors of fixed effects for the analysed traits which were herd, classifier, days in milk, age at first calving and region, α was the direct additive genetic effects; X and Z were the incidence matrices related to the respective effects and e was the vector of random residuals. All animals were scored in 2021, thus a year of classification effect was therefore not considered.

Results

Table 2 shows the heritabilities for the three traits as well as the genetic and phenotypic correlations between the scoring by classifiers and the image-based scoring. Heritabilities from classifier-scoring were higher than the image-based scoring. However, the standard error was in general high and slightly lower

Table 2. Heritabilities (h^2) of udder depth, rump angle, and rear leg set scored by classifiers and image-based scoring and the genetic (r_g) and phenotypic correlation (r_p) with SE in parentheses.

Conformation trait	h^2 class.	h^2 image	r_g	r_p
Udder depth	0.56 (0.11)	0.35 (0.09)	0.95 (0.05)	0.67 (0.02)
Rear leg set	0.31 (0.09)	0.20 (0.08)	0.90 (0.16)	0.48 (0.02)
Rump angle	0.48 (0.13)	0.25 (0.11)	0.86 (0.12)	0.56 (0.02)

for the image-based approach. Genetic correlations between both approaches were high for all three traits meaning that image and classifier-based scoring are genetically the same trait, although the standard error was relatively. Phenotypic correlations between both trait definitions were lower than the genetic correlations ranging between 0,49 and 0,67 with a very small standard error.

Discussion

High genetic correlations suggest that images and deep learning methods can be used to predict the conformation of an animal. Large heritability estimates and standard errors indicate that larger data sets are needed to compare the strengths and weaknesses of both scoring methods. Because field data was used, the dataset contained a relatively large number of different classifiers and herds. The results of this study are based on a relatively low number of observations per classifier and herd, which may reduce the reliability of heritability estimations. In tendency, it could be observed that image-based scoring is less confounded with the classifier and the region of scoring. Furthermore, only one image per region of the cow was used and the trait of interest was not always fully covered by the image. However, these promising results show that routine field data can be used for animal scoring and the deep learning methods can effectively extract region of interests from a large variety of different farms, light conditions and not highly standardised distances between camera and region of interest.

Acknowledgements

German breeding organisations MASTERRIND and RUW are kindly acknowledged for data collection.

References

Manzanilla Pech, Coralia Ines V; Stephansen, Rasmus Bak; Lassen, Jan (2021) AAABG Proceedings, Adelaide, Australia.

Nye J., Zingaretti L., and Perez-Enciso M. (2020) Frontiers in Genetics 11:513. http://dx.doi.org/10.3389/fgene.2020.00513

Poppe M., Mulder H.A., Ducro B.J., and de Jong G. (2019) Journal of Dairy Science 102 (2):1386-1396. https://doi.org/10.3168/jds.2018-14838

135. Prediction fat percentage and visceral weight from whole fish images with a multi-input neural network

Y. Xue*, J.W.M. Bastiaansen and H. Komen

Wageningen University & Research, P.O. Box 338, 6700 AH Wageningen, the Netherlands;
yuanxu.xue@wur.nl

Abstract

In aquaculture, high accuracy in trait measurements benefits the genetic progress from a breeding program. Breeding traits like fat percentage and visceral weight are related to feed/cost efficiency of growth and product quality, and important metabolism and health indicators. Problems concentrate on finding the proper methods to accurately measure or predict these traits, as most current approaches are invasive, labour-intensive or may disturb or damage the fish. Interior trait prediction from image analysis would allow a real-time, large-scale and non-invasive alternative for such traits. This study investigates using whole-fish images in combination with exterior traits to improve the prediction of fillet fat percentage and visceral weight. The result of including images as extra input shows improvement on the accuracy of fat percentage prediction. The neural network extracted contour-based features and brings into view several biological indicators that appear to be informative for prediction.

Introduction

In aquaculture, accurate measurements of breeding traits on individuals are essential for the assessment of breeding values and the progress of the breeding program. Modern breeding programs emphasize not only on production traits, but also traits that are important indicators of metabolism and health of the fish, such as fat percentage and visceral weight. However, these two traits are costly to measure for two reasons: the required measurement involves stress challenges and may require sacrificing the fish, risking the loss of a potential breeding candidate; these measurements are invasive and damaging, raising societal and ethical concerns (Seibel, Weirup and Schulz, 2020).

To answer the need for accurate assessment of these traits, researchers and breeders start to include indirect measurements. Popular approaches include fitting prediction models using easy-to-measure exterior traits to approximate the fat percentage and visceral weights (García-Celdrán *et al.*, 2015, Vandeputte *et al.*, 2017). Another approach, image analysis, has received increasing attention in recent years (Prchal *et al.*, 2020). However, image analysis often heavily relies on morphological annotation. This transfer of image data to interpretable numeric value inevitably discards other image features that are potentially informative to prediction. Color pattern, for instance, is difficult to be objectively quantified as a morphological reading. Yet the information is easy to obtain non-invasively and can be valuable to assess the health status of fish (Parolini *et al.*, 2018). Combined with the proper statistical methods, images might offer more than annotated predictors for invasive traits.

This study estimates the accuracy of predicting fat percentage and visceral weight by different models. It aims to elucidate if adding whole-fish image as extra inputs is beneficial to predict these two traits, and if so, how images contribute to the prediction.

Materials & methods

Data was collected on a total of 4766 gilthead seabream (*Sparus aurata*) and kept separated in two groups. Images for these seabreams were taken with a Logitech HD Pro Webcam C920.

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_135, © Y. Xue *et al.* 2022

Group A. Records of fat percentage and body weights on 3,069 individuals with their corresponding images. Data was modelled for the predictability of body fat percentage based on body weight with and without images.

Group B. Records of visceral weights, fat percentage and body weights on 1,697 individuals with their corresponding images. Data was modelled for the predictability of visceral weight based on body weight, body fat percentage, and with and without images.

For each group, data was divided into training, validation and test sets with a size ratio of 3:1:1. Training and validation sets were used for exploratory data analysis, cross validation and hyperparameter tuning. To avoid information leakage, test set was only used once for the final evaluation of the prediction. Three methods were included for comparison on their prediction performance.

Linear regression (LR). Although the assumed relationship between the variables and the response is simple, linear models are highly robustness with low computing cost. Given its simplicity and robustness, LR is chosen as the benchmark to use body weight for prediction of fat percentage and to use body weight and fat percentage for prediction of visceral weight. LR was combined with 5-fold cross validation to avoid overfitting.

Multilayer perceptron (MLP). Trait prediction sometimes requires more complex model as the relationship between the predictors and response are not always linear. Therefore, this study uses MLP to represent a simplified collection of all complex prediction model like polynomial regression, sinusoidal regression, etc. It is intended to capture the real relationship by going through all possible combinations of mathematical connections between body weight, fat percentage, and visceral weight to use body weight for prediction of fat percentage and to use body weight and fat percentage for prediction of visceral weight. Images are not included as predictors.

MLP is in essence a collection of neurons. Each neuron consists of a value and relates to either an input or an output neuron, or both. Neurons are grouped per layer. The connections between layers form a chain of mathematical functions that continuously manipulate the original input towards a final output. These functions can be as simple as a subtraction or addition or as sophisticated as a derivative. At the end of every round of calculation, these functions will self-update to minimize the difference between the calculated and observed output until the difference reaches the minimum. This is the learning process through which a MLP finally realizes a prediction.

Multi-input neural network (NN). To incorporate images to the prediction, a multi-input NN is applied. It is meant to include all pixel-level information from a whole-fish image, regardless of the biological interpretability of the information, and combine it with other measurement like body weights. A multi-input NN consists of two parts: a convolutional neural network (CNN) and an MLP. CNN scans the images repeatedly using different filters. These filters form different layers like neurons in MLP and translate the value of raw image pixels into imaginal features and concepts. In parallel, the MLP establishes a mathematical relationship between the numeric input from other traits like body weight. The outputs of these two compartments are combined and matched with the response variables. Multi-input NN uses the same 'learning process' as MLP to update all connected layers, until the difference between the observed output and the prediction stabilizes and reaches its minimum.

Feature visualization helps to visualize the learning process by displaying the filtered features from each convolutional layer. Mathematically, these features maximize the information that contributes to the

observed output. According to the human visual comprehension, these features can vary from edges and textures to patterns and parts of the objects.

The predictions were evaluated using two metrics: Mean squared error (MSE) and Pearson correlation coefficient (r). Lower MSE and higher r values imply that a prediction model has better prediction performance.

Results

Table 1 shows the results of prediction fitting three models on both test set. Accuracy of predicting visceral weight is high and similar for all methods (0.87-0.90). Prediction of fat percentage benefits from the image data which increases accuracy from 0.48 to 0.70.

To understand how images facilitate the prediction, the extracted features by each convolutional layer are displayed in Figure 1. Each row contains features obtained from the same layer. Moving to the next row, the layer goes deeper, which means that the features extracted will become more abstract but also more informative to the final prediction. The first row in Figure 1 illustrates that in the beginning layer, multi-input NN extracts different features, like whole-fish contours and contrasts between regions, as informative inputs for fat percentage prediction. Colour might also be considered as an informative feature given the visual inconsistency in the background. Visually dark regions on the fish correspond to locations like patch, eye and fin on the colour image. However, the multi-input NN quickly abandons this information and fixates on outlines and contours, starting from the second row of Figure 1. The final convolutional layer in the last row comprises features mostly highlighted in visually dark colour: the curvature of the dorsal side, together with the upper edge of pectoral fins.

Figure 2 compares the final convolutional features in both groups. Like its performance in group A on fat percentage, multi-input NN applied to visceral weights in group B has a strong preference for the contour.

Table 1. Evaluation of prediction methods on test sets.

	Group A – fat percentage		Group B – visceral weight	
Metrics	**MSE[1]**	**r**	**MSE[2]**	**r**
LR	4.86	0.48	63.26	0.90
MLP	5.26	0.48	75.41	0.89
Multi-Input NN (with images)	2.98	0.70	87.14	0.87

[1] Test set has a mean of 11.52 and a standard deviation of 2.5.

[2] Test set has a mean of 40 and a standard deviation of 18.

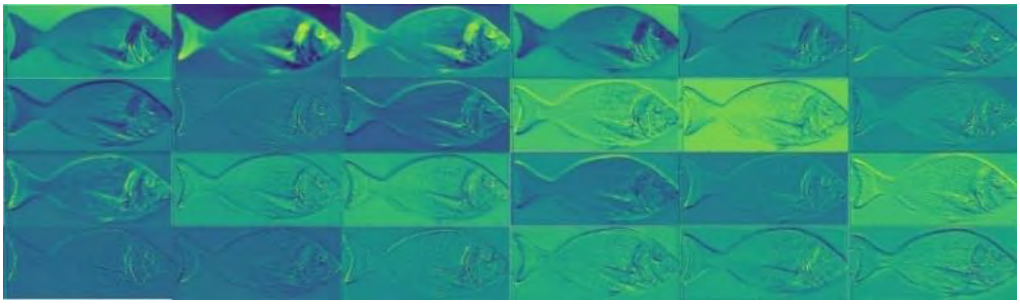


Figure 1. Visualized extracted features by each convolutional layer of the multi-input neural network from Group A, fat percentage.

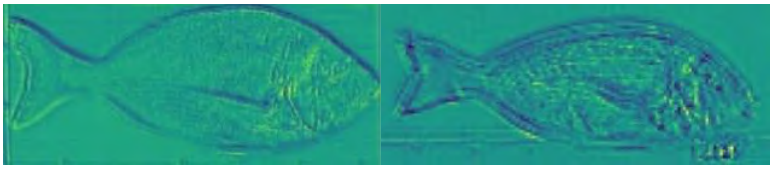


Figure 2. Visualized extracted features from the last convolutional layer of the multi-input neural networks. Left: Group A, fat percentage; Right: Group B, visceral weight.

The right image in Figure 2 shows an overlapping highlighted pattern around the edges, indicating that the contour has been read repetitively for visceral weight prediction. Compared to the result in fat percentage, however, there are additional patterns revealed on the body of the fish, especially around the operculum, dorsal edge and anal fin.

Discussion

For fat percentage prediction, accuracy of LR and MLP are similar. This similarity on accuracy indicates that the non-linear, complex MLP failed to capture more information compared to a linear model for predicting fat percentage. In contrast, multi-input NN uses the images and outperformed both LR and MLP. Features extracted suggest two important factors in predicting fat percentage: contour and pectoral fin. Contour can be interpreted as the size of body area, which is strongly correlated with body weight (Fernandes *et al.*, 2020). However, given the poor performance of LR with only body weight as predictor, the multi-input NN probably uses a different interpretation of the contour. Curvature is hypothesized to be the informative factor in fat percentage prediction. The result of MLP implies that prediction accuracy does not benefit from mathematically complex indicators. Including the images as inputs identifies curvature which is also a rather explicit morphological description. Further work could separate the contour and investigate a proper quantitative description of the curvature and of the location of pectoral fin and test the predictive value of these features. However, without a quantitative description of the contour, images can still improve the fat percentage prediction by the feature extraction of multi-input NN.

For visceral weight prediction, three models yielded similar results, suggesting that neither the complexity of the MLP method nor the information-concentrated image are good additions for visceral weight prediction. Future work may look into replacing body weight by body features detected from images as a fast, non-invasive alternative to assess the visceral weight.

Acknowledgements

This project receives financial support from the Topsector Agri & Food.

References

García-Celdrán, M., Ramis, G., Manchado, M., Estévez, A., Afonso, J. M. *et al.* (2015). Aquaculture 446:175–180. https://doi.org/10.1016/j.aquaculture.2015.04.028

Fernandes, A. F., Turra, E. M., de Alvarenga, É. R., Passafaro, T. L., Lopes, F. B. *et al.* (2020). Computers and electronics in agriculture, 170, 105274. https://doi.org/10.1016/j.compag.2020.105274

Parolini, M., Iacobuzio, R., Bassano, B., Pennati, R., and Saino, N. (2018). Physiological and Biochemical Zoology, 91(5): 1026–1035. https://doi.org/10.1086/699522

Prchal, M., Kocour, M., Vandeputte, M., Kause, A., Vergnet, A. *et al.* (2020) Aquaculture, 520. https://doi.org/10.1016/j.aquaculture.2020.734993

Seibel, H., Weirup, L., and Schulz, C. (2020). Food Ethics, (1–2). https://doi.org/10.1007/s41055-019-00063-3.

Vandeputte, M., Puledda, A., Tyran, A. S., Bestin, A., Coulombet, C. *et al.* (2017) Aquaculture 470:40–49. https://doi.org/10.1016/j.aquaculture.2016.12.014

136. A case-control GWAS for the chicken plumage colour using a computer vision approach

S. Heo[1*], S. Cho[1], J. Park[1], M. Kim[1], E. Cho[1], D. Jin[2], J. Cha[3], Y.J. Koh[1] and J.H. Lee[1]

[1]Chungnam National University, 99, Daehak-ro, Yuseong-gu, 34134, Daejeon, Republic of Korea; [2]National Institute of Animal Science, Animal Genetic Resources Research Centre, 224, Deogyuwolseong-ro, 50000, Seosang-myeon, Hamyang-gun, Gyeongsangnam-do, Republic of Korea; [3]National Institute of Animal Science, Animal of Genomics and Bioinformatics Division, 1500, Kongjwipatjwi-ro, 55365, Iseo-myeon, Wanju-gun, Jeollabuk-do, Republic of Korea; heoseonyeong9@gmail.com

Abstract

There has been a limitation in which ambiguity occurs in classifying qualitative traits such as colour. To solve it, a phenotyping approach of analysing RGB (Red, Green, and Blue) values of chicken plumage colour and assorting it into binary groups is proposed. With a total of 389 animals (100 high-pigmented as a case group and 289 low-pigmented as a control group), a case-control genome-wide association study was performed using F_2 offspring population of a cross between two different pure lines, White Leghorn and Yeonsan Ogye chicken. The result revealed 51 significant markers including 5 candidate genes for eumelanin pigmentation: *PMEL*, *SLC48A1*, *NEUROD4*, *KRT6A*, and *KRT80*. Therefore, this study can give sights to classify diverse phenotypic variance and to investigate genetic mechanisms related to eumelanin pigmentation in chicken plumage.

Introduction

Generally, the plumage colour variation in birds is mainly attributed to the total melanin concentration and the ratio of eumelanin to pheomelanin in the feather. The Eumelanin gives rise to black, brown, or grey colouration, while the pheomelanin makes it appear light red and yellow (McGraw *et al.*, 2005).

Many studies have been conducted to find the genetic basis related to chicken plumage pigmentation, colouration traits have been judged by human eyes (Kinoshita *et al.*, 2014, Yang *et al.*, 2017). A recent attempt has been made to objectively represent diverse colour variations using quantitative values for shank colours in chickens (Jin *et al.*, 2016). However, the method of classifying qualitative traits such as colour has the limitation that the criteria can still be subjective and ambiguous when there is a diverse variance in colour (Yang *et al.*, 2017; Jin *et al.*, 2016). Therefore, this study aimed to identify genes and genetic regions that affect eumelanin pigmentation in chicken plumage using an objective criterion for differentiating the degree of pigmentation.

Materials & methods

Animals. The F_2 cross-bred population was generated between Yeonsan Ogye (YO), a Korean native chicken breed and White Leghorn (WL) chicken. The YO has entirely black external features including plumage in contrast to the WL which has pure white plumage. Reciprocal mating between unrelated individuals, i.e. 1 WL male × 5 YO females were used to produce the F_1 generation. One F_1 male and 10 F_1 females were randomly selected to produce the F_2 generation. The distribution of body plumage colour in F2 population is as follows: all-black (10.0%), black and white barred (12.6%), black with brown-spotted (0.8%), black with white spotted (0.5%), all-white (35.7%), white with brown-spotted (9.5%), white with black-spotted (23.1%), bright grey (5.9%), dark grey (1.8%). In this study, a total of 389 F_2 chickens, which have both phenotypic and genotypic information was used as samples.

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_136, © S. Heo *et al.* 2022

Phenotyping. The process of collecting RGB (Red, Green, and Blue) values for classifying the phenotype is as follows: all photographs were acquired using a camera under the same background, lighting, angle, and position of the chickens. To accurately crop the chicken layer in the samples, the background was removed under a Python computing environment. Then, a segmented layer of body feathers excluding hackles, wings, and tail feathers in the sample was cropped out using Adobe Photoshop CC 2021. The RGB values of all pixels in the segmented layers were derived, and the average values were calculated. The distribution in a histogram of average RGB values of total samples was used for setting the criterion for differentiating two phenotypes, one with a high proportion of eumelanin pigmentation (HP) and the other with lower of it (LP).

Genotyping. Genomic DNA was extracted from blood samples collected from the chicken using the Wizard Genomic DNA Purification Kit (Promega, Wisconsin, USA). All DNA samples were genotyped using the Illumina chicken 60K SNP BeadChip.

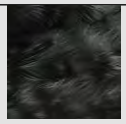
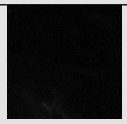
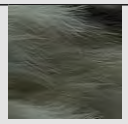
PLINK v1.9 software (Purcell *et al.*, 2007) was used for the quality control (QC) of genotyping. Only SNPs with a call rate >90% and a minor allele frequency (MAF) >5% were used for subsequent QC steps. Further, SNPs with the *P*-value of deviation from Hardy-Weinberg equilibrium (PHWE) $<1\times10^{-4}$ were removed. Ultimately, 389 individuals and 19,876 SNPs were retained for the association analysis.

Case-control GWAS. To identify loci conferring eumelanin pigmentation in chicken plumage, we performed the case-control GWAS through PLINK v1.9 software with a total of 389 chicken samples, including 289 LP (74.3%) and 100 HP (25.7%). The HP was used as a case group whereas the LP was used as the control group (Table 1). The genome-wide significance level was set based on the Bonferroni correction *P*-value as $0.05/19{,}876 = 2.52E^{-6}$.

Results

The distribution of average RGB values was divided into two: one below 50 and the other above 50 (Figure 1). The samples were divided into the former and the latter group, and the maximum and minimum values of each RGB were derived, which are [45.2, 47.8, 45.4] and [59.6, 65.5, 59.1]. The average of these two values was used as a criterion for dividing the distribution. Therefore, one case less than [52.4, 59.7, 52.3] is classified as group HP, and another exceeding it is classified as group LP.

Table 1. The average RGB values of each image layer represent body feathers excluding hackles, wings, and tail feathers.

Group[2]	HP			LP		
Image layer[1]						
R (Red)	45.2	33.2	8.5	152.9	89.6	91.3
G (Green)	47.8	34.6	8.8	155.3	91.9	90.3
B (Blue)	45.4	32.5	8.6	131.2	80.6	74.5

[1] A part of image layers from each sample.

[2] HP; high-pigmented as a case group and LP; low-pigmented as a control group.

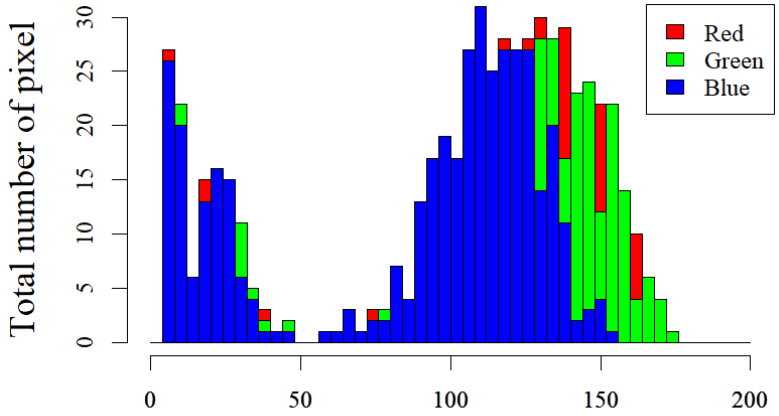


Figure 1. Histogram plot for the distribution of average RGB values of total samples.

We identified a total of 51 SNPs on chromosomes 1, 7, 18, 33 and linkage group E22C19W28_E50C23 (Figure 2). Specifically, the significant region (3.22-7.09 Mb) in chromosome 33 comprises 18 known genes and 5 anonymous genes. We identified the most significant SNP near genes, including *DGKA*, *CDK2*, *RAB5B*, and *PMEL* (P=5.369e^{-56}). Moreover, the GWAS results revealed *SLC48A1* and *NEUROD4*, as well as the keratin genes, *KRT6A* and *KRT80*.

Discussion

Genes involved in eumelanin pigmentation in many studies were reported as follows: the melanocortin 1 receptor (*MC1R*), *NUAK* family kinase 1 (*NUAK1*), sonic hedgehog signalling molecule (*SHH*), and premelanosome protein (*PMEL*) (Dávila *et al.*, 2014, Yang *et al.*, 2017; Kerje *et al.*, 2004). Of these genes, the *PMEL* gene, related to the dominant white allele, is a melanocyte protein responsible for eumelanin deposition and fibril formation in melanosomes (McGlinchey *et al.*, 2009; Kerje *et al.*, 2004). A recent study revealed that the downregulation of *PMEL* may lead to reduced eumelanin synthesis, which may alter eumelanin and pheomelanin content in chicken plumage (Zheng *et al.*, 2020).

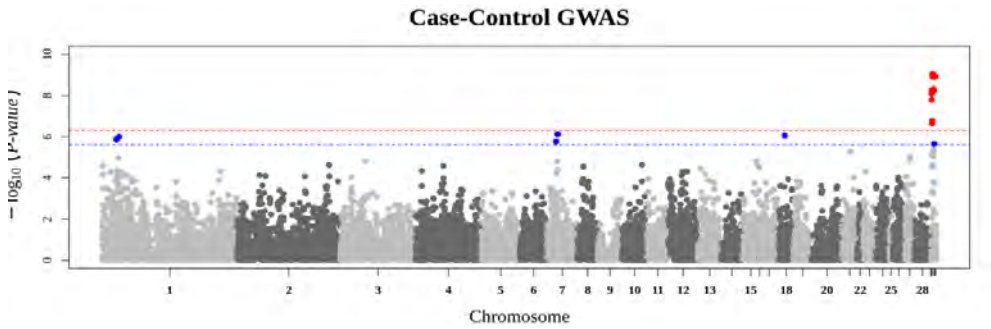


Figure 2. Manhattan plot representing the case-control GWAS result. The red and blue lines indicate Bonferroni corrected significance (α=0.01 and 0.05, respectively).

Other identified genes near the SNPs on chromosome 33 such as *SLC48A1*, *NEUROD4* have not been reported as relevant to eumelanin pigmentation. However, a previous study revealed the *SLC48A1* and *NEUROD4* in ~0.51 Mb region of the E22C19W28_E50C23 linkage group are related to *HOXC8* which is expressed in chicken cranial skin (Wang *et al.*, 2012). The homeobox genes including *HOXC8* turned out to be candidate genes involved in the plumage melanogenesis in black-bone chicken (Yu *et al.*, 2018). Therefore, *SLC48A1* and *NEUROD4* may be candidate genes for eumelanin pigmentation in chicken.

Of the candidate genes, the *KRT6A* and *KRT80* have clues that these are related to feather follicles in chicken. Previous studies found that the *KRT6A* gene, which is responsible for the development of hair follicles in mice, may also influence the development of feather follicles in chicken embryo fibroblasts (Chen *et al.*, 2017; Gu and Coulombe, 2007). In addition to that, *KRT6A* and *KRT80* were included in the results of GWAS for the frizzle feather trait in chickens (Chen *et al.*, 2021).

Previous studies reported that the crested and frizzle and dominant white traits were linked and located on linkage group E22C19W28 (Romanov *et al.*, 2004; Ruyter-Spira *et al.*, 1997). It implies various genes may influence eumelanin pigmentation in chicken feathers. Therefore, this study can be used as genetic clues to reveal the physiological mechanism of plumage pigmentation by indicating five candidate genes: *PMEL*, *SLC48A1*, *NEUROD4*, *KRT6A*, and *KRT80*.

References

Chen B., Xi S., El-Senousey H.K., Zhou M., *et al.* (2022) 53(1):101-107. https://doi.org/10.1111/age.13158

Chen S., Hu X., Zhu W., Jia C., *et al.* (2017) Poult. Sci. 96(4):1007-1014. https://doi.org/10.3382/ps/pew378

Dávila S.G., Gil M.G., Resino-Talaván P., Campo J.L. (2014) Poult. Sci. 9(5):1089-1096. https://doi.org/10.3382/ps.2013-03611

Gu L.H., and Coulombe P.A. (2007) J. Invest. Dermatol. 127(5):1061-1073. https://doi.org/10.1038/sj.jid.5700673

Jin S., Lee J.H., Seo D.W., Cahyadi M., *et al.* (2016) Asian Australas. J. Anim. Sci. 29(11):1555-1561. https://doi.org/10.5713/ajas.16.0183

Kerje S., Sharma P., Gunnarsson U., Kim H., *et al.* (2004) Genetics 168(3):1507-1518. https://doi.org/10.1534/genetics.104.027995

Kinoshita K., Akiyama T., Mizutani M., Shinomiya A., *et al.* (2014) PLoS One 9(1):e86361. https://doi.org/10.1371/journal.pone.0086361

McGlinchey R.P., Shewmaker F., McPhie P., Monterroso B., *et al.* (2009) Proc. Natl. Acad. Sci. U. S. A. 106(33):13731-13736. https://doi.org/10.1073/pnas.0906509106

McGraw K.J., Safran R.J., and Wakamatsu K. (2005) Funct. Ecol. 19(5):816-821. https://doi.org/10.1111/j.1365-2435.2005.01032.x

Purcell S., Neale B., Todd-Brown K., Thomas L., *et al.* (2007) Am. J. Hum. Genet. 81(3):559-575. https://doi.org/10.1086/519795

Romanov M.N., Sazanov A.A. and Smirnov A.F. (2004) Worlds Poult. Sci. J. 60(1):19-41. https://doi.org/10.1079/WPS20032

Ruyter-Spira C.P., Gu Z.L., Van der Poel, J.J., and Groenen M.A. (1997) Poult. Sci. 76(2):386-391. https://doi.org/10.1093/ps/76.2.386

Wang Y., Gao Y., Imsland F., Gu X., *et al.* (2012) PLoS One 7(4):e34012. https://doi.org/10.1371/journal.pone.0034012

Yang L., Du X., Wei S., Gu L, *et al.* (2017) Anim. Genet. 48(5):611-614. https://doi.org/10.1111/age.12573

Yu S., Wang G., Liao J., Tang M., *et al.* (2018) Cell. Physiol. Biochem. 46(6):2373-2384. https://doi.org/10.1159/000489644

Zheng X., Zhang B., Zhang Y., Zhong H., *et al.* (2020) Sci Rep 10(1):1-11. https://doi.org/10.1038/s41598-020-68931-1

137. Machine learning to predict somatic cell count at the subsequent test-day record in the Italian Mediterranean Buffaloes

T. Bobbo[1,2*], R. Matera[3], G. Pedota[4], J. Ramirez-Diaz[1,5], A. Manunza[1], A. Stella[1], A. Cotticelli[3], G. Neglia[3] and S. Biffani[1]

[1]Institute of Agricultural Biology and Biotechnology, National Research Council, Via Edoardo Bassini 15, 20133 Milan, Italy; [2]Department of Agricultural and Environmental Sciences, University of Milan, Via Celoria 2, 20133 Milan, Italy; [3]Department of Veterinary Medicine and Animal Production, Federico II University, Via Federico Delpino 1, 80137 Naples, Italy; [4]Associazione Regionale Allevatori della Basilicata, Via Dell'Edilizia, 85100 Potenza, Italy; [5]Department of Animal Science, Food and Technology, Università Cattolica del Sacro Cuore, Via Emilia Parmense 84, 29122 Piacenza, Italy; tania.bobbo@unimi.it

Abstract

Collection of a large amount of data during the routine milk recording procedures makes it possible to train machine learning algorithms for predicting traits of interest. In this study, we applied two different machine learning methods (General Linear Models and Random Forest) to predict the udder health status of Italian Mediterranean Buffaloes at the subsequent monthly test-day record, using milk data collected at the previous test-day. Prediction accuracies of both methods were above 73%. According to different metrics, General Linear Models had the best performance in predicting the udder health classes (high or low somatic cell count, based on a threshold of 200,000 cells/ml). Findings of this study highlight the promising role of machine learning algorithms to improve the surveillance of subclinical mastitis, helping farmers in the preventive identification of animals that would possibly have high somatic cell count the subsequent month. Such information may also be exploited for breeding purposes.

Introduction

Subclinical mastitis, i.e. the inflammatory condition of the mammary gland with no clinical signs, is a costly health issues of the dairy Mediterranean Buffaloes (*Bubalus bubalis*) (Alterisio *et al.*, 2021). Efforts have been made to improve udder health, exploring also novel indicators of mammary gland inflammation based on traditional somatic cell count (scc) (Costa *et al.*, 2021). However, there is still need to improve mastitis detection, management and selection. Collection of a large amount of data in the frame of the routine milk recording procedures and by automatic milking recording systems makes it possible to exploit such information to train machine learning algorithms for predicting specific traits of interest. Machine learning analyses have already been applied in different areas of dairy research to support farmers in decision making (Cockburn, 2020). Previous studies have applied machine learning methods to predict subclinical mastitis in dairy cows, some based on the presence of high scc in milk (Bobbo *et al.*, 2021; Ebrahimi *et al.*, 2019), others based on the presence of mastitis pathogens (Hyde *et al.*, 2020; Sharif *et al.*, 2018). Nevertheless, only few studies reported predictions based on data collected at different time points (Anglart *et al.*, 2020; Bobbo *et al.*, 2021). Following the approach reported by Bobbo *et al.* (2021), in the present study we exploit information collected during the routine milk recording to predict the presence of high or low scc level in milk collected the subsequent month in Italian Mediterranean Buffaloes.

Materials & methods

Data collection and editing. Test-day (TD) records were provided by the Italian Breeders Association (Rome, Italy). Data included information about herd, animals (ID, stage of lactation and parity), daily milk production, milk composition, scc, β-hydroxybutyrate (bhb), electrical conductivity, and milk coagulation properties [rennet coagulation time (r) and curd firmness 30 minutes after rennet addition (a30)]. The dataset was edited to include animals that calved in the years 2019 and 2020, with at least two TD within

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_137, © T. Bobbo *et al.* 2022

lactation, and with less than 360 days in milk (dim). In addition, only consecutive TD records with an interval lower than six weeks were selected. Among milk traits, outliers beyond four standard deviations were considered as missing values and only full records were considered. Average daily milk production (milk_htd) and scc (scc_htd) of contemporary groups, i.e. animals sampled in the same herd and day, were also determined. Finally, the two scc-related traits were log-transformed to somatic cell score (scs and scs_htd). A total of 18 features were considered: parity (from 1 to ≥6), stage of lactation (dim30:12 classes of 30 d each), year and month of calving (ymc: 24 levels), year and month of sampling (yms: 19 levels), daily milk production, fat, protein, casein, lactose, pH, urea, scs, bhb, electrical conductivity (cond), milk_htd, scs_htd, and the two milk coagulation properties. Udder health status at the subsequent TD (outcome) was coded as a binary trait: low (≤200,000 cells/ml) or high (>200,000 cells/ml) scc. After editing, the dataset included 3,939 records of 1,035 buffaloes in six herds. In each record, information of two subsequent TD was provided: 18 features of the previous TD and outcome at the subsequent TD. The prevalence of subclinical mastitis (SCC>200,000 cells/ml) was 41.2%.

Data processing and models building. Machine learning analyses were performed following Bobbo *et al.* (2021). Prediction models were developed to identify the best udder health prediction method [Generalized Linear Models (GLM) or Random Forest (RF)] using information recorded at the previous TD. The dataset was randomly split into two subsets: 80% of the data (3,152 records) to train and test the models, and 20% of the data as a validation set. To automatically select the most predictive features, we applied a recursive feature selection using a 10-fold cross validation (cv) repeated 10 times with RF method. A stratified 10-fold cv repeated 10 times was then applied to train and test the models. The train/test dataset was randomly divided into 10 subsets; prediction models were trained on 9 subsets and tested on the last one. Each 10-fold cv was repeated 10 time, for a total of 100 iterations. Data standardization was performed within cv. Data analysis was performed using Caret v.6.0-86 (Kuhn, 2021) and Tidyverse v. 1.3.1 (Wickham *et al.*, 2019) packages of R software v.4.1.2 (R Core Team, 2021).

Evaluation of methods performance on testing and validation set. Sensitivity, specificity, accuracy, Cohen's kappa value, F1 score and area under the receiver operating characteristic curve (AUC) were used to compare the predictive performance of the models on the testing set. Feature importance based on the prediction model with the greatest accuracy was then assessed. Finally, predictive ability of both models on the validation set were estimated and compared using the same metrics reported above.

Results

Results of the recursive feature selection applied before model training suggested that the most parsimonious model (with the highest accuracy of prediction using the lowest number of features) was the one that included all the features of interest (Figure 1a).

Evaluation of methods performance in predicting udder health classes (low or high scc) on testing and validation sets was based on several metrics (Table 1). Whereas similar results were obtained in the testing set, GLM showed a slightly better performance in the validation set in comparison to RF. Nevertheless, in both methods, accuracy was above 73%, F1 above 78%, and AUC above 80%.

According to considered metrics, GLM was the best method in predicting udder health classes. Feature importance using GLM for predicting udder health status on testing set suggested that scs-derived traits were the most important features, followed by electrical conductivity and milk production (Figure 1b).

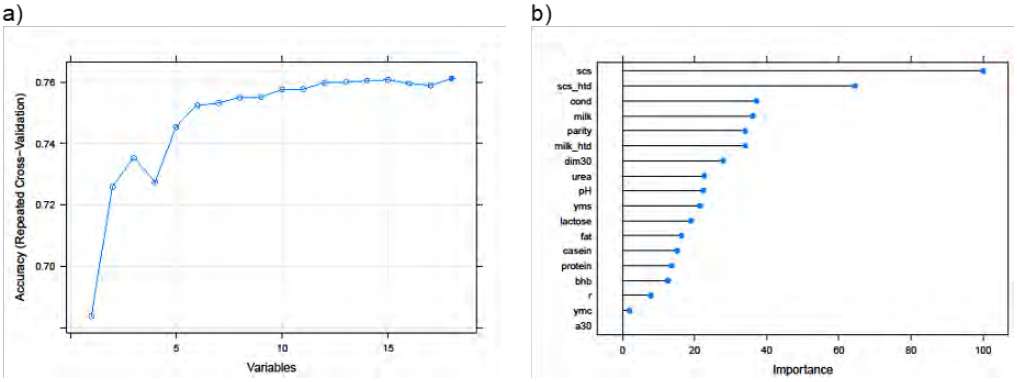


Figure 1. (a) Results of the recursive feature selection incorporating one to 18 features. (b) Plot of the feature importance showing the ranking for the prediction of udder health status in the testing set. Evaluated features, using Generalized Linear Models, are: individual somatic cell score and of contemporary group (scs and scs_htd), electrical conductivity (cond), milk production, parity, milk production of contemporary group (milk_htd), stage of lactation (dim30), urea, pH, year and month of sampling (yms), lactose, fat, casein, protein, β-hydroxybutyrate (bhb), rennet coagulation time (r), year and month of calving (ymc), and curd firmness 30 minutes after rennet addition (a30).

Table 1. Methods performance metrics [Sensitivity (Se), Specificity (Sp), Accuracy (Acc), Kappa value, F1 score, and area under the receiver operating characteristic curve (AUC)] on testing and validation sets.

Method	Testing set						Validation set					
	Se	Sp	Acc	Kappa	F1	AUC	Se	Sp	Acc	Kappa	F1	AUC
GLM[1]	0.84	0.62	0.75	0.47	0.80	0.82	0.84	0.62	0.75	0.47	0.80	0.81
RF[2]	0.83	0.64	0.75	0.47	0.79	0.82	0.81	0.63	0.73	0.44	0.78	0.80

[1] Generalized linear models.

[2] Random forest.

Discussion

In the present study, we applied two different machine learning methods (one linear and one based on decision trees) to predict the udder health status of Italian Mediterranean Buffaloes at the subsequent TD record, using already available milk data collected at the previous TD. The comparison of the performance of the two methods on both testing and external validation sets were based on several metrics. Besides accuracy, which is one of the most used metrics in classification problems, we included also F1 score and AUC, that have the advantage of being independent from outcome rate. Although GLM was the method with the best performance, similar results were obtained using RF. Findings of this study are in accordance with the results of Bobbo *et al.* (2021), who reported linear methods, RF and neural network to be the best methods for predicting the udder health status of Italian Holstein cattle. The scc-traits recorded at the previous TD were expected to be the most important features for making the prediction. The important role of electrical conductivity confirmed results of previous studies (Ebrahimi *et al.*, 2019, Ebrahimie *et al.*, 2018). In conclusion, machine learning methods applied to the large amount of data currently available represent a promising tool to improve the surveillance of subclinical mastitis, helping farmers in the preventive identification of animals that would possibly experience mammary gland issues. Further studies are required for a practical implementation of these methods in the management supporting systems as well as in the use of complex phenotypes in genetic and genomic evaluations.

Acknowledgements

This research was funded by Italian Ministry of Agriculture (MIPAAF – DISR 07). Project: 'BIG' Prot. N. 0215513 11/05/2021

References

Alterisio M.C., Ciaramella P., and Guccione J. (2021) Vet. Sci. 8(10):204. https://doi.org/10.3390/vetsci8100204

Anglart D., Hallén-Sandgren C., Emanuelson U., and Rönnegård L. (2020) J. Dairy Sci. 103(9):8433-8442. https://doi.org/10.3168/jds.2020-18320

Bobbo T., Biffani S., Taccioli C., Penasa M., and Cassandro, M. (2021) Sci. Rep. 11(1):1-10. | https://doi.org/10.1038/s41598-021-93056-4

Cockburn M. (2020) Animals 10(9):1690. https://doi.org/10.3390/ani10091690

Costa A., De Marchi M., Neglia G., Campanile G., and Penasa M. (2021) Ital. J. Anim. Sci. 20(1):548-558. https://doi.org/10.1080/1828051X.2021.1899856

Ebrahimi M., Mohammadi-Dehcheshmeh M., Ebrahimie E., and Petrovski K.R. (2019) Comput. Biol. Med. 114:103456. https://doi.org/10.1016/j.compbiomed.2019.103456

Ebrahimie E., Ebrahimi F., Ebrahimi M., Tomlinson S., and Petrovski K.R. (2018) J. Dairy Res. 85:193–200. https://doi.org/10.1017/S0022029918000249

Hyde R.M., Down P.M., Bradley A.J., Breen J.E., Hudson C., *et al.* (2020) Sci. Rep. 10(1):1-8. https://doi.org/10.1038/s41598-020-61126-8

Kuhn M. (2021). caret: Classification and Regression Training. R package version 6.0-90. Available at: https://CRAN.R-project.org/package=caret

R Core Team (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Available at: https://www.R-project.org

Sharif S., Pakdel A., Ebrahimi M., Reecy J.M., Fazeli Farsani S., *et al.* (2018) PloS one 13(2):e0191227. https://doi.org/10.1371/journal.pone.0191227

Wickham H., Averick M., Bryan J., Chang W., McGowan L.D.A., *et al.* (2019). J. Open Source Softw. 4(43): 1686, https://doi.org/10.21105/joss.01686

138. Computer vision systems to advance high-throughput phenotyping in livestock

J.R.R. Dorea[1,2*] and G.J.M. Rosa[1,3]

[1]Department of Animal and Dairy Sciences, University of Wisconsin-Madison, 1675 Observatory Drive, Madison-WI, 53712, USA; [2]Department of Biological Systems Engineering, University of Wisconsin-Madison, 1675 Observatory Drive, Madison-WI, 53712, USA; [3]Department of Biostatistics and Medical Informatics, University of Wisconsin-Madison, 1675 Observatory Drive, Madison-WI, 53712, USA; joao.dorea@wisc.edu

Abstract

Applications of computer vision are growing in importance in livestock systems due to their ability to generate accurate, real-time, and non-invasive animal-level phenotypes. However, the development of a computer vision system requires sophisticated statistical and computational approaches for efficient data management and appropriate data mining, as it involves large datasets. In this talk, we will discuss some of the challenges, applications, and potentials of computer vision systems in livestock phenomics. The development of computer vision technologies will potentially have a major impact in the livestock industry by providing real-time and accurate phenotypes, which can be used to enhance breeding for novel and hard-to-measure traits, as well as to assist animal research across various disciplines, and to improve management decisions in commercial herds.

Introduction

Computer vision and image analysis have been undergoing a great revolution in the last few years, mainly due to the advent of specific machine learning tools as deep learning. These methods show better performance than classic techniques for many challenging computer vision problems such as natural image segmentation, optical character recognition, and object classification (LeCun *et al.* 2015). Nowadays, computer vision systems using deep learning algorithms show comparable or better accuracy than humans in several visual recognition tasks, including recognizing traffic signs, faces, and handwritten characters. Among the various digital technologies in livestock, computer vision systems are emerging as a powerful solution for high-throughput phenotyping (HTP), which is crucial for the development of optimized farm management decisions and for the genetic improvement on novel traits, namely animal health and behavior traits.

High-throughput phenotyping and genetic improvement of livestock

HTP techniques based on computer vision can be used in animal breeding programs to provide more efficient and accurate phenotype scoring of traits traditionally used in breeding programs. For example, live body weight and average daily gain is nowadays measured at standard ages in beef cattle, such as birth, weaning, 365 and 550 days of age. With the implementation of computer vision, body weight can be estimated daily (even multiple times per day) for each animal, so selection can be applied not only for the traditional standard ages but also for the parameters describing the growth curve shape of each animal. In addition, computer vision tools can be used for scoring novel phenotypes or indicator traits to be used in advanced breeding programs (Silva *et al.*, 2021).

Phenotypes of economic or societal importance that are generally difficult to measure without digital technologies include individual feed consumption and feed efficiency, heat tolerance, mothering ability, aggressive behavior, disease resistance, among many others. Image analysis and computer vision systems (CVS) are extremely promising HTP tools as they do not require attaching sensors to the animals. In addition, camera sensors can be used for multiple animals and be reused across pens and over time, and

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_138, © J.R.R. Dorea and G.J. M. Rosa 2022

image or videos can be used to monitor many traits simultaneously. Some of the most desired applications of CVS for live animals correspond to animal identification, tracking, and monitoring, ultimately identifying changes in their daily behavior (Fernandes *et al.* 2020a).

Computer vision for livestock systems

Image analysis has been applied to various traits in different livestock species. In dairy cattle, for example, is has been explored for lameness classification, body condition score, and behavior monitoring (Fernandes *et al.*, 2020a). However, most of the studies published in the literature used manual feature extraction, such as animal biometric measurements (e.g. distance from the head to the tail, hip height, spine curvature, etc.), and then fit those features using linear models (Gomes *et al.* 2016) or other simple predictive approaches.

Recently, RGB-D sensors have gained attention due to the incorporation of the depth information to the digital color image. The image data from these sensors have been shown to be more informative for animal identification and for body parts recognition (Fernandes *et al.*, 2020a). The main applications have been on identification of animal locomotion and back posture, body condition score, leg angle and udder depth, and body weight (Nasirahmadi, *et al.*, 2017; Wurtz *et al.*, 2019; Fernandes *et al.*, 2020a; Oliveira *et al.*, 2021) of dairy cows. In addition, there are studies showing the applicability of thermal cameras for evaluation of lameness and mastitis (Hovinen *et al.*, 2008; Alsaaod *et al.*, 2015).

Mining large image datasets generated by computer vision systems is one of the most important challenges to deploy efficient systems in livestock operations. For example, the type and amount of feature extracted from 3D images (Fernandes *et al.*, 2019) and modeling approach (Fernandes *et al.*, 2020b) are important factors that may improve prediction quality in image analyses. Fernandes *et al.* (2020b) and Cominotte *et al.* (2020) evaluated the impact of different combinations of image features on the predictive model quality in pigs and cattle. The authors used features related to biometric body measurements (dorsal area and length, projected body volume, height, and width in 11 points across the dorsal area) and body shape discriminator (Fourier descriptors and eccentricity) and found that models including both shape descriptors and body measures presented the best performance.

Reducing data dimensionality in livestock image datasets

Recently, we implemented autoencoders to reduce data dimension from 2D depth images to 1D signals. The signal was built using two concatenated vectors. The first vector is a line crossing the center of the animal body, which represents the body length, and the second vector is a perpendicular line, crossing the center of animal body, and representing body width (Figure 1).

We explored one of the most straightforward autoencoder architectures to extract features from images. A sequence of fully connected layers composes the architecture: in the encoder, the number of nodes decreases at each layer, until we get to the most inner layer – the code, followed by the decoder, composed by a sequence of layers where the number of nodes increases at each layer symmetrically until we get the same data representation as observed in the input. In the model used in our experiments, we defined the hyperparameters as code size of 20; 3 layers in the encoder with 2,000, 2,000 and 500 nodes respectively, 3 layers in the decoder with 500, 2,000, 2,000 nodes respectively and the mean squared error as loss function. When we compared the predictive performance of algorithms trained with biological (BIO) features (i.e. body volume, area, length, width, eccentricity, and others, see Fernandes *et al.*, 2019) extracted through depth images and 20 autoencoder features (AUTO) from 1D signal, we did not observe important differences in R^2 (BIO=0.81, AUTO=0.78), Root mean squared error of prediction (RMSEP (kg), BIO=6.50, AUTO=6.90), and Concordance Correlation Coefficient (CCC, BIO=0.88, AUTO=0.88). However, the size of the dataset dramatically decreased as the input was reduced from approximately 600

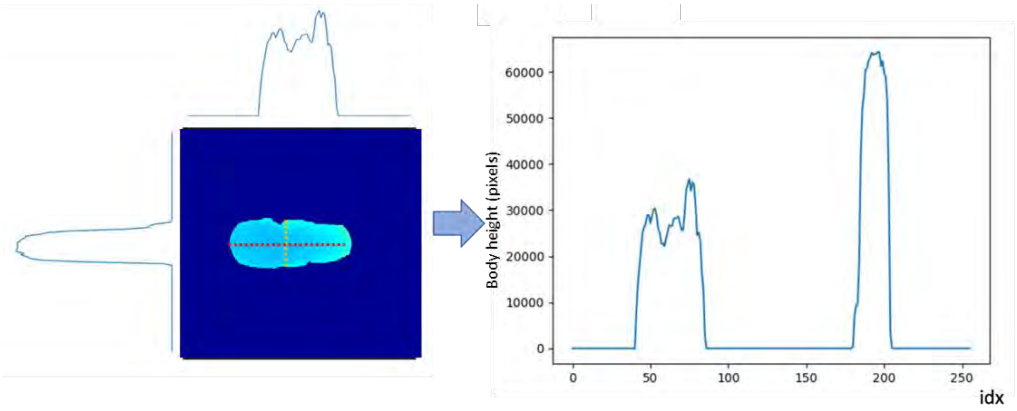


Figure 1. Pre-processing steps: from a calf depth image, vertical and horizontal projections create the 1D signal used for training the autoencoders.

KB (single depth image) to 0.21 KB (single 1D signal). Such example indicates that efficient analytical strategies can be adopted to reduce data dimensionality in large datasets, while keeping satisfactory predictive performance. In livestock farm conditions, where internet speed is generally a critical constraint to implement data collection and analyses in real-time, the development of efficient analytical systems play an important role for large scale deployment of phenotyping technologies, as well as to advance Machine Learning Development and Operations (DevOps or MLOps) in agricultural systems.

To highlight the recent growth and importance of Computer Vision in livestock, four recent review articles (Nasirahmadi, *et al.*, 2017; Wurtz *et al.*, 2019; Fernandes *et al.*, 2020a; Oliveira *et al.*, 2021) explored the potential of computer vision systems for HTP. Wurtz *et al.* (2019) and Nasirahmadi *et al.* (2017) discussed the advances on automated and high-throughput image detection of farm animal behavioral traits, focusing on animal welfare and production. Fernandes *et al.* (2020a) provided an overview of key concepts related to computer vision, image processing, image analyses, and the types of devices and image ranging systems. Fernandes *et al.* (2020a) also provided important key metrics for image-based prediction quality assessment, and a compilation of animal phenotypes used for management purpose (e.g. body condition score, body weight, animal behavior, etc.). An overview on how deep learning has been implemented and how it can be an effective tool to predict animal phenotypes, accelerating the development of predictive modeling for precise management decisions in livestock, was reported by Oliveira *et al.* (2021).

Final considerations

Systematic collection of animal-level phenotypes in large-scale operations is one of the main constraints to advance animal breeding programs and optimal farm management decisions, and HTP technologies are a key factor in this context. Despite all the challenges related to data collection, management, and analyses in livestock systems, the use of HTP technologies, such as computer vision, are crucial to generate cost-effective and reliable precise phenotypes.

References

Alsaaod M., Schaefer, A. L., Büscher, W., & Steiner, A. (2015). The role of infrared thermography as a non-invasive tool for the detection of lameness in cattle. *Sensors*, *15*(6), 14513-14525. https://doi.org/10.3390/s150614513.

Cominotte A., Fernandes A.F., Dorea J. R. R., Rosa G. J. M., Ladeira M. *et al.* (2020). Automated computer vision system to predict body weight and average daily gain in beef cattle during growing and finishing phases. Livest. Sci. 232:103904-. https://doi.org/10.1016/j.livsci.2019.103904.

Fernandes A.F.A, Dorea J.R.R., Rosa G.J.M. (2020a). Image Analysis and Computer Vision Applications in Animal Sciences: An Overview. Front. Vet. Sci. 21;7:551269. https://doi.org/10.3389/fvets.2020.551269.

Fernandes A.F.A., Dórea J.R.R., Fitzgerald R., Herring W., Rosa G.J.M. (2019). A novel automated system to acquire biometric and morphological measurements, and predict body weight of pigs via 3D computer vision. J. Anim. Sci. 97,496-508. https://doi.org/10.1093/jas/sky418

Fernandes A.F., Dorea, J.R.R., Valente, B.D., Fitzgerald, R., Herring, W. *et al.* (2020b). Comparison of data analytics strategies in computer vision systems to predict pig body composition traits from 3D images. J. Anim. Sci. 98 (8), 1–9. https://doi.org/10.1093/jas/skaa250

Gomes R.A., Monteiro G.R., Assis G.J.F., Busato K.C., Ladeira M.M. and Chizzotti M.L. (2016). Technical note: Estimating body weight and body composition of beef cattle trough digital image analysis. J. Anim. Sci. 94:5414–5422. https://doi.org/10.2527/jas.2016-0797

Hovinen M., Siivonen J., Taponen S., Hänninen L., Pastell M., Aisla A., and Pyörälä S. (2008). Detection of Clinical Mastitis with the Help of a Thermal Camera. J. Dairy Sci. 91:4592-4598. https://doi.org/10.3168/jds.2008-1218

LeCun Y., Bengio Y. and Hinton G. (2015). Deep learning. Nature 521:436–444. https://doi.org/10.1038/nature14539

Nasirahmadi A., Edwards S. A., Sturm B. (2017). Implementation of machine vision for detecting behaviour of cattle and pigs. Livest. Sci. https://doi.org/10.1016/j.livsci.2017.05.014.

Oliveira D.A.B., Pereira L.G.R., Bresolin, T., Ferreira R.E.P., Dorea J.R.R. (2021). A review of deep learning algorithms for computer vision systems in livestock. Livest. Sci., 253:104700. https://doi.org/10.1016/j.livsci.2021.104700.

Silva F. F., Morota G., and Rosa G. J. M. (2021). High-throughput phenotyping in the genomic improvement of livestock. Front. Genet. 12: 707343, https://doi.org/10.3389/fgene.2021.707343.

Wurtz K., Camerlink I., D'Eath R.B., Fernandez A.P., Norton T., Steibel J., Siegford J. (2019). Recording behaviour of indoor- housed farm animals automatically using machine vision tech- nology: A systematic review, PLOS ONE. https://doi.org/10.1371/journal.pone.0226669.

139. Tracking multiple cows simultaneously in barns using computer vision and deep learning

C. Kamphuis*, I. Adriaens, W. Ouweltjes and I. Hulsegge

Wageningen University & Research, P.O. Box 338, 6700 AH Wageningen, the Netherlands;
claudia.kamphuis@wur.nl

Abstract

This study investigated the automated tracking of multiple cows simultaneously using computer vision and deep learning. Video clips were collected in 2019 at Dairy Campus, where cows were housed in small groups (n=16). A systematic approach covering the true variability of barn circumstances eventually resulted in the selection of 159 frames that were annotated by drawing bounding boxes around each cow. These frames were used to retrain and test four You Only Look Once version 5 (YOLOv5) models to automatically detect cows. The weights of the best performing YOLOv5 model were used to parametrize the deep learning algorithm DeepSORT to track multiple cows simultaneously. This algorithm was applied to a 10 min timeframe of a randomly selected video clip and evaluated by computing the multi-object tracking accuracy, which was 92.8%. This outcome is a promising and essential step towards automated monitoring of individual behaviour of group-housed cows.

Introduction

Monitoring cow behaviour in research and in practice largely depends on visual observation. In research, experts usually visually observe a limited number of animals, for restricted periods of time and with long time intervals between observations (e.g. Mayo *et al.*, 2019). In practice, farmers visually observe their herd for altered behaviour related to e.g. lameness, with the risk that typically the more severe clinical abnormalities are detected, whereas mild abnormalities often remain unnoticed (Whay *et al.*, 2002). Despite previous efforts to automate continuous monitoring of behaviour, reliable and cost-efficient tools are still lacking. Recent advancements in computer vision and artificial intelligence opens up opportunity to automatically and continuously monitor behaviour of every animal (O'Mahony *et al.*, 2019), in a non-invasive manner and with relatively few sensors.

This study investigated the feasibility of detecting and automatically tracking multiple cows simultaneously with the use of computer vision and deep learning. In doing so, two challenges were addressed: the first was the detection of cows in video footage, and the second was the tracking of all detected cows through time and space. If this detecting and tracking of cows can be done with high accuracy levels, next steps involve the automated identification of cows and quantification of behaviour.

Materials & methods

Data. Data used in this study were collected as part of the Freewalk project (https://www.freewalk.eu/en/freewalk.htm), conducted at the Wageningen research facility Dairy Campus (Leeuwarden, the Netherlands). As part of that project, video footage was collected with four cameras placed in a freestall barn with a synthetic floor and with two cameras in a cubicle barn with a concrete slatted floor. Both barns housed a group of 16 Holstein Friesian cows. All cameras (HikVision DS-2CE16H5T-IT3E) generated footage with 15 frames per second with a resolution of 2,560×1,440 pixels per frame. From the collected video footage, 111 video clips of approximately 60 min each were selected for the current study. This selection included a number of clips recorded in summer (n=16, in July or August) and winter (n=95, in December) of 2019. In systematically selecting the clips we attempted to cover the true variability in barn circumstances in terms of coat patterns of cows, cow behavior (e.g. standing, lying, eating), camera

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_139, © C. Kamphuis *et al.* 2022

viewpoints and lighting conditions. The selected video clips were used to first automatically detect cows, and subsequently to automatically track these detected cows.

Automated detection of cows. The automated detection of cows was done with the latest version (version 5) of the unified, real-time object detection algorithm You Only Look Once (YOLO; Redmon *et al.*, 2016; Ultralytics, 2020). The YOLOv5 algorithm is pretrained with the COCO-dataset (Meel, 2021), a dataset including >200K annotated images of which also images of cows. We retrained four YOLOv5 models (YOLOv5s, -m, -l, and -x), each having different configuration and parameter sizes (Ultralytics, 2020). The YOLOv5s is the simplest model, requiring the least computation time; YOLOv5x is the most complex model, requiring the most computation time. Each of the models was retrained with 200 epochs. Retraining the YOLOv5 models was done with frames from the 111 selected video clips. These frames were selected by first dividing each video clip in 10 equal parts (based on the total number of frames in a clip), from which, in a second step,one frame per part was randomly selected. This resulted in 1,110 frames. From these, 159 frames were selected randomly for annotation. Annotation was done by manually drawing a bounding box (BB) around each cow visible in each frame (Figure 1), using a self-designed annotation tool implemented in Matlab 2018a (the Mathworks Inc., Natick, MA, USA). The 159 frames were subsequently divided to a training (n=127), validation (n=16), and test (n=16) dataset. The test dataset was used to assess the performance of the four YOLOv5 models by comparing the annotated (ground truth) and predicted BBs. Performance indicators were precision, recall, two mean average precision (mAP) indicators, and the F1-score. The mAP is a measure that explains how well the predicted and annotated BBs overlap. The first mAP indicator counts a predicted BB as true positive if the overlap between annotated and predicted BB is at least 50% (mAP@.5). The second mAP indicator is the average overlap from a threshold range (with steps of 5%) from 50-95% (mAP@.5;.95). The YOLOv5-model performing best on the test set was selected to be used for tracking the detected cows.

Automated tracking of cows. The weights of the best performing YOLOv5 model were used to parametrize the deep learning-based tracking algorithm DeepSORT (Brostrom, 2007; Wojke *et al.*, 2017). The performance of this DeepSORT tracking algorithm model in detecting and subsequently tracking multiple cows simultaneously was evaluated with one 10 min video clip. This video clip was randomly selected from the available 111 video clips. From this video clip, a 10 min timeframe was manually selected such that different behaviours, e.g. cows walking in or outside the camera view, cows walking around, cows standing up or lying down were present. The performance was evaluated by manually computing the multi-object tracking accuracy (MOTA; Bernardin and Stiefelhagen, 2008): from the 10 min video clip, the first frame of each second was selected resulting in 601 frames. The DeepSORT algorithm assigned BBs to each frame, and for each of the 601 frames, the number of false positive BBs (FP; predicted BB without an

Figure 1. Examples of annotated frames (left free-stall barn; right cubicle barn), with bounding boxes drawn around each cow visible in the frame.

annotated BB), false negative BBs (FN; annotated BB without a predicted BB), and identity switches (IDS; one annotated BB receives multiple predicted BB) were visually assessed. Each error was counted as one error. The MOTA was computed using formula 1

$$MOTA = (1 - ((\Sigma_t FN_t + FP_t + IDS_t) / (\Sigma_t \text{ annotated } BB_t))) \times 100 \quad (1)$$

Results

The subset of 159 randomly selected frames to retrain and test the YOLOv5 models for cow detection included frames from all the 111 video clips, with a maximum of two frames per videoclip. Table 1 summarizes the performance indicators of the four YOLOv5 models when tested on the frames in the test dataset: the four YOLOv5 models largely overlap in performance indicators. The $YOLOv5_m$ model and $YOLOv5_x$ perform similarly, but the first required less computation time. Therefore, the weights of the $YOLOv5_m$ model were used to parametrize the DeepSORT algorithm.

The selected 10 min timeframe to evaluate the parametrized DeepSORT algorithm had a total of 4,874 annotated BBs (cows), in which the number of BBs (cows) per frame varied between seven and nine. The DeepSORT algorithm applied to the 10 min video clip resulted in 32 FP BBs, 283 FN BBs, and 35 IDSs. These numbers resulted in a MOTA of 92.8%.

Discussion

This study retrained YOLOv5 models to detect cows in practical circumstances. This transfer learning approach has the advantage that it requires much less annotation. Moreover, a systematic annotation approach was applied to capture the true variability in practical circumstances. This resulted in good performing YOLOv5 models with just 143 (out of the 159) annotated frames that were used for retraining, comparable to performance results reported by Tassinari *et al.* (2021) who annotated more data (~25K frames) and applied YOLOv3.

The weights of $YOLOv5_m$ were used to parametrize DeepSORT, resulting in a MOTA of 92.8% when applied to 10 min of a single selected video clip. This is a promising high accuracy, but results have to be confirmed with more video clips that cover more variability in practical barn circumstances. The DeepSORT had low numbers of FP BBs and IDSs but the number of FN BBs was relatively high. A closer look into the data revealed that FN BBs often occurred when cows were overlapping each other in the camera view. In these cases, the algorithm often predicted one BB (one cow) instead of two. The FP BBs occurred often when

Table 1. Performance of four YOLOv5 models on the test set (n=16 frames), with YOLOv5s being the simplest model, and YOLOv5x the most complex model.

	YOLOv5			
Performance indicator (%)	**s**	**m**	**l**	**x**
Precision	95.9	95.9	95.7	97.5
Recall	95.1	96.7	91.0	93.5
mAP@.5[1]	97.5	97.1	95.9	98.0
mAP@.5:.95[2]	60.3	65.4	65.9	68.9
F1-score	95.5	96.3	93.3	95.5

[1] Mean average performance using a threshold of 50%: if the overlap between annotated and predicted bounding boxes is >50% the prediction is counted as a true positive prediction (cow detection).

[2] Mean average performance using a threshold of 50-95%: the average true prediction over a range of thresholds (with steps of 5%) running from 50 to 95% overlap between annotated and predicted bounding boxes.

cows had a 'strange' pose, e.g. when the cow was scratching her head with her hindleg. Both issues may be solved by adding more annotated frames with overlapping cows or unique poses to the training data. The IDSs often occurred when a second cow temporarily occluded the view of another cow. In these cases, the algorithm assigned a new identify to the occluded cow once the second cow passed-by completely. Building-in a memory function such that the DeepSORT algorithm remembers the assigned identities from previous images may help to solve this problem. This building in, however, has still to be developed.

The amount of data used in the current study was limited, and the evaluation of tracking cows was done on a single 10 min video clip only. Still, it confirmed the feasibility of detecting and tracking multiple cows simultaneously with a high level of accuracy (92.8%). This provides perspective to automatically continuously monitor individual group-housed cows. The use of BB to track cows, and the change in BB properties during tracking may be useful to detect specific cow behaviour as demonstrated by Adriaens *et al.* (2022).

Acknowledgements

This research was conducted by Wageningen Livestock Research, commissioned and funded by the Dutch Ministry of Agriculture, Nature and Food Quality, within the framework of Policy Support Research theme 'Data Driven & High Tech' (project number KB-38-001-002).

References

Adriaens I., Ouweltjes W., Hulsegge I., and Kamphuis C. (2022) Proc of the 12th WCGALP, Rotterdam, the Netherlands.

Bernardin K., and Stiefelhagen R. (2008) J. Image Video Process. Available at: https://link.springer.com/content/pdf/10.1155/2008/246309.pdf

Brostrom M. (2007) Yolov5_DeepSort_Pytorch. Available at: https://github.com/mikel-brostrom/Yolov5_DeepSort_Pytorch

Mayo L.M, Silvia W.J., Ray D.L., Jones B.W., Stone A.E. *et al.* (2019) J Dairy Sci 102(3):2645-2656. https://doi.org/10.3168/jds.2018-14738

Meel V. (2021) What is the COCO Dataset? What you need to know in 2021. Available at: https://viso.ai/computer-vision/coco-dataset/

O'Mahony N., Campbell S., Carvalho A., Krpalkova L., Riordan D *et al.* (2019) IFAC-PapersOnline 52(30):312-317. https://doi.org/10.1016/j.ifacol.2019.12.555

Redmon J., Divalla S., Girschick R., and Farhardi A. (2016) Proc of the IEEE conference on computer vision and pattern recognition, Las Vegas, USA.

Tasssinari P., Bovo M., Benni S., Franzoni S., and Poggi M. *et al.* (2021) CompAg 182(2021): 106030. https://doi.org/10.1016.j.compag.2021.106030

Ultralytics (2020) yolov5. Available at https://github.com/ultralytics/yolov5

Whay H.R., Main D.C.J., Green L.E., and Webster A.J.F. (2002) Proc. of the 12th International Symposium on Lameness in Ruminants, Orlando, USA.

Wojke N., Bewley A., and Paulus D (2017) Proc. of IEEE International Conference on Image Processing, Beijing, China.

140. A computer vision system for individual tracking of group housed rabbits

J.P. Sánchez[1*], I. Muñoz[2], O. González[1], M. Pascual[1], O. Perucho[3], P. Alsina[1] and M. Piles[1]

[1]Programa de Genética y Mejora Animal, IRTA, 08140 Caldes de Montbui, Spain; [2]Programa de Calidad y Tecnología Alimentaria, IRTA, 17121 Monells, Spain; [3]Programa de Infraestructuras Experimentales del Área de Producción Animal, IRTA, 08140 Caldes de Montbui, Spain; juanpablo.sanchez@irta.es

Abstract

We are developing a computer vision system for tracking growing rabbits' movements, in order to quantify individual variability on physical activity. The system relies on a Deep Convolutional Neural Network for the identification of the animals; and on the Hungarian and the Kalman Filter algorithms for tracking the animals across photograms. When the animals were sparsely distributed a precise identification was achieved with few images (mean Average Precision (mAP)≥0.95). When the rabbits were crowded within a given region, the quality of the identification dropped, but by increasing the size of the training set and by applying data augmentation techniques, we managed to improve the identification accuracy (mAP=0.83). In spite of these improvements, and the fact that good results were observed during some short tracking examples, the overall performance at tracking is still low and further training rounds are needed to properly define the aimed physical activity metrics.

Introduction

Our research group has been working on a line related to the genetic improvement of feed efficiency during fattening, aiming to propose genetic evaluation methods and tools to increase the amount of meat produced per kilo of feed supplied. In this context, we have designed and manufactured a device allowing for the individual registration of feed consumption in rabbit (Sánchez *et al.*, 2018). Information on consumption is essential for the implementation of genetic selection programs for the direct improvement of feed efficiency in group-housed animals.

An important percentage of the resources used by the growing animal is spent on physical activity. In pigs, it has been shown that selection for reducing residual feed intake (the amount of feed consumed above or below the needs of an animal) has led to less active animals (Gilbert *et al.*, 2019). To monitor the movements of the rabbits we have begun the development of a computer vision system using artificial intelligence techniques that will allow to quantify their physical activity. The ultimate goal of our development would be to record this information on physical activity for being considered in the selection processes for feed efficiency we are conducting (Pascual *et al.*, 2021). Once this computer vision system is ready, we also expect to train it for being used in other fields of animal production: behavior studies, interactions between group mates or early detection of pathological outbreaks.

Our purpose with this communication is to inform the animal breeding community about the development of this new computer vision system. We describe the methods that it implements and present some results related to its capacity of individual detection and tracking.

Materials & methods

Algorithms for artificial vision and individualized tracking of rabbits. The first step to achieve an individualized tracking of objects in a video signal is to have a good localization system for each of the objects that we want to track, in our case the rabbits. The problem of locating one or more objects of a certain class (rabbit) in an image is known, in the computer vision context, as instance segmentation. It

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_140, © J.P. Sánchez *et al.* 2022

consists in assigning each pixel of an image to one of the different classes we might have defined, and grouping together the nearby pixels belonging to the same class to define the different objects (instances) of that class in the image. In this way, it is possible to count the items belonging to each class in a single photogram or to track them across photograms. One of the most advanced and efficient tools to perform instance segmentation is the MaskR-CNN library (He *et al.*, 2018), which is based on deep convolutional neural networks (ResNet101). The result of applying MaskR-CNN to an image will be a table with as many rows as objects detected in the image. Each row includes the coordinates of the smallest rectangle that can be defined to cover the pixels corresponding to each detected object and the object class.

To be able to detect the rabbits, it is necessary to train the MaskR-CNN system. This training consists in obtaining the ResNet101 network parameters that allow identifying the rabbits in any image. This is done by cross-validation that consists in providing images in which the rabbits have been manually identified (training set). With this information, the values of the parameters of the ResNet101 network are iteratively adjusted.

Deep layers of the ResNet101 network are useful for the detection of general patterns, common to any type of image: angles, shadows, curvatures, etc. Thus, the coefficients of the deep layers of the ResNet101 network remained equal to the values previously trained within COCO project (Lin *et al.*, 2014), that aimed at detecting all types of common objects. The coefficients from this generalist training project will not only be valid for our specific problem, but will be probably better than those that we could obtain with our images. In COCO project hundreds of thousands of images have been used for training, whereas we will only have several hundred or a few thousand images with the identified rabbits. The outer layers of the ResNet101 network, those in charge of classifying the pixels into the defined classes, are the only ones specifically trained with the rabbit annotated images in the training set.

The detected rabbits were tracked across photograms using a library named Simple Online and Real-Time Tracking (SORT) (Bewley *et al.*, 2016); which implements two algorithms for object tracking: the Hungarian algorithm and the Kalman Filter. This library has been used in a similar project aiming to track pigs (Van der Zande *et al.*, 2021).

Collection of images: training and identification of the rabbits. The training of the ResNet101 network has been carried out using photos taken with a 2D RGB high resolution camera (3,280×2,464 pixels), placed on the ceiling of the barn at about 2 meters above the rabbits. The camera was positioned to record the area corresponding to 4 cages of 75×75 cm each. Each cage housed 6 rabbits.

Each of the images used for training had a size of 1,280×1,280 pixels, covering one single cage. The images for training were collected during 4 production batches, and within each of them, the images were taken on a few days chosen at random from the fattening period (35 to 56 days of rabbit's life). In the first two batches, the images were chosen so that the rabbits could be perfectly distinguished from each other (Figure 1A), while in the last two batches images corresponded to rabbits stacked next to each other, making them difficult to identify (Figure 1B). In addition, with the images from the last production batch, different data augmentation techniques were used, i.e. the images were subject to random twists and distortions, which allows increasing the available information without the need to carry out new manual identifications. For each training, a number of annotated images were reserved to validate the precision of the identifications.

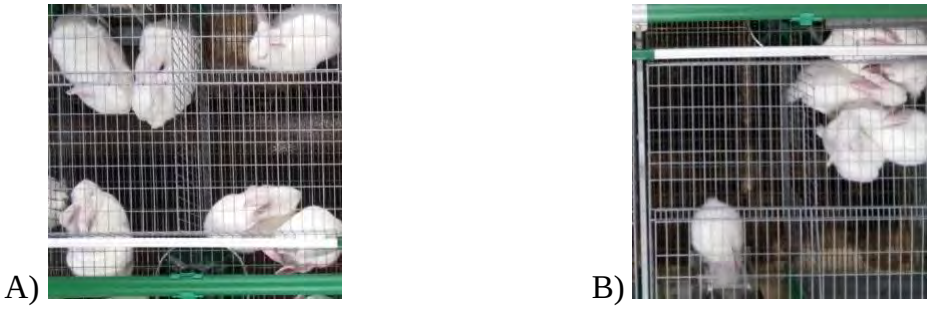


Figure 1. Examples of images: animals are easily differentiated (A) or animals are huddled on top of each other (B).

Results

The identification of the animals with a sparse distribution showed a high accuracy. With only a few training images a high precision of the location of the animals in the validation images was obtained (mean Average Precision (mAP)≥0.95) (Table 1). As expected, the use of images in which some individuals are huddled on top of others makes the identification notoriously difficult. It must be taken into account that in some of the images used it was difficult, even for the human eye, to determine where one individual ended and the next began. However, by increasing the number of training images and by including complex data augmentation procedures significant improvements were achieved.

Regarding tracking rabbits across consecutive photograms we have conducted two assessments. One has been done using standard central processing units (CPUs). The following link shows a video with the tracking results: https://www.youtube.com/watch?v=LVPEWSoh-zk. It can be seen how the identifications assigned in the first frame remain the same (in most of the cases) throughout the entire video, even when an animal passes by (or even over) some other cage mate. The other assessment has been done using one graphics processing unit (GPU). With this alternative system, it is possible to process a single-cage photogram every two seconds, i.e. a real-time processing is feasible. MaskR-CNN library does not support processing multiple cages at the time using a single GPU; thus, our strategy has been to process every cage, eight in total, for ten minutes during approximately 5.5 days; thus, a total of 777 10-minutes tracks were recorded. Figure 2 shows a graphical representation of the movements in three of these 10-minutes tracks. Panels A and B are two examples of good tracking performances; panel C, on the contrary, is an example of poor individual tracking.

Table 1. Performance of different trainings steps of the MaskR-CNN system.

Training set size	Validation set size	# Epochs[1]	# Steps / epochs[1]	Image type[2]	Data augmentation	mAP[3]
50	37	20	25	A	NO	0.95
156	44	20	100	A	NO	0.96
176	24	20	100	B	NO	0.69
331	24	10	50	B	Yes (Vertical)	0.76
331	24	20	100	B	Yes (45°)	0.76
331	24	30	150	B	Yes (Complex)	0.83

[1] At the end of each epoch all the training images would have been processed.

[2] See Figure 1.

[3] mAP=mean Average Precision in the validation set at the end of the training. It can be interpreted as a correlation between the class of the pixels in the annotated image used for validation and the class assigned by the algorithm.

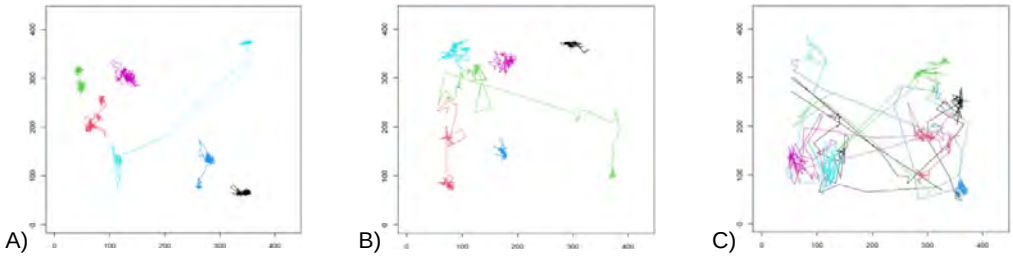


Figure 2. Graphical representation of three cage surfaces with the individual tracks (10 minutes) of the six rabbits in each cage. Each colour represents one animal.

Discussion

The preliminary results we present seem to be promising for achieving our objective, i.e. automatically monitoring individual activity. The key element for achieving such objective would be the real-time processing already accomplished, which would grant keeping animal identifications for long periods. The observed issues on limited accuracy (mAP=0.83) during the identification of the animals can be solved with additional training rounds. Similarly to Van der Zande *et al.* (2021) the performance of our system strongly depends on the quality of the identification of the individuals. Our ambition is to extend this system to growing or finishing pigs. In this species, we expect the animal recognition training to be easier since they do not pile up to each other as frequently as in rabbits (Figure 1 panel B). Another point yet to be solved is the definition of the metrics to quantify individual activity; we could not advance in this regard yet since better individual recognition is needed. Obtaining these metrics on the current tracks would be strongly influenced by the observed misidentifications (Figure 2C).

Acknowledgements

This research has been funded by GENEF2 project (RTI2018-097610-R-I00).

References

Bewley A., Ge Z., Ott L., Ramos F., Upcroft B. (2017) Available at: https://arxiv.org/pdf/1602.00763.pdf

Gilbert H., Billon Y., Brossard L., Faure J., Gatellier P. *et al.* (2019) Animal 11(9):1427-1439

He K., Gkioxari G., Dollár P., Girshick R. (2018). Available at: https://arxiv.org/pdf/1703.06870v3.pdf

Lin T.Y., Maire M., Belongie S., Hays J., Perona P., *et al.* (2014) Microsoft COCO: Common Objects in Context. In: Fleet D., Pajdla T., Schiele B., Tuytelaars T. (eds) Computer Vision – ECCV 2014. ECCV 2014. Lecture Notes in Computer Science, vol 8693. Springer, Cham.

Pascual M., Piles. M., Perucho O., Alsina P., Sánchez J.P. (2021) Proc. of Symposium de Cunicultura de ASESCU. Virtual

Sánchez J.P., Piles M., Pascual M., Rafel O. (2018) Proc. of the XLIII Symposium de Cunicultura de ASESCU. Calamocha, Spain, pp. 125-129.

Van der Zande L. E., Guzhva O., Rodenburg T. B. (2021) Frontiers in Animal Science 2-202110.

141. Clustering of Meishan and Large White sows for activity defined from image analysis, and influence on piglet mortality

O. Girardie[1*], M. Bonneau[2], Y. Billon[3], J. Bailly[3], I. David[1], L. Canario[1]

[1]*UMR1388 GenPhySE, INRAE/INPT ENSAT/INPT ENVT, 31326 Castanet, Tolosan, France;*
[2]*UR143 URZ Animal Production Research, INRAE, Paris, France,* [3]*UE GenESI, INRAE, 17700 Surgères, France;*
oceane.girardie@inrae.fr

Abstract

A computer-vision based procedure was developed to predict behaviour of 22 Meishan and 21 Large-White lactating sows kept in individual pens. A neural network analysis allowed detecting 5 postures (standing, sitting, lying ventrally, lying laterally without and with udder exposed) and 2 sow standing activities, i.e. drinking, eating. The accuracy was 31% for sitting and varied from 78 to 96% for the standing and lying postures. Sow behaviour was estimated from 3 days before to 9 days after farrowing. First, cluster analyses were carried out to identify patterns of activities over successive periods: pre-farrowing, the day of farrowing and post-farrowing. Second, a survival analysis was carried out using Cox proportional hazards model to study the influence of sow behaviour on piglet mortality. Piglets born from sows that were less active before farrowing and make more rooling and less changing standing to lying on the day of farrowing had a risk of dying up to 8 and 3 to 7 times higher than other piglets, respectively.

Introduction

More freedom is required for lactating sows. However, it is often reported that loose-housed sows may have higher piglet mortality than sows in crate but conclusions obtained from a few animals varied between studies (Baxter *et al.*, 2012). In addition, the analysis of relationships between sow behaviour and litter mortality is limited to short time periods and a few variables when it depends on human video observations. The recent and successful use of computer vision to analyse animal behaviour allowed addressing this issue with automated phenotyping over extended periods on large numbers of animals. In particular, methods based on convolutional neural networks (CNN) were recently used to estimate the posture and activity of a sow in an individual pen with its piglets around (Yang and Xiao, 2020). The first objective of our study was to use a CNN to predict sow postural activity from video recording. From the data produced, the second objective was to study differences between sows in time-budget activity (accounting for time elapsed from farrowing), and to identify sow behavioural traits related to piglet mortality.

Materials & methods

Animals and measurements. The study included 21 Large-White (LW) and 22 Meishan (MS) primiparous sows that were raised at the INRAE GenESI experimental farm (https://doi.org/10.15454/1.5572415481185847E12) in 5 batches from November 2010 to January 2011. Sows were moved to the farrowing unit about 1 week before farrowing. The pen size was 7.3 m^2 (286×255 cm) and included a creep area with a heat lamp. The concrete floor was covered with a thin layer of straw that was changed every morning. Sows were fed twice daily, at approximately 8am and 4pm and they had free access to water provided from a low-pressure water trough located on the wall opposite to the feeding trough. LW and MS sows were located alternately in adjacent pens. Farrowing was not induced, and interventions to save piglets were highly limited. Piglet death was recorded daily and causes of death were identified distinguishing stillborn piglets, piglets crushed by the sow, piglet that died from starvation and other causes.

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_141, © O. Girardie *et al.* 2022

Video recording and algorithm. A 2D camera was mounted on the ceiling above each pen to cover the entire surface of the pen. Sows were video recorded continuously 24h a day from 3 days before farrowing, to 9 days after farrowing. An ethogram was defined to analyse sow behaviour. It included six sow postures: sitting (A), standing (S), lying on the belly (LB), lying on the side (L), lying on the side with udder exposed (LU) and kneeling (K); and three sow activities while standing: drinking (D), eating (E) or other (N). From the videos, 8,400 images were selected randomly from all the sows (LW and MS) and the entire recording period. The images were manually labelled according to the ethogram. The location of the sow and piglets were also recorded. These images were used to train a YOLO detector (Redmon *et al.*, 2016), with ResNet50 as a feature extractor. YOLO was trained with 6 object classes, corresponding to the different postures. Thus, it was used to predict animal location and posture at the same time. The sow's centre of gravity was the estimated centre of gravity of the YOLO detection. The distance between sow head and each trough was calculated to detect drinking and eating activities. YOLO was finally used to predict sow posture and activity on 25,830 images, again selected randomly and manually annotated. The accuracy of prediction was 31% for A, 88% for S, 9% for K, 95% for LB, 96% for LU and 78% for L. The K and S postures were then grouped together for further analysis as they were highly confounded. YOLO training and predictions were done using Matlab (R2020b).

Statistical analyses. Sow activity was summarised per day, by % of time spent in each posture over 24h and % of time spent in each activity when the sow is standing. We considered 3 periods: the pre-farrowing period from D-3 to D-1, the farrowing day D0, defined as 24h after the birth of the first piglet, and the post-farrowing period from D1 to D9. A k-means clustering approach was carried out to identified posture and standing activity longitudinal patterns for the 3 periods (Genolini *et al.*, 2015). Obtained clusters represented different patterns of use of postures and standing activity. A survival analysis was carried out to evaluate the influence of sow posture (cluster posture) and standing activity (cluster standing activity; found during clusterisation procedure) on piglet survival. Deaths, stillborn excluded, were considered events. Piglets alive at weaning were censored. Survival time or time until censoring were defined from the farrowing day. First, survival probabilities were calculated by piglet genetic type using the Kaplan-Meier (KM) approach. The Cox proportional hazard model was used for each period to evaluate if cluster posture, cluster standing activity or number of postural changes influenced piglets mortality. In addition to sow clusters or number of postural changes (3 levels: I = below average, M = around average, S = upper average) effects included in each cox models were the piglet genetic type (4 levels), batch (5 levels), litter size (3 levels; defined as the total number of piglets born), and box (3 levels; defined by distance with maternity entrance). All statistical analyses were carried out using R software (version R 4.1.0).

Results

Clusters identified for postures and standing activity in each period are shown in Figure 1 and 2, respectively. Before farrowing (bf), clusters according to posture differed mainly according to the percentage of time spent lying on the side. The two clusters are defined according to the sow breed: A_{bf} included 17 LW sows and B_{bf} 11 MS sows.

At farrowing (af), posture clusters differed mainly according to the time spent lying (3 clusters: A_{af} including 5 LW, B_{af} 6 LW + 2 MS, and C_{af} 8 + 9), and after farrowing (pf) by the time spent lying on the belly (A_{pf} :10+16; B_{pf} :10+1). In the 3 periods under study, the standing activity clusters differed mainly by the percentage of time spent in other activity than drinking and eating.

Two, three and two clusters were obtained, respectively: Aa_{bf} (6+10) and Ba_{bf} (11+1) before farrowing, Aa_{af} (6+5), Ba_{af} (3+4) and Ca_{af} (10+2) at farrowing, and finally Aa_{pf} (19+6) and Ba_{pf} (1+12) after farrowing. Based on KM estimator, piglet survival differed markedly between the 4 piglet genetic types (P<0.0001; Figure 2a).

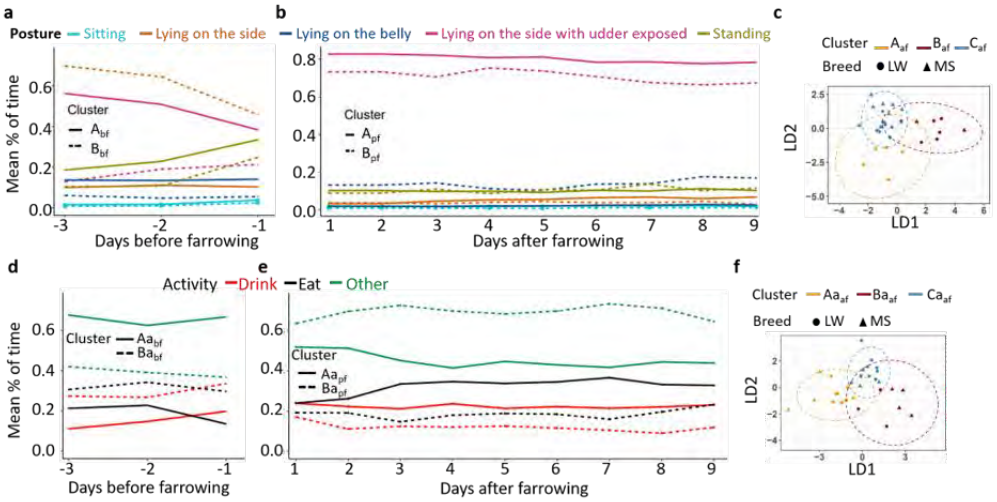


Figure 1. Clusters before farrowing, for sow postures (a) and for sow standing activity (d). Cluster after farrowing, for sow posture (b) and for sow standing activity (e). Cluster at farrowing obtained from discriminant analyses for sow posture (c) and sow standing activity (f).

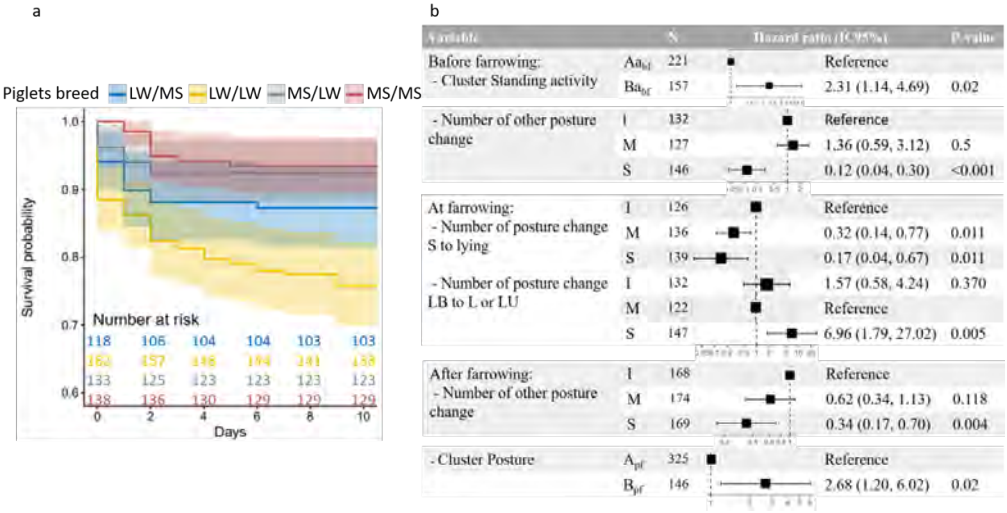


Figure 2. (a) Kaplan-Meier curves of piglet mortality after farrowing in Large White/ Large White (LW/LW), Meishan/Meishan (MS/MS), LW/MS and MS/LW genetic types of piglets; (b) Hazard ratios from the Cox models.

Piglet postnatal mortality was related significantly to sow standing activity before farrowing. Piglets from sows spending more time eating and drinking (in cluster Ba_{bf}) had a 2.3 times greater instantaneous risk of dying than piglets from sows in cluster Aa_{bf} ($P<0.02$). Piglets born from sows calmer before farrowing (I group) had an instantaneous risk of dying 8 times higher than other piglets. Piglets born from sows which made more rooling (change posture LB to L or LU) on the day of farrowing (S group) had an instantaneous risk of dying 7 times greater than other piglets. Conversely, if sow change less standing to lying (I group) on the day of farrowing, the piglets had an instantaneous risk of dying between 3 and 6 times greater

than other piglets (Figure 2b). In early lactation sow posture influenced piglet mortality; piglets from sows spending more time lying ventrally (B_{pf}) had an instantaneous risk of dying 3 times higher than piglets from cluster A_{pf} (P=0.02). Furthermore, piglets with sows performing fewer postural changes (I group) had a 2.5 higher risk of dying than other piglets (Figure 2b).

Discussion

The contrast between the two sow breeds in piglet mortality was confirmed. We studied how much maternal behaviour was involved in such difference. Clusters according to postures pre-farrowing were based solely on sow breed differences. A previous study showed that tethered MS sows spent twice less time lying on the belly before farrowing, and more time rooting especially 24h before farrowing (Meunier-Salaün *et al.*, 1991). In cluster Aa_{bf}, which was composed of 10 MS sows, 64% of standing time was allocated to other activity than eating and drinking. Presumably, it refers to nest-building which is positively correlated to piglet survival (Andersen *et al.*, 2005). All other clusters were a mix of sows from the 2 breeds. At farrowing, the risk of dying was higher in sows that performed more posture changes. Previous studies (e.g. Yun *et al.*, 2019) found that numerous posture changes during the first 24h after farrowing is positively correlated with crushing of piglets by the sow.

To conclude sow pattern of activity over the 3 studied periods influenced the risk of mortality of born alive piglets.

References

Andersen, I.L., Berg, S., Bøe, K.E. (2005) Appl. Anim. Behav. Sci. 93: 229–243. https://doi.org/10.1016/j.applanim.2004.11.015

Baxter, E.M., Lawrence, A.B., Edwards, S.A. (2012) Animal 6: 96–117. https://doi.org/10.1017/S1751731111001224

Genolini, C., Alacoque, X., Sentenac, M., Arnaud, C. (2015) J. Stat. Soft. 65. https://doi.org/10.18637/jss.v065.i04

Meunier-Salaün, M.C., Gort, F., Prunier, A., Schouten, W.P.G. (1991) Appl. Anim. Behav. Sci. 31: 43–59. https://doi.org/10.1016/0168-1591(91)90152-N

Redmon, J., Divvala, S., Girshick, R., Farhadi, A. (2016). Proc of IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, USA. https://doi.org/10.1109/CVPR.2016.91

Yang, Q., Xiao, D. (2020). A review of video-based pig behavior recognition. Appl.Anim. Behav. Sci. 233, 105146. https://doi.org/10.1016/j.applanim.2020.105146

Yun, J., Han, T., Björkman, S., Nystén, M., Hasan, S., *et al.* (2019) Animal 13: 1045–1053. https://doi.org/10.1017/S1751731118002549

142. Using machine learning to predict feed intakes of meat sheep from animal traits and ruminal microbiota

Q. Le Graverand[1*], C. Marie-Etancelin[1], J.L. Weisbecker[1], A. Meynadier[1], D. Marcon[2] and F. Tortereau[1]

[1]GenPhySE, Université de Toulouse, INRAE, ENVT, 24 Chemin de Borde-Rouge – Auzeville CS 52627, 31326 Castanet-Tolosan, France; [2]INRAE, Experimental Unit P3R, Domaine de la Sapinière, 18390 Osmoy, France; quentin.le-graverand@inrae.fr

Abstract

Animal traits, such as body weights, and rumen microbiota composition have been proposed as feed intake predictors. The present study assessed what are the best predictors out of animal traits, metabarcoding data or a combination of both. Predictions were carried out with sparse Partial Least Squares Regression (sPLSR), Support Vector Regression (SVR) and Random Forest Regression (RFR). With all three approaches, best feed intake predictions were obtained with animal traits only. The generalizability of models to animals of an independent year was assessed: negative (<-0.1) to high (>0.8) correlations between actual and predicted feed intakes were obtained. Finally, estimated breeding values (EBVs) were computed for actual and predicted feed intakes. These EBVs were highly correlated (>0.9) depending on the prediction approach. It mainly varied with proportions of true and predicted feed intakes used during the genetic evaluation.

Introduction

To assess feed efficiency of sheep, expensive feed intake recording is necessary. Thus, most meat sheep breeding companies cannot afford to select for feed efficiency. Predicting feed intake or feed efficiency could be one solution. As reviewed by Pittroff and Kothmann (2001), many sheep feed intake prediction models include live weight. Omics data were also used as predictors, such as 16S data to predict sheep Residual Feed Intake (RFI) with a general linear model (Ellison *et al.*, 2019), or rabbit feed intakes with mixed models and SPLSR (Velasco-Galilea *et al.*, 2021).

We have chosen to focus on the prediction of feed intake since it intervenes in the definition of many complex phenotypes such as feed efficiency or the environmental footprint. Moreover, to our knowledge, there is no comparison between accuracies of feed intake predictions from animal traits or microbiota compositions with a large number of sheep and machine learning.

The first goal of our study is to check whether microbial information improves feed intake predictions with different machine learning approaches. To do so, we compare accuracies of different machine learning approaches (sPLSR, SVR and RFR) applied to several sets of predictors (body weight, body composition traits, and/or ruminal microbiota composition). Then, the second objective is to see if a machine learning model could be generalized to new animals raised a different year. Finally, the last objective is to assess if a genetic selection with machine learning predictions is feasible.

Materials & methods

Animal husbandry and traits. Overall 277 Romane male lambs were reared at the INRAE Experimental Unit P3R between 2018 and 2020. Animals originated from two experimental divergent lines selected for a decreased or increased RFI. Lambs are part of the second generation of selection in 2018 (n=103) and the third generation in 2019 (n=101) or 2020 (n=73). The divergence represented 1.9 genetic standard deviations during the third generation. After weaning, lambs were accustomed to an *ad libitum* diet with low-energy concentrates. When animals were approximately 3 months old, feed intakes were recorded

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_142, © Q. Le Graverand *et al.* 2022

during six weeks. Average Daily Feed Intake (ADFI) was computed as the daily feed intake mean over six weeks. At the beginning and end of the trial period, body weights were recorded, then average daily weight gain and growth capacity (estimated body weight at 145 days old) were calculated. Back ultrasound measurements enabled to estimate the back-fat thickness and muscle depth at the trial end.

Ruminal microbiota composition. At the end of the six-weeks trial, ruminal fluid was collected with a medical gastric tube and DNA was extracted from samples. The V4-V5 region of 16S rRNA gene was amplified and then sequenced at the Genomic and Transcriptomic Platform (INRAE, Toulouse, France). Sequences were processed, cleaned and clustered into Operational Taxonomic Units (OTUs) with FROGS tools and pipeline (Escudié *et al.*, 2018). OTUs were pre-filtered according to their relative abundance: only OTUs representing at least 0.005% of all sequences were kept.

OTUs zero counts were imputed with Bayes-Laplace multiplicative replacement (Martín-Fernández *et al.*, 2015) before applying the Centered LogRatio (CLR) transformation. CLR coordinates were filtered to retain OTUs with less than 90% imputed values each year. Finally, the 496 retained OTUs coordinates were adjusted for the age, the year, the pen and sequencing plate effects with a robust MM linear regression (Maechler *et al.*, 2017).

Machine learning. Three different approaches of machine learning were carried out to predict ADFI: sparse Partial Least Squares (sPLSR, mixOmics R package), support vector (SVR, e1071 R package) and random forest (RFR, randomForest R package) regressions. Different sets of prediction features were considered: animal traits (body weights, growth traits and body composition), or adjusted 16S CLR coordinates. Those feature sets were used separately or were combined together by concatenation for ADFI predictions. Successively, animals of one year were used as a testing set. To avoid a bias, 2018 will not be used as a testing set because 2020 lambs are the progeny of 2018 individuals. Animals of the two remaining years were used as training/validation sets: 10-fold cross-validations repeated 50 times were used to tune the hyperparameters and maximize the coefficient of determination (R^2) between predicted and actual ADFI phenotypes. Models were then fitted without the testing dataset. Next, ADFI of testing set animals were predicted with each predictor set and machine learning model. Pearson correlations between predicted and actual ADFI phenotypes were computed. Within one testing set, significance of differences between machine learning approaches and features sets were tested with Dunn and Clark's z test (Dunn and Clark, 1969). *P*-values were adjusted with Bonferroni's procedure.

Breeding values estimation. Breeding values were estimated with PEST (Groeneveld *et al.*, 1990), considering a feed intake heritability of 0.28 (Tortereau *et al.*, 2020). Two sets of populations were used to estimate breeding values: an entire Romane population named E (born from 2009 to 2020), with 6,419 animals in the pedigree including 1,900 with ADFI records; one subset population named S (2018 to 2020), with 4,102 animals in the pedigree including 277 with records. The model included the fixed effects of year, pen, early life traits (litter size, suckling method), sex and body weight as a covariate. EBVs of actual ADFI and EBVs of predicted phenotypes are estimated. EBVs were only estimated for phenotypic ADFI predicted with accurate strategies (R^2>0.7).

Results

Comparison of different predictors and machine learning approaches for phenotypic ADFI. Table 1 details accuracies of ADFI predictions for the three machine learning models carried out with 16S data, animal traits and a concatenation of both. Considering correlations between actual and predicted ADFI phenotypes, there was no significant difference between sPLSR, SVR and RFR accuracies whatever the testing set or the predictors.

Table 1. Pearson correlations between predicted and actual ADFI phenotypes of testing set animals.

		Testing sets	
Models	**Features**	**2019**	**2020**
sPLSR	16S	-0.116[a]	0.191[a]
	Animal	0.763[b]	0.810[b]
	Animal+16S	0.726[b]	0.818[b]
SVR	16S	-0.040[a]	0.351[a]
	Animal	0.766[b]	0.802[b]
	Animal+16S	0.763[b]	0.817[b]
RFR	16S	0.031[a]	0.312[a]
	Animal	0.777[b]	0.773[b]
	Animal+16S	0.709[b]	0.737[b]

[a,b] Within one testing set, correlations with no common letter significantly differ (adjusted $P<0.05$, Dunn and Clark's z test).

With all testing sets and machine learning techniques, correlations were significantly lower when only 16S data was used as the feature set. Whatever the testing set and the machine learning model, combining animal traits and 16S data together as predictors did not significantly increased correlations.

Relationship between EBVs of predicted and actual ADFI. The quality of predicted ADFI EBVs is presented in Table 2. Within one machine learning approach and regardless of the testing set, correlations between EBVs are significantly higher when a full set of records (E) is used during the genetic evaluation compared to a partial set (S).

In 2020, no significant difference could be found between machine learning approaches. However, when only animal traits were used as 2019 ADFI predictors, sPLSR performed significantly better than SVR, with RFR having intermediate performances. Finally, the combination of animal traits and 16S data as predictors did not significantly improve correlations between actual and predicted ADFI EBVs.

Discussion

Rumen microbiota composition alone lead to poorly accurate ADFI predictions with sPLSR, SVR and RFR, which was also observed with sPLSR carried out with rabbit caecal microbiota compositions (Velasco-Galilea *et al.*, 2018). Predictions made with animal traits reached higher performances, whatever the

Table 2. Pearson correlations between predicted and actual ADFI estimated breeding values of testing set animals.

		Testing sets and populations[1]			
		2019		**2020**	
Models	**Features**	**S**	**E**	**S**	**E**
sPLSR	Animal	0.681[a]	0.868[c]	0.843[a]	0.954[b]
	Animal+16S	0.698[ab]	0.876[cd]	0.852[a]	0.956[b]
SVR	Animal	0.624[b]	0.814[d]	0.848[a]	0.962[b]
	Animal+16S	0.650[ab]	0.817[cd]	0.839[a]	0.956[b]
RFR	Animal	0.731[ab]	0.891[cd]	0.863[a]	0.952[b]
	Animal+16S	0.698[ab]	0.880[cd]	0.848[a]	0.959[b]

[1] Populations differ in the number of records used for EBVs estimation. S: subset population; E: entire population.

[a,b] Within one testing set (2019 or 2020), correlations with no common letter significantly differ (adjusted $P<0.05$, Dunn and Clark's z test).

machine learning model. The integration of animal traits and microbiota compositions did not improve predictions. Therefore, our results do not advocate for the use of metabarcoding data to predict feed intake. However, the microbiota composition could be more relevant when one is trying to predict feed efficiency. Ellison *et al.* (2019) managed to predict RFI of 20 animals with 16S data and achieved a correlation of 0.71 between actual and predicted feed efficiency. In the future, another comparison will be done with RFI to assess the potential of 16S and 18S data predictors and determine the relevance of archaeal, bacterial, fungal and protozoal compositions.

Depending on testing sets and predictors, no consistent difference was found between the three modelling approaches when it comes to predict ADFI and then estimate its EBVs. Only SVR performances seemed to be less stable across testing sets when predicted ADFI EBVs were compared to actual ADFI breeding values.

The moderate or high correlations achieved between EBVs of predicted and actual ADFI suggests that predictions could be used for genetic selection of feed intake. Further research in the phenotyping strategy should be carried out, especially about the ratio between predicted and recorded phenotypes. In the present study only two testing sets were used to assess performances of machine learning predictions. Those testing sets were constituted as independent cohorts, i.e. with animals raised a different year than the training set. When predictors are sensitive to environmental factors, such as microbiota compositions, the presence of contemporaneous animals in both training and testing sets should improve predictions.

Funding

This work received funding from the French funding agency (ANR) (GrassToGas project – ERA-GAS no. 39413) and from the European Union's Horizon 2020 research and innovation program under the Grant Agreement No. 772787 (SMARTER).

References

Dunn O. J., and Clark V. (1969). J. Am. Stat. Assoc. 64(325):366-377.

Ellison M. J., Conant G. C., Lamberson W. R., Austin K. J., Van Kirk E., *et al.* (2019). J. Anim. Sci. 97(7):2878–2888. https://doi.org/10.1093/jas/skz170

Escudié F., Auer L., Bernard M., Mariadassou M., Cauquil L., *et al.* (2018). Bioinformatics 34(8):1287–1294. https://doi.org/10.1093/bioinformatics/btx791t

Groeneveld E., Kovač M., and Wang T. (1990). In: World Congress on Genetics Applied to Livestock Production. Edinburg, UK, pp. 488–491.

Maechler M., Rousseeuw P., Croux C., Todorov V., Ruckstuhl A., *et al.* (2017). Robustbase: basic robust statistics R package.

Martín-Fernández J. A., Hron K., Templ M., Filzmoser P., and Palarea-Albaladejo J. (2015). Stat. Modelling. 15(2):134–158. https://doi.org/10.1177/1471082X14535524

Pittroff W., and Kothmann M. M. (2001). Livest. Prod. Sci. 71(2-3):131–150. https://doi.org/10.1016/S0301-6226(01)00218-4

Tortereau F., Marie-Etancelin C., Weisbecker J. L., Marcon D., *et al.* (2020). Animal 14(4):681-687. https://doi.org/10.1017/S1751731119002544

Velasco-Galilea M., Piles M., Ramayo-Caldas Y., and Sánchez J. P. (2021). Sci. Rep. 11(1):1–18. https://doi.org/10.1038/s41598-021-99028-y

143. The magic of predicting shapes from DNA

M. Pérez-Enciso[1*] and L.M. Zingaretti[2]

[1]ICREA, 08010 Barcelona, Spain; [2]Universidad Nacional de Villa María, 5900 Córdoba Argentina; miguel.perez@uab.es

Abstract

The generation of realistic images from marker information could be a main contribution of artificial intelligence to genetics. This paper is a proof of concept demonstrating its feasibility, since shapes and colors are heritable. We generated 2D and 3D images and 2D Cucurbitaceae shapes –which harbor the most extreme shape variation in the plant kingdom- assuming they are controlled by a set of QTLs. We use a class of deep models, called encoders, to reproduce the expected shapes given a linear vector of QTL and random SNPs. This novel theoretical framework can have an important impact in breeding, as studies so far focused on determining the genetics of shape related statistics, and not on predicting shape per se.

Introduction

Shape and color patterns of animals and plant fruits are not only aesthetic features; they convey essential information on animal welfare or fruit quality and are critical for consumer appreciation. Animal and fruit shapes are heritable. In fact, heritability of some phenotypes like human face features approaches one. It should be possible then to predict, i.e. to generate, shapes and appearance from DNA sequence. However, research so far show has focused on identifying quantitative trait loci of shape-derived statistics, such as elongation, which often explain only a percentage of observed variability (Monforte *et al.* 2014; Brewer *et al.* 2006). Further, the study of derived shape statistics constrains the type and role of genes to be discovered and does not allow the inference on shape as a whole.

We propose to approach this issue from a holistic, opposite angle. We aim to reproduce expected shapes and textures that would result from a given individual's DNA sequence. This paper is a proof of principle, and we show that this can be accomplished using decoder deep learning algorithms, at least in simplified scenarios. For simplicity, we focus on fruit shapes, although the same principles apply to animal images. Note we focus on generating images and not on predicting quantities, e.g. areas from the images.

Materials & methods

Generating simple 2D and 3D shapes. We generated images of size 64×64 pixels containing ellipses with different height / width ratios and sizes using OpenCV (https://opencv.org/). We assumed that shape and size was controlled by a set of QTLs with an environmental error added, together with a variable number of non-causative SNPs. To do this, a random shape (x / y axis ratio) and size (pixel number) were sampled, noise added to the causal loci, and an ellipse was drawn with these parameters. We employed pymrt library (Metere and Möller 2017) to generate 3D images of 128×128×128 size, where pixels take values 0/1 depending on whether the position is within the ellipsoid or not. To plot 3D images, we used marching cubes algorithm as implemented in scikit learn (Pedregosa *et al.* 2011), via pyplot projection='3d' option.

Cucurbit shapes. Cucurbitacea family displays, probably, the most extreme shape range in the whole plant kingdom, and *Cucurbit melo* is the most variable species within the family, in terms of shape (Paris 2001; Montero-Pau *et al.* 2017). *C. melo* comprises squashes, zucchini, scallop shapes, among others (Figure 1). We developed an algorithm to generate *C. melo* shapes and intermediate, unobserved variants. Starting from Figure 1, we segmented contours using a thresholding algorithm in openCV, next we generated 200 pseudo-landmarks using the algorithm in (Zingaretti *et al.* 2021), followed by a procrustes alignment (https://

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_143, © M. Pérez-Enciso and L.M. Zingaretti 2022

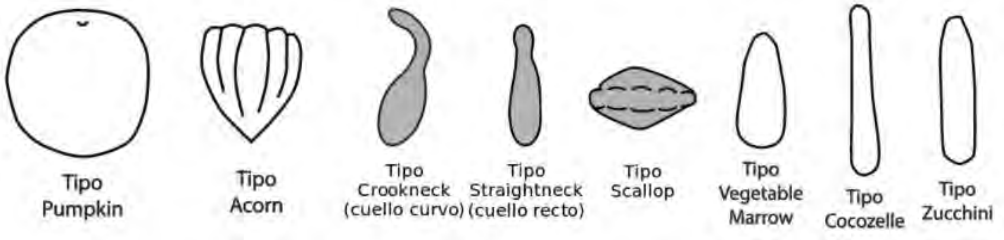


Figure 1. *C. melo* shapes used for simulation. Source: https://es.wikipedia.org/wiki/Cucurbita_pepo, based in Paris *et al.* (2001).

github.com/theochem/procrustes). Following Paris *et al.* (2001), we assumed a hypothetical evolutive path where the ancestral form is a small pumpkin that evolved into acorn → scallop, and a second evolutive path pumpkin → marrow → straightneck → zucchini → cocozelle. The different shapes were determined by a continuous breeding value (20 loci) and therefore intermediate shapes were possible (Figure 2). A random noise to the contour was finally added via an autoregressive process (https://stackoverflow.com/questions/33898665/python-generate-array-of-specific-autocorrelation). Finally, 100 random genotypes were added to the 20 causative loci.

Decoding algorithms. To generate images from linear genotypic data, we used the 'decoder' part from an 'autoencoder'. Very briefly, an autoencoder is an unsupervised deep learning algorithm where the inputs are equal to the outputs (Goodfellow *et al.* 2016). Penalization is employed to avoid overfitting, allowing the algorithm to retrieve repeatable patterns between images. Principal component analysis is equivalent to a linear autoencoder. An autoencoder consists of an 'encoder', which transforms image into a lower dimensional 'latent space' and a decoder that transforms back the vector into a reconstructed image. Since input and output dimensions are the same in an autoencoder, this is unsuited for our needs. However, we can use a decoder architecture and train it for our objective, since in this case input is linear (DNA information) and output is an image. We treated DNA as our latent space. As input, then, the model received non causative SNP and QTL genotypes, the output consisted of the corresponding shapes.

Implementation. About 1000 SNP lists and their corresponding images were generated, which were split into training (n=800), validation (n=100), and testing (n=100) subsets. Predictions presented are for test samples. All analyses were implemented in python 3.9 using keras, tensorflow, open-cv, scikit-learn, PIL libraries, among other packages.

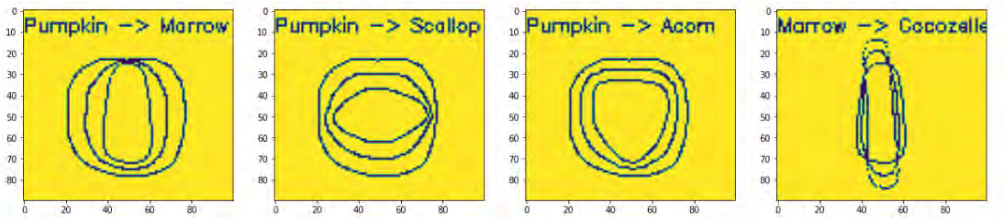


Figure 2. Examples of several pseudo-landmark profiles of *C. melo* shapes from Figure 1 and intermediate simulated shapes that represent hypothetical unobserved evolutive forms.

Results

Predicting 2D and 3D shapes and size. Top row in Figure 3 shows some of the ellipse shapes simulated, whereas the predicted shapes are in the lower row. The algorithm retrieves accurately size, orientation, and shape. Moreover, a similar algorithm using 3D convnet layers also accurately predicts 3D volumes (Figure 4).

Generating and predicting cucurbit shapes. Figure 5 shows a sample of simulated and predicted *C. melo* shapes. The use of a very simple and unoptimized network possibly prevented us from obtaining results as accurate as in the previous case. Yet, the algorithm provides quite reasonable outputs, except in a few cases (sample 999).

Discussion

This work opens new, unprecedented perspectives on the applications of deep learning to Agriculture. Although in simplistic scenarios, we show that it is possible to generate 2D and 3D images out of a linear vector, e.g. from DNA data. Undoubtedly, much work remains to be done. Improved algorithms should be trained with real image and SNP data and much more realistic images could be generated via, e.g. conditional generative adversarial networks or variational autoencoders (Goodfellow *et al.* 2016), but technology is there.

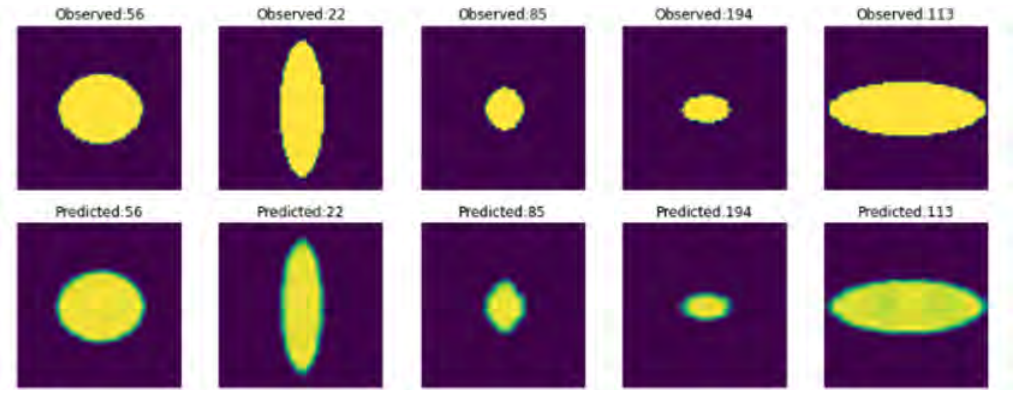


Figure 3. Top row: simulated ellipses; bottom row: predicted shapes.

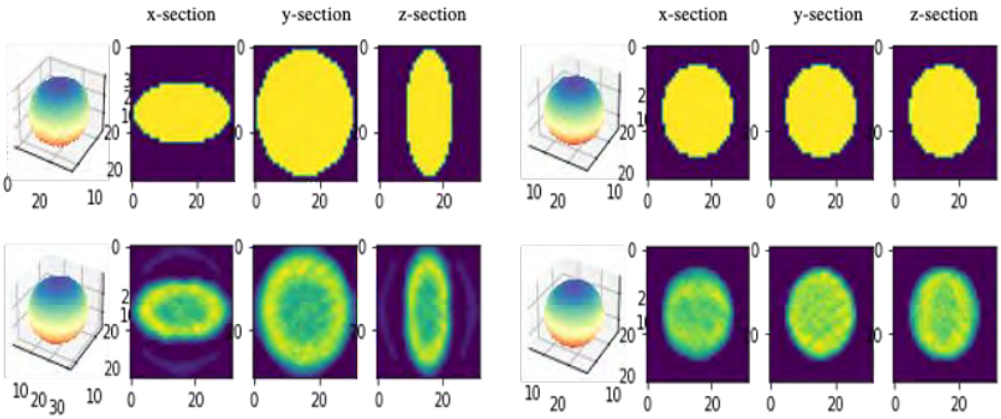


Figure 4. Top: sections of simulated 3D shapes. Bottom: sections of predicted 3D images.

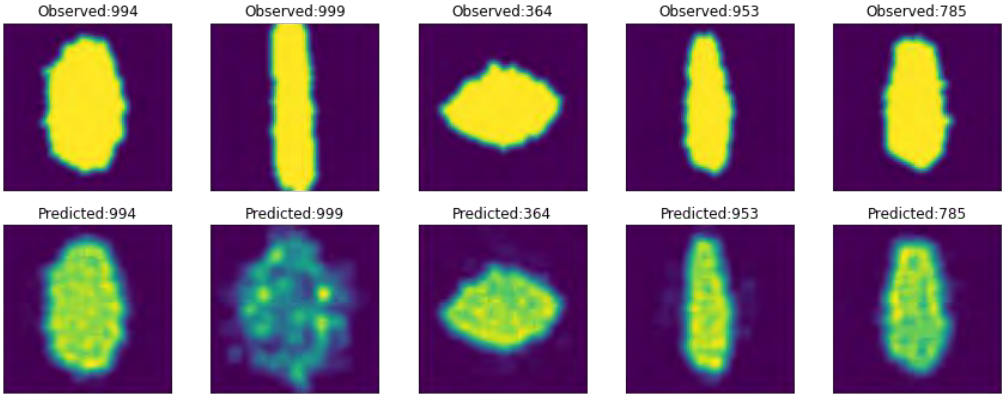


Figure 5. Top: Examples of simulated *C. melo* shapes after adding an autoregressive noise to the contours in Figure 1. Bottom: Predicted shapes based on genotype data.

References

Brewer M., Talbot L., Lang J., Fujimura K., Dujmovic N. *et al.* (2006) *Plant Physiology* 141: 15–25. https://doi.org/10.1104/pp.106.077867.

Goodfellow I., Bengio Y., Courville A. (2016) *Deep Learning*. MIT Press, Cambridge, USA.

Hinton G.E., and Salakhutdinov R.R. (2006) *Science* 313: 504. https://doi.org/10.1126/science.1127647

Metere R. and Möller H.E. (2017). *Proceedings of the International Society of Magnetic Resonance in Medicine*, 3816. Honolulu, Hawai. https://pypi.org/project/pymrt/.

Monforte A.J., Diaz A., Caño-Delgado A., and van der Knaap E. (2014). *Journal of Experimental Botany* 65: 4625–37. https://doi.org/10.1093/jxb/eru017.

Montero-Pau J., Esteras C., Blanca J., Ziarsolo P., Cañizares J., and Picó B. (2017) Genetics and Genomics of Cucurbita Spp. In *R. Grumet et al. (Eds), Genetics and Genomics of Cucurbitaceae*. Springer, New York, USA.

Paris H.S. (2001) *Horticultural Reviews* 25: 71–170. https://doi.org/10.1002/9780470650783.ch2

Pedregosa F., Varoquaux G., Gramfort A., Michel V. *et al.* 2011. *Journal of Machine Learning Research* 12: 2825–30. https://scikit-learn.org/

Zingaretti L.M., Monfort A., and Pérez-Enciso M. (2021) *Plant Phenomics 9812910*. https://doi.org/10.34133/2021/9812910

Part 10

Novel phenotyping tools: sensor devices

144. Estimates of genetic parameters for milkability traits derived from automatic milking systems in North American Holstein cattle

V.B. Pedrosa[1*], J.P. Boerman[1], L.S. Gloria[1], S.Y. Chen[1,2], M.E.M. Gonzalez[1], J.S. Doucette[3] and L.F. Brito[1]

[1]Department of Animal Sciences, Purdue University, West Lafayette, IN 47907, USA; [2]Farm Animal Genetic Resources Exploration and Innovation Key Laboratory of Sichuan Province, Sichuan Agricultural University, Chengdu, 611130, China; [3]Agriculture Information Technology (AgIT), Purdue University, West Lafayette, IN 47907, USA; vpedrosa@purdue.edu

Abstract

This study aimed to estimate genomic-based genetic parameters for milkability traits derived from automatic milking systems (AMS) in Holstein cattle. The traits analysed included average milk flow rate (FR, kg/min), maximum milk flow rate (FRM, kg/min), milking speed (MS, kg/min), and milking time (MT, min). The genetic parameters for FR, FRM, and MT were estimated based on random regression models while a repeatability animal model was used for MS. The daily heritability estimates ranged from 0.45 to 0.56 for FR, 0.47 to 0.58 for FRM, and 0.22 to 0.28 for MT. The heritability for MS was 0.39. The genetic correlation between all traits ranged from -0.85 to 0.96, demonstrating that FR, FRM, and MS are highly genetically correlated but a negative correlation with MT was observed. These results indicated that all traits evaluated can be improved through genetic selection and used when designing breeding schemes to improve milkability in AMS.

Introduction

Automatic milking systems (AMS; also known as milking robots) are revolutionizing the worldwide dairy cattle industry, by improving animal welfare (Jacobs *et al.*, 2012), reducing the need of human involvement in milking practices, and enabling large-scale automatized data collection from farms (Dechow *et al.*, 2020). Another advantage of using AMS is the continuous collection of milking information, including milk yield and other parameters such as milking speed and time (Wethal *et al.*, 2019). Automatic milking systems are equipped with sensors that can measure milk flow rate, milking time, and automatically store a variety of milking related variables. Gathering a large amount of daily phenotypic records enables the derivation of novel phenotypes for the genetic improvement of dairy-associated traits, and result in more accurate genetic parameters and genetic or genomic breeding values (Carlström *et al.*, 2013).

The use of random regression models (RRM) enables the estimation of genetic parameters for individual days of lactation as well as the genetic correlation coefficients between the daily measurements recorded on the same or on different days of lactation (Schaeffer, 2004). Furthermore, the use of RRM as well as individual genomic information in genetic evaluations can result in higher accuracy of estimated breeding values in comparison to other approaches (Oliveira *et al.*, 2019). Thus, the combination of advanced statistical models as RRM with genomic information can enhance the reliability of genetic results for traits with repeated records (longitudinal traits) in dairy cattle.

Evaluating the genetic background of milkability traits is paramount for the design and refinement of selection indexes, especially those focusing on precision dairy systems. To our knowledge, this is the first study using the GBLUP approach to estimate genetic parameters for flow rate and milking time across lactation in dairy cows, especially in North American Holsteins. The use of genomic information to predict genetic parameters can provide more reliable estimates, which are essential for designing or refining selection indexes for implementation in dairy cattle breeding programs. Thus, the main objective of this

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_144, © V.B. Pedrosa *et al.* 2022

work was to estimate heritabilities and genetic correlation for AMS-derived traits in North American Holstein cattle using genomic information.

Materials & methods

Dataset. A total of 1,752,713 daily records of milk flow rate (FR, kg/min), maximum milk flow rate (FRM, kg/min), milking speed (MS, kg/min), and milking time (MT, min) from 4,118 genotyped Holstein cows were collected from 2017 to 2021. The data was recorded by 36 Lely automatic milking systems (AMS) installed on a commercial dairy farm located in Plymouth, IN, USA. FR was calculated directly from the AMS, and MS was estimated by dividing the average daily milk production by the daily milking time. Phenotypic consistence was evaluated using the R software (R Core Team, v 4.1.2), with a threshold of 3.5 SD from the mean to exclude potential outliers. Genotype imputation was performed using the FImpute software (Sargolzaei *et al.*, 2014) to converge all SNP panels available to a common set of 72,820 SNPs for all individuals. Quality control (QC) was performed to remove individuals and SNPs with call rate lower than 0.90, minor allele frequency (MAF) lower than 0.05, and extreme deviation from Hardy-Weinberg equilibrium ($P<10^{-8}$) using the BLUPF90 family programs (Misztal *et al.*, 2014). After the QC, 57,600 SNPs and 4,118 animals remained for further analyses.

Models and parameters estimation. First, two-trait random regression models were used to estimate genetic parameters for FR, FRM, and MT, using a third order Legendre polynomial considering the number of days in milk from 5 to 305 days post-partum. A repeatability model was applied to estimate the heritability for MS. Variance components and genetic parameters were estimated using the AIREMLF90 software (Misztal *et al.*, 2014) and the following statistical model:

$$Y = \mu + YS + L + pe + ak + e;$$

Where μ is the overall mean; YS is the concatenation of year and season (March – May, Jun – Aug, Sep – Nov, Dec – Feb) of calving, fitted as a random effect; L is the fixed effect of the lactation number (1, 2, or 3+); pe and ak are the effects of the random regression coefficients of Legendre orthogonal polynomials for the additive genetic effect and permanent environmental effect, respectively. For MS, the same effects were considered, and DIM was fitted as a linear covariate (fixed effect).

Results

The average (±SD) phenotypic values for FR, FRM, MS, and MT were 1.55±0.54 kg/min, 2.40±0.75 kg/min, 2.16±0.54 kg/min, and 19.14±6.76 min per day, respectively. The heritability estimates for the traits analysed showed low variation throughout the lactation (Figure 1). The daily heritabilities ranged from 0.45 to 0.56 for FR, 0.47 to 0.58 for FRM, and 0.22 to 0.28 for MT. The shape of the heritability curves for FR and FRM was similar, with the highest heritability values around the lactation peak. For MT, the shape of the heritability curve presented a smoother wave, gradually increasing from the first days of lactation. The heritability (S.E.) estimate for MS was 0.39 (0.01) and the repeatability (S.E.) was 0.64 (0.02).

The phenotypic correlations between FR, FRM, MS, and MT ranged from -0.48 to 0.89 (Table 1). Positive and high phenotypic correlations were observed between traits related to milking speed, but on the other hand, the phenotypic correlations of all traits with MS were moderate and negative. The genetic correlation estimates between the milkability traits (Table 1) ranged from -0.85 to 0.96 and followed a similar pattern of the phenotypic correlations. The genetic correlations between milk flow rate traits (FR and FRM) and MS indicate that they are controlled by a similar set of genes. However, MT and the other studied characteristics showed an antagonist genetic correlation.

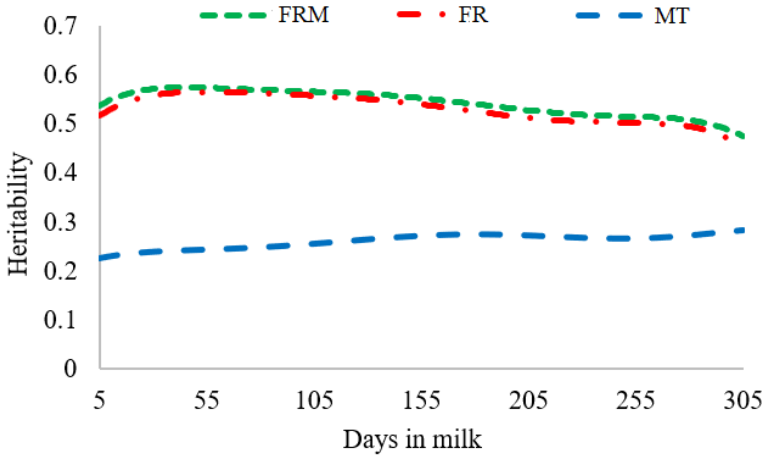


Figure 1. Heritability estimates for milk flow rate (FR), maximum milk flow rate (FRM), and milking time (MT) based on random regression models in North American Holstein cattle.

Discussion

The moderate to high heritabilities observed for FR (0.45-0.56), FRM (0.47-0.58), MS (0.39), and MT (0.22-0.28) indicate that all these traits are heritable and can be improved through genetic or genomic selection. These estimates agree with literature reports for other dairy cattle populations as demonstrated by Carlström *et al.* (2013), Marete *et al.* (2018), Wethal *et al.* (2019), and Dechow *et al.* (2020), with heritabilities ranging from 0.26 to 0.58 for the same traits evaluated in our study. High heritabilities for milkability traits are essential since milking flow/speed is a relevant aspect related to the adaptation of cows to automatic milking systems and can be associated with milking efficiency.

The genetic correlations among FR, FRM, and MS were high and positive, in part because these traits represent similar biological mechanisms related to milking speed. High genetic relationship between FR and FRM was also reported in the literature (0.95±0.01) in a study using different breeds milked in AMS (Edwards *et al.*, 2014). Due to the similar heritabilities and the high genetic correlation between FR and FRM, the use of one of these traits as a selection criterion is sufficient to improve milking flow in North American Holstein cattle. The same concept was observed for the genetic correlation between milking flow traits and MS (Table 1), where high genetic associations were observed. Usually, FR and MS are treated similarly in the literature, often considered as the same trait. However, in our study, these two traits cannot be interpreted as the same since they were measured differently, where FR was calculated directly from

Table 1. Estimates of genetic (above diagonal) and phenotypic (below diagonal) correlations (±S.E.) between milk flow rate (FR), maximum milk flow rate (FRM), milking speed (MS), and milking time (MT) for North American Holstein cattle.

Trait	FR[1]	FRM[2]	MS[3]	MT[4]
FR[1]	-	0.98 (±0.01)	0.95 (±0.02)	-0.75 (±0.02)
FRM[2]	0.89 (±0.01)	-	0.96 (±0.02)	-0.77 (±0.02)
MS[3]	0.85 (±0.01)	0.80 (±0.01)	-	-0.85 (±0.02)
MT[4]	-0.47 (±0.01)	-0.44 (±0.01)	-0.48 (±0.01)	-

the AMS, and MS was calculated by dividing the average daily milk production by the daily milking time. Nevertheless, the high genetic correlation between FR and MS indicates that only one of the measurement methods is necessary to obtain similar genetic progress in the breeding goal. Finally, the negative but moderate genetic correlation observed between the milking flow/speed traits and MT indicates that these traits have a different genetic control, and both should be included in breeding schemes. Genetic selection for FR, FRM, or MS, and MT can directly benefit farmers who seek to improve milking efficiency of herds managed in automatic milking systems. Our research group is actively investigating other novel traits derived from AMS data.

Acknowledgements

The authors would like to thank Homestead Dairy, LLC and especially Brian and Jill Houin for inviting us on their farm and for allowing our research group to utilize their farm records. We also thank Lely for giving us access to their datasets. This activity was funded by Purdue University as part of AgSEED Crossroads funding to support Indiana's Agriculture and Rural Development.

References

Carlström, C., Pettersson, G., Johansson, K., Strandberg, E., Stålhammar, H., *et al.* (2013). Journal of Dairy Science, 96(8):5324-5332. https://doi.org/10.3168/jds.2012-6221

Dechow, C. D., Sondericker, K. S., Enab, A. A., and Hardie, L. C. (2020). Journal of dairy science, 103(12):11503-11514. https://doi.org/10.3168/jds.2020-18786

Edwards, J. P., Jago, J. G., and Lopez-Villalobos, N. (2014). Journal of dairy science, 97(1):259-269. https://doi.org/10.3168/jds.2013-7051

Jacobs, J. A., and Siegford, J. M. (2012). Journal of dairy science, 95(5): 2227-2247. https://doi.org/10.3168/jds.2011-4943

Marete, A., Sahana, G., Fritz, S., Lefebvre, R., Barbat, A., *et al.* (2018). Journal of dairy science, 101(7):6205-6219.

Misztal, I., Tsuruta, S., Lourenco, D.A.L., Aguilar, I., Legarra, A., *et al.* (2014). Manual for BLUPF90 family of programs. University of Georgia, Athens, USA.

Oliveira, H.R., Brito, L.F., Lourenco, D.A.L., Silva, F.F., Jamrozik, J., *et al.* (2019). Journal of dairy science, 102(9):7664-7683. https://doi.org/10.3168/jds.2019-16265

Sargolzaei, M., Chesnais, J.P., and Schenkel, F.S. (2014). BMC Genomics, 15:478. https://doi.org/10.1186/1471-2164-15-478

Schaeffer, L.R. (2004). Application of random regression models in animal breeding. Livestock Production Science, 86:35-45. https://doi.org/10.1016/S0301-6226(03)00151-9

Wethal, K. B., and Heringstad, B. (2019). Journal of dairy science, 102(9):8221-8233. https://doi.org/10.3168/jds.2019-16625

145. Classifying the likelihood of conception in dairy cow with milk mid-infrared spectra before the first insemination

W. Lou[1,2], R. Shi[1,2], B. Ducro[2], A. van der Linden[2], H.A. Mulder[2], S.J. Oosting[2], L. Liu[3] and Y. Wang[1*]

[1]China Agricultural University, Beijing, 100193, China; [2]Wageningen University & Research, P.O. Box 338, 6700 AH Wageningen, the Netherlands; [3]Beijing Dairy Cattle Center, Beijing, 100192, China; wangyachun@cau.edu.cn

Abstract

Accurate and early identification of the likelihood of conception (LC) in cows is imperative for a profitable dairy farm. This study aims to use the milk mid-infrared (MIR) spectra in different intervals before the first insemination and partial least squares discriminant analysis (PLS-DA) to predict LC. The results show that the MIR data within 30 to 50 d after calving and close to insemination had a better prediction in LC (accuracy = 73.9 and 72.3%) than 0 to 30 d. And specificity (74.4 to 84.9%) was higher than sensitivity (67.8 to 73.1%). Once the expected date of insemination is given, the model can predict LC before the actual insemination, and intervene in advance for cows predicted poor LC such as delayed insemination and treatment can be initiated. The predicted LC also provide a novel and convenient way to accumulate extra reproductive phenotypes for genetic evaluation.

Introduction

As one of the most important factors in reproduction performance of dairy cow, the likelihood of conception (LC) has impact on the management and subsequent milk production of the herd because cows with more insemination or infertility are often culled and it does not fit the farms calving pattern or is not profitable, e.g. due to a prolonged dry period of the cow or lower milk revenues compared to a replacement heifer. Prior knowledge of the LC to a given insemination would enable farmers to optimize breeding decisions. The producer may decide not to use the cow where a poor LC is predicted but instead to wait for a subsequent estrus event or treatment. Sexed semen or expensive semen could be used where a greater LC is predicted. Thus, early diagnosis of LC in dairy cow needs to be investigated for defining more effective reproductive strategies. Dramatic changes in milk composition (e.g. sexual hormone) before insemination underlie the prediction of LC for the cows (Toledo-Alvarado *et al.*, 2018). And milk compositions are routinely quantified by mid-infrared (MIR) spectroscopy in many milk recording organizations (De Marchi *et al.*, 2014). In recent years, a few studies have attempted to calibrate milk spectra to predict LC in dairy cows before an insemination, reporting AUC of prediction results were 0.487 to 0.675 (Hempstalk *et al.*, 2015) and 0.66 to 0.75 (Ho *et al.*, 2019). The MIR predicted LC has greater heritability than true LC phenotypes (van den Berg *et al.*, 2021). These results highlighted the potential of milk MIR spectra as a predictor of LC, and MIR-predicted reproductive indicators can be used as a new trait for traditional reproductive phenotypes that are used in genetic evaluation. Nevertheless, there are no studies that determine which optimal time of milk testing after calving and before insemination to obtain the highest prediction accuracy. Therefore, this study aimed to: (1) using MIR spectra to classifying good and poor LC of the cow at the first insemination after calving and find the optimal test period; (2) identify the most relevant wavenumbers from MIR spectra that contribute to prediction accuracy of LC.

Materials & methods

Data. Original records of MIR spectra and reproduction from Chinese Holstein cows were obtained from December 2016 to March 2019 from 12 dairy farms located in Beijing, China. The MIR spectra were tested monthly using the machine FTS (Bentley, Chaska, USA) at the Beijing Dairy Cattle Centre. Spectra located in the infrared range from 649.03 to 3,999.59 cm^{-1} (i.e. containing 899 wavenumbers) were included. Only

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_145, © W. Lou *et al.* 2022

cows within the range of 1 to 375 days in milk and up to 4 parities were selected. To ensure the cow had normal fertility performance, the records of gestation length ranging from 268 to 296 d, interval from calving to first insemination ranging from 20 to 230 d were included in the analysis. In this study, three-class of LC at first insemination were grouped: good (cow that conceived at first insemination), average (cows that conceived at 2 or more inseminations), and poor (cows with no conception event recorded and that had only 1 insemination). From an economic and breeding perspective, it would be very valuable to be able to predict cows with poor LC, thus the cows with good and poor LC were considered in this study.

Methods. Savitzky-Golay second-order derivative algorithm was used to eliminate the chaos and perturbations from water and others in milk (Grelet *et al.*, 2020). Based on previous studies and known water related regions, 538 spectral points (928 to 1,596 cm^{-1}, 1,693 to 3,025 cm^{-1}) out of the 899 wavenumbers were kept for further analyses. A further detection of outliers was performed using the standardized Mahalanobis distance combined with principal component analysis (Grelet *et al.*, 2020). Finally, 95% of data was kept according to the standardized Mahalanobis distance. Here, the dates of calving and first insemination were set on day 0 and the last day. MIR data in different intervals after calving include 0 to 30 d (early stage), 30 to 50 d (middle stage), 50 d to last day (late stage), 0 to 50 d, and 0 to last day were considered. The two time points 30 d and 50 d were chosen to ensure data balance. Based on previous results, MIR spectra, days in milk, parity, somatic cell count, milk yield, and interval from calving to first insemination were selected as the independent variables in the model.

In this study, PLS-DA was used to develop the prediction model. When one herd was used for external validation, then the remaining data was used for training model and internal validation (random, 10-fold cross-validation). Cross-validation was used to evaluate model performance and enable comparison between the different models. The indicators accuracy, area under the curve (AUC), specificity, and sensitivity were used to define which model performed best.

$$\text{Accuracy} = (TP+TN)/(TP+TN+FP+FN) \quad (1)$$

$$\text{Sensitivity} = TP/(TP+FN) \quad (2)$$

$$\text{Specificity} = TN/(TN+FP) \quad (3)$$

Where *TP*, *TN*, *FP*, and *FN* represent total numbers of true positives, true negatives, false positives, and false negatives as predicted by the model, respectively. Here, positive is the cow with good LC and negative is the cow with poor LC.

Results

The prediction results of external validation at different intervals before first insemination are shown in Table 1. The accuracies and AUCs ranged from 70.2 to 77.3% and 0.698 to 0.775, respectively. Sensitivity and specificity respectively represent the probability of correct prediction of individuals with good LC and poor LC in the real data. Specificity (74.4 to 84.9%) was higher than sensitivity (67.8 to 73.1%). In addition, the MIR data within 30 to 50 d after calving had best prediction of poor LC (accuracy = 73.9%) cows compared to that within 0 to 30 d and 50 to last d (Table 1).

To find out the most relevant wavenumbers that contribute to prediction in LC of the cows in this study, the regression coefficient in PLS-DA after normalization was considered for the different intervals (0 to 30 d, 30 to 50 d, 50 to last d). The wavenumbers that were the most relevant to predict the LC were found in five small regions, names from 936 to 1,189, 1,395 to 1,596, 1,693 to 1,790, 2,346 to 2,364, and 2,857 to

Table 1. The prediction results of herd-by-herd external validation based on mid-infrared data at different intervals before first insemination.

Interval	Records	No. cow	Accuracy[1]	AUC	Sensitivity[2]	Specificity[3]
0[4]-last[5] d	2,766	2,140	77.3%	0.775	73.1%	84.9%
0-50 d	1,886	1,422	74.3%	0.746	70.8%	78.5%
0-30 d	967	724	70.2%	0.698	67.8%	74.4%
30-50 d	956	727	73.9%	0.741	71.2%	76.9%
50-last d	955	725	72.3%	0.731	69.5%	75.3%

[1] Accuracy = (TP+TN)/(TP+TN+FP+FN).

[2] Sensitivity = TP/(TP+FN).

[3] Specificity = TN/(TN+FP). TP, TN, FP, and FN represent total numbers of true positives, true negatives, false positives, and false negatives as predicted by the model, respectively.

[4] The 0 day = the date of calving. [5]the last day = the date of first insemination. See main text for other abbreviations.

3,017 cm^{-1} (Figure 1). In addition, the regression coefficients after normalization of the three intervals had moderate to low Spearman correlations, ranged from 0.604 (0 to 30 d and 30 to 50 d) to 0.284 (0 to 30 d and 50 to last d).

Discussion

The innovation of this study is to defining the optimal test period to better classify good and poor LC based on MIR spectra. In addition, the MIR region associated with this reproductive trait was identified. The ability to accurately predict the conception outcome for a future mating would be of considerable benefit for producers in deciding what mating plan to implement for a given cow. The results of this study can be used in the development of a novel tool for monitoring the fertility and health in dairy cows before insemination. And accurate prediction implemented in current study also provide a new trait to accumulate reproductive phenotypes for genetic evaluation, especially before routine collection of reproductive information is established.

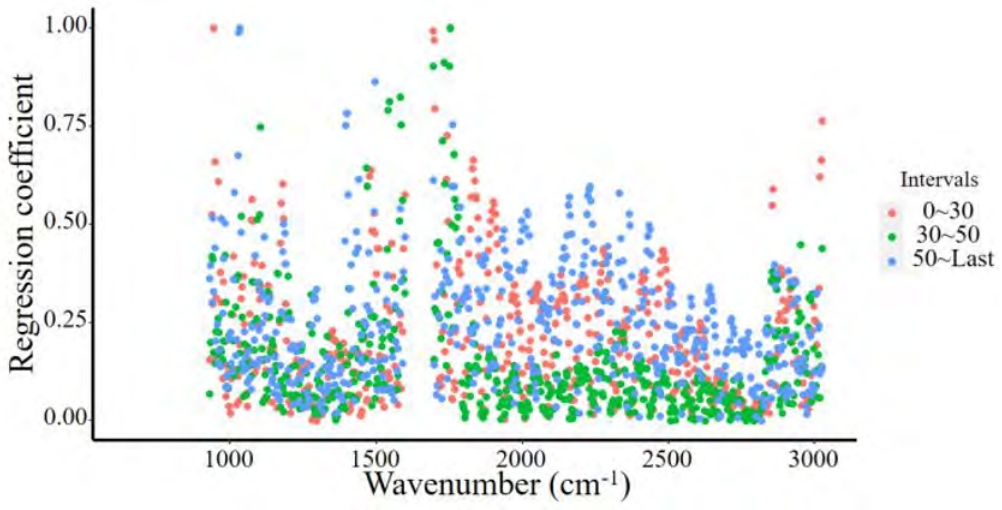


Figure 1. The regression coefficient after normalization for different wavenumbers in intervals of 0 to 30 (early stage), 30 to 50 (middle stage), and 50 to last (late stage) days.

Compared to the published studies, the accuracy (70.2 to 77.3%) and AUC (0.698 to 0.775) of external validation in this research are within these reported results (AUC: 0.487 to 0.750) (Hempstalk *et al.*, 2015; Ho *et al.*, 2019). Such differences among prediction is mainly due to differences in MIR data selection, spectral processing, and regression method employed. Specificity (74.4 to 84.9%) was higher than sensitivity (67.8 to 73.1%) in this study, which means the model has stronger ability to identify a cow with poor LC than a cow with good LC and this tendency is more important for reproductive management. The MIR data within 30 to 50 d after calving and close to insemination had the better prediction in LC (Accuracy = 73.9 and 72.3%) than 0 to 30 d. This indicated that collecting MIR spectra during these periods may obtain optimal prediction accuracy.

The most important wavenumbers of MIR may help in understanding why MIR spectra before insemination can predict LC. The regions from 936 to 1,189, 1,395 to 1,596, 1,693 to 1,790, 2,346 to 2,364, and 2,857 to 3,017 cm^{-1} are related to the chemical bonds including C-H, C-C, C=O, C-O-C, nitrogen-oxy compounds, aldehydes, and esters (Duffy, 1972). Many sex hormones (e.g. progesterone and follicle-stimulating hormone) contain the chemical bonds like esters, aldehyde, and nitrogen-oxy compounds. A previous study indicated that sex hormone exist in milk and that the level changes significantly before insemination (Toledo-Alvarado *et al.*, 2018), which means that MIR data before insemination contains information on sex hormone changes, especially within 30-50 d after calving and close to insemination. In addition, some studies found high concentration of blood ketone body before insemination will increase the risk of non-pregnancy (Walsh *et al.*, 2007), and MIR spectroscopy is a mature tool to qualify the concentration of ketone body in blood and milk (Benedet *et al.*, 2019). This indicated that the five spectral regions maybe reflect energy imbalance information, because the ketone body contain the chemical bonds like C-H, C-C, and C=O. The variation in other milk compositions could also be related to fertility and health of dairy cows. Given the most important wavenumbers in MIR data to predict LC, prediction based on MIR data before inseminations seems to be (partly) based on reflecting abnormal levels in sex hormones and energy imbalance. Meanwhile, the intervals of 30-50 d after calving and close to insemination might contain more changes in physiological state than the other periods, but it needs to be interpreted with caution.

Acknowledgements

The China Scholarship Council (No.201913043) is acknowledged for funding the PhD projects of Wenqi Lou and Rui Shi. This study is supported by China Agriculture Research System of MOF and MARA.

References

Benedet A., Manuelian C.L., Zidi A., Penasa M., and De Marchi M. (2019). Animal 13(8):1676-1689. https://doi.org/10.1017/S175173111900034X

De Marchi M., Toffanin V., Cassandro M., and Penasa M. (2014) J. Dairy Sci 97(3):1171-1186. https://doi.org/10.3168/jds.2013-6799

Duffy N.V. (1972) J. Chem Educ 49(9):30-45. https://doi.org/10.1021/ed049p652.1

Grelet C., Dardenne P., Soyeurt H., Fernandez J.A., Vanlierde A., and *et al.* (2020) Methods 186(3):97-111. https://doi.org/10.1016/j.ymeth.2020.07.012

Hempstalk K., Mcparland S., and Berry D.P. (2015) J. Dairy Sci 98(8):5262-5273. https://doi.org/10.3168/jds.2014-8984

Ho P.N., Bonfatti V., Luke T.D.W., and Pryce J.E. (2019) J Dairy Sci 102(11):10460-10470. https://doi.org/10.3168/jds.2019-16412

Toledo-Alvarado H., Vazquez A.I., Gustavo De L.C., Robert J.T., and Gianfranco G. (2018) J. Dairy Sci 101(10):9135-9153. https://doi.org/10.3168/jds.2018-14480

Van den Berg I., Ho P.N., Haile-Mariam M., and Pryce J.E. (2021) J. Dairy Sci 2(6):361-365. https://doi.org/10.3168/jdsc.2021-0141

Walsh R.B., Walton J.S., Kelton D.F., LeBlanc S.J., Leslie K.E, and *et al.* (2007) J. Dairy Sci 90(6):2788-2796. https://doi.org/10.3168/jds.2006-560

146. Use of tracking technology for phenotyping laying hens within cage-free housing

C.M. Montalcini, M.B. Petelle, S.G. Gebhardt-Henrich and M.J. Toscano*

Center for Proper Housing: Poultry and Rabbits (ZTHZ), Division of Animal Welfare, VPH Institute, University of Bern, Burgerweg 22, 3052, Zollikofen, Switzerland; michael.toscano@vetsuisse.unibe.ch

Abstract

Our group has been tracking individual hen movements within a commercial aviary and identified consistent within- and among-animal movement and location patterns (MLPs). To examine the potential for breeding selection and health and welfare, we conducted two trials that tracked either 80 (Trial A) or 1,125 (Trial B) Dekalb White hens over a subset of the entire laying period during which data on health was assessed and related to movement. Results from principal component (PC) analysis within Trial A based on 13 movement variables collected during the first seven weeks in the barn demonstrated moderate repeatability (0.44 [CI: 0.43, 0.48]), temporal plasticity, and predictability (0.25 [CI: 0.21,0.31]) and a negative relationship between predictability of the first PC score and keel fracture. Using a population of crossbreds with known sires for Trial B, a low but present variance attributed to individual (~13-19%) and sire (~1%) was identified. Our effort details novel means to collect movement data towards development of stable phenotypes that can be used within laying hen breeding.

Introduction

Cage-free housing systems for laying hens provide superior welfare over cage systems by providing a rich array of resources such as outside areas and litter for dustbathing that allow animals to express their natural behaviours (Weeks and Nicol, 2006). Despite these benefits, many problems exist that unintentionally compromise the welfare of housed animals including those of behavioural (e.g. feather pecking and toe pecking), physical (e.g. keel bone fracture), and disease (e.g. Erysipelas) natures.

The problem is likely rooted in current laying hen breeding paradigms that target performance within small groups of hens. Consequently, available genetic hybrids are sub-optimal for cage-free housing systems with thousands of animals leading to the discussed poor welfare. Given the increasingly global trend towards cage-free housing, it is important to develop appropriate methods to assess phenotypes in commercially relevant cage-free systems. By use of a uniquely commercial aviary outfitted for research, our group has been able to combine traditional animal assessments (e.g. manual evaluation of feather quality) with sensor tracking technology to provide an accurate representation of an individual hen's health and movement. We previously demonstrated highly consistent MLPs over time that were nonetheless starkly different between hens (Rufener *et al.*, 2018) and that birds with more severe types of keel bone fracture spent a greater amount of time in the aviary's upper tiers (Rufener *et al.*, 2019).

Materials & methods

To understand selection for movement and location patterns for laying hens within cage-free housing systems, the current effort involved two projects that sought to yield a prospective suite of MLP-based traits that could be used within breeding paradigms of cage-free housed laying hens. The first project (Trial A) explored creation of an index of phenotypically correlated traits within a composite movement variable and its repeatability, plasticity, and predictability, as well as their relationship to traditional welfare measures. A second effort (Trial B) used a population with known 28 known sires (8-11 animals of each sire /pen) to focus on a narrower subset of traits based on the animal's presence within three specific areas believed to have particular health and welfare importance.

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_146, © C.M. Montalcini *et al.* 2022

General. Two studies were conducted that tracked either 80 (Trial A) or 1,125 (Trial B) Dekalb White hens. In Trial A, hens were reared on site in one of four pens and in Trial B they were reared in eight pens with two different rearing aviaries (Rufener *et al*., 2019). At 17 weeks of age (WOA), hens were transferred to an on-site laying barn containing a Bolegg Vencomatic Terrace aviary (split into 20 identical pens separated by a vertical grid) and an outside winter garden accessible through a pop hole. Each pen contained 225 hens, of which eight were used for Trial A with 20 pre-selected focal hens. For Trial B, five pens were used of which all animals were tracked. Hens were a specialized crossbred where the sire of each was assigned a coded wingband at the hatchery. The wintergarden was initially closed (Trial A: 1 week; Trial B: 3 weeks) so that hens could acclimate to the nest boxes, perches, and the indoor area in general and thereafter opened and closed daily. Tracking data was collected near continuously by either a low-power active (Trial A) or radio frequency identification (Trial B) tracking system (Gantner Solutions GmbH, Schruns, Austria). Tracking tags were attached via a backpack or legband (Trial A and B, respectively). Each system was used to register the animal within five mutually exclusive zones containing key resources that together represented all places within the aviary: the three stacked tiers (top, middle, lower), as well as the littered floor and the wintergarden. For the work described here, data from between 19 to 25 WOA (Trial A) or 27 and 30 WOA was used for analysis.

Trial A. At 60 WOA, we assessed keel bone fractures based on radiographs with a scoring methodology used previously (Rufener *et al.,* 2018). Feather damage of the breast, tail, and neck was assessed using the Welfare Quality protocol (Welfare Quality Network, 2009). From the processed movement data, we extracted 13 daily variables per hen. Variables computed while artificial light was provided included: the percentage of time and the average number of stays per hour (to account for varying barn schedule) in each zone, the average realized vertical distance per hour, and whether the hen entered the winter garden within 15 min after access was possible. The sleeping height, measured by the number of stacked tiers underneath the zone where an animal was for most of the period when artificial light was off, was also calculated. We extracted a linear composite variable using PC analysis from psych package (Revelle, 2021). Linear mixed effect models were used to identify consistent individual movement differences in averages (individual intercept), temporal plasticity (individual slope), and predictability (individual variability). In a second step, we used bivariate models to evaluate if individual estimates of movements were associated with health assessments. Analyses were conducted using python for the data processing and R for statistics.

Trial B. Collected movement data was analysed using linear mixed models to relate daily duration in three of the five zones (continuous-Gamma distribution) with fixed (pen) and random (hen ID; sire without pedigree) effects. The three zones were targeted on a belief that they contain resources indicative of healthy behaviour (e.g. wintergarden and floor containing litter for dustbathing, foraging) and production (nestbox). Use of external areas (e.g. wintergarden) is also a trait of primary interest to consumers.

Results

Trial A. The first three PCs had eigenvalues >1. The first PC explained 41% of the variation with eight of the variables having a loading (in absolute value) >0.4 and six >0.7 (Table 1). The positive loadings on vertical distance/hour and number of stays/hour in zones other than the highest tier (where the majority of animals would sleep) suggested PC1 reflected general movement throughout the barn. A higher PC1 value would associate with animals that spent more time in the floor, lower tier, and wintergarden zones, but also transitioned more between these zones. In contrast, PC2, which explained 20% of total variation with four variables having an absolute loading value >0.4, had the greatest positive loadings on the middle tier which may reflect egg-laying behaviour. Lastly, PC3, which explained 14% of variation with three variables having an absolute loading value >0.4, was associated primarily with wintergarden presence. Focusing on the PC1, we found a repeatability estimate of 0.44 [CI: 0.43, 0.48] indicating that 44% of the remaining variance (after

Table 1. Loadings of a correlation-based principal components analysis within Trial A.

Variable	PC1	PC2	PC3
% Duration			
Highest Tier	-0.91	0.13	0.22
Lowest Tier	0.74	-0.23	-0.30
Floor	0.76	-0.25	-0.34
Middle Tier	0.12	0.69	-0.01
Wintergarden (when open)	0.40	-0.25	0.77
# Stays/hour			
Floor	0.92	-0.18	0.01
Lowest Tier	0.91	0.22	-0.18
Highest Tier	0.04	0.85	0.28
Middle Tier	0.27	0.90	0.13
Wintergarden (when open)	0.52	-0.25	0.69
Other			
Height at night (dark)	-0.35	0.09	0.25
Vertical distance/hour	0.81	0.51	-0.04
Wintergarden in 1st 15 min (Y/N)	0.39	-0.21	0.63

controlling for the external temperature downloaded from a local weather station and days post transfer) in PC1 can be attributed to differences between individuals. The degree of variation in predictability for PC1 among studied individuals was 0.25 [CI: 0.21, 0.31]. More predictable animals were associated with higher keel bone fracture scores.

Trial B. Examination of daily durations into each of the modelled aviary zones indicated that individual hens and sires composed a small but definitive amount of the phenotypic variance ranging from approximately 13-23 and 0.8-2.8%, respectively. Specifically for each zone (wintergarden, floor, and lowest, middle, and highest tiers) the variances for hens were 13.6, 15.0, 19.1, 23.4, and 28.2%, respectively. For sire, the variances (same zone order) were 1.1, 0.8, 0.8, 2.8, and 2.3%

Discussion

Our results demonstrate that continuous monitoring of individual hens' time entry/exit within a limited set of locations offers a valuable body of traits with relevance for breeding paradigms. While hens are remarkably homogenous in terms of phenotypes that were historically measured within cage housing (e.g. feed efficiency), they are relatively heterogenous in traits that are increasingly important for consumer demands (e.g. feather quality, presence outside). Our initial efforts indicate that some of the observed individual phenotypes can be quantified by average and temporal plasticity of early movements (PC score). In other words, our results offer a novel means to represent animals in a way that captures traits which previously could not be evaluated in a manner relevant for commercial housing. Sufficient sample sizes in future efforts will aim to link these phenotypes with pedigree information.

Individuals exhibited differences in the rate at which they adjusted movement over the observation period with a general increase in the PC1 value over the first weeks in the laying barn. For instance, some individuals exhibited an increased PC1 only after 35 days post-transfer while others did so almost immediately. Further examination suggested a variety of subtle MLPs that will need to be examined with larger datasets like that of Trial B. For instance, a degree of variation in predictability among studied individuals suggested that

some hens were more predictable (i.e. stronger daily routines) in their daily movements. Altogether, results found existence of individual differences in the variation of early movement.

In addition to the previously established link between daily duration in the upper tier and keel bone health, current results indicated a link between specific health indicators and behaviour shortly after transfer to the laying aviary. Therefore, behaviour shortly after transition may be useful as selection criteria to prevent development of poor health. For instance, we found that lower predictability of early movement was associated with reduced keel bone fracture severity later in life. Although such correlations would need to be investigated for causal relationships, animals exhibiting less predictability in movement may develop greater cognitive ability for movement leading to safer transitions between tiers. Alternatively, predictability and inactivity may be associated, although average PC1 value and predictability were only weakly correlated but need to be investigated further. Future analysis will seek to expand the movement variables identified in Trial A towards identifying variation across daughter-sires represented in Trial B.

References

Revelle W. 2021. Procedures for Psychological, Psychometric, and Personality Research. R package psych version 2.1.9. Available at: https://CRAN.R-project.org/package=psych Version = 2.1.9

Rufener C., *et al.* (2019) Appl. Anim. Behav. Sci. 217:48–56. https://doi.org/10.1016/j.applanim.2019.05.007

Rufener C., *et al.* (2018) Sci. Rep. 8:12303. https://doi.org/10.1038/s41598-018-29962-x

Weeks C.A., Nicol C.J. (2006) Worlds. Poult. Sci. J. 62:296–307.

Welfare Quality Network (2009) Welfare Quality assessment protocol for poultry (broilers, laying hens).

147. Learning behaviour of dairy cows in automated milking systems: genetic parameters and suggested candidate genes

R. Schafberg, F. Rosner, D. Oelschlägel and H.H. Swalve*

Institute of Agricultural and Nutritional Sciences, Martin-Luther University Halle-Wittenberg, 06099 Halle (Saale), Germany; hermann.swalve@landw.uni-halle.de

Abstract

Automated milking systems (AMS) are of ever increasing relevance in dairy cattle housing. The systems automatically provide measurements and parameters that can be used as traits for genetic analyses. The present study exhibited estimates of heritabilities of moderate size around 0.14 for the number of visits to the AMS unit, the number of successful milkings as well as the difference between these two parameters, the number of refusals. This latter trait is highly influenced by the ability of the cow to adapt to the time limits set for subsequent milkings and thus is strongly determined by elements of learning behavior thus allowing for genetic selection. A GWAS for the trait refusals revealed a number of interesting chromosomal regions with several genes well known from model animal or human genomics. The results underpin the hypothesis that highly conserved genomic regions affecting behavioural traits exist across species.

Introduction

The use of automated milking systems (AMS) in dairy cattle housing steadily is increasing and considered to be the future standard procedure for milking of dairy cows. The ability of a cow to cope with the technological environment is largely determined by learning behavior. In model animals is has been shown repeatedly that learning behavior is a trait highly influenced by genetics. An AMS may be viewed as a test system that automatically is documenting and recording behavior in general and specifically learning behavior of a dairy cow. Wethal and Heringstad (2019) as well as Chang *et al.* (2020) emphasized the relevance of AMS-recorded traits as the recording: (1) does not interfere with the standard routine on-farm; (2) provides objective measurements; and (3) yields observational values with high phenotypic variance. In our data, amongst other parameters, the system recorded 'visits' and 'milkings' of cows as well as technical 'errors'. 'Visits' and 'milkings' differ as the AMS rejects milking of cows when they visit the AMS-unit too frequently, the difference between the two parameters is denoted as 'refusal'. For the latter trait, it was hypothesized that it likely is affected by the 'ability to learn'. All traits considered were averaged on a daily basis. Aim of the present study was to estimate genetic parameters as well as to identify chromosomal regions with putative influence on the traits studied.

Materials & methods

Data used originated from a large dairy herd equipped with 27 AMS units, each serving around 40 to 50 cows and covered a time span of 25 months in years 2017 to 2019 starting with the entrance of the entire herd to the newly AMS-equipped barn. Data comprised 2,245 cows with a total of 4,190 lactations. SNP genotypes from 50k arrays and 10k arrays imputed to 50k by Vereinigte Informationssysteme Tierhaltung w. V., Verden (Germany), were available for 1,566 cows. Milk yield for each milking was available from the AMS on farm and was converted to daily milk yield by own programming according to standard ICAR procedures. After edits, a total of 949,990 daily phenotypes were available for all traits. As behavior and specifically learning behavior is highly influenced by adaptation, the focus of attention was given to the time period of familiarity or 'experience' with the AMS. For modelling purposes, four classes of experience defined as days at AMS (DAA) were formed with class levels of 0-75, 76-150, 151-250, 251+ days and

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_147, © R. Schafberg *et al.* 2022

combined with the effect of stage of lactation defined as days in milk (DIM) with three class levels of 0-100, 101-200, 201+ days, thus resulting in 12 class levels. The general mixed model applied was

$$y = Xb + Z_1u + Z_2pe + e \quad (1)$$

with y = vector of trait observations, X is the incidence matrix for fixed effects and b the vector of fixed effects to be estimated. X and b comprise the effect of season (4 classes), the combined DAA × DIM effect (12 classes), nested within parity class (4 levels), Z_1 is the incidence matrix for animals' additive genetic effects u, Z_2 is the incidence matrix for permanent environmental effects pe, and e is the vector of residuals. Estimation of LSMEANS for fixed effects and bivariate estimation of variance components for two traits at a time was conducted applying REML using ASREML. For these analyses, relationships between animals in u were considered via a pedigree-based relationship matrix comprising a total of 24,372 animals. Association analysis was carried out following the methodology of a single-step genomic association study (ssGWAS) approach proposed by Wang *et al.* (2012), based on the single-step genomic-BLUP method (ssGBLUP) (Aguilar *et al.*, 2010). The inverse of the H matrix was derived by combining inverses of the genomic relationship matrix G and the pedigree-based numerator relationship matrix A using the BLUPF90 family software as presented by Aguilar *et al.* (2018).

Results

LSMEANS for no. of visits per day initially, for low DAA, start at levels of around nine visits per day and drop to a level of around five for more experience while the number of successful milkings starts at a level of 3.6 and drops to around three for higher experience levels (not shown in table). Technical errors (not shown) occur at a low rate of around 0.14 and are unaffected by experience or stage of lactation or age. Refusals per day (RPD) start at a level of >5.0 for young cows without AMS experience and gradually drop to values around 2 for older and more experienced cows. It appears that learning to adapt to AMS ends with the midst of the second lactation. It should be noted that Table 1 also includes cows in later lactations but with little experience, i.e. they entered the AMS barn as older cows for the first time. For these cows, adaptation is relatively rapid.

Table 1. Least square means of refusals per day (RPD) by parity, stage of lactation, and experience level.

Parity class	DIM[1]	DAA[2]			
		0-75	76-150	151-250	251+
1	1-100	5.91	5.46		
	101-200	5.65	4.86	4.11	
	201+	5.57	4.38	3.77	3.18
2	1-100	2.80	3.05	2.33	2.77
	101-200	2.26	2.00	2.10	2.01
	201+	2.17	1.72	2.20	2.32
3	1-100	1.77	1.74	1.54	2.43
	101-200	2.13	1.41	1.57	1.73
	201+	2.70	1.79	1.98	2.06
4+	1-100	1.57	1.67	1.20	1.74
	101-200	2.04	1.16	1.51	1.46
	201+	1.82	1.43	1.91	1.95

[1] DIM = stage of lactation defined as days in milk.
[2] DAA = experience level defined as days at AMS.

Table 2 shows results from selected bivariate REML estimations of variance components. Estimates for the no. of visits and the no. of milkings per day as well as the no. of refusals per day (RPD) are on a moderate level of around 0.13 to 0.15. Technical errors exhibit very low heritabilities thus underpinning that their nature is purely technical and almost unrelated to genetic influences. However, errors appear to be cow-associated as indicated by the proportion of permanent environmental variance, c^2, of 0.49. Genetic correlations of visits × milkings as well as milkings × refusals are of substantial magnitude of around 0.80 due to the inherent autocorrelation.

Figure 1 displays results for GWAS as defined with -10log (pvalue) for SNP. Interesting chromosomal regions (-10log(pvalue) 4.0 to 6.0) appear to reside on chromosomes 3, 12, 18, and 19 with spurious elevated values for single SNP on other chromosomes.

Table 3 contains additional information for trait RPD and chromosomes 3, 12, 18, 19. It is striking to see that in all five regions genes are found that in mouse genetics and especially in human genetics are known to affect neuron development, intellectual ability or psychotic disorders. For sake of brevity, just a few publications in this regard will be mentioned here: Noh *et al.* (2019) emphasizes the role of NEGR1 in humans affecting hippocampal neurogenesis and behavior; Garcia-Corzo *et al.* (2013) underline the function of COCQ9 in encephalomyopathy in humans and a recent publication of a large international consortium highlights CNOT1 as a major factor in neuron development in humans (Vissers *et al.*, 2020).

Table 2. Estimates of heritabilities, proportion of permanent-environmental variance and genetic correlations from bivariate REML animal model analyses.[1]

Bivariate analysis		Trait 1				Trait 2				r_g (Trait 1 × Trait 2)	
Trait 1	Trait 2	h^2	SE	c^2	SE	h^2	SE	c^2	SE	r_g	SE
Vistits	Milkings	0.15	0.021	0.23	0.018	0.14	0.017	0.143	0.013	0.84	0.035
Visits	Errors	0.14	0.021	0.24	0.018	0.04	0.016	0.481	0.016	0.24	0.046
Milkings	Refusals	0.14	0.017	0.14	0.013	0.13	0.020	0.229	0.017	0.79	0.046
Milkings	Errors	0.15	0.017	0.14	0.014	0.03	0.015	0.485	0.016	0.60	0.183

[1] c^2 = permanent environmental effect

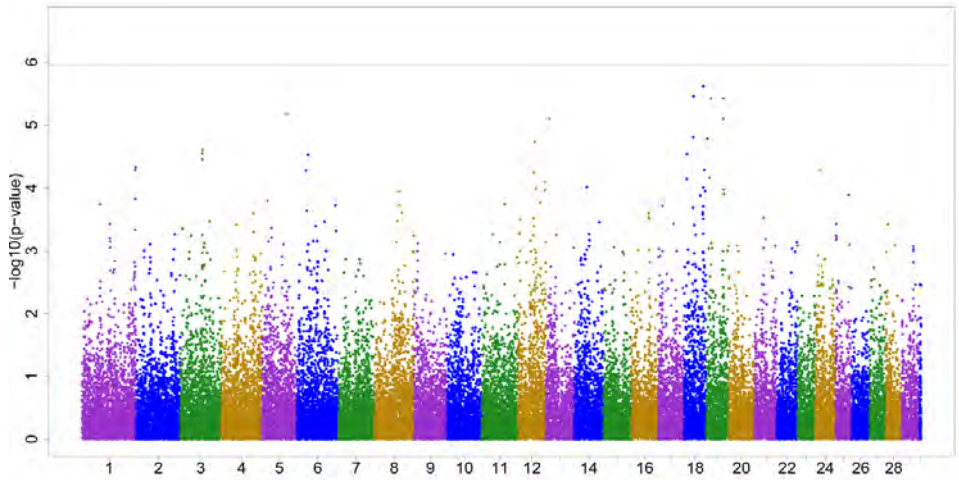


Figure 1. Manhattan plot of GWAS (-log10 *P*-value) for trait refusals per day (RPD).

Table 3. Chromosomal regions on BTA 3, 12, 18, and 19 each with several SNP≥4.0 – 10log(*P*-value).

Chromosome	Region of SNPs	Suggested candidate genes[1]
3	73.5-73.9 Mbp	**NEGR1**, **PTGER3**
12	53.0-55.0 Mbp	SPRY2, POU4F1
12	84.3 Mbp	**MYO16**, **CARS2**, IRS2
18	25.4-26.2 Mbp	**COQ9**, **CNOT1**
18	55.7-55.9 Mbp	**SLC6A16**, PTH2, ALDH16A1
18	60.0-63.2 Mbp	**MBOAT7**, **CNOT3**
19	50.5-50.6 Mbp	**TBCD**, **ACTG1**, **RAC3**

[1] Genes known from human genetics to affect intellectual ability or involved in psychotic disorders in bold.

Discussion & conclusions

As hypothesized by Chang *et al.* (2020) and in this study, parameters automatically recorded in automated milking systems are valuable traits that could be used for genetic selection of adaptation to the technical environment of AMS. Furthermore, the amount of evidence for highly conserved genomic regions across species appears to be increasing. Costilla *et al.* (2020) in a meta study analysed the trait flight time frequently recorded in beef cattle and found that the genetic control of temperament in cattle at least partly is similar to neuron development function and autism spectrum disorders in humans. Alvarenga *et al.* (2021) in a further large-scale meta-analysis of studies of behavioural traits in cattle, sheep, and pigs found numerous links of chromosomal regions identified in livestock when compared to studies of human disorders. They concluded that a number of highly conserved genomic regions exist across species that are related to behavioural traits. Further studies in this field seem to be warranted.

Acknowledgements

Gratitude is expressed to Vereinigte Informationssysteme Tierhaltung w.V. for supplying imputed SNP genotypes and to the participating farm for making data available.

References

Aguilar I., Misztal I., Johnson D.L., Legarra A., Tsuruta S., *et al.* (2010) J. Dairy Sci. 93: 743-752. https://doi.org/10.3168/jds.2009-2730

Aguilar I., Tsuruta S., Masuda Y., Lourenco D.A.L., Legarra A., *et al.* (2018) Proc. of the 11th WCGALP, Auckland, New Zealand.

Alvarenga, A.B.., Hinayah R.O., Chen S.-Y., Miller S.P., Marchant-Forde J.N., *et al.* (2021) Animals 11(3):715. https://doi.org/10.3390/ani11030715

Chang Y., Brito L.F., Alvarenga A.B., and Wang Y. (2020) Animal Frontiers 10(2). https://doi.org/10.1093/af/vfaa006

Costilla R., Kemper K.E., Byrne E.M., Porto-Neto L.R., Carvalheiro R., *et al.* (2020) Genet. Sel. Evol. 52:51. https://doi.org/10.1186/s12711-020-00569-z

Garcia-Corzo L., Luna-Sánchez M., Doerrier C., García J.A., Guaráset A., *et al.* (2013) Human Molecular Genetics 22(6):1233–1248. https://doi.org/10.1093/hmg/dds530

Noh K., Lee H., Choi T.-Y., Joo Y., Kim S.-J., *et al.* (2019) Molecular Psychiatry 24:1189–1205. https://doi.org/10.1038/s41380-018-0347-3

Wang H., Misztal I., Aguilar I., Legarra A., and Muir W.M. (2012) Genet. Res. (Camb) 94:73-83. https://doi.org/10.1017/S0016672312000274

Wethal K.B., and Heringstad B. (2019) J. Dairy Sci. 102:8221–8233. https://doi.org/10.3168/jds.2019-16625

Vissers L.E.L.M., Kalvakuri S., de Boer E., Geuer, S., Oud M., *et al.* (2020) Am. J. Human Gen. 107:164–172. https://doi.org/10.1016/j.ajhg.2020.05.017

148. Repeatability estimates of grazing and rumination activity of Merino sheep measured using wearable sensors

F. Almasi[1,2,3*], M. Khansefid[2], H. Nguyen[4,5], A. Desai[6], J.E. Pryce[2,3] and M. Stear[1]

[1]Department of Animal, Plant and Soil Sciences, La Trobe University, Victoria 3083 Australia; [2]Agriculture Victoria, AgriBio, Centre for AgriBioscience, Bundoora, Victoria 3083, Australia; [3]School of Applied Systems Biology, La Trobe University, Bundoora, Victoria 3083, Australia; [4]School of Mathematics and Physics, University of Queensland, St. Lucia 7072 Australia; [5]School of Engineering and Mathematical Sciences, La Trobe University, Victoria 3083 Australia; [6]Centre for Technology Infusion, La Trobe University, Victoria 3083 Australia; fazel.almasi@agriculture.vic.gov.au

Abstract

Accelerometer sensors were attached to 147 Merino sheep for 26 days on a commercial farm in Victoria, Australia. Using a Support Vector Machine learning algorithm, 6 behaviours were defined, including grazing and rumination. The overall repeatability of grazing and rumination were 0.44±0.03 and 0.49±0.03, respectively. When repeatability was estimated on 2 consecutive days, it was approximately 0.7 for each trait and by increasing the interval between days (over 22) dropped to below 0.1. The heritability estimates for all grazing behaviour traits were close to zero, which is likely to be associated with the small sample size in our study and weak genetic connections between individuals. Nevertheless, the repeatability results were promising and suggested that animals maintain their ranking at least over short periods of time. These results were preliminary and from a small sample size. Hence, we recommend a larger sample size for estimating heritability of grazing and rumination traits.

Introduction

Grazing is commonly defined as the act of placing the head down and grasping grass whilst on pasture, while rumination is the process of regurgitating, chewing, salivation, and re-swallowing feed (Welch *et al.*, 1970). The physiological and nutritional aspects of these two traits have been investigated in ruminants (Beauchemin, 2018), but the genetic aspects of these traits have been rarely investigated, as the most of data has been obtained by visual observation, which are usually too few for quantitative genetics analysis. Since sensor technology has been used to measure the activity of ruminants, with commercial installations now common for management purposes, there is a growing interest in its application to the genetics of previously uninvestigated traits, such as grazing and rumination time. As grazing and rumination are the main activities of ruminant livestock, they have a major impact on growth and health (Hancock, 1954). The heritability of rumination in dairy cattle using sensor data was estimated at 0.30 and 0.34 (Byskov *et al.*, 2017; Moretti *et al.*, 2018), while the repeatability was substantially higher (Byskov *et al.*, 2017). To the best of our knowledge, the heritability and repeatability of grazing and rumination activity has not been estimated in sheep. Using sensor data from a Merino flock, the aim of this research was to estimate: (1) the repeatability of grazing and rumination activities between days and during the whole experiment; and (2) the heritability of grazing and rumination activities.

Materials & methods

This study was carried out on a commercial farm at Edenhope, Victoria, Australia. All the experimental procedures were approved by the Animal Ethics Committee, La Trobe University, Victoria, Australia (AEC number 17-09).

Animal and activity recording. The study was conducted on 147 Merino sheep with the average liveweight of 45.8±14 kg (mean ± S.D.) from 3rd to 29th of May 2020. The ram (n=67) and hogget (n=80)

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_148, © F. Almasi *et al.* 2022

sheep were grazed on separated pastures, that were 16 and 30 ha, respectively. ActiGraph (wGT3X-BT; ActiGraph, LLC, Pensacola, FL) accelerometer sensors were attached to the left side of the muzzle to record activities. The animal management details of this experiment were described in Almasi *et al.* (2022, in press). The activities of grazing sheep, including grazing, rumination, walking, licking, idle and other behaviours, were recorded using video cameras and compared with predicted activity trained through a support vector machine (SVM learning) algorithm. SVM was used as our data were not linearly distinguishable, and thus kernel space embedding was appropriate. The SVM classification was based on 5 s epochs to distinguish the activity of animals. Only grazing and rumination activities were used in this study. The concordance between visual and predicted behaviours using SVM method for grazing and rumination were 90%±11 and 95%±10 (± SD), respectively. The SVM validation and its features are explained in detail in Sohi *et al.* (2022, in press).

Statistical analysis. The data were analysed using ASReml-R (Butler *et al.*, 2017). The sum of each trait per day (e.g. minutes grazing) of each animal was calculated and used in the model. In total, 3,607 and 3,604 observations for 147 animals were recorded for grazing and rumination activities, respectively. The repeatability and heritability were estimated using a univariate model:

$$y = Xb + Zu + Wc, \quad (1)$$

where y is a vector containing the response variable (grazing or rumination time), b is the vector of fixed effects of overall mean, sex (male and female) and day of experiment (day 1 to 26); u ~N(0, $A\sigma^2_a$), c ~N(0, $I\sigma^2_c$) and e ~N(0, $I\sigma^2_e$) are the vectors of random animal effects, permanent environment effects and residual effects, respectively. A is the relationship matrix calculated from a pedigree of 351 animals which was traced back for 5 generations and I is the identity matrix. X, Z, and W are design matrices relating observations to the corresponding fixed or random effects.

The heritability was estimated by dividing the estimated genetic variance (σ^2_a) on the phenotypic variance ($\sigma^2_a+\sigma^2_c+\sigma^2_e$). The overall repeatability was calculating by dividing the sum of σ^2_a and permanent environment variance (σ^2_c) on the total phenotypic variance.

For calculating the repeatability of a trait over two days, or the interclass correlation between measurements for the same animals on two different days, the observations for just two days were included in the model.

Results

The mean ± S.D. of time spent grazing and ruminating were 404.4±41.9 min/d and 330.6±25.1 min/d (Table 1). Both grazing and rumination had an overall moderate to high repeatability (0.44 and 0.48, respectively), while their heritabilities were not significantly different to zero. The rams grazed 333 min/d less ($P<0.001$) but their rumination time was 25 min/d more than hoggets ($P<0.05$).

Figure 1a shows the repeatability between days, from 1 to 24 days interval. The repeatability with one day interval (e.g. day 2 to 3) for both activities was over 0.68. By increasing the interval, the repeatability for both traits dropped to below 0.1. Figure 1b shows the daily mean ± SD of both grazing and rumination

Table 1. The mean, SD, and repeatability of grazing and rumination activities.

Trait	Number of records	Mean (min/d)	S.D.	Repeatability ± S.E.
Grazing	3,607	404.4	41.9	0.44±0.03
Rumination	3,604	330.6	25.1	0.48±0.03

activity. There was considerable variation among days in the amount of time spent grazing and rumination. Day 23 with 491 min/d had the highest grazing activity, while day 19 had the highest rumination activity with 371 min/d. The lowest grazing activity (309 min/d) occurred on day 1 and the lowest rumination activity was on day 25 with 272 min/d (Figure 1b).

Discussion

The current study showed that the repeatability of grazing and rumination activity of pasture kept sheep grazing was moderate to high, while heritability estimates were close to zero. These results imply that while the sensor technology was a reliable tool to measure feeding activities and animals had repeatable behaviour within the same week, but then the repeatability dropped to below 0.50 when the interval between measurements were longer than a week. Therefore, many measurements in a week could help reduce residual variation, while over a longer period of time changes in behaviour can be detected. The repeatability of grazing and rumination suggests that there is a role for assessing these traits for quantitative genetic analysis. Although there was a significant difference in grazing and rumination time between rams and hoggets, they were grazing on different paddocks, which could also affect their grazing and rumination time.

Byskov *et al.* (2017) showed that the repeatability and heritability of rumination traits in research and commercial dairy cows at four different stages of lactation were high and moderate, respectively. The repeatability results of consecutive days for rumination trait in the current study were close to the Research herd (0.76 vs 0.75) in Byskov *et al.* (2017), but the heritability estimation for the same trait was completely different. Whereas, in the current study, the permanent environment, a short period of measurement (26 days), and small sample size and low genetic connectedness between animals could be the main factors that reduced the estimated additive genetic variance. Nevertheless, the sheep showed similar grazing and rumination patterns the next day, which means the sensor technology and the SVM method were reliable for ranking animals according to their feeding activities. According to the reported results by Byskov *et al.* (2017) and Morretti *et al.* (2018) rumination is a moderately heritable trait that can be considered for selective breeding. Further, it is important to estimate and understand the genetic correlations between grazing behaviour traits and other economically important traits such as feed efficiency or methane emissions.

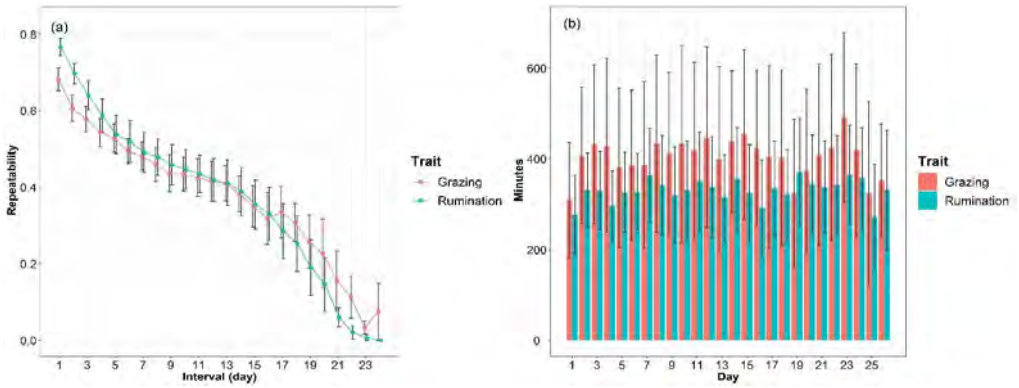


Figure 1. (a) The repeatability of grazing and rumination between days (from 1 to 24 days interval). The error bars represent SE. (b) The mean of grazing and rumination activity per day. The error bars represent SD.

Conclusions

Sensor technology and SVM method can be applied to determine grazing and rumination activities of sheep with potential application for breeding strategies. This is the first research study that has estimated the heritability and repeatability of these two traits in sheep. Further studies with a larger sample size are required to estimate the genetic variation in sheep in addition to correlations with traits of economic importance.

References

Almasi, F., Nguyen, H., Heydarian, D., Sohi, R., Nikbin, *et al.* (2022) Anim. Prod. Sci. In press.

Beauchemin, K.A., (2018) J. Dairy Sci 101(6): 4762-4784. https://doi.org/10.3168/jds.2017-13706

Butler, D.G., Cullis, B.R., Gilmour, A.R., Gogel, B.J., Thompson, R. (2017) VSN International Ltd, Hemel Hempstead, HP1 1ES, UK.

Byskov, M.V., Fogh, A., Løvendahl, P. (2017) J. Dairy Sci. 100(12):9635–9642. https://doi.org/10.3168/jds.2016-12511

Hancock, J., (1954) J. Agric. Sci. 45(1): 80–95. https://doi.org/10.1017/S0021859600045809

Moretti, R., de Rezende, M.P.G., Biffani, S., Bozzi, R. (2018) J. Anim. Breed. Genet. 135(4): 293–299. https://doi.org/10.1111/jbg.12346

Sohi, R., Almasi, F., Nguyen, H., Carroll, A., Trompf, J., *et al.* (2022) Anim. Prod. Sci. In press.

Welch, J.G., Smith, A.M., Gibson, K. S. (1970) J. Dairy Sci. 53(1):89–91. https://doi.org/10.3168/jds.S0022-0302(70)86153-7

149. Deep data recording for developing resilience traits in Swiss cattle populations: a proof of concept and an application

B. Bapst*, P. von Rohr and F.R. Seefried

Qualitas AG, Chamerstrasse 56, 6300 Zug, Switzerland; beat.bapst@qualitasag.ch

Abstract

The number of automatic milking systems in Switzerland is increasing. This possibility of in-depth phenotyping was used to investigate whether resilience traits could also be developed in Switzerland, analogous to Dutch studies. Based on a data set of Holstein, Simmental, Swiss Fleckvieh and Montbéliarde dairy cattle, similar heritabilities have been estimated and breeding values subsequently predicted. Breeding value comparisons indicate that dual-purpose breeds may be superior to dairy breeds in terms of resilience traits developed. But further research is needed to validate the results presented in this study. Generally, in-depth phenotyping with automatic milking systems offers new opportunities to develop new resilience traits.

Introduction

Breeding objectives have been changed from a pure yield oriented objective to a balanced one by the inclusion of functional traits (Miglior *et al.*, 2005). This trend has continued into the present day (Miglior *et al.*, 2017), with the addition of resilience traits (Brito *et al.*, 2021), even though this group of traits is not easy to define and to collect (Berghof *et al.*, 2019; Colditz and Hine, 2016). Because environmental changes are observed in small time windows, the apparent effects on yield, health, or welfare are difficult to capture. To address these information deficits, more frequent environmental information is necessary and in-depth phenotyping is essential. Automatic milking systems (AMS) provide in-depth phenotypes and can therefore support the development of resilience phenotypes. Poppe *et al.* (2020a) developed resilience traits based on daily milk yield by fitting lactation curves of individual cows and then measuring various deviation values from these curves as resilience traits. They observed medium to high heritabilities. Some of these new traits showed moderate to high positive correlations with other functional traits closely related to resilience.

Since the number of AMS is increasing rapidly in Switzerland, we investigated the following two aspects:

1. The first aim was to set up a pipeline which processes data recorded by AMS on a daily basis in order to analyse resilience traits developed by Poppe *et al.* (2020a) in Swiss dairy cattle populations.
2. Milk prices for producers have been decreasing in Switzerland. As a result, producers tend to switch to dual-purpose breeds to achieve higher income from meat products and to lower costs due to more resilient dual purpose breed cows. While the dairy cow population in Switzerland has been declining over the past few years, the number of dual-purpose cows remains constant or is gradually increasing. Since resilience traits belong to the group of functional traits, the question of interest is whether dual-purpose breeds are also genetically superior to dairy breeds with respect to the traits developed under (1)?

Materials & methods

Since 2018, it has been possible for Swiss farms using AMS to have their basic milking data, such as milk yield, automatically stored in a central database at Qualitas AG (https://qualitasag.ch).

While at the beginning the number of farms with AMS was rather low, now a strong increase can be observed and currently more than 1000 farms with at least one AMS unit are registered. Of these, 433 use the automatic data transfer system. This incoming data had to be processed and validated with a data

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_149, © B. Bapst *et al.* 2022

pipeline. In addition to routine plausibility checks, the following issues had to be solved: Data records that exceeded the day change had to be assigned to the individual days based on the assignment of the milking time length. Additionally, missing daily data were filled in based on previous and subsequent days, and outliers were corrected using the same principle.

A data set of Holstein cows (red and black) as dairy cattle (DC) and all Simmental, Swiss Fleckvieh, Montebéliarde, Normand cows and their crosses as dual-purpose cattle (2C) were included in this study. After data curation, a set of 11,427 phenotypic records was available from 9,744 cows (Table 1). First of all, a data pipeline was set up to calculate the daily milk yields and to filter based on data plausibility. After that, resilience traits were developed using the methods by Poppe *et al.* (2020a). To compare the EBV of DC and 2C, only the following resilience indicator was calculated because it showed the highest correlations to important functional traits in the mentioned study: Log-transformed variance (LVA) of the deviations in the daily milk yield from individually modelled cow lactation curves. As an additional control, the trait raw variance (RVA) was also evaluated. These cow individual curves were modelled with a polynomial (fourth order) quantile regression. To collect as much data as possible and to follow the Swiss regulations for performance tests (swissherdbook, 2021), only cows with lactations of at least 270 days in milk (DIM) were included in the study. Variance components and breeding value (EBV) estimations as well as the preceding pedigree preparation and renumbering were performed with programs from the BLUPF90 package (Misztal *et al.*, 2018). For routine genetic evaluation in Switzerland, the breeds from this study and other additional breeds are combined to form a common pedigree. Crossbreeds connect the different pure breeds. Therefore, for this study, we used the data here in a common genetic evaluation for DC and 2C. Traits were analysed univariately using the following animal model with repeated measurements. This model was derived from the study of Poppe *et al.* (2020a) and slightly modified:

$$y_{ijkl} = hys_i + agca_j + lac_k + a_l + pe_{kl} + e_{ijkl} \quad (1)$$

where y_{ijkl} is the resilience trait of animal l in lactation k with age at calving j in herd year season i. hys_i is the fixed effect of herd year season, $agca_j$ the fixed effect of age at calving, lac_k the fixed effect of the lactation, a_l the random animal effect, pe_{kl} the random permanent environmental effect and e_{ijkl} the random residual effect.

Since more resilient animals have smaller variances of deviations from the standard curve, a lower EBV represents reduced deviations and lower EBV are therefore desirable.

Table 1. Cows with phenotypes.

Breed	n	%[2]	Purpose[1]	%[2]
Simmental	16	0.2	2C	
Swiss Fleckvieh	549	5.6	2C	9.3
Montbéliarde	330	3.4	2C	
Normande	12	0.1	2C	
Holstein	8,837	90.7	DC	90.7
Total	9,744	100.0	DC	100.0

[1] 2C = Dual-purpose cattle; DC = Dairy cattle with phenotypes in %.
[2] Breed proportion of cows.

The breeding values were transformed to a mean value of 100 with a standard deviation of 12. The differences of the EBV variances between DC and 2C were evaluated with a F-test.

Results

For the first time, a workflow was developed to process Swiss AMS data on a daily basis. The variance components and the heritabilities from the evaluated traits are in Table 2. For both traits, the estimated heritabilities were almost identical, while the variances were different, as expected. The genetic correlation (r_a) between RVA und LVA was 0.69.

The comparison between EBV of DC and 2C shows that the distributions of the EBV are not equal (Figure 1). F-statistics (*P*-value=0.012) confirm this difference. The difference of the means between the EBV of 2C and DC is 4.2.

Discussion

The developed data pipeline is a valuable tool to aggregate individual AMS data sets into daily information. Based on this, evaluations can be developed, such as resilience traits. Although the data set is small, we have obtained preliminary evidence that resilience traits based on daily AMS records can be evolved and that the methods of Poppe *et al.* (2020a) are applicable to other cattle populations. Heritabilities were in a similar range, but variances are different. Poppe *et al.* (2020a) estimated larger variances. Reasons for this could be the different models. With the integration of 2C data, the opposite would have been expected. However, further research is needed to explain these differences in variances and in the genetic parameters

Table 2. Estimated variance components and heritabilities (SE in brackets).[1]

Trait	σ^2_a	σ^2_{pe}	σ^2_r	σ^2_p	h^2
LVA	0.021 (0.004)	0.046 (0.004)	0.049 (0.002)	0.116	0.181 (0.030)
RVA	23.749 (4.222)	49.051 (4.559)	61.484 (2.487)	134.284	0.177 (0.037)

[1] LVA = log-transformed variance; RVA = raw variance; σ^2_a = additive genetic variance; σ^2_{pe} = variance of the permanent environment; σ^2_r = residual variance; σ^2_p = phenotypic variance; h^2 = heritability.

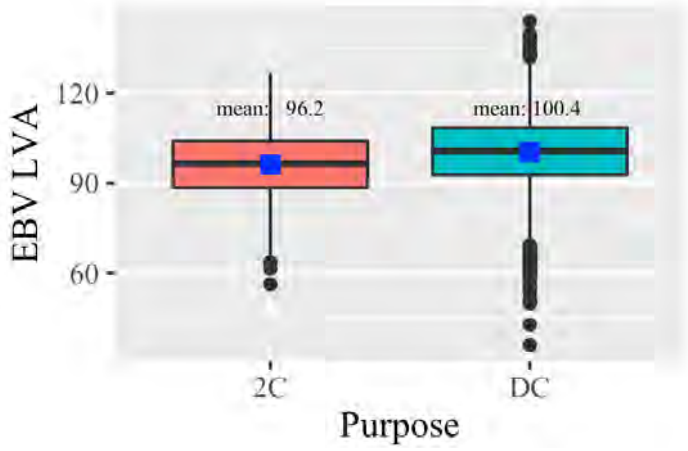


Figure 1. Comparison of the EBV distribution for 2C (dual-purpose cattle) and DC (dairy cattle) for the trait LVA; blue box: mean of the purpose groups.

derived from them. Additionally, it is surprising that r_a between RVA und LVA was larger in our study than in Poppe *et al.* (2020a). Therefore it needs to be verified that a model with repeated measurements gives the best fitting results. Poppe *et al.* (2020b) have found that LVA is not quite the same trait in second and higher lactations. Due to the limited data set, a repeatability model was used in this study. In addition, correlations with other functional traits need to be investigated. Cows with different breed proportions were included in the study. Therefore it must be checked whether non-additive effects should be taken into account (Van Der Werf and De Boer, 1989).

The current results suggest that 2C are better than DC in terms of the mean EBV of the resilience trait studied. Nevertheless, this needs to be validated with a larger data set and in addition, the differences between the levels of the EBV should be tested with a sensitive statistical procedure, such as a permutation test.

Acknowledgements

We would like to acknowledge Holstein Switzerland and swissherdbook for the permission to use their data for this study.

References

Berghof T. V. L., Poppe H. W. M. and Mulder H. A. (2019) Front. Genet. 9 (January): 1–15. https://doi.org/10.3389/fgene.2018.00692

Brito L. F., Bedere N., Douhard F., Oliveira H. R., Arnal M. *et al.* (2021) Animal 15(Supplement 1): 100292. https://doi.org/10.1016/j.animal.2021.100292

Colditz I. G. and Hine B. C. (2016) Anim. Prod. Sci *56*(12): 1961–1983. https://doi.org/10.1071/AN15297

Miglior F., Fleming A., Malchiodi F., Brito L. F., Martin P. *et al.* (2017) J. Dairy Sci. 100(12): 10251–10271. https://doi.org/10.3168/jds.2017-12968

Miglior F., Muir B. L. and Van Doormaal B. J. (2005) J. Dairy Sci. 88(3): 1255–1263. https://doi.org/10.3168/jds.S0022-0302(05)72792-2

Misztal I., Tsuruta S., Lourenco D. A. L., Masuda Y., Aguilar I. *et al.* (2018) Manual for BLUPF90 family programs. http://nce.ads.uga.edu/wiki/doku.php?id=documentation

Poppe M., Veerkamp R. F., van Pelt M. L. and Mulder H. A. (2020a) J. Dairy Sci. 103(2): 1667-1684. https://doi.org/10.3168/jds.2019-17290

Poppe M., Bonekamp G., van Pelt M. L. and Mulder H. A. (2020b) J. Dairy Sci. 104(3): 1967-1981. https://doi.org/10.3168/jds.2020-19245

swissherdbook (2021) Reglemente Leistungsprüfungen. Available at: https://www.swissherdbook.ch/fileadmin/Domain1/PDF_Dokumente/05-Statistiken-Formulare/58-Reglemente/Leistungspruefungen/3100.01_Reglement_Leistungspruefungen_2021-03-04_unterschrieben.pdf

Van Der Werf J. H. J. and De Boer, W. (1989) J. Dairy Sci. 72(10): 2606–2614. https://doi.org/10.3168/jds.S0022-0302(89)79401-7

150. Test-day genetic evaluations: a tool to measure herd resilience through monthly milk records

S. Mattalia[1,2*], M. Rocland[1,2], S. Aguerre[1,2] and M. Poppe[3]

[1]Institut de l'Elevage, UMT eBIS, 78350 Jouy-en-Josas, France; [2]Université Paris Saclay, INRAE, AgroParisTech, GABI, Domaine de Vilvert Bat 211, 78350 Jouy-en-Josas, France; [3]Wageningen University and Research, Animal Breeding and Genomics, P.O. Box 338, 6700 AH Wageningen, the Netherlands; sophie.mattalia@idele.fr

Abstract

This study aimed to explore a herd resilience indicator, LnVar, estimated from monthly test-day records. For each Holstein cow in first parity between 2014 and 2018, the expected milk yields were computed at each test-day, using the genetic and environmental effects estimated from a routine test-day genetic evaluation. LnVar was then estimated for each herd-year, as the natural logarithm of the variance within lactation of the deviation between the expected and the observed milk yields. LnVar presented a high variability and was quite repeatable between years. The herd-years with high LnVar were more productive and showed a greater risk of high somatic cell counts, clinical mastitis and short lactations. This study showed that LnVar is a relevant resilience indicator, even with monthly records. The genetic evaluations using test-day records could provide useful information for herd management.

Introduction

Cows are said to be resilient when they are able to maintain their performances throughout environmental disturbances or recover quickly if they are affected. Management practices can have an impact on these disturbances and the cows resilience. Poppe *et al.* (2019, 2020) showed that the variability of fluctuations around the expected lactation curve, measured through the natural logarithm of the variance of the deviation between the observed and the expected daily milk yields (LnVar) was a promising proxy to measure resilience at the cow and herd levels. However, these studies were based on daily records obtained by automatic milking systems, which limits the potential use of this indicator in commercial farms.

In France, expected lactations are provided by Institut de l'Elevage twice a year to the Milk Recording Organizations (MROs), in order to predict the production of each herd. Predictions use the genetic and environmental effects estimated from a genetic evaluation based on a test-day model. These effects can also be used to construct the expected lactation curve of the cows with past records and knowing their own parameters (year, age and season of calving).

In this study, a herd resilience indicator (LnVar) was estimated for each French Holstein herd using monthly test-day records. The objective was to assess to which extent the conclusions of Poppe's study were still valuable with less regular records, and whether a genetic evaluation could provide information to produce new herd management tools related to resilience.

Materials & methods

Description of the genetic evaluation. A genetic evaluation of Holstein cows was performed in October 2019 on milk yield, based on a test-day model described in Leclerc *et al.* (2008, 2009). This evaluation used all test-day records collected by MROs on Holstein cows in lactations 1 to 3 and having started their first lactation since September 1st, 1989 (more than 14 Mo cows). A random regression test-day model in which the first 3 lactations were considered as correlated traits was applied. In addition to a herd-test-day effect, the shape of the lactation curve defined by parity-region was taken into account, cumulating cubic splines of each level of the effects of calving month, calving age, length of dry period and gestation. Genetic and

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_150, © S. Mattalia *et al.* 2022

permanent effects were included using a reduced rank model. Days in milk, region-year-parity and herd-year were used to model the residual variance.

Selection of cows and herd-years. We selected test-day records of French Holstein cows in first parity and included in the genetic evaluation. The lactations should have started between September 1st, 2014 and August 31st, 2018 and have at least 2 test-days. Cows should have performed their lactation in one single herd. Only performances observed before 350 days in milk (DIM) were considered and were assigned to a given year corresponding to the calving year (Year n from September n-1 to August n). Selected herd-years should have recording intervals of less than 5 weeks and at least 10 cows in first lactation.

35,719 herds-years met all requirements, corresponding to 711,485 cows (on average 19.9±10 cows per herd-year). Each cow had on average 10.3±1.4 test-day records.

Estimation of LnVar. For each herd-year, LnVar was estimated as follows: first, for a given cow j in herd i and with a performance recorded during the test-day k, the expected milk yield was estimated by the sum of all estimated effects affecting the cow at this test-day. Then the deviation δ_{ijk} was estimated between the observed and the estimated milk yields. The variance of these deviations was used to estimate the individual fluctuation of the lactation curve around its expectation. It was log-transformed in order to be normally distributed:

$$LnVar_{ij} = Ln\ (Var(\delta_{ij}))$$

Lastly, the indicator LnVar of the herd-year i was estimated as the mean of the individual fluctuations $LnVar_{ij}$ of the cows with test-day records in this herd-year.

Herd-year descriptors. Several descriptors were computed to describe herd-years regarding their production level (average milk yield within 305 days MY) and their resilience and health status (percentage of short lactations of less than 90, 120 and 200 days PSL90, PSL120 and PSL200, percentage of cows with at least one test-day above 300,000 somatic cells/ml P300kcel, or with at least one clinical mastitis observed by the farmers PMacl).

Analyses. Pearson correlations were estimated between consecutive years. The association between LnVar and the herd-year descriptors was studied through Pearson correlations and using a one factor linear model (SAS® GLM procedure), in which we tested the effect of classes of LnVar (5 classes, from very low to very high LnVar) on each herd-year descriptor.

Results & discussion

LnVar presented a high variability, with extreme values at more than 4 standard deviations from the mean and a coefficient of variation of 0.26 (Table 1). LnVar was quite repeatable across years (Table 2), which means that the profile of fluctuations in lactation curves of a given herd was only partially affected by one-time events.

Correlations between herd-year descriptors and LnVar were low or moderate (Table 3). Herd-years with high LnVar tended to be more productive and more vulnerable to udder infections (correlations of 0.27 and 0.21 between LnVar and MY and P300kcel respectively).

Table 1. Descriptive statistics of the herd resilience indicator LnVar.

	N	mean	std	min	max
LnVar	35,719	1.26	0.33	-0.12	2.97

Table 2. Correlations between LnVar estimated in consecutive years.

2015-2016	2016-2017	2017-2018
0.49	0.50	0.52

Table 3. General characteristics of herd-year descriptors and correlations with LnVar.

	Abbreviation	Unit	mean	std	Cor. with LnVar
Milk Yield	MY	kg	8,719	1,413	0.27
% of short lact. (<90 days)	PSL90	%	1.4	3.3	0.03
% of short lact. (<100 days)	PSL120	%	2.3	4.5	0.05
% of short lact. (<200 days)	PSL200	%	4.9	7.5	0.08
% of lact. with ≥1 test day with ≥300,000 cells	P300kcel	%	39.1	16.5	0.21
% of cows with ≥1 clinical mastitis within 1st parity	PMacl	%	11.1	13.4	0.04

All effects of LnVar classes on the herd-year descriptors were significantly different from zero (α<1‰). The tendencies mentioned before were confirmed, with large differences between classes illustrated by the contrasts with the middle class (Table 4). The average MY of the highest class was 1,742 kg higher than that of the lowest one and 523 kg higher than that of the middle one. The risk of having at least one test-day above 300,000 cells/ml was 1.55 times higher in herd-years with very high LnVar than in those with very low LnVar (52.1 vs 33.6%). Differences were observed for PMacl, even though the correlation with LnVar was weaker. The percentage of short lactations was higher for both extreme classes, but the risk increased more after 120 DIM for classes with high LnVar than for those with low LnVar (+2.1 vs +0.8% between 90 and 200 days for herd-year classes with very high vs very low LnVar).

Discussion

Our indicator was based on the same principles as described in Poppe *et al.* (2020), with several differences. The first one was the use of monthly test-day records instead of daily data. Large differences could be observed

Table 4. Means of herd performance indicators in different herd-year classes based on LnVar (compared to the middle herd-year class, standard errors within brackets).[1]

	Classes of herd-years based on the LnVar level [2]				
	Very low	Low	Middle	High	Very high
N	643	4,700	25,022	4,345	1,007
MY	-1,219 (55)**	-685 (22)**	0	+466 (23)**	+523 (44)**
PSL90	+0.7 (0.13)**	+0.16 (0.05)**	0.0	+0.20 (0.05)**	+0.3 (0.11)**
PSL120	+0.8 (0.18)**	+0.2 (0.07)	0.0	+0.4 (0.07)**	+0.9 (0.07)**
PSL200	+1.5 (0.30)**	+0.0 (0.12)	0.0	+0.9 (0.12)**	+2.4 (0.23)**
P300kcel	-4.9 (0.64)*	-4.1 (0.26)**	0.0	+6.2 (0.27)**	+13.5 (0.52)**
PMacl	-2.8 (0.53)**	-1.5 (0.21)**	0.0	+0.9 (0.22)**	+0.5 (0.43)

[1] * $P<10^{-3}$; ** $P<10^{-4}$.

[2] Very low: ≤0.6; low: [0.6; 0.93]; Middle: [0.93; 1.59]; high: [1.59; 1.92]; very high: >1.92.

between herd-years, even with less regular performances, and these differences were quite repeatable across years. The second originality was that the expected milk yields at each test-day were based on the effects estimated within a test-day genetic evaluation. Thanks to this approach, expected milk yields were adjusted for the cow genetic abilities (individual peak and persistency), and for the environmental effects estimated at a given stage of lactation. These parameters are currently available for any herd participating to milk recording and can be updated at each new evaluation.

The relationship between LnVar and herd descriptors were consistent with Poppe *et al.* (2020): the greater the fluctuations in lactations (higher LnVar), the more productive the herds and the greater the risk of udder health problems (high somatic cell counts and more clinical mastitis (CM) events). For PMacl, the low correlation with LnVar and the fact that the effect did not increase between the two highest classes of LnVar was probably due to a lack of comprehensiveness of registered CM events in some herds (Govignon-Gion *et al.* 2012).

The relationship between the percentage of short lactations and LnVar was more complex to interpret, since lactations may have been closed for many reasons. A large part of very short lactations were probably due to a low production level, particularly in the herds with a low productivity, and thus in the class with very low LnVar. But Beaudeau *et al.* (1995) showed that early lactation (before the peak of lactation) and end of lactation were the main periods of decision making for culling. Thus short lactations closed after the pick (between 90 and 200 days) likely corresponded to involuntary culling, and the increase in occurrences of short lactations with high LnVar could be interpreted as a lack of resilience.

Conclusions

This study showed that even with monthly data, it is possible to build herd resilience indicators based on the fluctuations of lactation curves, and that the results of a test-day genetic evaluation can provide useful information for herd management. Further investigations should be undertaken, in order to assess the genetic variability of LnVar estimated at the cow level and whether this proxy could be used to select more resilient cows.

Acknowledgements

We acknowledge European Union's Horizon 2020 Research and Innovation program GENTORE under grant agreement No. 727213, for financial support. In addition, we acknowledge Claudia Kamphuis (Wageningen Livestock Research, Wageningen, Netherlands) for her intellectual input, and French MROs for providing the data.

References

Beaudeau F., Ducrocq V., Fourrichon C. and Seegers H. (1995) J. Dairy Sci. 78(1):103-117 https://doi.org/10.3168/jds.S0022-0302(95)76621-8

Govignon-Gion A., Dassonneville R., Baloche G. and Ducrocq V. (2012) Interbull Bull. 46;121-126. https://journal.interbull.org/index.php/ib/article/view/1276

Leclerc H., Duclos D., Barbat A., Druet T., and Ducrocq V. (2008) Animal 2(3):344–353. https://doi.org/10.1017/S175173110700119X

Leclerc H., Nagy I. and Ducrocq V. (2009) Interbull Bull. 40:42–46. https://journal.interbull.org/index.php/ib/article/view/986

Poppe M., Veerkamp R. F., van Pelt M. L. and Mulder H. A. (2019) J. Dairy Sci. 103(2):1667-1684. https://doi.org/10.3168/jds.2019-17290

Poppe M., Mulder H.A., Kamphuis C. and Veerkamp R.F. (2020) J. Dairy Sci. 104(1):616-627. https://doi.org/10.3168/jds.2020-18525

151. Genomic selection of postpartum anoestrus recorded with accelerometer collars to improve beef cattle fertility

E. Fontanillas[1*], V. Prieur[2], A. Michenet[1], G. Auvray[1] and J. Mante[2]

[1]IngenomiX: Pôle de Lanaud, 87220 Boisseuil, France; [2]France Limousin Sélection: Pôle de Lanaud, 87220 Boisseuil, France; eric.fontanillas@ingenomix.fr

Abstract

A key genetic component of the female fertility, the calving-to-first-heat interval (CFHI), was evaluated on Limousin cattle. In a large-scale experimental design, 3,923 cows spread over 35 farms were phenotyped during 4 years. Accelerometer collars were used to detect first heat after calving. After determining the non-genetic effects on CFHI, the evaluation of the genetic potential of the animals was then carried out using a two-step genomic selection model. Heritability of the CFHI was evaluated at 6% and predictive ability, evaluated on cross validation on direct genomic values, was around 0.16. Regarding these results, together with the high variability of the phenotype and the substantial size of our reference population, we are able to set up a genomic selection program for CFHI on French Limousin breed.

Introduction

Cows' fertility is of major importance for the breeders' income in both beef and dairy cattle (Berry *et al.*, 2014; Taylor *et al.*, 2018). The reproduction efficiency is related to two aspects: the age at first calving and the calving interval (Berry and Evans, 2014; Roughsedge *et al.*, 2005). The age at first calving is genetically related to the age at puberty of the heifer. The calving interval (CI) can be impacted by breeding practices, like the management of cows by groups for natural service (insemination rate is 10% in the French Limousin breed). The CI can be split into several components (Burns *et al.*, 2010; Flores, 1971). First, the calving to first heat interval (CFHI) corresponds to the postpartum anoestrus. Second, the interval from the first heat to first mating is mainly related to breeding practices (Donoghue *et al.*, 2004). Then, the number of reproduction events until success is crucial. It is a very low heritable trait which is already evaluated from the French national records of inseminations (Idele, 2014). At last, the gestation length leads to the calving.

The CFHI is the most heritable trait among fertility predictors (Berry *et al.*, 2014). However, as studied by Mialon *et al.* (2000) in the Charolais breed, detecting the heat requires human observation or progesterone dosage which cannot be routinely recorded in commercial farms. Sensors can help to get more accurate heat detection (Schöpke and Weigel, 2014).

Since the last decade, the French Limousin breeding society has been developing on-farm data recording in collaboration with voluntary breeders gathered in a farm network. Several traits of particular interest for the breeders, like pelvis opening or udder shape, were recorded and led to the development of genomic selection on these traits. In 2016, an original project called 'COWnect' was launched with data collection of postpartum first heat using sensors. We report here results for heritabilities and genomic predictive ability for CFHI.

Materials & methods

Phenotype records. A total of 3923 cows from Limousin breed were recorded on the 35 farms of the 'COWnect' farm network over 4 years. The first component of CFHI is the calving date which is considered as a reliable record. The second component is the first heat event consecutive to calving, which was recorded by analysing onsets of activity using the accelerometer system collar (Heatime; SCR Engineers Ltd., Netanya, Israel). Quality controls were performed to keep realistic records of CFHI based on the

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_151, © E. Fontanillas *et al.* 2022

cyclicity of oestrus and the average gestation length. Twin calving records were discarded. After filtering steps and quality controls, 5,265 records of CFHI were collected for 3,181 cows between May 2017 and May 2021 (Figure 1a.). Two levels of filtering the CFHI records were tested: no filter and filtering with a minimum value of 20 days and a maximum of 160 days (Figure 1b.).

Genotypes. A total of 2,429 animals with performance were genotyped on 11 various chips from 26,151 SNP to 138,894 SNP. After quality control (call rate of 1%, Hardy Weinberg Equilibrium *P*-value 5%), these animals were all imputed to a target map of 43,784 SNP using Beagle v5.1 (Browning *et al.*, 2018).

Genetic model. The genetic model included repeated records from the same animals for successive calving records. A five-generation pedigree was extracted for each animal with a performance record. AIREML was run to estimate genetic parameters with sample standard deviation estimation (sd). Sensibility of the CFHI to the extreme values (below 20 days or above 160 days) was tested with heritability estimation on a non-filtered or filtered dataset (Table 1). Estimated breeding values (EBV) were computed with a pedigree BLUP (best linear unbiased predictor) based on filtered phenotypes. The BLUPF90 software package (Misztal *et al.*, 2002) was used for both purposes.

Genomic model. The EBV from 2,429 animals with phenotype and genotype were used as a reference population. A BayesC approach (Habier *et al.*, 2011) was used for the genomic model with a proportion of markers π assumed to have an effect at each iteration was fixed or estimated by the model. The π value was either estimated by the model, or fixed to several values: 0.01, 0.1, 0.2 and 0.5. A total of 100,000 iterations were performed with a burn-in period of 20,000. The GS3 software (Legarra *et al.*, 2013) was used for the analyses.

Accuracy of the genomic model was assessed using cross validation k-fold procedure including 8 validation population samples randomly defined. Pearson correlations were computed between direct genomic value (DGV) and phenotypes corrected from environmental effects (averaged for repeated measures).

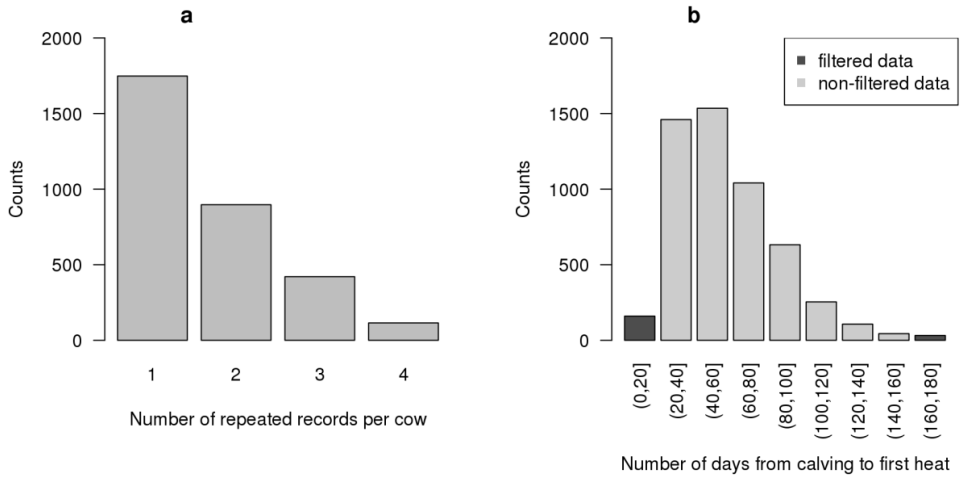


Figure 1. Description of the dataset: (a) counts of repeated records per cow; (b) distribution of calving to first heat intervals.

Results

Environmental factors detected as having a significant effect on CFHI were the contemporary group (made of herd × campaign of calving from September to August), the month of calving and the interaction of parity of the cow and its age at the first calving.

The removal of extreme CFHI values in the filtered dataset led to a lower phenotypic variance (Vp=706.2), a lower residual variance (Ve=546.2) and a slightly higher heritability (h^2=0.06) compared to the estimations made from non-filtered dataset (Vp=813.3; Ve=625.7 and h^2=0.05). The repeatability estimate was 0.07 in both cases. Moreover, the average reliability (i.e. CD) of the 2,429 EBV was 0.16 for the animals included in the genomic model.

Predictive ability of the genomic model, computed as correlation between predicted (DGV) and real information (deregressed phenotype) in cross-validation, was between 0.144 and 0.167 depending on the π parameter of the Bayes C model (Table 2.). The inclusion of the π as a random variable estimated by the Bayesian model gave the best accuracy. The posterior estimation of the π in this model is 0.29 with a sample standard deviation of 0.11.

Discussion

The CFHI trait definition implied several quality controls of the data which had a critical impact on heritability estimates. The assumption of non-realistic CFHI below 20 days related to the oestrous cycles (Forde *et al.*, 2011) led to slightly higher heritability but the estimates overlap given their standard deviations. Similarly, extremely long CFHI impacted the quantitative model while it might be more related to the breeding practices (based on breeder survey).

The best result of heritability estimated in this study is higher than the other traits of female fertility available until now for the French Limousin breed (Idele, 2014). The same trait studied in Holstein breed based on progesterone dosage as a proxy of commencement of luteal activity (Berry *et al.*, 2014) led to higher heritability (0.16-0.18). However, progesterone dosages are highly human-time consuming and are not sustainable for automated phenotypic recording. Similar results were observed by Mialon *et al.* (2000) on the Charolais breed, where CFHI (full-time human supervision) and postpartum-interval-to-

Table 1. Sensibility to extreme values of calving-to-first-heat interval (CFHI). In the filtered dataset, CFHI lower than 20 days and higher than 160 days were discarded from the analysis.

Datasets	Non-filtered	Filtered
Sample size	5,265	5,097
Heritability (standard deviation)	0.05 (0.02)	0.06 (0.02)
Repeatability	0.07	0.07

Table 2. Correlations from cross-validation for CFHI with a Bayes C model.

π value	Correlation from cross-validation (standard deviation)
0.01	0.144 (0.06)
0.1	0.162 (0.04)
0.2	0.162 (0.05)
0.5	0.159 (0.04)
Estimated	0.167 (0.04)

first-positive-progesterone test were measured on the same animals with respective heritability of 0.12 and 0.35. Heritability obtained in this study was lower but showed that the accelerometer system collar allows recording this performance in an automated way.

The precision of genomic selection is in the range of what is expected considering the heritability of the trait and the size of the reference population. Toghani *et al.* (2017), studied three fertility traits on composite beef cattle. Using the same Bayes Cπ model, validation correlations were equivalent for a trait with similar heritability.

These results are promising for the implementation of the genomic selection for female fertility which will become available in the next release for the French Limousin breed. This will be the first fertility trait genomic evaluation which exploits data from the 'COWnect' project. With low accuracy, extreme animals could be detected and culled. More data is needed to increase the reference population size. The development of sensors in commercial farms is a way to build up a larger database, considering that the long and thorough work of setting up the procedure of data exchange and collection has already been done.

To follow up on the topic, studying other components of the calving interval such as the gestation length or precocity is scheduled to build a package on genomic prediction for female reproduction.

References

Berry D.P. and Evans R.D. (2014) J Anim Sci 92(4):1412-22. https://www.doi.org/10.2527/jas.2013-6723

Berry D.P., Wall E., and Pryce J.E. (2014) Animal 8:s1:105-121. https://www.doi.org/10.1017/S1751731114000743

Browning B.L., Zhou Y., and Browning S.R. (2018) Am J Hum Genet 103(3):338-348. https://www.doi.org/10.1016/j.ajhg.2018.07.015

Burns B.M., Fordyce G. and Holroyd R.G. (2010) Anim Reprod Sci 122(1-2):1-22. https://www.doi.org/10.1016/j.anireprosci.2010.04.010

Donoghue K.A., Rekaya R., Bertrand J.K, and Misztal I. (2004) J Anim Sci 82(2):362-7. https://www.doi.org/10.2527/2004.822362x

Flores. (1971) A study of calving interval. PhD Dissertation from Iowa State University, USA.

Forde N., Beltman M. E., Lonergan P., Diskin M., Roche J. F. *et al.* (2011). Anil Repro Sci 124(3-4), 163-169. https://doi.org/10.1016/j.anireprosci.2010.08.025

Habier D., Fernando R.L., Kizilkaya K., and Garrick D.J. (2011). BMC Bioinformatics. 12:186. https://www.doi.org/10.1186/1471-2105-12-186

Idele, French Livestock Institute. Note IBOVAL 68. (2014) Available at: https://idele.fr/?eID=cmis_download&oID=workspace://SpacesStore/9c1536a9-e7cc-4f14-947f-3b45b5f28354

Legarra, A., Ricard, A., and Filangi, O. (2013). GS3, software. Available at: https://github.com/alegarra/gs3

Mialon M.M., Renand G., Krauss D., and Ménissier F. (2000) GSE 32:403-14. https://www.doi.org/10.1186/1297-9686-32-4-403

Misztal I., Tsuruta S., Strabel T., Auvray B., Druet T. *et al.* (2002) Proc. of the 7th WCGALP, Montpellier, France.

Roughsedge T., AmerP.R., Thompson R., and Simm G. (2005) J Anim Sci 83(10):2319-29. https://www.doi.org/10.2527/2005.83102319x

Schöpke K. and Weigel K.A. (2014). Interbull Bulletin 48:68-72. https://journal.interbull.org/index.php/ib/article/view/1348/1419

Taylor J.F., Schnabel R. D., and Sutovsky P. (2018) Animal 12:S1:s172-s183. https://www.doi.org/10.1017/S1751731118000599

Toghiani S., Hay E.,Sumreddee P., Geary T.W., Rekaya R. *et al.* (2017). J Anim Sci 95(11):4787-4795 https://www.doi.org/10.2527/jas2017.1944

152. Developing precision livestock farming in practice: using sensor time series data for breeding decision support systems

J.M.E. Statham[1,2,3*], K.L. Burton[1], I. Adriaens[1,4,5,6], I. Lora[7], G. Cozzi[7], Y. de Haas[6], C. Kamphuis[6], L. Vedder[8], B. Loke[8] and N.C. Friggens[9]

[1]RAFT Solutions Ltd., Sunley Raynes Farm, Galphay Road, Ripon, HG4 3AJ, United Kingdom; [2]Harper and Keele Veterinary School, Keele University and Harper Adams University, Staffordshire, United Kingdom; [3]Bishopton Veterinary Group, Mill Farm, Studley Road, Ripon HG4 2QR, United Kingdom; [4]KU Leuven, Dept of Biosystems, Biosystems Technology Cluster, Campus Geel, Kleinhoefstraat 4, 2440 Geel, Belgium; [5]KU Leuven, Dept of Biosystems, Division Mechatronics, Biostatistics and Sensors, Kasteelpark Arenberg 30, 3001 Leuven, Belgium; [6]Wageningen Research, Breeding and Genomics, Droevendaalsesteeg 1, 6708 PB Wageningen, the Netherlands; [7]Department of Animal Medicine, Production and Health, University of Padova, Viale dell'Università 16, 35020 Legnaro (Padova), Italy; [8]Noldus Innovation Works, Nieuwe Kanaal 5, 6709 PA Wageningen, the Netherlands; [9]UMR Modélisation Systémique Appliquée aux Ruminants, INRAE, AgroParisTech, Université Paris-Saclay, 16 Rue Claude Bernard, 75005 Paris, France; jonathan.statham@raftsolutions.co.uk

Abstract

The objective of GenTORE is to develop innovative genome-enabled selection and management tools to optimise cattle resilience and efficiency in widely varying environments. These tools incorporate both genetic and non-genetic variables, aiming to increase the economic, environmental and social sustainability of European cattle meat and milk production systems. Using available on-farm technology allows large-scale phenotyping of resilience and efficiency that can be applied to evidence-based management, breeding and culling decisions. Veterinarians and other farm advisors are engaged with farm business drivers that are influenced by consumer and societal demands including the environment, human health concerns regarding antimicrobial resistance and animal welfare. Conflicts exist in balancing these factors. Evidence-based tools to support herd-level strategy are lacking. This work describes how multiple streams of sensor data can be combined to inform herd-level strategy in a time-efficient, automated and objective system to support advisor input to herd health.

Introduction

The GenTORE project aims to develop innovative genome-enabled selection management tools to optimise cattle resilience and efficiency in widely varying environments.

Current decision tools for breeding often fail because they ignore the complex interactions between production, health, genotypic predisposition, and reproduction performance at a specific herd population level; they also require significant time and are not user friendly for advisors or at farm level.

This paper outlines how the management tools developed through GenTORE can help farmers, veterinarians and other advisors to balance the four key drivers in sustainable dairy and beef farming (Figure 1).

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_152, © J.M.E. Statham *et al.* 2022

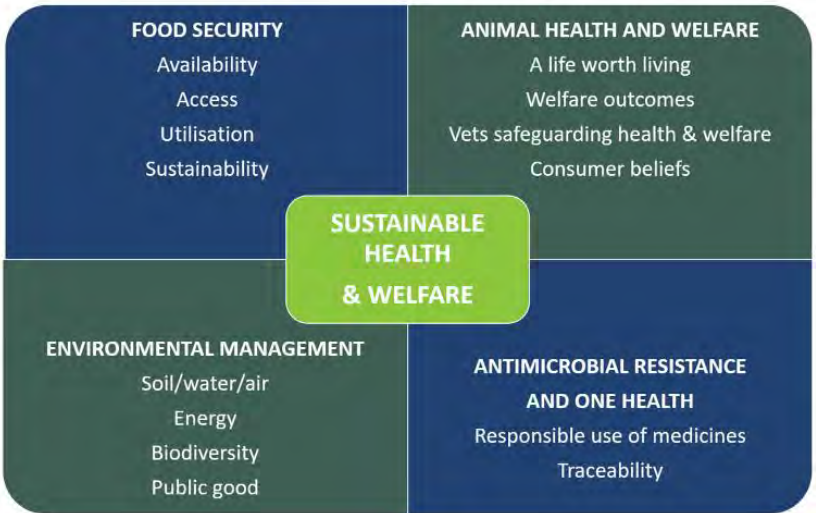


Figure 1. The sustainable herd health balance for farm advisors.

Balancing the drivers for evidence-based herd decision-making

Veterinarians and other farm advisors are engaged with the farm business drivers described below that are influenced by consumer and societal demands. Conflicts exist in balancing these factors. Precision Livestock Farming (PLF) offers opportunities for innovation in evidence-based decision-making for farmers, veterinarians and advisors. A balance must be achieved in sustainable dairy and beef farming between four key drivers (Figure 1):

1. 'One Health';
2. environmental impact;
3. animal health and welfare;
4. food security.

'One Health'. The concept of 'One Health' encompasses animal health, human health and the environment. One Health is ever present in the global fight against the development of antimicrobial resistance. It is already the cause of 700,000 human deaths every year (O'Neill 2016). Decisions to administer antimicrobials to cattle are critically important, but often highly subjective and not risk-based.

Environmental impact. The importance of mitigation of the impacts of dairy farming on the environment are increasingly recognised (Statham *et al.* 2017; Statham *et al.* 2020). In 2010 the Food and Agriculture Organization of the United Nations (FAO) produced a 'lifecycle assessment' of the dairy sector which was proposed to contribute 4% to the total global man-made greenhouse gas emissions. Improved efficiency of milk production will reduce negative environmental effects; if fewer cows (and replacements) are required and there are fewer 'lost' litres of milk and a reduction in greenhouse gases and resources per litre of saleable milk. 'Lost' milk here includes milk that does not enter the food chain following animal treatment or a reduction in yield that occurs following clinical or subclinical disease or poor reproductive performance. The difference in environmental impact between the best and poorest performing herds in terms of health and fertility is likely to be very large; in terms of fertility alone reductions in methane emissions of the order of 25% appear to be possible between the best and poorest herds (Garnsworthy 2004; ADAS 2013).

Animal health and welfare. It is important that the connections are clearly made for welfare as well as health in managing the sustainability of food production (Broom *et al.* 2013).

Food security. Food security is not limited solely to the quantitative aspects of food supply. Three fundamental dimensions are applicable to products of animal origin: availability of food, access to food, and effective and safe utilisation of food (Bonnet *et al.* 2011). Sustainability refers to the capacity for sustainably maintaining agricultural production (including animal production), national and international trade, storage and supply and food consumption that will meet the demand in the long term and even in the context of new constraints.

Herd level breeding and health decisions can be supported by on-farm sensor time series data presented as proxies of resilience and efficiency

Evidence-based tools to support herd-level strategy setting are lacking. However multiple streams of sensor data can be combined in a time-efficient, automated and objective system to support advisor input to herd health. Climate change, with its increasing frequency of environmental disturbances puts pressures on the livestock sector (Hansen *et al.* 2012). More complex traits such as resilience must be considered in our management strategies and breeding programs (Friggens *et al.* 2022). Resilient animals respond well to environmental challenges, are more likely to overcome them without assistance and therefore have better health and welfare. In practice there remains a lack of operational measures of resilience that can be deployed at large scale across different farm types and livestock species. Such measures are needed to provide more precise phenotypes of resilience for use in farm management and animal breeding decisions. Decisions can be made at individual animal level, but when combined can inform critical changes in herd level policies and protocols.

Cow resilience and efficiency are two key traits that can be used for decision making. Sensor-based proxies for phenotypes offer an important dimension to application of genomics data. Novel proxies for complex phenotypes were developed in Horizon 20-20 project 'GenTORE' for resilience using already existing at-market technology. Definitions and ways to calculate resilience outcomes were lacking; methodology had to be objective and valid so appropriate ranking of the cows for these traits on farm was possible (Adriaens *et al.* 2020). For resilience, a combination of longevity, health, robustness, production, and reproduction performance had to be considered. This resulted in the following definition for ranking cows: 'resilience is the cumulative result of a cow's ability to re-calve' (and thus extend her productive lifespan), supplemented with secondary corrections for age at first calving, calving intervals, 305-day milk yield, health events and number of inseminations. Different weightings of the included variables were proposed and cross-validated to produce a common equation to calculate the resilience scores. The highest weight was given to the number of parities. One of the main challenges in this work was the availability, completeness and reliability of the data (Adriaens *et al.* 2020).

Sensor technologies record parameters reflecting animals' performance, behaviours and physiological state. Besides the classical use for event detection and 'alerts', time series data from sensors can be characterized to provide information on how animals have performed over time. Two types of sensor features were defined. Type A sensor features generally characterize the sensor measurements relative to the herd peers, and include milk yield, body weight, rumination and activity. Type B sensor features specifically target the identification of dynamics in the time series indicating certain health or fertility events. Several primary algorithms to predict the resilience and efficiency rankings from the sensor features were tested. These algorithms were deployed to design software that could signal herd level change in resilience and thus provide an evidence base for practical changes in herd management policies.

Combining individual level data from multiple technologies in a form that is useable in real world situations in real time requires an additional integration step to deal with issues such as synchrony of multiple proxy indications at animal level. Statistical methods to deal with these issues and combine proxies into one reliable indicator have been proposed (Adriaens *et al.*, 2020). GenTORE partner 'Noldus' have developed a demonstration web framework using the software Blazor (https://demos.telerik.com/blazor-ui/; Microsoft, Redmond, USA) that outlines the potential of this combination of time-series sensor data as herd level resilience and efficiency based tools: Comparing herd resilience and efficiency score against a group benchmark; Visualization of four main component scores (Health, Fertility, Production and Herd Inventory) and a detailed view of underlying data for farm advisor or researcher.

Developing an advanced dairy cow decision-support tool prototype to optimize breeding decisions for dairy cows is a vital missing link in the industry today. So many PLF technologies operate in separate silos; opportunities to make evidence based breeding or culling decisions are essentially missed on many farms. GenTORE looks to integrate multiple sources of information including PLF sensor data, genomics and germplasm quality data (Spilman *et al.* 2017) to drive an opportunity to realise the wasted potential in commercial herds by targeting scarce resources and optimizing herd management decisions. The aim is to establish a practical approach to how the genetic potential offered by genomics can be optimally exploited in each individual herd environment.

Conclusions

Sustainable herd health would be supported by objective advice. Current decision tools for breeding often fail because: (1) they ignore the complex interactions between production, health, genotypic predisposition and reproduction performance at a specific herd population level; and (2) they require significant time and are not easy to use at farm level or for advisors. GenTORE aims to integrate multiple sources of information including PLF sensor data and genetics in order to deliver a herd level advisory tool that can inform a balanced herd sustainable management strategy.

References

Adriaens, I., Friggens N.C., Ouweltjes W., Scott, H., *et al.* (2020) J. Dairy Sci. 103(8):7155-7171. https://doi.org/10.3168/jds.2019-17826

Bonnet, P., Lancelot, R., Seegers, H., Martinez, D. (2011) Proc. of the 79th General Session of the World Organisation of Animal Health, Paris, France

Broom, D.M., Galindo, F.A., Murgeitio, E. (2013) Proc.of *R. Soc. B* 280(1771): 20132025. https://doi.org/10.1098/rspb.2013.2025

ADAS. Defra Report AC0120 (2013) Available at: http://sciencesearch.defra.gov.uk/Default.aspx?Menu=Menu&Module=More&Location=None&Completed=0&ProjectID=17791

Food and Agriculture Organization of the United Nations (FAO) Greenhouse gas emissions from the dairy sector: A life cycle assessment (2010). Available at: https://www.fao.org/3/k7930e/k7930e00.pdf

Friggens N.C., Adriaens I., Boré, R., Cozzi, G. *et al.* (2022) Zenodo, 5215797, ver. 5 peer-reviewed and recommended by Peer community in Animal Science. https://doi.org/10.5281/zenodo.6046479

Garnsworthy, P.C. (2004) Animal feed science and technology 112(1-4): 211-223. https://doi.org/10.1016/j.anifeedsci.2003.10.011

Hansen, J., Sato, M., Ruedy, R. (2012) PNAS 109(37): 14726-14727.

O'Neill, J. (2016) Tackling drug-resistant infections globally: final report and recommendations. Available at: https://amr-review.org/sites/default/files/160518_Final%20paper_with%20cover.pdf

Spilman, M., Burton, K., Statham, J. (2017) Reprod. Fertil. Dev. 29(1): 116. https://doi.org/10.1071/RDv29n1Ab1

Statham, J., Green, M., Husband, J., Huxley, JN., (2017) In Practice 39(1): 10-19. https://doi.org/10.1136/inp.j195

Statham, J., Scott, H., Statham, S., Williams, A., Sandars, D. (2020) Dairy Cattle Health and Greenhouse Gas Emissions Study: UK, Chile and Kenya. Available at: http://www.dairysustainabilityframework.org

Part 11

Challenges – genetic control of infectious disease transmission and impact

153. Molecular mechanisms involved with influenza A nanovaccine immunogenicity in pigs

A.M.G. Ibelli[1,2*], J.O. Peixoto[1,2], L.T. Fernandes[1], V. Haach[3], D. Gava[1], A.P. Bastos[1,2], F.N. Fonseca[4], M A.Z. Morés[1], M.E. Cantão[1], R. Schaefer[1] and M.C. Ledur[1,5]

[1]Embrapa Suínos e Aves, caixa postal 321, 89715-899, Concórdia, Santa Catarina, Brazil; [2]Programa de Pós-Graduação em Ciências Veterinárias, Universidade Estadual do Centro-Oeste, R. Salvatore Renna, 875, 85015-430, Guarapuava Paraná, Brazil; [3]Departamento de Microbiologia, Imunologia e Parasitologia, Instituto de Ciências Básicas da Saúde, Universidade Federal do Rio Grande do Sul, Av. Paulo Gama, 110, Porto Alegre, Rio Grande do Sul, Brazil,[4]Secretaria de Inovação e Negócios, Embrapa, Parque Estação Biológica – PqEB s/n, 70770-901 Brasília, Distrito Federal, Brazil; [5]Programa de Pós-Graduação em Zootecnia, UDESC-Oeste, Rua Beloni Trombeta Zanin, 680E, 89815-630 Chapecó, SC, Brazil; adriana.ibelli@embrapa.br

Abstract

Influenza A virus (IAV) is one of the most important primary agents in porcine respiratory disease complex (PRDC). The gene expression profile was evaluated by RNA-Seq analysis in mediastinal lymph nodes from non-vaccinated and vaccinated pigs with a nanovaccine for IAV to clarify molecular mechanisms and genes involved in the protective immunity conferred by influenza vaccination. A total of 14,381 genes were expressed in the analysed tissue. From those, 564 were differentially expressed (DE) between the two groups. Eighteen biological processes (BP) were enriched in immune response, cell cycle regulation and reproductive processes. The defence response was the BP with more gene counts, including chemokines (*CXCL9, CXCL10, CCL8, CXCL8*), interleukin (*IL21, IL26*), interferons (*ISG20, IFN-alpha-8*), within others. The transcriptome analyses provided a better understanding of the immunogenicity in pigs induced by nanovaccination and could support the development and evaluation of better vaccines for swine and other animal species.

Introduction

Influenza A virus (IAV) is one of the most important primary agents in porcine respiratory disease complex (PRDC), causing a major economic concern for the swine industry and a pandemic threat for humans. Although the economic impact of influenza infection in pig farms is difficult to estimate, an influenza outbreak in a non-immune pig herd causes an acute respiratory disease (up to 100% morbidity), increased number of abortions and other bacterial secondary infections. In general, influenza outbreaks observed in pig farms over time increase the mortality rates (up to 2%) and cause a 10 to 30% reduction in piglet weaning weight (Gillespie *et al.* 1999; Torremorell *et al.* 2009).

In Brazil, IAV is the main virus detected in PRDC (Silva *et al.*, 2013). The genetic diversity of IAV in swine has increased since the emergence of pandemic H1N1 (pdm) virus in 2009 (Schaefer *et al.* 2015), introducing new challenges for the diagnostic and for the development of cross-protective vaccines. H1N1, H1N2 and H3N2 virus subtypes are prevalent in pig herds in Brazil, and they have distinct genetic and antigenic profiles. Hence, vaccination of swine with IAV strains that match the strains currently circulating in swine is the main strategy to control the disease in pig herds (Salvesen;Whitelaw, 2021). Although several influenza vaccines for pigs have been tested, and their efficacy proved in field trials, very few data are available on the immunological mechanisms underlying nanovaccine-induced protection. Previous study has shown a robust humoral immune response in pigs vaccinated with an influenza nanovaccine (Haach *et al.*, 2021). Therefore, we evaluated the gene expression profile in mediastinal lymph nodes (LMD) of pigs non-vaccinated and vaccinated with a nanovaccine for IAV containing the hemagglutinins of H1N1pdm, H1N2 and H3N2 viruses, to clarify the molecular mechanisms and genes involved with the immune response for a virosome-based influenza vaccine.

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_153, © A.M.G. Ibelli *et al.* 2022

Materials & methods

Animals and sample collection. A total of 40 specific pathogen free (SPF) pigs were raised at the Embrapa Swine and Poultry National Research Centre, Concórdia, SC, Brazil, in a standard pig production system with water and feed *ad libitum*. The 28 days-old pigs were randomly distributed into two groups: G1: 10 non-vaccinated pigs, which received PBS injection and G2: 30 vaccinated pigs which received two doses (with 14 days of interval) of an IAV nanovaccine containing the hemagglutinins of H1N1pdm, H1N2 and H3N2. Animals were monitored for clinical signs, as well as for IAV or other virus and bacteria infection. Fifteen days after the second dose, all pigs were euthanized for mediastinal lymph node collection. LMD were stored in liquid nitrogen for gene expression analysis. The sample collection was performed according to the ethical guidelines of the Embrapa Swine and Poultry Ethics Committee on Animal Utilization, under the protocol number 001/2017.

RNA extraction, library preparation and sequencing. For the gene expression analysis, a total of 24 pigs (8 non-vaccinated, 5 females and 3 males and 16 vaccinated, 8 females and 8 males) were selected. The total RNA was extracted from approximately 100 mg of frozen mediastinal lymph node tissue using Trizol (Life Technologies), according to the manufacturer's protocol, followed by an RNA cleanup using the RNeasy mini kit (Qiagen). The RNA concentration was measured with Biodrop spectrophotometer (Biodrop) and the RNA integrity was verified using Bioanalyzer 2100 equipment (Agilent). The libraries were prepared using the Illumina Truseq Stranded mRNA (Illumina), with 1µg of total RNA, following the manufacturer's protocol. The sequencing was performed in Illumina NextSeq 2000 sequencer (lllumina, USA), using 2×100 bp paired-end reads, at the NGS Soluções Genômicas facility, in Piracicaba, São Paulo, Brazil.

Data analysis and functional annotation. The data were analysed using the Bioinformatics tools for quality control (QC) and mapping reads using BAQCOM pipeline available in the Github repository (Oliveira, 2022), which uses Trimmomatic (Bolger; Lohse; Usadel, 2014) to remove short reads (<70 bp), reads with low quality (QPhred <20) and adapter sequences, STAR (Dobin *et al.*, 2013) for mapping against the swine reference genome (Sscrofa 11.1, Ensembl release 104) and the HTseq-count (Anders *et al.*, 2015) for reads counting. The differentially expressed (DE) genes were obtained using the limma package (Ritchie *et al.*, 2015) and considered DE when a false discovery rate (FDR) was <0.05 and a logFC >|1|. Negative and positive fold-changes indicate down and upregulation of the genes in the vaccinated compared to non-vaccinated groups. The functional annotation of the DE genes was performed with clusterProfiler (Wu *et al.*, 2021) package from R language using gene ontology (GO), MSigdb and Reactome databases. Biological processes (BP) with FDR<0.05 were considered enriched. A Gene-Concept Network, which considers potentially biological complexities within genes and BP, was also constructed in the clusterProfiler.

Results

A total of 397,856,687 paired end reads were obtained for the 24 LMD samples. After QC, an average of 31.1±6.5 million paired-end reads per sample were kept, and 94.42% were uniquely mapped across the swine reference genome. Out of those, 80.63% were mapped in genes. A total of 14,381 genes were expressed in the swine mediastinal lymph node. From those, 564 were DE between groups: 316 were down and 248 were upregulated in the vaccinated group. A total of 481 genes had gene names assigned and were considered annotated, and therefore used for the enrichment analysis. After removing redundant GO terms, 18 BP were enriched in this dataset, where most of them were related to immune response, cell cycle regulation and reproductive processes. The defence response was the BP with more gene counts (25 genes), including chemokines (*CXCL9, CXCL10, CCL8, CXCL8*), interleukin (*IL21*), interferons (*ISG20, IFN-alpha-8*). In this BP, the activation induced cytidine deaminase gene (*AICDA*) gene was included,

which was the most upregulated gene in response to vaccination. Furthermore, it was possible to observe different BP that have genes with multiple functions, such as *CDC6, CDC20, CCNB1, RMI2, SGO1, TOP2A, BIRC5, ZWINT, AURKB, KIF18A, NUF2* and *HJURP*.

Using the immunologic signature of gene sets (C7) available in MSigdb, that are based on human and mouse datasets, 32 genes DE in our study have already been associated to the immune response to IAV: *MYBL2, RRM2, UBE2T, ASF1B, E2F2, CDC20, KIF2C, CDCA8, CDCA5, TOP2A, POLQ, HJURP, CENPE, BIRC5, NCAPG, KNL1, MKI67, ZWINT, GTSE1, AURKB, DLGAP5, KIF18A, E2F8, ESPL1, NCAPH, CDCA2, BUB1, CDCA7, LGI4, CKB, LPL* and *AZU1*.

Discussion

Nanovaccines can potentialize the immune response for a given antigen and could be one of the strategies to improve the performance of swine IAV vaccines (Salvesen; Whitelaw, 2021). In this study, the LMD gene expression profile in response to an IAV polyvalent nanovaccine developed by Embrapa Swine and Poultry was evaluated, evincing genes in several BPs related to the regulation of humoral immune response, immune cell migration and cell differentiation. These results corroborate with Haach *et al.* (2021), since a robust humoral immune response has already been described after vaccination of pigs with a polyvalent nanovaccine against IAV. Genes such as *AMCF-II, CXCL9, CXCL10* and *CXCL8*, known by their function in humoral response, were within the 40 genes most upregulated in the vaccinated group. The *AMCF-II*, an alveolar macrophage-derived chemotactic factor has antiviral activity (Borca *et al.*, 2021), and the *CXCL9* and *CXCL10* are activated in response to IAV in humans (Wang *et al.*, 2011).

The *AICDA* was the most upregulated (logFC=4.49) in response to the IAV nanovaccination. This gene is essential to generate high-affinity antibodies, robust humoral immunity and immune memory, and its activation in response to influenza vaccination has been observed in humans (Frasca *et al.*, 2015). It has been suggested that this gene could be a predictive marker of individual response to IAV vaccination (Frasca *et al.*, 2015). Other highly expressed genes, such as *IL26* and *IL21* are also involved in host defence and inflammatory response (Che *et al.*, 2020). Besides the functional candidate genes, which are potential markers for immune response, several genes that are important for the activation/proliferation of B and T cells were also highly expressed. Therefore, the transcriptome analyses provide a better understanding of the immunogenicity in pigs induced by nanovaccination and could support the development and evaluation of better vaccines for swine and other animal species.

Acknowledgements

This study was supported by the Brazilian Agricultural Research Corporation (EMBRAPA project # 22.16.05.004.00.06. LTF is supported by a visiting specialist fellowship from the National Council of Scientific and Technological Development (CNPq) grant # 380282/2021-6 and VH by a Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) scholarship. MCL is a CNPq productivity fellow. We thank the CAPES, Brazil – Finance Code 001 for the free access to the journals used in the literature review.

References

Borca. M.V. *et al.*, 2008. Virus Res. 2008. 38(1-2):89-96. http://doi.org/10.1016/j.virusres.2008.08.009.

Che KF, Tengvall S, Lindén A. I.2021. Cytokine Growth Factor Rev.57:1-10. http://doi.org/10.1016/j.cytogfr.2020.10.003.

Dobin, A. *et al.* 2013.Bioinformatics 2013, 29, 15–21. https://doi.org/10.1093/bioinformatics/btu170

Frasca, D., Diaz, A., Blomberg, B.B. 2015. Methods Mol Biol. 1343:107-14. http://doi.org/10.1007/978-1-4939-2963-4_9

Gillespie, T.G. 1999. The Journal of Swine Health and Production.7(2):81–8.

Haach, V. 2021. Annals of the XXXII Brazilian Congress of Virology. Online.

Kosmider B, Hartshorn K, Mason RJ.2011. Am J Respir Cell Mol Biol. 5(3):582-91. http://doi.org/10.1165/rcmb.2010-0108OC.

Oliveira, H.C. 2022. Github Repository. Available online: https://github.com/hanielcedraz/BAQCOM (accessed on 12 January 2022).

Ritchie, M.E. *et al.* 2015. Nucleic Acids Research, 43(7), e47. http://doi.org/10.1093/nar/gkv007.

Salvesen, H.A., Whitelaw, B.A. 2021. Porcine Health Manag. 7(1):23. http://doi.org/10.1186/s40813-021-00196-0.

Schaefer, R. *et al.*2015. Archives of Virology. 160(1): 29-38. http://doi.org/10.1007/s00705-014-2223-z

Silva, M.C. *et al.*. 2013. Proceedings of the American College of Veterinary Pathologists (ACVP and ASVCP Annual Meeting), Montreal, Canada. 2013.

Torremorell, M. *et al.* 2009. Veterinary Record. 165: 74-77. http://doi.org/10.1136/vetrec.165.3.74

Wu, T. *et al.* 2021. The Innovation, 2(3), 100141. http://doi.org/10.1016/j.xinn.2021.100141.

154. Host variation in susceptibility to lumpy skin disease: gene expression analyses of experimentally infected cows

S. Van Borm[1], T. Klingström[2], A. Niazi[2], K. De Clercq[1], L. Mostin[1], A. Haegeman[1] and D.J. de Koning[2*]

[1]Sciensano, Rue Juliette Wytsmanstraat 14, 1050 Brussels, Belgium; [2]Department of Animal Breeding and Genetics, Swedish University of Agricultural Sciences, Box 7023, 750 07 Uppsala, Sweden; dj.de-koning@slu.se

Abstract

Lumpy Skin Disease (LSD) is becoming a global threat to cattle and related species. It requires annual vaccination to prevent further outbreaks. There is host variation in the response to infection both in field studies and under carefully controlled artificial infections. Here, we study the gene expression response in whole blood at four time points before and after experimental infection with LSD virus in five cows with divergent disease outcomes. Differentially expressed genes (FDR<0.05) were identified for comparisons between LSD symptomatic and asymptomatic animals at all time points as well as between time points for symptomatic animals. Gene set enrichment showed a role of oxidative phosphorylation as well as a range of disease pathways. While preliminary, these results are the first ever study of host transcriptomes related to LSD.

Introduction

Lumpy skin disease (LSD) is endemic in Africa and an immediate and serious threat to cattle in Europe and eastern and central Asia. The disease is caused by the capripoxvirus Lumpy Skin Disease Virus (LSDV). The disease has spread steadily north over the past 10 years from Africa through the Middle East and the eastern Mediterranean region before entering Europe, the Balkans and Caucasus in 2014-16 (EFSA *et al.*, 2020). Several Balkan countries currently vaccinate against LSD but this is costly (reviewed by Tuppurainen *et al.*, 2021)). There is clear host variation in response to LSD infection: in field studies about 50% of the animals develop no clinical signs when a herd is infected with LSDV (EFSA *et al.*, 2020). This is confirmed by experimental challenges where animals are infected by direct injection of LSDV into the vena jugularis and the neck (Haegeman *et al.*, 2021). Here we show a first attempt to explore mechanisms behind divergent disease outcomes by studying gene expression after experimental LSDV infection in whole blood.

Materials & methods

Five approximately 6-month-old male Holstein bulls which were tested free of BVD and IBR were experimentally infected at Sciensano with LSDV via injection in the vena jugularis and the neck with a LSDV strain derived from Israel (Haegeman *et al.*, 2021). Blood samples were taken five days prior to infection, at day 1 post infection and subsequently at least every other day until day 15. Three animals were symptomatic for LSD while the other two were asymptomatic. Here we used whole blood stored in Tempus™ Blood RNA Tubes (Applied Biosystems™) from days -5, 3, 7, and 15. RNA was isolated using the Tempus™ Spin RNA Isolation Reagent Kit according to the manufacturer's instructions. RNA integrity was validated using high sensitivity RNA Screentape reagents (Agilent) yielding integrity scores (RQN)≥9.6 for all samples.

RNA Sequencing. RNA sequencing was performed at the Neuromics Support Facility of VIB University of Antwerp, Center for Molecular Neurology. RNA sequencing followed a QuantSeq approach where sequencing libraries were constructed near the 3′ end of polyadenylated RNAs (Moll *et al.*, 2014). This results in only a single fragment per transcript thus simplifying the quantification of gene expression. Libraries were sequenced with Illumina NextSeq500 using 150 HO sequencing kit.

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_154, © S. Van Borm *et al.* 2022

RNA analysis. Bioinformatics analyses were performed by the SLU Bioinformatics Infrastructure (SLUBI). Quality control of sequencing data was performed using FastQC v0.11.8. Quality trimming and adapter removal was performed using bbduk.sh script available in BBMAP suite v38.94. After data pre-processing, the filtered data was input to STAR aligner v2.7.9a (Dobin *et al.,* 2013) for read mapping to *Bos taurus* reference genome. Top level assembly of the genome Bos_taurus.ARS-UCD1.2.105 and annotation in annotation v105 in GTF format was downloaded from ENSEMBL database. Differential Expression analysis was performed in R version 4.1.2 using DESeq2 v1.32.0 package (Love *et al.*, 2014). Comparisons were made between symptomatic and asymptomatic animals at each time points as well as within symptomatic animals between time points. Gene Set Enrichment Analyses were done for both gene ontology terms (GO) and KEGG pathways.

Results

All 20 libraries were sequenced successfully with an alignment rate of 90-93% of reads uniquely mapped to the reference genome. One sample of a symptomatic animal at day 3 was shown to be an outlier when doing a PCA plot of all the results. This sample also had the lowest RNA concentration and was hence removed from further analysis. When comparing the symptomatic animals to the asymptomatic animals at different time points the smallest number (17) of differentially expressed (false discovery rate (FDR)<0.05) genes was at 5 days prior to infection with the highest number (376) at 7 days post infection. At 3 days and 15 days post infection, we detected 36 and 40 differentially expressed genes, respectively. When comparing gene expression patterns over time within the symptomatic animals, the largest number of differentially expressed genes was between day 3 and day 7 post infection (81). The Venn diagrams in Figure 1 show limited overlap between different analyses, in particular when comparing symptomatic and asymptomatic animals. Table 1 shows the results for the functional annotation of the differentially expressed genes. While day 7 after infection showed the largest number of differentially expressed genes, day 3 after infection gives the clearest gene set enrichment related to disease response, both for GO terms and KEGG pathways. The GO terms point towards several metabolic processes while the KEGG pathways highlight a broad spectrum of disease-related pathways. While only nominally significant, the GO terms for day 7 post infection included terms like coagulation, platelet activation, wound healing and hemostasis. Both GO and KEGG point towards a role of oxidative (de-)phosphorylation.

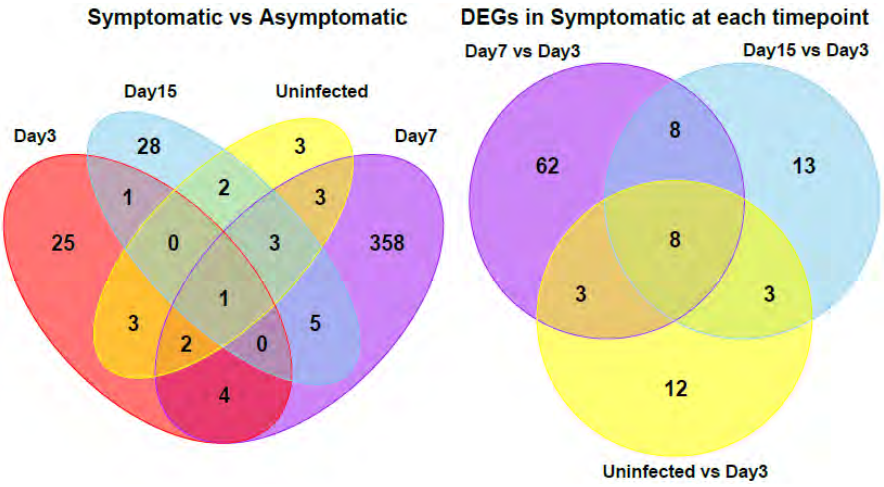


Figure 1. Differentially expressed genes between experimentally infected cows with different clinical outcomes (left), and within symptomatic animals over time (right).

Table 1. Functional annotation of differentially expressed genes.

Comparison	GO_Terms	FDR	KEGG Pathways	FDR
Symptomatic versus asymptomatic, day 3	ATP metabolic process	0.021	Oxidative phosphorylation	<0.001
	Oxidative phosphorylation	0.038	Chemical carcinogenesis – reactive oxygen species	<0.001
	Generation of precursor metabolites and energy	0.054	Prion disease	0.001
			Parkinson disease	0.001
			Diabetic cardiomyopathy	0.002
			Thermogenesis	0.004
			Amyotrophic lateral sclerosis	0.004
			Huntington disease	0.01
			Pathways of neurodegeneration – multiple diseases	0.011
			Alzheimer disease	0.021
Symptomatic versus asymptomatic, day 7	-	-	Vitamin digestion and absorption	0.08
Symptomatic versus asymptomatic, day 15	-	-	Complement and coagulation cascades	0.03
			Hematopoietic cell lineage	0.087
Symptomatic, before infection versus day 3	-	-	Systemic lupus erythematosus	0.002
			Transcriptional misregulation in cancer	0.007
			Alcoholism	0.033
			Neutrophil extracellular trap formation	0.038
Symptomatic, day 7 versus day 3	Nervous system process	0.042	-	-
	Sensory perception	0.08		
	Positive regulation of dephosphorylation	0.08		
	Regulation of organ growth	0.08		
	Behavior	0.08		
	Bone morphogenesis	0.08		

Discussion

To the best of our knowledge this is the first study of host transcriptome response to infection with LSDV. While modest in terms of sample size, we could show clear differences between animals with different disease outcomes for all time points. The finding of differentially expressed genes, while only a handful, between symptomatic and asymptomatic animals at five days before infection is intriguing. While it may simply be an artifact, it could also reflect the prior immune status of animals that has a subsequent effect on the outcome of the infection. The most significant gene prior to infection, Interleukin 1 Receptor Accessory Protein (IL1RAP) is also significant at other time points, together with other genes related to interleukin. In further analyses we also look at genet works and pathways over time using the whole data. In order to make efficient use of limited blood volumes and other resources, we opted for a QuantSeq approach in this study. While this is an efficient way for gene counting, we did not capture any splice variants or get a complete whole blood transcriptome of LSDV response. In ongoing studies, we will perform whole transcriptome sequencing in a cohort of 15 animals which will facilitate a comparison with more statistical power and a more detailed analysis of the transcriptome itself.

Acknowledgements

This research was funded by the European Union's Horizon 2020 Research and Innovation Program, under grant agreement No. 773701, DEFEND project (www.defend2020.eu). The authors would like to thank the animal caretakers of the Experimental Centre of Sciensano. We are very grateful to the Kimron Veterinary Institute (Israel) and the Field Israeli Veterinary Services for providing us with the LSDV strain LSD/OA3-Ts. MORAN. M. seed pass.4. 155920/2012.20.1.13.

References

European Food Safety Authority (EFSA), Calistri, P., De Clercq, K., Gubbins, S., Klement, E., *et al.* (2020) EFSA Journal: 18(2), p.e06010. https://doi.org/10.2903/j.efsa.2020.6010

Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., *et al.* (2013) Bioinformatics 29(1):15–21. https://doi.org/10.1093/bioinformatics/bts635

Haegeman A, De Leeuw I, Mostin L, Van Campe W, Aerts L, *et al* (2021). Vaccines 2021, 9(5), 473. https://doi.org/10.3390/vaccines9050473

Love, M.I., Huber, W. and Anders, S. (2014) Genome Biol 15, 550. https://doi.org/10.1186/s13059-014-0550-8

Moll, P., Ante, M., Seitz, A., and Reda, T. (2014) Nat. Methods 11: i–iii. https://doi.org/10.1038/nmeth.f.376

Tuppurainen, E., Dietze, K., Wolff, J., Bergmann, H., Beltran-Alcrudo, D. *et al.* (2021) Vaccines 9:1136. https://doi.org/10.3390/vaccines9101136

155. Relationship between resistance to a *Pasteurella multocida* experimental infection and production traits in rabbits

H. Garreau[1*], F. Lantier[2], B. Bed'hom[3], E. Guitton[4], V. Helies[5], E. Helloin[2], C. Herbert[6], M. Maupin[7], R. Robert[7] and M. Gunia[1]

[1]GenPhySE, Université de Toulouse, INRAE, F-31326 Castanet-Tolosan, France; [2]ISP, INRAE Centre Val de Loire, Université François Rabelais de Tours, UMR 1282, 37380 Nouzilly, France; [3]GABI, INRAE, AgroParisTech, Université Paris-Saclay, 78352 Jouy-en-Josas, France; [4]PFIE, INRAE Centre Val de Loire, 37380 Nouzilly, France; [6]HYCOLE, Route de Villers-Plouich, 59159 Marcoing, France; [7]HYPHARM SAS, La Corbière, Roussay, 49450 Sèvremoine, France; herve.garreau@inrae.fr

Abstract

Nine hundred and fifty five rabbits were inoculated at 6 weeks of age with a pyogenic strain of *Pasteurella multocida* and were scored for resistance from 0 to 4 by taking into account post-mortem examinations. Fifty five sires having at least 10 inoculated offspring produced also 9,943 rabbits tested for health and growth and 1,468 females tested for reproduction in commercial farm conditions. Total number born and number born alive per litter were significantly lower in daughters of resistant sires (-0.27 and -0.41). Prevalence of digestive diseases and infectious diseases were significantly lower in the resistant growing rabbits than in the susceptible ones (-1.7 point and -3.0 points, respectively). This result suggests that resistance to pasteurellosis assessed by an experimental infection is favorably correlated to resistance to other infectious diseases.

Introduction

Pasteurellosis is one of the most common bacterial infection in commercial rabbit farms. The RELAPA project (Genomics for the Genetic Resistance of Rabbits to Pasteurellosis) aims at identifying genomic regions associated with the resistance to pasteurellosis. To establish a successful commercial breeding program to improve resistance to pasteurellosis, we need to evaluate the relationship between this trait and resistance to other diseases and production traits. This study aims at comparing production and health traits of growing rabbits and breeding does raised in commercial farms and produced from the resistant and susceptible sires having offspring tested through experimental inoculation with *Pasteurella multocida* in the RELAPA project.

Materials & methods

Sires evaluation. In the RELAPA program, 955 crossbred rabbits were inoculated at 6 weeks of age with a pyogenic strain *of Pasteurella multocida* and the response to inoculation was assessed during the following 14 days, the animals being sacrificed after 14 days (Shrestha *et al.*, 2018). These crossbreds are progenies from six sire lines (two lines each from breeding companies Eurolap, Hycole, and Hypharm) and one dam line (INRA 1777). The rabbits were scored for resistance to pasteurellosis from 0 to 4, by taking into account scores for extent of abscesses (0 to 4), scores of *P. multocida* presence (0 to 4), and status of the rabbits (dead/alive) at the end of the experiment. (Shrestha *et al.*, 2018). Among the sires used for RELAPA, 55 having produced at least 10 inoculated animals were classified as resistant (28 sires) or susceptible (27 sires) according to the mean resistance value of their inoculated offspring (Table 1).

Table 1. Means and standard deviation of the resistance score of resistant and susceptible sires calculated from their inoculated offspring score.

Group of sire	N	Mean resistance score	Standard deviation of resistance score
Resistant	28	1.57	0.16
Susceptible	27	2.03	0.26

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_155, © H. Garreau *et al.* 2022

Animals and measured traits. The 28 resistant sires produced 4,991 growing rabbits (RG) and 884 breeding does (RD) while the 27 susceptible sires produced 4,952 growing rabbits (SG) and 712 breeding does (SD). Growing rabbits and does were raised in six commercials farms belonging to the three breeding companies partners of the RELAPA project. The does were raised in conventional housing, inseminated for the first time at 19 weeks of age and then every 42 days. Females produced 6,676 litters, or 4.2 litters per female on average. For each insemination, fertility was scored one (AI followed by a kindling) or zero (AI not followed by a kindling). For each litter, the number born alive and dead has been recorded. Stillbirth is calculated as the difference between the total number born and the number born alive divided by total number born. Nest mortality was calculated as the difference between the litter balance after adoption and the number of weaned rabbits divided by the litter balance.

The rabbits were weighed at weaning and at the end of the test. The weaning age ranged from 28 to 35 days, while the age of the end of fattening ranged from 66 to 70 days depending on the rearing farm. Mortality and clinical signs of diseases occurring in growing rabbits were recorded at the end of the test. The most likely cause of death of rabbits that died before the end of the test was also recorded. Disease traits were coded as 0 (absence) or 1 (disorder= morbidity or mortality). They were the following: (1) digestive disease, which includes diarrhea, bloated abdomen, and any form of digestive symptoms; (2) respiratory diseases which includes nasal discharge, pneumonia; (3) all infectious diseases, which combines digestive and respiratory diseases, abnormal low weights and other clinical signs of infectious origin. The basic statistics of growth and disease traits are given in Table 2.

The total prevalence of digestive, respiratory, infectious diseases and mortality recorded during the growing period are given in Table 3.

Table 2. Number of animals, mean and standard deviation (Std) for growth and reproduction traits.

Trait	Number of animals	Mean	Standard deviation
Weaning weight (g)	8,025	866	232
Final weight (g)	7,477	2,292	258
Fertility (0/1)	6,676	0.87	-
Total number born / litter	5,798	10.80	3.10
Number born alive / litter	5,798	10.09	3.57
Stillbirth (%)	5,137	7.35	-
Nest mortality (%)	5,069	9.81	-

Table 3. Mean prevalence (in %) of digestive, respiratory, all infectious diseases and mortality (%) recorded during the growing period.

Trait	Mean
Digestive diseases	8.7
Respiratory diseases	7.1
All infectious diseases	21.3
Mortality	5.3

Statistical analysis. The growth performances were analyzed with the PROC MIXED procedure of the SAS software (2008). The significant fixed effects ($P<0.05$) retained in the model were: the sire's resistance group (n=2), the combined effect batch × genotype (n=30), the litter size at birth (n=12), the litter size at weaning (n=10), the parity of the birth litter (n=4). The effect of weaning age was added to the weaning weight model while the end of fattening age effect was added to the final weight model. Health traits were analysed with the same procedure and the same fixed effects but without the litter size and the parity of the birth litter, which were not significant for this model. The random common environment of the litter effect was added for growth and health traits to the model.

The reproduction performances were analyzed with the PROC MIXED procedure of the SAS software (2008). The fixed effects retained in the model were: the sire's resistance group (n=2), the combined effect year × season × farm (n=26), the combined effect parity × physiological stage (lactating or not) (n=9). The random effect of the female was added to the model to take into account the repetition of female performances.

Ethics. All experiments were conducted in accordance with the guidelines of the Directive 2010/63/EU of the European Parliament and of the Council and were approved by the ethical review board (APAFIS#3866-2016020113447262).

Results & discussion

Differences in least squares means of traits are presented in Table 4. Weaning and final weight were not significantly higher in RG than in SG (+7 g, P=0.26 and +14 g, P=0.10 respectively). However Shrestha *et al.* (2018) estimated a favorable genetic correlation between resistance to pasteurellosis and average daily gain from birth to weaning recorded in inoculated animals of the same experiment. A moderate and favourable correlation between natural resistance to infectious diseases and weaning weight was also observed in another study carried in a commercial population (-0.34±0.12) (Gunia *et al.*, 2018). All together these findings suggest that breeding for pasteurellosis resistance is not detrimental to breeding for growth traits. On the contrary, total number born and born alive were significantly lower in litters of RD than in litters of SD (-0.27, P=0.02 and -0.41 P=0.003). These results suggest that breeding for resistance to pasteurellosis could decrease the prolificacy of does. However birth and nest mortality were not significantly different between litters of RD and SD.

Table 4. Differences in least squares means between offspring of resistant and susceptible sires for growth, diseases and reproduction traits, *P*-value of the sire group effect and standard error of least square means (LSM) differences.

Trait	Difference resistant-susceptible	*P*-value	Standard error of LSM difference
Weaning weight (g)	7	0.264	6
Final weight (g)	15	0.106	9
Digestive diseases (%)	-1.673	0.031	0.776
Respiratory diseases (%)	-0.698	0.284	0.652
Infectious diseases (%)	-2.991	0.008	1.121
Mortality (%)	-1.084	0.066	0.588
Fertility (0/1)	0.003	0.708	0.008
Total number born / litter	-0.270	0.023	0.119
Number born alive / litter	-0.410	0.003	0.137
Stillbirth (%)	1.547	0.056	0.809
Nest mortality (%)	-0.901	0.706	0.202

Prevalence of digestive diseases and infectious diseases were significantly lower in RG than in SG (-1.7 point, P=0.03 and – 3.0 points, P=0.008 respectively). This interesting result suggests that resistance to pasteurellosis assessed by an experimental infection using a single strain of *Pasteurella multocida* is favorably correlated to resistance to other infectious diseases. Gunia *et al.* (2015) demonstrated that a composite trait grouping all disease syndromes was heritable in rabbit. However, there was almost no genetic correlation between digestive and respiratory disease traits. This result can be partly explained by the fact only one syndrome was recorded per animal. The difference in prevalence of respiratory disease between RG and SG is not significant (-0.70 point, P=0.28). This could be explained by a lack of homogeneity in the recording of the trait between the different farms: some breeders considered only true respiratory symptoms (nasal discharges, pneumonia) while some others included some other symptoms related to pasteurellosis (abscesses, wryneck, eye infections).

Funding

This work was funded by the Institut Carnot Santé Animale 2015 and by the CLIPP, HYCOLE and HYPHARM.

References

Gunia M., David I., Hurtaud J., Maupin M., Gilbert H., Garreau H. (2015) J. Anim. Sci. 93:5631-5638. https://doi.org/10.2527/jas.2015-9377

Gunia M., David I., Hurtaud J., Maupin M., Gilbert H., Garreau H. (2018) Front. Genet. [Internet] 9:467. https://doi.org/10.3389/fgene.2018.00467

SAS. 2008. SAS Inst.Inc.,Cary,NC,v9.

Shrestha M., Garreau H., Balmisse E., Bed'hom B., David I. *et al.* (2018) Proc. of the 11th WCGALP, Auckland, New-Zealand.

156. Quantification of environmental management systems for U.S. Katahdin sheep producers

B.C. Arisman[1*], J.M. Burke[2], J.L.M. Morgan[3] and R.M. Lewis[1]

[1]Department of Animal Science, University of Nebraska-Lincoln, P.O. Box 830908, Lincoln, NE 68588, USA; [2]ARS, USDA, Dale Bumpers Small Farm Research Center, 6883 South State Highway 23, Booneville, AR 72927, USA; [3]Round Mountain Consulting, 18235 Wildlife Road, Fayetteville, AR 72701, USA; barisman2@huskers.unl.edu

Abstract

Gastrointestinal nematodes negatively affect the performance and wellbeing of sheep, encouraging producers to choose breeds that are naturally resistant. An example is Katahdin sheep. Katahdin are raised in various climates and management systems in the U.S., which can be combined into eco-management clusters. We hypothesized that gastrointestinal nematode challenge varies predictably based on the characteristics of the eco-management cluster. Katahdin producers from across the country were surveyed for management information. Body weights and faecal egg counts (FEC) were collected. Management and climate data were combined using a Factorial Analysis of Mixed Data approach to visualize ecomanagement clusters. The milder, drier, low input clusters had higher FEC while the colder, wetter, low input, clusters had a lower FEC ($P<0.05$). Through formation of ecomanagement clusters, an environmental variable can be specified for study of interactions or sensitivities of genotypes to their environment, potentially useful for identifying parasite-resistant sheep.

Introduction

The sheep industry contributed $5.8 billion to the United States (U.S.) economy in 2017 (Shiflett, 2017). Given the monetary impact, and for animal wellbeing, ensuring the good health and sound management of sheep is of utmost importance. A main factor negatively affecting the growth and health of sheep managed in a forage-based system is gastrointestinal nematode (GIN) parasites. GIN infection can cause weight loss, decreased appetite, and even death (Lôbo *et al.*, 2009). The inability to control GIN infection due to the prevalence of anthelmintic resistance has increased the popularity of breeds recognized as naturally resistant to parasites. Katahdin are a relatively prolific maternal composite hair breed that are comparable to other medium-sized maternal breeds in adult bodyweight (BW) and lamb growth (Ngere *et al.*, 2018). This breed is also known to potentially exhibit resilience and tolerance to GIN infection.

One approach to defining robustness is to characterize the sensitivity of individuals or families to environments varying in GIN infection challenge. The environment can be defined by differences among flocks in geographic location, with distinctive climates, and in management. Management differences among flocks, however, are often not considered when defining an environment due to the difficulty in quantifying management practices. The combination of the climate and management data can be referred to as ecomanagement groups or clusters.

Katahdin producers from across the U.S. were surveyed for management information, which was combined with U.S. Climate Normals data that captures precipitation, snowfall, and temperatures for the different flock locations. Performance data, including BW and faecal egg counts (FEC), were also available on the flocks. The aim of this study was to quantify ecomanagement clusters in Katahdin sheep with respect to their association with GIN challenge.

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_156, © B.C. Arisman *et al.* 2022

Materials & methods

Over three years (2017-2019), data were collected on 4,645 lambs from 17 flocks from across the U.S. as part of the National Sheep Improvement Program (Notter, 1998). The data considered included pedigree, BW (kg) and FEC (eggs/g) collected around 90-d of age [91.4 (SD 8.3) d], and drenching protocols. Only FEC from animals for which no anthelmintic was given 30 d prior to measurement were used, yielding 4,232 lambs with a valid FEC record. The FEC were transformed as log(FEC+25). Birth and rearing type measures were concatenated into a single variable (6 levels), and dam ages classified (5 levels). Contemporary groups (CG) were assigned based on birth year, management group, record date, sex, and age slice. Age slice was defined in 35-d intervals from the birth of the first lamb. CG with fewer than five individuals were discarded.

Environmental data. Climate data were obtained from the weather station closest to each flock's geographical location using the U.S. Climate Normals (1991 to 2020) published by the National Weather Service (2021). The weather data captured were monthly and yearly averages for precipitation, snowfall, and minimum and maximum temperature. Site elevation was also included.

Management data. A Qualtrics survey was distributed to 21 producers; 17 responded (81%), defining the flocks and data used in the study. The intent of the survey was to quantify differences in management practices. Practices of interest were grazing system, GIN impact, selection strategies to mitigate parasitism, feeding regime, and ewe flock and lamb management strategies. Survey responses were combined with climate data per a given flock's location.

Data analysis. The data were analysed using the Factoextra and FactoMineR packages in R (Lê *et al.*, 2008; Mundt and Kassambara, 2020). Using a Factorial Analysis of Mixed Data (FAMD) approach, a type of principal component analysis, groups of flocks with similar management practices and climate variables were grouped into ecomanagement clusters that define an environment. Using ANOVA (Fox and Weisberg, 2019), the systematic influences of ecomanagement cluster, CG, birth/rearing type, and dam ages in defining variation in FEC and BW were tested (α=0.05). Least square means were obtained for ecomanagement clusters and compared with the Fisher's least significant difference test (pairwiset.t.test; R Core team, 2020).

Results

The 17 producers answered all questions on the questionnaire. The greatest dissimilarity in responses were in feeding regime (from 0 to 100% pasture based) and sheep density per acre (from 2 to 27 ewes per acre). The greatest similarity in responses were regarding lambing season and selection for GIN strategy, where 94% of flocks lambed in the spring and used a combination of estimated breeding values and phenotypic performance records to select for GIN resistance.

Based on the FAMD analysis, ecomanagement clusters of flocks were formed (delineated by colour ellipses in Figure 1), where principal components 1 and 2 captured 13.5 and 11.2% of the variation, respectively. Any flocks not in an ellipse formed a standalone ecomanagement cluster. The three main factors discriminated the clusters were temperature, precipitation, and grain supplementation.

All systematic effects fitted in the statistical model defined variation in FEC and BW (P<0.05). In Table 1, mean FEC and BW are summarized relative to ecomanagement cluster. The FEC means ranged from 2.34 to 2.96 eggs/g while BW ranged from 17.0 to 27.3 kg.

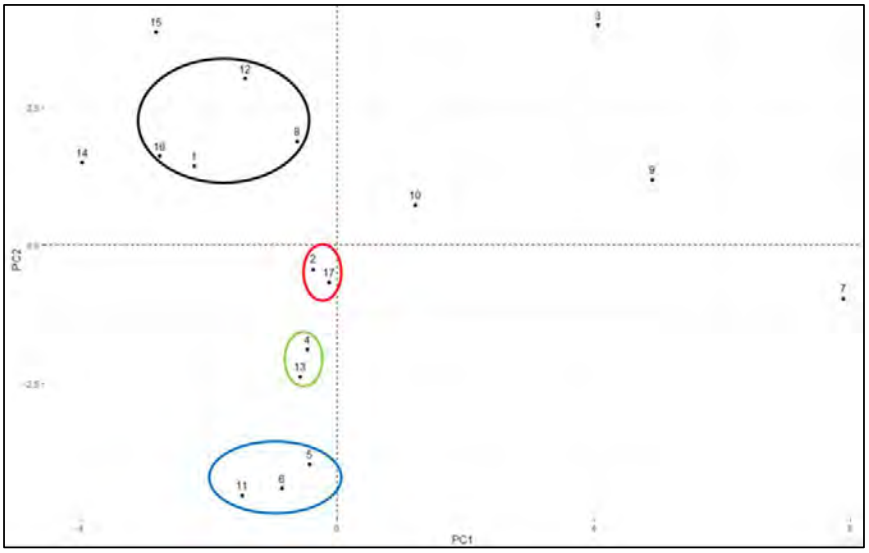


Figure 1. Principal component analysis plot derived from the Factorial Analysis of Mixed Data approach (PC1 – first principal component; PC2 – second principal component).

Table 1. Descriptions and least square means (± s.e.) for log transformed faecal egg count [(Log(FEC+25)] and bodyweight for ecomanagement clusters.

Ecomanag. cluster	Climatic and management variables defining cluster	Log(FEC+25)		Bodyweight	
		No.	Mean, eggs/g[1]	No.	Mean, kg[1]
15	Colder, wetter, no grain suppl.	260	2.88±0.010^{abc}	316	17.0±0.05^{g}
10	Colder, wetter, low grain suppl.	315	2.34±0.008^{g}	333	25.9±0.10^{bcd}
14	Colder, drier, no grain suppl.	461	2.71±0.008^{bcd}	461	23.7±0.08^{e}
Blue	Colder, drier, low grain suppl.	518	2.63±0.012^{de}	741	20.4±0.11^{f}
Green	Hotter, drier, high grain suppl.	339	2.69±0.012^{cd}	390	27.3±0.11^{a}
9	Hotter, wetter, high grain suppl.	405	2.58±0.011^{e}	425	24.5±0.09^{cde}
7	Hotter, wetter, low grain suppl.	187	2.88±0.011^{ab}	89	26.3±0.09^{abc}
Black	Milder, wetter, no grain suppl.	235	2.48±0.013^{f}	245	24.0±0.08^{de}
3	Milder, drier, low grain suppl.	123	2.72±0.009^{bcd}	184	25.0±0.13^{bcde}
Red	Milder, drier, low grain suppl.	137	2.96±0.006^{a}	88	17.6±0.01^{g}

[1] Means within a column with different superscripts differ ($P<0.05$).

Discussion

In previous research, climate and management practices were considered separately to quantify the environment. U.S. Climate Normals were used to define ecoregions in beef cattle, while surveys were used to quantify management practices in sheep (Rowan *et al.*, 2021; Stewart *et al.*, 2020). However, our combining climate with management data allowed identification of ecomanagement clusters improving the description of the overall environment. Of the factors considered in the FAMD analysis, temperature appeared to affect the pattern of the clustering more substantially. This is logical since GIN development relies on temperature and humidity (Lôbo *et al.*, 2009). The clusters with milder and drier climates, and

low grain supplementation, suffered from a greater FEC. Conversely, the colder and wetter climates, with low grain supplementation, benefitted from a lower FEC. There was no clear association of parasite load with BW.

Although the climate and management practices of only 17 flocks were considered in the study, they represented the full geographic distribution of Katahdin flocks in the U.S. The ecomanagement clusters were defined using management practices present during the time span of the study. Re-analysis of the clusters would need to be completed on a periodic basis, accounting for changes in management practices that may occur over time. The next stage of this research is to determine if sufficiently strong familial relationship exist across ecomanagement clusters to connect them genetically. If so, the clusters could be used as an environmental variable to investigate genotype by environment interactions, or in a reaction norm analysis to identify superior individuals or families able to tolerate variable GIN challenges.

Acknowledgements

This project is supported by the Agriculture and Food Research Initiative Grant number 2016-51300-25723 from the USDA National Institute of Food and Agriculture, Organic Agriculture Research and Extension Initiative program.

References

Fox J., and Weisberg S. (2019). An {R} Companion to Applied Regression. Available at https://socialsciences.mcmaster.ca/jfox/Books/Companion/

Lê S., Josse J., and Husson F. (2008) J. Stat. Softw. 25(1):1-18. https://doi.org/10.18637/jss.v025.i01

Lôbo R.N., Vieira L.S., de Oliveira A.A., Muniz E.N., and da Silva J.M. (2009) Genet. Mol. Biol. 32(2):288-94. https://doi.org/10.1590/S1415-47572009005000032

Mundt F., and Kassambara A. (2020) factoextra: Extract and Visualize the Results of Multivarate Data Analysis. Available at https://rpkgs.datanovia.com/factoextra/index.html

National Weather Service (2021) U.S. Climate Normals. Available at https://www.ncei.noaa.gov/products/land-based-station/us-climate-normals

Ngere L., Burke J.M., Morgan J.L.M., Miller J.E., and Notter D.R. (2018) J. Anim. Sci. 96:1590–9. https://doi.org/10.1093/jas/sky064

Notter, D.R. (1998) J. Anim. Sci. 76(9):2324–2330. https://doi.org/10.2527/1998.7692324x

R Core Team (2020). R: A language and environment for statistical computing. Available at https://www.R-project.org/

Rowan T.N., Durbin H.J., Seabury C.M., Schnabel R.D., and Decker J.E. (2021) PLoS Genet. 17(7):e1009652. https://doi.org/10.1371/journal.pgen.1009652

Shiflett J.S. (2017) Sheep Industry Economic Impact Study. Available at https://d1cqrq366w3ike.cloudfront.net/http/DOCUMENT/SheepUSA/ASI%202016%20Sheep%20Industry%20Economic%20Impact%20Analysis.pdf

Stewart W.C., Scott D.M., Howell S.B., Kaplan R.M., Roeder B.L., *et al.* (2020) Sheep and Goat Res. J. 35:30-37.

157. Bias and heritability of the autocorrelation based on longitudinal data used as resilience indicator

H.A. Mulder

Wageningen University & Research Animal Breeding and Genomics, P.O. Box 338, 6700 AH Wageningen, the Netherlands; han.mulder@wur.nl

Abstract

To genetically improve resilience, the autocorrelation between subsequent deviations from fitted curves of longitudinal data was proposed. Autocorrelation is related to the speed of recovery after a disturbance: a high autocorrelation means slow recovery, while a low autocorrelation means fast recovery. It is known that estimated autocorrelations are biased, while the genetic properties of the autocorrelation are largely unknown. The aims of this research were to investigate the bias and the heritability of the autocorrelation. Deterministic expressions were derived and evaluated with Monte Carlo simulation. The estimated autocorrelation was 0.2-0.3 lower than the true value when the number of records was 10. The heritability was between 0.05 and 0.10 in most situations. It is recommended to have at least 50 records per animal. This shows good opportunities for genetic improvement. This study is a first step towards better understanding of the mathematical and genetic properties of the autocorrelation.

Introduction

In recent studies on genetics of resilience indicators, the lag-one autocorrelation, or simply called autocorrelation, based on subsequent deviations from fitted curves of longitudinal data from an animal, was found to be a promising resilience indicator. Berghof *et al.* (2019) and Poppe *et al.* (2020, 2021a) showed that the autocorrelation is a heritable trait, with a heritability between 0.05 and 0.1. Autocorrelation based on milk yield deviations was related to recovery from environmental perturbations, e.g. from a heat wave or an unspecified herd disturbance (Poppe *et al.*, 2021b). Cows with genetically a low autocorrelation had a faster recovery than cows with a high autocorrelation. This means that cows with a low autocorrelation need less time until their milk yield reaches normal levels after a disturbance than cows with a high autocorrelation. Genetic correlations between the autocorrelation based on milk yield deviations and health and longevity traits were weak in dairy cattle (Poppe *et al.*, 2020). No relationships were found between the autocorrelation based on body weight deviations and lesion scores or mortality upon a disease challenge in chickens (Berghof *et al.*, 2019). These weak or absent relationships may be partly due to properties of the autocorrelation.

The autocorrelation is new to quantitative geneticists, although it has been studied for many years in resilience research (Scheffer *et al.* 2015) or in research using time series (Arnau and Bono, 2001). It is already known for many years that estimates of autocorrelation are biased, especially with a small number of records (Kendall, 1954; Marriott and Pope, 1954; Arnau and Bono, 2001). The mathematical properties of the autocorrelation are largely unknown for geneticists. Furthermore, there is no equation to understand the relationship between the heritability of the autocorrelation and the number of records. Such an equation would help in optimizing breeding schemes to genetically improve resilience.

The aim of this study was to investigate the bias and the heritability of the estimated autocorrelation. Monte Carlo simulation was used to evaluate the deterministic equations.

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_157, © H.A. Mulder 2022

Materials & methods

Genetic model and Monte Carlo simulation. Phenotypes were generated following a Markov process with autoregressive environmental effects. For simplicity, the phenotype was equal to the environmental effects and did not include additive genetic effects on the level of the phenotype, i.e. the normal additive genetic variance utilized in breeding programs:

$$P_{i,t} = \mu_P + E_{i,t} = \mu_P + r_{a_i} E_{i,t-1} + e_{i,t} \quad (1)$$

where $P_{i,t}$ is the phenotype at time t of individual i, μ_p is the mean of the trait and assumed to be zero, r_{a_i} is the lag-one autocorrelation, or simply called autocorrelation, of individual i, $E_{i,t} = r_{a_i} E_{i,t-1} + e_{i,t}$ is the environmental effect at time t and $e_{i,t}$ is the random environmental effect specific to time t. Since $\mu_p = 0$, $E_{i,t} = P_{i,t}$ here. The initial environmental variance σ^2_E, e.g. for the first environmental effect to be sampled $E_{i,l}$, was 1.0. The variance of $e_{i,t}$ was $1 - r^2_{ai}$, so that the phenotypic variance is one. The r_{a_i} had a mean μ_{r_a} a random additive genetic effect A_{r_i} ($N(0, \sigma^2_{A_r})$) and a random environmental effect E_{r_i} ($N(0, \sigma^2_{E_r})$):

$$r_{a_i} = \mu_{r_a} + A_{r_i} + E_{r_i} \quad (2)$$

Monte Carlo simulations were used to evaluate the deterministic expression for the estimated autocorrelation $\widehat{r_a}$. In the simulations to evaluate the expression for the estimated autocorrelation, $\sigma^2_{A_r} = 0$ and $\sigma^2_{E_r} = 0$, while μ_{r_a} was varied between -0.9 and 0.9 and the number of records was varied between 10 and 300. In the simulations to evaluate the simple expression for heritability, the maximum heritability was set to 0.1 and $\sigma^2_{A_r} = 0.003$ and $\sigma^2_{E_r} = 0.027$, which approximately resemble estimated parameters by Poppe *et al.* (2020). The results were averages of 100 replicates.

Deterministic equation for the estimated autocorrelation. The estimated autocorrelation $\widehat{r_a}$ can be approximated with a Taylor series approximation (Lynch and Walsh, 1998):

$$\widehat{r_a} \cong \frac{\sigma_{E_t \widehat{E_{t-1}}}}{\widehat{\sigma^2_E}} \left(1 + \frac{var(\widehat{\sigma^2_E})}{\widehat{\sigma^2_E}} - \frac{cov(\widehat{\sigma^2_E}, \sigma_{E_t \widehat{E_{t-1}}})}{\widehat{\sigma^2_E} \sigma_{E_t \widehat{E_{t-1}}}}\right) \quad (3)$$

where $var(\widehat{\sigma^2_E})$ is the variance of $\widehat{\sigma^2_E}$, i.e. the variance of a variance estimate, $cov(\widehat{\sigma^2_E}, \sigma_{\widehat{E_t E_{t-1}}})$ is the covariance between $\widehat{\sigma^2_E}$ and $\sigma_{\widehat{E_t E_{t-1}}}$ i.e. a covariance between a variance and a covariance estimate. The required expressions for $\widehat{\sigma^2_E}$, $\sigma_{\widehat{E_t E_{t-1}}}$, $var(\widehat{\sigma^2_E})$ and $cov(\widehat{\sigma^2_E}, \sigma_{\widehat{E_t E_{t-1}}})$ were derived and will be shown in a full paper.

Simple equation for the heritability. The heritability of the autocorrelation $h^2_{r_a}$ is a function of the number of records. The number of records determines the sampling variance of the autocorrelation, which can be approximated as $\sigma^2_{r_a} \cong \frac{(1 - r^2_a)^2}{n - 2}$ (Lynch and Walsh, 1998), where n is the number of records. The heritability can therefore be approximated as:

$$h^2_{r_a} = \frac{\sigma^2_{A_r}}{\sigma^2_{A_r} + \sigma^2_{E_r} + \sigma^2_{r_a}} = \frac{\sigma^2_{A_r}}{\sigma^2_{A_r} + \sigma^2_{E_r} + \frac{(1 - r^2_a)^2}{n-2}} \quad (4)$$

Equation 4 makes it feasible to quickly evaluate the effect of the number of records on the heritability of the autocorrelation and therefore also the accuracy of selection when used as selection criterion.

Results

The estimated autocorrelation was substantially lower than the true autocorrelation, indicating bias, as shown in Figures 1A and 1B. The curves show that the bias was as large as 0.2-0.29 when the number of records was 10. The bias decreased slowly when the number of records increased, but even with 200-300

records the bias was 0.005-0.013 when the autocorrelation was 0.3 or 0.5. The bias was larger when the autocorrelation was larger. The deterministic prediction using Equation 3 resembled results from Monte Carlo simulation, while a small deviation was observed when the number of records was 10. If an acceptable bias is smaller than 0.05, the minimum number of records is 40 when the autocorrelation is 0.3 and >50 when the autocorrelation is 0.5. In summary, the estimated autocorrelation can be substantially lower than its true value; therefore it is recommended to have at least 50 records per individual.

The simple equation for heritability (Equation 4) accurately predicted the heritability when the number of records was large, e.g. 80-100, or when the autocorrelation was 0.1 (Figure 2A). However, when the number

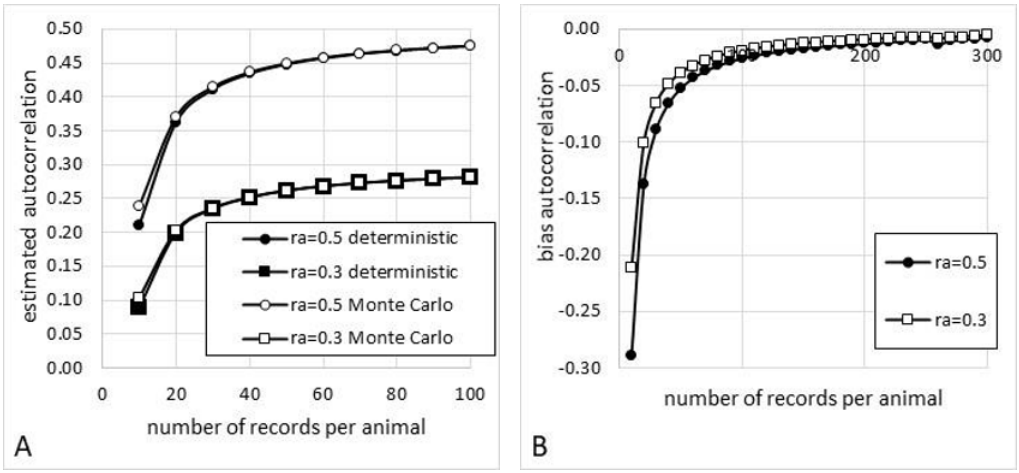


Figure 1. The estimated autocorrelation (panel A) and its bias (estimated − true autocorrelation; panel B) as a function of the number of records per animal when the true autocorrelation (ra) is 0.3 and 0.5 using deterministic Equation 3 (panel A and B) or Monte Carlo simulation (panel A).

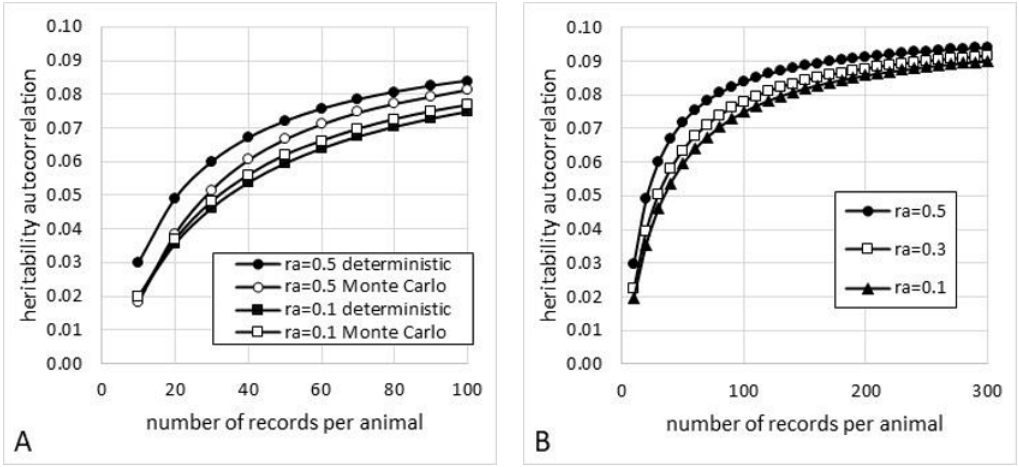


Figure 2. The heritability of the autocorrelation as a function of the number of records when the mean autocorrelation is 0.1, 0.3, 0.5 and the maximum heritability is 0.1 using the deterministic Equation 4 or Monte Carlo simulation.

of records was smaller than 50 and the autocorrelation was 0.5, the simple equation overpredicted the heritability by 8-65%. Figure 2B shows that the heritability was larger when the mean autocorrelation was higher. Given a number of records, the heritability was lowest when the mean autocorrelation was zero. The increase in heritability was small when the number of records was larger than 200. Figure 2B shows that the heritability was between 0.06 and 0.1 when the number of records was at least 50 and when the maximum heritability was 0.1.

Discussion

In this study, the bias in the autocorrelation and the heritability of the autocorrelation were investigated and deterministic equations were presented. The bias in the autocorrelation was already studied in detail by Marriot and Pope (1954) and Kendall (1954). The bias is due to the dependency between the estimated covariance and the estimated variance, which is shown in Equation 3 by the term $\frac{cov(\widehat{\sigma_E^2},\widehat{\sigma_{E_{t,E_{t-1}}}})}{\widehat{\sigma_E^2}\widehat{\sigma_{E_{t,E_{t-1}}}}}$. The estimated autocorrelation may be corrected by adding the predicted bias based on the number of records (Arnau and Bono, 2001). Alternatively, a fixed polynomial regression on the number of records may be included in the model to account for the non-linear effect of number of records on the estimated autocorrelation.

The bias in the mean autocorrelation also affects the heritability of the autocorrelation, because the genetic covariance between the breeding value for the autocorrelation and the estimated autocorrelation, i.e. the numerator of the heritability, was smaller than the true value. Furthermore, the variance of the autocorrelation, the denominator, was underestimated in Equation 4. However, the simple equation gave a very good prediction of the heritability when the number of records was larger than 50. The equation can be used to get more insight in the accuracy of selection for different types of longitudinal data in animal breeding. In general, given the range of heritabilities between 0.05 and 0.10, there are good opportunities for genetic improvement of resilience by selecting on a lower autocorrelation. This study is a step forward in quantitative genetic understanding of the autocorrelation used as resilience indicator in breeding.

References

Arnau, J., Bono, R. (2001) Qual. Quant. 35:365-387. https://doi.org/10.1023/A:1012223430234

Berghof, T.V.L., Bovenhuis, H., Mulder, H.A. (2019) Front. Genet. 10:1216 https://doi.org/10.3389/fgene.2019.01216

Kendall, M.G. (1954) Biometrika 41:403-404. https://doi.org/10.1093/biomet/41.3-4.403

Lynch, M., Walsh, B. (1998) Genetics and analysis of quantitative traits. Sinauer Associates, Inc. Publishers, Sunderland, Massachusetts, US.

Marriott, F.H.C., Pope, J.A. (1954) Biometrika 41:390-402. https://doi.org/10.2307/2332719

Poppe, M., Bonekamp, G., van Pelt, M.L., Mulder, H.A. (2021a) J. Dairy Sci. 104(2):1967-1981. https://doi.org/10.3168/jds.2020-19245

Poppe, M., Mulder, H.A., Veerkamp, R.F. (2021b) J. Dairy Sci. 104(7):8094-8106. https://doi.org/10.3168/jds.2020-19817

Poppe, M., Veerkamp, R.F., Van Pelt, M.L., Mulder, H.A. (2020) J. Dairy Sci. 103(2):1667-1684. https://doi.org/10.3168/jds.2019-17290

Scheffer, M., Carpenter, S.R., Dakos, V., van Nes, E.H. (2015) Annu. Rev. Ecol. Evol. Syst. 46:145-167. https://doi.org/10.1146/annurev-ecolsys-112414-054242

158. The possibilities of estimating relationships between traits with and without micro-genetic environmental sensitivity

M.D. Madsen*, J.H.J. van der Werf and S. Clark

[1]*School of Environmental and Rural Science, University of New England, Armidale, NSW 2351, Australia; mmadsen3@myune.edu.au*

Abstract

Genotype by environment interaction due to short term, localised differences in environmental conditions is caused by the animal's underlying micro-genetic environmental sensitivity (GES). Many traits across a variety of livestock species exhibit micro-GES, however little is known about the genetic relationship between environmental sensitivity in a statistical sense e.g. estimated micro-GES and more direct measure of sensitivity such as disease resistance. This paper investigated the possibility of using multi-trait models including a double hierarchical generalised linear model to examine the genetic relationships between a trait with and a trait without micro-GES. A simulation study was conducted with varying genetic parameters and data structures. Results showed that non-zero genetic correlations had the greatest impact on the accuracy of the estimated parameters, and resulted in an overestimation of the non-zero genetic correlations themselves.

Introduction

Genetic variation in micro-genetic environmental sensitivity (GES) is the underlying cause of genotype by environment interactions (G×E) in response to short-term or localized environmental differences impacting individual animals (termed micro-environments) (Hill and Mulder, 2010). Micro-GES results in variability of phenotype, e.g. a sire with high micro-GES would have offspring with more variable phenotypes than the offspring of a sire with low micro-GES when reared in the same macro-environment (definable environments often containing multiple animals). Micro-GES have been found in multiple species and traits (see e.g. Rönnegård *et al.*, 2010; Felleki *et al.*, 2012). It remains to be determined whether there is a genetic relationship between micro-GES of routinely measured traits and traits, such as disease resistance, which is more expensive to measure and less extensively measured. If such relationships exists, micro-GES of routinely measured traits could potentially be used as indication of for direct measurements of disease resistance. Obtaining the genetic relationship would require estimating the genetic correlation between the breeding values for micro-GES of one trait and the additive genetic value of a second trait, which can be estimated by fitting a multi-trait model with a double hierarchical generalised linear model (DHGLM) fitted for one trait. We propose to investigate the accuracy of this estimate and how this estimate depends on genetic parameters and data structure.

Materials & methods

Simulations. In all scenarios, two traits were simulated. Trait 1 (T1) could be considered a disease resistance trait or a similar trait that did not exhibit micro-GES. Trait 2 could be considered a production trait or similar trait with both a mean genetic potential (T2) and genetic component due to micro-GES (T2d) expressed as a genetic heterogeneity of the environmental variance. Phenotypes were simulated following

$$y_{\mathrm{T1},i} = \mu + a_{\mathrm{T1},i} + e_{\mathrm{T1},i} \quad (1)$$

$$y_{\mathrm{T2},i} = \mu + a_{\mathrm{T2},i} + \exp\left(0.5\, ln\left(\sigma^2_{e_{\mathrm{T2}_d}}\right) + 0.5 a_{T2_d,i}\right)\varepsilon_i \quad (2)$$

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_158, © M.D. Madsen *et al.* 2022

where $y_{T1,i}$ and $y_{T2,i}$ were the phenotypes of animal i for T1 and T2 respectively, $a_{x,i}$ was the true breeding value (TBV) of animal i for trait x ($x \in \{T1, T2, T2d\}$). The residual of T1 ($e_{T1,i}$), and the animal specific random environmental variable (ε_i) were independently distributed following $N(0,1)$. The environmental variance of the exponential model ($\sigma^2_{e_{T2_d}}$) was 1.

The TBVs were drawn from $MVN(0, \begin{bmatrix} \sigma^2_{a_{T1}} & r_{a_{T1}a_{T2}}\sigma_{a_{T1}}\sigma_{a_{T2}} & r_{a_{T1}a_{T2_d}}\sigma_{a_{T1}}\sigma_{a_{T2_d}} \\ & \sigma^2_{a_{T2}} & r_{a_{T2}a_{T2_d}}\sigma_{a_{T2}}\sigma_{a_{T2_d}} \\ & & \sigma^2_{a_{T2_d}} \end{bmatrix} \otimes A)$, where A is the additive genetic relationship matrix derived from the pedigree. $\sigma^2_{a_{T1}}$, $\sigma^2_{a_{T2}}$ and $\sigma^2_{a_{T2_d}}$ were the additive genetic variance of T1, T2 and T2d, respectively, while $r_{a_{T1}a_{T2}}$, $r_{a_{T1}a_{T2_d}}$ and $r_{a_{T2}a_{T2_d}}$ were the genetic correlations between the TBVs for T1 and T2, T1 and T2d and T2 and T2d, respectively.

For both traits 20,000 phenotypes were simulated on the offspring of 200 sires (100 offspring per sire). The percentage of records kept for analysis of T1 ($n_{T1\%}$) was varied to investigate the model's sensitivity to unequal information availability amongst the traits. All combinations of $\sigma^2_{a_{T1}} = \{0.5, 0.3, 0.2\}$, $r_{a_{T1}a_{T2}} = \{0, 0.5\}$, $r_{a_{T1}a_{T2_d}} = \{0, 0.5\}$ and $n_{T1\%} = \{100\%, 50\%, 10\%\}$ were simulated, resulting in a total of 36 scenarios. For all scenarios $\mu_{T1} = \mu_{T2} = \mu_{T2_d} = 0$, $\sigma^2_{a_{T2}} = 0.3$, $\sigma^2_{a_{T2_d}} = 0.1$ and $r_{a_{T2}a_{T2_d}} = 0.3$ were used.

Statistical analysis. For all scenarios the phenotypes were analysed using a multivariate model, combining a linear mixed model and a DHGLM, both fitted at the sire level.

$$\begin{bmatrix} \boldsymbol{y}_{T1} \\ \boldsymbol{y}_{T2} \\ \boldsymbol{y}_{T2_d} \end{bmatrix} = \begin{bmatrix} \boldsymbol{\mu}_{T1} \\ \boldsymbol{\mu}_{T2} \\ \boldsymbol{\mu}_{T2_d} \end{bmatrix} + \begin{bmatrix} \boldsymbol{Z}_{T1} & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{Z}_{T2} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{Z}_{T2_d} \end{bmatrix} \begin{bmatrix} \boldsymbol{s}_{T1} \\ \boldsymbol{s}_{T2} \\ \boldsymbol{s}_{T2_d} \end{bmatrix} + \begin{bmatrix} \boldsymbol{e}_{T1s} \\ \boldsymbol{e}_{T2s} \\ \boldsymbol{e}_{T2s_d} \end{bmatrix} \quad (3)$$

where $\boldsymbol{y}_{T1}$, $\boldsymbol{y}_{T2}$ and $\boldsymbol{y}_{T2_d}$ were vectors containing the response variables for T1, T2 and T2d, respectively, $\boldsymbol{\mu}_x$, $\boldsymbol{Z}_x$, $\boldsymbol{s}_x$ and $\boldsymbol{e}_x$ were the population mean, design matrices relating phenotypes to sires, additive genetic effect of sires, and residuals of trait x, respectively. The distribution assumptions for the additive genetic sire effects were:

$$\begin{bmatrix} \boldsymbol{s}_{T1} \\ \boldsymbol{s}_{T2} \\ \boldsymbol{s}_{T2_d} \end{bmatrix} \sim MVN\left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma^2_{s_{T1}} & \sigma_{s_{T1}s_{T2}} & \sigma_{s_{T1}s_{T2_d}} \\ & \sigma^2_{s_{T2}} & \sigma_{s_{T2}s_{T2_d}} \\ & & \sigma^2_{s_{T2_d}} \end{bmatrix} \otimes \mathbf{A} \right)$$

where A is the additive genetic relationship matrix among sires and ⊗ was a Kronecker product. The distribution assumptions for the residuals were:

$$\begin{bmatrix} \boldsymbol{e}_{T1s} \\ \boldsymbol{e}_{T2s} \\ \boldsymbol{e}_{T2s_d} \end{bmatrix} \sim MVN\left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \mathbf{I_{T1}} & 0 & 0 \\ 0 & \boldsymbol{W}^{-1} & 0 \\ 0 & 0 & \boldsymbol{W}_d^{-1} \end{bmatrix} \right),$$

$\mathbf{I_{T1}}$ was a $n_{T1} \times n_{T1}$ identity matrix ($n_{T1} = \frac{n_{T1\%}}{100} * 20000$), $\boldsymbol{W} = diag(\widehat{\boldsymbol{y_d}})^{-1}$ and $\boldsymbol{W}_d = diag\left(\frac{1 - \boldsymbol{h}_{T2}}{2}\right)$. $\boldsymbol{h}_{T2}$ was the diagonal element of the part of the hat-matrix corresponding to $\boldsymbol{y}_{T2}$. The response variables in $\boldsymbol{y}_{T1}$ and $\boldsymbol{y}_{T2}$ were the simulated phenotypes and $y_{T2_d} = \frac{\hat{e}^2_{T2}}{1 - h_{T2}}$, where $\hat{\boldsymbol{e}}_{T2}$ was the estimated residual of $\boldsymbol{y}_{T2}$ (Rönnegård *et al.*, 2010). The AI-REML package of DMU (Madsen and Jensen, 2013) was used to estimate both variance components. The model in (3) was iterated until convergence and $\boldsymbol{y}_{T2_d}$, $\boldsymbol{W}$ and $\boldsymbol{W}_d$ was updated in each round.

Post-analysis calculations. The additive genetic variances and estimated breeding values (EBVs) were converted from the level of sire to the animal level as described by Madsen *et al.* (2021). The sensitivity of the model to the varied parameters were assessed as the magnitude of any deviation in the means of the estimated variance parameters across converged replicates from their true values and whether the means deviated significantly from the true values. Significances were based on the 95% confidence interval derived from the empirical SEM across converged replicates. The convergence percentage was also assessed.

Results

The alternative genetic parameters used to simulate each traits had minimal impact on the estimated genetic variance of T1 and T2, and the estimated additive genetic variances of T1 was within -0.01 to 0.02 of the true value (corresponding to 2 to 10% of the true value) across all scenarios. The estimated additive genetic variance of the mean of T2 was within ±0.01 (3%) of the true value. There was a tendency for the estimated genetic variance of the dispersion of T2 to be overestimated in scenarios where the true genetic variance of T1 was 0.2 or 0.5 and the true genetic correlation between T1 and T2d was 0.5. The overestimation of the estimated additive genetic variance of the dispersion of T2 was at most 0.01 (10%) of the true value for the dispersion. The SEMs of the estimated additive genetic variances ranged from 0.00 to 0.01, regardless of the varied parameters. In the scenarios where the true genetic correlation between T1 and the mean of T2 or between T1 and the dispersion of T2 was 0, the estimated genetic correlation ranged from -0.04 to 0.03. If the true genetic correlation was 0.5 instead, the estimates ranged from 0.50 to 0.55. The true genetic correlation between the mean and dispersion of T2 was fixed at 0.3 across all scenarios and the estimates ranged from 0.30 to 0.35. The number of records used for analysis of T1 did not systematically impact any of the estimated parameters, except the convergence percentage. 50% of the scenarios where only 10% of the records were used for analysis had <100% convergence (range 93-99) and one scenario where 50% of the records were used had 99% convergence.

Table 1. Estimated genetic correlations (SEM) from the scenarios where the true additive genetic variance of trait 1 was 0.3.

$r_{a_{T1}a_{T2}}$[1]	$r_{a_{T1}a_{T2_d}}$[1]	$n_{T1\%}$[2]	$\hat{r}_{a_{T1}a_{T2}}$[3]	$\hat{r}_{a_{T1}a_{T2_d}}$[3]	$\hat{r}_{a_{T2}a_{T2_d}}$[3]
0	0	100	0.01 (0.01)	0.02 (0.01)	0.34
		50	-0.01 (0.01)	0.00 (0.01)	0.33 (0.01)
		10	0.00 (0.01)	0.02 (0.02)	0.33 (0.01)
	0.5	100	-0.02 (0.01)	0.51 (0.01)	0.33 (0.01)
		50	0.00 (0.01)	0.54 (0.01)	0.30 (0.01)
		10	-0.01 (0.01)	0.50 (0.02)	0.33 (0.01)
0.5	0	100	0.52 (0.01)	0.01 (0.01)	0.32 (0.01)
		50	0.51 (0.01)	0.00 (0.01)	0.33 (0.01)
		10	0.54 (0.01)	-0.01 (0.02)	0.31 (0.01)
	0.5	100	0.52 (0.01)	0.51 (0.01)	0.32 (0.01)
		50	0.52 (0.01)	0.51 (0.01)	0.34 (0.01)
		10	0.53 (0.01)	0.51 (0.02)	0.31 (0.01)

[1] The true genetic correlations between trait 1 and the mean potential ($r_{a_{T1}a_{T2}}$) and dispersion ($r_{a_{T1}a_{T2_d}}$) of trait 2.

[2] The percentage of records kept for trait 1 ($n_{T1\%}$)

[3] The estimated genetic correlations between trait 1 and the mean potential of trait 2 ($\hat{r}_{a_{T1}a_{T2}}$), trait 1 and the dispersion of trait 2 ($\hat{r}_{a_{T1}a_{T2_d}}$) and mean and dispersion of trait 2($\hat{r}_{a_{T2}a_{T2_d}}$).

Discussion

In general, the presence of non-zero genetic correlations between T1 and either mean or dispersion of T2 caused the estimates of non-zero genetic correlations to be up to 10% overestimated. Estimation of the additive genetic variance of T1 and T2 was generally accurate. The additive genetic variance for dispersion (T2d) tended to be overestimated when the true genetic variance of T1 was not equal to 0.3 (i.e. $\sigma^2_{s_{T1}} \neq \sigma^2_{s_{T2}}$), and the genetic correlation between T1 and T2d was 0.5. However, this overestimation was slight and the estimates ranged from 0.10 to 0.11. The convergence percentage was generally high, except in a proportion of scenarios where not all of T1 records were used. In summary, it is possible to obtain accurate measurements of additive genetic variances from a multi-trait model with a routinely measured trait exhibiting micro-GES and less extensively measured trait such as disease resistance. In the presence of a genetic correlations between the non-GES trait and the mean or dispersion of the micro-GES trait, the model will estimate a genetic correlation, though it may be slightly overestimated. Similarly, if there is no genetic correlations the model will indicate so. This allows for identifying traits that could be used to shed light on traits, like disease resistance, which are harder to obtain.

The family structure in the current study was very well balanced with sires having 200 offspring each and thus the same amount of information to base their EBVs on. Further research should be done to investigate the impacts of unbalanced data structures. However it has been shown that DHGLMs are more sensitive to average family size than variability in family size (Madsen *et al.*, 2021). Similar results would be expected for multi-trait models including DHGLMs.

References

Falconer, D. S., and T. F. C. Mackay. 1996. Introduction to quantitative genetics. Pearson Education Limited, Essex, United Kingdom.

Felleki, M., D. Lee, Y. Lee, A. R. Gilmour, and L. Rönnegård. 2012. Estimation of breeding values for mean and dispersion, their variance and correlation using double hierarchical generalized linear models. 94(6):307-317. https://doi.org/10.1017/S0016672312000766

Hill, W. G., and H. A. Mulder. 2010. Genetic analysis of environmental variation. Genet. Res. 92(5-6):381-395. https://doi.org/10.1017/S0016672310000546

Madsen, M. D., J. van der Werf, V. Börner, H. A. Mulder, and S. Clark. 2021. Estimation of macro- and micro-genetic environmental sensitivity in unbalanced datasets. Animal (Cambridge, England) 15(12):100411-100411. https://doi.org/10.1016/j.animal.2021.100411

Madsen, P., and J. Jensen. 2013. A user's guide to DMU. A package for analysing multivariate mixed models version 6:1-33.

Rönnegård, L., M. Felleki, F. Fikse, H. A. Mulder, and E. Strandberg. 2010. Genetic heterogeneity of residual variance - estimation of variance components using double hierarchical generalized linear models. Genetics, Selection, Evolution: GSE 42(1):8-8. https://doi.org/10.1186/1297-9686-42-8

159. Selective breeding for survival to *Vibrio harveyi* in European seabass (*Dicentrarchus labrax*) is robust to different feeding environments

M. Besson[1*], R. Morvezen[1], A. Bestin[1], Y. François[1], A. Bajek[2], J. Brunier[2], T. Morin[3], M. Vandeputte[4], M. Dupont-Nivet[5], M. Izquierdo[6], D. Barkas[7] and P. Haffray[1]

[1]SYSAAF, Campus de Beaulieu, Rennes, France; [2]EMGI, Route de Enrochements, Gravelines, France; [3]ANSES, Viral Fish Diseases Unit, Plouzané, France; [4]MARBEC, Univ Montpellier, CNRS, Ifremer, IRD, INRAE, Palavas-les-Flots, France; [5]Université Paris-Saclay, INRAE, AgroParisTech, GABI, Jouy-en-Josas, France; [6]Grupo de Investigación en Acuicultura, Universidad de Las Palmas de Gran Canaria, Las Palmas, Spain; [7]Nireus Aquaculture, Koropi, Greece; mathieu.besson@inrae.fr

Abstract

Vibrio harveyi is a major pathogen affecting aquaculture. Selective breeding could be a strategy to improve European sea bass resistance but genetic parameters are unknown. Additionally, the on-going replacement of marine-based ingredients by plant-based ingredients in fish diets could affect the efficiency of a breeding program due to genotype-by-diet interaction (G×D). Hence, we estimated the genetic parameters and G×D of survival to *V. harveyi*. Two groups of 1,200 seabass were fed (A) a custom feed enriched with plant-based ingredients and (B) a commercial feed. Then, the fish were injected with a *V. harveyi* suspension, mortality was recorded and fish were genotyped. With both diets, survival was moderately heritable (0.16 and 0.18, respectively), and the genetic correlation for survival between diets was 0.88, indicating that G×D was low. These results are encouraging to develop robust breeding program for survival to *V. harveyi*.

Introduction

Vibrio harveyi is considered as a highly problematic pathogen in aquaculture across species and countries as it cannot be efficiently controlled by commercial vaccines, due to low antigen specificity. In commercial farms, the mortality caused by *V. harveyi* can reach 50% in case of an outbreak. In this context, selective breeding could be an effective strategy to improve fish resistance to *V. harveyi*. Indeed, selective breeding has become a prime strategy to tackle disease resistance in fish species and many breeding programs include disease resistance in their breeding goals (Chavanne *et al.*, 2016).

In parallel, fish farming is facing another challenge: the replacement of fish meal and fish oil in the diets by plant-based ingredients (soy, corn gluten, etc.). Such replacement could have several impacts on farmed fish. First, it could affect the sensitivity of fish to pathogens. Second, it could decrease the efficiency of fish breeding programs due to genotype-by-diet interactions (G×D). When G×D occurs, the ranking of families fed a diet A differs from the ranking of families fed a diet B. More generally, genotype-by-environment interactions (G×E) have already been observed on fish reared in different farming systems (e.g. Gulzari *et al.*, 2022). In their review paper, Sae-Lim *et al.* (2016) reported substantial re-ranking for growth-related traits (average genetic correlation=0.72) and for survival-related traits (average genetic correlation=0.54). The consequence of re-ranking due to G×D is of great concern because the selection of the nucleus would lead to lower genetic responses in commercial environments where fish are reared and grow with a different feed than in the nucleus. Our goal was therefore to estimate the genetic parameters of survival to *V. harveyi* and to study potential G×D in seabass to develop a robust and efficient breeding program improving the survival to *V. harveyi*.

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_159, © M. Besson *et al.* 2022

Materials & methods

Animals. Seabass from 225 families were produced from three full factorial mating design (5 dams crossed with 15 sires each) using artificial fertilization by the EMGI breeding company. The fish were split in two groups and fed with two different iso-lipidic, iso-energetic and iso-proteic feeds. The first group (A) was fed a custom feed, produced by Nireus, using a formulation developed in the Arraina EUproject, with a high level of vegetal ingredients, including wheat protein extract to limit gut injury often caused by soy-based ingredients. The second group (B) was fed a commercial feed (Biomar).

***Vibrio harveyi* challenge.** 1,186 and 1,211 fish from group A and B were challenged to *V. harveyi* in the SYSAAF-ANSES Fortior Genetics platform (Plouzané, France) at 345 days post fertilization. The fish were split into 12 tanks (5 test tanks and 1 control for each group A and B). All fish were injected with 100 µl of bacterial suspension. The negative controls were injected with a solution containing sterile cell culture medium. The mortality was recorded each day during the 14-days challenge period.

Statistical analysis of phenotypes. To test whether the type of feed influences mortality due to *V. harveyi*, we applied a binary logistic regression model using glm function in R. The following model was fitted:

$$log\left[\frac{P(y_i = 1)}{1 - P(y_i = 1)}\right] = \mu + \text{Group}_i + \beta_0 \times \text{BW}_i + \beta_1 \times (\text{BW}_i \times \text{Group}_i) \quad (1)$$

Where y_i is the variable of interest, dead or alive, for fish i. Group_i is the fixed effect of the type of feed (A or B) and BW_i is the body weight of fish i at the start of the challenge.

Genetic analysis. A total of 942 fish from group A and 895 fish from group B were genotyped the 57k SNP array DlabChip (Griot *et al.*, 2021). The genomic data were then used to compute variance components and estimated breeding values for survival using THRGIBBSF90 (Tsuruta and Misztal, 2006). The survivals in each group A and B were considered as different traits to estimate the genetic correlation between both groups. Variance components were estimated using the following bi-variate model:

$$y = Xb + Zu + e \quad (2)$$

Where y is the vector of phenotypes (survival as binary trait; 1 = dead, 2 = alive), X is the design matrix for intercept and regression on initial body weight, b is the vector of intercept and regression coefficient estimates, u is the vector of breeding values and Z the corresponding incidence matrix. It is assumed that u follows a multivariate normal distribution $N(0, G\sigma^2_g)$ with G the genomic relationship matrix proposed by VanRaden, (2008) and σ^2_g is the additive genetic variance. e is the vector of the random residual errors that follows a normal distribution $N(0, I\sigma^2_e)$ with σ^2_e the residual variance and I the identity matrix. The variance components (σ^2_g and σ^2_e) were estimated using a Gibbs sampler with 500,000 iterations, 100,000 of burn-in and one sample was kept every 20 iterations for posterior analysis.

Results

First significant mortality occurred the day after the injection and reached a peak 2 days after the injection (Figure 1). At the end of the challenge, the mortality rate reached 38% for the group fed diet A (dead 468 fish) and 48% for the group fed diet B (dead 546 fish) (Figure 2a).

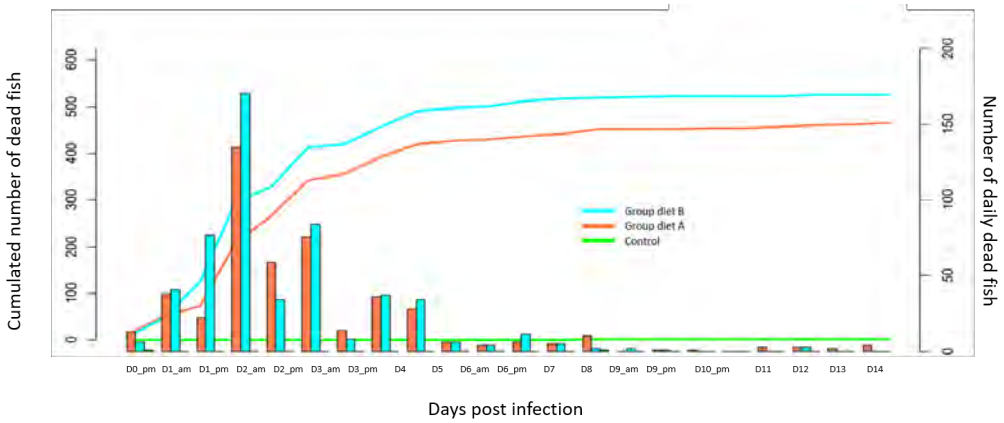


Figure 1. Barplot of the number of dead fish per day (or half day) after *V. haveyi* injection for the group fed diet A and the group fed diet B. The lines represent the cumulated mortality for both groups.

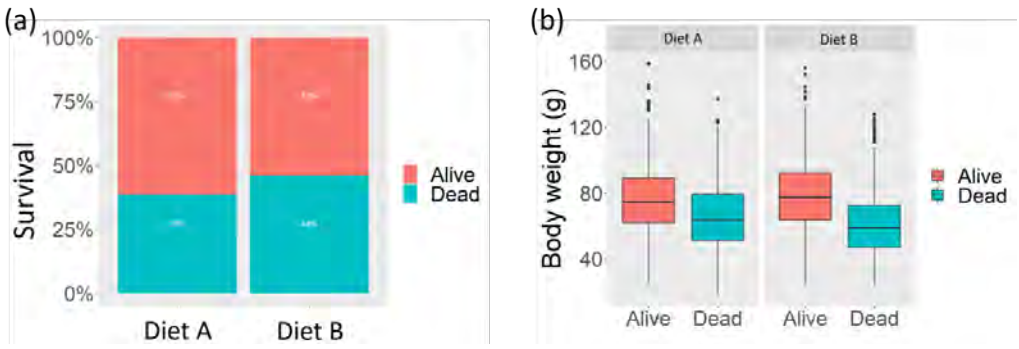


Figure 2. Panel (a) represents the mortality and survival rate at the end of the challenge for both diets A and B. Panel (b) is the boxplot of the body weight at the start of the challenge for surviving and dead fish in both groups.

The logistic regression showed that feed and body weight significantly affected the survival rate of the fish. Fish fed diet B were more likely to die from the infection compared to fish fed diet A (P<0.0001). Also, bigger fish were less likely to die from the infection compared to smaller fish (P<0.0001). The interaction between the type of feed and the weight was also significant (P=0.0001). Survival to *V. harveyi* was moderately heritable, h^2=0.18 (0.07) in the group fed diet B and 0.16 (0.07) in the group fed diet A. Furthermore, we found that the genetic correlation of survival to *V. harveyi* between diet A and B was high (0.88).

Table 1. Heritabilities (on the diagonal) of survivals to *Vibrio harveyi* when fed diet A and when fed diet B. Above the diagonal is the genetic correlation between the two traits.

	Survival with diet A	**Survival with diet B**
Survival with diet A	0.16 (0.07)	0.88 (0.10)
Survival with diet B		0.18 (0.07)

Discussion

G×D has already been investigated in aquaculture (e.g. Le Boucher *et al.*, 2013) but to our knowledge, this study is the first investigating the interaction between the type of feed and the genetic component of resistance to a specific pathogen, i.e. *V. harveyi*. First, we observed differential mortality between small and large fish that could be explained by the injection of the same bacterial dose to all fish. This is why body weight was included as covariable in the genetic model. Then, we found that the type of feed significantly affected the mortality at the challenge. However, the genetic correlation of survival between each diet group was high, indicating that G×D was low and that the type of feed only marginally affected the ranking of families. Additionally, the heritabilities of survivals to *V. harveyi* in both feeding environments were moderate, and substantial variation in the survival of the progeny of the 45 sires used in the mating was observed. These results are encouraging to develop a breeding program specifically targeting survival to *V. harveyi* in seabass. The results are also encouraging for breeding companies that could select fish for better resistance to *V. harveyi* without losing efficiency across feeding environments, which is essential to design an efficient and robust breeding program.

References

Chavanne, H., Janssen, K., Hofherr, J., Contini, F., Haffray, P., *et al.* (2016) Aquaculture international 24, 1287–1307.

Griot, R., Allal, F., Phocas, F., Brard-Fudulea, S., Morvezen, R., *et al.* (2021) Aquaculture 530, 735930. https://doi.org/10.1016/j.aquaculture.2020.735930

Gulzari, B., Komen, H., Nammula, V.R., Bastiaansen, J.W. (2022) Aquaculture 548, 737555.

Le Boucher, R., Vandeputte, M., Dupont-Nivet, M., Quillet, E., Ruelle, F., *et al.* (2013) Journal of animal science 91, 44–56.

Sae-Lim, P., Gjerde, B., Nielsen, H.M., Mulder, H., Kause, A. (2016) Reviews in Aquaculture 8, 369–393.

Tsuruta, S., Misztal, I. (2006) Threshold 3, 4.

VanRaden, P.M. (2008) Journal of dairy science 91, 4414–4423.

160. Can we prevent pathogen adaptation when breeding disease resistant livestock?

A.D. Hulst[1,2*], P. Bijma[2] and M.C.M. De Jong[1]

[1]Quantitative Veterinary Epidemiology, Wageningen University & Research, P.O. Box 338, 6700 AH Wageningen, the Netherlands; [2]Animal Breeding and Genomics, Wageningen University & Research, P.O. Box 338, 6700 AH Wageningen, the Netherlands; dries.hulst@wur.nl

Abstract

Recent research shows that genetic selection has high potential to reduce the prevalence of infectious diseases in livestock. However, as with all interventions targeted at infectious diseases, pathogens might evolve strategies to escape the resistance. Thus, for sustainable breeding, strategies that prevent the invasion of these escape mutants are needed. Here we use a mathematical model of infection transmission that accounts for genetic selection in hosts and pathogens to investigate the conditions under which escape mutants can invade. The results show that genetic selection for resistance typically leads to an 'invasion window', the range in host resistance in which an escape mutant can invade. This window is smallest when resistance is strong. To prevent the invasion of escape mutants, host resistance should be increased fast through the invasion window. This raises questions about the sustainability of the common multi-trait approach when the breeding goal includes infectious disease resistance.

Introduction

Genetic selection has long been considered a potential strategy to combat infectious diseases in livestock (Bishop *et al.,* 2010). Recent theoretical studies show that the response to genetic selection to reduce the prevalence of infectious diseases should be much larger than predicted by common quantitative genetic models (Bijma *et al.,* 2021; Hulst *et al.,* 2021). With each intervention targeted at infectious diseases, however, comes the risk that pathogens evolve strategies to escape the effects of that intervention, with the widespread antibiotic resistance as the most prominent example (Davies and Davies, 2010). Thus, questions arise whether and when pathogens can evolve escape strategies to genetically selected animals, and how this can be prevented.

For some other interventions, strategies exist to prevent the evolution of pathogen escape. An example is the restricted use of antibiotics. antibiotics are preferably used as little as possible, and when used, this should be in sufficiently high dose to kill all bacteria. In this way development of bacterial escape is prevented or at least largely delayed (Walsh, 2003).

For genetic selection of livestock for increased resistance against infectious diseases, however, strategies to prevent pathogen escape have not been investigated as far as we know. In this paper, we use a mathematical model of infection transmission that accounts for artificial genetic selection in the host (i.e. livestock) population and for the evolution of escape by natural selection in the pathogen population, to investigate the conditions under which escape mutants of a pathogen can invade a livestock population. Furthermore, we investigate how such invasion of an escape mutant is influenced by the strength of resistance of the host. Ultimately, we try to find strategies for genetic selection of livestock that limit or prevent the risk of pathogen escape.

Methods & results

In this section we briefly describe and analyse a mathematical model of infection transmission that is tailored to the invasion of escape mutants in a host population under genetic selection for increased resistance (for details see Hulst *et al.,* 2022, in preparation). The starting point of our analysis is a local host population

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_160, © A.D. Hulst *et al.* 2022

endemically infected with the wild type pathogen. We first describe how we model genetic selection for resistance in this host population and show how this affects the prevalence of the wild-type infection for different strength of resistance. Next, we model the invasion of escape mutants in the endemically infected host population and show how the opportunity for invasion depends on the frequency of resistant hosts and the strength of resistance.

Genetic variation in host resistance to wild-type infection. We assume that host resistance is defined by a single locus, which is either fully dominant or fully recessive, such that the host population consists of two types: resistant (with fraction f_R) and non-resistant ($1 - f_R$). Here resistant merely means that animals are less susceptible to the infectious disease, not that they cannot get infected at all.

We model differences in resistance between the two host types through the basic reproduction ratio, which is a key parameter in epidemiology and is defined as the average number of secondary infections caused by a single infected individual in an otherwise fully susceptible population (Diekmann *et al.*, 1990). R_0 has a threshold function; when it is above one an infection may persist in the population, while when R_0 is below one an infection will die out with certainty. Thus, to eradicate an infectious disease, it is essential that genetic selection reduces R_0 below a value of one.

In our case, each of the two host types has its own R_0. For infection with the wild-type (W) pathogen, these are denoted R_{NW} for non-resistant hosts, and R_{RW} for Resistant hosts. Because resistant hosts are less likely to become infected, $R_{RW}<R_{NW}$. The R_0 of the population is a weighted average of these the values, $R_0 = (1 - f_R)R_{NW} + f_R R_{RW}$.

The model tends to an equilibrium, where the number infected individuals of both host types do not change. Figure 1 shows the prevalence of the infection in this equilibrium over the frequency of resistant hosts (f_R), for R_{RW} of 1.1, 0.8 and 0.1. For R_{RW} of 0.1 and 0.8, the prevalence decreases to 0 at a certain f_R, meaning that the disease dies out. This point of dying out occurs at lower f_R for stronger resistance (i.e. lower R_{RW}). When R_{RW} is 1.1, the infection can persist in a fully resistant population, reflected in the nonzero prevalence at f_R=1. In the next section we investigate whether an escape mutant can invade a population that is endemically infected with the wild type pathogen, assuming this population is in the equilibrium as shown in Figure 1.

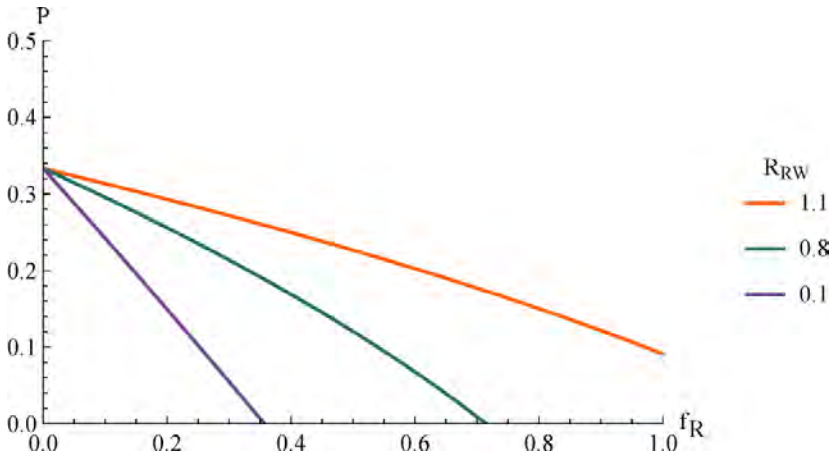


Figure 1. Equilibrium prevalence as a function of the frequency of resistant hosts. R_{NW}=1.5.

Escape mutant invasion. Escape mutants are, by definition, better able to infect resistant hosts than the wild-type pathogen. Furthermore, we assume that the escape mutation comes with a cost in pathogen fitness for infection of non-resistant hosts. We model this again through the reproduction ratios; R_{NE} for infection of non-resistant host with the escape mutant and R_{RE} for infection of resistant hosts with the escape mutant. The conditions above imply that $R_{NE}<R_{NW}$ and $R_{RE}>R_{RW}$.

We assume that escape mutants arise continuously as long as the wild type is present, so we look at their invasion rather than their emergence. Furthermore, we will only look at the invasion of a single type of mutant, with certain R_{NE} and R_{RE}, at a time. Since the mutant arises in a host population that is already infected with the wild type, to determine whether it can invade we need to derive a reproduction ratio for invasion (R_{INV}) in a host population in which the wild type is at endemic equilibrium. This can be done by multiplying the reproduction ratios of the escape mutant with the fraction of non-infected individuals of each type in the equilibrium,

$$R_{INV} = R_{NE}[(1 - f_R) - P_{NW}] + R_{RE}(f_R - P_{RW}) \quad (1)$$

where P_{NW} and P_{RW} denote the endemic prevalence of the wild type in non-resistant and in resistant hosts. If $R_{INV}>1$ the escape mutant can invade, whereas if $R_{INV}<1$, the escape mutant can emerge, but not invade because it is outcompeted by the wild type.

Figure 2 shows R_{INV} as a function of f_R (dashed line) for R_{RW} of 0.8 (green lines) and 0.1 (purple lines). The prevalence of the wild type over f_R is also shown (similar to Figure 1). The threshold value for R_{INV} of 1 is indicated by the dotted line. The invasion possibility of the escape mutant is restricted by two bounds: (1) the lower bound where R_{INV} becomes above 1, the escape mutant can invade from that point; and (2) the upper bound at which the wild type dies out (P=0), from that point the escape mutant cannot emerge anymore because there are no wild-type pathogens to mutate from. We call the range in f_R between these bounds the 'invasion window'. It is clearly visible in Figure 2 that this window is smaller with lower R_{RW} (i.e. stronger resistance). Thus, stronger resistance not only results in faster eradication of the infection, but also corresponds to a lower risk of escape mutant invasion.

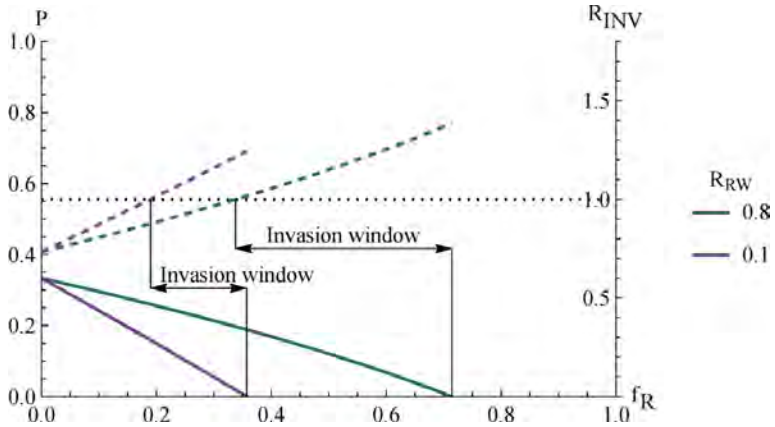


Figure 2. Invasion window for R_{RW} of 0.8 and 0.1. R_{NE}=1.1, R_{RE}=1.5, R_{NW}=1.5.

Discussion

We have demonstrated that an infectious disease can be eradicated from a local population by increasing the frequency of genetically resistant hosts sufficiently. For eradication, it is necessary that the reproduction ratio of the infection in resistant hosts (R_{RW}) is lower than one. Eradication occurs as a consequence of herd immunity, not because resistant individuals cannot become infected. As long as the infection is not eradicated, escape mutants of the pathogen might emerge and invade the host population. The invasion window is bounded by the frequency of resistant hosts from which the escape mutant can outcompete the wild-type pathogen ($R_{INV}>1$) and the frequency of resistance at which the wild type dies out. This window is smaller with lower R_{RW}.

For animal breeding, the above implies that to increase infectious disease resistance sustainably, without escape mutants invading and diminishing the effects of selection, the frequency of resistant hosts in the local population should be increased as fast as possible through the invasion window. A possible strategy to achieve that is by basing herd composition on the resistance status of the animals, such that herds are either on the left (no invasion) or on the right (eradication) side of the invasion window. This is most feasible in species where the animals in production herds are replaced all at once, such as in poultry or fattening pigs.

In genetic improvement of livestock, multi-trait selection has become the standard, because it is superior to other selection strategies for ordinary quantitative traits (Hazel *et al.*, 1994). However, improvement happens typically in small steps for each trait with multi-trait selection, which creates ample opportunity for escape mutants to invade. A relevant question is thus whether the current approach in animal breeding of gradual weak selection for infectious disease resistance is sustainable. On the one hand, it might be more difficult for pathogens to evolve escape strategies when resistance is polygenic, especially if this entails multiple resistance mechanisms. While on the other hand there are clear parallels with the long-term use of low-doses of antibiotics, which has strongly supported the evolution of antibiotic resistant bacteria (Olofsson and Cars, 2007). Further research into this is needed, but if the latter holds, it might be better to use a high dose for a short time, which corresponds to applying a large selection differential for infectious disease resistance within a single or a few generations.

References

Bijma P., Hulst A.D., and De Jong M.C.M. (2021) Genetics. https://doi.org/10.1093/genetics/iyab141

Bishop, S. C., Axford R.F., Nicholas F.W., and Owen J.B. (2010) Breeding for disease resistance in farm animals. CABI, Wallingford, UK.

Davies, J., and Davies D. (2010) Microbiology and Molecular Biology Reviews 74: 417-433. https://doi.org/10.1128/MMBR.00016-10

Diekmann, O., Heesterbeek J. A. P., and Metz J. A. (1990) Journal of mathematical biology 28: 365-382. https://doi.org/10.1007/BF00178324

Hazel, L. N., Dickerson G. E., and Freeman A. E. (1994) Journal of dairy science 77: 3236-3251. https://doi.org/10.3168/jds.S0022-0302(94)77265-9

Hulst, A. D., De Jong M. C. M., and Bijma P. (2021) Genetics 217. https://doi.org/10.1093/genetics/iyab024

Olofsson, S. K., and Cars O. (2007) Clinical Infectious Diseases 45: S129-S136. https://doi.org/10.1086/519256

161. Indicators of disease resilience from complete blood count and *in vitro* immunoassays data from young-healthy pigs

X. Bai[1,2*], J. Cheng[3], F. Fortin[4], J.C.S. Harding[5], M.K. Dyck[1], J.C.M. Dekkers[3], C.J. Field[1], C. Rogel-Gaillard[6], F. Blanc[6] and G.S. Plastow[1]

[1]Livestock Gentec, Department of Agricultural, Food and Nutritional Science, University of Alberta, Edmonton, AB, Canada; [2]Current: Hendrix Genetics, Boxmeer, the Netherlands; [3]Department of Animal Science, Iowa State University, Ames, IA, USA; [4]Centre de Développement du Porc du Québec, Inc., Quebec City, QC, Canada; [5]Department of Large Animal Clinical Sciences, University of Saskatchewan, Saskatoon, SK, Canada; [6]Université Paris-Saclay, INRAE, AgroParisTech, GABI, France; xuechun.bai@hendrix-genetics.com

Abstract

This study estimates heritability of complete blood count (CBC) and *in vitro* immunoassay traits. The traits were measured from blood samples of young-healthy nursery animals that then survived to slaughter during a natural polymicrobial disease challenge in the grow-to-finish stage but had variable production performance. Genetic correlations of CBC and immunoassay traits with performance traits collected from grow-to-finish pigs exposed to the challenge were also estimated, including carcass, average daily gain, and feed intake traits. Most CBC and several immunoassay traits had heritability estimates ranging from 0.11 to 0.46. Several CBC and immunoassay traits showed moderate to high positive (0.25 to 0.51) or negative (-0.92 to -0.37) genetic correlations with performance traits under disease. These results suggest that some CBC and immunoassay traits collected from young-healthy animals before exposure to disease challenge are potential indicator traits for selecting disease resilient pigs that can maintain better performance under pathogen exposure.

Introduction

Infectious diseases have steadily increased morbidity and mortality in multiplication and production herds of the pig industry, which causes significant productivity and economic losses. The current challenge in the pig industry is the hard-to-manage disease caused by a multitude of infectious pathogens that exist around the world. Thus, there is increased interest in selecting disease resilient pigs that can maintain undepressed performance under pathogen exposure, resulting from a combined optimal balance of disease tolerance and resistance mechanisms (Albers *et al.*, 1987, Mulder and Rashidi, 2017). Improving disease resilience can deliver multiple benefits to the pig industry, including more sustainable production, improved animal health, and the potential for managing the polymicrobial challenge common in pig production to reduce antimicrobial use.

Multiple immune traits collected from vaccinated or disease challenged pigs, including antibody titers, immune cell counts, and cytokine levels, have shown relatively high estimates of heritability and genetic correlations with performance (Clapperton *et al.*, 2009; Flori *et al.*, 2011; Bai *et al.*, 2020; Blanc *et al.*, 2021). However, the value of using immune traits collected from young healthy pigs as indicators of disease resilience remains largely unknown. Thus, the objective of this study was to investigate CBC traits and *in vitro* immunoassay traits collected from young-healthy pigs before exposure to disease challenge as potential indicator traits for the selection for disease resilience.

Materials & methods

The study was carried out in accordance with the Canadian Council on Animal Care guidelines. The protocol was approved by the Animal Protection Committee of the Centre de Recherche en Sciences

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_161, © X. Bai *et al.* 2022

Animales de Deschambault (15PO283) and the Animal Care and Use Committee at the University of Alberta (AUP00002227).

Natural polymicrobial disease challenge model. This study is derived from a research project that aimed to investigate indicators and the genetic basis of disease resilience in a wean-to-finish polymicrobial natural disease challenge model (NDCM). The NDCM was established to mimic severe disease challenge in a commercial production environment, as a combination of the circulating pathogens, together with the environment, management, and veterinary strategies; details can be found in Putz *et al.* (2019). Briefly, the NDCM consisted of three phases, the pre-challenge quarantine nursery (19 days on average beginning at ~21 days of age), the challenge nursery (~27 days) and the finishing (~100 days) phases. Every three weeks, a batch of 60 or 75 weaned crossbred (Landrace × Yorkshire) barrows from a high-health multiplier farm were introduced into the NDCM in a continuous flow manner, with approximately 3,300 pigs entering between 2015 and 2019.

Common disease-causing pathogens found in Canadian commercial farms were the primary target of the NDCM, including multiple strains of porcine reproductive and respiratory syndrome virus-2 (PRRSV2), *Mycoplasma hyopneumoniae* (MHP), and various other systemic, respiratory, and enteric pathogens. The challenge nursery was established by introducing pigs from commercial units with known diseases together with the first few batches of healthy pigs. Each subsequent batch of pigs was challenged by exposure to pigs introduced three weeks earlier. After one week of exposure, the older pigs were moved to the finisher. Not every batch was exposed to the same types or doses of pathogens, as disease pressure varied by batch and season, except every batch was confirmed to have been exposed to PRRSV2 and MHP.

Performance traits. Performance traits were collected from grow-to-finish barrows that survived to slaughter from all batches, including average daily gain in the challenge nursery (cNurADG) and in the finisher (FinADG); slaughter traits of carcass weight (CWT), carcass backfat (CBF), carcass loin depth (CLD), dressing percentage (DRS), and lean yield (LYLD); finishing feed intake traits of average daily feed intake (ADFI) and duration (ADFD), feed conversion ratio (FCR), residual feed intake (RFI), the proportion of off-feed days based on feed intake (OFF_{FI}) or feed intake duration (OFF_{DUR}), and day-to-day variation in feed intake (VAR_{FI}) or feed intake duration (VAR_{DUR}). Details of the traits collection, number of records, derivation, and genetic parameters of these traits can be found in Cheng *et al.* (2020).

CBC and *in vitro* immunoassay traits. CBC (n=2,036) from all batches and *in vitro* immunoassays of phagocytosis (n=736) and blood cell stimulation (n=775) from selected batches (Lim *et al.* 2021) were collected from barrows that survived to slaughter when they were in the pre-challenge quarantine nursery at ~26 days of age. Details of the CBC traits are in Bai *et al.* (2020). The phagocytosis assay quantified the percentage of leukocytes showing phagocytosis of FITC-labelled opsonized *E. coli* bacteria by flow cytometry (Phagotest, Glycotope). The percentages of phagocytosis of leukocyte subpopulations were also quantified based on cell morphology. Blood cell stimulation detected and quantified nine cytokines (IL-4, IL-2, IL-8, IL-13, IL-6, IFNg, IL-10, IL-17 and IL-1b) under non-stimulation, and LPS (lipopolysaccharide at 1 μg/ml) and PI (phorbol myristate acetate at 50 ng/ml and ionomycin at 1 μg/ml) 24h-stimulation conditions.

Genotyping and quality control. Genotyping used the 650 K Affymetrix Axiom® Porcine Genotyping Array. Details of genotype quality control are in Putz *et al.* (2019) and the genomic relationship matrix construction are described in Cheng *et al.* (2020).

Variance component analyses. Variance components were estimated by GBLUP using ASReml 4.1 (Gilmour *et al.*, 2015). The general models used in univariate and bivariate analyses were described in Chen *et al.* (2020).

Results

Most CBC and immunoassay traits were estimated to be low to moderately heritable: estimates for CBC traits ranged from 0.11±0.05 to 0.40±0.06, phagocytosis test traits ranged from 0.19±0.09 to 0.48±0.10, and blood cell stimulation traits ranged from 0.01±0.08 to 0.46±0.13. Some CBC and immunoassay traits showed significant (P<0.05) genetic correlations with performance traits and were summarized in Table 1. CBC and phagocytosis test traits were primarily found to be genetically correlated with carcass and feed intake traits, respectively. However, blood stimulation with LPS was genetically correlated with multiple categories of performance traits, including DRS, cNurADG, and ADFI.

Discussion

The significant heritabilities and genetic correlations for CBC and immunoassay traits collected from young-healthy animals with performance under pathogen exposure suggest that they can be indicators for selection of disease resilience in commercial farms. Higher CBC traits of LYM and HCT associated with improved DRS and CWT under disease. However, the greater CBF associated with higher HCT must be noted, indicating decreased lean meat yield. Putz *et al.* (2019) and Cheng *et al.* (2020) showed

Table 1. Significant (P<0.05) estimates of heritability (h^2) and genetic correlations for CBC and immunoassays traits with performance traits (SE in parentheses).

Traits[1]	h^2 (SE)	r_g (SE) with performance traits	
CBC			
LYM	0.28 (0.05)	DRS	0.48 (0.16)
HCT	0.12 (0.05)	CWT	0.45 (0.23)
		CBF	0.37 (0.17)
		LYLD	-0.37 (0.17)
PLT	0.27 (0.05)	ADFD	0.25 (0.12)
Phagocytosis test			
PHAGO	0.48 (0.10)	ADFD	0.32 (0.15)
GRAN_PHAGO	0.33 (0.10)	ADFD	-0.48 (0.16)
		OFF_{DUR}	0.51 (0.22)
NEU_PHAGO	0.19 (0.09)	ADFD	-0.61 (0.22)
LYM_PHAGO	0.48 (0.10)	ADFD	0.31 (0.15)
EOS_PHAGO	0.29 (0.09)	OFF_{DUR}	-0.49 (0.25)
Blood cell stimulation			
IL8_LPS	0.19 (0.11)	DRS	-0.74 (0.32)
		cNurADG	-0.92 (0.34)
		ADFI	-0.49 (0.25)
IFNg_LPS	0.30 (0.13)	cNurADG	-0.46 (0.24)
		ADFI	-0.44 (0.22)
IL6_LPS	0.21 (0.10)	cNurADG	-0.72 (0.26)

[1] LYM (10^3/μl): lymphocyte concentration; HCT (%): hematocrit; PLT: platelet concentration; PHAGO (%): phagocytizing cells among total cells; GRAN/NEU/LYM/EOS_PHAGO (%): phagocytizing cells among granulocytes/neutrophils/lymphocytes/eosinophils; IL8/IFNg/IL6_LPS (pg/ml): IL8/IFNg/IL6 concentration under LPS stimulation.

that higher ADFD and ADFI, but lower OFF_{DUR} were proxy traits of improved disease resilience in the NDCM. Therefore, the higher CBC trait of PLT and phagocytosis test traits of PHAGO, LYM_PHAGO and EOS_PHAGO, but lower GRAN_PHAGO or NEU_PHAGO tended to be associated with greater disease resilience. For blood stimulation traits, lower cytokine levels of IL8, IFNg, and IL6 after LPS stimulation correlated with greater DRS, cNurADG, and ADFI, and therefore, tended to be associated with improved disease resilience.

Immune traits have the potential to be further explored as indicators of disease resilience because the ability to produce effective and appropriate immune responses is closely associated with disease severity and tissue damage, which may be a key player in balancing between immune responses and tissue damage to maintain an animal's performance in response to the disease challenge (Hine *et al.* 2014). The present work contributes to our knowledge of the genetic parameters of immune traits collected from young healthy pigs and their relationships with multiple production performance traits under disease. Further studies are warranted to validate these traits as genetic indicators of resilience and further determine the role of these immune traits in disease resilience.

Acknowledgements

Funding from Genome Canada, Genome Alberta, PigGen Canada, and USDA-NIFA grant # 2017-67007-26144 is acknowledged.

References

Albers G.A.A., Gray G.D., Piper L.R., Barker J.S.F., Le Jambre L.F., *et al.* (1987) Int J Parasitol 17(7):1355–1363. https://doi.org/10.1016/0020-7519(87)90103-2.

Bai X., Putz A.M., Wang Z., Fortin F., Harding J.C.S., *et al.* (2020) Front Genet 11: 216. https://doi.org/10.3389/fgene.2020.00216.

Blanc F., Maroilley T., Revilla M., Lemonnier G., Leplat J.J., *et al.* (2021) Genet Sel Evol 53:24. https://doi.org/10.1186/s12711-021-00614-5.

Chen Y., Cortes L.E.T., Ashley C., Putz A.M., Lim KS., *et al.* (2020) BMC Genomics 21 (648). https://doi.org/10.1186/s12864-020-06994-0.

Cheng J., Putz A.M., Harding J.C.S., Dyck M.K., Fortin F., *et al.* (2020) J Anim Sci 98(8). https://doi.org/10.1093/jas/skaa244.

Clapperton M., Diack A.B., Matika O., Glass E.J., Gladney C.D., *et al.* (2009) Genet Sel Evol 41(1): 54. https://doi.org/10.1186/1297-9686-41-54.

Flori L., Gao Y., Laloë D., Lemonnier G., Leplat JJ., *et al.* (2011) PLoS ONE 6(7): e22717. https://doi.org/10.1371/journal.pone.0022717.

Gilmour A.R., Gogel B.J., Cullis B.R., Welham S.J., and Thompson R. (2015). Hemel Hempstead: VSN International. Available at: http://www.vsni.co.uk/.

Hine B., Mallard B., Ingham A., and Colditz I. (2014). Immune competence in livestock. In: Hermesch S. and Dominik S. (eds) Breeding Focus 2014 – Improving Resilience. Animal Genetics and Breeding Unit, University of New England, Armidale, Australia, pp. 49 – 64.

Lim KS., Cheng J., Putz A., Dong Q., Bai X., *et al.* (2021) BMC Genomics. 22(1):1-18 https://doi.org/10.1186/s12864-021-07912-8.

Mulder H.A. and Rashidi H. (2017) J Anim Sci 95(8):3346-3358. https://doi.org/10.2527/jas.2017.1479.

Putz A.M., Harding J.C.S., Dyck M.K., Fortin F., Plastow G.S., *et al.* (2019) Front Genet 9:660. https://doi.org/10.3389/fgene.2018.00660.

162. Evidence of naturally occurring genetic variation in host response to challenge with a highly pathogenic PRRSV strain

J.R. Dunkelberger[1,2*], P.K. Mathur[1], E. Little[3], D. Hanson[3], S. Dee[3] and E.F. Knol[1]

[1]Topigs Norsvin, Topigs Norsvin Research Center, Schoenaker 6, 6641 SZ, Beuningen, the Netherlands; [2]Topigs Norsvin USA, 12750 Nicollet Ave. S. Suite 300, 55337, Burnsville, USA; [3]Pipestone Applied Research, 1300 S. Hwy 75, 56164, Pipestone, USA; jenelle.dunkelberger@topigsnorsvin.us

Abstract

The objective of this study was to estimate genetic parameters for traits recorded on finishing pigs reared under commercial conditions and experimentally inoculated with porcine reproductive and respiratory syndrome virus (PRRSV). Pigs (n=1,446) were vaccinated for PRRS and inoculated with PRRSV four weeks later. Individual body weights were recorded at 0, 13, and 42 days post-infection (dpi) and at marketing. Individual clinical scores were recorded at 0, 13, and 42 dpi and mortality was recorded from 0 dpi to market. Heritability estimates for average daily gain, clinical score at 13 dpi, clinical score at 42 dpi, and mortality were 0.26(0.05), 0.15(0.04), 0.09(0.04), and 0.10(0.07), respectively. In conclusion, data collected from this study can be used as a training data set for immediate implementation of genetic selection to breed pigs for improved, natural robustness to disease challenge in high disease pressure environments.

Introduction

Disease has a major impact on the cost, efficiency, and welfare of pig production. Biosecurity practices and vaccination may be used to control the incidence of disease outbreaks or severity of infection upon exposure. However, these approaches are not effective for all diseases or in all situations. Therefore, using genetic approaches to produce animals with enhanced robustness to disease is an attractive, additional disease control strategy.

One approach is to use genetic engineering to introduce novel genetic variation associated with enhanced robustness to disease. Research conducted by Whitworth and others (2015) showed that gene-editing technology can be used to introduce a novel mutation within the *CD163* gene, which confers complete resistance to PRRSV-infection. An alternative genetic solution is to use existing genetic variation within a population to breed pigs for enhanced robustness to disease challenge. Evidence of naturally occurring genetic variation in response to challenge has already been detected for purebred and crossbred sows following a PRRS outbreak (Herrero-Medrano *et al.*, 2015; Mathur *et al.*, 2014; Serão *et al.*, 2014) and commercial pigs when housed in a closed research facility and inoculated with PRRSV (Boddicker *et al.*, 2012; Dunkelberger *et al.*, 2017; Hess *et al.*, 2016). Results from these studies suggest that genomic selection, which capitalizes on existing genetic variation within a population, can be used to breed pigs for enhanced, natural robustness to disease as a viable, alternative PRRS control strategy. However, naturally occurring genetic variation in host response to experimental PRRSV-challenge for pigs reared under commercial conditions has not yet been evaluated. Therefore, the objective of this study was to estimate genetic parameters for host response to disease challenge for finishing pigs reared under commercial conditions and experimentally inoculated with PRRSV.

Materials & methods

The South Dakota State University (SDSU) and Pipestone Applied Research (PAR) institutional animal care and use committees (SDSU IACUC 18-030A and PAR IACUC 1-18) reviewed and approved the trial protocol, mortality standards, and caretaker handling certification.

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_162, © J.R. Dunkelberger *et al.* 2022

Animals. Commercial crossbred (Synthetic boar × [Landrace × Large White]) (n=1,446) pigs were used for this study. Ear tissue was collected from each pig at birth and used for genotyping with a custom single nucleotide polymorphism (SNP) panel. Pigs were transported to a commercial research barn at weaning and placed into pens balanced by sex, with 27 pigs housed per pen. Post-arrival, each pig was vaccinated for PRRS using a PRRS modified live virus vaccine (IngelVac ATP, Boehringer Ingelheim), according to the label instructions. Four weeks later, each pig was experimentally inoculated with $1\times10^{3.5}$ TCID50 of the 1-7-4 PRRSV isolate via intramuscular injection (SDSU, South Dakota, USA) and followed until marketing.

Traits. Various phenotypes were recorded throughout the study, including individual body weights at 0, 13, and 42 dpi, which were used to calculate average daily gain (ADG) as the slope of body weight regressed on day from 0 to 13 dpi, 13 to 42 dpi, 0 to 42 dpi, and 0 dpi to market. Clinical scores were recorded for each individual at 0, 13, and 42 dpi using a subjective, 6-point scoring system, based on the method described by Pantoja *et al.* (2013). Mortality was recorded from 0 dpi to slaughter.

Genotypes. Genotypes across 45,577 SNPs were available on 1,419 pigs to construct a genomic relationship matrix according to the method of VanRaden (2008).

Statistical analyses. All statistical analyses were conducted using ASReml 4.0 (Gilmour *et al.*, 2015). Variance components and heritability estimates were derived for each trait using a single record per individual. Average daily gain and clinical score were analyzed using a mixed linear model with sex (barrow or gilt), room (East or West), age at 0 dpi (covariate), and weight at 0 dpi (covariate) fitted as fixed effects. Animal, litter, and pen were fitted as random effects to account for genetic, maternal, and environmental effects, respectively. Mortality was analyzed as a binary variable (0/1) using both a LOGIT and linear model.

Results

Although pigs tested positive for swine influenza virus at weaning and for *Salmonella* and rotavirus post-placement, most pigs (69.05%) appeared physically healthy and without signs of disease (clinical score = '1') at PRRSV-inoculation (Figure 1). This percentage dropped to 47.77% at peak challenge (13 dpi) and returned to baseline by 42 dpi (70.08%) (Figure 1).

Overall mortality rate (from wean to market) was 13.5%, indicating severe disease, thus maximizing phenotypic variation and, therefore, expression of genetic variation, in host response to challenge. The high mortality rate observed for this trial was likely due to a combination of factors, including: flu-positive status at weaning, inoculation with a highly pathogenic PRRSV strain, co-infection with numerous, opportunistic pathogens post-challenge, and intentionally withholding antibiotic treatment from 0 to 21 dpi, in adherence of microbiome analyses to be conducted using fecal samples collected during this time period. Examples of opportunistic pathogens detected post-challenge include: *Haemophilus parasuis, Actinobacillus suis, Streptococcus suis, Salmonella, Pasteurella multocida,* and *Escherichia coli.*

Results of genetic parameter estimation show that breeding pigs for improved overall response to disease challenge is possible. Heritability of ADG post-challenge was estimated at 0.27(0.05), 0.16(0.05), and 0.24(0.05) when calculated from 0 to 13 dpi, 13 to 42 dpi, and 0 to 42 dpi, respectively (Table 1). Similar heritability estimates were obtained for ADG from birth to market, wean to market, and 0 dpi to market at 0.24(0.05), 0.25(0.05), and 0.26(0.05), respectively (Table 1). Clinical score was not heritable when evaluated prior to challenge (0.03[0.03]). Post-challenge, heritability was estimated at 0.15(0.04) for clinical score evaluated at 13 dpi, 0.09(0.04) for clinical score evaluated at 42 dpi, and 0.10(0.07) for mortality from 0 dpi to market (Table 1).

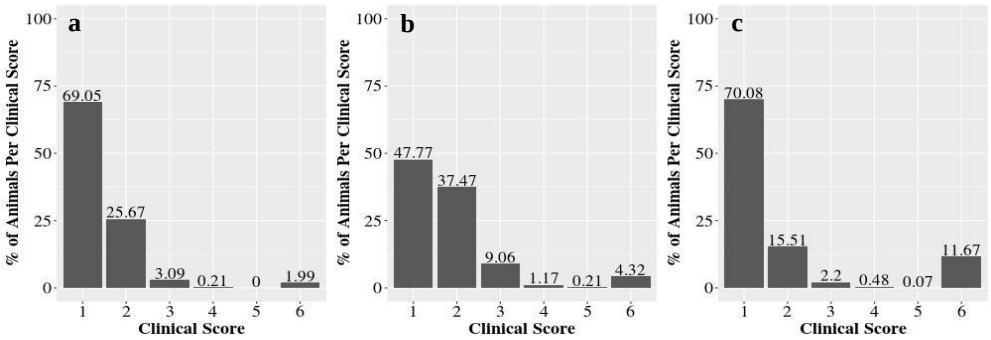


Figure 1. Percent of pigs per clinical score assigned at (a) 0, (b) 13 and (c) 42 days post-infection with porcine reproductive and respiratory syndrome virus. Individual clinical scores were assigned according to the following scale: '1' = healthy, no signs of disease; '2' = early signs of disease; '3' = moderate signs of disease; '4' = advanced clinical disease; '5' = candidate for euthanasia; and '6' = mortality.

Table 1. Variance components and heritability estimates (SE) for traits evaluated following experimental infection with porcine reproductive and respiratory syndrome virus.

Trait	Genetic variance	Litter variance	Pen variance	Residual variance	Heritability (SE)
Average daily gain					
0 to 13 dpi[1]	0.0136	0.0005	0.0076	0.0291	0.27 (0.05)
13 to 42 dpi	0.0054	0.0004	0.0021	0.0254	0.16 (0.05)
0 to 42 dpi	0.0056	0.0002	0.0009	0.0163	0.24 (0.05)
0 dpi to market	0.0022	0.0002	0.0006	0.0056	0.26 (0.05)
Wean to market	0.0014	0.0002	0.0004	0.0036	0.25 (0.05)
Birth to market	0.0009	0.0001	0.0002	0.0025	0.24 (0.05)
Clinical score					
0 dpi	0.0070	0.0105	0.0191	0.1694	0.03 (0.03)
13 dpi	0.1183	0.0000	0.0329	0.6564	0.15 (0.04)
42 dpi	0.1944	0.0743	0.0370	1.7776	0.09 (0.04)
Mortality (LOGIT)	0.3825	0.1670	0.0278	3.2899	0.10 (0.07)
Mortality (observed scale)	0.0117	0.0036	0.0012	0.0968	0.10 (0.04)

[1] Days post-infection.

Discussion

Genetic parameters for traits evaluated in this study provide evidence of substantial, naturally occurring genetic variation in host response to challenge with a highly pathogenic PRRSV strain. Further, genetic correlations between traits evaluated post-challenge (results not shown) were in the expected direction. Previous studies conducted to investigate the role of genetics on host response to PRRS challenge used data collected following an outbreak or following experimental infection with PRRSV in a closed research facility. Therefore, this is the first study to evaluate host response to experimental PRRSV-challenge for pigs reared under commercial conditions. Detection of numerous opportunistic pathogens throughout this study, a reflection of the immunosuppressive effect of PRRSV, is consistent with presentation of PRRS in the field. An experimental challenge model was used for this study to maximize the accuracy of genetic

parameters estimated for traits recorded under challenge. Variability in exposure, which is typical during an outbreak, leads to underestimation of heritability for traits recorded under challenge (Bishop, 2014). Experimental inoculation, which standardizes the level, timing, and route of pathogen exposure across individuals, can be used to avoid this issue.

In conclusion, the combination of genotypic and phenotypic data collected from this study can be used as a training data set for immediate implementation of genetic selection to breed pigs for improved, natural robustness to disease challenge in commercial, high disease pressure environments. Genomic selection is the ideal methodology to facilitate genetic improvement of traits controlled by multiple genes, which is typical of disease-related traits. Therefore, breeding pigs for enhanced, natural robustness to challenge should be considered a viable, alternative PRRS control strategy to introducing novel genetic variants for PRRS resistance using genetic engineering technologies.

References

Bishop S. C. (2014) Proceedings of the 10th World Congress on Genetics Applied to Livestock Production. Aug 17- 22, Vancouver, Canada

Boddicker N., Waide E. H., Rowland R. R. R., Lunney J. K., *et al.* (2012) J. Anim. Sci. 90:1733–1746. https://doi.org/10.2527/jas.2011-4464

Dunkelberger J. R., Serão N. V. L., Niederwerder M. C., Kerrigan M. A., *et al.* (2017). J. Anim. Sci. 95(2):584–598. https://doi.org/10.2527/jas2016.1071

Gilmour A. R., Gogel B. J., Cullis B. R., Welham S. J., *et al.* (2015). VSN International Ltd, Hemel Hempstead, UK. www.vsni.co.uk

Herrero-Medrano J. M., Mathur P. K., ten Napel J., Rashidi H., *et al.* (2015). J. Anim. Sci. 93:1494–1502. https://doi.org/10.2527/jas2014-8583

Hess, A. S., Islam Z., Hess M. K., Rowland R. R. R., *et al.* (2016). Genet. Sel. Evol. 48(43):1–20. https://doi.org/10.1186/s12711-016-0222-0

Mathur P. K., Herrero-Medrano J. M., Alexandri P., Knol E. F., *et al.* (2014) Proceedings of the 10th World Congress on Genetics Applied to Livestock Production. Aug 17- 22, Vancouver, Canada

Pantoja L. G., Kuhn M., Hoover T., Amodie D., *et al.* (2013). J. Swine Health. Prod. 21:188–194

Serão N. V. L., Matika O., Kemp R. A., Harding J. C. S., *et al.* (2014). J. Anim. Sci. 92:2905–2921. https://doi.org/10.2527/jas2014-7821

VanRaden P. M. (2008). J. Dairy Sci. 91(11):4414–4423. https://doi.org/10.3168/jds.2007-0980

Whitworth K. M., Rowland R. R. R., Ewen C. L., Trible B. R., *et al.* (2015). Nat. Biotechnol. 34(1):20–22. https://doi.org/10.1038/nbt.3434

163. A quantitative genetic theory for infectious diseases

P. Bijma*, A.D. Hulst and M.C.M. de Jong

Wageningen University & Research, P.O. Box 338, 6700 AH Wageningen, the Netherlands; piter.bijma@wur.nl

Abstract

We integrated quantitative genetics and epidemiology to develop a quantitative genetic theory of the prevalence of endemic infectious diseases. Results show that infectious diseases respond very differently to selection than common non-communicable traits, and strongly suggest that the genetic variance determining the potential response of prevalence to selection must be much larger than currently believed. Moreover, heritable variation and response to selection increase significantly when the prevalence of the infection decreases, ultimately leading to local extinction of the infection due to herd immunity. These results change our perspective on the prospects of genetic selection against infectious diseases.

Introduction

A reduction of the prevalence of infectious diseases in livestock is highly desirable. Individual disease status, however, measured as $y = 0$ or 1 indicating non-infected vs infected, typically shows low heritability (h^2). Moreover, the classical threshold model predicts that h^2 goes to zero when a disease becomes rare. Hence, genetic selection against infectious diseases seems difficult. This perspective, however, ignores that infectious diseases rely on transmission between individuals. Here we present a quantitative genetic theory for response to selection and heritable variation in the prevalence of endemic infectious diseases.

Methods & results

The endemic prevalence (P) of an infectious disease is the average fraction of the population that is infected, and is equal to the mean of individual disease status (P = avg(y)). Standard epidemiological theory shows that the endemic prevalence follows from the basic reproduction number (R_0; Weiss and Dishon 1971; Figure 1),

$$P = 1\text{-}1/R_0 \quad (1)$$

R_0 is the number of individuals that gets infected by a single typical infected individual in a naive population. For example, with $R_0 = 3$, an infected individual would on average infect three others. However, if 2/3 of the population is infected already, then only a single new infection will occur on average. Hence, prevalence reaches an equilibrium at $P = 1 - 1/3 = 2/3$. To reduce prevalence, therefore, breeders should lower R_0. At low R_0, prevalence is very sensitive to changes in R_0 (Figure 1), which has significant implications for response to selection.

Breeding values for R_0, prevalence and disease status. R_0 is the product of the contact rate between individuals, the *sus*ceptibility of individuals to become infected, the propensity of individuals to infect others (*inf*ectivity) and the duration of the infectious period (*t*). Because R_0 is a product and because it takes strictly positive values, we define an additive model with normally distributed genetic effects for the logarithm of R_0,

$$A_{\ln(R0)} = A_{\ln(sus)} + A_{\ln(inf)} + A_{\ln(t)} \quad (2)$$

We assume that additive genetic (co)variances are constant on this log scale, i.e. are independent of the level. (This implies that additive genetic variance in R_0 decreases as R_0 decreases). Connecting Equation 1

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_163, © P. Bijma *et al.* 2022

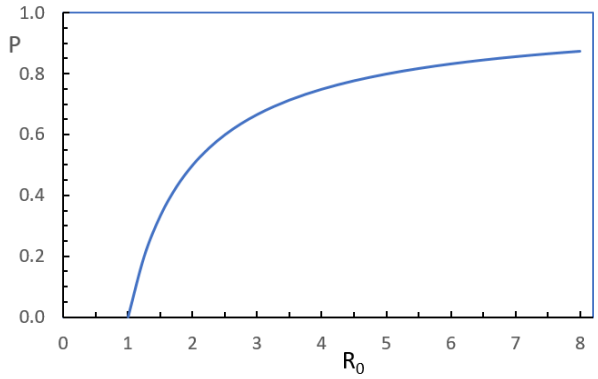


Figure 1. Prevalence as a function of R_0 (from Bijma *et al.* 2022).

and 2 leads to a breeding value for individual disease status (A_y) and a breeding value for prevalence (A_p; Bijma *et al.* 2022; here assuming no genetic variation in infectivity),

$$A_y = P(1\text{-}P)\, A_{\ln(R0)} \quad (3a)$$

$$A_P = (1\text{-}P)\, A_{\ln(R0)} \quad (3b)$$

A_y is the ordinary ('direct') breeding value for the disease status of the individual itself, while A_p is a total breeding value that measures the total effect of an individual's genes on the prevalence in the population. Hence, response of prevalence to selection equals the per generation change in mean A_P. Note that A_p is a factor 1/P greater than A_y. Moreover, for a constant $A_{\ln(R0)}$, A_y has a maximum for P=0.5, while A_p increases when P decreases. Thus, the common h^2 of individual disease status (y=0,1) has a maximum at P=0.5, similar to the threshold model (Robertson 1950), but the additive genetic variance that breeders can use for response to selection increases significantly as prevalence falls (Figure 2). This result is a consequence of the increasing slope of the relationship between P and R_0 at lower values of R_0 (Figure 1).

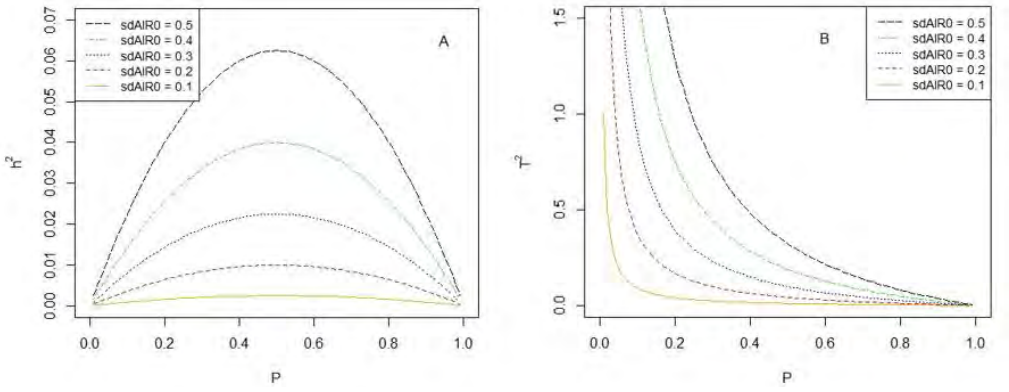


Figure 2. Heritability of individual disease status (h^2, panel A), and additive genetic variance in the endemic prevalence as a fraction of phenotypic variance (T^2, panel B), as a function of prevalence (P). For different additive genetic standard deviations in $\ln(R_0)$, and without genetic variation in infectivity (from Bijma *et al.* 2022).

Figure 2 shows that common heritabilities of disease status correspond to a large additive genetic variance in prevalence, particularly when prevalence is small. For example, for P=0.2 and sd($A_{ln(R0)}$)=0.4, the common observed-scale heritability of binary disease status is only 0.026, while the additive genetic variance that can be used for response to selection is 64% of phenotypic variance (T^2=0.64). Hence, for this scenario, the additive genetic standard deviation in prevalence is sqrt[0.64×0.2×(1-0.2)] = 0.32, which is very large.

Response to selection. Figure 2B shows a strong increase in additive genetic variance as prevalence falls, suggesting that response to selection for lower prevalence should accelerate over generations. Figure 3 shows response to selection observed in simulations following standard methods in epidemiology, not using any of the above theory. Figure 3 also shows the mean breeding values for prevalence (A_p) and for individual disease status (A_y), expressed as deviations from prevalence in the first generation. As expected based on Figure 2, the response accelerates as prevalence decreases. In the final generations (t=16, and t=70) the infection disappears due to herd immunity in the local population, not because individuals are fully resistant to infection. The mean breeding value for prevalence (Equation 3b) closely matches the observed response, while the mean breeding value for individual disease (Equation 3a) deviates more and more from the observations as prevalence decreases. Hence, response in Figure 3 is significantly greater than the common breeding values for individual disease status suggest.

Discussion

We developed a quantitative genetic theory of infectious diseases, focussing on R_0 and on the endemic prevalence of the infection. The breeding value for the logarithm of R_0 is central to the theory, and is the sum of the breeding values for the logarithms of susceptibility, infectivity and duration of the infectious period. From the breeding value for ln(R_0) we derived a breeding value for individual binary disease status

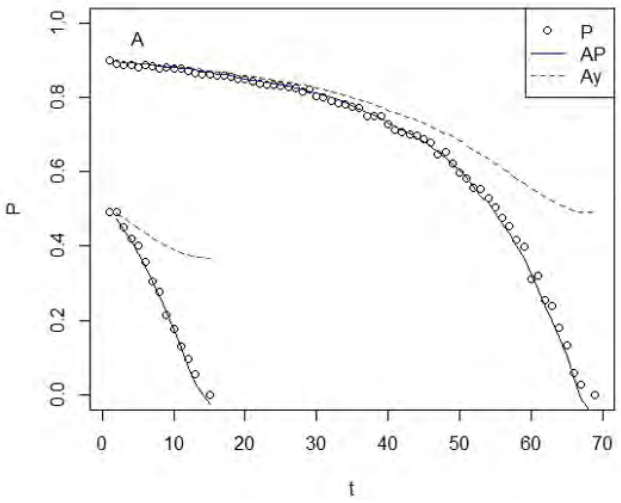


Figure 3. Response to mass selection against an infectious disease; observed response (small circles), mean breeding value for prevalence in (A_p, solid blue line) and mean breeding value for individual diseases status (A_y, dashed red line). For two populations, one starting at a prevalence of 50%, and the other at 90%. With genetic variation in susceptibility only, with var($A_{ln(R0)}$) = var($A_{ln(sus)}$) = 0.3^2, and a selected proportion of 0.5. The common observed-scale heritability for individual disease status in any generation can be read from Figure 2A, using an x-axis value corresponding to the prevalence in that generation, and does not exceed 0.022 (from Bijma *et al.* 2022).

and a breeding value for the prevalence of the infection in the population. The first relates to ordinary h^2 of individual disease status, while the second relates to response to selection. Without genetic variation in infectivity, the breeding value for prevalence is a factor $1/P$ greater than the breeding value for individual disease status. This is a general result, and does not rely on the assumption of an additive model on the log scale. For this reason, both the additive genetic variance for prevalence and the response of prevalence to selection increase when prevalence goes down (Figure 2B, 3).

The large difference between the h^2 of individual disease status and the total additive genetic variance in prevalence as a fraction of phenotypic variance (T^2) shown in Figure 2 indicates the presence of considerable indirect genetic effects (IGE). The h^2 reflects the variance of only the direct genetic effect of individuals on their own disease status. The T^2 reflects the variance of the full effect, i.e. of the sum of the effects on the disease status of the individual itself and on the disease status of its herd mates. Diseases that rely on transmission generate considerable IGE because an individual who does not become infected itself also cannot infect other individuals (simply because it is not infected). Hence, the direct effects for susceptibility and for duration of the infectious period are fully correlated to the corresponding indirect effects, resulting in positive feedback and an increase in the total additive genetic variance. This positive feedback effect increases strongly when prevalence decreases, which causes IGE to make up an increasing proportion of the total effect, and explains the difference between h^2 and T^2 in Figure 2 A vs B.

The results shown here apply to endemic microparasitic infections, and where the pathogen relies on the host individual (i.e. the livestock) for its replication. Hence, we assume absence of an external reservoir of pathogens, such as a second host species (for example badgers in the case of bovine Tb). Note that temporary survival of the pathogen in the environment does not violate this assumption. This assumption implies that a reduction in, e.g. susceptibility translates fully into reduced exposure in the local population (because of fewer infected individuals). Example traits include mastitis, infectious claw disorders, respiratory infections in young animals (young replacement stock, meat calves, fattening pigs), faecal-oral transmitted infections causing gastro-intestinal diseases, and several zoonosis.

In conclusion, our results show that the quantitative genetics of infectious diseases is very different from the quantitative genetic theory for ordinary, non-communicable, traits, and that prospects for genetic improvement are much better than currently believed, as was also proposed by Bishop and Stear (1997). These findings, coupled with the high importance of infectious diseases for animal health and welfare, and also for human health in the case of zoonoses, suggest that breeders should become familiar with basic concepts in epidemiology, such as R_0. Together with recent advances in sensor technology and AI, which will enable us to collect large-scale high-quality data on individual disease status, these findings facilitate artificial selection against infectious diseases.

References

Bijma, P., Hulst, A. D., and de Jong, M. C. M. (2022) Genetics iyab141. https://doi.org/10.1093/genetics/iyab141

Bishop, S. C., and Stear, M. J. (1997) Anim. Sci. 64(3):469-478.

Robertson, A. (1950) Genetics 35:234-36.

Weiss, G. H., and Dishon, M. (1971) Math. Biosci. 11(3-4):261-265.

164. Breeding for general disease resistance: a selection experiment in rabbits

M. Gunia[1*], J. Ruesche[1], P. Aymard[1], E. Gillet[2], C. Herbert[3], V. Helies[1], D. Savietto[1], R. Robert[4], L. Warin[5], H. Gilbert[1] and H. Garreau[1]

[1]GenPhySE, INRAE, Université de Toulouse, ENVT, 31326, Castanet Tolosan, France; [2]CLIPP, 7 rue du Fg. Poissonnière, 75009 Paris, France; [3]HYCOLE, Route de Villers-Plouich, 59159 Marcoing, France; [4]HYPHARM SAS, La Corbière, Roussay, 49450 Sèvremoine, France; [5]ITAVI, L'Orfrasière, 37380 Nouzilly, France; melanie.gunia@inrae.fr

Abstract

We performed a selection experiment to investigate the possibility to improve resistance to non-specific diseases in rabbits. The selection criterion was a binary trait (healthy versus sick or dead) obtained from clinical signs of diseases occurring naturally on farm. A herd of 116 to 132 does and 30 bucks was bred based on this selection criterion. The heritability of the trait was 0.035±0.009. After four generations of selection, the genetic progress was 1.5 genetic standard deviation, indicating that breeding for general disease resistance is feasible.

Introduction

Livestock populations are exposed to a variety of pathogens, not always identified, resulting in multifactorial diseases that are difficult to eradicate. For this reason, animal breeders and farmers are searching for robust strains of animals, able to cope with environmental stresses and to resist epizootics. However, a measure of individual responses to diseases resistance that is accurate, cheap, heritable and easy to perform on farm is difficult to obtain (Davies *et al.*, 2009; Bishop and Woolliams, 2014). In this study, we performed a selection experiment to improve the general resistance of animals to the various diseases they may encounter. The selection criterion was a binary trait (healthy versus sick or dead) based on clinical signs of diseases occurring naturally on farms. The aim of our project was to investigate if selection for non-specific disease is possible based on this criterion in rabbits.

Materials & methods

Trait. The selection criterion was a binary trait (0 = morbid, sick or dead from digestive, respiratory, or other diseases, mostly from infectious origin; 1 = healthy). Clinical signs were recorded at the INRAE experimental rabbit farm at a single time point toward the end of growth (63 days of age). The most likely cause of death was also recorded for rabbits that died between weaning and the end of the test. Clinical signs of diseases were recorded once in the young rabbit life, at 63 days of age, but not between weaning and the end of the test. Thus, rabbits categorized as healthy at 63 days of age might have been sick (exposed) individuals which recovered before recording (resistant or tolerant individuals). During the two years of the experiment, 10,169 rabbits were recorded individually.

Animals. A foundational herd of 132 females and 30 males was constituted from the INRA1777 rabbit line, a maternal line selected for litter size and for direct and maternal effects on weaning weight. Throughout the experiment, animals were not exposed to any sort of antibiotic treatments. Reproducers were vaccinated against myxomatosis and Rabbit Hemorrhagic Disease (RHD). The herd was conducted with overlapping generations. The number of reproducing does varied between 116 to 132, while the number of available bucks was kept to 30. The reproduction interval was of 42 days. Between 300 to 900 kits were born in each batch. They were weaned at 35 days old and reared in groups of 5 kits during the fattening period. Litters

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_164, © M. Gunia *et al.* 2022

were split after weaning to mix the genetic origins in each group. Weaned rabbits were fed *ad libitum* a commercial diet for growing rabbits (STABI-GREEN G, Terrya, Rignac, France).

The first 2 generations (10 batches) were reared in old facilities with standard cages (38×90×32 or 40×100×38). Then, the herd was moved to new facilities (renovated buildings, new housing systems: 46×90×70 or 46×90×60) from the eleventh batch.

Every 42 days, the best 30 females and 10 males were selected based on our selection criterion to become reproducers, with a constraint to limit inbreeding increase at 1% per generation, by using the Gencont software (Meuwissen, 2002). The selection was based on the genetic merit of the individuals, estimated from the information collected on all relatives tested in the pedigree to achieve the best possible accuracy.

Genetic analyses. Variance components were computed with a restricted maximum likelihood method, using a single-trait animal linear model and a threshold model with a logit link function. Breeding values were estimated with the BLUP method using the linear model. Variances components and genetic values were estimated together with the ASReml 3.0 software (Gilmour *et al.*, 2009). The models included a random additive polygenic effect and a random common litter effect. The significant fixed effects were the batch, parity of the dam, and gestation length. The estimated breeding values were used to estimate the genetic gain on the selection criterion. The linearity of the relationship between the means of the estimated breeding value per batch and the batch number was tested using a Fischer test from the Proc Reg of the SAS software.

Ethics. All experiments were conducted in accordance with the guidelines of the Directive 2010/63/EU of the European Parliament and of the Council and were approved by the ethical review board (no. 115, under the code APAFIS#1812-201812121652792 v4).

Results & discussion

In total, 14.9% of the rabbit showed clinical signs of diseases. The proportion of the main syndromes are shown in Table 1. Animals mostly suffered from digestive symptoms (in 68% percent of the cases), which could be related to Epizootic Rabbit Enteropathy (a gastrointestinal disorder of unknown aetiology) or to other digestive diseases caused by various pathogens.

The heritability of the selection criterion was low (Table 2), with a low genetic variance. This value was in the same range as previous studies in other rabbit lines (Gunia *et al.*, 2015, 2018), where the heritability estimates ranged from 0.03±0.003 to 0.10±0.02.

Table 1. Main disease syndromes recorded.

Syndromes	Number of rabbits	Proportion
Digestive symptoms	1,024	68%
Poor body condition	324	21%
Rabbit Haemorrhagic Disease (RHD)	139	9%
Other clinical signs	28	2%
Total	1,512	100%

Table 2. Genetic components of the non-specific disease status.

	Threshold model (underlying scale)		Linear model (observed scale)	
	Value	Standard-error	Value	Standard-error
Genetic variance	0.140	0.068	0.003	0.001
Heritability	0.027	0.013	0.032	0.009
Common litter effect	0.152	0.012	0.128	0.006

We observed a linear (P<0.001) increase of the estimated selection response (Figure 1), reaching 1.5 genetic standard deviation after 18 batches (approximately 2 years, a batch being 42 days). The first generation (batch 1 to 5) was a transition from the foundational herd (batch 0) conducted with separated generation to the herd selected on disease resistance with overlapping generation. It explains why the genetic level remained stable from batch 1 to 5, before increasing at batch 6.

The percentage of sick or dead rabbits was on average 14.9% per batch. The very low percentage of sick rabbit in batch 11 (0.1%) was related to the installation of the herd to the renovated facilities. In batch 13, an outbreak of RHD occurred. This disease is lethal and highly contagious. Sick rabbits were euthanised and all the other rabbit of this batch were culled to prevent contamination risk between batches and herds at the experimental facility. The culled rabbits were not considered in the experiment. The room and parcs were disinfected before introducing the next batch, and all fattening rabbits were vaccinated against RHD from batch 14.

Phenotypically, we did not see any progress or reduction in the number of sick rabbits (Figure 1). It is probably be due to the various pathogens that rabbits encountered in each batch. In addition, the type of pathogen and the pathogen loads may vary from batch to batch.

The first results of the selection experiment are promising. We did not perform a divergent section for ethical reasons. However, a comparison with a control herd is planned by the end of 2022. Embryos of

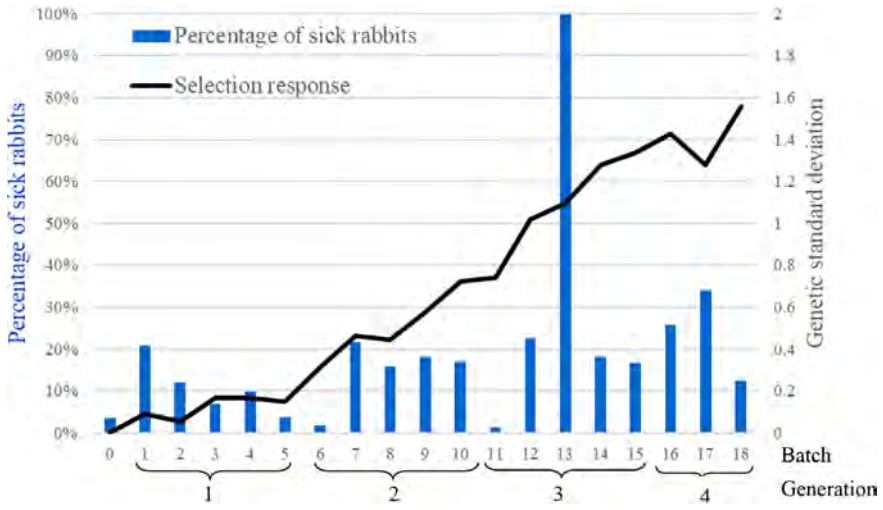


Figure 1. Proportion of sick rabbits (in percentage) per batch and predicted selection response for the proportion of healthy rabbit (in genetic standard deviation units).

the founder herd (before selection) were frozen. This final experiment will allow assessing the effect of selection by comparing the selected and control groups together in the same environment. Two contrasted environments (indoor in cages and outdoor on pasture) will be tested. Selected and control groups will be naturally exposed to the same pathogens within the same environment and thus will be able to express their various degree of disease resistance and/or tolerance.

Our results show that selection for health, without targeting any specific disease, is possible. We will assess if the results of selection in a specific environment are also validated in a different environment with different pathogens challenges.

Analyses on immune response traits and gut microbiota will now be performed to get a first insight into the mechanisms underlying this general disease resistance.

Funding

This work was funded by the Institut Carnot France Futur Elevage 2018 and by the French Ministry of Agriculture, Ecoantibio 2 Plan 2018.

References

Bishop,S.C., and Woolliams. J. A. (2014). Genomics and disease resistance studies in livestock. Livest. Sci. 166:190–198. https://doi.org/10.1016/j.livsci.2014.04.034

Davies G., Genini. S.,Bishop S.C., and Giuffra E. (2009). An assessment of opportunities to dissect host genetic variation in resistance to infectious diseases in livestock. Animal. 3:415–36. https://doi.org/10.1017/S1751731108003522.

Gilmour A.R., Gogel B.J., Cullis B.R. and Thompson R. (2009). ASREML user guide release 3.0. VSN International, Hemel Hempstead. http://www.ncbi.nlm.nih.gov/pubmed/22444313

Gunia M., David I., Hurtaud J., Maupin M., Gilbert H., *et al.*, 2015. Resistance to infectious diseases is a heritable trait in rabbits. J. of Anim Sci. 93: 5631-5638. https://doi.org/10.2527/jas.2015-9377

Gunia M, David I, Hurtaud J, Maupin M, Gilbert H, Garreau H. (2018) Genetic parameters for resistance to non-specific diseases and production traits measured in challenging and selection environments; Application to a rabbit case. Front Genet;9:467. https://doi.org/10.3389/fgene.2018.00467

Meuwissen, T. (2002). Proc. of the 7th WCGALP, Montpellier, France.

SAS Institute Inc. (2013). Base SAS® 9.4 Utilities: Reference. Cary, NC, USA.

165. A single step genetic evaluation including causal candidate SNPs for resistance to paratuberculosis in Holstein cattle

M.P. Sanchez[1*], T. Tribout[1], S. Fritz[1,2], R. Guatteo[3], C. Fourichon[3], L. Schibler[2], A. Delafosse[4] and D. Boichard[1]

[1]Université Paris Saclay, INRAE, AgroParisTech, GABI, 78350 Jouy-en-Josas, France; [2]Allice, 75595 Paris, France; [3]INRAE, Oniris, BIOEPAR, 44300 Nantes, France; [4]GDS Orne, 61000 Alençon, France; marie-pierre.sanchez@inrae.fr

Abstract

Bovine paratuberculosis, or Johne's disease (JD), is a contagious, incurable and very difficult to control disease due to *Mycobacterium avium* subsp. *paratuberculosis* (MAP). Using a single step GBLUP, including candidate SNPs for resistance to JD, applied to a pedigree of 161,253 Holstein animals (56,766 with serological phenotypes and 12,431 with genotypes), we estimated a moderate heritability value (0.14) of this trait and detected genomic regions associated to resistance to JD on chromosomes 23, 21, 12, 3, 7, 27, 5, 10, and 20. We also estimated a slightly favorable genetic trend of resistance to JD over the last two decades and relatively reliable genomic predictions (CD=0.55) in a validation population allowing the identification of cows at high risk of infection. Genomic predictions should therefore rapidly become an effective tool for controlling paratuberculosis in French Holstein cow farms.

Introduction

Bovine paratuberculosis, also referred as Johne's disease (JD), is a contagious and incurable disease, caused by *Mycobacterium avium* subsp. *paratuberculosis* (MAP). After being contaminated *in utero* or more frequently via the intake of MAP-infected feces as a young calf, animals undergo a long latency phase that can last several years. Subclinical symptoms of JD are weight loss and reduced milk production together with an inconsistent humoral immune response and fecal shedding. Subclinical cases can therefore contaminate their environment. Clinical cases finally develop chronic diarrhea and severe emaciation, ending in death. JD, which has adverse effects on animal welfare and serious economic consequences, has no effective treatment and access to vaccines is restricted, it is therefore very difficult to control. A number of studies have shown the role of host genetics in resistance to MAP (Brito *et al.*, 2018) but their results are not fully consistent mainly due to the various definitions of the MAP-resistant phenotype.

In genome-wide association studies (GWAS) conducted on imputed whole genome sequences of 1,644 Holstein cows with an accurately defined status for MAP, we previously identified various candidate SNPs explaining around 20% of the phenotypic variance of resistance to MAP infection (Sanchez *et al.*, 2020; 2021). Using these data together with additional data collected in infected herds and a ssGBLUP model including the best candidate SNPs, we investigate here the genetic determinism of resistance to JD and estimate genetic trends, reliability and risk factors associated to genomic predictions.

Materials & methods

Phenotypes and genotypes. Data of 247,375 Holstein cows were collected in northwestern French herds enrolled in JD control plans. These data included MAP statuses of 4,100 cows obtained from serum ELISA and PCR on faeces in the PARADIGM project (see details in Sanchez *et al.*, 2020) and MAP statuses of 243,275 cows deduced from serological data (Idexx or Idvet ELISA test, Montpellier, France) routinely recorded since 2015. These cows were non-infected (NI; n=228,337) or infected (I; n=19,038) by MAP. We applied different filters to these data. First, to exclude cows still potentially in the latency period, I and NI cows younger than 24 and 36 months old, respectively, were discarded. Then, to maximize the probability

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_165, © M.P. Sanchez *et al.* 2022

of exposure to MAP, we kept only cows in herds with at least one I cow and one NI cow born in the same year. These filters resulted in a drastically reduced dataset, which ultimately contained 56,766 Holstein cows (42,829 NI and 13,937 I) from 3,114 herds. The pedigree was traced over four generations and contained 161,253 animals, including 12,431 genotyped individuals, among which 4,031 had phenotypes (2,787 NI and 1,244 I). The 12,431 animals were genotyped with different SNP Beadchip versions, including the EuroGMD Beadchip that contains an add-on for selected SNPs predictive for resistance to JD. In total, 53,469 SNPs were imputed using FImpute (Sargolzaei *et al.*, 2014).

Model and methods. To estimate ssEBVs, resistance to JD was modelled as follows:

$$y = X\beta + Za + e \quad (1)$$

where y is the vector of phenotypes, i.e. 1 for NI cows and 0 for I cows, $a \sim N(0, H\sigma_a^2)$ is the vector of random additive genetic effects and $e \sim N(0, I\sigma_e^2)$ is the vector of random residual effects. The β vector included the fixed effects of herd × birth year, birth month, and ELISA test to account for variability of exposure to MAP in the environment between farms and in time, and for differences in test sensitivities and specificities. X and Z are incidence matrices, H is the relationship matrix among individuals and I the identity matrix, σ_a^2 and σ_e^2 are the additive genetic and residual variances, respectively. Model (1) was applied to the Hybrid Single Step method, proposed by Fernando *et al.* (2016) and implemented in the HSSGBLUP software (Tribout *et al.*, 2020). This method directly provides SNP effect estimates and ssEBVs of all animals in the pedigree, genotyped or not. ssEBVs were then expressed in genetic standard deviation units by dividing predictions on the raw scale by 0.15.

Variance components were estimated with WOMBAT software (Meyer, 2007) using a similar model but only with pedigree information to build the relationship matrix A. Heritability h^2 for resistance to JD was calculated as follows: $h^2=\sigma_a^2/(\sigma_a^2+\sigma_e^2)$.

Genetic trends of resistance to JD in females and males were assessed separately by the annual average ssEBVs of cows with phenotypes born between 2008 and 2018 and of bulls born between 2000 and 2014 with at least 10 daughters with MAP statuses.

Reliability of ssEBVs was estimated in a validation (VAL) population of 907 cows with phenotypes (715 NI and 192 I) and genotypes born between 2015 and 2018. For this study, model (1) was run on a truncated dataset constituted by phenotypes of the 41,774 cows (31,115 NI and 10,659 I) born before 2015, among which 3,118 had genotypes (2,069 NI and 1,049 I). Reliability was estimated by $CD=r^2/h^2$, with r the correlation between phenotypes adjusted for non-genetic effects estimated with model (1) on the complete dataset and genomic predictions (GP) of VAL animals predicted using the truncated dataset. Risk factor was also assessed in the VAL set. GP were grouped in 0.5 point classes and the relative risk factor in class i was estimated by the ratio of the proportion of infected animals in class i to that in the central 0 class.

Finally, considering SNP effects (ŝ) estimated in HSSGBLUP, the % of genetic variance explained by the i^{th} genomic region of 25 adjacent SNPs was calculated as $var(\hat{a}_i)/\sigma_a^2$ where $\hat{a}_i = M_i\hat{s}_i$ is the vector of genomic values for the i^{th} region, M_i the matrix of SNP content in region i, and $\hat{s}_i$ the effects of the 25 SNP of the region.

Results

Estimates of genetic and residual variances were 0.019 and 0.113, respectively, corresponding to a moderate h^2 estimate of 0.143±0.01 for resistance to JD.

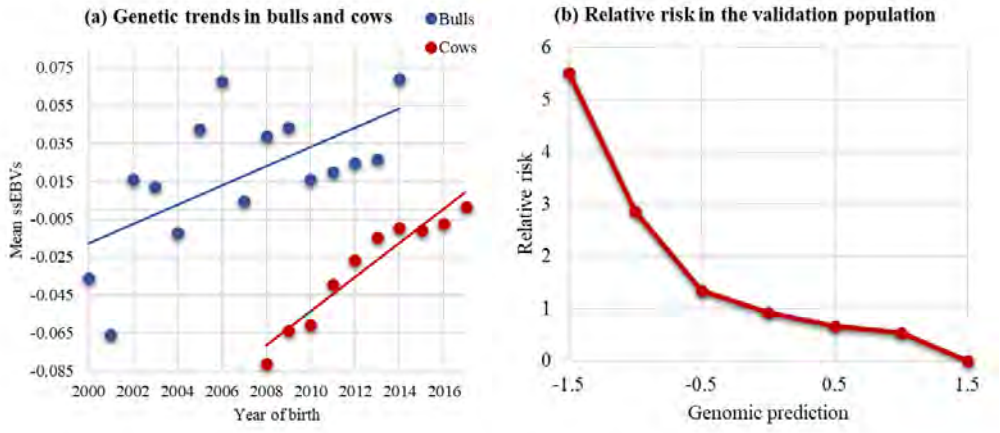


Figure 1. Evaluation results: (a) genetic trend of resistance to JD and (b) relative risk of being infected by MAP.

Estimated genetic trends of resistance to JD in bull and cow populations (31 to 117 bulls and 2,356 to 8,674 cows depending on the year of birth) were favourable (+0.6 σ_a and +0.8 σ_a, respectively over the period considered), the younger animals being genetically more resistant than the older ones (Figure 1a). Correlation between (GP) and phenotypes adjusted for fixed effects in VAL was equal to 0.28 (SE=0.03); the corresponding reliability was therefore ~0.55.

To illustrate the effect of genetics on the phenotype, we used the VAL population to estimate the relative risk for cows to be infected based on their GP (Figure 1b). A strong effect of genetics was observed in this validation dataset. Assuming the relative risk equal to 1 for GP=0, this risk was half as big and three times higher for cows with GP=1 and -1, respectively. Although numbers are limited, effects were even much stronger for cows with GP equal to ±1.5.

The estimation of percentage of genetic variance explained by overlapping genomic regions of 25 SNPs revealed a high peak on chromosome 23 (max. 0.32%) and lower peaks explaining between 0.10 to 0.15% of the total genetic variance, on chromosomes 21, 12, 3, 7, 27, 5, 10, and 20 (Figure 2).

Discussion

This study, based on a single step SNP GBLUP including candidate SNPs for resistance to JD, provides new insights into the genetic determinism of this trait and gives very encouraging results for the implementation of a genomic evaluation in Holstein cows using MAP statuses deduced from serological data routinely recorded.

With respect to the genetic determinism of resistance to JD, we herein obtain results quite different from those found in our previous study conducted on imputed whole-genome sequences of 1,644 Holstein cows with accurate extreme phenotypes (Sanchez *et al.*, 2020; 2021). In the present study, we first estimated a moderate h^2 value (0.14) much lower than the one we reported before (0.57) but which agrees with h^2 values reported in various studies (0.03-0.27; Brito *et al.*, 2018). The multi-SNP GWAS approach taking into account a total of 161,253 animals in the pedigree – 56,766 with phenotypes and 12,431 with genotypes – confirms effects of genomic regions located on chromosomes 23, 12, 3, and 5 and revealed new regions affecting resistance to JD on chromosomes 21, 7, 27, 10, and 20. However, the major effect detected on chromosome 13 in our previous single-SNP GWAS is not confirmed here.

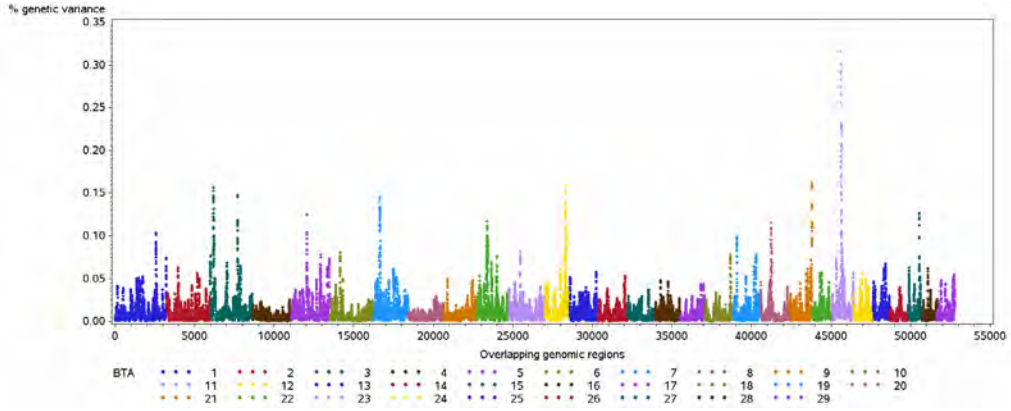


Figure 2. Percentage of genetic variance explained by overlapping regions of 25 adjacent SNPs.

Trends of genomic predictions, estimated over the last two decades, show a favourable genetic evolution of resistance to JD. In a validation population, we highlight relatively reliable genomic predictions for resistance to JD (CD=0.55) estimated from a reduced reference population and we show that genomic predictions can already be of great help in identifying cows at high risk of infection. These breeding values can be used at two levels: (1) in infected herds, this information will allow orientating early culling of highly susceptible animals and planning matings with resistant bulls; (2) at the breeding scheme level, these breeding values can be integrated into the breeding goal and, more importantly, used to produce resistant artificial insemination bulls for matings in infected herds. We can anticipate that as the number of genotyped and phenotyped cows increases continuously, genomic predictions will rapidly become an effective tool for controlling paratuberculosis in French Holstein farms. The same strategy is currently being developed in the Normande breed and could be extended to other populations in the future.

References

Brito L., Mallikarjunappa S., Sargolzaei M., Koeck A., Chesnais J., *et al.* (2018) J. Dairy Sci. 101(11):10062-10075. https://doi.org/10.3168/jds.2017-14250

Fernando R., Cheng H., Golden B., and Garrick D. (2016) Genet. Sel. Evol. 48:96. https://doi.org/10.1186/s12711-016-0273-2

Meyer K. (2007) J. Zhejiang Univ. Sci. B 8:815-821. https://doi.org/10.1631/jzus.2007.B0815

Sanchez M.P., Guatteo R., Davergne A., Saout J., Grohs C., *et al.* (2020) Genet. Sel. Evol. 52:14. https://doi.org/0.1186/s12711-020-00535-9

Sanchez M. P., Guatteo R., Delafosse A., Davergne A., Grohs C., *et al.* (2021) Proc. of the 72nd EAAP, Davos, Switzerland.

Sargolzaei M., Chesnais J., and Schenkel F. (2014) BMC Genomics 15(1):478. https://doi.org/10.1186/1471-2164-15-478

Tribout T., Ducrocq V., and Boichard D. (2020) Proc. of the 6th ICQG, Brisbane, Australia.

166. Genetic parameters and QTL mapping for antibody response to paratuberculosis in a natural infected flock of sheep

M.G. Usai[1*], S. Casu[1], T. Sechi[1], M.G. Cancedda[2], D. Pintus[2], S. Salaris[1], G. Mulas[1] and A. Carta[1]

[1]*AGRIS SARDEGNA, Research Unit Genetics and Biotecnology, Località Bonassai, 07100 Sassari, Italy;* [2]*Istituto Zooprofilattico Sperimentale della Sardegna G. Pegreffi, Via Duca degli Abruzzi 8, 07100 Sassari, Italy;* gmusai@agrisricerca.it

Abstract

The aims of this study were to estimate genetic parameters and to identify genomic regions associated to the host's immune response against *Mycobacterium avium* subsp. *paratuberculosis* (MAP). Three thousand eighty eight ewes, genotyped with the OvineSNP50 Beadchip, were subjected to repeated serum ELISA tests over time in an experimental flock. Two different variables were derived from the original ELISA results. A genomic relationship matrix was used to estimate variance components with variable-specific animal models. QTL mappings were performed by a combined linkage and linkage disequilibrium method using a principal component analysis to synthesize identity-by-descent matrix information. Results were similar for both the variables. Heritability of MAP status was approximately 0.2 and a significant QTL was detected on OAR 20 at about 24.8 Mb for both the variables. The QTL region overlaps some MHC II genes and will be further investigated by using whole genome sequence data.

Introduction

Paratuberculosis (PTB), is a contagious disease caused by infection with the Gram-positive bacteria *Mycobacterium avium* subsp. *paratuberculosis* (MAP). It affects ruminant species and can manifest as chronic granulomatous enteritis. PTB infection, results in significant negative effects both on the animal welfare and on farm profitability. Moreover, several studies suggested the implication of MAP in the Crohn's disease (Sechi *et al.* 2005). There are no effective treatments against MAP infection and vaccination does not guarantee the elimination of MAP shedding. Control strategies based on culling seropositive animals are very expensive and have a limited effectiveness. The role of host genetics in the susceptibility to MAP infection have been largely studied in dairy cattle, where low-to-moderate heritabilities were estimated and numerous Quantitative Trait Loci (QTL) were mapped. In sheep, the few studies on the role of genetics on resistance to PTB, suggest a moderate heritability with differences between breeds (Hick*ey et al.*, 2003; Sechi *et al.*, 2011). Moreover, studies on association between genetic polymorphisms and PTB susceptibility are very rare and usually based on experiments of limited size. In this work the genetic resistance to PTB in sheep was studied on the basis of data collected on a large sheep experimental flock. Ewes were genotyped with the OvineSNP50 Beadchip and repeated enzyme-linked immunosorbent assay (ELISA) tests have been performed over time. The main goals were to estimate genetic parameters of proxy traits for PTB resistance and to identify genomic regions associated to the host's immune response against MAP. Selected regions will be further investigated through association and functional annotation analysis of polymorphisms identified by whole genome sequence data (Casu *et al.* this congress).

Materials & methods

Experimental population. The experimental flock was set in 1999 with the first generation of ewes consisting in 928 back-cross Lacaune × Sarda. The following generations were obtained by mating adult ewes with rams from the Sarda Herd Book. The average size of the flock has been about 800 ewes with a replacement rate of 25%. The flock is managed by the farming system commonly adopted by commercial farms and is raised in an experimental farm in the south of Sardinia. Animals of the nucleus flock were

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_166, © M.G. Usai *et al.* 2022

naturally exposed to the MAP infection and no control practice, treatment or vaccination against PTB have been applied.

Phenotypes. From 2001 to 2012, 3,088 ewes from the experimental population, born from 1999 to 2011 were blood sampled with lags between consecutive samplings ranging from 5 to 12 months. On the whole, 14,482 blood samples were collected: on average 4.7 per head ranging from 1 to 8. Blood samples were tested using ELISA screening test (IDvet, France or Pourquier ELISA, France). Test results were expressed on the base of the percentage of the sample-to-positive ratio (SP) calculated as 100 × [OD of the sample – OD of the negative control] / [OD of the positive control – the OD of the negative control], where OD is the optical density. A test was definite positive when SP was ≥70%.

Genotypes. All the 3,088 phenotyped ewes, their 100 sires and 10 Sarda grandsires were genotyped with the Illumina Inc. OvineSNP50 Beadchip. SNP editing was performed using call rate and MAF thresholds of 95 and 1%, respectively. The ovine genome assembly v4.0 was used to construct the genetic map by assuming 1 Mb=1 cM. Unmapped SNPs and SNPs on sex chromosomes were not included. Finally, 43,390 SNPs were retained.

Genetic analyses. The definition of the susceptible/resistant status with the antibody levels assessed by the ELISA test is questionable, due to the lack of sensitivity (Mathevon *et al.*, 2019). Thus, to check the overall consistency of results, we performed the genetic analyses with two different variables. The first one assigns the susceptible status to animals showing at least one positive test along the recording period (CAR). However, CAR roughly approximates the status for those animals that showed at least one negative test after a positive one (41%) or those with just one test in the recording period (5%). Thus, the quantitative variable SP was considered to avoid the use of thresholds to define the PTB status. Heritability estimate for the binary variable CAR (coded as 1 for susceptible and 0 for resistant) was realized through an animal model with no fixed effect other than the mean and the random animal genetic effect. The quantitative variable SP was analyzed with a single trait repeatability animal model including the fixed effect of the age of the animal at the date of sampling as covariate and the ELISA plate, the animal non genetic effect and the animal genetic effect as random. In both the analyses, genetic relationships were taken into account by calculating the genomic relationship matrix (Van Raden, 2008) among genotyped animals. Variance components were estimated by a REML procedure through airemlf90 software (Misztal *et al.*, 2002). For the QTL detection on CAR, the original variable was used. For the QTL detection on SP, pseudo-phenotypes were obtained by averaging individual animal residuals and summing-up the genetic and non genetic animal effects of the previous model. QTL mappings were performed by the method proposed by Usai *et al.* (2019) which uses principal components to summarize Identity-by-Descent probabilities (IBD) between gametes of genotyped individuals. IBD were estimated at each position *l* of the OvineSNP50 SNP (IBD_l) by combining linkage disequilibrium and linkage analyses. In addition, the QTL detection model included a genome-wide IBD matrix calculated by averaging all the IBD_l to account for the polygenic effects. In this study, only QTL segregating within gametes of Sarda origin were tested by F-test at each SNP position. The 5% genome-wide significance threshold was determined by a Bonferroni correction for the 43,390 tests, which resulted in $-\log_{10}$ (*P*-value) = 6. Moreover, the ratio between the ANOVA estimators of the QTL variance and the total phenotypic variance was calculated (Casu *et al.*, 2022).

Results

Sixteen % of the ELISA tests were positive and 29.7% of the 3,088 ewes investigated in this study had at least one positive test during the observation period. Heritabilities estimated with the specific animal models were 0.20 for both CAR and SP with standard errors of 0.029 and 0.024 respectively. Repeatability estimate for SP was 0.68. Significance profiles obtained for CAR (Figure 1) and SP (Figure 2) were very similar. SNPs

exceeding the 0.05 GW significance threshold were 116 and 156 for SP and CAR respectively and 112 were significant for both the variables. All the significant positions were located on OAR 20 within a region from 23 to 35 Mb. For both the variables, the phenotypes predicted by the QTL effects at each significant SNP positions were strongly correlated (r>0.45) to the phenotypes predicted by the QTL effects at the position of the peak. The most significant position for the two variables corresponds to two close SNPs located at around 24.8 Mb (Table 1). ANOVA estimators of QTL variances at the peak positions were 4.1 and 4.7% of the phenotypic variance for SP and CAR respectively. Moreover, further positions close to WGS significance threshold were detected: for CAR on OAR18 at ~30 Mb; for SP on OAR1 at ~200 Mb and on OAR15 at ~62 Mb for both the variables.

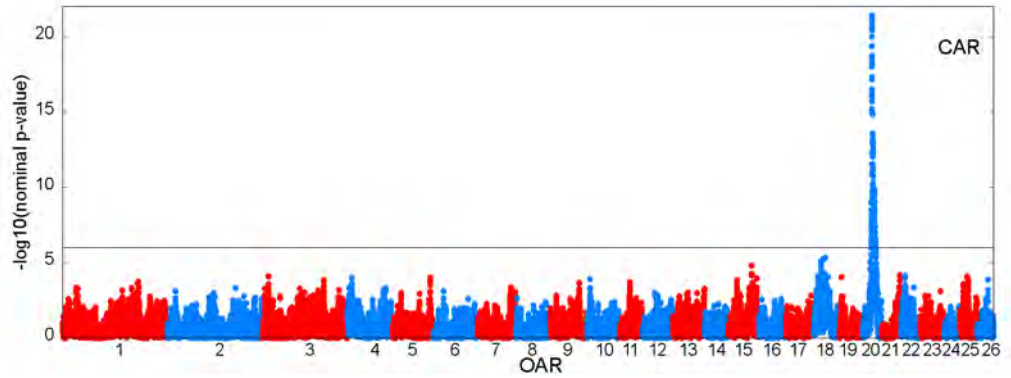


Figure 1. Manhattan plots corresponding to the QTL mapping carried out on CAR.

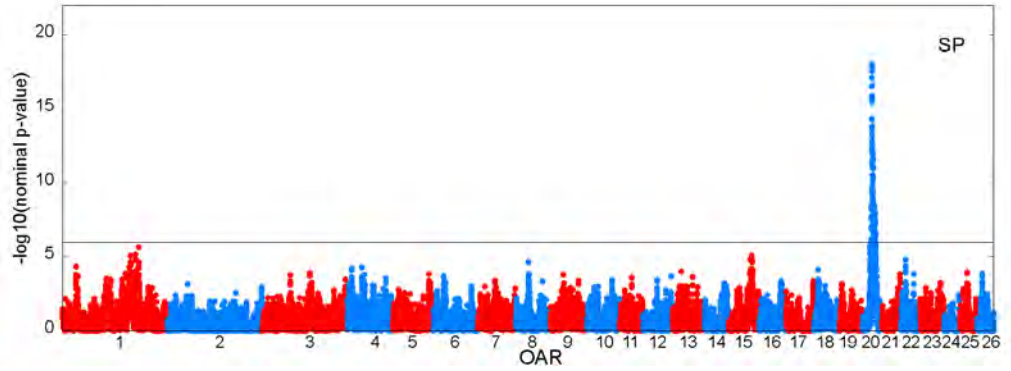


Figure 2. Manhattan plots corresponding to the QTL mapping carried out on SP.

Table 1. Details of the QTL regions that include SNP positions exceeding the 0.05 genome-wide significance threshold.

Variable	OAR	Significant SNP (n)	Range (Mb)	Highest peak		
				SNP name	Position (bp)	– log10 (*P*-value)
CAR	20	156	23.12-35.13	rs413015672	24860179	21.396
SP	20	116	23.49-35.13	rs419287784	24791184	18.025

Discussion

In our study, the two different variables led to very similar results in terms of heritability estimates and QTL detection profiles. CAR slightly outperformed SP in the significance level of the QTL on OAR 20 and further few differences were detected for some positions close to significance. Overall, heritabilities of serological response to PTB estimated by a genomic relationship matrix were moderate and consistent with previous studies, although they may be biased upwards by vertical transmission (Hickey *et al.*, 2003; Sechi *et al.*, 2011). This result suggests that selective breeding based on large scale ELISA tests may be an option to enhance resistance of sheep to MAP infection. The QTL on OAR20 identified in this study maps close to a region affecting the resistance to gastro-intestinal nematodes identified in the same experimental population (Casu *et al.*, 2022). The phenotypic variance explained by the QTL is equivalent to about 20% of the heritability of the two variables. This result seems promising for the potential application of a marker-assisted selection approach. The QTL region on OAR20, and especially its most significant segment (about 23-25 Mb), overlaps locations of some genes belonging to MHC II that are already known as involved on the genetic resistance to PTB (Kravitz *et al.*, 2021). With the aim of searching for potential causative mutations, this region was deeply investigated by using whole genome sequence data (Casu *et al.* this congress).

References

Casu S., Usai M.G., Sechi T., Salaris S.L., Miari S., *et al.* (2022) Genet Sel Evol 54, 2. https://doi.org/10.1186/s12711-021-00690-7

Hickey S.M., Morris C.A., Dobble J.L., and Lake D.E. (2003) Proc. of the NZSAP, Queenstown, New Zealand.

Kravitz A, Pelzer K. and Sriranganathan N. (2021) Front. Vet. Sci. 8:721706. https://doi.org/10.3389/fvets.2021.721706

Mathevon Y, Foucras G, Corbière F (2019) PLoS ONE 14(12). https://doi.org/10.1371/journal.pone.0226246

Misztal, I., Tsuruta, S., Strabel, T., Auvray, B., Druet, T., *et al.* (2002) Proc. of the 7th WCGALP, Montpellier, France.

Sechi L.A., Scanu A.M., Molicotti P., Cannas S., Mura M., *et al.* (2005) Am J Gastroenterol 100:1529–1536. https://doi.org/10.1111/j.1572-0241.2005.41415.x

Sechi S, Salaris S., Ligios C., Ponti N., Foucras., *et al.* (2011) Proc. of the 19th ASPA Congress, Cremona, Italy.

Usai M.G., Casu S., Sechi T., Salaris S.L., Miari S., *et al.* (2019) Genet Sel Evol 51(1):1-19. https://doi.org/10.1186/s12711-019-0508-0

VanRaden, P. M. (2008) J dairy sci 91(11):4414-4423. https://doi.org/10.3168/jds.2007-0980

167. The genetic basis of sex differences in immune responses in wild Soay sheep (*Ovis aries*)

E.A. Wimborne[1], A.M. Sparks[1,2], K. Watt[1], R. Sinclair[1], J.G. Pilkington[1], J.M. Pemberton[1], T.N. McNeilly[3], D.H. Nussey[1] and S.E. Johnston[1*]

[1]Institutes of Evolutionary Biology and Immunology and Infection Research, University of Edinburgh, Edinburgh, United Kingdom; [2]School of Biosciences, University of Sheffield, Sheffield, United Kingdom; [3]Moredun Research Institute, Pentlands Science Park, Midlothian, United Kingdom; susan.johnston@ed.ac.uk

Abstract

Sex differences in immune responses are widely observed in nature. Females have stronger innate and adaptive immune responses with higher rates of autoimmunity, whereas males have reduced resource allocation to immunity. If immune traits share a genetic basis between the sexes, sexual conflict may prevent each sex from reaching their optimal phenotype, maintaining variation at immune loci. We investigated sex differences in genetic basis of immune traits in wild Soay sheep (*Ovis aries*) using sex-stratified genome-wide analyses of IgA, IgE and IgG antibody levels against the nematode *Teladorsagia circumcincta*. We found a strong genetic constraint between the sexes (r_A=0.87-0.99), indicating a strong potential for sexual conflict. We identified increased maternal effects in female lambs and sex-specific variants in the male lamb genome near the candidate genes *SHARPIN*, *LRRC14* and the IL-9 cytokine gene, indicating moderate effect alleles contributing to sex differences in immune trait variation in Soay sheep.

Introduction

Sex differences in immune responses are widely observed in nature and are likely to be source of evolutionary trade-offs (van der Most *et al.* 2011; Metcalf *et al.* 2020). Females typically have stronger innate and adaptive immune responses, with higher rates of autoimmunity due in part to the immune-enhancing effects of oestrogen (Klein and Flanagan 2016). This may enhance female reproductive success through longevity, whilst investing heavily in each reproductive effort (Zuk 2009). Conversely, males tend to have reduced resource allocation to immunity, instead improving their fecundity through mating frequency, with testosterone having a suppressive effect on immune responses (Kanda *et al.* 1996). If immune responses are heritable and share a genetic basis between the sexes, then these differences can create an opportunity for sexual conflict to prevent each sex from reaching their optimal immune phenotype (Rowe *et al.* 2018). Indeed, despite infection imposing strong selection pressures, substantial genetic variation underpinning immune responses often persists in human and livestock studies (Liston *et al.* 2016; Denholm *et al.* 2017). Yet, few studies have specifically investigated sexual conflict in immunity in vertebrate populations. Soay sheep are a primitive breed of domestic sheep that have lived wild on the St Kilda archipelago in Scotland since the Bronze age and have been intensively studied since 1985. They are infected with several gastrointestinal strongyle nematodes common to domestic sheep, predominantly *Teladorsagia circumcincta*, *Trichostrongylus axei* and *Trichostrongylus vitrinus* (Craig *et al.* 2006). A recent genome-wide association study found large-effect loci associated with anti-*Teladorsagia circumcincta* IgA, IgE and IgG at the MHC, *CIITA*/*CLEC16A* and the IGH complex loci (Sparks *et al.* 2019). Here, we investigate the potential for sexual conflict by characterising genetic correlations between the sexes and conducting sex-stratified variance component and genome wide association analyses of Ig levels in this population.

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_167, © E.A. Wimborne *et al.* 2022

Materials & methods

Study population and data sets. The Soay study population is in the Village Bay of the island of Hirta, St Kilda, Scotland (57°49'N, 8°34'W). There is a comprehensive genomic pedigree and 7,691 individuals have been genotyped at 417,373 polymorphic SNP loci (see Stoffel *et al.* 2021 for full details). Antibody levels were previously assayed and quantified by Sparks *et al.* (2019); IgA, IgE and IgG levels were measured against *Teladorsagia circumcincta* third larval stage (L3) somatic antigens using direct (IgA, IgG) and indirect (IgE) ELISA. This was done using blood samples taken each August between 1990 and 2015.

Genetic architecture of Ig levels. Sex-specific restricted maximum likelihood animal models were run in in ASReml-R v4.1 for IgA, IgE and IgG levels in lambs and adults. Fixed effects included age in days for lambs and years for adults. Random effects included the additive genetic component, birth year, mother identity, individual identity (permanent environment) ELISA plate number and run date. Genetic correlations between the sexes were fit in bivariate models using the CORGH and IDH functions for each age class using the same fixed and random effect structures. Sex-specific genome-wide association studies (GWAS) were carried out using the rGLS function in RepeatABEL v1.133, including a genomic kinship matrix and the same model structure defined as above. This approach fit SNPs as an additive effect (0, 1 and 2) for each genotype as a fixed covariate. The significance threshold at α=0.05 was set at P=2.245×10^{-6} as in Johnston *et al.* (2016). Gene by sex interactions were investigated using bivariate linkage disequilibrium score regression analysis (Bernabeu *et al.* 2021):

$$t = \frac{b_{males} + b_{females}}{\sqrt{SE^2_{males} + SE^2_{females} - 2r \cdot SE_{males} \cdot SE_{females}}}$$

where b is the estimated SNP effect sizes from the GWAS, SE is the standard error of b and r is the Spearman rank correlation (r).

Results

IgA, IgE and IgG levels were significantly heritable and were similar between the sexes within each age class (Figure 1 and 2). IgA had the greatest additive genetic effects (h^2=0.45-0.61), whereas IgE ranged from 0.19-0.46 and IgG from 0.18-0.26. The presence of high inter-individual variation, demonstrated by the permanent environmental and additive genetic effects, suggests repeatability of the antibody levels across the individual's lifetimes (repeatabilities of IgA: 0.74 and 0.77, IgE: 0.70 and 0.80, IgG: 0.51 and 0.48 for male and female adults; Figure 1 and 2). Generally, there were no significant maternal effects with the exception of female lamb IgA antibody levels. Other random effects including birth year, capture year, ELISA plate identity and run date explained between 1-8% of the phenotypic variance. Genetic correlations were strongly positive between the sexes for all three antibody traits in both lambs and adults, with rA ranging between 0.87 and 0.99.

Both male and female lambs confirmed previous associations with *CIITA/CLEC16A* (females: oar3_OAR24_10616039, P=1.18×10^{-17}; males: oar3_OAR24_10712874, P=1.70×10^{-15}). Male lamb IgE levels were significantly associated with *oar3_OAR9_12632265* on chromosome 9 (P=1.35×10^{-06}), ~1.05 Mb upstream from the candidate genes *SHARPIN* with mutations associated with IgA, IgE and IgG level variation in mice (Wang *et al.* 2012), and ~0.75 Mb upstream *LRRC14*, which is involved in NF-kB and TLR immune signalling (Wu *et al.* 2016). Male lamb IgG levels were associated oar3_OAR5_45528930 on chromosome 5 (P=1.67×10^{06}), located ~0.8 Mb downstream of the IL-9 cytokine gene, which is associated with IgG antibody production and decreased parasitic infection susceptibility in mice and humans (Petit-Frere *et al.* 1993; Dugas *et al.* 1993). No significant sex by genotype interactions effects (P<0.05) were observed for any Ig traits in either age or sex class.

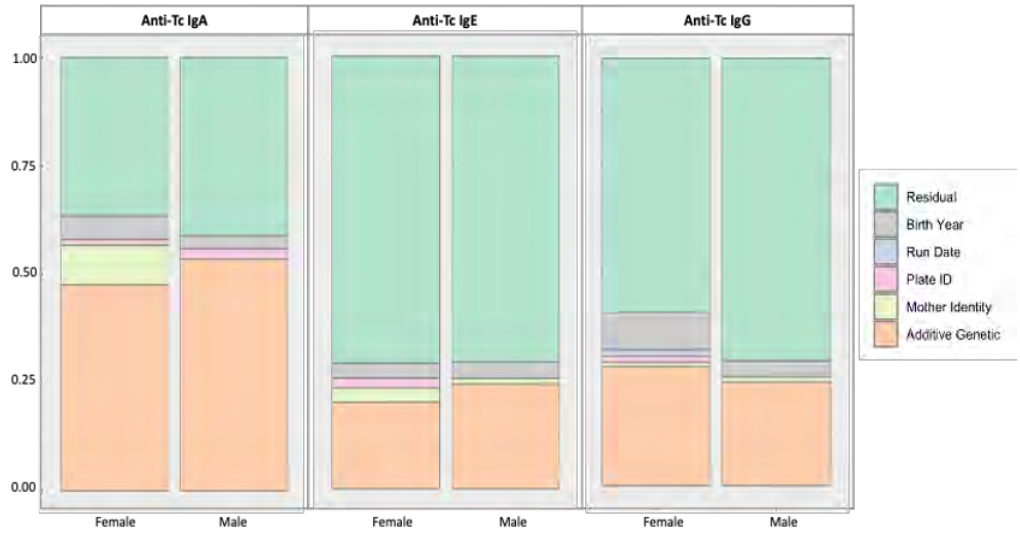


Figure 1. Proportion of phenotypic variance attributed to each random effect in lambs.

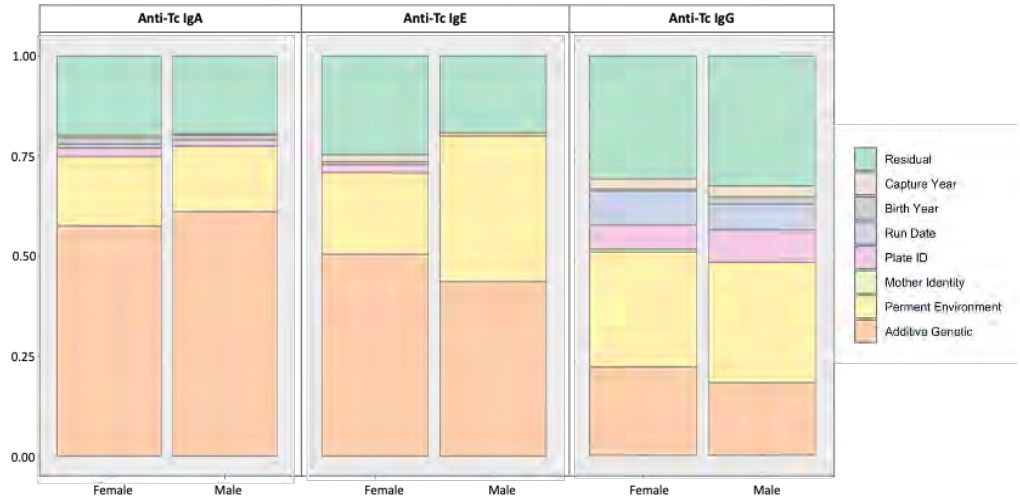


Figure 2. Proportion of phenotypic variance attributed to each random effect in adults.

Discussion

Soay females were previously shown to have higher IgA, IgE and IgG antibody levels, following trends of enhanced female innate and adaptive responses in other vertebrate species (Sparks *et al.* 2019). Nevertheless, we observed strong genetic correlations (r_A<0.867) and similar levels of trait heritability between males and females for IgA, IgE and IgG levels in Soay sheep, indicating a similar genetic basis between the sexes. High genetic correlations imply a genetic constraint, reducing the evolvability of sexual dimorphism in these traits and potentially explaining the maintenance of genetic variation (Rolff *et al.* 2005). However, the contributions of environmental heterogeneity should also not be overlooked in this analysis, especially considering the extreme variation in weather, population density and nutritional supply on St Kilda.

The loci *SHARPIN*, *LRRC14* and *IL-9* were associated with male Ig levels. It should be noted that despite occurring up to 1 Mb away, high linkage disequilibrium in this population indicates potential for these candidate regions to have an effect on immune phenotype (Sparks *et al.*, 2019), but should be confirmed with functional studies. Future work will also investigate associations between previously and currently identified immune loci with sex-specific reproductive success and survival, and explore how further mechanisms maintain variation, such as frequency-dependent selection, age, and parent-offspring conflict.

References

Bernabeu, E., Canela-Xandri, O., Rawlik, K., Talenti, A., Prendergast, J. *et al.* (2021) Nat. Genet. 53(9):1283–1289. https://doi.org/10.1038/s41588-021-00912-0

Craig, B. H., Pilkington, J. G., & Pemberton, J. M. (2006) Parasitology 133(Pt 4):485–496. https://doi.org/10.1017/S0031182006000618

Denholm, S. J., McNeilly, T. N., Banos, G., Coffey, M. P., Russell, G. C., *et al.* (2017) J. Dairy Sci. 100(4):2850–2862. https://doi.org/10.3168/jds.2016-11679

Dugas, B., Renauld, J. C., Pène, J., Bonnefoy, J. Y., Petit-Frère, C., Braquet, P. *et al.* (1993) Eur. J. Immunol. 23(7):1687–1692. https://doi.org/10.1002/eji.1830230743

Johnston, S. E., Bérénos, C., Slate, J., & Pemberton, J. M. (2016) Genetics, 203(1):583–598. https://doi.org/10.1534/genetics.115.185553

Kanda, N., Tsuchida, T., & Tamaki, K. (1996) Clin. Exp. Immunol. 106(2):410–415. https://doi.org/10.1046/j.1365-2249.1996.d01-842.x

Klein, S. L., & Flanagan, K. L. (2016) Nat. Rev. Immunol. 16(10):626–638. https://doi.org/10.1038/nri.2016.90

Liston, A., Carr, E. J., & Linterman, M. A. (2016) Trends Immunol. 37(10):637–646. https://doi.org/10.1016/j.it.2016.08.002

Metcalf, C. J. E., Roth, O., & Graham, A. L. (2020) Funct. Ecol. 34(1): 129–140. https://doi.org/10.1111/1365-2435.13458

Petit-Frere, C., Dugas, B., Braquet, P., & Mencia-Huerta, J. M. (1993) Immunology 79(1):146–151

Rolff, J., Armitage, S. A. O., & Coltman, D. W. (2005) Evolution, 59(8):1844–1850. https://doi.org/10.1111/j.0014-3820.2005.tb01831.x

Rowe, L., Chenoweth, S. F., & Agrawal, A. F. (2018) Am. Nat 192(2):274–286. https://doi.org/10.1086/698198

Sparks, A. M., Watt, K., Sinclair, R., Pilkington, J. G., Pemberton, *et al.* (2019) PLOS Genet. 15(11):e1008461. https://doi.org/10.1371/journal.pgen.1008461

Stoffel, M. A., Johnston, S. E., Pilkington, J. G., & Pemberton, J. M. (2021) Nat. Comm., 12(1):2972. https://doi.org/10.1038/s41467-021-23222-9

Van der Most, P. J., de Jong, B., Parmentier, H. K., & Verhulst, S. (2011) Funct. Ecol. 25(1):74–80. https://doi.org/10.1111/j.1365-2435.2010.01800.x

Wang, Z., Potter, C. S., Sundberg, J. P., & Hogenesch, H. (2012). J. Cell. Mol. Med. 16(10):2271–2279. https://doi.org/10.1111/j.1582-4934.2012.01574.x

Wu, C., Yang, Y., Ou, J., Zhu, L., Zhao, W., & Cui, J. (2016). Exp. Cell Res. 347(1):65–73. https://doi.org/10.1016/j.yexcr.2016.07.011

Zuk, M. (2009) PLOS Pathogens, 5(1):e1000267. https://doi.org/10.1371/journal.ppat.1000267

168. New tools and insights to enable breeding for reduced disease transmission

A. Doeschl-Wilson[1*], G. Marion[2] and C.M. Pooley[1,2]

[1]The Roslin Institute, University of Edinburgh, Easter Bush EH25 9RG United Kingdom; [2]Biomathematics and Statistics Scotland, The King's Buildings, Edinburgh EH9 3FD, United Kingdom; andrea.wilson@roslin.ed.ac.uk

Abstract

Theoretical and empirical evidence indicates there is far more host genetic variation underlying infectious disease prevalence than currently utilized in breeding programmes. In particular, genetic variation in host propensity to transmit infection is rarely fully captured. This paper describes methods now available for estimating genetic effects underlying infectious disease transmission, and novel insights regarding genetic (co-) regulation of the corresponding host epidemiological traits from empirical studies. Results show we now have the tools available to obtain reliable estimates for genetic effects for host susceptibility, infectivity and recovery or death rates from individual disease data, given the appropriate experimental or field study design. Substantial genetic (co-)variation appears to exist for all three epidemiological host traits and previously identified resistance genotypes may also confer genotypic differences in at least one if not all of these traits, thus holding great promise to drastically improve the effectiveness of genetic disease control.

Introduction

Reducing the transmission of infectious pathogens is a critical component of sustainable livestock production (Doeschl-Wilson *et al.*, 2021). Yet, it is currently not known how to breed animals for low disease transmission, as quantifying the underlying host genetic contributions remains a long-standing challenge. Epidemiological theory points to three key host traits regulating infectious disease transmission: susceptibility (propensity to acquire infection), infectivity (propensity to pass on infection to others) and recovery or death rate post infection, which affects the duration that an infected animal can transmit infection (Pooley *et al.*, 2020; Bijma *et al.,* 2021). Despite increasing empirical evidence for substantial genetic variation in these traits (e.g. Welderufael *et al.*, 2018, Anacleto *et al.,* 2019; Chase-Topping *et al.*, 2021; Prentice *et al.,* 2022), genetic parameter estimates for these traits rarely exist. Furthermore, the effects of previously identified resistance genotypes (e.g. based on single SNPs or estimated breeding values) on these epidemiological traits are largely unknown. Therefore, it is difficult to predict to what extent genetic selection for disease resistance also reduces disease transmission and prevalence (Hulst *et al.*, 2021; García-Ballesteros *et al.,* 2022).

In this paper we give an overview of recently developed methods and data requirements for estimating genetic parameters and genotypic effects for these three host epidemiological traits. We collate key insights about their genetic (co-)regulation gained from experimental and field studies in different host and pathogen species. We conclude by outlining further research needs to transform livestock breeding programmes to not only reduce the direct impact of infectious disease on individuals' production and health, but also their transmission and prevalence.

Materials & methods

Methods and software for estimating genetic effects underlying disease transmission. Compared to conventional genetic analyses of disease resistance phenotypes, models for estimating genetic effects underlying disease transmission need to incorporate epidemiological concepts (Pooley *et al.*, 2020; Bijma *et al.*, 2021). These concepts have been recently implemented in the open-source Bayesian inference software called SIRE (Susceptibility-Infectivity-Recoverability Estimator) which provides simultaneous estimates of

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_168, © A. Doeschl-Wilson *et al.* 2022

genetic effects on host susceptibility, infectivity and recovery / death rates from various types of individuals' disease records from epidemics that can be represented by an epidemiological susceptible-infectious-recovered (SIR) model (Pooley *et al.*, 2020, 2022a). SIRE implements three hierarchical models: (1) a genetic model for the three host epidemiological traits, which incorporates either the effects of a specific SNP or genotype (e.g. known to be associated with resistance) under consideration (Pooley *et al.*, 2020) or additive genetic (co)-variance for polygenic traits (SIRE 2.0; Pooley *et al.*, 2022b), together with other fixed and random effects; (2) an epidemiological model that links the three epidemiological traits to individuals' (usually unobservable) infection and recovery/death times; and (3) an observational model that links individuals' infection status to their disease records. The software has been validated with different types of data from a large range of simulated epidemics, which also provided insights into data requirements for obtaining sufficiently precise estimates for the genotypic effects (Pooley *et al.*, 2020) or genetic parameter estimates, and for achieving high prediction accuracies for the epidemiological traits (Pooley *et al.*, 2022a).

Methods and software to aid experimental designs. First empirical estimates of genetic effects for the epidemiological host traits may be best derived from controlled disease transmission experiments (Doeschl-Wilson *et al.*, 2021). We recently developed an open source online software tool SIRE-PC, (SIRE Precision Calculator) to assist with the design of these in the case of a single SNP under investigation (e.g. one already known to be associated with disease resistance) (Pooley *et al.*, 2022b). SIRE-PC provides precision estimates for SNP substitution and dominance effects for all three host epidemiological traits depending on experimental designs specified by the user, and optimal solutions for different types of experimental designs. These estimates are based on analytical expressions for the precision of estimated SNP substitution and dominance effects for the three epidemiological traits, and have been validated with simulated experiments using the above described SIRE software (Pooley *et al.*, 2022b).

Results

Software validation and insights from simulation studies. Extensive validations of both versions of the SIRE software show that it is possible to obtain accurate and unbiased estimates of genetic effects (e.g. SNP effects and genetic parameter estimates) associated with all three epidemiological host traits from individual temporal epidemic data, even when that data is incomplete (Pooley *et al.,* 2020, 2022b) for detailed assessment of influencing factors on these estimates). To summarize: inference of genetic parameter estimates and SNP effects for all three traits is possible even when only recovery (or death) times of individuals are known, or when measures of individuals' infection status are infrequent or censored. Unsurprisingly this requires larger datasets to gain equivalent precision and prediction accuracies compared with perfect data. For genetic infectivity effects, a large number of contact groups with few individuals generally provides better estimates than the converse.

Optimal designs for transmission experiments. Maximising analytical expressions for the precision on SNP effects led to three distinct 'experimental' designs (potentially consisting of several replicates), each specifying a different set of ideal SNP genotype compositions across contact groups: (1) appropriate for a single contact-group per replicate; (2) a multi-group 'mixed' design; and (3) a multi-group 'pure' design, where 'mixed' and 'pure' refer to contact groups consisting of contact individuals with different (mixed) or same genotypes, respectively (Figure 1). Precision estimates for infectivity were found more sensitive to the experimental design than those for susceptibility or recovery/death rate. The multi-group designs provide more precise estimates for infectivity effects than the single-group design containing the same number of individuals and records (e.g. standard deviation for the optimal single group design in Figure 1 was 1.5 higher than for the multi-group designs). Whilst the multi-group pure and mixed designs are predicted to yield similar precisions (results not shown), the pure design may be more intuitive and robust to stochastic

effects. In contrast, the mixed design may be preferable if the number of contact groups is limited or if artificially infected seeder animals are used.

Insights from disease transmission experiments and field studies. Table 1 shows summary results from various experiments that assessed the effects of known disease resistance genotypes on disease transmission. Most of these applied the 'multi-group pure design' (Figure 1). The studies confirm the predictions from simulation studies, that it is possible to obtain informative estimates for the genetic effects for all three epidemiological traits even from relatively small scale transmission experiments and incomplete data. There was a large difference between estimated genotypic effects on the diverse epidemiological host traits across studies, with e.g. genotypic differences in host infectivity ranging from negligible to large (Table 1). The same is true for the first published genetic parameter estimates for polygenic epidemiological host traits obtained from field and experimental studies (Welderrufael *et al.*, 2018; Biemans *et al.*, 2019; Prentice *et al.*, 2022; Sanchez-Molano *et al.*, 2022), which indicate that substantial genetic variation and co-variation could exist in any or all epidemiological traits.

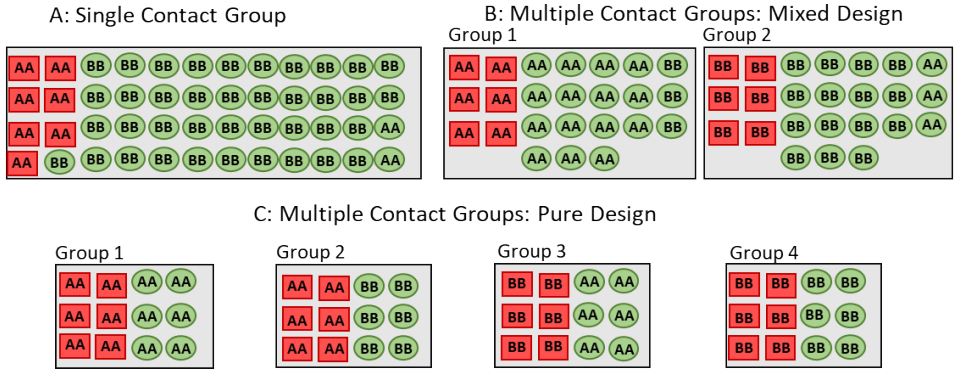


Figure 1. Optimal experimental designs for estimating genotype effects on host infectivity without compromising estimation of susceptibility and recovery rate. Assumptions underlying the analysis shown here: bi-allelic SNPs (A,B), only additive genotypic effects, random distribution of individuals regarding other influencing factors (e.g. family structure), fixed number of individuals (here n=48) per replicate. Red boxes: individuals infected at the start of the experiment; green circles: initially uninfected individuals.

Table 1. Empirical estimates of resistance genotype effects on transmission parameters.

Disease / species[Ref] [nr of animals]	Resistance genotype	Resistance phenotype	Effect on susceptibility	Effect on infectivity	Effect on recovery / death rate
PRRS /pigs[1] [164]	Genotype at *GBP5* resistance gene	Viral load	None found	None found	Not assessed
Marek's disease / chicken[2] [370]	Inbred lines with low / high genetic resistance	Tumour formation / viral load	Not assessed	None found	Not assessed
IPN/ Atl. Salmon[3] [580]	Genotype at 3 IPN resistance markers	Time to death	Favourable[5]	Favourable[5]	Favourable[5]
ISA\ Atl. Salmon[4] [420]	High / low resistance GBV[6]	Time to death	None found	Favourable[5]	Un-favourable[5]

[1] Chase-Topping *et al.*, 2022.
[2] Dunn *et al.*, 2020.
[3] Doeschl-Wilson *et al.*, 2018.
[4] Chase-Topping *et al.*, 2021.
[5] Favourable in the epidemiological sense, i.e. high resistance confers low susceptibility, low infectivity or fast recovery or mortality post infection.
[6] GBV: Genomic Breeding Value.

Discussion

Genetic infectious disease control would strongly benefit from shifting the current focus on improving individuals' disease resistance towards reducing disease transmission in populations (Bijma *et al.* 2021; Doeschl-Wilson *et al.* 2021). Estimating genetic effects for the three epidemiological host traits is now possible and constitutes an important step towards this goal. Necessary future developments include improvements in the existing computational inference methods to accommodate a larger variety of genomic and epidemic models and data, as well routine implementation of genetic-epidemiological models to predict response to selection (Bijma *et al.*, 2021; Doeschl-Wilson *et al.*, 2021; García-Ballesteros *et al.,* 2022).

References

Anacleto O., Cabaleiro S., Villanueva B., Saura M., Houston, R. D., *et al.* (2019) Sci Rep, 9(1): 1-12. https://doi.org/10.1038/s41598-019-40567-w

Biemans F., de Jong M., and Bijma P. (2019) Genet Sel Evol 51(1): 1-13. https://doi.org/10.1186/s12711-019-0505-3

Bijma P., Hulst A. D., and de Jong M. (2021) Genetics https://doi.org/10.1093/genetics/iyab024

Chase-Topping M. E., Pooley C., Moghadam H. K., Hillestad B., Lillehammer M. *et al.* (2021). Aquaculture, 535: 736365. https://doi.org/10.1016/j.aquaculture.2021.736365

Chase-Topping M., Plastow G., Dekkers J., Fang Y., Gerdts V., *et al.* (2022) Proc. of the 12th WCGALP, Rotterdam, the Netherlands

Doeschl-Wilson A., Anacleto O., Nielsen H.M., Karlsson-Drangshold T., Lillehammer M. *et al.* (2018) Proc. of the 11th WCGALP, Auckland, New Zealand.

Doeschl-Wilson A., Knap P. W., Opriessnig T., and More S. J. (2021) Animal, 15, 100286. https://doi.org/10.1016/j.animal.2021.100286

Dunn J.R., Mays J., Bailey R., Chase-Topping M., Cheng H.H., *et al.* (2020) Proc. Annual Conference of Research Workers in Animal Diseases (CRWAD), Chicago, IL, USA.

García-Ballesteros S., Fernández J., Doeschl-Wilson A. and Villanueva B., Proc. of the 12th WCGALP, Rotterdam, the Netherlands.

Hulst A. D., de Jong M. C., and Bijma P. (2021) Genetics, 217(4): iyab024. https://doi.org/10.1093/genetics/iyab024

Pooley C. M., Marion G., Bishop S. C., Bailey R. I., and Doeschl-Wilson, A. B. (2020) PLOS Comp. Biol. 16(12): e1008447. https://doi.org/10.1371/journal.pcbi.1008447

Pooley C. M., Marion G., Bishop S. C., and Doeschl-Wilson, A. B. (2022a) Proc. of the 12th WCGALP, Rotterdam, the Netherlands

Pooley C. M., Marion G., Bishop S. C., and Doeschl-Wilson, A. B. (2022b) bioRxiv 2022.01.10.475628; https://doi.org/10.1101/2022.01.10.475628

Prentice J., Pooley C.M., Tsairidou S., Pong-Wong R., Marion G., *et al.* (2022) Proc. of the 12th WCGALP, Rotterdam, the Netherlands.

Sánchez-Molano E., Madenci D., Tsairidou S., Banos G., and Doeschl-Wilson A. Proc. of the 12th WCGALP, Rotterdam, the Netherlands

Welderufael B. G., Løvendahl P., De Koning D. J., Janss L. L., and Fikse, W. F. (2018) Front. Genet. 9: 141. https://doi.org/10.3389/fgene.2018.00141.

169. Transmission experiment in turbot shows high genetic variation in host infectivity affecting disease spread and survival

J.C. Prentice[1*], C.M. Pooley[1,2], S. Tsairidou[1,3], R. Pong Wong[1], O. Anacleto[1], M. Saura[4], R.I. Bailey[1,5], G. Marion[2], B. Villanueva[4] and A.B. Doeschl-Wilson[1]

[1]The Roslin Institute, University of Edinburgh, Easter Bush, EH25 9RG, United Kingdom; [2]BioSS, The King's Buildings, Edinburgh, EH9 3FD, United Kingdom; [3]GAAFS, University of Edinburgh, Easter Bush, EH25 9RG, United Kingdom, [4]INIA-CSIC, Ctra. de La Coruña, km 7.5, 28040 Madrid, Spain; [5]Department of Ecology And Vertebrate Zoology, University of Lodz, 90-237 Łódź, Poland; jamie.prentice@ed.ac.uk

Abstract

Evolutionary theory predicts substantial genetic variation in host infectivity influences infectious disease spread and mortality rates in populations. This study combined data from a bespoke transmission experiment in turbot, and novel inference software (SIRE 2.1) to provide empirical genetic parameter estimates for host infectivity, in addition to susceptibility and infection induced mortality. Validation of SIRE 2.1 with simulated disease and survival data mimicking the experimental design provided reliable (co-) variance estimates and high prediction accuracies for all three traits. Confirming theoretical expectations, posterior means for genetic variance in host infectivity, estimated from the experimental data, were at least twofold higher than those for susceptibility and mortality respectively. Infectivity was positively genetically correlated with mortality. The results highlight the huge potential to reduce infectious disease transmission in livestock through genetic selection.

Introduction

Infectious diseases can cause high loss of animals in livestock production. To mitigate its impact on animal health and welfare, many breeders of aquatic and terrestrial livestock species tend to select animals that are genetically more likely to survive when exposed to infectious pathogens (Saura *et al.*, 2019). However, it is currently unknown to what extent this selection strategy reduces disease transmission, as survivors may still become infected and transmit infections, becoming 'tolerant super-spreaders' in the worst case scenario.

Determining animals' genetic contribution to infectious disease transmission, requires understanding the genetic (co-)regulation of three key epidemiological host traits: susceptibility (propensity to acquire infection), infectivity (propensity to pass on infection to others), and recovery or mortality post infection, which affects the duration that an infected animal can transmit infection (Pooley *et al.*, 2020). In line with the common genetic architecture of conventional disease traits, these traits are also expected to be polygenic. Furthermore, evolutionary theory suggests that genetic variation may be higher for host infectivity than for the other epidemiological host traits (Lipschutz-Powell *et al.*, 2012). However, to date no empirical genetic parameter estimates for these three traits exist despite strong evidence that the traits are under genetic control (Doeschl-Wilson *et al.*, 2022).

This study uses data from a bespoke transmission experiment for parasite infections in turbot fish (*Scophthalmus maximus*) that had been specifically designed to provide empirical genetic parameter estimates for all three epidemiological host traits (Anacleto *et al.*, 2019). The parasite was *Philasterides dicentrarchi*, which is the causative agent of scuticociliatosis, which manifests itself with the onset of visual signs, and is a common cause of mortality in turbot farming. Previous analyses of this experimental data provided evidence that genetic variation in host infectivity, in addition to variation in the time to onset of clinical signs (termed resistance) and the time from onset of signs to death (termed endurance) exists and

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_169, © J.C. Prentice *et al.* 2022

affects disease spread and survival of fish (Anacleto *et al.*, 2019). Furthermore, estimates for heritabilities and candidate SNPs for resistance and endurance traits by Saura *et al.* (2019), showed no significant genetic correlation between them.

The aim of this study was to estimate genetic parameters for the three unobservable host epidemiological traits (susceptibility, infectivity, and mortality post infection) from these data. For this purpose, we adapted our recently developed Bayesian inference software SIRE (Susceptibility-Infectivity-Recoverability Estimation) (Pooley *et al.*, 2022) to fit the data. To validate the empirical results, we evaluated the software first on simulated data.

Methods & materials

Experimental design and data. We refer to Anacleto *et al.* (2019) and Saura *et al.* (2019) for a detailed description of the experimental design and data. Briefly, the experiment consisted of two consecutive trials, each comprising ~900 turbot from 22 (Trial 1) and 23 (Trial 2) full-sibling families generated from 29 sires and 25 dams, respectively. Fish in each trial were distributed into 36 isolated tanks, each comprising 25 fish (5 initially infected seeder fish, 20 naive contact fish; 5 individuals per full-sib family; families were assigned to tanks to maximise statistical power for parameter estimation). In each tank the infection was seeded by inoculating the seeder fish at the start of the experiment, with the 50,000 (Trial 1) and 56,000 (Trial 2) ciliates of *Philasterides dicentrarchi* diluted in 200 µl saline (Anacleto *et al.*, 2019). Infected fish that developed scuticociliatosis were identified by the development of visible signs such as changes in colour or behaviour (Anacleto *et al.*, 2019). The phenotypic data used for the genetic analyses in this and previous studies consisted of individual records of the time (days post exposure) at which visual signs were first observed (t_S, which is different to the time of infection), and time of death (t_D). Sire and dam information was available for all individuals, and used to generate the numerator relationship matrix (A).

Bayesian inference of genetic parameter estimates for the epidemiological host traits. The recently developed Bayesian inference software SIRE 2.0 (Pooley *et al.*, 2022) was adapted to infer genetic parameter estimates for the three epidemiological host traits from this data. The software has been specifically designed to produce posterior parameter estimates for these traits from a wide range of individual disease records (Pooley *et al.*, 2020, 2022). Central to the inference process is to specify the underlying epidemiological model that links the epidemiological host traits to individual infection status, combined with the genetic models for each trait. The infection profiles of the seeder and contact fish strongly indicated that fish were infectious before they developed visual signs. Hence, an epidemiological model that fits this infection pattern is the SIDR model, where *S* refers to the susceptible state, *I* is infectious but undetectable, *D* is infectious and detectable, and *R* is removed (dead).

The rate of new infections S→I is described by the force of infection $\lambda_i = \beta \exp(g_i) \sum_j \exp(f_j)$ (summed over all infectious individuals *j* in the same group as *i*). The times taken for individual *i* to transition I→D and D→R are gamma distributed with means $1/\rho$ and $\exp(r_i)/\mu$, and shape parameters k_{ID} and k_{DR} respectively. Here β is the mean *per capita* transmission rate for the disease, ρ is the mean *per capita* detection rate, and μ is the mean *per capita* infection induced mortality rate, g_i, f_i, and r_i represent fractional deviations in the susceptibility, infectivity, and infection induced mortality rates of individual *i* compared to a population-wide reference. The phenotypic fractional deviations in susceptibility, infectivity, and infection induced mortality were defined as x = a + e, where x refers to the phenotypic vector for the three epidemiological traits, the additive genetic effect $a \sim \text{MVN}(0, \text{A}\otimes\Omega)$, where A and Ω are the numerator relationship matrix and is a 3×3 genetic trait covariance matrix, and the random environmental effects $e \sim \text{MVN}(0, \text{I}\otimes\Psi)$, where I and Ψ denote the 3×3 identity and environmental trait covariance matrices.

Validation of SIRE 2.1 with simulated data. To validate SIRE 2.1's ability to provide reliable estimates for genetic parameters and high prediction accuracies for the epidemiological traits, we first simulated the above experimental design (using the same pedigree, group layout, assignments, and population size), and generated simulated data for t_S and t_D using a stochastic implementation of the SIDR epidemiological model described above (algorithm described in Pooley *et al.* 2020). The values for the epidemiological trait means were chosen to provide similar population profiles for t_S and t_D as those observed in the experiment, the genetic and environmental trait variances were arbitrarily assumed to be 0.5 (h^2=0.5), and genetic correlations were set to 0.4. A total of 20 epidemiological replicates were simulated and the A-matrix, tank ID, and simulated times t_S and t_D for each individual were passed to SIRE 2.1 for inference. SIRE 2.1 was run to generate 10^7 samples, sufficient to ensure convergence (estimated sample size >200 for each parameter). Precision for variance and covariance estimates was assessed by comparing the posterior means and the 95% credible intervals (C.I.) with the true values, and prediction accuracies were estimated as the Pearson correlation between the true and estimated breeding values.

Application of SIRE 2.1 to experimental data. The procedure described above was applied to the experimental data for individuals' t_S and t_D. These were rounded to the nearest day, and t_S was adjusted to be one day earlier if both days coincided. Both trials were analysed separately, as their epidemic profiles differed substantially (Anacleto *et al.*, 2019).

Results

Validation of SIRE 2.1 with simulated data. SIRE 2.1 was able to provide reliable estimates for the genetic variances and correlations for all three epidemiological host traits (Figure 1). The results suggest that despite high uncertainties in the variance component estimates, the experimental design would allow detection of genetic variance and correlations between the traits, if they exist. Similarly, reasonably high prediction accuracies (95% C.I.) were obtained for all three traits. These were 0.594 (0.314, 0.795) for susceptibility *g*, 0.455 (0.209, 0.770) for infectivity *f*, and 0.772 (0.659, 0.871) for mortality *r*.

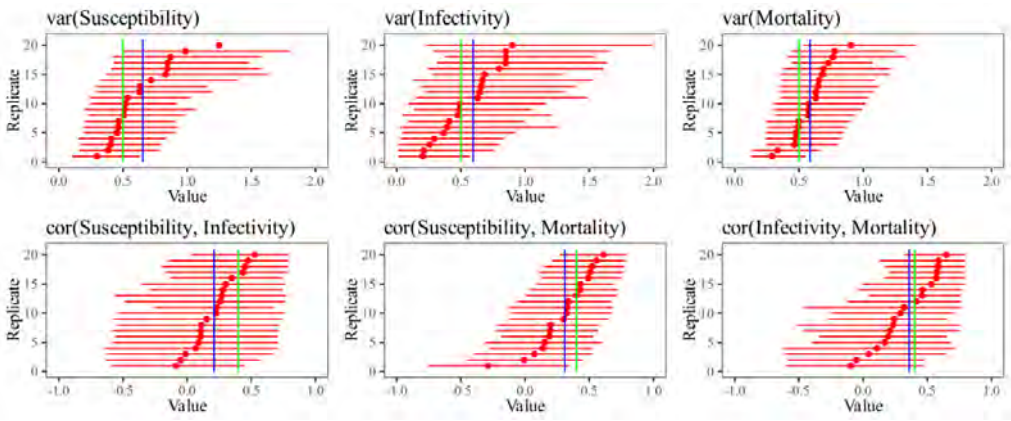


Figure 1. Validating estimated trait variance-covariances: Tornado plots showing estimated posterior means (red dots) with 95% C.I. (red horizontal lines) for additive genetic variances and correlations for the three epidemiological traits. Green vertical lines show the true parameter value, blue lines refer to the estimated mean across all 20 replicates. Each line refers to one replicate.

Genetic parameter estimates for host susceptibility, infectivity, and mortality from experimental data. Significant positive additive genetic variance was found for all three traits, with genetic variance in infectivity substantially larger than for susceptibility or mortality, respectively (Table 1). Despite substantial uncertainty in the estimates, genetic correlation between the traits seems to be weak to moderate.

Table 1. Estimated Ω: genetic variance (along diagonal) and correlation (off diagonal) estimates (95% C.I.) for log-normally distributed host traits, for Trials 1 and 2.

		Susceptibility	Infectivity	Mortality
Trial 1	Susceptibility	0.08 (0.00, 0.24)	-0.29 (-0.50, 0.26)	0.39 (0.01, 0.50)
	Infectivity		1.51 (0.63, 2.67)	0.24 (-0.16, 0.49)
	Mortality			0.66 (0.39, 0.96)
Trial 2	Susceptibility	0.06 (0, 0.24)	-0.07 (-0.48, 0.45)	0.08 (-0.46, 0.49)
	Infectivity		2.78 (2.28, 2.99)	-0.29 (-0.49, 0.17)
	Mortality			0.05 (0.00, 0.19)

Discussion

This study demonstrates that it is possible to obtain reliable genetic parameter estimates for three epidemiological traits controlling disease transmission given appropriate experimental design, and produces the first empirical estimates for all three traits. These findings support previous hypotheses that there is substantially more host genetic variation underlying disease prevalence than currently utilised in breeding programmes, and in particular in host infectivity (Bijma *et al.*, 2021; Lipschutz-Powell *et al.*, 2012). Further validation of these results is required by including additional effects and fitting a wider range of genetic and epidemiological models. However these preliminary results suggest that it may be possible to select turbot that not only have higher genetic resistance and endurance to scuticociliatosis infection, as previously suggested (Saura *et al.*, 2019), but are also less likely to transmit it.

Acknowledgements

The study was funded by the EU FP7 (KBBE.2013.1.2-10), the Scottish Government RESAS programme, and the Roslin ISP grant (BBS/E/D/20002172).

References

Anacleto O., Cabaleiro S., Villanueva B., Saura M., *et al.* (2019) Sci Rep 9:4924. https://doi.org/10.3389/fgene.2019.00539

Bijma P., Hulst A.D., and de Jong M. (2021) Genetics 217(4):iyab024. https://doi.org/10.1093/genetics/iyab024

Doeschl-Wilson A.B., Marion G., and Pooley C.M. (2022) Proc. of the 12th WCGALP, Rotterdam, the Netherlands.

Lipschutz-Powell D., Woolliams J.A., Bijma P., and Doeschl-Wilson A.B. (2012) PLOS ONE 7(6):e39551. https://doi.org/10.1371/journal.pone.0039551

Pooley C.M., Marion G., Bishop S.C., Bailey R.I., and Doeschl-Wilson A.B. (2020) PLOS Comput Biol 16(12):e1008447. https://doi.org/10.1371/journal.pcbi.1008447

Pooley C.M., Marion G., Bishop S.C., and Doeschl-Wilson A.B. (2022) Proc. of the 12th WCGALP, Rotterdam, the Netherlands.

Saura M., Carabaño M.J., Fernández A., Cabaleiro S., *et al.* (2019) Front Genet 10:539. https://doi.org/10.3389/fgene.2019.00539

170. Impact of selection for disease resistance on the spread of infection in a simulated aquaculture population

S. García-Ballesteros[1*], J. Fernández[1], A. Doeschl-Wilson[2] and B. Villanueva[1]

[1]INIA-CSIC, Crta. La Coruña km 7.5, 28040 Madrid, Spain; [2]The Roslin Institute, University of Edinburgh, EH25 9RG, Midlothian, United Kingdom; silvia.garciab@inia.es

Abstract

Disease resistance is a common breeding objective in aquaculture selection programmes. The main phenotype recorded is the time of death, which depends on three underlying epidemiological host traits: susceptibility, infectivity and infection induced mortality. The aim of this simulation study was to predict the impact of selection for time of death on these underlying traits and on the spread of the disease in a fish population. The time of death phenotypes were obtained by simulating virus transmission in a disease challenge experiment for Atlantic salmon with genetic variation assumed in the three epidemiological traits. Three scenarios varying in the point where the challenge test was stopped were simulated. Selection for time of death reduced disease spread evaluated through the basic reproductive ratio in all scenarios. However, the greatest decrease in the basic reproductive ratio (reaching values <1) was when the challenge test continued until mortality naturally levels off.

Introduction

Infectious disease outbreaks are a major threat to the economic and environmental sustainability of the aquaculture industry and it is accepted that selective breeding could be a key factor to prevent these outbreaks (Houston *et al.* 2020). In fact, disease resistance (the ability to survive when exposed to the pathogen) is included as a breeding objective in species with advanced selective breeding programmes.

In aquaculture programmes, disease resistance is often recorded on sibs of selection candidates in challenge tests, and the main phenotype is time of death (Doeschl-Wilson *et al.* 2021). However, it is unknown whether selecting directly on time of death reduces pathogen transmission and, thus, the incidence and severity of outbreaks (Doeschl-Wilson *et al.* 2021). This is because the spread of the disease is affected by three underlying epidemiological host traits, which are difficult to record: susceptibility (the propensity of an uninfected individual to become infected when exposed to infectious material), infectivity (the ability of an individual, once infected, to transmit the infection) and mortality (the propensity of an infected individual to die).

The objective of this study was to predict, through computer simulations, the impact of selection for disease resistance measured as time to death on the three underlying traits and, therefore, on the spread of the disease by mimicking a selective breeding programme for Atlantic salmon.

Materials & methods

Genetic and population models. The genome simulated was composed of 30 chromosomes of 1 Morgan each, approximating the salmon genome. A total of 31,500 evenly spaced biallelic loci (including 1,500 SNP markers to be used in genomic evaluations) were simulated per chromosome. The base population was created in a two-step process. Firstly, a large population in mutation-drift equilibrium was generated by randomly selecting 500 males and 500 females across 4,000 discrete generations. Secondly, once the equilibrium was reached, 100 males were randomly mated with 200 females (200 full sib families) in order to simulate a base population with family structure. Each mating produced 30 fish (10 selection candidates and 20 sibs to be used in the challenge test).

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_170, © S. García-Ballesteros *et al.* 2022

Underlying traits. Three uncorrelated traits (susceptibility, infectivity and mortality) were considered. A total of 1000 loci (randomly sampled from the 31,500 loci) affected simultaneously the three traits for which the initial phenotypic variance and the heritability (h^2) were 1 and 0.5, respectively, leading to a h^2 for the observed trait (time of death) of 0.2. The phenotypic value was obtained by adding to the true breeding value an environmental deviation sampled from a normal distribution with mean 0 and variance 0.5.

Observed trait. The observed trait was time of death (day when an individual dies from the beginning of the challenge test) which depends on infection and, once infected, death times. These times were obtained assuming a stochastic compartmental Susceptible-Infected-Removed epidemiological model. The algorithm used (based on Anacleto *et al.* 2015) specifies inter-event times which determine the transition rates of individuals to move between susceptible to infected determined by the force of infection (the infection rate of an individual j) $\lambda_j = \beta \exp(S_j) \sum_i \exp(I_i)$ and to move between infected to dead $w_j = (\gamma \exp(M_j))$, where β is a mean *per capita* transmission rate for the disease, γ is a mean *per capita* infection induced death rate, S_j and M_j represent fractional deviations in the susceptibility and infection induced mortality rates of individual j compared to a population-wide reference, and I_i represents the fractional deviation in infectivity of individual i (the sum goes over all infected individuals sharing the same contact group as j at any given point in time). Estimates for β and γ were obtained by matching simulated infection and survival profiles to those from published studies of ISA virus challenge experiments in Atlantic salmon with fast transmission (Ødegård *et al.* 2007; Chase-Topping *et al.* 2021). Simulated challenge tests consisted of 20 groups, each of 200 fish (one from each family). The infection was seeded by randomly choosing five fish per group as initially infected. The rest of the individuals were considered susceptible. Three scenarios varying in the point where the challenge test was stopped were simulated. The test was stopped when the cumulative mortality reached an asymptote, which occurred at ~60 days with 80% of dead fish (scenario A), when 50% of the fish per group were dead, which occurred on day ~16 (scenario B) or when 25% of the fish per group were dead, which occurred on day ~7 (scenario C). The survivors' phenotype was assumed to be the last day of the challenge test plus one (i.e. day 61, 17 and 8 for A, B and C, respectively).

Genetic evaluation and selection. Standard genomic evaluation (GBLUP) using 1,500 SNPs/Morgan was implemented to obtain estimates of breeding values for the observed trait. The BLUPf90 software (Misztal *et al.* 2015) was used for carrying out GBLUP evaluation. The 100 males and the 200 females with the highest estimated breeding value for time of death were selected each generation. Each selected male was randomly mated with two females and each mating produced 30 fish (10 selection candidates and 20 sibs to be used in the challenge test). Selection was practised for 10 discrete generations.

Parameters evaluated. For each generation t, the cumulative genetic gain for each underlying trait was calculated as the true breeding value averaged across all fish born in t. The basic reproductive ratio (R_0) was also evaluated as a measure of the impact of selection for disease resistance (time to death) on the spread of the disease. Extrapolating from Hyman and Li (2000) the expected theoretical R_0 was estimated for each group by randomly choosing an individual as infected index case (IC) in an otherwise susceptible population:

$$R_0 = (\beta/\gamma) * \left(\sum_{i=1}^{n} (\exp(S_i)) \Big/ n \right) * \left(\exp(I_{IC}) / \exp(M_{IC}) \right),$$

where β and γ (defined above) are 0.8 and 0.4 respectively, n is the total number of susceptible fish per group (excluding the fish chosen as IC), S_i is the susceptibility phenotype of fish i and I_{IC} and M_{IC} are respectively the infectivity and mortality phenotypes of the fish chosen as IC. For each group, R_0 was computed as the average R_0 calculated with the five initially infected fish as IC. Finally, for a given generation, R_0 was calculated as the average R_0 across groups. A total of 25 replicates were simulated and the results presented are averages over replicates.

Results

Selection for disease resistance (time of death) led to a reduction in all three underlying traits although the magnitude of the response for each trait differed across the scenarios considered (Figure 1). In scenario A, the genetic gain was mainly for susceptibility. As the duration of the challenge test decreased, the gain of the three traits became more similar, although in the three scenarios the lowest response was obtained for infectivity. At t=10, the genetic gain for susceptibility was 30% (scenario B) and 47% (scenario C) lower than in scenario A. However, the gain for mortality and infectivity was about double in scenarios B and C than in A.

Selection for time of death decreased R_0 in all scenarios (Figure 2). However, only scenario A led to an R_0 value below 1 and this was reached at t=5. Reducing R_0 below 1 is an important objective, as it implies that the disease cannot spread.

Discussion

In this study, we have simulated a disease leading to high mortality within a 60-day time window with mortality profiles similar to those found in some diseases of great importance such as infectious salmon anaemia (Ødegård *et al.* 2007, Holborn *et al.* 2020). We show that selection for disease resistance (measured as time of death) not only increases survival (Houston *et al.* 2020, Doeschl-Wilson *et al.* 2021), but also contributes to reducing disease transmission in a severe outbreak when the underlying traits are assumed to be uncorrelated.

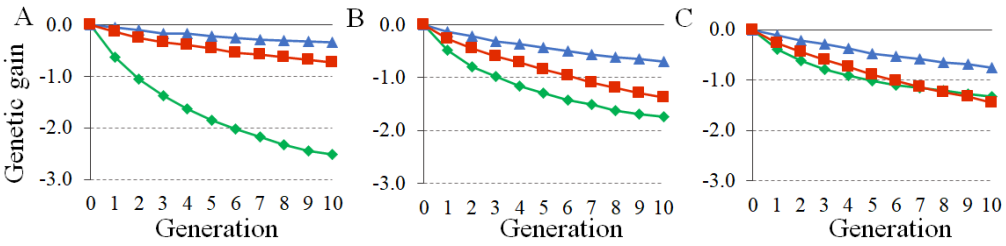


Figure 1. Cumulative genetic gain across generations for each trait (susceptibility in diamonds, infectivity in triangles and mortality in squares) for scenarios varying in the proportion of cumulative mortality at the end of the challenge (scenarios A, B and C).

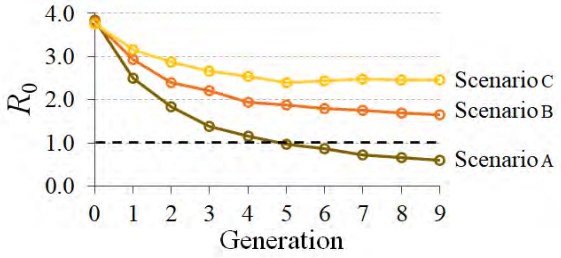


Figure 2. Basic reproductive ratio (R_0) across generations for scenarios varying in the proportion of cumulative mortality at the end of the challenge (scenarios A, B and C).

The genetic response for each underlying trait and R_0 depended on when the challenge test was stopped. The closer the end of the challenge test was to the point where cumulative mortality rate reached an asymptote, the greater the response in susceptibility and the greater the decrease in R_0. In this case, the survivors at the end of the test were mainly fish with low susceptibility (they were not infected or they were infected late and did not die before the end of the test). However, when the challenge test was stopped before the asymptote in the cumulative mortality rate was reached, the genetic response in susceptibility and the decrease in R_0 were lower and the response in infectivity and infection induced mortality was higher. In this case, the survivors were fish with low susceptibility, but also fish with medium and high susceptibility and medium and low mortality (i.e. fish that were infected but survived with the infection for a long time and had not yet died when the challenge test was stopped) and also fish with high and medium susceptibility that coexist with groups of infected fish with less infectivity (in scenarios B and C, fish with low infectivity and low mortality have less chance/time of infecting).

In summary, our results highlight the fact that, if the selection objective is to control the spread of the disease, the challenge test should continue until mortality naturally levels off.

Acknowledgements

This work was funded by INIA-CSIC (FPI-SGIT2016-02) and by MCIN/ AEI /10.13039/501100011033 (Project PID2020-114426GB-C2), and the Roslin ISP grant (BBS/E/D/20002172).

References

Anacleto O., Garcia-Cortés L.A., Lipschutz-Powell D., Woolliams J.A., Doeschl-Wilson A. B. (2015). Genetics, 201(3), 871-884. https://doi.org/10.1534/genetics.115.179853

Chase-Topping M.E., Pooley C., Moghadam H.K., Hillestad B., Lillehammer M. *et al.* (2021) Aquaculture 535:736365. https://doi.org/10.1016/j.aquaculture.2021.736365

Doeschl-Wilson A., Knap P.W., Opriessnig T., and More S.J. (2021) Animal 15:100286. https://doi.org/10.1016/j.animal.2021.100286

Holborn M.K., Ang K.P., Elliott J.A.K., Powell F., and Boulding E.G. (2020) Aquaculture 514:734514. https://doi.org/10.1016/j.aquaculture.2019.734514

Houston R.D., Bean T.P., Macqueen D.J., Gundappa M.K., Jin Y.H. *et al.* (2020) Nat. Rev. Genet. 21(7):389-409. https://doi.org/10.1038/s41576-020-0227-y

Hyman J.M., and Li J. (2000) Math Biosci. 167(1):65-86. https://doi.org/10.1016/S0025-5564(00)00025-0

Misztal I., Tsuruta S., Lourenco D., Aguilar I., Legarra A. *et al.* (2015) Available online at http://nce.ads.uga.edu/wiki/lib/exe/fetch.php?media=blupf90_all2.pdf

Ødegård J., Olesen I., Gjerde B., and Klemetsdal G. (2007) Aquaculture 266:70-76. https://doi.org/10.1016/j.aquaculture.2007.02.012

Part 12

Challenges – across species and breed genomics genetics

171. Application of haplotype relationship matrices for genomic prediction in purebred and crossbred cows

M. Khansefid[1*], M.H. Ferdosi[2], M.E. Goddard[1,3], M. Haile-Mariam[1], C. Schrooten[4], G. de Jong[4], E. O'Connor[5], H.D. Daetwyler[1,6], J.E. Pryce[1,6] and I.M. MacLeod[1]

[1]Agriculture Victoria, AgriBio, Centre of AgriBioscience, Bundoora, Victoria 3083, Australia; [2]AGBU, a joint venture of NSW Department of Primary Industries and University of New England, 2351, Armidale, Australia; [3]Faculty of Veterinary and Agricultural Sciences, University of Melbourne, Parkville, Victoria 3010, Australia; [4]CRV, P.O. Box 454, 6800 AL Arnhem, the Netherlands; [5]CRV Ambreed, Hamilton, Waikato, 3214, New Zealand; [6]School of Applied Systems Biology, La Trobe University, Bundoora, Victoria 3083, Australia; majid.khansefid@agriculture.vic.gov.au

Abstract

We assessed the usefulness of defining kinship based on the combinations of alleles (haplotypes) instead of individual alleles for genomic predictions of milk traits. The accuracy and bias of predictions were calculated when pure Holstein (H), pure Jersey (J) and/or H×J crossbred cattle were included in the reference set. For within-breed predictions, using haplotypes marginally improved the accuracy by 1%. For across-breed predictions, using haplotypes of 5 adjacent markers increased the accuracy of prediction for fat yield for H and J by 4 and 2%, respectively. The most consistent gains for H (1.8%) and J (1.5%) were achieved when all pure and crossbred animals were included in the reference. The change in accuracy of predictions for crossbred animals was not consistent, which could be associated with shorter haplotypes and lower accuracy of phasing. Generally, the bias of predictions was lower when haplotypes were used in the relationship matrix.

Introduction

Genomic prediction relies on linkage disequilibrium (LD) between single nucleotide polymorphism (SNP) markers and causal variants. However, this LD could be different across breeds mainly due to crossing over and change in allele frequency after breed formation. Consequently, the accuracy of across-breed predictions is lower than within breed predictions (Khansefid *et al.*, 2020). To avoid this problem, the causal mutations, or the SNPs close to them should be used to build the genomic relationship matrix (GRM). However, finding all of the causal mutations is difficult because most of them have very small effects (Xiang *et al.*, 2021). Using haplotypes instead of SNPs might better describe the relationship between individuals of different breeds in admixed populations because it could distinguish identity by descent from identity by state to some extent (Ferdosi *et al.*, 2016). In this study, we calculated the accuracy and bias of genomic estimated breeding values (GEBV) for milk traits by replacing the GRM with a haplotype relationship matrix (HRM). We assessed the usefulness of the HRM for within-breed, across-breed and crossbred genomic prediction.

Materials & methods

Phenotypes and genotypes. The data used in this study were described by Khansefid *et al.* (2020). The animals were recorded for milk traits in New Zealand. For milk, fat and protein yields, the de-regressed proofs (DRP) for bulls and their equivalent records for cows (adjusted fixed effects) were calculated by DataGene (all cow phenotypes were processed on the Australia scale). The animals had XT_50k imputed genotypes which consisted of 46,516 SNPs enriched for causal variants (Xiang *et al.*, 2021). In total, 14,987 pure H and 5,016 pure J bulls and cows as well as 20,281 crossbred cows were used in this study. The crossbred cows were divided into three subgroups according to their breed proportions (~75%H:25%J,

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_171, © M. Khansefid *et al.* 2022

~50%H:50%J, and ~25%H:75%J). The genotypes of all animals were phased using Eagle v2.4.1 with default parameters (Loh *et al.*, 2016).

Reference and validation sets. Four reference sets in our study were R_1 (13,985 pure H), R_2 (4,484 pure J), R_3 ($R_1 + R_2$) and R_4 ($R_1 + R_2$ + 18,469 H×J crossbred). The validation sets consisted of 1,002 pure H, 863 75%H:25%J, 868 50%H:50%J, 324 25%H:75%J and 532 pure J cows which were selected to avoid high relationships with animals in the reference sets.

Prediction model. MTG2 (Lee and van der Werf, 2016) was used to calculate GEBV when the same validation set but different reference sets were included in the model (Equation 1).

$$y = Xb + Zu + e \quad (1)$$

where y contains DRP for milk, fat, or protein yields for the animals in the reference; b, $u \sim N(0, A\sigma^2_a)$ and $e \sim N(0, E\sigma^2_e)$ are the vectors of fixed effects (mean and sex-breed group), random effects (GEBV) and random residual effects, respectively; X and Z are design matrices. The genetic variance (σ^2_a) in this model was estimated when the relationships between animals in matrix A were defined by either a GRM (Yang *et al.*, 2010) or HRM. The HRM was calculated using the sliding window algorithm (Ferdosi *et al.*, 2016) and three haplotype sizes (5, 10 and 15). The error variance (σ^2_e) was estimated considering the weighting coefficients in diagonal E matrix (Garrick *et al.*, 2009).

Accuracy and bias of genomic predictions. The accuracy (correlation coefficient between GEBV and DRP), and the bias (regression coefficient of DRP on GEBV) of genomic predictions were calculated and reported separately for each of the validation breed groups.

Results

The accuracy and bias of the GEBV were shown in Figure 1 and Figure 2, respectively. Replacing the GRM with the HRM: (1) marginally improved the accuracy of within-breed predictions for milk traits (i.e. using R_1 to predict H and R_2 to predict J) and generally reduced the bias except for fat yield in J. (2) The accuracy and bias of across-breed predictions (i.e. using R_1 to predict J and R_2 to predict H) was only improved for fat yield. (3) In R_3 (i.e. H+J purebred reference), the accuracy and bias of predictions generally improved for milk traits in purebred cows. (4) In R_4, (i.e. purebred and crossbred animals in the reference), the accuracy of predictions improved for milk traits in purebred cows but marginally improved just for protein yield in crossbred cows.

Using HRM instead of GRM in the prediction model increased the estimated σ^2_a because HRM absorbed some of the variance explained by breed groups which was fitted as a fixed effect. However, the estimated σ^2_e was reduced by 1 to 27%, where larger reductions were associated with the size of the haplotypes used to build HRM. The average reduction in the estimated σ^2_e was smaller in R_2 (6%) compared to others references (16%).

The goodness of fit of the models (i.e. log likelihood) was only improved for protein yield and in R_1, R_3 and R_4 ($P<0.01$). Furthermore, haplotype sizes of 5 and 10 outperformed haplotype size of 15 in terms of the model goodness of fit.

Discussion

The usefulness of replacing the GRM with HRM in the prediction model was not consistent across different traits and haplotype sizes. In a within breed context Ferdosi *et al.* (2016) also reported that optimum

haplotype length for each trait and dataset could be different. This could be associated with the distribution of causal mutations and SNPs across the genome as well as the relationship between animals. Hence, for the closer relatives, the larger haplotype sizes may perform better. As we selected the validation set to be least related to the reference set, the improvement in predictions was limited. Nevertheless, there was a small consistent improvement in accuracy for within-breed genomic prediction. We found that using HRM was not promising for crossbred prediction. This could be associated with different optimum length of haplotype to define relationships in crossbred and purebred animals. Further, the phasing in crossbred animals could be less accurate compared to purebred animals because of expected higher imputation errors and larger effective population size in crossbred animals.

The application of HRM was most useful for genomic prediction of protein yield according to the goodness of fit in the model. However, the accuracy of across-breed prediction was almost zero when GEBVs of pure J were calculated from R_1 (i.e. pure H). Haile-Mariam *et al.* (2019) also reported the lowest correlation between milk traits across different breed groups was for protein yield. Hence, probably the effect of haplotypes on protein yield was largely breed specific. As identity by descent is considered in defining the relationships in HRM, the accuracy of across-breed predictions was reduced but was generally improved in R_4 (i.e. multibreed reference) for all validation breed groups.

In conclusion, HRM has limited benefits such as marginal increase in accuracy and reduction of bias of predictions compared to GRM. The advantages of using haplotypes instead of SNPs to define relationships could be potentially higher using standard Bovine 50k instead of a chip already enriched for causative mutations. However, XT_50k genotypes were used in this study because we wanted to find whether replacing the GRM with HRM could improve the accuracy and reduce the bias of prediction when the best

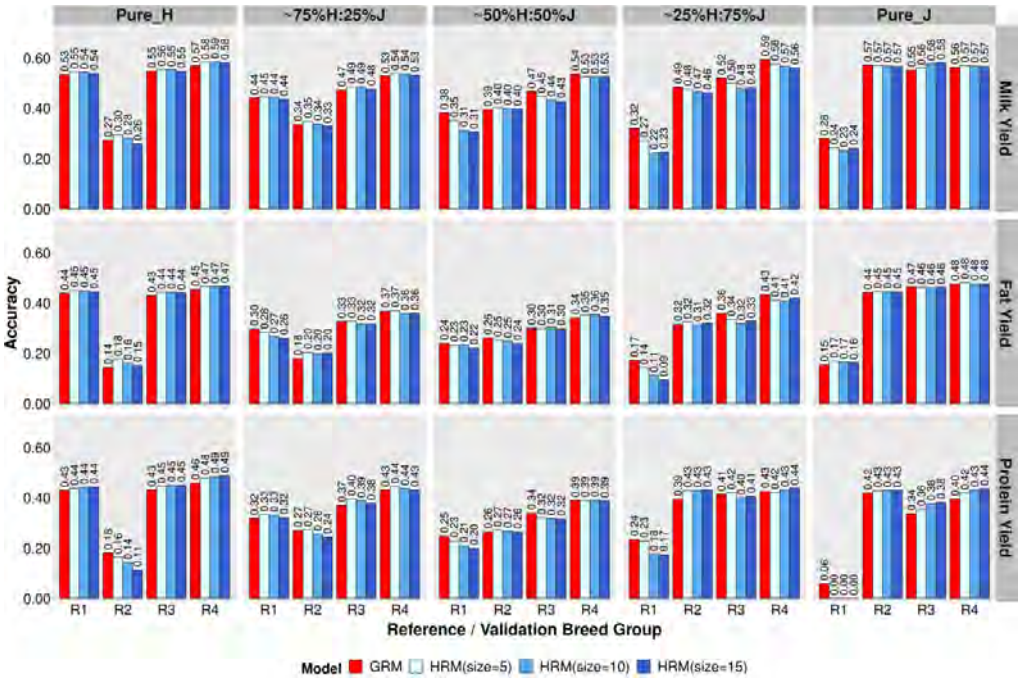


Figure 1. Accuracy of genomic prediction for milk, fat and protein yields in five validation breed groups using different reference sets and relationship matrices.

available set of SNPs were used in genomic prediction (Khansefid *et al.*, 2020). Considering the haplotype frequency in calculating HRM may improve the accuracy of predictions and reduce the inflated estimation of σ^2_a. Otherwise, HRM could explain the variance associated with population structure better than the fixed effects such as breeds groups. Nevertheless, the reduction in σ^2_e and improved log likelihood could be still useful to assess the improvement in goodness of fit of the model.

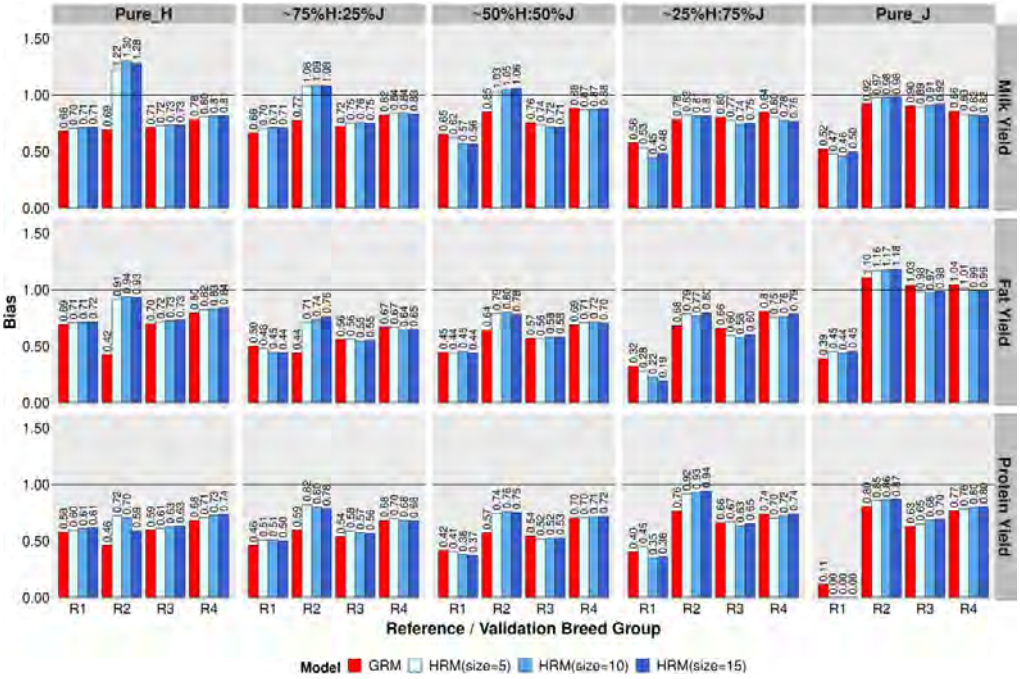


Figure 2. Bias of genomic prediction for milk, fat and protein yields in five validation breed groups using different reference sets and relationship matrices.

References

Ferdosi M.H., Henshall J. and Tier B. (2016) Genet. Sel. Evol. 48(1):1-14. https://doi.org/10.1186/s12711-016-0253-6

Garrick D.J., Taylor J.F. and Fernando R.L. (2009) Genet. Sel. Evol. 41(1):1-8. https://doi.org/10.1186/1297-9686-41-55

Haile-Mariam M., MacLeod I., Khansefid M., Schrooten C., O'Connor E., *et al.* (2019) Proc. of the 23rd Assoc. Advmt. Anim. Breed. Genet., Armidale, Australia.

Khansefid M., Goddard M.E., Haile-Mariam M., Konstantinov K.V., Schrooten, *et al.* (2020) Front. genet. 11(1):1-14. https://doi.org/10.3389/fgene.2020.598580

Lee S.H. and Van der Werf J.H. (2016) Bioinformatics 32(9):1420-1422. https://doi.org/10.1093/bioinformatics/btw012

Loh P.R., Danecek P., Palamara P.F., Fuchsberger C., Reshef Y.A., *et al.* (2016) Nat. Genet. 48(11):1443-1448. https://doi.org/10.1038/ng.3679

Yang J., Benyamin B., McEvoy B.P., Gordon S., Henders A.K., *et al.* (2010) Nat. Genet. 42(7):565-569. https://doi.org/10.1038/ng.608

Xiang R., MacLeod I.M., Daetwyler H.D., de Jong G., O'Connor E., *et al.* (2021) Nat. Commun. 12(1):1-13. https://doi.org/10.1038/s41467-021-21001-0

172. *Bos taurus* and *Bos indicus* reduced representation bisulfite sequencing comparison showed high epigenetic variability in *Bos* subspecies

E. Capra[1*], B. Lazzari[1], M. Milanesi[2,3], G.dP. Nogueira[2], J.F. Garcia[2,3], P. Ajmone Marsan[4] and A. Stella[1]

[1]Istituto di Biologia e Biotecnologia Agraria, Consiglio Nazionale delle Ricerche IBBA CNR, Milan, Italy, [2]School of Veterinary Medicine, Araçatuba, Department of Production and Animal Health, São Paulo State University (unesp), Araçatuba, Brazil, [3]International Atomic Energy Agency, Collaborating Centre on Animal Genomics and Bioinformatics, Araçatuba, Brazil, [4]Università Cattolica del Sacro Cuore, Piacenza, Italy; emanuele.capra@ibba.cnr.it

Abstract

Bos indicus and *Bos taurus* subspecies possess distinct morphological traits as a consequence of artificial selection and environmental adaptation. However, the impact of epigenetic contribution in relation to breed adaptation in the two subspecies has not been explored. In this work, Reduced Representation Bisulfite Sequencing (RRBS) profiling of the indicine Nellore (N) and the taurine Angus (A) cattle breeds was compared to identify CpG methylation differences in the two subspecies. A total of 134,553 methylated regions (MRs) were detected across all 20 samples (5 N and 5 A in two seasons). Approximately one third of them show differences in the two breeds in genes which function is related to sensory perception of chemical stimulus, development process and regulation of cellular organization. Further investigation is necessary to understand the real impact of epigenetic differences in the two breeds and how they can regulate genes important for cattle adaptation.

Introduction

Cattle belong either to the Indicine or the Taurine subspecies, the results of two independent domestication processes. Indicine breeds are more adapted to semi-arid, arid and tropical wet/dry climates and to parasites, and possess specific distinct morphological traits, as a cervico-thoracic hump, large ears and excess skin (Utsunomiya *et al.*, 2019, Verdugo *et al.*, 2019). European taurine breeds are characterized by high milk production potential or excellent carcass and meat yield and quality, but poor adaptation to hot environments (Manirakiza *et al.*, 2017; Rodrigues *et al.*, 2017). Both cattle subspecies have been extensively characterized at the genome-wide level and at specific loci (Bolormaa *et al.*, 2013; Pérez O'Brien *et al.*, 2014; Hu Y *et al.*, 2020; Barbato *et al.*, 2020) but few studies have focused on epigenomic diversity. Recently, we reported genome-wide DNA methylation changes in blood samples from indicine (Nellore) and Taurine (Angus) breeds under heat stress (Del Corvo M, *et al.*, 2021). After stressful period, a breed specific response was observed. Significant changes in methylation were observed in genes related to stress response and cellular defence in Nellore, whereas the response was less focused in Angus. Although heat stress seems induces changes in methylation patterns at specific loci, a remarkable diversity in methylation pattern was observed comparing the overall methylation of the two subspecies independently on the environmental challenge, presumably related to their origin and breed characteristic. In this paper, reduced representation bisulfite sequencing (RRBS) data from Nellore and Angus breed collected in different seasons, were analysed to investigate in more detail the epigenetic differences between the two subspecies.

Materials & methods

Data were obtained for ten half-sibs males of 15 months of age: an indicine breed (5 Nellore) and a Taurine breed (5 Angus) were reared at the UNESP Aracatuba (Sao Paulo state, Brazil) experimental station (located at -21.186244 latitude and -50.439053 longitude) during the hot Winter (C= Challenge period, heat stress)

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_172, © E. Capra *et al.* 2022

and until the following cooler Spring (R=Recovery period) (del Corvo *et al*., 2021). DNA was isolated with QIAamp DNA Blood Midi Kit (Qiagen) following manufacture instruction. One µg of genomic DNA was used for RRBS libraries preparation. Illumina sequence reads were analysed using nf-core (Ewels *et al*., 2020), via the nf-core/methylseq v1.5 pipeline (https://github.com/nf-core/methylseq) selecting Bismark v0.22.3 (https://www.bioinformatics.babraham.ac.uk/projects/bismark/) as aligner. To evaluate the epigenetic variation in the two sub-species, sequence reads were aligned to the bisulfite-converted *Bos taurus* reference genome (ARS-UCD1.2: GCF_002263795.1), and methylation calls were extracted using the Bismark methylation extractor v0.22.3 function. The Seqmonk (https://www.bioinformatics.babraham.ac.uk/projects/seqmonk/) software v1.47.1 was used for visualization and analysis of the Bismark output. Methylated regions (MRs) were detected by dividing the genome in 400bp tiles with sliding windows of 200bp and analyzing average methylation for tiles containing at least three mutually covered 5× Cs in CpG context per position. Differentially methylated regions (DMRs) between Angus and Nellore were calculated for 5 A and 5 N in the two seasons using the logistic regression filter in R (FDR≤0.05, absolute cut-off of 10%). MRs and DMRs were investigated for 5' UTR, 3' UTR, exons and introns. Hierarchical clustering was calculated for all MRs detected. Methylation percentage was used for clustering using the Genesis software (Sturn *et al.*, 2002). Gene Ontology analysis was performed with PANTHER (Mi *et al.*, 2021).

Results

Sequencing statistics are reported in Table 1. The average number of reads per sample was 14.8 M (range: 9.2-25.M). About 73.7 and 77.6% cytosines in the CpG enriched regions were methylated in *B. taurus* and *B. indicus*, respectively. A slight difference was observed in the two season with 3 out of 5 A showing a methylation decrease and 3 out of 5 N a methylation increase from challenge to recovery period. A total

Table 1. Mapping and methylation statistics. A=Angus; N=Nellore; C=Challenge; R=Recovery; meth = methylation, SD=Standard Deviation.

Sample	Mapping %	Total Cs	CpGs meth	CpGs not meth	CpGsmeth %	Mean %	SD %
A1_C	43.0	2.71E+08	22936201	8354660	73.3		
A1_R	44.3	3.85E+08	31736484	11383739	73.6		
A2_C	45.2	2.75E+08	23112566	7472972	75.6		
A2_R	46.4	3.04E+08	24891807	8509203	74.5		
A3_C	40.3	3.09E+08	25311972	8655726	74.5	74.1	0.93
A3_R	46.0	3.51E+08	28420620	10233846	73.5		
A4_C	46.6	3.18E+08	26122712	8442877	75.6		
A4_R	45.7	4.5E+08	36141144	13506272	72.8		
A5_C	45.1	3.05E+08	25218653	9002866	73.7		
A5_R	45.6	3E+08	24518583	8679181	73.9		
N1_C	44.0	2.25E+08	18624964	7331023	71.8		
N1_R	42.2	2.76E+08	22917615	8592611	72.7		
N2_C	44.8	3.17E+08	25742207	9334119	73.4		
N2_R	44.8	5.41E+08	44264805	15546828	74.0		
N3_C	45.8	3.08E+08	25440027	8982612	73.9	73.4	0.90
N3_R	43.0	4.34E+08	35860258	13026295	73.4		
N4_C	42.3	2.71E+08	22324516	8335052	72.8		
N4_R	45.7	3.77E+08	31985936	10570562	75.2		
N5_C	42.8	1.88E+08	15624915	5711561	73.2		
N5_R	48.2	2.85E+08	21337132	7823402	73.2		

of 134,553 methylated regions (MRs) were detected across all 10 samples in two conditions. Hierarchical clustering of MRs in Angus and Nellore animals considering the heat-stress treatment and recovery period (C= Challenge period, heat stress and R=Recovery period) is in Figure 1.

About 30.22% of MRs were differentially methylated (DMRs n=51,421) between Angus and Nellore (LR FRD<0.05). Compared to MRs, the number of DMRs was slightly higher in non-coding regions and slightly lower in exons and 5'UTRs. GO analysis of a subset of genes (n=606) showing the higher methylation variation (DMRs with FDR<10^{-40}) showed enrichment in biological processes related to sensory perception of chemical stimulus, development process and development and regulation of cellular organization and signalling.

Discussion

In the current study, blood collected from taurine and indicine cattle subspecies was profiled by RRBS. Blood from Angus and Nellore showed an average level of CpG methylation of about 74%, not very consistent with previous data obtained by RRBS profiling of blood from tropical bovine breeds and their main ancestors, ranging from 51 to 57% (Sevane *et al.*, 2018), probably due to differences in the protocol procedure in size selection during library preparation. Although, the epigenomic comparison of Colombian Creole cattle breeds and their putative Spanish ancestors identified only 334 DMRs by RRBS (Sevane *et al.*, 2018), in our study the CpG methylation comparison between A and N showed a high diversity (DMRs n=51,421), with

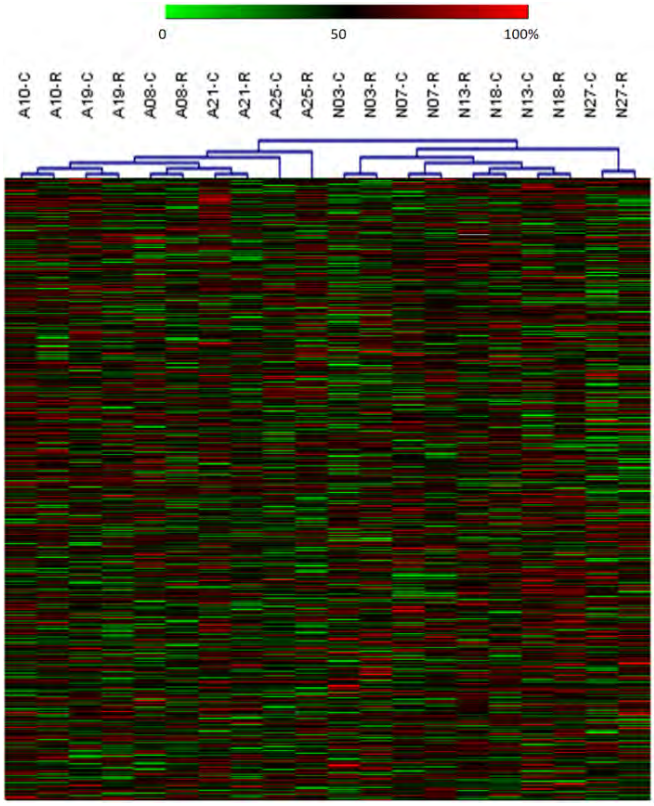


Figure 1. Hierarchical Clustering for 134,553 methylated regions (MRs) in the 20 samples mapped on *Bos taurus*. A=Angus; N=Nellore; C=Challenge; R=Recovery.

about 30% of the total identified MRs that differ between the two subspecies. The inconsistency between these two studies can probably be related to differences in experimental design. Only 2 animals from Costeño con Cuernos and one from San Martinero from Colombian Creole cattle breeds and one animal from each Asturiana de los Valles, Lidia and Retinta breeds from Iberian cattle were compared in Sevane *et al.*'s study, whereas we compared five animals from same indicine and taurine breeds. Important changes in the level of muscle methylation between different breeds was also observed in pigs, probably indicating the result of long-term selection for quantitative traits involving a very high number of genes (Ponsuksili *et al.* 2019). In conclusion, RRBS profiling of A and N animals showed a high epigenetic diversity between two subspecies, certainly not influenced by external environmental condition but related to evolutionary trajectories, that probably affects genes implicated in important pathways for cattle adaptation.

References

Bolormaa S, Pryce JE, Kemper KE, Hayes BJ, Zhang Y, *et al.* (2013). Genet Sel Evol. 45(1):43. https://doi.org/10.1186/1297-9686-45-43

Barbato M, Hailer F, Upadhyay M, Del Corvo M, Colli L, *et al.* (2020) Sci Rep. 10(1):1279. https://doi.org/10.1038/s41598-020-57880-4

Del Corvo M, Lazzari B, Capra E, Zavarez L, Milanesi M, *et al.* (2021) Front Genet. 12:633132. https://doi.org/10.3389/fgene.2021.633132

Ewels PA, Peltzer A, Fillinger S, Patel H, Alneberg J, *et al.* (2020) Nat Biotechnol. 38(3):276-278. https://doi.org/10.1038/s41587-020-0439-x

Hu Y, Xia H, Li M, Xu C, Ye X, *et al.* (2020) BMC Genomics. 21(1):682. https://doi.org/10.1186/s12864-020-07097-6

Ma B, Wilker EH, Willis-Owen SA, Byun HM, Wong KC, *et al.* (2014) Nucleic Acids Res. 42(6):3515-28. https://doi.org/10.1093/nar/gkt1380

Manirakiza J, Hatungumukama G, Thévenon S, Gautier M, Besbes B, *et al.* (2017) Anim Genet. 48(5):544-550. https://doi.org/10.1111/age.12578

Mi H, Ebert D, Muruganujan A, Mills C, Albou LP, et al PD. (2021) Nucleic Acids Res. 49(D1):D394-D403. https://doi.org/10.1093/nar/gkaa1106

Okonechnikov K, Conesa A, García-Alcalde F. (2016) Bioinformatics. 32(2):292-4. https://doi.org/10.1093/bioinformatics/btv566

Pérez O'Brien AM, Mészáros G, Utsunomiya YT, Sonstegard TS, Fernando Garcia J *et al.* (2014) Livest Sci. 2014;166:121. https://doi.org/10.1016/j.livsci.2014.05.007

Ponsuksili S, Trakooljul N, Basavaraj S, Hadlich F, Murani E, *et al.*(2019) BMC Genomics. 20(1):492. https://doi.org/10.1186/s12864-019-5880-1

Rodrigues RT, Chizzotti ML, Vital CE, Baracat-Pereira MC, Barros E, *et al.* (2017) PLoS One. 12(1):e0170294. https://doi.org/10.1371/journal.pone.0170294

Sevane N, Martínez R, Bruford MW. (2019) Anim Genet. 50(1):15-26. https://doi.org/10.1111/age.12731

Sturn A, Quackenbush J, Trajanoski Z. (2002) Bioinformatics. 18(1):207-8. https://doi.org/10.1093/bioinformatics/18.1.207

Utsunomiya YT, Milanesi M, Fortes MRS, Porto-Neto LR, Utsunomiya ATH, *et al.* (2019) Anim Genet. 50(6):557-568. https://doi.org/10.1111/age.12836

Verdugo MP, Mullin VE, Scheu A, Mattiangeli V, Daly KG, *et al.* (2019). Science. 365(6449):173-176. https://doi.org/10.1126/science.aav1002

173. Evaluating the use of genotype information for the prediction of breed composition in purebred and crossbred cattle

C.A. Ryan[1,2*], D.P. Berry[1], T. Pabiou[3] and D.C. Purfield[2]

[1]Teagasc, Moorepark, Fermoy, Co. Cork, Ireland; [2]Munster Technological University, Cork, Co. Cork, Ireland; [3]Irish Cattle Breeding Federation, Bandon, Co. Cork, Ireland; cliona.ryan@teagasc.ie

Abstract

The ability to accurately predict breed composition using genomic information has many potential uses including increasing the accuracy of genetic evaluations, designing mating plans, traceability and genotype quality control. While software exists to estimate breed composition, the ability to exploit existing pipelines used in genomic evaluations for breed prediction could be advantageous. Utilising 49,213 single nucleotide polymorphisms (SNPs), the breed composition of 3,146 purebred and 4,330 crossbred cattle was estimated using both SNP-BLUP and Admixture. The accuracy of low-density panels to predict breed composition was also evaluated. The average mean difference in predicted breed composition between SNP-BLUP and Admixture was 0.000 and 0.048 for purebreds and crossbreds, respectively. Both methods accurately assigned ≥98% of the purebred validation population to the correct breed. SNP-BLUP required 2,000 and 3,000 SNPs to accurately predict purebred and crossbred breed composition, respectively; the respective values for Admixture was 500 and 1000 SNPs.

Introduction

In the absence of genomic data, an animal's breed proportion is assumed to be the average breed composition of both parents (Sölkner *et al.*, 2010). However, due to parental chromosomal recombination during gametogenesis, the breed composition of the offspring from a crossbred parent may differ from the expectation. Genomic information has the potential to more precisely determine the breed composition of animals due to its capacity to establish parental contribution, and therefore can help correct pedigree errors and estimate kinships when pedigree data is missing. While software exists to estimate breed composition (e.g. Admixture; Alexander *et al.*, 2009) the potential to exploit existing pipelines used in genomic evaluations for breed prediction could be more computationally efficient and be utilized for quality control, ensuring no mislabelled genotypes enter the database. The objective, therefore, of this study was to use a large database of genotyped purebred and crossbred cattle to compare breed composition predictions using a freely available software, Admixture, with those from SNP-BLUP. An additional objective was to determine the accuracy and general robustness of low-density genotype panels for breed prediction which was achieved by varying: (1) the SNP density; and (2) the selection strategy of SNP for alternative custom-derived low-density panels.

Materials & methods

Genotypes from 49,213 single nucleotide polymorphisms (SNPs) were available post quality control for 703,078 dairy and beef cattle; sporadically missing genotypes were imputed using FImpute (Sargolzaei *et al.*, 2014). A total of 13 breeds were included in this study, and purebreds from each breed were verified using both a principal component analysis and having a breed composition of >0.9 assigned in an unsupervised Admixture analysis (Alexander *et al.*, 2009). Within breed identity-by-state (IBS) clustering was undertaken on all purebred animals to identify the 500 least similar animals within each breed to represent the training population. A validation population of 3,146 verified purebred and 4,330 two and three-way crossbred animals with no direct relationship to the training population was chosen at random. An additional purebred independent population of 9,955 animals was also created to determine, and

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_173, © C.A. Ryan *et al.* 2022

subsequently rank, SNPs by informativeness for the development low-density panels; these animals were not included in either the training or the validation populations.

The allele substitution effect per SNP for each breed was estimated using SNP-BLUP (MIX99 Development Team, 2017), where all SNPs were fitted as random effects within a linear mixed model:

$$\mathbf{y}_i = \mu + \sum_{i=1}^{n} \mathbf{X}_j \mathbf{g}_{ij} + \mathbf{e}_i \quad (1)$$

where the dependant variable y_i was recorded as either one if animal i had a breed proportion of >0.9 in unsupervised admixture and was in the training population for the breed under investigation, zero if the animal was in the training population but not for the breed under investigation, or -99 if the animal was in the validation population; μ refers to the overall mean, X_j is the allele substitution effect of SNP_j; g_{ij} is the genotype of animal i at locus j and e_i is the residual term for animal i. The phenotypic variance for the dependent variable varied per breed and was estimated as pq, where p was the proportion of animals which were verified to be the breed under investigation (i.e. coded as 1); q was 1 minus this proportion. The resulting allele substitution effects were then summed to generate an estimate of the proportion of each breed in each animal. A supervised analysis in Admixture with 13 breed clusters (K=13) was then performed using the same training and validation animals and their estimated breed composition were compared with estimates from SNP-BLUP.

Development of low-density genotype panels. Seven lower-density panels (i.e. 500, 1000, 2,000, 3,000, 5,000 and 7,500 SNPs) were generated using six different SNP selection strategies. All low-density panels were compared to SNP-BLUP breed estimations based on all 49,213 SNPs, which were considered the gold standard prediction due to the fact that SNP-BLUP was computationally less intensive than Admixture and breed predictions from both software's were very similar. For purebred animals, breed assignment was assumed to have occurred if the prediction of breed composition was ≥0.90 for any single breed. The six alternative methods used to rank SNPs for the generation of the low-density panels were:

1. *Fixation Index* (F_{st})*:* The global F_{st} was estimated across all 13 breeds from the independent population using all 49,213 SNPs.
2. *PCA:* SNP weightings for each principal component were calculated using the 'smartpca' algorithm in Eigensoft v7.2.1 (Patterson *et al.*, 2006) on the independent population. The greater the difference in allele frequency between the breeds the greater the SNP weighting.
3. *SNP-BLUP variances:* SNP effects for breed prediction were estimated from SNP-BLUP using all 49,213 SNPs in the independent population. The standard deviation of these SNP effects were estimated across breeds and this was used to rank the SNPs.
4. *Random Forest (RF):* The RF analysis was conducted in the R package randomForest (Liaw and Wiener, 2002) using the genotypes of the purebred independent population. SNPs were ranked according to their relevance for breed prediction using the Mean Decrease in Gini Index. The higher the value of each SNP's Mean Decrease in Gini score, the more informative the SNP should be for breed prediction.
5. *Partitioning-around-mediods (PAM):* The partitioning-around-mediods (PAM) algorithm clusters SNPs on each chromosome together based on their proximity in genomic position. The number of clusters created per chromosome was set to the number of predefined SNPs for that chromosome. The SNP located in the middle of each cluster was selected.
6. *Random Selection:* The number of predefined SNP required per chromosome was randomly selected until each of the respective panel densities was obtained.

The number of SNP chosen per chromosome remained constant for each selection methods evaluated and was proportional to the genome length of each chromosome. The F_{st}, SNP-BLUP variances and RF methods

all ranked SNPs for the generation of low-density panels, and SNPs in the nth highest ranking for each SNP selection method were chosen per chromosome where n was the number of predefined number of SNPs for that chromosome.

Results

As there is no gold standard for genomic breed prediction, we tested the precision of SNP-BLUP by comparing the breed estimates with breed estimates from the well-established Admixture method. When 49,213 SNPs were included in the analysis, both SNP-BLUP and Admixture accurately assigned ≥98% of the purebred validation population to the correct breed. For crossbreds, the average mean difference in breed proportion predictions using SNP-BLUP and Admixture was 0.060 with a root mean square error of 0.08. For the low-density panels, the mean accuracy of breed composition predictions improved with increasing panel density (Figure1). In general, purebreds and crossbreds required 2,000 and 3,000 SNPs, respectively for accurate prediction of breed composition when using SNP-BLUP. There was a minimal difference in prediction accuracy across the various SNP selection methods when employing SNP-BLUP at panel densities ≥3,000 SNPs, with a mean difference of 0.021 between gold standard predictions and those made at panel densities ≥3,000 SNPs (Figure 1), but at panel densities below this, the F_{st} method was the most accurate at predicting breed composition. For comparison purposes, the SNPs selected using the three most accurate SNP selection methods (i.e. F_{st}, PCA and PAM) were used for breed prediction in Admixture and the predictions of breed composition were compared to those from SNP-BLUP. Using the same low-density panels, Admixture predicted breed composition more accurately than SNP-BLUP, with 500 and 1000 SNPs sufficient to accurately predict the breed composition of purebred and crossbred cattle, respectively (Figure 2).

Discussion

The root mean square error for breed predictions made using Admixture and SNP-BLUP across purebreds and crossbreds when 49,213 SNPs were used was 0.08, indicating a high degree of similarity between the two methods. As suggested by others (Bjørnstad and Røed, 2002), when predicting breed composition, some animals may never be correctly assigned regardless of the number of SNPs used because individuals are

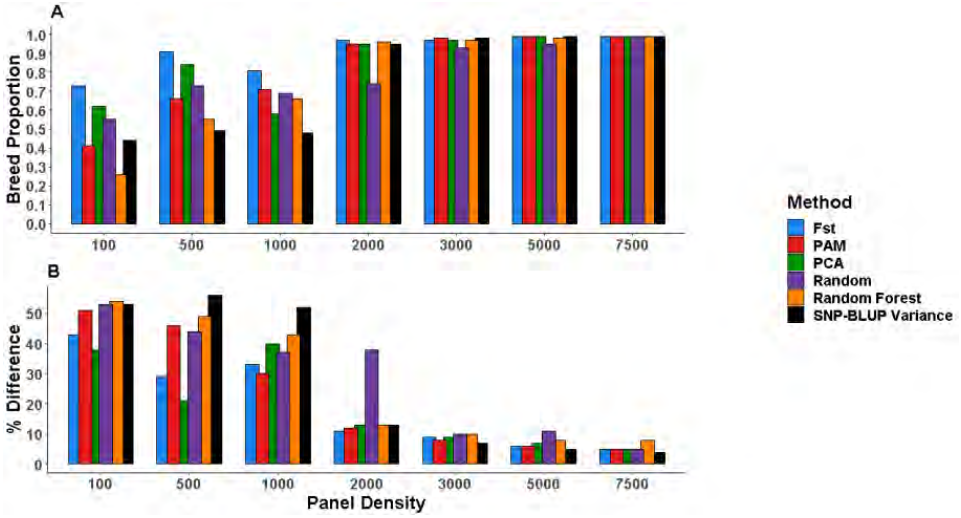


Figure 1. (A) Purebred breed proportion estimates across all low-density panels, (B) The ±% difference between gold standard and low-density breed estimates for crossbreds.

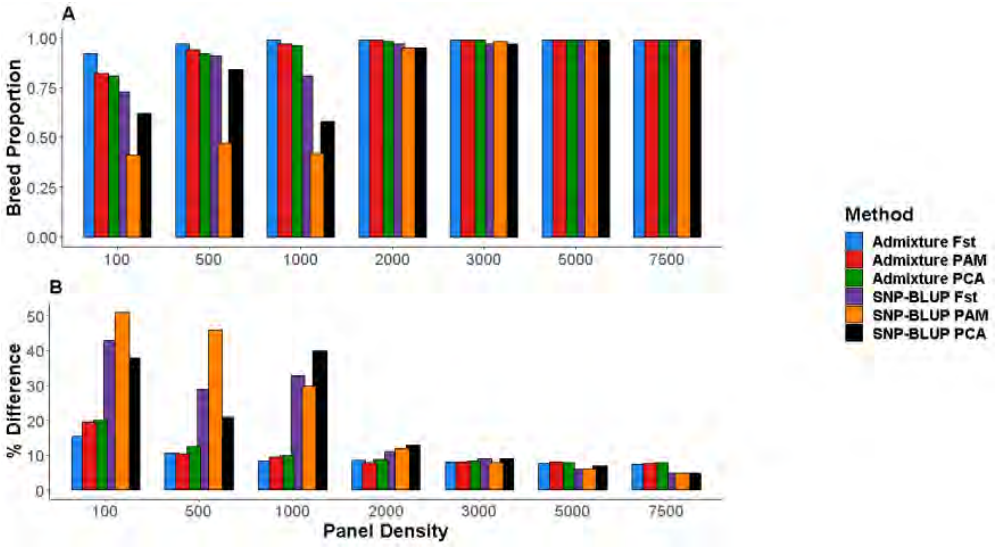


Figure 2. (A) Purebred breed proportions estimates from Admixture and SNP-BLUP, (B) The ±% difference between gold standard and low-density breed proportion estimates from Admixture and SNP-BLUP.

genetically uncharacteristic of their breeds. To avoid the latter from occurring, the training population in the present consisted of a large population of animals, thus increasing the within-breed variability. Purebreds required 2,000 SNPs to accurately predict their breed composition, whereas 3,000 SNPs were required to predict animals in the crossbred validation population, corroborating results of Bjørnstad and Røed (2002) in that predicting the breed composition of crossbreds is more challenging than that for purebreds. Despite choosing SNPs located in close proximity to each other on each autosome, the Fst method proved to be the best SNP selection method for low-density panels. This suggests that informative SNPs for the prediction of breeds may cluster together, which could potentially be reflective of positive selection as result of modern breeding programmes. The strong similarity between SNP-BLUP and Admixture breed predictions when panel density ≥3,000 SNPs suggests that breed prediction could be readily integrated into the SNP-BLUP pipelines used for genomic evaluations, thus replacing stand-alone software.

References

Alexander D. H., Novembre J., and Lange K. (2009) Genome Res *9*(9):1655-64. https://doi.org/10.1101/gr.094052.109.

Bjørnstad G., and Røed K. H. (2002) Animal Genet *33*(4): 264-270. https://doi.org/10.1046/j.1365-2052.2002.00868.x.

Liaw A., and Wiener M. (2002) R News 2(3):18-22. Available at: https://CRAN.R-project.org/doc/Rnews/.

MIX99 Development Team. (2015) Mix99. Available at: https://www.luke.fi/en/expert-services/mix99-solving-large-mixed-model-equations/.

Patterson N., Price A. L., and Reich D. (2006) PLoS Genet *2*(12): e190. https://doi.org/10.1371/journal.pgen.0020190.

Sargolzaei M., Chesnais J. P., and Schenkel F. S. (2014) BMC Genom *15*(1): 478. https://doi.org/10.1186/1471-2164-15-478.

Strucken E. M., Al-Mamun H. A., Esquivelzeta-Rabell C., Gondro C., Mwai O. A., *et al.* (2017) Genet Sel Evol 49(1):67. https://doi.org/10.1186/s12711-017-0342-1.

174. Genomic prediction using breed origin of alleles model accounting for probabilities in the assignment of the alleles

A. Guillenea*, G. Su, M.S. Lund and E. Karaman

Center for Quantitative Genetics and Genomics, Aarhus University, 8830 Tjele, Denmark; ana.guillenea@qgg.au.dk

Abstract

This study investigates genomic prediction using a breed origin of alleles (BOA) model which accounts for BOA with a (1) definitive assignment to a breed; or (2) probabilities of assignments to each breed. Our BOA model estimates breed-specific marker effects based on genotypic and phenotypic information from the purebred and crossbred animals. A traditional combined analysis of all breeds' data implicitly assumes a correlation of one between the marker effects of the breeds, whereas our BOA model allows the estimation of these correlations. We used de-regresed proofs of production traits from the admixed Nordic Red Cattle population to evaluate the model and performed the analysis assuming marker effects between breeds are correlated or assuming they are uncorrelated. We found that using probabilities in the BOA model outperformed the BOA model, assuming the origin of each marker is known with certainty, especially when the marker effects were assumed uncorrelated between breeds.

Introduction

The implementation of genomic prediction (GP) is challenging in admixed populations since the same allele of a given marker might have a different effect on the phenotype according to its breed of origin. Accuracy of GP relies on the linkage disequilibrium (LD) between single nucleotide polymorphisms (SNP) and quantitative trait loci (QTL). In multi-breed reference populations, accuracy of GP relies on the LD phase between SNP and QTL being conserved across breeds, but LD may be different, or phase may be even reversed across breeds. QTL effects might also vary between breeds if there are interactions between QTL and genetic background (Calus *et al.*, 2018). Several methods have been proposed recently to account for breed origin of alleles (BOA) in GP, and several methodologies/software have been developed to infer BOA (Lawson *et al.*, 2012; Vandenplas *et al.*, 2016; Eiriksson *et al.* 2021). For GP, Karaman *et al.* (2021) developed a method that allows to include purebred and crossbred animals, estimating SNP effects considering correlation of SNP effects between the breeds. This method includes breed-specific matrices and thus, relies on the correct assignment of the alleles. Some of the existing software to infer BOA provide the probabilities of belonging to each breed for each allele, and others provide the probabilities of belonging to each breed for alleles that could not be assigned to a specific breed. Including these probabilities in the models, instead of assigning an allele to one particular breed, could improve reliabilities because they may account for some uncertainty in the assignments, especially for highly admixture populations. The objective of this study was to compare GP for a BOA model using either assignment of an allele to a specific breed or probabilities of an allele from the possible breeds, based on data from the Nordic Red dairy cattle (RDC) population, which is an admixture of Danish Red (RDM), Swedish Red (SRB) and Finnish Ayrshire (FAY) dairy cattle, which have strong ties due to the use of common bulls.

Materials & methods

Genotypic and phenotypic data. We used de-regressed proofs (DRP) of 51,336 animals (5,104 bulls and 46,232 cows) of the RDC population for milk, fat and protein as response variables to predict direct genomic breeding values (DGV). The Nordic Cattle Genetic Evaluation provided the data. The bulls in the data set were genotyped using the Illumina Bovine SNP50 chip (50K). For cows, 23,245 were genotyped with the standard Illumina Bovine low-density Chip and 22,987 with the EuroGenomics customized chip,

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_174, © A. Guillenea *et al.* 2022

which, in addition to the previous, includes SNPs selected from imputed whole-genome sequencing data that were not used in this study. In addition, 3,000 HOL bulls were included as an ancestral population as a preliminary pedigree analysis showing ties of HOL with RDM. The genotypes of all animals (RDC and HOL) were imputed to 50K and phased. Finally, 42,470 SNPs common to RDC and HOL were used in this study.

Estimation of breed origin. First, we used all genotyped RDC bulls to perform an ancestry analysis to identify purebred animals, regardless of the presence of DRP. In total, 8,356 bulls were included in this analysis (5,356 RDC and 3,000 HOL). The bulls were allocated to purebred and admixed groups based on the breed proportions predicted using ADMIXTURE v1.30 software with k=4 (k = ancestral populations). In total, in this step 1,820 bulls were selected as purebred (751 HOL, 244 RDM, 760 FAY and 65 SRB). Second, BOA was inferred for the rest of the RDC animals with ChromoPainterV2 software (Lawson *et al.*, 2012) using the haplotypes of purebred animals of each breed, selected in the ADMIXTURE analysis, as donor haplotypes of their corresponding breed. The expected probability of each SNP coming from each of the purebred populations was used to infer the BOA in two ways: (1) each allele was assigned to one purebred population for which the expected probability was the highest, or (2) each allele was assigned to each population as a product of the allele count (0, 1 or 2) and the expected probabilities. Breed proportions were inferred by summing the alleles from each breed.

Statistical model. The BOA model to estimate breed-specific SNP effects is as follows (Karaman *et al.*, 2021):

$$y = 1\mu + Xb + Z_D u_D + Z_F u_F + Z_S u_S + Z_H u_H + e$$

where y is the vector of DRP of milk, fat or protein for the reference animals, 1 is the vector of 1s, μ is the overall mean, X is the matrix of breed proportions, b is the vector of fixed breed effects, and e is the vector of random residuals. Z_D, Z_F, Z_S, and Z_H are the matrices of breed-specific content of SNP for RDM, FAY, SRB, and HOL, u_D, u_F, u_S and u_H are vectors of SNP effects for RDM, FAY, SRB, and HOL, respectively.

Two approaches to construct Z matrices were used. For the first approach, the Z matrices were formed by assigning each allele to a certain purebred population for which the expected probability is the highest (BOAg). The entry at a locus in an Z matrix, for instance, Z_D, were the number of '*A*' alleles (0, 1 or 2) originating from RDM for an animal. Consequently, if an animal had '*aa*' genotype or had no allele originating from RDM, the corresponding entry in Z_D was zero. The same applies to other Z matrices for each breed. For the alternative approach, each allele was assigned to all original populations according to expected probabilities (BOAp). Thus, the Z matrices were achieved by taking the sum of the product of the allele count (0 or 1) and the expected probabilities over the two alleles at each locus. The Z matrices were centered by subtracting the mean of the column prior to analysis.

For the analysis assuming breed-specific SNP effects uncorrelated across the breeds (Uncor), each vector of SNPs was assigned a prior of a normal distribution separately for each breed (i = D,F,S,H): $u_i|\sigma^2_i \sim N(0, \sigma^2_i)$. Here, σ^2_i is the SNP variance for breed i, obtained from a preliminary analysis using only bulls' data. For the analysis assuming correlated breed-specific SNP effects (Cor), a multivariate normal distribution was assigned for the vectors of SNP effects: $[u'_D, u'_F, u'_S, u'_H]' \mid B \sim N(0, B \otimes I)$, where I is an identity matrix, and B is as in (Guillenea *et al.*, 2022). Briefly, the diagonals are the breed-specific SNP variances and off-diagonals of B are covariances. Random residuals were assumed to follow a normal distribution, $e \sim N(0, D\sigma^2_e)$, where σ^2_e is the residual variance, and D is a diagonal matrix with elements $d_{ii} = 1/w$, w is the weighting factor for the ith DRP. More details of priors are given in Guillenea *et al.* (2022). Fixed effects were assigned to

having flat priors. The analyses were written and carried out in Julia programming language (Bezanson *et al.*, 2017).

Model validation. The validation consisted of 11,786 cows, while reference consisted of 5,104 bulls and 34,446 cows. Both populations were made based on groups of half-sibs with a cut-off birth date July 1, 2015. The DGVs of validation animals were calculated by multiplying the breed-specific matrices with breed-specific SNP effects and adding a fixed breed contribution. The prediction reliability was calculated as the squared correlation between DGV and DRP divided by the reliability of DRP for the validation animals. We used a bootstrap procedure sampling 10,000 times to calculate reliabilities and the standard errors of each model.

Results

Reliabilities of GP are presented in Figure 1. Using probabilities of assignments improves prediction reliabilities for the three traits in Uncor, but not in Cor. On average, the increase of BOAp over BOAg was 1% for the correlated analysis and 33% for the uncorrelated analysis. Regarding the production traits under study, milk presented the highest reliabilities and the patterns over the traits were similar.

Discussion

The performance of models based on the breed origin of the alleles depends, among other factors, on the precise estimation of the origin of each allele, which is used to construct the breed-specific allele matrices. The RDC is a highly admixed breed, which made it difficult to define purebred animals of the three Nordic original breeds, and consequently, the assignments of the alleles in ChromoPainterV2 showed a distribution of probabilities, in which only 40% were assigned to their breed origin with probability >0.9. This means that in our first approach (BOAg) if one allele had a probability of, for instance, 30% of coming from RDM and the probabilities for the other breeds three were lower, the allele was assigned to RDM. In the second approach (BOAp), the same allele was assigned to all breeds with the expected probabilities. The

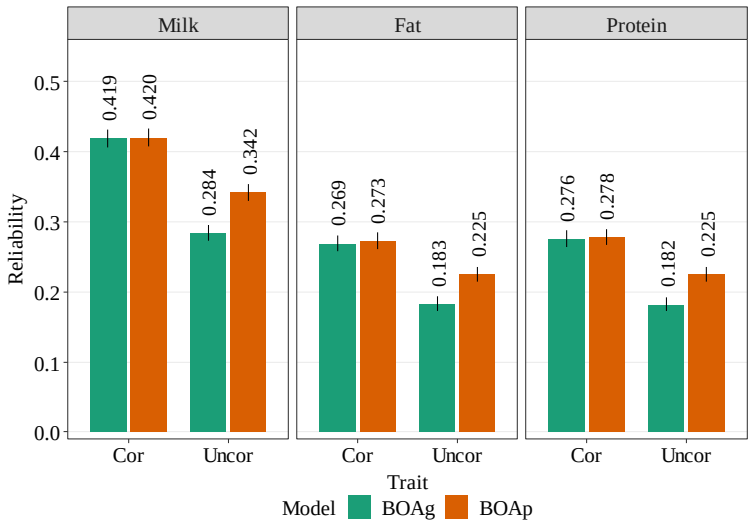


Figure 1. Reliabilities and standard errors of BOA with definitive assignment to a breed (BOAg) (Guillenea *et al.*, 2022) or probabilities of assignments to possible breeds (BOAp) for a scenario of correlated (Cor) or a scenario of uncorrelated (Uncor) SNP effects between breeds.

uncorrelated analysis obtains more gains for using the probabilities because it depends more on intra-breed information. In this population, in which breeds were found to be highly related, the impact of accounting for probabilities was small for the correlated analysis. However, it is expected that the improvement of predictions by using probabilities would be more in a correlated analysis for the breeds which are more distantly related. For such breeds the use of BOA models would also become more important.

Conclusions

This study compared GP of BOA with two types of breed-specific genotypic matrices: (1) each allele was assigned to only one specific breed; and (2) probabilities in the assignments were taken into account in the matrices. We conclude that the BOA model benefits from the use of probabilities in the construction of the breed-specific matrices, so it is a more precise approach to assign alleles that can be implemented in models with breed-specific matrices.

Acknowledgements

We thank GentTORE and ANII Uruguay (POS_EXT_2018_1_154296) for financial support.

References

Bezanson J., Edelman A., Karpinski S., and Shah. V. (2017) SIAM rev. 59(1):65-98. https://doi.org/10.1137/141000671

Calus M., Goddard M., Wientjes Y., Bowman P., and Hayes. B. (2018) J Dairy Sci 101(5):4279-4294. https://doi.org/10.3168/jds.2017-13366

Eiríksson J.H., Karaman E., Su G., and Christensen O.F. (2021). GSE 53(1), 1-13. https://doi.org/10.1186/s12711-021-00678-3

Guillenea A., Su G., Lund M.S., Karaman E. (2022). J. Dairy Sci. (in press)

Karaman E., Su G., Croue I., and Lund M.S. (2021) GSE 53(1), 1-13. https://doi.org/10.1186/s12711-021-00637-y

Lawson D.J., Hellenthal G., Myers S., and Falush D. (2012). PLoS Genet 8(1):e1002453. https://doi.org/10.1371/journal.pgen.1002453

Vandenplas J., Calus M.P., Sevillano C.A., Windig J.J. and Bastiaansen J.W. (2016). GSE 48(1):61. https://doi.org/10.1186/s12711-017-0350-1

175. Making the most of genetic relationships between Iberian dairy sheep breeds: exploring multi-breed genomic evaluation

I. Granado-Tajada[1*], C. Pineda-Quiroga[1], E. Ugarte[1], M. Serrano[2], M.A. Jiménez[2], H. Marina[3], J.J. Arranz[3] and M. Ramón[4]

[1]Animal Production Department, Neiker-BRTA, Campus Agroalimentario de Arkaute, 01080 Vitoria-Gasteiz, Spain; [2]Instituto Nacional de Investigación y Tecnología Agraria y Alimentaria CSIC-INIA, Ctra. de la Coruña km 7.5, 28040 Madrid, Spain; [3]Universidad de León, Campus de Vegazana s/n, 24071 León, Spain; [4]Instituto Regional de Investigación y Desarrollo Agroalimentario y Forestal de Castilla-La Mancha (IRIAF), Av. del Vino s/n, 13300 Valdepeñas, Spain; igranado@neiker.eus

Abstract

Combining data from different breeds into a genetic evaluation can have several benefits, especially when there is a close evolutionary relationship as in Iberian dairy sheep breeds. This work explores the benefits on prediction reliability of combining genomic and phenotypic data into a multi-breed genomic evaluation for milk yield in four Iberian dairy sheep breeds: Assaf, Churra, Latxa, and Manchega. The phenotype was the estimated breeding value by classic BLUP evaluation and together with genomic information, single- and multi-breed evaluations were run. No gain in prediction reliability of multi-breed genomic evaluations was found, probably related to differences between linkage disequilibrium patterns and genetic differentiation between concerned breeds. The relationship among these breeds is probably not close enough to increase prediction reliability into a multi-breed genomic evaluation. Additional studies need to be done to confirm that.

Introduction

Multi-breed genetic evaluations can improve prediction reliability due to a bigger reference population (RP) size that increases the accuracy of determining the relationship between phenotypes and SNP markers (Lund *et al.*, 2014). Dairy sheep breeds are characterized by the presence of a wide range of breeds, with an adapted production system, small population size, heterogeneous data recording system and lower economic value per individual, that difficult a join evaluation, but at the same time makes it more interesting. The relationship among Iberian dairy sheep breeds has been previously studied based on morphological characters and molecular data (microsatellites and mitochondrial DNA), and although genetic differentiation exists, a close relationship has been described (Jordana and Ribo, 1991; Pedrosa *et al.*, 2007; Legaz *et al.*, 2008). Currently, the implementation of genomic schemes into these breeds is ongoing work at different stages, while farmers' attitude towards genomic breeding tools seems to be positive (Martin-Collado *et al.*, 2021). However, the size and composition of their RP are highly varying, which makes comparison difficult between breeds. This situation provided the stimulus for our work to explore the impact on prediction reliability of joining phenotypic and genomic data of four Iberian dairy sheep breeds (Assaf, Churra, Latxa, and Manchega) into a multi-breed genomic evaluation for the main breeding objective, milk yield.

Materials & methods

Phenotypic, genealogic and genomic data. The concerned Assaf, Churra, Latxa, and Manchega breeds, have considered 120 days standardized milk yield records as initial phenotype; the number of records and animals in the pedigree is highly different across breeds, as well as the size and sex composition of the genotyped population, as shown in Table 1. The genotyping was carried out using different platforms: Affymetrix 50K, Custom Affymetrix 50K ULE, AxiomTM Ovine Genotyping Array (Thermo Fisher Scientific Inc., Waltham, MA, USA), and Illumina Ovine SNP50 Bead Chip (Illumina Inc., San Diego, CA,

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_175, © I. Granado-Tajada *et al.* 2022

Table 1. Description of the phenotypic, pedigree and genomic data included in the study.

Breed[1]	No. of records	No. of animals in the pedigree	No. of genotyped individuals			No. of common markers
			AI males[2]	NSV males[2]	Females	
Assaf	899,155	348,040	561	3,285	3,109	21,787
Churra	671,796	246,269	254	-	3,120	
LCNEUS	664,878	272,049	446	195	321	
LCR	431,692	163,737	476	254	265	
LCNNAF	197,081	73,033	211	162	77	
Manchega	6,853,108[3]	531,464	530	2,975	1,599	

[1] Latxa breed is divided in three subpopulations: Latxa Cara Negra from Euskadi (LCNEUS), Latxa Cara Rubia (LCR), and Latxa Cara Negra from Navarre (LCNNAF).

[2] Artificial insemination (AI) and natural service (NSV) males.

[3] Milk daily controls.

USA); that brings between 49,702 and 63,679 markers, and after quality controls the common markersacross breeds were included (21,787 SNP).

Based on the available genomic data, to explore the existing relationship across breeds, we applied principal component analysis (PCA) on the VanRaden (2008) matrix of genomic relationships, which was computed as $G = ZZ' / 2\Sigma p_i q_i$, where Z is a matrix of centered genotypes and $p_i = 1 - q_i$ is the allelic frequency for the i^{th} marker based on observed genotypes. Moreover, we studied the conservation of linkage disequilibrium (LD) patterns across these sheep breeds, estimated as the set of correlations of genotypes across all possible pairs of SNP loci (Hill and Robertson, 1968), considering a ranges of distance of 0.05, 0.5 and 5 Mb.

Single breed estimates. Classic BLUP evaluations considered the phenotypic information and routinely used models to obtain estimated breeding values (EBV) of genotyped individuals. No information was available to estimate de-regressed proofs, so EBV provided by breeders associations were used as pseudo-phenotypes despite their limitations. These EBV were used as phenotypes for genomic BLUP evaluations (GBLUP; Habier *et al.*, 2007; VanRaden, 2008) to obtain single-breed genomic estimated breeding values (GEBV). For all breeds Blupf90 family programs (Misztal *et al.*, 2002) were used.

Multi breed pooled estimates. As phenotype, the estimated classic EBV was used. It was combined with the genomic information of concerned breeds into a multi-breed GBLUP evaluation (multiGBLUP), where the breed was considered as a fixed effect. Variance component estimates and evaluations in all cases were run with Blupf90 family programs (Misztal *et al.*, 2002). We estimated correlations among GEBV obtained from different analysis and we also compared mean reliabilities from each method calculated as $\sqrt{1 - \frac{PEV}{\sigma_u^2}}$, being PEV the prediction error variance and σ_u^2 the genetic variance.

Results

The projection of each individual on the three main principal components is shown in Figure 1. The two first principal components (PC) explained 38 and 14% of the covariance across individuals, while the PC3 explained 10%. We can see that the PC1 distinguished Churra, Latxa and Manchega populations from the Assaf population, whereas the PC2 distinguished Assaf and Churra from Manchega and Latxa and PC3 distinguished Latxa subpopulations.

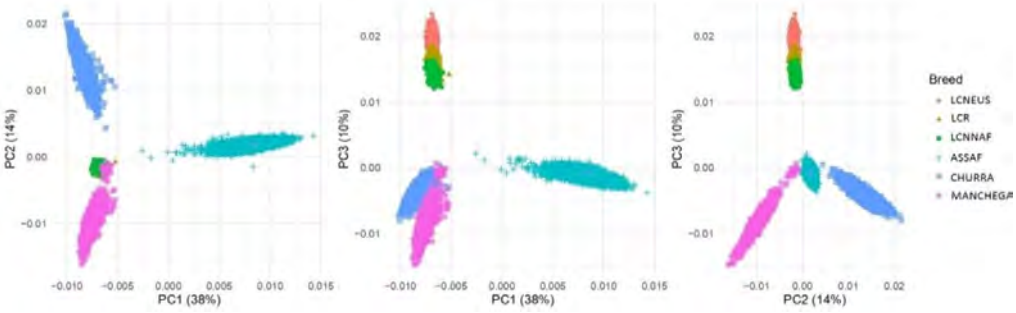


Figure 1. Principal component analysis based on common SNPs across breeds.

The linkage disequilibrium (LD) patterns varied across breeds. At 0.05 Mb the mean LD value was similar for all the studied breeds (~0.16), but when the distance increased, there was a significant drop in mean LD in Assaf, Churra and Manchega breeds (0.06 and 0.03, at 0.5 Mb and 5 Mb) but not in Latxa populations (0.15 and 0.13, at 0.5 Mb and 5 Mb). Regarding the correlation between breeding values estimated by classic BLUP, GBLUP and multiGBLUP, all concerned breeds showed a higher correlation between BLUP and GBLUP estimates (~0.98) than between BLUP-MultiGBLUP (~0.92) or between GBLUP-MultiGBLUP (~0.91) based estimates. Finally, mean reliability values for each genetic prediction method are shown in Table 2. Reliabilities increased to almost one in all breeds with genomic predictions (GBLUP). When the information was combined into a multi-breed evaluation (multiGBLUP), Assaf, Churra and Manchega breeds had higher reliabilities than BLUP evaluations. However, they were lower than GBLUP evaluations in all cases. Latxa populations showed similar reliabilities in BLUP and multiGBLUP evaluations.

Discussion

The lack of gain in prediction reliability of multi-breed genomic evaluations could be related to using EBVs as phenotypes, which were estimated differently in each of the breeds. Even though EBVs represent the ability of animals to transmit their milk yield ability, the variability across breeds is highly different, and the additive variances result in accuracy loss. Possibly using pseudo-phenotypes as daughter yield deviations (VanRaden and Wiggans, 1991) or de-regressed proofs (Garrick *et al.*, 2009) appropriately weighted would be more suitable. Moreover, the different number of genotyped animals across breeds could have affected results, obtaining higher prediction reliabilities of multiGBLUP evaluations within-breed with more genotypes (Assaf, Churra, and Manchega) and being less gain in Latxa populations. Also,

Table 2. Mean reliability for the three evaluation methods: classic BLUP, single breed genomic BLUP (GBLUP) and multi breed genomic BLUP (MultiGBLUP).

Breed[1]	BLUP	GBLUP	MultiGBLUP
Assaf	0.673	0.995	0.809
Churra	0.363	0.999	0.795
LCNEUS	0.738	1.000	0.729
LCR	0.732	0.995	0.724
LCNNAF	0.735	0.988	0.718
Manchega	0.510	0.952	0.761

[1] Latxa Cara Negra from Euskadi (LCNEUS), Latxa Cara Rubia (LCR), and Latxa Cara Negra from Navarre (LCNNAF).

the number of common markers fulfilling all requirements in all breeds could be small to pick up the LD that goes across breeds and increases reliabilities. The PCA has shown that the concerned breeds are genetically differentiated and display differences between LD patterns. Therefore, the relationship among these breeds is probably not close enough to increase prediction reliability into a multi-breed genomic evaluation. Similar results have been previously reported in related sheep breeds (Legarra *et al.*, 2014), highlighting the necessity of a closer genetic relationship to take advantage of a common RP (Lund *et al.*, 2014). The exception could be Latxa populations; given the closer relatedness, it might be worth analysing if combining only them improves prediction accuracy.

In conclusion, we did not find the expected gain in prediction reliability from a multi-breed genomic evaluation. This might be related to the genetic differences between breeds, LD differences and the use of EBVs as phenotypes. Moreover, breeding programs have varying breeding criteria and follow different strategies, making a multi-breed evaluation less suitable. Further studies are needed to test the use of more suitable phenotypes to take advantage of a joint RP of Iberian dairy sheep breeds.

Acknowledgements

The authors thank the breeders' associations for providing phenotypic and genomic data. This work would not have been possible without the funding of the Spanish Ministry of Agriculture, Fisheries and Food by the proposed ENCOMIENDA DE GESTION (2017-2020) and the support of the project ARDI (EFA208/16), funded by INTERREG POCTEFA with the European Fund for Regional Development (FEDER).

References

Garrick, D. J., J.F. Taylor, R. L. Fernando. 2009. Genet. Sel. Evol. 41:55. https://doi.org/10.1186/1297-9686-41-55

Habier, D., R.L. Fernando, and J.C. M. Dekkers. 2007. 177(4):2389–2397. https://doi.org/10.1534/genetics.107.081190

Hill, W.G., and A. Robertson. 1968. Theor. Appl. Genet. 38:226–231

Jordana, J., and O. Ribo. 1991. Investig. Agrar., Prod. Sanid. Anim. 6(3):225–237.

Legarra, A., G. Baloche, F. Barillet, J.M. Astruc, C. Soulas, *et al.* 2014. J. Dairy Sci. 97:3200–3212. https://doi.org/10.3168/jds.2013-7745

Legaz, E., I. Álvarez, L. J. Royo, I. Fernández, J.P. Gutiérrez, *et al.* 2008. Small Rumin. Res. 80:39–44. https://doi.org/10.1016/j.smallrumres.2008.09.001

Lund, M.S., G. Su, L. Janss, B. Guldbrandtsen, and R.F. Brøndum. Livest. Sci. https://doi.org/10.1016/j.livsci.2014.05.008

Martin-Collado, D., C. Díaz, G. Benito-Ruiz, D. Ondé, A. Rubio, *et al.* 2021. Animal. 15(2): 100062. https://doi.org/10.1016/j.animal.2020.100062

Misztal, I., S. Tsuruta, T. Strabel, B. Auvray, T. Druet, *et al.* 2002. Proc. of 7th World Congress on Genetics Applied to Livestock Production. Montpellier. France.

Pedrosa, S., J. J. Arranz, N. Brito, A. Molina, F. SanPrimitivo, *et al.* 2007. Genet. Sel. Evol. 39:91–103. https://doi.org/10.1051/gse:2006034

VanRaden, P. M., and G. R. Wiggans. 1991. J. Dairy Sci. 74:2737–2746. https://doi.org/10.3168/jds.S0022-0302(91)78453-1

VanRaden, P. M. 2008. J. Dairy Sci. 91(11):4414–4423. https://doi.org/10.3168/jds.2007-0980

176. Online Mendelian Inheritance in Animals (OMIA) – future proofing of a globally used animal genetics knowledgebase

I. Tammen[1*], M. Mather[2], Z. Li[2], J. Nothman[2], D.P. Vanichkina[2] and F.W. Nicholas[3]

[1]The University of Sydney, Faculty of Science, Sydney School of Veterinary Science, PMB 4003, Narellan, NSW 2567, Australia; [2]The University of Sydney, Sydney Informatics Hub, Darlington, NSW 2008, Australia; [3]The University of Sydney, Faculty of Science, Sydney School of Veterinary Science, Sydney, NSW 2006, Australia; imke.tammen@sydney.edu.au

Abstract

Online Mendelian Inheritance in Animals (OMIA) is a freely available curated knowledgebase that contains information and facilitates research on inherited traits and disorders across more than 300 species. It is used by animal geneticists, breeders and veterinarians worldwide as a definitive source of information on single-locus traits (especially disorders) in all relevant animal species. Increase in curation capacity and funding for software engineering support have resulted in software upgrades and commencement of new initiatives. We provide here an overview of information contained in OMIA, routine curation activities and key user statistics. A review and update of hyperlinks between OMIA and Online Mendelian Inheritance in Man (OMIM) is reported. Ongoing initiatives including updates of likely causal variant table information, submission of variant information to the European Variation Archive (EVA), introduction of Mammalian Phenotype Ontology (MPO) categories and planned introduction of breed and phene ontologies into OMIA are briefly discussed.

Introduction

OMIA (Online Mendelian Inheritance in Animals 2022, https://omia.org) is a freely available, curated, online knowledgebase which provides users with up-to-date summary information on the known harmful and beneficial variants in animals, together with background information on known inherited disorders and beneficial traits. OMIA is modelled on and reciprocally hyperlinked to OMIM (https://www.omim.org), and provides links to the National Center for Biotechnology Information (NCBI, https://www.ncbi.nlm.nih.gov/), and the European Bioinformatics Institute's Ensembl (http://asia.ensembl.org/index.html).

To facilitate across species analysis, OMIA has a hierarchical structure of traits ('phenes') in one or more species. This structure is reflected in the OMIA ID, which has the form OMIAxxxxxx-yy…yy, where xxxxxx is the 6-digit phene ID and yy…yy is NCBI's species taxonomy ID. Thus an OMIA ID signifies a phene-species, i.e. a particular phene in a particular species. Phenes that have been reported in multiple species, often using different synonyms for the trait or disease, can thus be easily explored using comparative biology approaches.

OMIA focuses on phenes with confirmed and suspected Mendelian modes of inheritance. However, several phenes with unknown or complex modes of inheritance and phenes caused by somatic mutations, genetic modifications or genome editing are also included. Furthermore, OMIA provides references for landmark and review articles and for papers describing genetic maps and reference genomes in animals. While the vast majority of OMIA entries are for the major domesticated animal species, entries for all animal species, except for humans and common model organisms such as mice, rats, and zebrafish, are included.

Since the beginning of 2021, the curation team has increased from one to two main curators and bequest funding has enabled software engineering support. Ongoing curation work, software updates, and several new initiatives are discussed below.

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_176, © I. Tammen *et al.* 2022

Materials & methods

Routine curation activities. A daily automated PubMed (https://pubmed.ncbi.nlm.nih.gov/) search including key genetics terms, all publications of key animal genetics journals and of key authors with a track record in the field is conducted. Identified papers are reviewed based on title and abstract information, and selected papers are added to OMIA via PMID upload as references to existing or new phene-species, and/or added as landmark, mapping or review papers. If relevant, text fields in phene-species entries are updated, hyperlinks to NCBI genes are added, OMIM phene and/or gene hyperlinks are added or updated and likely causal variant information is added to variant tables using Human Genome Variation Society (HGVS) nomenclature (http://varnomen.hgvs.org/).

OMIA software upgrade. Since August 2010, the OMIA database and website have been using Django software, a high-level Python Web framework. In July 2021, software packages were upgraded from Python 2 to 3 (https://www.python.org/) and from Django 1.9 to 3.2 (https://www.djangoproject.com/), and coding was refined with the aim to future-proof OMIA and improve homepage response times. Several additional changes were implemented: new site administration tools were introduced to facilitate 'liftover' of variant information to newer reference genome assemblies, and to enable export of variant information for submission to EVA (https://www.ebi.ac.uk/eva/). A new literature search tool was trialled; separate fields for 'OMIM phene' and 'OMIM gene' hyperlinks were included; and 'Variant ID' and 'Source of Genetic Variant' fields were added to provide unique numerical variant identification and to facilitate inclusion of variant information for genome-modified or edited variants.

Review of OMIA-OMIM hyperlinks. Since 1997, OMIA has been reciprocally hyperlinked to OMIM. OMIM links are identified by OMIA curators when new phenes are entered into OMIA. OMIM downloads OMIA phene IDs with an OMIM ID link once a week and updates OMIM accordingly. We reviewed OMIA-OMIM hyperlinks for all OMIA phenes for which a likely causal variant has been identified. OMIM links were confirmed, deleted, or added.

Results

At the beginning of 2022, OMIA included information on 2,195 phenes across 319 species, contained 4,030 phene-species entries and included a total of 28,184 references. The references include 90 landmark publications, 816 review publications and 428 reference genome or mapping publications. Core statistics for key livestock species are summarised in Table 1.

On average, each phene has reports in 1.9 species, ranging from 1 to 47 species per phene. Of the 33 phenes with reports in 10 or more species,10 phenes are pigmentation traits, which are easy to phenotype. Susceptibility or resistance to spongiform encephalopathy, a zoonotic disease, is the phene reported in the greatest number of species (n=47).

Table 1. Summary of OMIA information relating to key livestock species (9/1/2022).

	Cattle	Pig	Sheep	Horse	Chicken	Goat	All
Total phenes	583	329	291	251	232	99	4,030
Mendelian phenes	272	118	114	61	134	23	1,680
Phenes with at least one known likely causal variant	175	57	48	47	56	17	1,012
Likely causal variants known	235	65	80	101	70	30	1,368
Likely causal variants with EVA ID[1]	73 (134)	18 (20)	26 (41)	16 (50)	15 (12)	8 (10)	197 (506)

[1] Number of variants submitted to EVA in October 2021 for assignment of EVA IDs shown in brackets.

Google analytics (https://analytics.google.com) data for 2021 identified 36,786 users from 161 countries. The majority of users were located in the USA (20%), followed by Russia (5%), Spain (5%), Brazil (5%), Mexico (4%), Australia (4%), Germany (3%) and the UK (3%).

During 2021 the daily automated PubMed literature search resulted in 16,081 hits, of which 778 papers were identified to be added to OMIA (average daily hits =44, and average daily additions = 2.1). Additional references were added as part of other curation activities. We trialled other literature search strategies, including use of the machine learning tool LitSuggest (Allot *et al.* 2021), but further work is required to optimise the search strategy.

A performance review after completion of the software upgrades and refinements to coding confirmed vast improvements to search and homepage response times.

The review of the OMIA-OMIM hyperlinks for phenes in OMIA that had at least one causal variant identified, resulted in confirmation of 607 OMIM links, addition of 683 OMIM links and deletion of 46 OMIM links. Most of the added OMIM links were OMIM gene IDs (n=493), as these were in the past not routinely added to OMIA. OMIA currently lists 2,283 models of human traits. 1,369 OMIM phenes and 645 OMIM genes (14.3 and 3.9% of all OMIM phene and gene entries, respectively) have a link to OMIA.

Discussion

OMIA is a globally used knowledgebase. Until recently, curation was predominately conducted single-handedly by one curator, and limited funding restricted access to urgently needed software upgrades and modifications. Increased curation capacity and bequest funding support to upgrade and refine the underlying software is improving curator and user experiences and has resulted in the commencement of several new innovations while maintaining ongoing curation activities.

We reported the introduction of variant tables to OMIA (Tammen and Nicholas 2018) and indicated that the aim would be to provide EVA IDs to provide standardised nomenclature for all variants. However, EVA (https://www.ebi.ac.uk/eva/), the global resource for animal variant information since NCBI's dbSNP phased out support for non-human animals, does not accommodate all types of variants and only very few authors of OMIA-relevant papers submit variant information to EVA. We therefore have started to review and standardise all historic variant information in OMIA using HGVS nomenclature, with the aim to report location information based on a recent reference genome assembly (Shields *et al.* 2021; Tammen *et al.* 2021a). Software upgrades to site administration tools have further facilitated this process by inclusion of an integrated 'liftover' pipeline. In October 2021, variants listed in OMIA that were lacking an EVA ID but had standardised location information were submitted to EVA using a new OMIA pipeline for export of variant information to EVA. EVA RS IDs for these variants are expected to be released in 2022 and will further facilitate integration between OMIA and other online resources. Variant tables and variant information in VCF format can be downloaded from OMIA, providing valuable resources for researchers and genotyping service providers and enable use of OMIA data as a reference when interpreting high-throughput genomic studies. The need for more standardised nomenclature for variants has been recently discussed to ensure greater transparency in relation to DNA testing. The constantly rising number of causal variants, the increase in genes which are reported to have multiple causal variants, the fact that variants that are initially reported as breed-specific are later found across several breeds and in cross bred animals, as well as use of different nomenclature to describe a variant, can create challenges in how to identify which variants are included in various diagnostic panels. To this end, OMIA numerical variant IDs are now presented in the first column of all OMIA variant tables. An OMIA variant ID provides a unique unchanging ID for each likely causal variant, including those complex variants for which there is no EVA ID.

The revision of OMIA-OMIM hyperlinks will facilitate comparative-medicine-related research approaches. However, a large list of OMIA phenes without known likely causal variants have not been reviewed so far, as it is more speculative to identify homology between human and animal phenes if the underlying genetic cause is unknown. Similarly, due to lack of standardised nomenclature, newly reported animal diseases or traits, for which the underlying genetic cause is unknown, can sometimes be difficult to allocate to a specific phene in OMIA.

Two major current projects relate to the introduction of breed and phene ontologies to OMIA. This work is being conducted in collaboration with colleagues from the Monarch initiative (https://monarchinitiative.org/), as previously outlined (Tammen *et al.* 2021b). The vertebrate breed ontology (VBO) will combine breed information from various resources using the Domestic Animal Diversity Information System (https://www.fao.org/dad-is/en/) as a starting point, with the aim of harmonizing breed definitions (while allowing for the use of country or language specific synonyms) in a computer-accessible format. VBO will improve the currently limited options to analyse breed predispositions in OMIA, and will be publicly available as a resource for any across breed analysis. The phene ontology will enable integration of OMIA with other global online resources and ultimately enable mechanistic discovery and diagnostics of Mendelian diseases. As an interim step, Mammalian Phenotype Ontology (Smith and Eppig 2009) 'major biological system headers' have replaced existing OMIA 'categories' in advanced searches and in the OMIA browse page (https://omia.org/browse/).

Sustainability of curation efforts is an ongoing concern, and we encourage authors to use HGVS nomenclature wherever possible, and to include EVA IDs when publishing new variants, to facilitate the curation process. OMIA curators (frank.nicholas@sydney.edu.au or imke.tammen@sydney.edu.au) welcome feedback on current information presented in OMIA.

Acknowledgements

The Acknowledgements tab on the OMIA home page lists the many people who contributed to OMIA. Recent software engineering support has been possible due to support of the Sydney Informatics Hub, a core research facility of the University of Sydney, and funding from the Ronald Bruce Anstee Bequest to the Sydney School of Veterinary Science.

References

Allot A., Lee K., Chen Q., Luo L., and Lu Z. (2021) Nucleic Acids Research. 49(W1):W352–W358. https://doi.org/10.1093/nar/gkab326

Online Mendelian Inheritance in Animals, OMIA {9/01/2022}. Available at: https://omia.org/

Shields S., Hermesch S., Nicholas F.W., and Tammen, I. (2021). Proc. of the 24th AAABG Conference, Adelaide, Australia.

Smith C.L., and Eppig J.T. (2009) Wiley Interdiscip. Rev. Syst. Biol. Med. 1(3):390-399. https://doi.org/10.1002/wsbm.44

Tammen I., and Nicholas F.W. (2018) Proc. of the 11th WCGALP, Auckland, New Zealand.

Tammen I., Eager K.L.M., Woolley S.A., Shields S.M.Y., Hermesch S., *et al.* (2021a). Proc. of the 38th ISAG Conference, Virtual Conference.

Tammen I., Vasilevsky N., Park C.A., Hu Z., Haendel M., *et al.* (2021b). Proc. of the 38th ISAG Conference, Virtual Conference.

177. Predicting the purebred-crossbred genetic correlation from phenotype and genotype data of parental lines in pigs

P. Duenk[1*], Y.C.J. Wientjes[1], P. Bijma[1], M.S. Lopes[2,3] and M.P.L. Calus[1]

[1]Wageningen University & Research, P.O. Box 338, 6700 AH Wageningen, the Netherlands; [2]Topigs Norsvin Research Center, P.O. Box 43, 6640 AA Beuningen, the Netherlands; [3]Topigs Norsvin, 80420-210 Curitiba, Brazil; pascal.duenk@wur.nl

Abstract

In previous work, we theoretically derived expressions for an upper and lower bound of the correlation between purebred and crossbred performance (r_{pc}), using only variance components of the parental purebred lines. In the current study, we validated these expressions in real data of pigs by comparing predicted bounds of r_{pc} with the estimated r_{pc}. We compared three methods to approximate the required variance components. The results suggest that the most useful method is to use ordinary REML estimates. If confirmed in other datasets, this approach may help breeders to predict the value of r_{pc} based only on parental line information, or to determine the relative contributions of genotype by genotype and genotype by environment interactions to the value of r_{pc}. We therefore advise studies estimating r_{pc} with genotype data to also estimate and report genetic variance components within and between the parental lines, estimated as described in this study.

Introduction

The genetic correlation between purebred (PB) and crossbred (CB) performance (r_{pc}) is an important parameter in crossbreeding programs, because the r_{pc} determines the response in CB performance when selection is based on PB performance in the parental lines (Smith 1964; Wei *et al.* 1991). A low value of r_{pc} may therefore indicate the need for collecting CB information when the aim is to improve CB performance.

The r_{pc} can be estimated from phenotypes measured on PB and CB animals that are genetically related, and when the pedigree-based relationships between these animals are known (Wei and van der Werf 1995; Lutaaya *et al.* 2001). In practice, however, collecting CB phenotypes can be costly, and the pedigree of CB animals is often difficult to record. The need for a pedigree can be alleviated by genotyping the PB and CB animals that have phenotypes. Although this approach can lead to accurate estimates of r_{pc} (Duenk *et al.* 2019), it may not justify the large investment that is needed for phenotyping and genotyping the required number of CB animals, in particular because the value of CB information for genetic improvement has not yet been determined. Breeders may therefore benefit from predicting the r_{pc} beforehand, based only on information from the parental lines, which is readily available in ongoing breeding programs.

In a previous study, we derived expressions for bounds of r_{pc} based on true variance components computed from QTL effects and genotypes in the parental lines (Duenk *et al.* 2021). These expressions predict the bounds of r_{pc} when only genotype by genotype interaction (G×G) due to dominance and epistasis, and no genotype by environment interaction (G×E) is present. Although obtaining the required variance components is straightforward when QTL effects and genotypes in the parental lines are known (Duenk *et al.* 2021), it is not yet clear how they can be estimated from phenotypes and marker genotypes in the parental lines. In this study, we compare three methods to approximate the variance components required to predict bounds of r_{pc}. For this purpose, we compared predicted bounds of r_{pc} with the estimated r_{pc} in empirical data of pigs.

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_177, © P. Duenk *et al.* 2022

Materials & methods

Data. We used phenotypic and genotypic data from three PB lines and their three-way CB, provided by Topigs Norsvin and Norsvin. In total, we used 17,100 animals from a synthetic sire line (S), 6,611 animals from a Landrace (LR) line, 8,587 animals from a Large-White (LW) line, and 4,173 three-way CB. The LR and LW lines were crossed to produce F1 sows, which were crossed with the sires from S to produce the three-way CB. Purebreds were housed in a nucleus environment, and crossbreds in a commercial environment. Phenotypes were pre-corrected for fixed effects using a larger dataset during the routine genetic evaluation of Topigs Norsvin. Traits included were test growth (TGR), lifetime daily gain (LGR), daily feed intake (DFI), backfat (BFE) and loin depth (LDE). All animals were genotyped with the Illumina 50K SNP chip or the Illumina Custom 25K SNP chip. After quality control, all genotypes were imputed to 50K using Fimpute v2.2 (Sargolzaei *et al.* 2014). Markers with MAF<0.01 in any of the lines were excluded, yielding a total of 35,595 markers that were used for analyses.

Estimating r_{pc}. We estimated r_{pc} in line S using a bivariate model that treats PB and CB performance as different, but genetically correlated traits. The statistical model was:

$$\begin{bmatrix}\mathbf{y}_{\mathrm{PB}}\\ \mathbf{y}_{\mathrm{CB}}\end{bmatrix} = \begin{bmatrix}\mathbf{1}_{\mathrm{PB}} & 0\\ 0 & \mathbf{1}_{\mathrm{CB}}\end{bmatrix}\begin{bmatrix}\mu_{\mathrm{PB}}\\ \mu_{\mathrm{CB}}\end{bmatrix} + \begin{bmatrix}\mathbf{Z}_{\mathrm{PB}} & 0\\ 0 & \mathbf{Z}_{\mathrm{CB}}\end{bmatrix}\begin{bmatrix}\mathbf{u}_{\mathrm{PB}}\\ \mathbf{u}_{\mathrm{CB}}\end{bmatrix} + \begin{bmatrix}\mathbf{e}_{\mathrm{PB}}\\ \mathbf{e}_{\mathrm{CB}}\end{bmatrix} \quad (1)$$

where y is a column vector of corrected phenotypes, μ is the mean, 1 is column vector of 1's, u is a column vector of additive genetic values with incidence matrix Z, and e is a column vector of random residuals. The distribution of additive genetic values for PB and CB performance was:

$$\begin{bmatrix}\mathbf{u}_{\mathrm{PB}}\\ \mathbf{u}_{\mathrm{CB}}\end{bmatrix} \sim N\left(\begin{bmatrix}\mathbf{0}\\ \mathbf{0}\end{bmatrix}, \begin{bmatrix}\sigma^2_{u_{PB}} & \sigma_{u_{PB}u_{CB}}\\ \sigma_{u_{PB}u_{CB}} & \sigma^2_{u_{CB}}\end{bmatrix} \otimes \mathbf{G}\right) \quad (2)$$

where G is the genomic relationship matrix, constructed following Wientjes *et al.* (2017). Estimated variance components were obtained using restricted maximum likelihood (REML) in MTG2 (Lee and van der Werf 2016).

Predicted bounds of r_{pc}. Bounds of r_{pc} were predicted based on parental line information only, using the expressions derived in Duenk *et al.* (2021). The expression for the lower bound of r_{pc} in line S for a three-way CB is:

$$r^L_{pc} = \frac{\sigma_{S(S),S(LW)} + \sigma_{S(S),S(LR)}}{\sigma_S\sqrt{\left(\sigma^2_{S(LW)} + \sigma^2_{S(LR)} + 2\sigma_{S(LW),S(LR)}\right)}} \quad (3)$$

where σ_S is the standard deviation of additive genetic values of individuals in line S for the trait expressed in line S, $\sigma^2_{S(b)}$ is the variance of additive genetic values of individuals in line S for the trait expressed in line *b*, where *b* is LW or LR, and $\sigma_{S(b),S(c)}$ is the covariance between additive genetic values of individuals in line S for the trait expressed in line *b* and *c*. The expression for the upper bound of r_{pc} uses the same parameters, and is

$$r^U_{pc} = \frac{\sigma^2_S + 0.5\sigma_{S(S),S(LW)} + 0.5\sigma_{S(S),S(LR)}}{\sigma_S\sqrt{\left(\sigma^2_S + 0.25\sigma^2_{S(LW)} + 0.25\sigma^2_{S(LR)} + \sigma_{S(S),S(LW)} + \sigma_{S(S),S(LR)} + 0.5\sigma_{S(LW),S(LR)}\right)}} \quad (4)$$

Approximating variance components. Equations (3) and (4) show that we need genetic variance components of parental lines that are usually not available, such as the additive genetic standard deviation in line S for the trait expressed in line LW ($\sigma_{S(LW)}$). Such a component cannot be estimated directly, because there are no phenotypes of individuals from line S for the trait expressed in line LR. Instead, we approximated these components using three different methods. With the first method (M-GEBV), we obtained genomic estimated breeding values (GEBV) of individuals in line S using univariate genomic analyses for each line separately, and computed the (co)variances between these GEBV to approximate the parameters in (3) and

(4). For example, $\sigma_{S(LW)}$ was approximated as the standard deviation of GEBV of individuals in line S for the trait expressed in line LR, and $\sigma_{S(LW),S(LR)}$ was approximated as the covariance between GEBV of individuals in line S for the traits expressed in line LW and line LR. With the second method (M-GEBV-S), we used the same approach, but we corrected for shrinkage of GEBV by dividing the GEBV by the square-root of their reliabilities before using them to approximate the variance components. With the third method (M-REML), we used ordinary genomic REML estimates of variance components within and between lines. For example, $\sigma_{S(LW)}$ was approximated as the ordinary REML estimate of the additive genetic standard deviation in line LW, and $\sigma_{S(LW),S(LR)}$ was approximated as the ordinary REML estimate of the additive genetic covariance between line LW and LR.

Results

The estimated r_{pc} ranged from 0.76 (LGR and DFI) to 0.79 (BFE), with standard errors that ranged from 0.04 (DFI and BFE) to 0.07 (LDE) (Figure 1). With M-GEBV, the estimated r_{pc} was always between the predicted lower and upper bound. However, the difference between the lower and upper bound of r_{pc} (i.e. the range) was large (0.57 on average). Correcting GEBV for shrinkage (M-GEBV-S) resulted in a smaller range (0.40 on average), but the estimated r_{pc} was sometimes higher than the upper bound. Finally, with M-REML, the estimated r_{pc} was usually between the bounds, except for BFE. Furthermore, M-REML resulted in the smallest range (0.22 on average).

Discussion

In this study, we predicted bounds of r_{pc} based on approximated genetic variance components of parental lines. We compared three methods to approximate the required variance components. The results suggest that the most useful method is M-REML, because this method resulted in the smallest range between the lower and upper bound, and in correct bounds for four out of five traits. For one trait (BFE), the estimated r_{pc} was lower than the predicted lower bound, but this difference was smaller than the standard error of r_{pc}.

Based on theory, the variances and covariances in (3) and (4) should be obtained from true breeding values of S individuals, for the trait expressed in the three parental lines. We therefore expected that method M-GEBV-S would yield the most useful predictions, because this method ensures that (1) approximated variance components refer to the individuals in the focal line, and (2) the GEBV are corrected for shrinkage. The difficulty of this method lies in accurate estimation of GEBV (and their reliabilities) for traits expressed in the mated lines. As a result, estimates of variances and covariances may be inaccurate. In contrast, with M-REML, variance components are estimated directly from phenotypes of the trait of interest, making them more accurate. However, with M-REML, the variance components do not refer to the individuals in

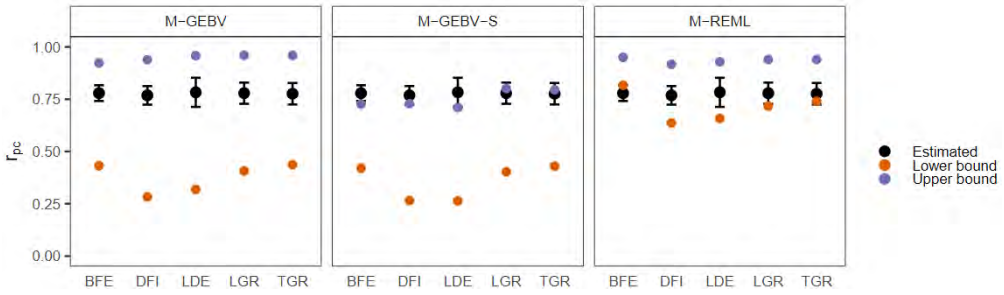


Figure 1. Estimated r_{pc} and predicted bounds of r_{pc} in line S. Facets indicate the method that was used to approximate variance components. BFE=backfat, DFI=daily feed intake, LDE=loin depth, LGR=lifetime daily gain, TGR=test growth rate.

the focal line, but to the individuals in the mated line. For the line and traits that were used in this study, M-REML seemed to give more accurate predictions than M-GEBV-S, suggesting that the assumptions of M-REML have a smaller impact on predicted r_{pc} than those of M-GEBV-S.

The expressions to predict bounds of r_{pc} assume that the value of r_{pc} is determined only by non-additive effects in combination with allele frequency differences between parental lines (i.e. G×G interaction), and not by G×E interaction. In this study, G×E interaction may be present, because the PB animals were housed in a nucleus environment, while the CB animals were housed in a commercial environment. Hence, the value of r_{pc} may be lower than 1 due to both G×E and G×G.

In conclusion, we empirically validated that bounds of r_{pc} can be predicted from ordinary REML estimates of variance components in the parental lines. If confirmed in other datasets, this approach may help breeders to predict the value of r_{pc} based only on parental line information, or to determine the relative contributions of G×G and G×E to the value of r_{pc}.

Acknowledgements

The authors thank the Netherlands Organisation of Scientific Research (NWO) and the Breed4Food consortium partners Cobb Europe, CRV, Hendrix Genetics, and Topigs Norsvin for their financial support. Special thanks go out to Norsvin and Topigs Norsvin for providing the data.

References

Duenk, P., P. Bijma, Y.C.J. Wientjes, and M.P.L. Calus, 2021 Genet Sel Evol 53(1):10. https://doi.org/10.1186/s12711-021-00601-w

Duenk, P., M.P.L. Calus, Y.C.J. Wientjes, V.P. Breen, J.M. Henshall *et al.*, 2019 Genet Sel Evol 51(1):6. https://doi.org/10.1186/s12711-019-0447-9

Lee, S.H., and J.H.J. van der Werf, 2016 Bioinformatics 32(9):1420-1422. https://doi.org/10.1093/bioinformatics/btw012

Lutaaya, E., I. Misztal, J.W. Mabry, T. Short, H.H. Timm *et al.*, 2001 J Anim Sci 79(12):3002-3007. https://doi.org/10.2527/2001.79123002x

Sargolzaei, M., J.P. Chesnais, and F.S. Schenkel, 2014 BMC Genomics 15(478. https://doi.org/10.1186/1471-2164-15-478

Smith, C., 1964 Anim Sci 6(3):337-344. https://doi.org/10.1017/S0003356100022133

Wei, M., H.A.M. van der Steen, J.H.J. van der Werf, and E.W. Brascamp, 1991 J Anim Breed Genet 108(1-6):253-261. https://doi.org/10.1111/j.1439-0388.1991.tb00183.x

Wei, M., and J.H. van der Werf, 1995 J Anim Sci 73(8):2220-2226. https://doi.org/10.2527/1995.7382220x

Wientjes, Y.C.J., P. Bijma, J. Vandenplas, and M.P.L. Calus, 2017 Genetics 207(2):503-515. https://doi.org/10.1534/genetics.117.300152

178. Multi-breed genomic predictions from commercial crossbred data

B.J Hayes*, J. Copley, E. Dodd, E.M. Ross and G. Fordyce

Centre for Animal Science, Queensland Alliance for Agriculture and Food Innovation, University of Queensland, St Lucia, QLD 4067, Australia; b.hayes@uq.edu.au

Abstract

Ideally genomic estimated breeding values would be accurate regardless of breed or breed composition (crossbreds, composites) of the selection candidates. Additionally, it would be an advantage if commercial (potentially crossbred) animals could be used in the reference population, as this would potentially allow for very large reference populations. Here we propose a model for such data, using marker heterozygosity (to model heterosis) and breed composition derived from markers, as covariates in the model. Using a tropical beef data set of close to 30k purebred, crossbred and composite commercial animals, we demonstrate moderately-accurate (0.3-0.43) genomic estimated breeding values can be predicted from this model. Finally, a method for estimating the number of effective genome comparisons for each breed combination within contemporary groups is derived.

Introduction

Multi-breed genomic evaluations are desirable as they allow producers to select sires and dams across breeds, crossbreds and composites, increasing selection intensity. The advantage of multi-breed evaluations for increasing accuracy of genomic prediction, particularly for breeds with smaller reference set sizes, has been demonstrated in dairy cattle (Khansefid *et al.* 2020), sheep (Brito *et al.* 2017, Moghaddar *et al.* 2013) and beef cattle (Hayes *et al.* 2019).

There are a number of challenges to overcome for multi-breed predictions. In particular, multiple breeds must be measured in the same contemporary groups in order to disentangle contemporary group effects from breed effects.

Using commercial data from purebred, crossbred and composite animals (rather than data collected on studs) in multi-breed evaluations is attractive, as such data may be abundant. Such data is also recorded in the target environment for production, so that genotype by environment interaction between stud and commercial environments is no longer an issue. However often little is known about breed composition or heterosis of these animals.

Here, we propose a model for multi-breed genomic prediction based on large scale commercial data that overcomes at least some of these problems.

Methods and materials

Animals and phenotypes. Fifty-four collaborator herds from across northern Australia participated in data collection (the 'Northern Genomics project'). The 28,250 heifers enrolled included crossbred and in approximately 8,000 purebred heifers from at least 14 breeds. Importantly, producers in this project committed to having bulls in the paddock for approximately 6 weeks. Traits measured on the heifers included weight, height and body condition score at an average of 600 days, and a heifer puberty trait. The heifer puberty trait was cycling or not cycling by an average of 600 days (CLscore) assessed by ovarian scanning, as described by (Corbet *et al.* 2018). To maximise genetic variation, the trait is actually measured when an estimated 50% of heifers are pubertal, i.e. at 1.0-2.5 years of age.

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_178, © B.J Hayes *et al.* 2022

Genotypes. All heifers were genotyped with the 35k or 50k tropBeef SNP array by Neogen, Australasia. Genotypes were imputed up to 709,768 SNP (Bovine HD array) using the findhap software (VanRaden *et al.* 2013), and a panel of 4,650 cattle of relevant breeds genotyped for the Bovine HD array.

Statistical analysis. We first estimated breed proportions of each heifer for each of the 14 breeds known to be in the data set (using the 35K array data only). Previously, a separate large data set consisting of only purebred cattle was used to estimate SNP effects for breed composition. A GBLUP model was fitted, where the phenotype was 1 if the animal was of that breed and 0 if not (Dodds *et al.* 2014). The effect of each SNP for the proportion of each breed was then derived by back-solving for the SNP effects (Yang *et al.* 2011), and the resulting prediction equations for each breed were used to estimate breed proportions in the heifers.

Then the model fitted to the CLScore, height, body condition score and weight data was:

$y = \mu + cohort + year + het + breedprop + animal + \varepsilon$

Where: y is a vector of trait records (CLscore, weight, hip height or body condition score); μ is the population mean; cohort is the property+yeardrop+paddock that the heifers were in prior to mustering for trait recording; year is the year of recording; het is the heterozygosity of each heifer as measured by the proportion of markers that were heterozygous (to capture heterosis effects) fitted as a liner effect; breedprop is a series of 15 covariates (14 breeds and *Bos indicus* content), measuring the proportion of each breed in the heifers as described above; animal is a vector of random effects ~ $N(0, \mathbf{G}\sigma_g^2)$, with G the genomic relationship matrix among all heifers (Yang *et al.* 2011) and σ_g^2 the genetic variance captured by the SNP markers; ε is a vector of random deviations ~ $N(0, \mathbf{I}\sigma_e^2)$. There were at least 50 animals in each contemporary group. To report heterosis, which ranges in the data from marker heterozygosity ~0.25 (purebred) to 0.5 (F1), we have re-scaled the estimate to be on the scale of purebred to F1 (e.g. ¼ of the actual estimate of marker heterozygosity which is scaled 0 to 1). Variance components were estimated in GCTA (Yang *et al.* 2011), and the heritability of the traits (actually the proportion of phenotypic variance captured by the SNP) was estimated as $h^2=\widehat{\sigma_g^2}/(\widehat{\sigma_g^2}+\widehat{\sigma_e^2})$.

The accuracy of GEBV was evaluated by dropping out 10 of the 2018 cohorts (the latest birth year of heifers for which data was captured) at random. The breed composition within the 10 herds (3,790 heifers) ranged from purebred Brahman to crossbreds of *Bos taurus* breeds. There were 24,561 heifers in the reference population. GEBV were predicted for the heifers in the 10 excluded herds, then the GEBV were correlated with the actual phenotypes (adjusted for fixed effects) of the heifers within each herd. This correlation was divided by the square root of the heritability of the trait to get the accuracy of genomic prediction. We also assessed accuracy of genomic prediction in a completely independent data set, the Beef CRC cohort of 894 and 1,088 Brahman and Tropical composite cattle (respectively). These cattle were phenotyped for a different fertility trait than we have used, age at first corpus luteum, assessed by ultrasound scanning at six-week intervals (Johnston *et al.* 2013). Genotypes for these cattle were imputed to the same 709,768 SNP used in the genomic predictions.

Ultimately for multi-breed evaluations, head-to-head comparisons of breeds in the same herd/environment are necessary. We assessed how many head-to-head comparisons as $\sum_i^n X_i' X_i$ where for each of *n* cohorts in the data set, X_i is a matrix of breed proportions in cohort *i*, of dimensions number of heifers in the cohort × number of breeds.

Results

Heterosis had a positive effect on all traits, increasing height, weight, body condition and a substantial effect on the chance of being pubertal at scanning, Table 1. The heritability of the traits estimated from the genomic data was moderate for heifer puberty and body condition score, and higher for weight and hip height (Table 1).

Table 1. Number of phenotypes, heterosis effect (scaled purebred to F1 of *Bos indicus* × *Bos taurus*; trait range for reference in body condition score is 1 to 5, for heifer puberty 0 to 1), trait genomic heritabilities (± standard errors) and accuracy of genomic predictions (across 10 herds).

Trait	Number of records	Heterosis	Heritability	Accuracy of genomic prediction
Weight	24,528	40.5 kg	0.29±0.02	0.31±0.04
Hip height	24,261	43.8 mm	0.39±0.02	0.43±0.04
Body condition score	24,537	0.25 units	0.22±0.02	0.33±0.04
Heifer Puberty	28,351	0.27 units	0.22±0.03	0.42±0.05

Accuracies of GEBV in the 10 validation herds were moderate, and fairly consistent across herds, reflected in the relatively low standard error on these accuracies.

The number of head-to-head comparisons possible from the data set, which enables estimates of breed effect, was reasonable for Angus versus Brahman (337), Brahman versus Droughtmaster (924) and Brahman versus Santa Gertrudis (421), but was lower for other breed combinations.

Discussion

This study demonstrates that moderate accuracies of multi-breed genomic prediction can be achieved from reference sets derived from large scale commercial data. Genomic heritabilities were consistent with previous estimates for these traits in tropical beef cattle data sets derived from pedigree (e.g. Corbet *et al.* 2018), as were heterotic effects (e.g. Frisch and Vercoe 1984). The challenges of unknown breed composition and level of heterosis can be solved using estimates of these parameters obtained directly from the genotype data (e.g. Dodds *et al.*, 2014, Akanno *et al.* 2016). Using crossbred data allows head-to-head comparison of chromosome segments derived from different breeds, disentangling breed effects from contemporary group effects, at least to some extent. This suggests the data set will contribute to multi-breed genomic evaluations for many, but not all breeds used in northern Australia.

Acknowledgements

The authors gratefully acknowledge the participation and efforts of the 54 northern herds in the Northern Genomics Project, and assistance from various DAF staff at various stages of the project. We are grateful for funding from the MLA Donor Company and University of Queensland through project P.PSH.P0833.

References

Akanno E.C., Chen, L., Abo-Ismail, M.K., Crowley, J.J. *et al.* (2016) Canadian Journal of Animal Science. 97, 431-438. https://doi.org/10.1139/cjas-2016-0124

Brito, L.F., Clarke, S.M., McEwan, J.C. *et al.* (2017) BMC Genet 18, 7. https://doi.org/10.1186/s12863-017-0476-8

Corbet, N.J., Allen, J.M., Laing, A.R., Fordyce, G., McGowan, M.R., *et al.* (2018) Anim. Prod. Sci. Vol.58, 1735–1742. https://doi.org/10.1071/AN16616

Dodds, K.G., Auvray, B., Newman, SA.N. McEwan, J.C. (2014) BMC Genet. 15:92 https://doi.org/10.1186/s12863-014-0092-9

Frisch JE. and Vercoe JE. (1984) J. Agric. Sci. Camb. 103, 137. https://doi.org/10.1017/S0021859600043409

Hayes, B.J., Corbet, N.J., Allen, J.M., Laing, A.R., Fordyce G. *et al.* (2019). J. Anim. Sci. 97, 55–62. https://doi.org/10.1093/jas/sky417

Johnston D. J., Barwick S.A., Fordyce G., Holroyd R. G., Williams P. J. *et al.* (2013). Australia. Anim. Prod. Sci. 54, 1–15. https://doi.org/10.1071/AN13043

Khansefid M, Goddard ME, Haile-Mariam M, Konstantinov KV, Schrooten C, *et al.* (2020) Front Genet. 11, 598580 https://doi.org/10.3389/fgene.2020.598580

Moghaddar N., Swan A. A., van der Werf J. H. J. (2013) Animal Production Science 54, 544-549. https://doi.org/10.1071/AN13129

VanRaden, P.M., Null, D.J., Sargolzaei, M., Wiggans, G.R., Tooker, M.E. *et al.* (2013) J Dairy Sci. 96, 668-78. https://doi.org/10.3168/jds.2012-5702

Yang., J., Lee, S.H, Goddard, M.E. Visscher, P.M. (2011) Am. J. Hum. Genet. 88, 76-82. https://doi.org/10.1016/j.ajhg.2010.11.011

179. Simultaneous mapping of quantitative trait loci in multiple subspecies using hybrid populations

C.L. Warburton[1*], R.C. Costilla[2], B.N. Engle[1] and B.J. Hayes[1]

[1]Centre for Animal Science, Queensland Alliance for Agriculture and Food Innovation, University of Queensland, St Lucia, QLD 4067, Australia; [2]Agresearch Limited, Ruakura Research Centre, Hamilton 3214, New Zealand; c.warburton@uq.edu.au

Abstract

Beef cattle breeds in North Australia originate from two subspecies, *Bos indicus* and *Bos taurus*. Development of accurate genomic evaluations for this industry will require models that can accurately predict across all breeds and cross-breeds of these two subspecies in this industry. As these subspecies are genetically diverse, it will be essential to understand if there are differences in genomic architecture of quantitative traits between these subspecies. We propose a novel method of using haplotypes to define subspecies of origin in a multi-breed, multi-subspecies population of heifers with no available pedigree. Results from this method show that subspecies specific genomic architecture exists in this population of heifers, and further research is required to develop methods to account for this population structure in genomic evaluations.

Introduction

Many livestock breed and crop varieties are actually hybrids of species or subspecies. For example, tropical beef and dairy cattle are often hybrids of *Bos indicus* and *Bos taurus* (Achilli *et al.*, 2008, Bolormaa *et al.*, 2011). Accurate genomic selection and QTL mapping relies upon sufficient linkage disequilibrium (LD) between SNP on the marker panel and the quantitative trait loci (QTL) affecting the trait of interest (Meuwissen *et al.*, 2001, de Roos *et al.*, 2008). For subspecies that diverged many generations ago, it is unlikely that the same QTL segregate in both subspecies (Bolormaa *et al.* 2011). When genomic selection and or QTL mapping is applied in such populations, ghost QTL, where a SNP tracks a QTL in a subset of the population but not the rest of the population, may result (Kemper *et al.* 2015). This reduces the accuracy of genomic predictions and reduces mapping precision. Here we propose a new method for genomic prediction that accounts for the possibility of different LD between SNP and QTL in haplotypes derived from the different subspecies that make up a hybrid population.

Methods and materials

Genotypes and phenotypes. In order to assign SNP alleles in the reference population to subspecies of origin, genotypes from 1,182 purebred *B. indicus* and *B. taurus* animals were obtained from the 1000 Bulls Genome Project Run 6 (Hayes and Daetwyler, 2019). In total there were 868 *B. indicus* animals from a single breed, Brahman, and 313 *B. taurus* animals from five breeds, Angus (n=100), Hereford (n=43), Limousin (n=62), Shorthorn (n=95) and Charolais (n=13).

The genomic prediction reference population consisted of 3695 heifers from three breeds. There were 979 Brahman heifers, which is a *B. indicus* breed, 914 Santa Gertrudis and 1802 Droughtmaster heifers, which are both stabilised composites of *B. indicus* × *B. taurus* origin. All heifers were genotyped with the Geneseek GGP-LD array consisting of 21,121 SNP (Hayes *et al.*, 2019) that was then imputed up to the Bovine HD array of 728,785 SNP (800K panel) using the FImpute software (Sargolzaei *et al.*, 2014). All heifers were phenotyped for reproductive maturity score (RMS) which is a single-ultrasound measure of puberty that has been shown to be moderately heritable (20%) in this population of heifers (Warburton *et al.*, 2020). Full details of heifer management and recording have been described previously (Corbet *et al.*, 2018).

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_179, © C.L. Warburton *et al.* 2022

Subspecies of origin assignment. Both datasets were phased using Eagle version 2.4.1 (Loh *et al.*, 2016) and haplotypes were generated using fixed, non-overlapping windows of 100 kb from in-house scripts using Julia (Bezanson *et al.*, 2017). A second Julia function was developed to calculate subspecies specific haplotype frequencies from the reference population. Subspecies of origin was assigned using a method derived from Bolormaa *et al.* (2011). For each animal in the validation population, each haplotype in the genome was assigned a b-value, which is the probability that the haplotype is *B. indicus* in origin (Bolormaa *et al.* 2011).

$$b = \frac{pBi_j}{pBi_j + pBt_j} \quad (1)$$

Where *pBi* is the frequency of the j^{th} haplotype in the *B. indicus* 1000 Bulls Genomes data and *pBt* is the frequency of the j^{th} haplotype in the *B. taurus* 1000 Bulls Genomes data. If a $b \geq 0.5$ the haplotype was considered to be *B. indicus* in origin and if $b < 0.5$ it was considered *B. taurus*. In the instance a haplotype from the validation population did not exist in the reference, Hamming distance methods were used to calculate *b* in that subspecies using Julia (Gomez, 2015, Van der Loo, 2014). Each animal has two haplotypes per locus *j* and b-values were used to estimate if an animal is homozygous *B. indicus* (BiBi), homozygous *B. taurus* (BtBt) or B. *indicus* × *B. taurus* (BiBt) composite for each SNP within a haplotype. After assigning subspecies of origin to each animals SNP's, *B. indicus* content was averaged for all animals within a breed. The estimated *B. indicus* content from haplotypes was compared to theoretical *B. indicus* content for each of the breeds, derived from known breed history, to assess the accuracy of subspecies assignment using haplotypes.

X-matrix construction. We designed a subspecies-specific X-matrix, with dimensions of $n \times 3m$ where *n* is the number of animals in the validation population, and *m* is the number of SNP on the marker panel used to construct haplotypes. Each SNP from the 800K panel has three columns in the X-matrix, one for BiBi, one for BtBt and one for BiBt. After subspecies assignment, the X matrix was populated with phased genotype data (0, 1 or 2) using subspecies of origin assignment (BiBi, BtBt, BiBt) to allocate the appropriate X-matrix column for each SNP and animal.

Bayesian analysis. BayesR analyses used the model in Equation 2.

$$RMS = 1n\mu + cg + age + Xg + e \quad (2)$$

Where RMS is a vector of reproductive maturity score phenotypes, 1n is a vector of ones, cg is contemporary group defined as herd, year and season fitted as a fixed effect and age is age at measurement fitted as a covariate. As there were no mixed breed contemporary groups in this dataset, breed was not fitted as a covariate in the model as contemporary group described all variation due to breed. X is the X-matrix described above and g is a vector of SNP effects with a distribution $g \sim N(\mathbf{0}, \sigma_i^2$ where σ_i^2 belongs to one of four distributions: $\sigma_i^2 = \{0, 0.0001, 0.001, \text{or } 0.01\} \times \sigma_g^2$ and σ_g^2 is the genetic additive variance. Full details on the BayesR model can be found in Erbe *et al.* (2012). We used 50,000 Gibbs chain iterations, with the first 20,000 iterations discarded as burn-in.

Subspecies genome mapping. The X-matrix used in this BayesR analysis allowed us to simultaneously estimate subspecies effects for each SNP on the 800K marker panel. Subspecies specific genomic architecture for the RMS trait was investigated by identifying differences in subspecies specific SNP effects for each of the SNP on the 800K marker panel. The estimated subspecies SNP effects were plotted using the CMplot package in R (Yin 2021).

Results

Results in Table 1 show the average *B. indicus* content of each of the breeds of heifers, estimated using haplotypes (observed) and theoretical known breed history (expected). These results show that there are no significant differences in *B. indicus* content between theoretical estimates and haplotype subspecies assignment. This suggests that haplotype assigned subspecies of origin is able to detect and trace subspecies specific genomic architecture accurately throughout the genome.

Figure 1 shows that there are distinct subspecies differences in the genetic architecture for RMS in this population of heifers. There are very few instances where SNP are explaining variation in both subspecies. Furthermore, in the few regions where SNP are explaining variation in multiple subspecies, the estimated SNP effects are often quite different between the subspecies.

Discussion

To our knowledge, this is the first attempt using BayesR to map QTL in multiple subspecies in a hybrid population. Our results suggest the method can identify distinct subspecies differences. In some instances, the estimated subspecies SNP effects showed opposing effects with other subspecies, which may suggest that associations (phase) between SNP and QTL differ between the subspecies. It has been estimated that the *B. indicus* and *B. taurus* subspecies diverged from a common ancestor approximately 332,000 years ago (Achilli *et al.*, 2008). Thus, these observed differences in genomic architecture are not unexpected.

Table 1. Average *Bos indicus* content of Brahman, Droughtmaster and Santa Gertrudis heifers estimated from known breed origins (expected) and from haplotype subspecies mapping (observed).

Breed	Expected *B. indicus* %	Observed *B. indicus* %
Brahman	90	89
Droughtmaster	50	54
Santa Gertrudis	34	39

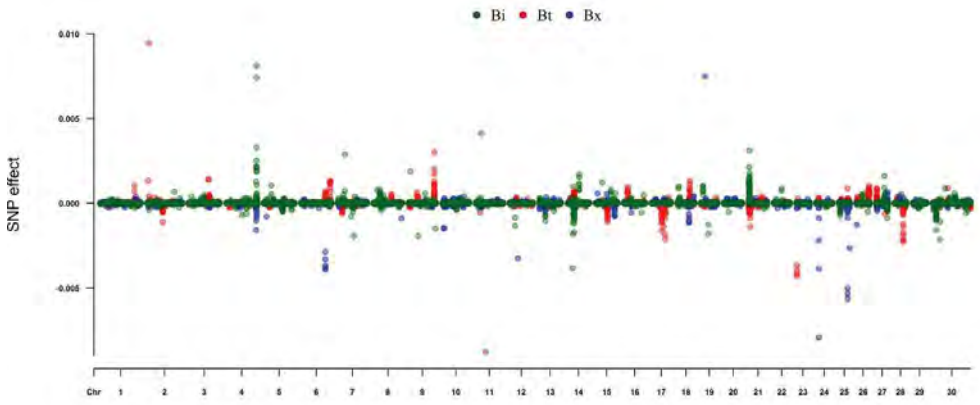


Figure 1. Estimated SNP effects (y-axis) from BayesR analysis using haplotypes to allocate subspecies of origin across all chromosomes (x-axis). Estimated homozygous *Bos indicus* (Bi) SNP effects in green, homozygous *Bos taurus* (Bt) effects in red, and *B. indicus* × *B. taurus* (Bx) effects in blue.

Accurate genomic evaluations rely upon maintenance of LD-phase between SNP and QTL (de Roos *et al.*, 2008). In a study of multi-breed *B. taurus* dairy and beef cattle, de Roos *et al.* (2008) showed that high density (~300,000) marker panels would be required to accurately perform genomic evaluations between these breeds due to differing LD associations. It must be noted that in the paper by de Roos *et al.* (2008), all of the breeds studied were from a single subspecies, *B. taurus*. In our study we have shown that LD phase between SNP and QTL is unlikely to be conserved across subspecies, *B. indicus* and *B. taurus*. Failure to take this into account may generate ghost QTL, and decrease the precision of QTL mapping. More work is required to investigate methods to use subspecies specific SNP effects in genomic evaluations, with the aim of improving prediction accuracy in hybrid populations.

References

Achilli, A., Olivieri, A., Pellecchia, M., Uboldi, C., Colli, L., *et al.* (2008) Curr. Biol. 18:R157–R158. https://doi.org/10.1016/j.cub.2008.01.019

Bezanson, J., Edelman, A., Karpinski, S., Shah, V.B. (2017) SIAM Rev. 59:65–98. https://doi.org/10.1137/141000671

Bolormaa, S., Hayes, B.J., Hawken, R.J., Zhang, Y., Reverter, A. *et al.* (2011) J. Anim. Sci. 89:2050–2060. https://doi.org/10.2527/jas.2010-3363

Corbet, N.J., Allen, J.M., Laing, A.R., Fordyce, G., McGowan, M.R., *et al.* (2018) Anim. Prod. Sci. 58:1735–1742. https://doi.org/10.1071/AN16616

De Roos, A.P.W., Hayes, B.J., Spelman, R.J., Goddard, M.E. (2008) Genetics 179:1503. https://doi.org/10.1534/genetics.107.084301

Erbe, M., Hayes, B.J., Matukumalli, L.K., Goswami, S., Bowman, P.J. *et al.* (2012) J. Dairy Sci. 95:4114–4129. https://doi.org/10.3168/jds.2011-5019

Gomez, M., (2015) String Distances, Available at: https://juliapackages.com/p/stringdistances

Hayes, B.J., Corbet, N.J., Allen, J.M., Laing, A.R., Fordyce *et al.* (2019). J. Anim. Sci. 97: 55–62. https://doi.org/10.1093/jas/sky417

Hayes, B.J., Daetwyler, H.D. (2019) Annu. Rev. Anim. Biosci. 7:89–102. https://doi.org/10.1146/annurev-animal-020518-115024

Kemper KE, Hayes BJ, Daetwyler HD, Goddard ME. (2015). J Anim Breed Genet. 132:121-34. https://doi.org/10.1111/jbg.12152

Loh, P.-R., Danecek, P., Palamara, P.F., Fuchsberger, C., Reshef, Y.A. *et al.* (2016) Nat. Genet. 48:1443–1448. https://doi.org/10.1038/ng.3679

Meuwissen, T.H., Hayes, B.J., Goddard, M.E. (2001) Genetics 157:1819–1829. https://doi.org/10.1093/genetics/157.4.1819

Sargolzaei, M., Chesnais, J., Schenkel, F. (2014) BMC Genomics 15:478–478. https://doi.org/10.1186/1471-2164-15-478

Van der Loo, M.P., 2014. The stringdist package for approximate string matching. R J 6, 111. https://doi.org/10.32614

Warburton, C.L., Engle, B.N., Ross, E.M., Costilla, R., Moore, S.S. *et al.* (2020) Genet. Sel. Evol. 52:28. https://doi.org/10.1186/s12711-020-00547-5

Yin, L., Zhang, H., Tang, Z., Xu, J., Yin, D. (2021) Genomics Proteomics Bioinformatics. https://doi.org/10.1016/j.gpb.2020.10.007

180. Strategies to find predictive variants that improve multi-breed genomic prediction in dairy cattle

I. van den Berg[1*], R. Xiang[1,2], M. Khansefid[1], M.E. Goddard[1,2], J.E. Pryce[1,3], C.P. Prowse-Wilkins[1,2], A.J. Chamberlain[1,3] and I.M. MacLeod[1]

[1]Agriculture Victoria, AgriBio, Centre of AgriBioscience, 5 Ring Road, Bundoora, Victoria, 3082, Australia; [2]Faculty of Veterinary & Agricultural Science, University of Melbourne, 3010 Parkville, Victoria, Australia; [3]School of Applied Systems Biology, La Trobe University, Bundoora, Victoria, 3083, Australia; irene.vandenberg@agriculture.vic.gov.au

Abstract

The lack of large reference populations has limited the accuracy of genomic prediction for dairy cattle breeds with small populations. We review strategies to improve multi-breed genomic prediction. Several studies show that multi-breed reference populations can increase accuracy for minor breeds. Evidence suggests sequence and functional genomics can get us closer to causal variants, and thus improve multi-breed prediction. However, we know little about the sharing of causal variants across breeds. We carried out a case study comparing direction of effect of variants selected from within-breed genome-wide association studies (GWAS), a meta-GWAS, or a combination of functional, evolutionary, and pleiotropic information. Variants selected from a meta-GWAS were most likely to have the same direction of effect in different populations. Selecting variants from a meta-GWAS with the same direction of effect in multiple breeds may be a suitable strategy to select sequence variants that increase the accuracy of multi-breed genomic prediction.

Introduction

Genomic prediction has revolutionised the dairy cattle industry in the last decade. This is especially true for breeds with large reference populations, such as Holstein cattle. However, the uptake of genomic prediction has been challenging for breeds with smaller populations. The size of the reference population is a major factor determining the accuracy of genomic prediction (Liu *et al.*, 2011), and the lack of sufficiently large reference populations has been an obstacle for some breeds. For example, in Australia, genomic prediction for Holstein has been implemented in 2011, yet genomic estimated breeding values (GEBV) for Australian Red dairy cattle (a minor breed) were first released only in August 2021, 10 years later. Furthermore, even though Australian Reds now have GEBV for milk, fat, protein and somatic cell count; Holsteins have GEBV for a wider range of traits, such as fertility, survival, feed efficiency and heat tolerance. Hence, for breeds with smaller population sizes obtaining a sufficiently large reference population for all traits, the dairy industry does not have access to the same level of genomic and phenotypic information as breeds with larger datasets. For closely related populations, sharing reference populations can increase prediction accuracy. However, the benefit declines as breeds become more distantly related (Lund *et al.*, 2016). In this paper, we reviewed factors impacting the accuracy of multi-breed genomic prediction in dairy cattle, and future opportunities. Furthermore, we carried out a case study comparing the direction of effect in three dairy cattle breeds of preselected sequence variants.

Composition of multi-breed reference population

Larger reference populations tend to increase genomic prediction accuracy (Liu *et al.*, 2011). However, this is not necessarily true for multi-breed prediction. A multi-breed reference population dominated by one breed can actually reduce the prediction accuracy in smaller breeds (Khansefid *et al.*, 2020; van den Berg *et al.*, 2020a). In Australian Reds, prediction accuracies increased substantially when a small number of Holstein bulls was included in the reference population, but further increasing the number of Holsteins

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_180, © I. van den Berg *et al.* 2022

reduced the accuracy, sometimes below that achieved within-breed (van den Berg *et al.*, 2020a). Similar results were obtained for genomic prediction of crossbred dairy cattle (Khansefid *et al.*, 2020). Hence, rather than using all available data, multi-breed reference populations need to be designed cautiously.

Selection of sequence variants

Various studies have shown that using preselected sequence variants can help to increase the accuracy of genomic prediction (e.g. Brøndum *et al.,* 2015; VanRaden *et al.*, 2017), and can be especially beneficial for across-breed prediction (Raymond *et al.*, 2018). The optimal set of prediction markers may, however, vary between traits and breeds (van den Berg *et al.*, 2016). Additionally the increasing availability of 'omics phenotypes, such as metabolomics and transcriptomics, provides opportunities to detect likely causal variants. For example, Prowse-Wilkins *et al.* (2021) showed that variants under ChIP-seq peaks were enriched for quantitative trait loci (QTL) associated with milk production traits, as well as for expression QTL (eQTL). Littlejohn *et al.* (2016) reported an eQTL in *MGST1* associated with milk production traits. Xiang *et al.* (2019) developed the Functional and Evolutionary Trait Heritability (FAETH) score, that integrates functional, evolutionary, and pleiotropic information of variants and can be used to select prediction markers than can increaseprediction accuracy in multiple breeds (Xiang *et al.*, 2019). A custom chip of approximately 50,000 variants, of which three-quarters were selected using sequence and multi-omic approaches, increased the accuracy of genomic prediction in a multi-breed setting (Khansefid *et al.*, 2020).

Multi-breed GWAS

While long-range linkage disequilibrium (LD) can be beneficial for genomic prediction, it reduces precision of within-breed GWAS. When a large section of the genome is in high LD, many sequence variants are highly significant in a GWAS. While many of these variants could be good prediction markers within-breed, this is not necessarily true across breeds, where LD is only conserved over a shorter distance. Including multiple breeds in the GWAS to select sequence variants may, therefore, help to identify variants that are in consistent LD with the causal variant in multiple breeds. Even for within-breed prediction, variants selected from multi-breed GWAS resulted in a higher prediction accuracy than variants selected from within-breed GWAS (van den Berg *et al.*, 2016). In practice, sharing genotypes and phenotypes for a multi-breed GWAS can be challenging. However, a multi-breed GWAS can be approximated with a meta-GWAS, by combining only summary statistics of multiple GWAS, and increase power and precision compared to within-breed GWAS (van den Berg *et al.*, 2020b).

Models suited for multi-breed prediction

While GBLUP assumes all variants have a small effect, using a weighted genomic relationship matrix (GRM) (Zhang *et al.*, 2010), or fitting multiple GRMs (van den Berg *et al.*, 2016) allows preselected sequence variants to have larger effects than standard prediction markers. However, the effects are often assumed to be the same across breeds. Using haplotypes instead of variants to define the relationships between individuals within and across breeds in multi-breed predictions was found to marginally improve the accuracy of predictions (Khansefid *et al., 2022*). Bayesian variable selection models assume most variants have no or very small effects and allow some markers to have larger effects. BayesRC is especially suitable for the analyses of preselected sequence variants, as it uses prior biological information to define classes of variants that are expected to be enriched for causal variants (MacLeod *et al.*, 2016). Alternatively, intermediate omics data can be integrated directly in genomic prediction models alongside prediction markers (Christensen *et al.*, 2021). Multi trait genomic prediction models that consider the same trait in different breeds as different but correlated traits can result in higher accuracies than single trait models and are less sensitive to the composition of the reference population (van den Berg *et al.,* 2020a).

Case study: causal variants sharing across breeds

A potential explanation for the limited success of multi-breed prediction to date could be inconsistent LD between prediction markers and causal variants across breeds. Identifying causal variants, and using models that can allocate an appropriate effect to the variants closest to causal variants, may therefore increase the accuracy of multi-breed genomic prediction. However, this assumes that the majority of causal variants are shared across breeds. Only a few causal variants are identified in dairy cattle, making it challenging to estimate how many causal variants are shared across breeds. The large variation in sample size across breeds makes it challenging to set a threshold to detect QTL and determine whether QTL are shared or not.

Here we evaluated differences in variant effects across breeds by comparing the direction of effect in GWAS in Holstein, Jersey and Australian Red cows. To this end, we undertook within-breed and sex GWAS and a multi-breed meta-GWAS using imputed whole genome sequence data and protein percentage records of 6,323 Holstein bulls, 56,843 Holstein Cows, 1,476 Jersey bulls, 10,210 Jersey cows and 3,080 Australian Red cows. The top 1000/500/100 variants were selected as follows: (1) variants were ranked based on *P*-value or FAETH score; (2) the most significant variant or variant with highest FAETH score genome-wide was selected; (3) variants within 1 Mb of selected variant were discarded; (4) repeat until the desired number of variants was selected. We then estimated the proportion of variants with the same direction of effect (pSDE) in different populations for all variants and the top variants. While pSDE in different breeds increased when variants were more significant in within-breed GWAS (Table 1), pSDE remained substantially lower than for bulls and cows of the same breed. The pSDE across populations was larger for top variants selected from the meta-GWAS than the within-breed GWAS, because *P*-values in the meta-GWAS decrease when the direction of effect is consistent in all populations. Variants selected based on their FAETH score tended to have a higher pSDE than all variants, though the difference was smaller than for variants selected from within-breed GWAS or meta-GWAS. This is not surprising, given that the FAETH variants are not specifically selected for protein percentage. Selecting variants based on their *P*-value in a meta-GWAS and the direction of effect in different populations showed promising results for genomic prediction of milk urea nitrogen using a multi-breed and multi country reference population (van den Berg *et al.*, 2022), and may be a suitable method to select variants for a specific multi-breed reference population and trait. The effect direction meta-analysis developed by Xiang *et al.* (2020) uses direction of effect in multiple GWAS to select pleiotropic variants rather than trait specific variants.

Table 1. Proportion of variants with the same direction of effect across populations[1] for different sets of variants2 (effect direction assessed based on within-breed GWAS).

			GWAS			meta-GWAS			FAETH		
pop1	**pop2**	**all**	**1000**	**500**	**100**	**1000**	**500**	**100**	**1000**	**500**	**100**
HB	HC	0.58	0.79	0.85	0.99	0.98	0.99	1.00	0.60	0.58	0.62
JB	JC	0.57	0.68	0.73	0.85	0.80	0.84	0.92	0.58	0.57	0.58
HC	JC	0.51	0.54	0.57	0.73	0.93	0.94	0.94	0.54	0.57	0.51
HC	RC	0.52	0.58	0.61	0.71	0.82	0.85	0.94	0.53	0.55	0.61
JC	RC	0.50	0.52	0.54	0.66	0.80	0.83	0.93	0.51	0.47	0.46

[1] Populations tested were Holstein bulls (HB), Holstein cows (HC), Jersey bulls (JB), Jersey cows (JC) and Australian Red cows (RC); [2] the 1000/500/100 most significant QTL regions in GWAS or meta-GWAS, or with highest FAETH score

Conclusions

To date, implementing routine multi-breed genomic prediction remains challenging. Further identification of causal variants and integration of multi omics data looks promising to increase the accuracy of multi-breed prediction, but we need to better understand the proportion of causal variants shared across breeds. Selecting sequence variants based on multi-breed GWAS or meta-GWAS with consistent direction of effect in different breeds may be a useful strategy to select variants for multi-breed genomic prediction.

References

Brøndum R.F., Su G., Janss L., Sahana G., Guldbrandtsen B. *et al.* (2015) J. Dairy Sci. 98(6):4107-4116. http://dx.doi.org/10.3168/jds.2014-9005

Christensen O.F., Bömer V., Varona L., and Legarra A. (2021) Genetics 219(2):iyab130. https://doi.org/10.1093/genetics/iyab130

Khansefid M., Goddard M.E., Haile-Mariam M., Konstantinov K.V., Schrooten C. *et al.* (2020) Front. genet. 11:1588. https://doi.org/10.3389/fgene.2020.598580

Khansefid M., Ferdosi M.H., Goddard M.E., Haile-Mariam M., Schrooten C. *et al.* (2022) Submitted to WCGALP 2022.

Littlejohn M.D., Tiplady K., Fink T.A., Lehnert K., Lopdell T. *et al.* (2016) Sci. Rep. 6:25376. https://doi.org/10.1038/srep25376

Liu Z., Seefried F.R., Reinhardt F., Rensing S., Thaller G. *et al.* (2011) Genet. Sel. Evol. 43:19. https://doi.org/10.1186/1297-9686-43-19

Lund M.S., van den Berg I., Ma P., Brøndum R.F., and Su G. (2016) animal 10(7):1042-1049. https://doi.org/10.1017/s1751731115003031

Prowse-Wilkins C.P., Wang J., Xiang R., Garner J.B., Goddard M.E. *et al.* (2021) Front. genet. 12:664379. https://doi.org/10.3389/fgene.2021.664379

Raymond B., Bouwman A.C., Schrooten C., Houwing-Duistermaat J., Veerkamp R.F. (2018) Genet. Sel. Evol. 50:27. https://doi.org/10.1186/s12711-018-0396-8

Van den Berg I., Boichard D., and Lund M.S. (2016). Genet. Sel. Evol. 48:83. https://doi.org/10.1186/s12711-016-0259-0

Van den Berg I., MacLeod I.M., Reich C.M., Breen E.J., and Pryce J.E. (2020a) J. of Dairy Sci. 103(7):6276-6298. https://doi.org/10.3168/jds.2019-17914

Van den Berg I., Xiang R., Jenko J., Pausch H., Boussaha M. *et al.* (2020b) Genet. Sel. Evol. 52:37. https://doi.org/10.1186/s12711-020-00556-4

Van den Berg I., Huong P.H., Nguyen T.V., Haile-Mariam M., MacLeod I.M. *et al.* (2022). Genet. Sel. Evol. 54:15. https://doi.org/10.1186/s12711-022-00707-9

VanRaden P.M., Tooker M.E., O'Connell J.R., Cole J.B. and Bickhart D.M. (2017). Genet. Sel. Evol. 49:32. https://doi.org/10.1186/s12711-017-0307-4

Xiang R., van den Berg I., MacLeod I.M., Hayes B.J., Prowse-Wilkins C.P. *et al.* (2019) PNAS 116(39):19398-19408. https://doi.org/10.26188/5c5617c01383b

Xiang R., van den Berg I., MacLeod I.M., Daetwyler H.D., and Goddard M.E. (2020) Commun. Biol. 3:88. https://doi.org/10.1038/s42003-020-0823-6

Zhang Z., Liu J., Ding X., Bijma P., de Koning D.-J. *et al.* (2010) PLoS ONE 5:e12648. https://doi.org/10.1371/journal.pone.0012648

181. Impact of breed ancestry on the performance of crossbred dairy cattle

M.A. Jaafar[1*],B.J. Heins[2], C. Dechow[3], C.P. Van Tassell[4] and H.J. Huson[1]

[1]Cornell University, Animal Science Department, 149 Morrison Hall, Ithaca NY 14850, USA; [2]University of Minnesota, West Central ROC, 46352 State Hwy 329, Morris, MN 56267, USA; [3]Penn State University, Department of Animal Science, University Park, State College, PA 16802, USA; [4]United States Department of Agriculture, Animal Genomics and Improvement Lab, 10300 Baltimore Ave, Beltsville, MD 20705, USA; maj223@cornell.edu

Abstract

Crossbred dairy cattle populations have different proportions of parental breeds. However, individuals who display the same breed composition could exhibit remarkable differences in the distribution of local ancestry across the genome. Thus, this study examined the distribution of local ancestry blocks using Illumina BovineSNP50 genotypes and correlated them to key production traits. Two rotational crossbred populations, referenced as ProCROSS and Grazecross were assessed. The results showed higher performance ProCROSS animals in milk fat production more commonly have Holstein ancestry on BTA 2, 6, 13, 22 and 23, where the genes Prion, Lactotransferrin, and Cytochrome P450 reside. We found evidence of an excess proportion of Viking Red ancestry and decreased Jersey ancestry on BTA 1, 10, 12, 14 and 25 at known health QTL in the Grazecross population. In conclusion, we believe that localized breed ancestry plays an important role; thus, selecting and maintaining particular ancestry blocks is crucial to ensure optimal trait performance in crossbred cattle.

Introduction

Generally, crossbreeding is the breeding among individuals from at least two distinct populations. The admixed individuals will inherit haplotype blocks that contain genetic variants of one or more populations. The melding of these haplotype blocks increases the heterozygosity level that contributes to heterosis or hybrid vigour in the admixed offspring. However, this heterozygosity may not always exist, as recombination events occur randomly in gamete formation, creating unique combinations of chromosome segments from each population. Due to this, admixed individuals may display the same breed composition but exhibit differences in the distribution of local ancestry blocks that directly affect complex production trait performance. Mapping local ancestry composition instead of global ancestry could aid in identifying causal variants in complex traits, especially ancestry-specific traits in admixed individuals (Ding *et al.*, 2011). Therefore, our primary goals were to describe the distribution of local ancestry blocks across the genome of the ProCROSS and Grazecross dairy cattle populations. Furthermore, we aimed to explore the correlation of the local ancestry blocks with various production traits.

Materials & methods

Study population. The University of Minnesota (UMN) granted access to ProCROSS and Grazecross datasets. The datasets included 378 ProCROSS and 229 Grazecross cattle genotypes and four traits (milk yield (MY), fat yield (FY), protein yield (PY) and somatic cell score (SCS). To compare ancestry variation for trait performance, the top 20% high-performance animals were compared to the lowest 20% low-performance animals for each trait. To investigate admixture of the ProCROSS and Grazecross, genotypes representing the purebred and admixed breeds used in the two rotational systems were needed for comparison. ProCROSS are a product of rotational crossbreeding of Viking Red (VKR), Holstein (HOL), and Montbeliarde (MON). In contrast, Grazecross consists of VKR, Normande (NOR), and Jersey (JER). VKR breeds were represented by 3 genotype breeds Danish Red (DNR, n=4), Swedish Red

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_181, © M.A. Jaafar *et al.* 2022

(SWD, n=32) and Finnish Ayrshire (FAY, n=27) (Upadhyay *et al.*, 2017; Iso-Touru *et al.*, 2016, Viking Genetics, Denmark). HOL genotypes (n=92) were provided by UMN obtained from animals directly used in the development of these two admixed populations. Jersey (JER, n=73), Montbeliarde (MON, n=34), and Normande (NOR, n=35) genotypes were obtained from the Gautier *et al.* (2010) publication and represented purebred animals from the USA (JER) and Europe (MON & NOR), respectively. SNP data of all animals were obtained through bovine genotyping array kits BovineHD DNA Analysis Kit (HD150K), BovineSNP50 DNA Analysis BeadChip (50K) and CLARIFIDE® 50k (ZL5) (Illumina- Neogen, Lansing, MI, USA; Clarifide-Zoetis, San Diego, CA, USA).

Filtering and quality control of genomic data. Quality control (QC) analyses for both sets of autosomal SNPs were calculated using SNP and Variation Suite (SVS) v8.x (Golden Helix, Inc., Bozeman, MT, USA). SNPs were excluded if they were unmapped to the UMD 3.1 bovine genome assembly (Zimin *et al.*, 2009) or mapped to the sex chromosomes, had a call rate less than 0.95, or a minor allele frequency less than 0.05. SNPs were further pruned for linkage disequilibrium (LD) using a threshold of r^2>0.75, leaving 9,792 SNPs and 10,795 SNPs for ProCROSS and Grazecross analysis respectively. Overall, SNPs used provided relatively clear separation between ancestral and admixed populations.

Admixture analysis. The global breed composition was estimated from genomic data using a maximum likelihood model implemented in ADMIXTURE 1.23 software. The datasets were analyzed using unsupervised clustering analysis with K=3, where K represented the expected number of ancestral populations in the 3-breed rotational crossbreeding systems. Average global ancestry was determined for each extreme trait performance group with significance in ancestry variation between performance groups assessed using a T-test.

Local ancestry. The pipeline established by Martin *et al.* (2017) was used in developing the local ancestry inference. Crossbred individuals were first phased based on their respective reference population using SHAPEIT2 (v.2) before all haplotypes were merged using custom scripts. These combined phased haplotypes were used as input into the PopPhased version of RFMix v.1.5.471 with the following flags: -w 0.2, -e 1, -n 5, – use-reference-panels-in-EM, and – forward-backward EM. Node size was set to five to reduce bias in random forests resulting from unbalanced reference panel sizes (HOL panel n=92, MON panel n=34, and VKR panel n=63). Several methods were assessed to determine the best representation of local ancestry for each extreme performance group. The method shown here calculated the number of occurrences of the particular breed ancestry for each SNP within each performance group. The ancestry having the highest number of occurrences was considered the ancestry origin for a SNP within a specific performance group.

Results & discussion

Global ancestry composition. Admixture analysis was run to assess the global ancestry composition of ProCROSS and Grazecross. The admixture plots (Figure 1) for both populations revealed proportions of ancestral breeds between 23 and 34%, expected with the 3-breed crossbreeding system. Comparing the average of global ancestry composition between high- and low-performance animals showed that HOL composition plays a significant role in MY and FY in ProCROSS populations, while in the Grazecross population, VKR and NOR composition play a significant role in SCS performance (*P*-value<0.05). This finding provided the basis for understanding the impact of breed-specific ancestry on the performance of crossbred cattle.

Local ancestry. Local ancestry karyograms representing each extreme group were plotted for significant production traits in each admixed population (Figure 2) so that a comparison can be made to identify several potential regions correlated to the performance. A set of breed-specific QTL was extracted from

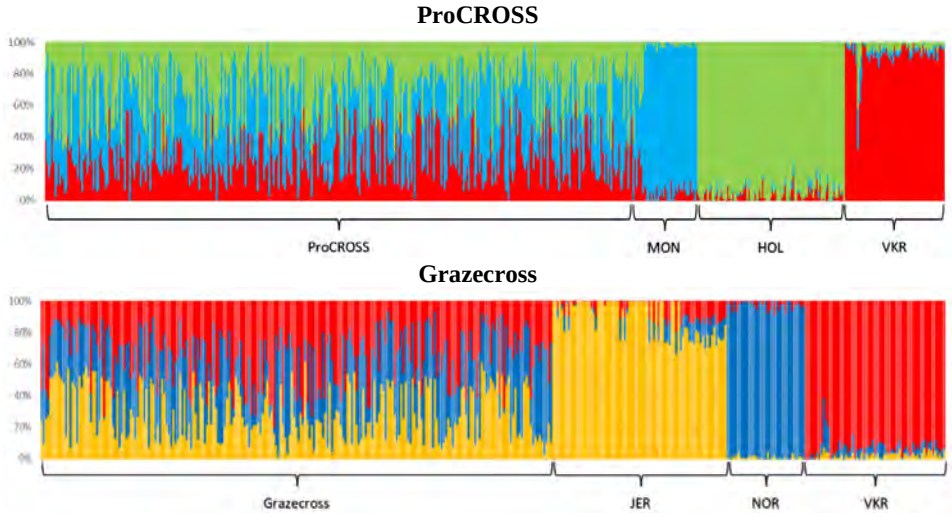


Figure 1. Admixture analysis reflecting global ancestry of breeds for both ProCROSS and Grazecross admixed individuals and all parental populations for K=3.

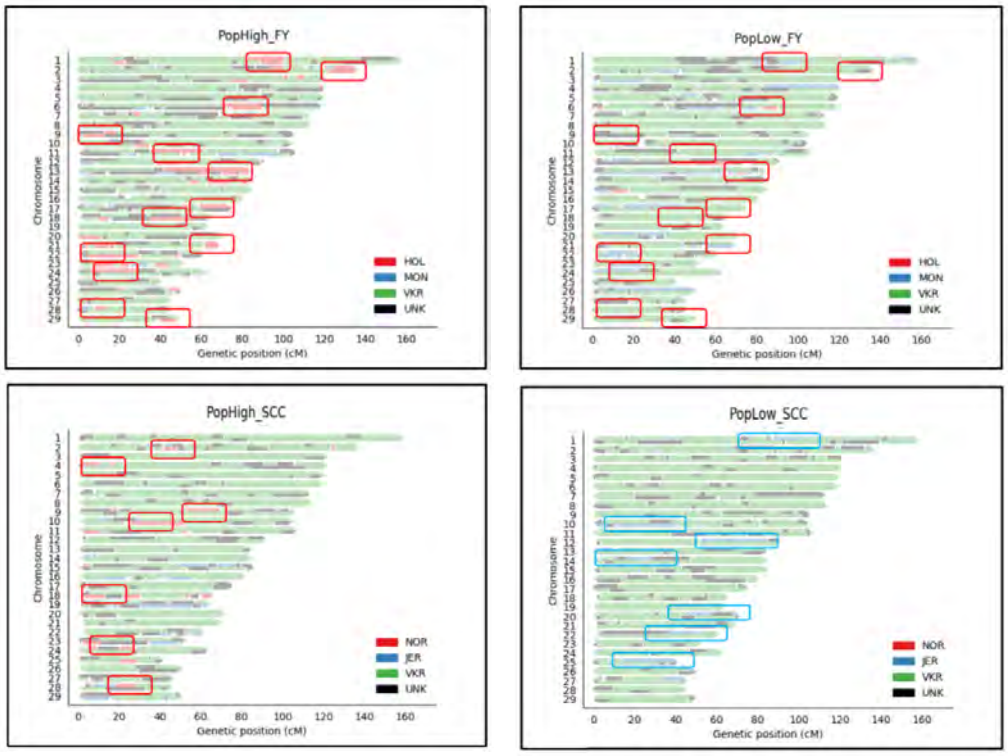


Figure 2. Local ancestry analysis compares high-performance groups (Left side diagrams) and low-performance groups (Right side diagrams) for both ProCROSS (Top) and Grazecross (Bottom) populations.

the Animal QTL database (https://www.animalgenome.org) and several recent publications. Higher ProCROSS FY production animals more commonly have visually larger HOL ancestry on BTA 2, 6, 13, 22 and 23, where genes for Prion, Lactotransferrin, and Cytochrome P450 reside and are denoted by the red box in Figure 2 (top). On the other hand, we found evidence of an excess proportion of VKR ancestry (red box) together with a decrease of Jersey ancestry (blue box) on BTA 1, 10, 12, 14 and 25 at known health QTL for high- and low-performance Grazecross animals respectively (Figure 2 bottom). These findings agreed with the global ancestry results but warrant further investigation to determine if the local ancestry variation visually seen is significantly different between performance groups. In addition, approaches for defining local ancestry for a performance group is still being assessed so that haplotype blocks define group ancestry as opposed to individual SNPs.

References

Ding L., Wiener H., Abebe T., Altaye M., *et al.* (2011) BMC Genom., 12, 622. https://doi.org/10.1186/1471-2164-12-622

Gautier M., Laloë D., and Moazami-Goudarzi K. (2010). PloS One, 5(9), e13038. https://doi.org/10.1371/journal.pone.0013038

Iso-Touru T., Tapio, M., Vilkki, J., Kiseleva T., *et al.* (2016). Animal Genetics, 47(6), 647-657. https://doi.org/10.1111/age.12473

Martin A. R., Gignoux C. R., Walters R. K., Wojcik G. L., *et al.* (2017). Am. J. Hum. Gen., 100(4), 635-649. https://doi.org/10.1016/j.ajhg.2017.03.004

Upadhyay M. R., Chen W., Lenstra J. A., Goderie C. R. J., MacHugh D. E., *et al.* (2017). Heredity, 118(2), 169-176. https://doi.org/10.1038/hdy.2016.79

182. DNA pooling is a cost effective method of including commercial crossbred data in selection of purebreds

M.N. Aldridge[1*], J. Marjanovic[1], J.M. Henshall[2], B. de Klerk[2], K. Peeters[3] and Y. de Haas[1]

[1]Wageningen University & Research, P.O. Box 338, 6700 AH Wageningen, the Netherlands; [2]Cobb Vantress B.V. Koorstraat 2, 5831 GH Boxmeer, the Netherlands; [3]Hendrix Genetics B.V., P.O. Box 114, 5830 AC Boxmeer, the Netherlands; michael.aldridge@wur.nl

Abstract

It is generally economically unviable to individually genotype large quantities of commercial crossbred broilers to predict purebred breeding values for crossbred performance. Pooling DNA of crossbreds with similar phenotypes, and genotyping the pool, could be a cost-effective alternative. To test this hypothesis, we used a dataset from a broiler three-way cross experiment consisting of ~9,000 individual genotypes and constructed pools of four different sizes (5, 10, 25, and 50 individuals). Estimated SNP effects and predicted sire breeding values from pooled genotypes were compared to results from individual genotypes, where number of individual genotypes was equal to number of pools. The pool size of 50 and 58 pools had a reasonable accuracy and bias (0.45 and 0.76, respectively) compared to 58 genotyped individuals (0.14 and 0.24). Our results indicate that DNA pooling may be used as a cost-effective means to obtain information on commercial crossbreds for selection in purebreds.

Introduction

Production, health, and welfare traits are routinely recorded on commercial crossbred broilers. Due to cost and logistics, it is not viable to link information from individual commercial crossbred animals to purebred selection candidates. Nevertheless, it is desirable to include these traits in selection of purebreds for crossbred performance. Particularly traits measurable at point of slaughter, observable only on animals raised in a commercial environment, or traits for which genetic correlation between purebred and crossbred performance is low. A potential solution is DNA pooling, where individuals with a similar phenotype are grouped together, and then genotyped as one sample. DNA is extracted from each individual, and amplified to a known concentration. A pool is formed by combining equal contributions of DNA, and then the pool is genotyped. In our previous study we showed that the genotypes of DNA pools are a good representation of individuals within the pool (Marjanovic *et al.*, 2020). DNA pooling can be used to estimate sire breeding values (GEBVs) by building a pseudo genomic relationship matrix (Bell *et al.*, 2017). However, this approach would be difficult to implement with complex crossbreeding structures, as relationships between pools and each sire would be relatively equal. Alternatively, SNP effects could be estimated, from which GEBVs of purebred selection candidates can be predicted. Our objective was to determine if DNA pooling of crossbred genotypes and estimation of SNP effects could be used to accurately and unbiasedly predict GEBVs for crossbred performance in purebred animals.

Materials & methods

Available data. Cobb Vantress provided a dataset from a broiler three-way cross experiment. From a single generation and across five batches, 156 genotyped purebred sires (A) were mated with 1,027 crossbred dams (BC). In total there were 9,262 crossbred progeny (A(BC)) that had a body weight recorded at 35 days after hatching (BW35) and each crossbred animal was genotyped (50,961 SNPs). The crossbred progeny were raised in an environment with 'commercial-like' properties, and each batch was divided across five pens.

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_182, © M.N. Aldridge *et al.* 2022

Sampling scenarios. In total, 40 contemporary groups were formed based on the fixed effects batch×pen×sex×age, with between 108 and 408 individuals in each. The majority of a sire's offspring within a batch were housed in the same pen, and each pen represented multiple sires, this means sires were not equally distributed across all contemporary groups. Each contemporary group was divided into a high end and a low end (top and bottom 25% performing animals based on BW35, respectively). From the high end of one contemporary group, five individuals were randomly sampled, forming one high pool. Each individual could only appear in one pool, and a pool would only be formed if it could be filled completely (each pool must have 5 individuals). This sampling was repeated for the low end and all contemporary groups. In total, 890 pools were formed (445 high pools and 445 low pools). Pool sizes of 10, 25, and 50 were tested with the same sampling method, which resulted in 426, 158, and 58 pools respectively. For each pool a pseudo genotype was created by calculating the allele frequency of individual genotypes in that pool (pools were not actually genotyped). The phenotype for the corresponding pool was calculated as the mean BW35 for the individuals within the pool. To compare pooling and individual genotyping, three sampling methods were tested. Individuals were selected randomly from the full population (Full), randomly from animals in the high end (High only), or randomly from animals in either the high or low end (High/Low). The number of individuals genotyped was equal to the total number of pools. For the larger individual sampling sizes (890, 426, 158) each contemporary group was represented. In addition, one analysis used all 9,262 individual crossbred genotypes as the golden standard.

Estimating SNP effects. Previously parameter estimates for BW35 were estimated with the same dataset (Duenk *et al.* 2019). They were used as input for the program MiXBLUP with the 'hpblup solver', which was used to estimate SNP effects (Ten Napel *et al.*, 2017). For scenarios that included pools, the pseudo genotypes were treated as if they were imputed genotypes. For scenarios that used the individual genotypes (no pooling), the actual genotypes were used. For the 156 line A sires, the SNP genotypes were used to predict GEBVs following Meuwissen (2001). For each scenario with pooling (5, 10, 25, and 50) and individual genotyping (Full, High only, High/Low), the sampling of individuals and estimation of SNP effects and GEBVs were repeated ten times. In addition, GEBVs were estimated using SNP effects for the golden standard.

Accuracy, bias, and cost. Accuracy and bias were calculated for each scenario and presented as means of the ten replicates. Accuracy was estimated as a weighted correlation between the predicted sire GEBV and the mean offspring performance (Calculated from the full dataset for each sire). Bias was calculated using R (Team R Core, 2013), as a weighted regression of the mean offspring performance on the sire GEBVs, and multiplied by two (for an expectation of 1.00, rather than 0.5 of just the sire). Weighted correlations and regressions were used because some of the sires had significantly more crossbred offspring than others. The weights used, were reliabilities of corrected offspring performance, estimated following Cameron (1997), as:

$$(1/4nh^2) / (1+1/4(n\text{-}1)h^2), \quad (1)$$

where n is the number of offspring and h^2 is the estimated heritability.

The cost for each sampling method, was set at € 28 per sample to be genotyped and an additional 20 cents per DNA extraction.

Results

When the full population of crossbred progeny (9,262) were genotyped, the highest accuracy (0.88) and no bias (1.00) were achieved, but at the highest cost (€ 259,336). With less than 10% of those costs, a reasonable accuracy (0.62) and bias (0.98) was achieved by genotyping 890 individuals (Table 1), but only if genotyped animals were sampled from both the high and low performing end (High/Low). Sampling from the Full population or just the High end, reduced accuracy (Full: 0.49 and High: 0.29) and increased bias (Full: 0.85 and High: 0.51). The High/Low sampling was optimal regardless of the number of individuals genotyped. Therefore, the comparison between pooling and individual sampling is made with High/Low.

With a pool size of 5 (4,450 individuals in 890 pools), the accuracy and bias (0.51 and 0.87) was lower than High/Low sampling. The pool size of 10 (4,260 individuals in 426 pools) outperformed individual High/Low sampling for both accuracy (Pool size 10: 0.49 and High/Low 426: 0.42) and bias (Pool size 10: 0.83 and High/Low 426: 0.74). Due to the cost of extracting additional DNA samples, the cost of 426 pools (€ 12,780) is slightly higher than individual genotyping (€ 11,928). With large pool sizes (25 and 50) and small number of pools (158 and 58), pooled sampling significantly outperformed accuracy and bias of the equivalent number of individual genotypes, at a lower cost (€5,214 and €2,204). With pools of 25, the accuracy was higher (Pool size 25: 0.48 and High/Low: 0.24) and bias lower (Pool size 25: 0.82 and High/Low: 0.42) than 158 individual genotypes. The extreme pooling of 50 individuals in 58 pools had a higher accuracy than 58 individual genotypes (Pool size 50: 0.45 and High/Low 58: 0.14) and less bias (Pool size 50: 0.76 and High/Low 58: 0.24).

Discussion

Pooling individuals based on phenotypic performance within contemporary groups, and then estimating SNP effects from the DNA pools, is an effective method of estimating GEBVs for a sire line. This is in line with the results of Bell *et al.* (2017) with a different approach where a pseudo G-matrix is formed instead. If there is sufficient budget, individual genotyping is preferred. However, to maximise accuracy and reduce bias, animals from both the low and high performing end should be included. Pool size and the number of pools had limited impact on accuracy and bias. The purpose of DNA pooling is cost saving, when the budget is limited, maximising pool size and minimizing number of pools is preferential.

Table 1. Estimated accuracy, bias, and economic cost of individual or pooled genotyping.

Sampling method	Number of individuals	Accuracy[1]	Bias[2]	Cost per sample genotyped	Number of samples genotyped	Total cost
Population	9,262	0.88	1.00	€ 28	9,262	€ 259,336
Full 890	890	0.49	0.85	€ 28	890	€ 24,920
High 890	890	0.29	0.51	€ 28	890	€ 24,920
High/Low 890	890	0.62	0.98	€ 28	890	€ 24,920
High/Low 426	426	0.42	0.74	€ 28	426	€ 11,928
High/Low 158	158	0.24	0.42	€ 28	158	€ 4,424
High/Low 58	58	0.14	0.24	€ 28	58	€ 1,624
Pool size 5	4,450	0.51	0.87	€ 29	890	€ 25,810
Pool size 10	4,260	0.49	0.83	€ 30	426	€12,780
Pool size 25	3,950	0.48	0.82	€ 33	158	€ 5,214
Pool size 50	2,900	0.45	0.76	€ 38	58	€ 2,204

[1] SE for individual genotyping methods ranged between 0.01 and 0.03, and for pooling were all <0.01.

[2] SE for individual genotyping methods ranged between 0.01 and 0.04, and for pooling were all <0.01.

It is worth noting that we formed pseudo genotypes as allele frequencies, rather than actually genotyping pools as it would be required in practice. Previously we showed small differences depending on how the pseudo genotype is calculated and some differences compared to real genotyping of pools (Marjanovic *et al.*, 2020). Therefore, our method of obtaining pseudo genotypes may have influenced our findings. Furthermore, our cost analysis only considered the additional cost of extracting DNA, while there would be other additional costs. Finally, we acknowledge that the accuracies are likely to be spurious (important for EBVs of the high performing end), as the relationship between sires and crossbred progeny are likely to be different across strategies. A cross-validation will be performed to confirm our findings.

Regardless of limitations, in the most extreme scenario (58 pools with 50 individuals in each pool), the accuracy and bias outperformed most individual genotyping scenarios. The accuracy and bias from 58 pools suggests substantial genetic progress could be made using predicted GEBVs in the sire line. On the other hand, selection based on information from 58 individual genotypes would result in poor outcomes due to low accuracy. DNA pooling and estimating SNP effects is a promising method for situations with limited resources but there is a desire to include traits in selection, which is often the case with commercial crossbreds.

Funding

This study was financially supported by the Dutch Ministry of Economic Affairs (TKI Agri and Food project 16022) and the Breed4Food partners Cobb Europe, CRV, Hendrix Genetics and Topigs Norsvin.

References

Bell A.M., Henshall J.M., Portro-Neto L.R., Dominik S., McCulloch R., *et al.* (2017) Genet. Sel. Evol. 49(28):1-7. https://doi.org/10.1186/s12711-017-0303-8

Cameron N.D. (1997) Selection indices and prediction of genetic merit in animal breeding. Wallingford: CAB International; 1997. pp. ix + 203.

Duenk P., Calus M.P.L., Wientjes Y.C.J., Breen V.P., Henshall J.M., *et al.* (2019) Genet. Sel. Evol. 51(6):1-11. https://doi.org/10.1186/s12711-019-0447-9

Marjanovic J., de Klerk B., de Haas Y., Dibbits B.W., and Aldridge M.N. (2020) Proc. of the 71st Annual Meeting of the European Federation of Animal Science, Online conference.

Meuwissen T.H.E., Hayes B.J., and Goddard M.E. (2001). Genetics 157:1819–1829. https://doi.org/10.1093/genetics/157.4.1819

Team R Core. (2013). R: A language and environment for statistical computing. http://r.meteo.uni.wroc.pl/

Ten Napel J., Vandenplas J., Lidauer M., Stranden I., Taskinen M., *et al.* (2017) MiXBLUP, user-friendly software for large genetic evaluation systems – Manual Wageningen, the Netherlands V2.1-2017-08. http://www.mixblup.eu/documents/Manual%20MiXBLUP%202.1_June%202017_V2.pdf

Part 13

Challenges – understanding and utilizing genetic diversity

183. Genetic variability in the individual inbreeding load: genomic prediction for artificial purging

L. Varona[1*], D. López-Carbonell[1], H. Srihi[1], M. Ramírez[1], P. Martínez[2], M. Hermida[2], J. Altarriba[1] and J. Casellas[3]

[1]Instituto Agroalimentario de Aragón (IA2). Facultad de Veterinaria. Universidad de Zaragoza. 50013. Zaragoza, Spain; [2]Facultad de Veterinaria, Universidad de Santiago de Compostela, 27002 Lugo, Spain; [3]Universitat Autònoma de Barcelona, 08193 Bellaterra, Spain; lvarona@unizar.es

Abstract

Inbreeding is caused by mating between related individuals and it is associated with inbreeding depression. Recently, some models have been described to predict individual inbreeding depression loads with the aim of avoiding undesirable mating and to perform artificial purging. In this study, those models were implemented using a single-step GBLUP approach with two main objectives: (1) to evaluate the increase of accuracy of the prediction of inbreeding loads thanks to genomic information; and (2) to use the equivalence between GBLUP and SNP-BLUP to estimate the SNP effects and to identify the genomic regions associated with the largest variance of inbreeding loads. The results of this study revealed that the increase of accuracy was only relevant for genotyped individuals, and that the inbreeding load variance is heterogeneously distributed along the genome, indicating that the procedure may be adequate to detect potential candidate genes which explain the phenomenon of inbreeding depression.

Introduction

Inbreeding is caused by mating between related individuals and it is associated with inbreeding depression, understood as a reduction in the phenotypic yield (Charlesworth and Willis, 2009) that has been widely observed in animals, plants and humans (Charlesworth and Willis, 2009; Leroy, 2014). Inbreeding depression is brought about by the higher degree of homozygosity in inbred individuals that lose the advantage of over-dominance at heterozygous loci or increase the probability of the presence of recessive alleles. Over the last few years, some models (Casellas, 2018, Varona *et al.*, 2019) have been developed for the prediction of individual hidden inbreeding depression loads that can be used to predict the expected inbreeding depression of each potential inbred cross in the population and to perform individual purging by selecting those with the best inbreeding loads. One limitation to this approach is that most of the inbreeding is caused by a small number of ancestors. Therefore, the accuracy achieved for most of the individuals in the population is very low. However, the approach described above has been developed using only genealogical and phenotypic information. Here we include the genotypic information provided by SNPs device using a single-step approach (Aguilar *et al.*, 2010) with the aim of: (1) evaluating the increase of accuracy for the prediction of the inbreeding loads of younger genotyped animals; and (2) identifying the genomic regions associated with the largest variance of inbreeding loads in order to disentangle the genetic basis of the inbreeding depression.

Materials & methods

Data. The dataset used in this analysis included 91,412 records for cold carcass weight and a pedigree of 141,399 individual-sire-dam entries for the Rubia Gallega beef cattle breed. The average weight was 223.5 kg. with a standard deviation of 40.38 kg. Additionally, 3,441 individuals were genotyped with Axiom Bovine platform (*Axiom_BovMDv3*) from ThermoFisher (Waltham, MA USA). After a standard filtering (95% of individual and population call rate and MAF>0.5), 36,429 SNP for the autosomal chromosomes were used.

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_183, © L. Varona *et al.* 2022

Inbreeding. The average inbreeding in the 91,412 phenotyped individuals was 0.027. The great majority of them has a null (35,572) or below 0.05 (42,222) inbreeding coefficient. However, 12,608 of them have an inbreeding coefficient greater than 0.05 and 2,015 of them larger than 0.25. The maximum inbreeding coefficient was 0.46875. The total inbreeding coefficient was decomposed using the Mendelian decomposition proposed by Caballero and Toro (2000) as well as the algorithm described by García-Cortés *et al.* (2010). We were able to decompose the total inbreeding in 1,646,121 partial inbreeding coefficients generated by 2,136 ancestors. The contributions of these ancestors to the total inbreeding were very heterogeneous as 104 ancestors contributes to more than 1000 inbred descendants. Among the ancestors generating inbreeding 475 of them were genotyped.

Model. The model of analysis was:

$$\boldsymbol{y} = \boldsymbol{f}g + \boldsymbol{t}c + \mathbf{Xb} + \mathbf{Wh} + \mathbf{Zu} + \mathbf{Ki} + \boldsymbol{e} \quad (1)$$

Where y is the vector of phenotypic records, *g* is a covariate with total inbreeding, *c* is a covariate with age of recording, b is the vector of systematic effects than includes sex -2 levels- and slaughterhouse –29 levels-, h is the random herd-year-season effects with 14,687 levels, u is the vector of the additive genetic effects, i is the vector of individual inbreeding loads and e is the vector of residuals. The vectors f and t include the genealogical inbreeding and the age of recording, respectively. Lastly, X, W, Z and K are the corresponding incidence matrices. The prior distribution of the herd-year-season was univariate Gaussian and the prior distribution of the additive breeding values was defined with two alternatives: model 1) $\begin{pmatrix}\boldsymbol{u}\\ \boldsymbol{i}\end{pmatrix} \sim N\left(\begin{matrix}0\\ 0\end{matrix}, \boldsymbol{G} \otimes \boldsymbol{A}\right)$ and model 2) $\begin{pmatrix}\boldsymbol{u}\\ \boldsymbol{i}\end{pmatrix} \sim N\left(\begin{matrix}0\\ 0\end{matrix}, \boldsymbol{G} \otimes \boldsymbol{H}\right)$ with $\boldsymbol{G} = \begin{pmatrix}\sigma_u^2 & \sigma_{ui}\\ \sigma_{ui} & \sigma_i^2\end{pmatrix}$. σ_u^2 is the additive genetic variance, σ_i^2 is the inbreeding load variance and σ_{ui} is the covariance between them. Furthermore, A is the numerator relationship matrix and H is the matrix that combines the genomic relationship matrix with A as described by Aguilar *et al.* (2010). The analysis was performed through a Bayesian approach with the Gibbs sampler using a single long chain of 500,000 iterations, having discarded the first 50,000.

Once the analysis was performed, the accuracy (r_i) of the predicted inbreeding loads was approximated in both models by:

$$r_i = \sqrt{\frac{(1+F_i)\hat{\sigma}_i^2 - \mathrm{PSD}(i_i)^2}{(1+F_i)\hat{\sigma}_i^2}} \quad (2)$$

Where F_i is the total inbreeding coefficient of the i^{th} individual and $PSD(i_i)$ is the posterior standard deviation of its inbreeding load. Lastly, we used the predictions of breeding values under model 2 to estimate the SNP effects for the individual inbreeding load following Wang *et al.* (2012). We later used the predictions of SNP effects to estimate the percentage of the total variance explained by segments of 50 consecutive SNP along the autosomal chromosomes.

Results

The estimates of the variances components with model 1 and 2 were very similar. The posterior means (and posterior standard deviations) estimates under model 2 were -56.66 (3.50), 731.55 (17.47), 6,829.71 (2,005.91), -1,531.59 (263.38), 180.79 (4.25) and 223.50 (10.57) for *g*, σ_u^2, σ_i^2, σ_{ui}, σ_h^2 and σ_e^2 respectively. The predictions of the individual inbreeding loads under model 2 are presented in Figure 1, where they are compared with minus the posterior mean estimate of the average inbreeding depression (*-g*).

It must be noted that the model provides estimates of the inbreeding loads that must be understood as the inbreeding depression achieved by a fully inbred descendant. It is worth to highlight that 9.92% of individuals get a prediction of inbreeding load bigger than *–g*, suggesting that inbreeding generated by them will increase the phenotypic performance of their inbred descendants.

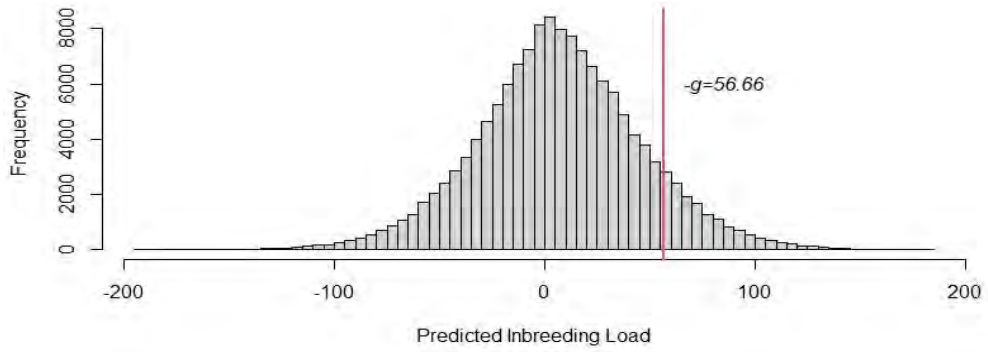


Figure 1. Distribution of the prediction of inbreeding loads compared with minus the posterior mean estimate of the average inbreeding depression (in red).

Later, we investigated the accuracy of the predicted inbreeding loads under models 1 and 2. The average repeatability achieved for all individuals in the pedigree was very similar (0.558 vs 0.562). Only an increase of 2.6% was detected for the genotyped individual (0.606 vs 0.632). Moreover, the difference increased up to 3.7% (0.548 vs 0.585) for genotyped individuals born after 2016.

Finally, Figure 2 presents a Manhattan plot of the percentage of the inbreeding load variance explained by genomic segments of 50 consecutive SNP. The distribution of the inbreeding load variance is heterogeneous along the autosomal genome, and two genomic regions in chromosomes 21 and 28 can be highlighted.

Discussion

The model proposed by Varona *et al.* (2019) allows to estimate the individual inbreeding loads using genealogical and phenotypic information. In this study, it is confirmed that the model is able to detect genetic variation on the inbreeding load and to predict phenotypic consequences in future inbred individuals. Those predictions can be used to avoid undesirable mating and to perform artificial purging by selecting the individuals with the best inbreeding loads. Moreover, we proved that the genomic information can be used to improve the accuracy in the predicted inbreeding loads. However, the use of genomic information has only provided a relevant increase of accuracy in genotyped individuals, with more evidence in young

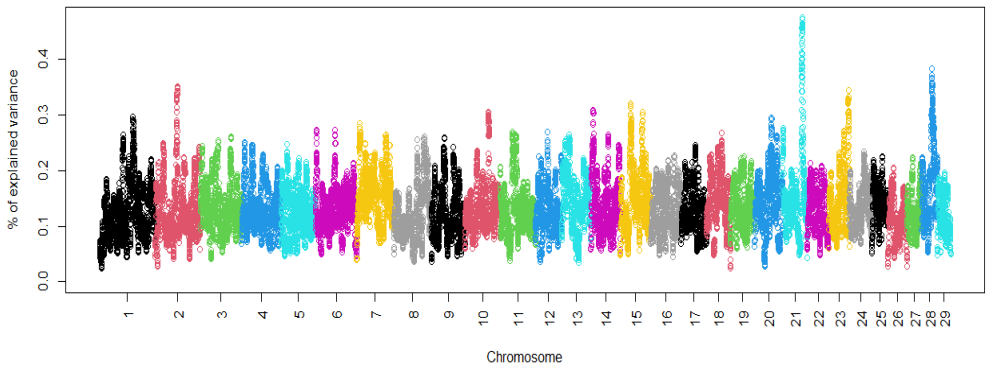


Figure 2. Manhattan plot of the percentage of the inbreeding load variance explained by genomic segments of 50 SNP.

individuals that have no inbred descendent. Additionally, the equivalence between GBLUP and SNP-BLUP models (Wang *et al*., 2012) can be used to estimate the SNP effects associated with the inbreeding load and to calculate the amount of inbreeding load variance explained by each specific genomic segment. As an illustrative example, we calculated the amount of variance explained by genomic segments defined by 50 consecutive SNP and we were able to identify two genomic regions in chromosomes 21 and 28. Nevertheless, these results must be taken as preliminary, and further research has to be done to define more precisely the genomic regions and to identify potential candidate genes within these regions.

Acknowledgements

The authors are grateful to ACRUGA (Asociación de Criadores de Ganado Vacuno Selecto de Raza Rubia Gallega) for providing the data, pedigree and genotypic information. This research was funded by FEADER, grant number 2019/053A.

References

Aguilar I., Misztal I., Johnson D., Legarra A, Tsuruta S, Lawlor T. (2010). J. Dairy Sci. 93; 743–752, https://doi.org/10.3168/jds.2009-2730

Caballero A., and Toro M. A. (2000). Genet. Res. 75:331–343. https://doi.org/10.1017/S0016672399004449.

Casellas J. (2018). J Anim Breed Genet. 2018;135:37–44. https://doi.org/10.1111/jbg.12308.

Charlesworth D, and Willis JH. (2009). Nat. Rev. Genet.;10:783–796. https://doi.org/10.1038/nrg2664.

García-Cortés L. A., Martínez-Ávila J. C., and Toro M. A. (2010). Conserv. Genet. 11:1945-1952. https://doi.org/10.1007/s10592-010-0084-x

Leroy G. (2014). Animal Genetics. 2014;45:618–628. https://doi.org/10.1111/age.12178

Varona L., Altarriba, J., Moreno, C., Martínez-Castillero, M., Casellas, J. (2019). Genet. Sel. Evol. 51:78. https://doi.org/10.1186/s12711-019-0521-3

Wang H, Misztal I, Aguilar I, Legarra A., Muir, W.M. Genet. Res. 2012, 94, 73–83. https://doi.org/10.1017/S0016672312000274

184. Inbreeding depression in livestock: comparing trait groups and inbreeding measures

H.P. Doekes*, P. Bijma and J.J. Windig

Wageningen University & Research, P.O. Box 338, 6700 AH Wageningen, the Netherlands; harmen.doekes@wur.nl

Abstract

Inbreeding depression has been documented for many livestock populations. We compared inbreeding depression estimates (1) between trait groups, and (2) between different pedigree-based and genomic inbreeding measures. Inbreeding was found to affect all sorts of traits to a similar extent, independent of heritability and of whether a trait is a primary fitness trait or not. Inbreeding depression estimates were moderately to highly correlated between different pedigree-based and genomic inbreeding measures, but those based on genomic measures tended to have lower *P*-values, suggesting more power to detect inbreeding depression with genomic measures (assuming there is a true inbreeding depression effect). Among different genomic measures (based on genomic relationship matrix, regions of homozygosity, or percentage of homozygous markers), there were no consistent differences in *P*-values.

Introduction

Inbreeding reduces the mean performance of animals, a phenomenon known as inbreeding depression. Based on theory and laboratory experiments, inbreeding mostly affects primary fitness traits such as survival and fertility (Lynch and Walsh, 1998). These traits are under directional selection and the average dominance effect is expected to be favourable, because fixation occurs more quickly for loci with an unfavourable dominance effect (Lynch and Walsh, 1998). For livestock populations, one could argue that traits related to e.g. production may equally qualify as fitness traits, since there has been (artificial) selection for these traits for many generations. Hence, one may expect a similar degree of inbreeding depression for a wide range of traits in livestock.

The degree of inbreeding depression is typically quantified by linear regression of phenotypes on inbreeding coefficients. Inbreeding coefficients can be estimated from pedigree data, but are increasingly derived from genomic information, particularly SNP data. An advantage of genomic measures is that they provide estimates of realized inbreeding, instead of genealogical probabilities (Wang, 2016). Therefore, using genomic inbreeding coefficients may result in more accurate estimates of inbreeding depression (Caballero *et al.*, 2021). Among SNP-based measures, some may better capture inbreeding depression than others.

Many studies have reported (pedigree-based, and more recently genomic) estimates of inbreeding depression. In this paper, we summarize our recently published meta-analysis (Doekes *et al.*, 2021) and extend it with some additional literature and analyses. The aim was to determine and compare inbreeding depression estimates between different trait groups, as well as between different pedigree-based and SNP-based approaches.

Materials & methods

Data. A literature search in Web of Science and Scopus on inbreeding depression in cattle, pig, chicken, sheep, goat, horse and rabbit resulted initially in 766 studies published between 1990 to 2020. After filtering (for details, see Doekes *et al.* 2021) 154 studies remained with inbreeding depression estimated as linear regression coefficients (*b*-values), i.e. as the change in phenotypic mean per 1% inbreeding. From these studies, 2321 *b*-values were extracted. For each trait, the favourable phenotypic direction was determined,

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_184, © H.P. Doekes *et al.* 2022

and the sign of *b*-values was changed if needed to ensure that negative values reflected unfavourable effects. Estimates for traits with an optimum or an unclear favourable direction were removed (n=197).

Comparison between trait groups. Traits were divided into six groups: reproduction/survival, weight/growth, production, conformation, health and other traits (details on classification, see (Doekes *et al.* 2021)). For the comparison between traits and trait groups, *b*-values were divided by the corresponding trait mean (to get b_m) or trait standard deviation (to get b_s) and only pedigree-based estimates were used (n=1,818 for b_m and n=1,259 for b_s). A general linear model was run, in which a population effect (combined effect of study and breed) was fitted in addition to the trait group effect. Estimated marginal means (EMMs) for the different trait groups were obtained and compared. In addition, h^2-estimates were extracted when available (n=490), compared across trait groups and correlated to b_m- and b_s-estimates.

Comparison between inbreeding measures. Inbreeding depression estimates were compared within 11 studies and within traits between pedigree inbreeding (F_{PED}), inbreeding based on ROH (F_{ROH}), inbreeding derived from a GRM (F_{GRM}, from VanRaden's method 2 and Yang's method), inbreeding derived from a GRM computed with allele frequencies of 0.5 ($F_{GRM0.5}$) and the percentage of homozygous SNPs (*HOM*). To account for scale differences of inbreeding measures, b_m and b_s were standardized, by multiplying them by the SD of the inbreeding measure in the corresponding population, such that they were expressed per 1 SD increase in inbreeding (rather than per 1%). Significance of *b*-values was also compared across inbreeding measures, with smaller *P*-values indicating a more significant association between inbreeding and phenotypic value.

Results

Across all studies and traits, the mean (median) pedigree-based b_m and b_s were -0.22 (-0.13) and -0.71 (-0.59), respectively. Distributions for b_m and b_s were roughly similar for different trait groups (Figure 1). The mean and median b_m and b_s were negative for each trait group, except for 'other traits'.

When correcting for population and ignoring the 'other traits', the EMM of b_m was most negative for production traits (-0.31) and reproduction/survival traits (-0.30), and least negative for conformation traits (-0.14). For b_s, the EMM was most negative for weight/growth traits (-1.07) and least negative for

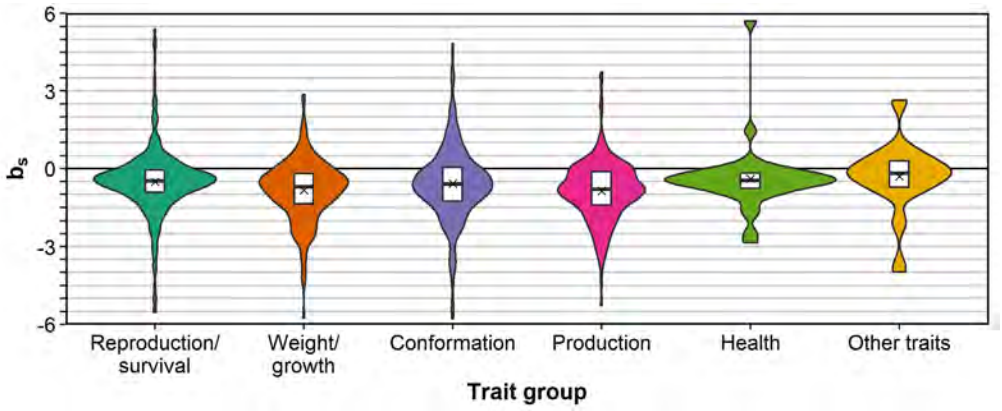


Figure 1. (adapted from Doekes *et al.*, 2021) Violin plots of inbreeding depression estimates per trait group, expressed as percentage of a trait's SD (b_s). Boxplots and the mean (×) are also shown. There were 39 outliers outside the range of this figure.

reproduction/survival traits (-0.41). Pairwise differences of EMMs across trait groups were generally not significant.

Heritability estimates were on average 0.07 for reproduction/survival traits (n=141), 0.30 for weight/growth traits (n=126), 0.29 for production traits (n=109), 0.27 for conformation traits (n=95) and 0.17 for health traits (n=9). Correlation coefficients between inbreeding depression estimates and heritability estimates were close to zero (Figure 2).

Standardized b_m- and b_s-estimates generally showed moderate to (very) high correlations between different inbreeding measures. Correlations of b-values between $F_{GRM0.5}$ and *HOM*, between F_{PED} and F_{GRM}, between F_{ROH} and *HOM*, and between F_{GRM} and $F_{GRM0.5}$ were all ≥0.9. Correlations were lowest between F_{ROH} and F_{GRM} (0.27 for b_m and 0.52 for b_s) and between F_{GRM} and *HOM* (0.24 for b_m and 0.47 for b_s).

Inbreeding depression estimates for SNP-based inbreeding measures had lower *P*-values than estimates based on pedigree inbreeding in more than 80% of direct comparisons (Table 1). Among the SNP-based measures, there were no consistent trends, i.e. which measure had the lowest *P*-value largely differed across studies and also across traits within studies.

Discussion

The mean (median) b_m of -0.22% (-0.13%) and b_s of -0.71% (-0.59%) are similar to the mean b_m of -0.14% and b_s of -0.56% that were previously reported by Leroy (2014) for a subset of 57 studies. Across trait groups, there were some small differences, but those differences were not consistent for b_m and b_s and did not suggest that primary fitness traits such as survival and reproduction exhibit more inbreeding depression than other traits.

Inbreeding depression estimates for SNP-based inbreeding were generally more significant than those based on pedigree inbreeding. Some other studies, which were not included in our analysis, also support this finding. Baes *et al.* (2020), for example, reported more significant inbreeding depression for genomic inbreeding than for pedigree inbreeding for six traits in a sample of ~15,000 Holstein cows (genomic inbreeding was most likely F_{GRM}, though this is not explicitly stated). Lozada-Soto *et al.* (2021) recently estimated inbreeding depression for heifer pregnancy and weight/growth traits in American Angus cattle and found smaller credible intervals for F_{GRM} and F_{ROH} than for F_{PED}. Thus, SNP-based measures appear promising for accurately estimating inbreeding depression.

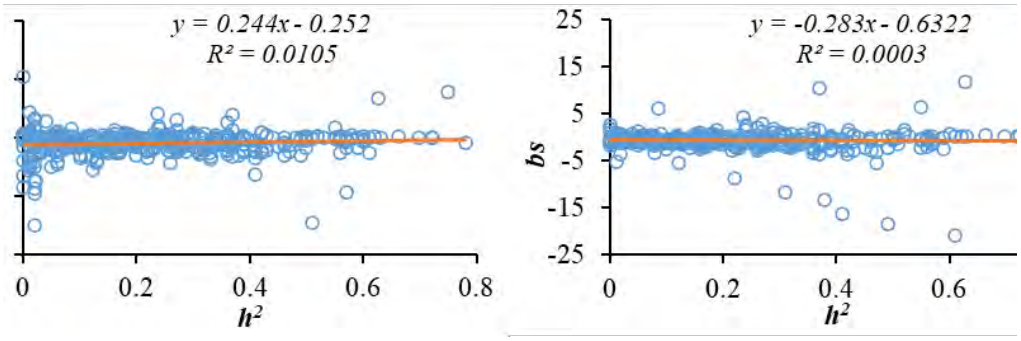


Figure 2. Relationship between inbreeding depression estimates expressed as percentage of a trait's mean (b_m; left) or SD (b_s; right) and heritability estimates (h^2).

Table 1. (adapted from Doekes *et al.*, 2021) Comparison of *P*-values of inbreeding depression estimates within 11 studies and within traits for each combination of two inbreeding measures, where both inbreeding measures had an unfavorable effect on the phenotype. There were no comparisons between F_{GRM} and $F_{GRM0.5}$ in the data set.[1]

				Comparisons where *P2*<*P1*	
Measure 1	**Measure 2**	**N studies**	**N comparisons**	**N**	**% of total**
F_{PED}	F_{ROH}	8	38	30	79
F_{PED}	F_{GRM}	1	7	7	100
F_{PED}	$F_{GRM0.5}$	4	26	21	81
F_{PED}	*HOM*	4	18	15	83
F_{ROH}	F_{GRM}	1	7	4	57
F_{ROH}	$F_{GRM0.5}$	5	40	15	38
F_{ROH}	*HOM*	4	34	19	56
F_{GRM}	$F_{GRM0.5}$	1	3	3	100
F_{GRM}	*HOM*	2	14	4	29
$F_{GRM0.5}$	*HOM*	1	20	10	50

[1] *P*1 = *P*-value of measure 1; *P*2 = *P*-value of measure 2.

Among the SNP-based measures, there were no consistent trends with regard to which measure yielded the lowest *P*-values. Caballero *et al.* (2021) performed a set of computer simulations to shed more light on the question which SNP-based measures are more accurate for estimation of inbreeding depression. They found that the answer largely depends on the population parameters. For example, while F_{ROH} is appropriate for populations with a small effective population size (in which relationships are tight and there is relatively less intense selection), it may be less appropriate for populations of larger effective population size.

Overall, we conclude that inbreeding depression in livestock populations is present for all sorts of traits to a similar extent and that SNP-based inbreeding measures tend to have more predictive power than pedigree-based measures.

References

Caballero A., Villanueva B., and Druet T. (2021). Evol. Appl. 14(2):416-428. https://doi.org/10.1111/eva.13126

Doekes H.P., Bijma P., and Windig J.J. (2021). Genes 12(6):926. https://doi.org/10.3390/genes12060926

Leroy G. (2014). Anim. Genet., 45(5):618-628; https://doi.org/10.1111/age.12178

Lozada-Soto E.A., Maltecca C., Lu D., Miller S, Cole J.B, *et al.* (2021). Genet Sel Evol 53:50. https://doi.org/10.1186/s12711-021-00644-z

Lynch M., and Walsh B. (1998). Inbreeding depression. In: Genetics and Analysis of Quantitative Traits, 1st ed. Sinauer, Sunderland, MA, USA.

Maltecca C., Tiezzi F., Cole J.B. and Baes C. (2020). J. Dairy Sci., 103(6):5302-5313. https://doi.org/10.3168/jds.2019-17846

Wang, J. (2016). Theor. Popul. Biol., 107:4-13. https://doi.org/10.1016/j.tpb.2015.08.006

185. Global and local genome ancestry reveals signal of Merinization in Noire de Thibar sheep

I. Baazaoui[1,2*], E. Ciani[3], S. Mastrangelo[4] and S. Bedhiaf-Romdhani[1,2*]

[1]Faculty of Science of Bizerte, University of Carthage, 7021 jarzouna,Tunisia; [2]National Institute of Agricultural Research of Tunisia (INRAT), University de Carthage, 2049 Ariana, Tunisia; [3]Dipartimento di Bioscienze, Biotecnologie, Biofarmaceutica, Università degli Studi di Bari 'AldoMoro', Via Amendola 165/A 70126 Bari, Italy; [4]Dipartimento Scienze Agrarie, Alimentari e Forestali, Università degli Studi di Palermo, 90128 Palermo, Italy; imenbaazaoui18@gmail.com; bedhiaf.sonia@gmail.com

Abstract

Information about patterns of genetic ancestry in admixed livestock populations provide insights on past introgression events and genomic regions under selection. The Tunisian Noire de Thibar sheep is an ideal model of composite breed between local and European stocks. Using genome-wide SNP data, we evaluated patterns of global and local ancestry in this breed. Global ancestry analysis confirmed the introgression of Merino d'Arles to improve the wool fiber quality and at a lower extent, the Black Brown Swiss, used to fix the black color. Therefore, we estimated marker average ancestry at each locus, to identify genomic regions shaped by ancestral Merinos. Surprisingly, we detected a high signal of 'Merinization' on chromosome nine harbouring the *SULF1* gene involved in the pathway of primary wool follicle induction. These results provide new insights of exotic allele's diversity in local sheep, which could be utilized for rapid introgression of novel alleles through genomics-assisted breeding.

Introduction

Information about patterns of genetic ancestry in admixed livestock populations provide insights on past introgression events and genomic regions under selection. Recent advances in genomic technologies and molecular data analysis allow for accurate detection of ancestry at the genome level by identifying genomic fractions originating from ancestral populations. There are currently two different approaches to determine ancestry inference: 'global ancestry' and 'local ancestry'. The main objective of global ancestry inference is to estimate the average ancestry contributed by each population, over the whole genome, useful for studying admixture populations. In contrast, local ancestry inference provides the same information but for each locus of the genome in order to detect ancestry at each locus (Maples *et al.* 2013). Different methods to detect local inference have been used mainly to study recent admixed populations (Tang *et al.* 2006; Sankararaman *et al.* 2008; Maples *et al.* 2013). The Tunisian Noire de Thibar sheep, also known as 'Black Thibar', a local synthetic breed, created through a reciprocal cross between an Algerian thin tailed, known as Queue Fine de l'Ouest in Tunisia and the imported Merino d'Arles (Porter *et al.* 2016), which is originated from local ewes crossed with Spanish Merino rams. In 1945, Noire de Thibar sheep was officially registered in the national Herd book and was recognized as a local breed in Tunisia. In the 1970s, a new gene pool was introduced into the breed using the Swiss Brown-Black sheep breed, mainly to avoid consanguinity, fix the black coat colour and improve meat and wool quality (Kallal 1968). Indeed, the black coat Noire de Thibar animals are tolerant to toxic '*Hypericum perforatum*' weeds, which causes skin photosensitization in white sheep, causing a major economic loss to sheep farmers (Baazaoui *et al.* 2020). Now, Noire de Thibar flocks represents only 2% of national sheep flock and are characterized by lower fat meat and high wool quality. A study of genetic structure of Tunisian sheep at global scale highlighted a clear differentiation of the Noire de Thibar breed from local sheep populations and a genetic proximity to Northern Mediterranean breeds explained by introgression of European genetic stocks (BAAZAOUI *et al.* 2018). However, this European ancestry is still unknown at the genome level, just as its contribution to shaping regions under selection. Thus, the aim of this study is to estimate global and local ancestry of Noire de Thibar breed to infer patterns of European (Merinos d'Arles and Swiss Black Brown breeds) ancestry using genome-wide SNP data.

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_185, © I. Baazaoui *et al.* 2022

Materials & methods

Animals and data genotyping. Animals were genotyped with the Illumina Ovine BeadChip (Kijas *et al.* 2009) according to the manufacturer's protocol. Raw material of 13 Tunisian Noire de Thibar (NDT) animals was genotyped in a previous study (Bedhiaf-Romdhani *et al.* 2020). SNP data of 18 French Merinos d'Arles (MER) and 23 Swiss Black Brown (SBB) were obtained respectively from previous published studies (Tortereau *et al.* 2017) and (Kijas *et al.* 2012). Genome coordinates for all SNP were based on the OAR version 3.1 ovine genome assembly.We used PLINK 1.9 (http://www.coggenomics.org/plink 2) (Purcell *et al.* 2007) for data merging and quality checks. The combined raw dataset was filtered to exclude poorly genotyped individuals (<90% of SNPs). SNP markers located on sex chromosomes, genotyped in <90% of individuals and with rare alleles (MAF<0.01). We retained 37,776 SNPs after linkage disequilibrium filtering using the following parameters: 50 SNPs per window, a shift of 5 SNPs between windows and a variation inflation factor's threshold of two (corresponding to r^2>0.5).

Genetic relationship among breeds. To better visualize the genetic relationship and complex population network of NDT animals and its three parental breeds (QFO, MER and SBB), we used a network-based approach implemented in NetView v.1.1 software (Neuditschko *et al.* 2012) using identity by state (IBS) distance matrix. Thus, we adopted an increasing K (parameter determining the number of mutual nearest neighbors) values up reaching an arbitrarily sufficient level of connectedness among samples. Global patterns of ancestry and admixture was assessed using unsupervised and supervised hierarchical clustering of NDT individuals, at K=2, with prior information given to its European ancestors (MER and SBB) implemented in the Admixture software v1.3.0 (Alexander and Lange 2011) by applying the default settings.

Inference of European local ancestry in Noire de Thibar sheep. The local ancestry was inferred using LAMP (Local Ancestry in adMixed Populations) software (Sankararaman *et al.* 2008) to estimate ancestries at each locus in a population of admixed individuals. On the light of previous results, we assumed that NDT animals were the result of crossbreeding between local QFO and MER. Thus, we estimated the individual's local ancestry of MER breed proportion in NDT animals. For each locus, LAMP analysis provides the marker average ancestry (MAA) related to considered ancestral populations (QFO and MER). The software operates on sliding windows of contiguous SNPs and assigns ancestries by combining the results with a majority vote. The following default settings were adopted: number of generations since admixture (g) = 32 and recombination rate (r) =10^{-8}. The fraction of global admixture (α) related to MER was assigned to α=0.1, using the Admixture software (Alexander and Lange, 2011). The highest MAA values related to MER are considered as significant genomic regions supposed to be under selection. We conducted a comprehensive search in the available literature and public databases to investigate the biological function of each annotated gene.

Results

Analysis of ancestry using Network-based approach (Figure 1A) has shown that MER is strictly linked to NDT and the influence of SBB is just manifested through two NDT samples at K=25 and 30, probably *via* recent crossbreeding. These findings are confirmed by results of supervised and unsupervised admixture results at K=2 (Figure 1B) where samples from NDT breed show high proportions of MER ancestry.

The results of MER marker-ancestry repartition on autosomes of NDT animals revealed that of MAA values ranged between 0.07 and 0.33. Therefore, the highest signal defined a putatively regions under selection, located on chromosome 9 spanning the genomic region 45.9-46.3 Mb including eight adjacent SNPs (MAA=0.33) harboured two candidate genes (*SULF1* and *SLCO5A1*) (Figure 2).

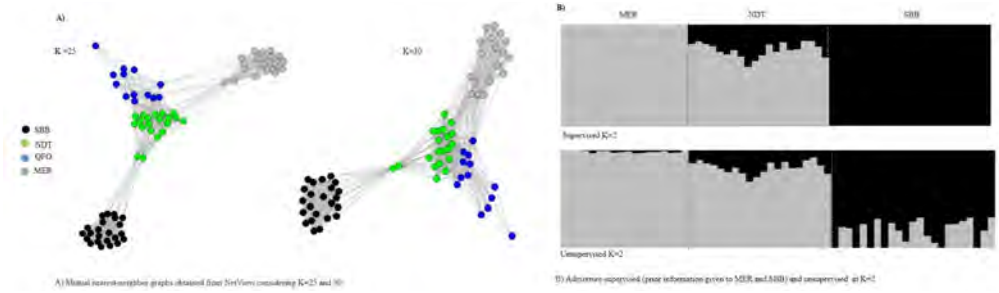


Figure 1. Inference of global ancestry of Noire de Thibar animals.

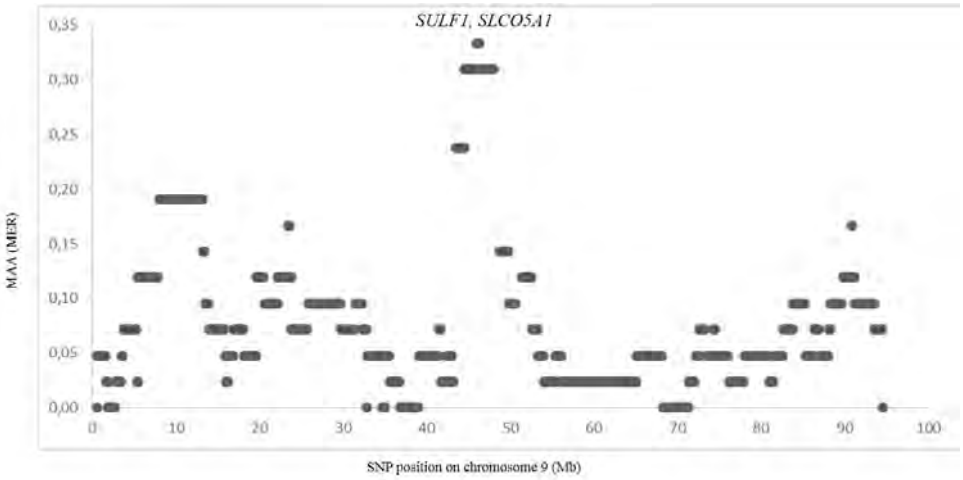


Figure 2. MAA of MER on chromosome 9.

Discussion

Our results are in line with historical origin of NDT evidenced by past introgression of Merinos gene flow through artificial selection for black coat colour (Kallal 1968). These findings could explain the Merinos ancestry from the Spanish Merino ancestors from which MER was derived (Ben Jemaa *et al.* 2019). Local ancestry analysis revealed a high signal of MER ancestry in the genomic region containing the *SULF1* gene which is a candidate gene involved in primary wool follicle induction and skin development (Nie *et al.* 2018). Indeed, *SULF1* is an enzyme of the sulfatase class able to alter the sulfation patterns of proteoglycans and consequently the binding site of many growth factors, among them WNT, BMP and FGF, which are known to play major roles during embryogenesis, and among others, during the hair follicle development. Therefore, the assumed origin of NDT as a composite breed was supported here, with clear influence of MER breed, likely introduced at the beginning of the last century to improve wool quality. Our results provide a genetic base reference to establish molecular marker-assisted selection to improve the production performance of natural coloured wool, which represent a niche market for top quality fleeces in the world. We should mention that wool from coloured sheep is a valuable genetic resource in high demand by handcrafts and wool crafters for its natural look and its freedom from dye chemicals.

References

Alexander D.H., and Lange K. (2011) BMC bioinformatics 12(1)**:** 1-6. https://doi.org/10.1186/1471-2105-12-246

Baazaoui I., Mcewan J., Anderson R., Brauning R., Mcculloch A. *et al.* (2020) Animals 10(1)**:** 5. https://doi.org/10.3390/ani10010005

Baazaoui I., Mastrangelo S., Ben Sassi M., Bedhiaf-Romdhani S., Ciani E. (2018) Proc. of 12th Convegno Nazionale Biodiversità, Ambienti, Salute, Teramo, Italy. https://www.biodiversita2018.it/

Bedhiaf-Romdhani S., Baazaoui I., Ciani E., Mastrangelo S. and Sassi M.B. (2020) Small Ruminant Research 191(106192. https://doi.org/10.1016/j.smallrumres.2020.106192

Ben Jemaa S., Kdidi S., Gdura A., Dayhum A., Eldaghayes I. *et al.* (2019) Animal Genetics 50(5)**:** 526-533. https://doi.org/10.1111/age.12831.

Kallal A. (1968) Le mouton noir de Thibar, Thèse pour le Doctorat Vétérinaire, Ecole Nationale Vétérinaire de Toulouse, Toulouse, France. p62.

Kijas J.W., Lenstra J.A., Hayes B., Boitard S., Porto Neto L.R. *et al.* (2012) PLoS biology 10(2)**:** e1001258. https://doi.org/10.1371/journal.pbio.1001258

Kijas J.W., Townley D., Dalrymple B.P., Heaton M.P., Maddox J.F. *et al.* (2009) PloS one 4(3)**:** e4668. https://doi.org/10.1371/journal.pone.0004668

Maples B.K., Gravel S., Kenny E.E. and Bustamante C.D. (2013) The American Journal of Human Genetics 93(2)**:** 278-288.

Neuditschko M., Khatkar M.S. and Raadsma H.W. (2012) PloS one 7(10)**:** e48375. http://dx.doi.org/10.1371/journal.pone.0048375.

Nie Y., Li S., Zheng X., Chen W., Li X. *et al.* (2018) Frontiers in Physiology 9(446). https://doi.org/10.3389/fphys.2018.00446

Porter V., Alderson L., Hall S.J. and Sponenberg D.P. (2016) Mason's world encyclopedia of livestock breeds and breeding, 2 Volume Pack. Cabi.

Purcell S., Neale B., Todd-Brown K., Thomas L., Ferreira M.A. *et al.* (2007) The American journal of human genetics 81(3)**:** 559-575.

Sankararaman S., Sridhar S., Kimmel G. and Halperin E. (2008) The American Journal of Human Genetics 82(2)**:** 290-303. https://doi.org/10.1016/j.ajhg.2007.09.022

Tang H., Coram M., Wang P., Zhu X. and Risch N. (2006) The American Journal of Human Genetics 79(1)**:** 1-12. https://doi.org/10.1086/504302

Tortereau F., Moreno C., Tosser-Klopp G., Servin B. and Raoul J. (2017) BMC genetics 18(1)**:** 1-11. https://doi.org/10.1186/s12863-017-0518-2

186. Common heterozygosity-rich regions across the genomes of commercial and local goat breeds

F. Biscarini*, A. Manunza, P. Cozzi and A. Stella

Institute of Agricultural Biology and Biotechnology (IBBA), National Research Council, Via Edoardo Bassini 15, 20133 Milan, Italy; filippo.biscarini@ibba.cnr.it

Abstract

Heterozygosity-rich regions (HRRs) are still under-characterized in animal genomes. In this work, we detected HRRs in goat samples from 6 breeds, both commercially widespread (ALP, BOE, SAA) and local (BRK, CRE, LNR), specifically looking for HRRs common across breeds. In total, 79 common HRRs have been detected; in particular, a region on chromosome 12 (49.9-51.6 Mbps) was shared by over 35% of the goats in 5 breeds. The imputation of missing SNP genotype data did not have a significant impact on results.

Introduction

Contiguous segments ('runs') of diploid genomes may be characterised by higher than average values of homozygosity or heterozygosity. Runs of homozygosity (ROH) have been extensively studied in animal populations, e.g. for the quantification of inbreeding or the detection of selection signatures (Peripolli *et al.* 2017). Runs of heterozygosity, better known as heterozygosity-rich regions (HRRs), are a more recent concept (Williams *et al.* 2016), and are still largely uncharacterized in animals. A handful of studies described HRRs in cattle (Ferenčaković *et al.* 2016, Biscarini *et al.* 2020), sheep (Tsartsianidou *et al*., 2021), turkeys (Marras *et al.* 2018) and horses (Bizarria dos Santos *et al.* 2021). In this study we looked for HRRs in the goat genome, specifically looking for HRRs common across commercial and local breeds. The impact of imputing missing SNP genotypes was also assessed.

Materials & methods

SNP genotype data. For this study, we used SNP genotype data provided by the AdaptMap (Stella *et al.*, 2018) and SMARTER (https://www.inrae-transfert.fr/en/showroom/439-smarter) projects. With the aim of including both widely distributed commercial breeds ('commercial') and locally adapted breeds ('local'), samples from the Barki (BRK), Creole (CRE) and Landrace (LNR) (local), and from the Alpine (ALP), Boer (BOE) and Saanen (SAA) (commercial) breeds were selected, for a total of 1,072 goats. SNPs were obtained from the GoatSNP50 BeadChip (Illumina Inc., San Diego, CA) which includes 53,347 SNPs (mapped on ARS1.2). Only SNP loci on the goat autosomes (chromosomes 1-29) were used, and they were filtered for MAF>0.05 and call-rate (>95% locus-wise, >80% sample-wise). After filtering, 1,051 goats and 47,689 SNPs were left for the analysis (per-breed data breakdown is reported in Table 1). Plink v.1.9 was used for data management and filtering (Chang *et al.* 2015). To evaluate the impact of SNP genotype imputation on HRR detection, missing data were imputed in a separate run using the LHCI (localized haplotype clustering imputation) method implemented in Beagle v.5.1 (Browning and Browning 2009).

Detection of heterozygosity-rich regions (HRR). HRR were detected in the goat genome by scanning the ordered sequence of SNP loci with the window-free method described by Marras *et al.* (2015). The following parameters, derived from Biscarini *et al.* (2020), were used for the detection of HRR: minimum 15 SNPs and minimum length of 250 kbps to define a HRR; maximum gap 1e+03 kbps between adjacent SNPs in a HRR; maximum 3 homozygous SNPs and 2 missing SNPs in a HRR. The R package detectRUNS was used for the detection of HRR (https://cran.r-project.org/web/packages/detectRUNS/). HRR islands have been defined as HRR found (identical) in at least 20% of the samples (within breed).

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_186, © F. Biscarini *et al.* 2022

Table 1. Descriptive statistics of filtered genotype data per breed.

	Breed	n_ind	n_snp	missing_rate	MAF
Commercial	ALP	211	47,689	0.003	0.372
	BOE	251	47,689	0.002	0.368
	SAA	145	47,689	0.005	0.374
Local	BRK	146	47,689	0.019	0.363
	CRE	179	47,689	0.012	0.353
	LNR	120	47,689	0.001	0.356

[1] n_ind: number of goats per breed; n_snp: number of SNPs; MAF: minor allele frequency.

Results

MAF was slightly higher in commercial goat breeds compared to local breeds (0.371 vs 0.357); conversely, missing rate was higher in local breeds (e.g. 1.9% in BRK, 1.2% in CRE; see Table 1). Both findings likely relate to the ascertainment bias of the GoatSNP50 array, developed based mainly on commercial breeds. From non-imputed data, a total of 73,173 HRRs were detected in the 1,051 goats remaining after filtering (69.6 HRRs per goat, on average). More HRRs have been detected in commercial breeds than in local breeds, both overall (50,477 vs 22,696) and per goat (83.3 vs 51). The average length of detected HRRs was 805.7 kbps in commercial breeds and 794.5 kbps in local breeds. With imputation of missing SNP genotypes, similar results in terms of HRRs have been obtained: overall, slightly fewer and shorter HRRs have been detected in all breeds. Descriptive statistics of HRR per breed are reported in Table 2.

Looking at HRRs shared by individual goats, 79 HRR islands (HRRs in 20% of the samples) have been identified. Figure 1 shows HRR islands in chromosomes harbouring more than one island. Common HRR islands are found on chromosomes 1, 10 and 14 (across 3 breeds), chromosomes 3, 13 and 18 (across 4 breeds), and on chromosome 12 (across 5 breeds). The HRR islands most shared across breeds is the one found on chromosome 12 at 49.884-51.611 Mbps. Figure 2 shows a visualization of this common HRR island across samples (B, right), and the corresponding peak of SNP loci found to be lying inside it (A, left).

Discussion

This is the first report on HRR in the goat genome. Like ROH, HRRs can provide insights into the demographic history of populations (e.g. admixture or introgression events) or help detect genomic regions under balancing selection. In particular, HRR common across breeds may indicate regions which

Table 2. Summary of detected HRR in the base scenario (left: selected parameters, no imputation), and with imputed data (same parameters, right).

Group	Breed	Non-imputed[1]			Imputed[1]		
		n_hrr	avg_n	Length	n_hrr	avg_n	Length
Commercial	ALP	15,439	73.17	807.20	15,419	73.08	803.78
	BOE	24,611	98.05	802.85	24,384	97.15	802.13
	SAA	10,427	72.41	807.97	10,323	71.69	806.38
Local	BRK	6,481	44.39	789.39	6,019	41.23	787.95
	CRE	9,725	54.32	795.17	9,339	52.17	792.50
	LNR	6,490	54.08	798.17	6,403	53.36	797.51

[1] n_hrr: total number of of HRRs per breed; avg_n: average number of HRR per goat; length: average length of detected HRRs (in kbps).

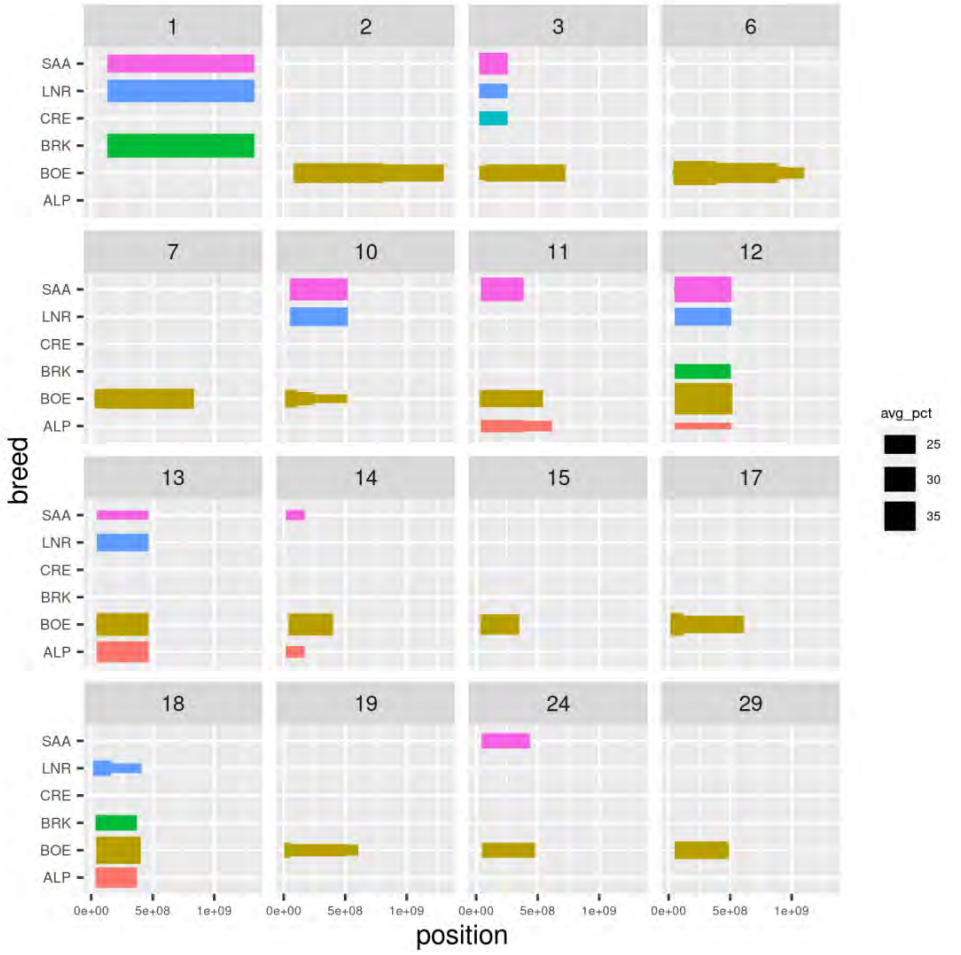


Figure 1. Visualization of HRR islands detected in the genome of commercial (ALP, BOE, SAA) and local (BRK, CRE, LNR) goat breeds. Chromosomes with only one HRR have been excluded. The thickness of the HRR is proportional to the percentage of samples in which it has been detected (avg_pct). The length of the HRRs has been artificially extended (multiplied by 10) for the sake of visualization.

are conserved for biological or evolutionary reasons. This may be the case for the 49.9-51.6 Mbps region on chromosome 12, which harbours about 20 annotated genes, among them IL17D known to be associated with the immune response in humans (Huang *et al.* 2021). Overall, the imputation of residual missing SNP genotypes does not appear to influence results, though this may depend on the initial missing rate and the accuracy of imputation. A systematic analysis of the parameters used to detect HRRs and of the threshold to identify HRR islands still needs to be carried out.

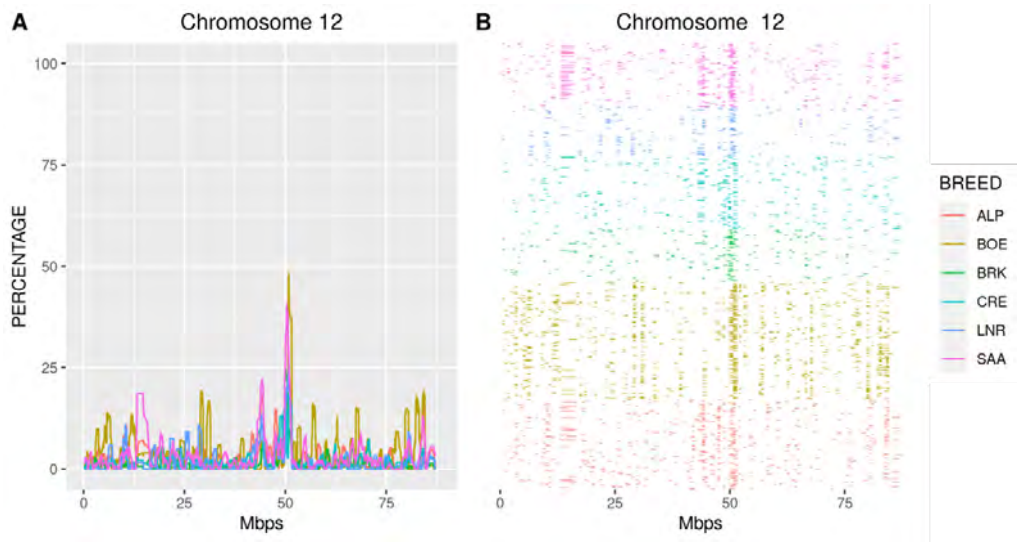


Figure 2. HRR island on goat chromosome 12. (A) proportion of times SNP loci fall within HRRs in the analysed goat samples (across breeds); (B) distribution of HRRs on chromosome 12 in goat samples (a clear signature of heterozygosity is visible at 49.9-51.6 Mbps).

Acknowledgements

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 772787 (SMARTER).

References

Biscarini F *et al.* (2020) Animals 10(12):2285. https://doi.org/10.3390/ani10122285
Bizarria dos Santos W *et al.* (2021) J. Anim. Breed. Genet. 138(2):161-173. https://doi.org/10.1111/jbg.12508
Browning BL *et al.* (2009) Am. J. Hum. Genet. 84(2):210-223. https://doi.org/10.1016/j.ajhg.2009.01.005
Chang CC *et al.* (2015) Gigascience 4(1):s13742-015. https://doi.org/10.1186/s13742-015-0047-8
Ferenčaković M *et al.* (2016) Acta Agric. Slov. 5:S41-4.
Huang J *et al.* (2021) Immunity 54(4):673-686. https://doi.org/10.1016/j.immuni.2021.03.018
Marras G *et al.* (2015) Anim. Genet. 46(2):110-121. https://doi.org/10.1111/age.12259
Marras G *et al.* (2018) Proceed. of the 11th WCGALP (pp. 11-16).
Peripolli E *et al.* (2017) Anim. Genet. 48(3):255-271. https://doi.org/10.1111/age.12526
Stella A *et al.* (2018) Genet. Sel. Evol. 50(1):1-7. https://doi.org/10.1186/s12711-018-0427-5
Tsartsianidou V *et al.* (2021) Genet. Sel. Evol. 53(1):1-17. https://doi.org/10.1186/s12711-021-00682-7
Williams JL *et al.* (2016) Anim. Genet. 47(1):19-27. https://doi.org/10.1111/age.12376

187. Markhor-derived introgression of *PAPSS2* confers high-altitude adaptability in Tibetan goats

C. Li[1,2], Y. Wu[3], B. Chen[1], Y. Cai[1], J. Guo[4], P. Kalds[1], Y. Chen[1], H. Pausch[2], J. Han[5,6], Y. Jiang[1] and X. Wang[1*]

[1]Key Laboratory of Animal Genetics, Northwest A&F University, China; [2]Animal Genomics, ETH Zürich, Switzerland; [3]Institute of Animal Sciences, Tibet Academy of Agricultural and Animal Husbandry Sciences, China; [4]College of Animal Science and Technology, Sichuan Agricultural University, China; [5]CAAS-ILRI Joint Laboratory on Livestock and Forage Genetic Resources, Chinese Academy of Agricultural Sciences (CAAS), China; [6]International Livestock Research Institute (ILRI), Nairobi, Kenya; xiaolongwang@nwafu.edu.cn

Abstract

Understanding the genetic mechanism of how animals adapt to extreme environments is fundamental to unriddle the relationship between molecular evolution and continuous climate change. After domestication, goats have developed strong adaptation to various environments, including harsh conditions at high-altitudes with low temperature and reduced oxygen concentrations. Here, we utilized 331 genomes of goats living at varying altitudes (high >3,000 m above sea level and low <1,200 m) as well as wild caprid species to precisely dissect genetic determinants underlying the adaptation to high altitudes on the Qinghai-Tibetan Plateau (QTP). Population genomic analyses combined with genome-wide association testing revealed that the *PAPSS2* locus confers adaptability to high altitudes in Tibetan goats. Additionally, this gene harbours significant signatures of interspecies introgression from a wild caprid species, markhor (*Capra falconeri*). We further demonstrated a functional role for *PAPSS2* in response to hypoxia using *PAPSS2*-deficient cells. In conclusion, our results suggest a hitherto unknown contribution of *PAPSS2* to high-altitude adaptation and showed that interspecific introgression contributed to the adaptation of goats to the harsh conditions at the Tibetan Plateau.

Introduction

Goats (*Capra hircus*) are believed to be one of the first domesticated livestock species, dating back to ~10,000 years ago (Daly *et al.*, 2021; Zeder, 2000). Goats have developed rapid and strong physiological and genetic adaptation to various environments along with human civilization and migration (Daly *et al.*, 2018). This includes adaptation to some challenging agro-ecological conditions, such as high-altitude environments. Adaptive introgressions had occurred in humans (Huerta-Sánchez *et al.*, 2014) and other mammals native to the QTP, such as Tibetan Mastiff (Miao *et al.*, 2017), cattle (Chen *et al.*, 2018) and sheep (Hu *et al.*, 2019). These studies focus on the genetic basis of adaptive traits. However, how Tibetan goats acquired this adaptability is still unknown. In this study, large-scale genomic and transcriptomic data were systematically compared between goats from the Tibetan Plateau and low-altitude regions to provide new insights into the evolutionary origins of high-altitude adaptability in mammals.

Methods

Blood samples from 91 goats (83 from Tibet, 4 from low altitude, and 4 Siberian ibex) were collected for short-read genome sequencing.

The genomic short-read sequencing data were mapped to the goat reference genome using BWA-MEM (Heng L. 2013), and SNPs were discovered and genotyped using GATK (McKenna *et al*., 2010). The F_{ST}, *XP-EHH*, and GWAS statistics were calculated between 76 high- and 80 low-altitude animals. The *D*-statistics (Martin *et al.* 2015) and Q95 (Racimo *et al.* 2017) analyses were used to conduct a genome-wide screening of putative introgression events from the markhor.

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_187, © C. Li *et al.* 2022

We used CRISPR/Cas9 to knock out the murine *papss2* gene in NIH/3T3 cells. The *papss2* mRNA expression in the NIH/3T3 cells was determined by quantitative PCR (qPCR). The *papss2* deficient cells were incubated in a humidified atmosphere under normal (21% O_2) and hypoxic (3% O_2) conditions at 37 °C (Thermo Scientific HERAcell 150i), respectively. Then the expression of *papss2* was measured alongside changes in cell phenotype using quantitative PCR (qPCR). Differentially expressed genes (DEGs) by RNA-sequencing reads were calculated using cufflinks (Trapnell *et al.* 2012).

Results

We used three complementary approaches (GWAS, F_{ST}, *XP-EHH*) to identify genes potentially involved in high-altitude adaptation (Figure 1A-B). Positively selected haplotypes encompassing *PTEN*, *ATAD1*, *PAPSS2,* and *DSG3* were significantly associated with high-altitude adaptation (Figure 1C). *PTEN* contributes to regulating hematopoiesis and leukemogenesis (Wu *et al.*, 2020). A previous study suggested *PTEN* to likely be under selection in the Tibetan human (Simonson *et al.*, 2010). *ATAD1* plays a critical role in regulating synaptic plasticity, learning, and memory (Ahrens-Nicklas *et al.*, 2017). *PAPSS2* was also previously identified as the major selective gene for high-altitude and cold environment adaptation in horses (Librado *et al.*, 2015; Shi *et al.*, 2019). *DSG3* was previously highlighted as a candidate gene for hypoxia adaptation in Tibetan goats (Kumar *et al.*, 2018). *PAPSS2* has not been implicated in high-altitude adaptability so far.

The *D*-statistic was used to detect the ancestral origin of the selected region and to screen the Tibetan goat genomes for signatures of markhor introgression. The analysis revealed a strong signature of introgression at Chr26:42,070,001-42,120,000 (f_d=0.7634; Figure 2A) which overlapped with *PAPSS2* locus and which was also supported by the Q95 (Figure 2A) and a phylogenetic analysis. The genomic region encompassing

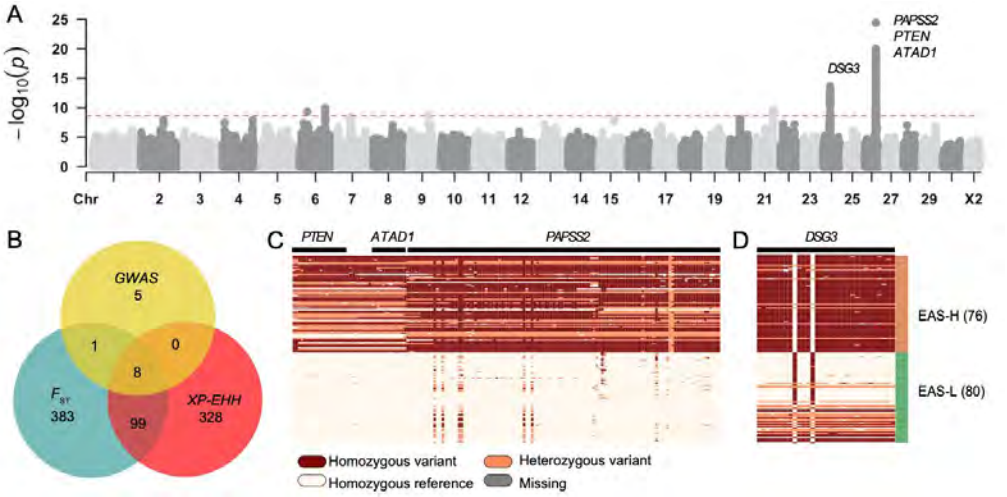


Figure 1. Gene regions responsible for adaptation to high-altitude in goats. (*A*) Manhattan plot of genome-wide association results for high-altitude adaptation in goats (n=156). The dotted horizontal line indicates the threshold (-Log10(*P*)=8.65). (B) The strategy used to identify a list of genes related to high-altitude adaptation. It relies on three sets of genes, derived from the F_{ST} (blue), *XP-EHH* (red), and GWAS (yellow) analyses. F_{ST} and *XP-EHH* selection candidate sets include genes in the top 0.5% of the empirical distributions of F_{ST} and *XP-EHH* results, respectively. (C-D) Comparison of genomic regions identified in selection scans between high- and low-altitude goats. EAS-H, East Asian high-altitude goats, and EAS-L, East Asian low-altitude goats.

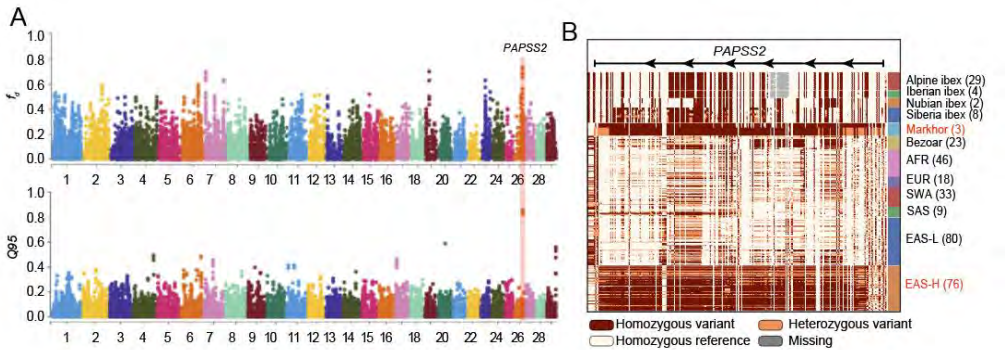


Figure 2. A genomic region encompassing *PAPSS2* was introgressed into Tibetan goats from markhor. (A) Genome-wide distribution of *D*-statistics (top) and Q95 (bottom) values calculated for 50-kb windows with a 10-kb step across the genome. Each dot represents a 50-kb window. (B) Haplotype structure for *PAPSS2*. Rows represent individuals and columns represent SNPs positions in the genome. AFR, African goats; EUR, European goats; SWA, Southwest Asian goats; SAS, South Asian goats.

PAPSS2 had a high similarity between Tibetan goats and markhor but was substantially different in other domestic goat populations and wild caprid species (Figure 2B). The functional role of *PAPSS2* in the adaptation to hypoxia was further confirmed with the *papss2*-deficient cells.

Discussion

In this study, we leveraged a large caprine whole-genome and transcriptome data set encompassing 331 goats and 13 tissues to localize signatures of high-altitude adaptation in Tibetan goats. Collectively, both positively selected haplotypes and expression patterns suggest genomic regions encompassing *PAPSS2* and *DSG3* to be significantly associated with high-altitude adaptation in Tibetan goats.

Although harsh environments may promote the accumulation of adaptive mutations in animals, it is unlikely that many de novo mutations can emerge at the same site during a short period of time (Hu *et al.*, 2019; Jeong *et al.*, 2016; Zhao *et al.*, 2017). Adaptive introgressions from other geographically close species may contribute to the process of adaptation that occurred in humans and many animals native to the QTP. In this study, we revealed that Tibetan goats carry specific adaptive introgressions derived from markhor (*C. falconeri*), a wild goat species native to high Himalayan mountains, including a haplotype encompassing *PAPSS2*. This region showed the strongest signature of both adaptive introgression and selection. A relatively long haplotype (chr26:42,077,488-42,179,191) also indicated that the area was introgressed rather recently.

Several studies have shown that the kidneys and the spleen may play a critical role in human oxygen-sensing (Schodel and Ratcliffe, 2019) and mammalian physical activity (Storz and Cheviron, 2021). Through multi-tissue transcriptome data comparison, we found that the *PAPSS2* gene was differentially expressed in the kidney and spleen between Tibetan and other domestic goats. The functional role of *PAPSS2* in the adaptation to hypoxia was further confirmed by our functional validation assay. Thus, it is plausible that the unique but recent introgression from the markhor may have conferred the adaptability of Tibetan goats to survive over the harsh environment on the QTP.

Our data provide new insights into the role of genetic introgression from wild relatives on environmental adaptation in livestock. We showed that the establishment of the genetic mechanism of plateau adaptation in Tibetan goats had occurred by the combined effects of gene introgression and natural selection. This

enabled Tibetan goats to adapt to the hypoxic environment in a relatively short period of time and continues to strengthen this ability until now.

References

Ahrens-Nicklas R.C., *et al.* (2017) Neurol Genet 3(1), e130. https://doi.org/10.1212/NXG.0000000000000130.

Chen N., *et al.* (2018) Nat. Commun 9(1):1-13. https://doi.org/10.1038/s41467-018-04737-0.

Daly K.G., *et al.* (2018) Science 361(6397):85-88. https://doi.org/10.1126/science.aas9411.

Daly K.G., *et al.* (2021) Proc. Natl. Acad. Sci 118(25). https://doi.org/10.1073/pnas.2100901118.

Heng L. (2013) arXiv Prepr. 1303.3997. arXiv:1303.3997.

Hu X.J., *et al.* (2019) Mol. Biol. Evol 36(2):283-303. https://doi.org/10.1093/molbev/msy208.

Huerta-Sánchez E., *et al.* (2014) Nature 512(7513):194–197. https://doi.org/10.1038/nature13408.

Jeong C., *et al.* (2016) Proc. Natl. Acad. Sci 113(27):7485-7490. https://doi.org/10.1073/pnas.1520844113.

Kumar C., *et al.* (2018) Front. Genet 553. https://doi.org/10.3389/fgene.2018.00553.

Librado P., *et al.* (2015) Proc. Natl. Acad. Sci 112(50):E6889-6897. https://doi.org/10.1073/pnas.1513696112.

McKenna A., *et al.* (2010) Genome Res 20:1297–1303. https://doi.org/10.1101/gr.107524.110.

Miao B., *et al.* (2017) Mol. Biol. Evol 34(3):734-743. https://doi.org/10.1093/molbev/msw274

Schodel J., and Ratcliffe P.J. (2019) Nat Rev Nephrol 15(10):641-659. https://doi.org/10.1038/s41581-019-0182-z.

Shi Y., *et al.* (2019) Nat. Commun 10(1):4892. https://doi.org/10.1038/s41467-019-12801-6.

Simonson T.S., *et al.* (2010) Science 329(5987):72-75. https://doi.org/10.1126/science.1189406

Storz J.F., and Cheviron Z.A. (2021) Annu Rev Anim Biosci 9:149-171. https://doi.org/10.1146/annurev-animal-072820-102736.

Trapnell C., *et al.* (2012) Nat. Protoc 7:562–578. https://doi.org/10.1038/nprot.2012.016.

Wu Y., *et al.* (2020) Cold Spring Harb. Perspect. Med 10(10). https://doi.org/10.1101/cshperspect.a036244.

Zeder M.A., and Hesse B. (2000) Science 287(5461):2254-2257. https://doi.org/10.1126/science.287.5461.2254.

Zhao Y.X., *et al.* (2017) Mol. Biol. Evol 34(9):2380-2395. https://doi.org/10.1093/molbev/msx181.

188. Patterns of inbreeding and selection using runs of homozygosity in North American dairy cattle

E.A. Lozada-Soto[1*], F. Tiezzi[2], J.B. Cole[3], P.M. VanRaden[4] and C. Maltecca[1]

[1]Department of Animal Science, North Carolina State University, Raleigh, NC 27607, USA; [2]Department of Agriculture, Food, Environment and Forestry (DAGRI), University of Florence, 50144 Florence, Italy; [3]URUS Group LP, Madison, WI 53718, USA; [4]Animal Genomics and Improvement Laboratory, Henry A. Wallace Beltsville Agricultural Research Service, USDA, Beltsville, MD 20705, USA; ealozada@ncsu.edu

Abstract

The main objective of this study was to leverage genomic information to ascertain patterns of inbreeding and selection in five North American dairy cattle populations. We obtained genotypes for over 4 million individuals of the Ayrshire, Brown Swiss, Guernsey, Holstein, and Jersey breeds. Inbreeding based on runs of homozygosity was calculated in each population. The average inbreeding ranged from 0.11 for Ayrshire to 0.17 for Jersey. We calculated a coefficient of homozygosity for each marker. Highly homozygous markers were joined into larger genomic segments of interest that ranged from 0.08 to 7.83 Mb in length and spanned 14 chromosomes across breeds. Annotation of genes and QTLs in the highly homozygous regions revealed selection for economically important traits, notably for udder and cow health, productive life, and reproductive traits. We found differences across breeds on inbreeding load, genomic regions of high inbreeding, and selection signatures.

Introduction

Runs of homozygosity (ROH) are continuous stretches of markers in a homozygous state in the genome of an individual. The characterization of the length, number, and genomic distribution of ROH allow us to ascertain the level of inbreeding in individuals and decipher the inbreeding history of populations (Gibson *et al.*, 2006). The proportion of the autosomal genome covered by ROH segments has been routinely used as a measure of inbreeding in cattle populations (Howard *et al.*, 2015; Makanjuola *et al.*, 2020) due to its high correlation with other inbreeding coefficients and autozygosity (Peripolli *et al.*, 2017). We can also use the level of homozygosity at the marker level to identify regions of selection within a population and uncover differences in the selection history of multiple populations. Therefore, our aim was to characterize ROH segments, calculate the level of inbreeding, and identify regions of high homozygosity as selection signatures in the five major breeds of North American dairy cattle.

Materials & methods

Animals and genotypes. Imputed genotypes were obtained for 76,389 autosomal markers of 4,173,679 animals of the Ayrshire (AY), Brown Swiss (BS), Guernsey (GU), Holstein (HO), and Jersey (JE) breeds. Quality control of the genomic data was first performed within each breed, including the removal of animals with a marker call rate lower than 95%, markers with a minor allele frequency lower than 0.01%, and markers with Mendelian error rates larger than 10%. Subsequently, we removed markers with an across-breed call rate lower than 95%. After genomic quality control, a total of 4,010,718 animals (8,413 AY, 39,040 BS, 3,834 GU, 3,525,992 HO, and 433,439 JE) and 62,546 markers remained for subsequent analyses.

Detection of ROH and calculation of segment-based inbreeding. Runs of homozygosity (ROH) were detected using a sliding window approach with the PLINK v.1.9 software (Chang *et al.*, 2015). Parameters to define windows and detect ROH were the following: a minimum window size of 20 markers, at most 1 heterozygote call in a window, at most 2 missing calls in a window, a minimum physical length of 1 Mb for ROH, a maximum gap of 0.5 Mb between consecutive markers, a minimum marker density of 1 marker

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_188, © E.A. Lozada-Soto *et al.* 2022

per 0.1 Mb, a minimum number of 60 markers to declare a ROH. Parameters were chosen according to previous results (Purfield *et al.*, 2012; Forutan *et al.*, 2018; Meyermans *et al.*, 2020). Descriptive statistics of the detected ROH were calculated for each breed. A segment-based inbreeding coefficient (F_{ROH}) was calculated for each animal as the proportion of the autosomal genome covered by ROH.

Region-specific homozygosity and ROH islands. A coefficient of homozygosity ($F_{ROH.i}$) was calculated for each i^{th} marker as the proportion of animals (within breed) for which the marker is within a ROH segment. Afterwards, we identified markers that fell at or above the 99.5th percentile for $F_{ROH.i}$ and joined significant adjacent markers into larger segments, henceforth identified as ROH islands.

Annotation of genes and QTLs located within ROH islands. Annotation of genes located within the identified ROH islands was done using the 'btaurus_gene_ensembl' (ARS-UCD1.2 genome assembly) database with the 'biomaRt' R package (v.2.48.3) (Durinck *et al.*, 2009). We utilized the GALLO R package (Fonseca *et al.*, 2020) to query the Animal QTLdb (https://www.animalgenome.org/cgi-bin/QTLdb/index) for quantitative trait loci (QTL) that have been previously identified in the regions of interest. Furthermore, a trait enrichment analysis was performed on the annotated QTLs to identify traits whose presence is not due to chance (FDR adjusted *P*-value<0.05).

Results & discussion

Descriptive statistics of detected ROH per individual are presented in Table 1. The average number of ROH ranged from 39.20 for AY to 65.87 for JE, the average length ranged from 6.48 Mb in GU to 7.31 Mb in BS, and the average combined length ranged from 268.39 Mb in AY to 431.77 Mb in JE. The inbreeding followed a similar trend with AY having the lowest average inbreeding (0.11) and JE having highest average inbreeding (0.17).

Region-specific homozygosity for every marker across the genome for the five breeds is presented in Figure 1. We found significant markers in chromosomes 4, 5,6, and 8 for AY; 5, 6, and 16 for BS; 6, 11, 13, 15, and 19 for GU; 10 and 20 for HO; and 2, 3, 7, and 20 for JE. ROH islands ranged from 0.08 to 7.83 Mb in length. We found two overlaps between ROH islands in different breeds, a 3.44 Mb overlap in chromosome 6 (BTA6: 86.76-90.1) between AY and BS, and a 2.57 Mb overlap in chromosome 20 (BTA20: 25.61-28.17) between HO and JE. We found a total of 26 genes shared between AY and BS, most notably *ADAM metallopeptidase with thrombospondin type 1 motif 3* (ADAMTS3). The ADAMTS3 gene has been linked to longevity (Mészáros *et al.*, 2014), embryonic development (Janssen *et al.*, 2016), and milk and milk component yield (Costa *et al.*, 2019) in cattle. The region that overlapped between HO and JE contained 5 genes, most notably *ISL LIM homeobox 1* (ISL1) and *pelota mRNA surveillance and ribosome rescue factor* (PELO). Both of these genes have been associated with spermatogenesis (Fernández *et al.*, 2014). The most

Table 1. Descriptive statistics of detected ROH per individual and average F_{ROH} in 5 breeds.[1]

Breed[2]	Avg. n.	Avg. length (Mb)	Avg. combined length (Mb)[3]	F_{ROH}
AY	39.20 (8.47)	6.78 (1.36)	268.39 (85.11)	0.11 (0.03)
BS	49.58 (8.72)	7.31 (1.08)	363.55 (87.12)	0.15 (0.04)
GU	57.39 (7.71)	6.48 (1.03)	371.99 (79.22)	0.15 (0.03)
HO	42.03 (8.66)	7.24 (1.17)	305.88 (87.68)	0.12 (0.04)
JE	65.87 (8.52)	6.55 (0.90)	431.77 (83.49)	0.17 (0.03)

[1] Standard deviation shown within parenthesis.

[2] AY= Ayrshire; BS= Brown Swiss; GU=Guernsey; HO= Holstein; JE= Jersey.

[3] Average combined length of genome in ROH.

gene-dense island was found for Jersey cattle in chromosome 7 (BTA7: 39.76-45.56), where we identified a total of 167 genes. A notable gene found in this region was *NLR family pyrin domain containing 3* (NLRP3), this gene has been previously associated with resistance to *Mycobacterium avium* ssp. paratuberculosis (MAP) infection (Mallikarjunappa *et al.*, 2018).

The top 10 (based on *P*-value) enriched traits from the annotated QTLs are presented for each breed in Figure 2. Milk production traits were found to be enriched in all five breeds, including milk or milk protein yield in AY, BS, HO, and JE. Other traits that were found enriched in two or more breeds included productive life, immunoglobin G level, clinical mastitis, and ketosis. This highlights the high selective pressure that has been put on traits that increase the health, reproductive ability, and length of productive life of U.S. dairy cows. Interestingly, 7 out of the top 10 enriched traits in the Guernsey breed related to meat and carcass traits, mainly content of saturated and monounsaturated fatty acids.

In summary, the differences observed in the genomic patterns of homozygosity suggest substantial heterogeneity in the management and selection strategies implemented in the different populations of U.S. dairy cattle.

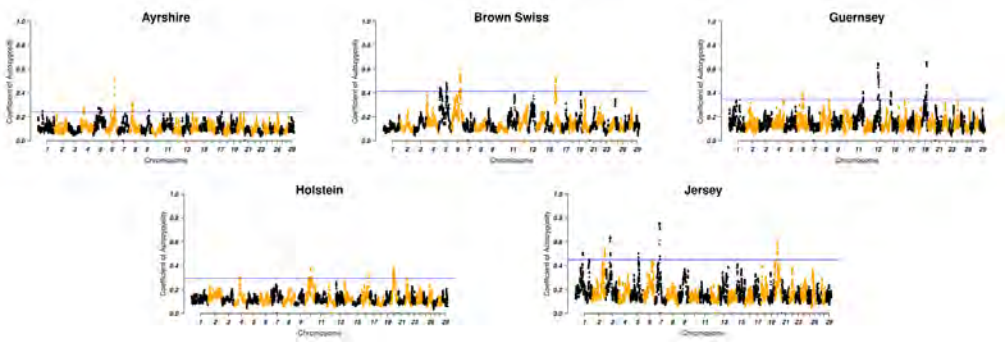


Figure 1. Manhattan plot of region-specific homozygosity ($F_{ROH,i}$) for Ayrshire (A), Brown Swiss (B), Guernsey (C), Holstein (D), and Jersey (E) cattle.

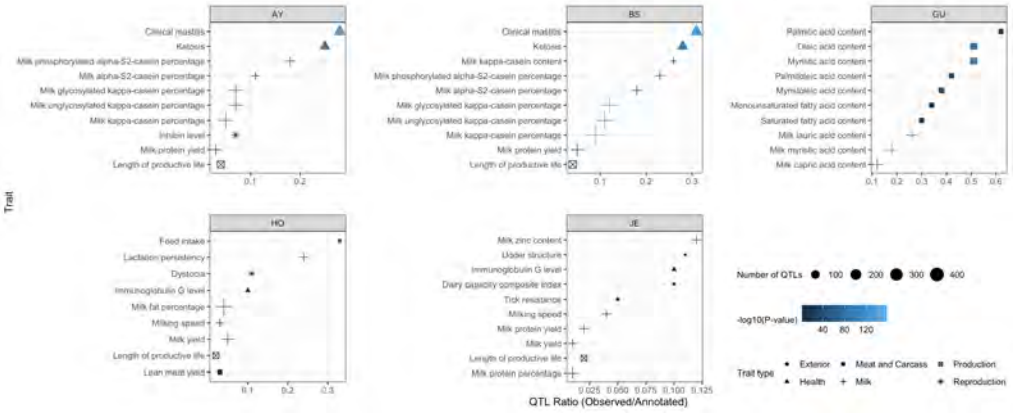


Figure 2. Top 10 enriched traits for annotated QTL's mapped to ROH islands in 5 breeds.

References

Chang C.C., Chow C.C., Tellier L.C., Vattikuti S., Purcell S.M., *et al.* (2015) Gigascience 4:7. https://doi.org/10.1186/S13742-015-0047-8.

Costa A., Schwarzenbacher H., Mészáros G., Fuerst-Waltl B., Fuerst C., *et al.* (2019) J. Dairy Sci. 102:10088-10099. https://doi.org/10.3168/JDS.2019-16663.

Durick S., Spellman P.T., Birney E., and Huber W. (2009) Nat. Protoc. 4:1184-1191. https://doi.org/10.1038/nprot.2009.97.

Fernández M.E., Lirón J.P., Prando A., Rogberg-Muñoz A., Peral-García P., *et al.* (2014) Livest. Sci. 167:9–18. https://doi.org/10.1016/J.LIVSCI.2014.05.009.

Fonseca P.A.S., Suárez-Vega A., Marras G., and Cánovas Á. (2020) Gigascience 9. https://doi.org/10.1093/GIGASCIENCE/GIAA149.

Forutan M., Ansari Mahyari S., Baes C., Melzer N., Schenkel F.S., *et al.* (2018) BMC Genomics 19:98. https://doi.org/10.1186/s12864-018-4453-z.

Gibson J., Morton N.E., and Collins A. (2006) Hum. Mol. Genet. 15:789–795. https://doi.org/10.1093/HMG/DDI493.

Howard J.T., Maltecca C., Haile-Mariam M., Hayes B.J, and Pryce J.E. (2015) BMC Genomics 16:1–13. https://doi.org/10.1186/s12864-015-1352-4.

Janssen L., Dupont L., Bekhouche M., Noel A., Leduc C., *et al.* (2016) Angiogenesis 19:53-65. https://doi.org/10.1007/S10456-015-9488-Z.

Makanjuola B.O., Miglior F., Abdalla E.A., Maltecca C., Schenkel F.S., *et al.* (2020) J. Dairy Sci. 103:5183–5199. https://doi.org/10.3168/jds.2019-18013.

Mallikarjunappa S., Sargolzaei M., Brito L.F., Meade K.G., Karrow N.A., *et al.* (2018) J. Dairy Sci. 101:7280–7286. https://doi.org/10.3168/JDS.2018-14388.

Mészáros G., Eaglen S., Waldmann P., and Sölkner J. (2014) Open J. Genet. 2014:46–55. https://doi.org/10.4236/OJGEN.2014.41007.

Meyermans R., Gorssen W., Buys N., and Janssens S. (2020) BMC Genomics 21:1-14. https://doi.org/10.1186/S12864-020-6463-X.

Peripolli E., Munari D.P., Silva M.V.G.B, Lima A.L.F., Irgang R., *et al.* (2017) Anim. Genet. 48:255–271. https://doi.org/10.1111/AGE.12526.

Purfield D.C., Berry D.P., McParlan S., and Bradley D.G. (2012) BMC Genet. 13:1-11 https://doi.org/10.1186/1471-2156-13-70.

189. Exploring genetic diversity and selection signatures in indigenous African cattle breeds

J. Friedrich[1*], R. Bailey[2], A. Talenti[1], K. Marshall[3], J. Prendergast[1] and P. Wiener[1]

[1]Division of Genetics and Genomics, The Roslin Institute and Royal (Dick) School of Veterinary Studies, University of Edinburgh, Midlothian, EH25 9RG, United Kingdom; [2]Department of Ecology and Vertebrate Zoology, University of Łódź, 90-237 Łódź, Poland; [3]Centre for Tropical Livestock Genetics and Health (CTLGH), ILRI Kenya, P.O. Box 30709, Nairobi, Kenya; juliane.friedrich@roslin.ed.ac.uk

Abstract

African cattle are a fascinating resource for exploration of genetic diversity due to their complex history and their adaptation to diverse environments. As a consequence of taurine and indicine ancestries and selection pressure from various environments as well as from humans, over 150 African cattle breeds exist today with remarkable genotypic and phenotypic diversity. We analysed whole-genome sequence data of over 100 African cattle comprising 9 breeds and calculated within- and between-breed population statistics to characterise genetic diversity and identify selection signatures. We found overlapping selection signatures between indigenous African breeds, with the longest selection signature on BTA10. The greatest genetic differentiation from other breeds was observed for the N'Dama and Boran breeds, potentially mirroring their high proportion of taurine and indicine ancestry, respectively. Genomic regions that are highly differentiated between breeds suggest candidate regions for associations with important adaptation and production traits.

Introduction

Indigenous African cattle with their complex evolution, migration and admixture history are a unique resource to explore genetic diversity and identify selection signatures for adaptation to diverse environments. Indigenous African cattle descended from *Bos taurus taurus* and *Bos taurus indicus* subspecies: from their respective domestication centres, taurine cattle migrated to Europe and subsequently first entered the African continent from the North, while indicine cattle migrated later to Africa from the East (Mwai *et al.*, 2015). The ancestors of indigenous cattle were exposed to considerable environmental and disease challenges that led to evolutionary genetic adaptations. Nowadays, it is estimated that around 150 indigenous cattle breeds inhabit the African continent (Gebrehiwot *et al.*, 2020) that can be roughly grouped into breeds with high proportions of indicine background (Zebu), high proportions of taurine background ('African taurine') and cattle with a mixture of taurine and indicine ancestries (Kim *et al.*, 2020). A number of traits have been associated with genetic ancestry, including the hump, large ears and excess skin indicative of indicine ancestry. These characteristics are thought to provide heat tolerance and thus contribute to adaptation to semi-arid and desert climates (Hansen, 2004). Disease resilience is another ancestry-related factor: e.g. African taurine cattle are considered to be more resistant to trypanosomiasis (Berthier *et al.*, 2015) than those with substantial indicine ancestry. Adaptation to these environmental challenges can improve animal performance and fertility and hence, indigenous African cattle are an important socio-economical resource for the African population, particularly for smallholder farmers. The aims of this study are to use genetic diversity patterns of African cattle breeds to improve understanding of the genetic basis of important traits and to contribute to conservation and breeding strategies.

Materials & methods

Whole-genome sequence data of African cattle. Illumina sequencing data for multiple global cattle genomes was prepared as described in Dutta *et al.* (2020). We retained all African cattle samples with a sequencing depth of ≥8. The biallelic variants were filtered for a call rate ≥95% and genotyping quality

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_189, © J. Friedrich *et al.* 2022

(QG) >20. Samples with a call rate <75% or a 1st-degree relationship to another sample were excluded. After removal of breeds with fewer than 7 samples, the following samples remained: 7 Ankole, 7 Baoule, 47 Boran, 7 Djakkore, 9 Kenana, 27 N'Dama, 8 Ogaden and 7 Zebu Gobra.

Determining the genetic structure of African cattle breeds. The genetic structure and admixture of samples was analysed using principal component analysis (PCA) in Plink (Chang *et al.*, 2015) and the ADMIXTURE (Alexander *et al.*, 2009) software. The genotype data was pruned for high linkage disequilibrium in Plink, leaving a pruned dataset of 4,902,838 variants. For the ancestry estimation in ADMIXTURE, the best number of clusters (K) was determined by comparing 5-fold cross-validation (CV) errors for K=2-10.

Comparative analysis of selection signatures. Signatures of selection *within* breeds were identified using the integrated haplotype score (iHS) statistic. Genotypes were phased in Beagle version 4.173 (Browning and Browning, 2007) and marker-wise iHS were calculated according to Voight *et al.* (2006) using the hapbin software (Maclean *et al.*, 2015). Correlations between marker-wise iHS statistics across breeds were calculated and the top 0.01% iHS per breed were selected to identify signatures shared across breeds. To determine signatures of selection *between* breeds, a windows-based fixation index (F_{ST}) was estimated between each breed and all others. The mean F_{ST} for windows of 50 kb with an overlap of 25 kb was determined using vcftools (Danecek *et al.*, 2011). Windows with fewer than 30 variants were excluded.

Results

The underlying genomic population structure for all African cattle breeds was explored using PCA and Admixture analysis. The PCA indicated a separation according to the longitudinal geographical origin of the samples, which also correlates with the occurrence of more taurine (West Africa) or more indicine (East Africa) breeds (Figure 1A). The lowest CV error in the Admixture analysis was calculated for K=3. Breeds with the greatest ancestry proportions for the three clusters were N'Dama and Baoule (red cluster), Kenana (dark blue cluster) and Boran (light blue cluster) (Figure 1B). To detect recent and incomplete selective sweeps within breeds, we calculated the iHS score for each marker and identified the markers belonging to the top 0.01% of the iHS for each breed. The correlation between iHS scores for the African breeds was highest for Kenana and Ogaden (corr=0.27) and Kenana and Djakkore (corr=0.22). The lowest correlation was observed between N'Dama and Boran (corr=0.03). We grouped variants within 50 kb of each other that were detected as within the top 0.01% in at least two breeds into windows. We detected 78 windows including at least two selection candidate variants; these spanned 51.8 kb on average. The longest candidate window was found on BTA10 (233-237 kb; 54 SNPs), with top 0.01% variants for all African breeds except Boran. This region included one gene: the T cell receptor alpha variable 24 (*TRAV24*) gene.

genetic differentiation between African breeds, we calculated window-based F_{ST} between each breed and all others. The genome-wide average F_{ST} was highest for N'Dama (0.10), followed by Boran (0.07), Baoule (0.06), Kenana (0.04), Ogaden (0.03), Ankole (0.02), Zebu Gobra (0.02) and Djakkore (0.01). The highest windows-based F_{ST} differed across breeds and included regions on BTA6, 7, 9, 15 and 20 (Figure 2). The window with the highest F_{ST} across all breeds was identified for N'Dama on BTA7, and included the coding region for one gene, the pleckstrin and Sec7 domain containing 2 (*PSD2*) gene.

We checked for coincident variants between top 0.01% F_{ST} windows and top 0.01% iHS for every breed, but detected no overlaps.

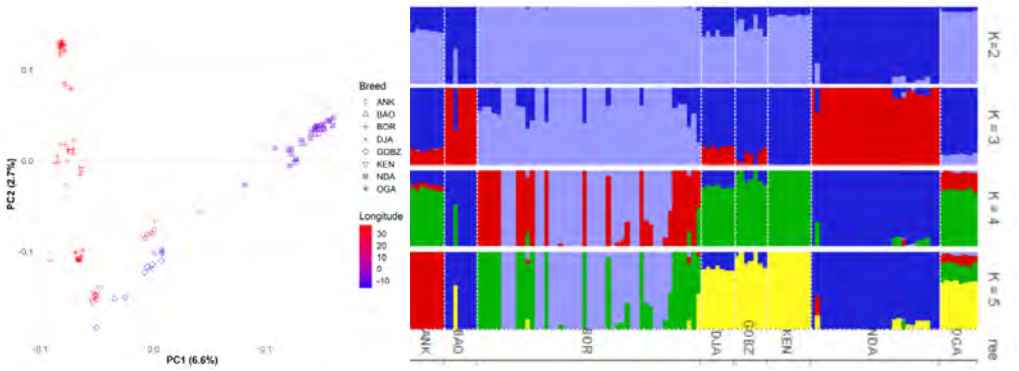


Figure 1. Genomic population structure analysed by PCA (A) and Admixture analysis (B). Samples are coloured according to the longitude of the capital of the country of origin. ANK = Ankole; BAO = Baoule; BOR = Boran; DJA = Djakkore; GOBZ = Zebu Gobra; KEN = Kenana; NDA = N'Dama; OGA = Ogaden.

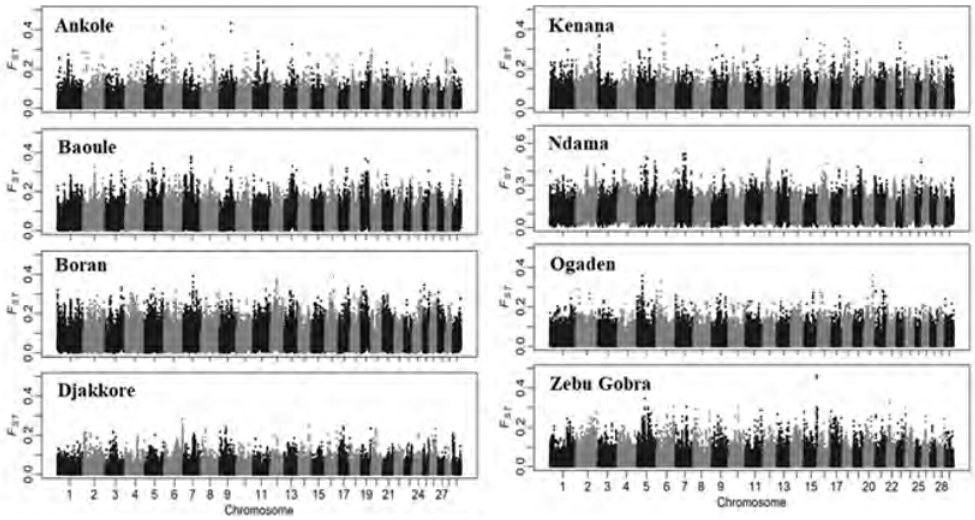


Figure 2. Window-based F_{ST} across the genome for African breeds.

Discussion

African cattle represent a valuable genetic resource to study both artificial and natural selection. Characterising the selection signatures and genetic differences between breeds will provide insights into genomic regions associated with environmental adaptation. PCA and admixture analyses showed a clear separation between cattle with higher proportions of taurine and indicine ancestry, reflecting the domestication and evolutionary history of indigenous African cattle and their geographical origins (Mwai *et al.*, 2015). The highest genetic differentiation from the remaining breeds was observed for N'Dama and Boran, which had the highest proportion of taurine and indicine ancestry, respectively. Furthermore, the selection signature results suggest the potential to identify regions associated with breed-specific adaptation or production traits, e.g. a genomic region on BTA10 harbouring the *TRAV24* gene emerged from the iHS results as a candidate region under selection in all African breeds except Boran. *TRAV24* is involved in adaptive immunity and the wider region on BTA10 overlaps with the multi-gene TRA/TRD

cluster (Connelley *et al.*, 2014). Furthermore, the F_{ST} results highlighted the locus on BTA7 (harbouring the *PSD2* gene) as a region for which N'Dama is highly differentiated from all other breeds. Previously, Kim *et al.* (2020) identified this as a region showing an excess of indicine ancestry in admixed African cattle. *PSD2* has also been suggested as a selection target related to fat deposition in sheep (Mastrangelo *et al.*, 2019). These two examples demonstrate that identification of selection signatures across multiple breeds has the potential to shed further light into gene functions and the genetic basis of phenotypic variability in indigenous African cattle breeds.

References

Alexander D.H., Novembre J., and Lange K. (2009) Genome Res 19: 1655–1664. https://dx.doi.org/10.1101%2Fgr.094052.109

Berthier D., Peylhard M., Dayo G.-K., Flori L., Sylla S. *et al.* (2015) PLoS One 10: e0126498. https://doi.org/10.1371/journal.pone.0126498

Browning S.R., and Browning B.L. (2007) Am J Hum Genet 81: 1084–1097. https://dx.doi.org/10.1086%2F521987

Chang C.C., Chow C.C., Tellier L.C., Vattikuti S., Purcell S.M. *et al.* (2015) GigaScience 4: 7. https://doi.org/10.1186/s13742-015-0047-8

Connelley T.K., Degnan K., Longhi C.W., and Morrison W.I. (2014). BMC Genomics 15: 994. https://doi.org/10.1186/1471-2164-15-994

Danecek P., Auton A., Abecasis G., Albers C.A., Banks E. *et al.* (2011) Bioinformatics 27: 2156–2158. https://dx.doi.org/10.1093%2Fbioinformatics%2Fbtr330

Dutta P., Talenti A., Young R., Jayaraman S., Callaby R. *et al.* (2020) Nat Commun 11: 4739. https://doi.org/10.1038/s41467-020-18550-1

Gebrehiwot N.Z., Strucken E.M., Aliloo H., Marshall K., and Gibson J.P. (2020) BMC Genomics 21: 869. https://doi.org/10.1186/s12864-020-07270-x

Hansen P.J. (2004) Anim Reprod Sci 82–83: 349–360. https://doi.org/10.1016/j.anireprosci.2004.04.011

Kim K., Kwon T., Dessie T., Yoo D., Mwai O.A. *et al.* (2020).Nature Genetics 52: 1099–1110. https://doi.org/10.1038/s41588-020-0694-2

Maclean C.A., Chue Hong N.P., Prendergast J.G.D. (2015) Mol Biol Evol 32: 3027–3029. https://doi.org/10.1093/molbev/msv172

Mastrangelo S., Bahbahani H., Moioli B., Ahbara A., Abri M.A. *et al.* (2019) PLOS ONE 14: e0209632. https://doi.org/10.1371/journal.pone.0209632

Mwai O., Hanotte O., Kwon Y.-J., Cho S. (2015) Asian-Australas J Anim Sci 28: 911–921. https://doi.org/10.5713/ajas.15.0002r

Voight B.F., Kudaravalli S., Wen X., and Pritchard J.K. (2006) PloS Biol 4. https://doi.org/10.1371/journal.pbio.0040072

190. Understanding and utilizing genetic diversity in Dual-Purpose Blue: genome-wide association for type traits

N. Gengler[1*], H. Wilmot[1,2], S. Vanderick[1] and H. Atashi[1,3]

[1]Gembloux Agro-Bio Tech, University of Liège, Passage des Déportés 2, 5030 Gembloux, Belgium; [2]National Fund for Scientific Research (F.R.S.-FNRS), Rue d'Egmont 5, 1000 Brussels, Belgium; [3]Jomhouri Eslami Blvd, Shiraz University, 71441-65186 Shiraz, Iran; nicolas.gengler@uliege.be

Abstract

Dual-Purpose Blue in Belgium and northern France is a local breed that is currently in the focus in order to understand its genetic distinctiveness and its use for local products. Its genomic evaluation systems are evolving, and its specific genetic distinctness was here studied for type traits. Results showed that throughout the 23 traits, there were two distinguishable genomic regions (BTA2 and BTA26) where BTA26 was associated with development traits and BTA2 was associated with muscularity traits, but also with associated traits as overall rump, udder depth, overall udder, and rear udder. Given these results future genomic evaluation systems should consider the segregating *MSTN* (myostatin), which is located on BTA2.

Introduction

Blue cattle in Belgium, but also in neighbouring northern France, originated from the crossbreeding between local dairy cattle and imported Durham (Shorthorn) bulls during the 1850-1880. Even if the presence of blue cattle in the low countries is attested by paintings from the 17th century, this crossbreeding mostly created the blue colour pattern. This pattern is a variant of the Shorthorn pattern where black replaces red leading to three colours: black (and white), white and blue roan. During the early 1900's this breed evolved into a local dual-purpose breed. In Belgium, since the 1950's, most breeders favoured the beefier type leading to the Beef Belgian Blue (BBB) breed. In France political circumstances let few farmers continue keeping blue cattle local known as the dual-purpose 'Bleue du Nord'. For many years, French breeders have been joining forces with fellow dual-purpose breeders in the 'Belgian Blue' Herdbook in Belgium who have recently changed the name of their Belgian Blue strain to 'Dual Purpose Blue' (DPB), a name used hereafter for the whole population. Several projects have led to common genetic evaluations for several traits and a common understanding of the needs in this breed (BlueSter 2021). General efforts will be reported to exploit genomics as a tool and its use in the of context of DPB cattle. The specific objective will be to illustrate this by estimating genomically enhanced estimated breeding values (GEBV) and by performing genome wide association studies (GWAS) for linear type traits.

Materials & methods

DPB populations. The DPB are present in the southern Belgian (Walloon) part (~80%), as mentioned in France (~15%) and the remaining animals are in Flanders, around Brussels. This third population is not providing data to the joint DPB breeding program but selected young bulls are entering it, also based on genomic predictions. Flemish breeders can also access DPB AI sires. In total approximately 3,000 cows are (milk) performance recorded.

Genotyping. Extensive efforts are ongoing to genotype a large part of the currently active DPB population. This study, which used data available in May 2021, had access to 1,772 genotypes. The DPB population is rather unique as it has a tradition of testing for the *MSTN* (myostatin) mutation that is segregating in its population since it was discovered (Grobet *et al.* 1997). Also, given this fact, there is a large acceptance for genomic testing, it is expected that genotyping will cover nearly the whole population in a few years.

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_190, © N. Gengler *et al.* 2022

Performance recording and type data. Milk performance recording is well established, and individual data was available down to the 1970s in Wallonia. This is relevant as the DPB population was larger at that moment and semen of AI sires from this period is still being used. Recently, a transnational type recording program was initiated addressing the complexity of the needs for this dual-purpose breed. Therefore, the 23 traits scheme includes 18 individual traits: five udder traits, three muscularity traits, but also body development, including stature, rump, feet and legs and other traits and five additional overall traits that are scored and not computed from individual traits: udder, development, feet and legs, muscularity, and rump. These traits allowed the inclusion of less recent data from a simplified system previously used in France which handled only stature and these overall traits. Results reported in this study labelled udder, development, feet and legs, muscularity, and rump refer always to those 5 overall traits.

A total of 1,746 French and 10,701 Belgian records from 10,134 cows scored until the beginning of 2021 were included in this study. All traits are scored from 1 to 9 except stature which is measured. Normally, cows needed to be scored at least once in first or second lactation but could then be rescored. For this reason, 50.7 and 37.6% of the records were scored in first or second lactation, respectively. The remaining were animals that were rescored, mostly in the third lactation (10.0%), however late rescoring was also used because DPB breeding strives to obtain a high longevity.

Generating genomically enhanced estimated breeding values (GEBV). A transnational genomic evaluation system is currently under development for type traits. This system is based on two steps. In a first step, as there were missing traits, it was preferable to use a multiple-trait model allowing missing values. Moreover, tests showed that a full multiple-trait model had convergence issues. However, the use of a canonical transformation adopted for multiple diagonalization (Misztal *et al.* 1995) was able to handle easily the required 23 trait repeatability model. Associated (co)variance components were estimated as outlined by Misztal *et al.* (1995) on a subset with no missing values. These computations were based on 9,777 Walloon and 934 French records of 7,514 DPB cows. The used model was very close to the model used for Holstein cattle in the Walloon region of Belgium (Croquet *et al.* 2006) but without heterogeneous variance adjustments. The model had five fixed effects with a herd-year-month of classification x classifier-system contemporary group effect where classifier-system distinguished between classifiers working in the three recognized systems (Walloon, France-old and France-new). Also included were age and lactation stage effects distinguishing between fine overall effects and larger classifier-system specific effects. A direct implementation of a genomic single-step model in this setting was however not available. Therefore, in a second step, using estimated breeding values (EBV) and associated reliabilities (REL) generated in the first step, the Bayesian approach as proposed by Vandenplas *et al.* (2014) and used by Colinet *et al.* (2018), but only limited to local data, was used. Conceptually, this is equivalent to replacing the pedigree-based relationship matrix A by a combined matrix H transforming EBV into GEBV and REL into genomically enhanced REL (GREL) thus approximating closely an ssGBLUP.

Genome wide association study (GWAS). The SNP effects were estimated individually for each of the 23 traits using the postGSf90 software (Aguilar *et al.* 2014). The vector animal effects (GEBV) was separated in genotyped (a_g) and ungenotyped animals (a_{ng}). The GEBV of genotyped animals were considered being a function of the SNP effects, $a_g = Zu$, where Z is a matrix relating genotypes of each locus to SNP marker effect and u is a vector of the SNP marker effect. The variance of animal effects was assumed as:

$$\mathrm{Var}\left(\mathbf{a}_g\right) = \mathrm{Var}\left(\mathbf{Zu}\right) = \mathbf{ZDZ}'\sigma_u^2 = \mathbf{G}\sigma_a^2$$

where D is a diagonal matrix of weights for variances of markers (D = I), σ_u^2 is the genetic additive genetic variance captured by each SNP marker when the weighted relationship matrix (G) was built with equal weights.

The SNP effects were obtained using the following equation:

$$\hat{\mathbf{u}} = \lambda \mathbf{DZ}'\mathbf{G}^{-1}\hat{\mathbf{a}}_g = \mathbf{DZ}'\left[\mathbf{ZDZ}'\right]^{-1}\hat{\mathbf{a}}_g$$

where λ was defined as a normalizing constant (VanRaden 2008) and the percentage of genetic variance explained by the ith genomic region was estimated as:

$$\left(\mathrm{Var}\left(\mathbf{a}_i\right)/\sigma_a^2\right)\times 100 = \left(\mathrm{Var}\left(\sum_{j=1}^{25}\mathbf{Z}_j\hat{\mathbf{u}}_j\right)/\sigma_a^2\right)\times 100$$

where a_i is the genetic value of the ith region that consists of 25 adjacent SNPs, σ_a^2 is the total genetic variance, Z_j is the vector of the SNP content of the jth SNP for all individuals, and $\hat{\mathbf{u}}_j$ is the marker effect of the jth SNP within the ith region. The results were presented by the proportion of variance explained by each window of 25 adjacent SNPs and windows explaining for at least 1.0% of the total additive genetic variance were identified.

Results

Heritabilities, repeatabilities and correlations. Heritabilities across all 23 traits ranged from 0.15 for foot angle to 0.74 for stature. Repeatability ranged from 0.28 for foot angle to a very high 0.84 for stature. Table 1 shows heritabilities and correlations among the 5 overall traits.

Genome wide association studies (GWAS). Throughout the 23 traits, there were two important genomic regions. The genomic region located between 20.8 to 22.9 Mb on BTA26 was associated with traits including chest width, overall development (Figure 1), and stature. A region located between 5.3 to 7.3 Mb on BTA2 was associated with muscularity traits, but also with overall rump, overall udder (Figure 1), udder depth, and rear udder.

Table 1. Genetic (above diagonal) and phenotypic (below diagonal) correlations, and heritabilities (diagonal) for overall traits.

	Udder	Development	Feet and legs	Muscularity	Rump
Udder	0.37	0.15	0.10	-0.66	0.48
Development	0.08	0.63	0.26	-0.19	0.47
Feet and legs	0.10	0.15	0.19	0.00	0.12
Muscularity	-0.36	-0.08	0.01	0.60	-0.65
Rump	0.28	0.28	0.06	-0.39	0.42

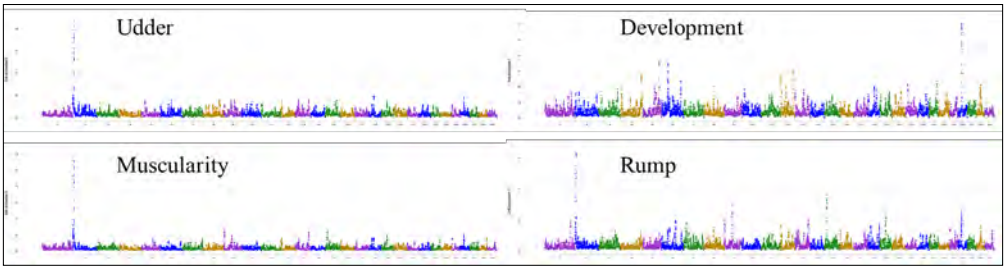


Figure 1. Additive genetic variance explained by windows of 25 adjacent SNP across chromosomes for the 4 overall traits udder, development, muscularity, and rump in Dual-Purpose Blue cows.

Discussion

Heritabilities, repeatabilities and correlations. Heritabilities were, as expected, with a tendency on the upper bound compared to those reported in literature. In general, repeatability tended to be high too. On a genetic level, an expected strong opposition between udder and muscularity appeared clearly, underlying the importance to include both in an adapted breeding objective for this dual-purpose breed.

Genome wide association studies (GWAS). At least one study in a completely different breed (Holstein) has associated BTA26 with an overall development trait (body form composite index) (Ashwell *et al.* 2005). Previous studies showed that SNPs inside the region on BTA2 position 5.3 to 7.3 Mb in various breeds of dairy and beef cattle were linked to mutations in *MSTN* (myostatin), the cause of the double-muscled phenotype in cattle, in general considered as the most important functional gene in the region found on BTA2. A visible distinct muscular hypertrophy (mh) occurs with high frequency in the DPB breed, linked to a specific deletion (Grobet *et al.* 1997), without being homozygous as in BBB. As explained previously, a large part of the DPB population is genotyped for this mutation. Given the large variance on BTA2, strategies may be useful to include mh in the future genomic evaluation system as already proposed for calving difficulty and associated traits (Mota *et al.* 2017).

Acknowledgements

Support of the Walloon Government and the INTERREG FWV program are acknowledged. H. Wilmot, as research fellow, acknowledges the support of the FNRS (Brussels, Belgium).

References

Aguilar I., Misztal I., Tsuruta S., Legarra A., and Wang, H. (2014) Proc. of the 10th WCGALP, Vancouver, Canada.

BlueSter. (2021). BlueSter. Available at: https://www.projet-bluester.eu/

Ashwell M, Heyen D, Weller J, Ron M, Sonstegard T, *et al.* (2005) J. Dairy Sci. 88(11):4111-4119. https://doi.org/10.3168/jds.S0022-0302(05)73095-2

Colinet F.G., Vandenplas J., Vanderick S., Hammami H., Mota R.R., *et al.* (2018) Animal 12(5):898-905. https://doi.org/10.1017/S1751731117002324

Croquet C., Mayeres P., Gillon A., Vanderick S., and Gengler N. (2006) J. Dairy Sci. 89(6):2257-2267. https://doi.org/10.3168/jds.S0022-0302(06)72297-4

Grobet L., Royo Martin L.J., Poncelet D., Pirottin D., Brouwers B., *et al.* (1997) Nat. Genet. 17(1):71-74. https://doi.org/10.1038/ng0997-71

Misztal I., Weigel K., and Lawlor T.J. (1995) J. Dairy Sci 78(8):1862-1872. https://doi.org/10.3168/jds.S0022-0302(95)76811-4

Mota R.R., Mayeres, P., Bastin, C., Glorieux, G., Bertozzi, C., *et al.* (2017) J. Anim. Sci. 95(10):4288-4299. https://doi.org/10.2527/jas2017.1748

Vandenplas J., Colinet F.G., and Gengler N. (2014) Genet. Sel. Evol. 46(1):59. https://doi.org/10.1186/s12711-014-0059-3

VanRaden P.M. (2008) J. Dairy Sci. 91(11):4414-4423. https://doi.org/10.3168/jds.2007-09801

191. Genetic background of productive and adaptive features in the West African indigenous cattle breeds in Benin

S.F.U. Vanvanhossou[1*], T. Yin[1], I.J. Giambra[1], K. Brügemann[1], L.H. Dossa[2] and S. König[1]

[1]Institute of Animal Breeding and Genetics, Justus-Liebig-University Gießen, 35390 Gießen, Germany; [2]School of Science and Technics of Animal Production, Faculty of Agricultural Sciences, University of Abomey-Calavi, Abomey-Calavi, 03 BP 2819 Jéricho Cotonou, Benin; seyi.f.vanvanhossou-2@agrar.uni-giessen.de

Abstract

This study aimed to investigate the genetic patterns underlying productive and adaptive features in West African indigenous cattle breeds. Multi-breed genome-wide association studies for two morphometric traits (height at withers and sacrum height) were performed considering 449 animals from four indigenous cattle breeds (Lagune, Somba, Borgou and Pabli) in Benin. In addition, 67 animals were selected from the previous sample for next-generation sequencing targeting six milk protein genes (*CSN1S1, CSN2, CSN1S2, CSN3, LALBA, LGB*). A total of eight single nucleotide polymorphisms were associated with the morphometric traits. The loci and the potential candidate genes were not only involved in conformation, growth or carcass traits, but also in adaptive trait mechanisms including response to heat stress, immune functions or disease resistances. Furthermore, high genetic polymorphism was observed in the milk protein genes, and a total of 17 milk protein variants including four previously undescribed variants ($CSN3^{K}$, $LALBA^{F}$, LGB^{B1}, LGB^{K}) were identified.

Introduction

African animal genetic resources are highly diverse and represent a valuable asset for food security and the livelihood of farmers (Mwai *et al.,* 2015). Unfortunately, they are increasingly threatened by several factors including climate change, diseases and lack of adequate management policies (Mwai *et al.,* 2015). In addition, the lack of proper genetic characterization impedes their promotion, genetic amelioration and sustainable use. Recent genetic studies on some African breeds confirmed their richness and valuable features including resistance to disease, heat stress as well as their economic potential (e.g. Kim *et al.,* 2020). Novel genomic statistical methods offer new possibilities to investigate African breeds and to unravel their genetic potential.

The aim of this study was to use up-to-date genomic tools to characterise West African indigenous cattle breeds in Benin. For this purpose, morphometric traits, known as proper early indicators for animal growth, health, welfare and longevity (Kominakis *et al.,* 2017) were targeted, and their genomic architectures were investigated in genome-wide association studies (GWAS). In addition, four casein genes (*CSN1S1, CSN2, CSN1S2, CSN3*) on *Bos taurus* autosome (BTA) 6, and two whey protein genes (*LALBA* and *LGB*) on BTA5 and BTA11, respectively, were characterized, because they are associated with milk production traits in several cattle breeds (Caroli *et al.,* 2009). Genetic polymorphisms in milk proteins genes and milk protein variants were detected using next-generation sequencing.

Materials & methods

Animal genotyping and multi-breed GWAS for morphometric traits. In total, 449 animals from four indigenous cattle breeds (Lagune, Somba, Borgou and Pabli) were sampled in Benin and genotyped using the Illumina BovineSNP50 BeadChip. SNP with a minor allele frequency <5%, genotyping call rate <90% and which are in Hardy-Weinberg equilibrium ($P \geq 10^{-6}$), as well as animals with genotyping call rate <95%, were excluded. The final genotype data were used to perform GWAS for the morphometric traits in PLINK (Purcell *et al.,* 2007), including height at withers (HAW) and sacrum height (SH). The genetic structure

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_191, © S.F.U. Vanvanhossou *et al.* 2022

of the animal population was estimated applying a discriminant analysis of principal components. The genetic-statistical model was defined as follows:

$$y = Xb + Wg + e$$

where y was a vector of observations for HAW and SH; b was a vector of fixed effects including age, sex, the origin (agro-ecological zone), breed, and linear discriminant functions; g was a vector for the SNP fixed effects; e was a vector of random residual effects with $\mathbf{e} \sim N(0, \mathbf{I}\sigma_e^2)$; and X, W were incidence matrices for b and g, respectively.

Significantly and suggestively associated SNP were identified using the Bonferroni-corrected significance threshold (P=1.55×10^{-6}) and chromosome-wide Bonferroni-corrected significance thresholds (pc = 0.05/m_c, with m_c = effective number of SNP for each chromosome).

Animal sequencing and detection of milk protein variants. Next-generation sequencing was performed using 67 animals selected from the previous samples. For this purpose, genotyping-in-thousands by sequencing (GT-seq) using the Illumina MiSeq V3 platform targeted the exon sequences, flanking intron sequences and parts of the 5′-upstream regions of milk protein genes (*CSN1S1, CSN2, CSN1S2, CSN3, LALBA, LGB*). The generated raw sequence data were processed to identify SNP, insertions and deletions (InDel). Variants detected in the coding regions were considered to annotate milk protein variants following the existing nomenclature (Caroli *et al.,* 2009; Gallinat *et al.,* 2013) and referring to the amino acid (AA) position in the mature proteins.

Results

Genomic regions associated morphometric traits. The multi-breed GWAS for HAW identified one significantly associated SNP (rs109126926 on BTA1, P=1.12×10^{-6}) and five chromosome-wide suggestively associated SNP (on BTA3, 17, 19 and 21, P=6.44×10^{-6}-5.04×10^{-5}) (Figure 1). The significantly associated SNP for HAW on BTA1 is located in the *VEPH1* gene. The SNP rs110369628 and rs109889052 on BTA19 are positioned close to each other (7.41 Mb), within the genes *PIK3R6* and *SSH2*, respectively. In addition, the SNP rs4163436 is positioned in the gene *CCDC117* on BTA17, whereas the two remaining associated SNP are not located in the vicinity of any gene. Sacrum height was associated with two suggestive SNP (P=2.35×10^{-5}-3.86×10^{-5}) and each is located in close distance to a gene (Figure 1). The SNP rs111001850 is positioned near the *LYPD8* gene on BTA7, and the SNP rs110441360 is mapped in relative proximity to the *PIK3R1* gene on BTA20.

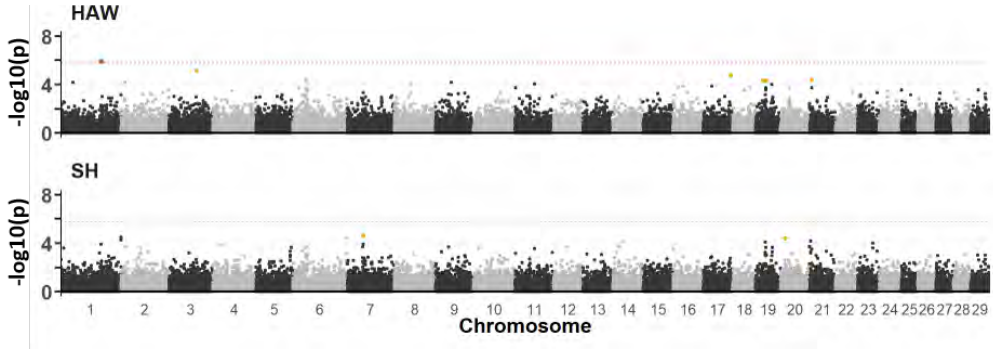


Figure 1. Manhattan Plots displaying GWAS results for height at withers and sacrum height in Beninese indigenous cattle breeds. Genowe-wide significant SNP are in red and chromosome-wide suggestive SNP are in yellow.

Milk protein variants in the Beninese indigenous cattle breeds. Within the coding regions, a total of 12 DNA polymorphisms caused 17 milk protein variants. The distribution of the milk protein variants in the investigated breeds is presented in Table 1. Two casein variants were identified for *CSN1S1 (CSN1S1^{B}* and *CSN1S1^{c}).* They were due the SNP rs43703010A>G located in exon 17. The variant *CSN1S1^{B}* and *CSN1S1^{C}* were associated with the reference alleles rs43703010A and rs43703010G, respectively. For *CSN2,* the casein variants *CSN2^{A1}*, *CSN2^{A2}* and *CSN2^{L}*, were identified. *CSN2^{A1}* corresponds to the reference sequence and was differentiated from *CSN2^{A2}* with the SNP rs43703011A>C. *CSN2^{L}* is due to the SNP rs715383373T>C in exon 7. In the *CSN1S2* gene, only one missense variant (rs441966828C>T) was identified, and is associated with the *CSN1S2^{B}* and *CSN1S2^{A}* protein variants. With regard to *CSN3,* three milk protein variants (*CSN3^{A,} CSN3^{B}, CSN3^{K}*) were detected. The *CSN3^{A}* and *CSN3^{B}* were associated with the SNP rs43703015T>C and rs43703016C>A in exon 4. The third detected milk protein variant was previously undescribed and was named *CSN3^{K}*, following the previous nomenclature in cattle. *CSN3^{K}* is caused by a novel SNP (BTA6:85656526C>T) in exon 4. Moreover, three known (*LALBA^{A}*, *LALBA^{B}*, *LALBA^{E}*) and one novel (*LALBA^{F}*) milk protein variants were detected in *LALBA*. *LALBA^{A}* differs from the reference variant *LALBA^{B}* with the SNP rs722550244G>C in exon 1. Near this SNP, another SNP rs7144688595C>T induces a new milk protein variant previously undescribed and therefore named *LALBA^{F}*. In addition, the variant *LALBA^{E}* due to the SNP rs465119286A>G was idenified. A missense SNP (rs109625649C>T) and a frameshift insertion induced one known (*LGB^{B}*) und two novel milk protein variants (*LGB^{B1}*, *LGB^{K}*) for *LGB*. *LGB^{B}* is associated with the reference allele rs109625649C, whereas the alternative allele rs109625649T is responsible for a novel milk protein variant called *LGB^{B1}*. *LGB^{K}* is induced by a frameshift insertion causing the emergence of a premature stop codon, and resulting in the reduction of the coding sequence of the *LGB* gene from a total of 178 to 104 AA with a complete exchange of the protein sequence from AA 92 to 104.

Discussion

The rather small number of detected SNP for the moderately to highly heritable polygenic traits is in line with our small sample size and its diversity (multi-breed), and is most likely due to a lack of power to detect variants with smaller effects. However, the functional annotation of the identified genes and comparison with previous studies confirm the association of the detected loci with conformation, growth and carcass traits in cattle or in other species. For instance, the *VEPH1* gene affects rump fat thickness in Nellore

Table 1. Distribution of milk protein variants in the Beninese indigenous cattle breeds.

Genes	Variants [1] (Ref. / Alt.)	Protein exchange	Variant Frequency (Ref. / Alt.)			
			Cross-breed (n=27)	Lagune (n=20)	Somba (n=20)	Total (n=67)
CSN1S1	B / C	p.Glu192Gly	0.62 / 0.38	0.97 / 0.03	0.93 / 0.08	0.81 / 0.19
CSN2	A1 / A2	p.His67Pro	0.39 / 0.61	0.80 / 0.20	0.63 / 0.37	0.57 / 0.43
	A1 / L	p.Val197Ala	0.97 / 0.03	1.00 / 0.00	1.00 / 0.00	0.99 / 0.01
CSN1S2	A / B	p.Ser8Phe	0.98 / 0.02	1.00 / 0.00	1.00 / 0.00	0.99 / 0.01
CSN3	B / A	p.Ile136Thr	0.50 / 0.50	0.75 / 0.25	0.80 / 0.20	0.72 / 0.28
		p.Ala148Asp	0.50 / 0.50	0.50 / 0.50	0.80 / 0.20	0.69 / 0.31
	B / **K**	p. Ala66Val	1.00 / 0.00	0.75 / 0.25	1.00 / 0.00	0.97 / 0.03
LALBA	B / A	p. Arg10Gln	0.88 / 0.12	1.00 / 0.00	1.00 / 0.00	0.95 / 0.05
	B / E	p. Ile41Val	0.88 / 0.12	1.00 / 0.00	0.95 / 0.05	0.93 / 0.07
	B / **F**	p. Arg10Trp	0.96 / 0.04	1.00 / 0.00	0.88 / 0.13	0.94 / 0.06
LGB	B / **B1**	p. Ala118Val	0.89 / 0.11	0.50 / 0.50	0.81 / 0.19	0.80 / 0.20
	B / **K**	p.Thr92Asnfs*13	0.96 / 0.04	1.00 / 0.00	1.00 / 0.00	0.98 / 0.02

[1] The variants in bold have been newly annotated.

cattle (Santana *et al.,* 2015). The *PIK3R6* gene was previously related to body size in sheep (Kominakis *et al.,* 2017). In addition, the *CCDC117* gene associated with HAW is implicated in feed intake and heat stress regulation in cattle (Lindholm-Perry *et al.,* 2016). Similarly, *PIK3R1* is associated with resistance to disease infections in cattle (Marino *et al.,* 2017). The observed overlap of genomic regions associated with morphometric and adaptive traits is reported in previous studies, and is in line with the selection strategies for animal adaptation using morphometric traits in African cattle herds.

The results of this study also show that the milk protein gene regions display several known and previously unknown polymorphisms in the Beninese cattle breeds (Caroli *et al.,* 2009; Gallinat *et al.,* 2013). The identified variants confirm the distribution of previously described milk protein alleles in Beninese and African cattle breeds (Moazami Goudarzi *et al.,* 2001). However, the detection of several previously unknown milk protein variants (*CSN3*^{K}, *LALBA*^{F}, *LGB*^{B1}, *LGB*^{K}) support the unicity and genetic diversity of the West African cattle breeds. The identified variants are expected to be further investigated for potential association with milk quantity and quality in the Beninese indigenous breeds. This study corroborates the necessity to enhance genetic studies in local breeds from tropical countries, contributing to a deeper understanding of genetic mechanisms for adaptation and the identification of novel genes for productive performances.

References

Caroli A.M., Chessa S., and Erhardt G.J. (2009) J. Dairy Sci. 92(11):5335-5352. https://doi.org/10.3168/jds.2009-2461

Gallinat J.L., Qanbari S., Drögemüller C., Pimentel E.C.G., Thaller G., *et al.* (2013) J. Dairy Sci. 96(1): 699–709. https://doi.org/10.3168/jds.2012-5908

Kim K., Kwon T., Dessie T., Yoo D., Mwai O.A., *et al.* (2020) Nat. Genet. 52(10), 1099–1110. https://doi.org/10.1038/s41588-020-0694-2

Kominakis A., Hager-Theodorides A.L., Zoidis E., Saridaki A., Antonakos G., *et al.* (2017) Genet. Sel. Evol. 49(1): 1–16. https://doi.org/10.1186/s12711-017-0316-3

Lindholm-Perry A.K., Butler A.R., Kern R.J., Hill R., Kuehn L.A., *et al.* (2016) Anim. Genet. 47(4): 408–427. https://doi.org/10.1111/age.12440

Marino R., Capoferri R., Panelli S., Minozzi G., Strozzi F., *et al.* (2017) Immun. 91: 259–271. http://dx.doi.org/10.1016/j.molimm.2017.08.017

Moazami Goudarzi K., Belemsaga D.M., and Ceriotti G. *et al.*, (2001) Rev. Elev. Med. Vet. Pays Trop., 54(2):129-138. https://doi.org/10.19182/remvt.9791

Mwai O., Hanotte O., Kwon Y.-J., and Cho S. (2015) Asian-australas J. Anim. Sci. 28(7), 911–921. http://dx.doi.org/10.5713/ajas.15.0002R

Purcell S., Neale B., Todd-Brown K., Thomas L., Ferreira M.A.R, *et al.* (2007) Am. J. Hum. Genet. 81(3): 559–575. https://doi.org/10.1086/519795

Santana M., Ventura R.V., Utsunomiya Y.T., Neves H.H.R., Alexandre P.A. *et al.* (2015) J. Anim. Breed. Genet. 132, 420–427. https://doi.org/10.1111/jbg.12167

192. Genetic diversity and population structure of Slovenian local breed Drežnica goat

N. Pogorevc[1*], M. Simčič[1], N. Khayatzadeh[2], J. Sölkner[2], B. Berger[3], D. Bojkovski[1], M. Zorc[1], P. Dovč[1], I. Medugorac[4] and S. Horvat[1]

[1]University of Ljubljana, Biotechnical Faculty, Jamnikarjeva 101, 1000 Ljubljana, Slovenia; [2]BOKU, Gregor Mendel Str. 33, 1180 Vienna, Austria; [3]AREC Raumberg-Gumpenstein, 4601 Thalheim b. Wels, Austria; [4]LMU Munich, Veterinary Faculty, Lena-Christ-Straβe 48, 8215 Martinsried/Planegg, Germany; neza.pogorevc@bf.uni-lj.si

Abstract

Genomic analyses in livestock animals, beside the obvious contribution to applications in agriculture, provide important insights into biodiversity research and species conservation. The Drežnica goat (*Capra hircus*) is the only Slovenian local goat breed with critically endangered status. Its robustness and adaptability to the hard Alpine environment are important traits worth preserving especially in light of climate changes to even harsher conditions. The aim of the study was to analyse the genetic structure of the breed and its independent origin within other goat breeds. We genotyped Drežnica goats with an Illumina 50K SNP chip and analysed the genetic diversity and population structure within the breed and genetic relatedness to other breeds from the Alpine region. The results showed that the breed had a large number of private alleles and retained moderate to high genetic variability. Estimates of genetic distances showed Drežnica goat's distinct genetic origin and the admixture with foreign breeds is extremely low.

Introduction

In the era of preferential rearing of high-performance breeds, local breeds are slowly being recognised as an important source of phenotypes with good production potential in environments with harsh conditions. Due to high selection pressure for economically important traits associated with indoor housing system, cosmopolitan goat breeds have diminished or lost alleles for adaptation to extreme environments like high-altitude areas. On the other hand, low pressure of artificial selection and high pressure of natural selection are two possible reasons why local goat breeds maintain those adaptive traits (Marsoner *et al.*, 2018). In Slovenia, the only local goat breed is called Drežnica goat. Because of its small population size, the breed is at high risk of extinction. Based on production purpose, the population is divided into dairy and meat production type. The dairy subpopulation is mainly widespread around the Bovec region (upper dot in Figure 1), while the meat type is concentrated in the Drežnica-Kobarid region (bottom dot in Figure 1). Drežnica goat is reared according to traditional alpine production system with short span of indoor rearing in the winter and free-range grazing on mountain pastures from spring to autumn. Adaptation to the hard conditions of the Alpine environment allows the breed to survive without human help, which are important traits worth conserving. Nevertheless, this breed lacked genetic analyses on which breeding and conservation programmes could be based. Additionally, the variability in coat colour raised doubts that the breed is rather a genetically mixed population of different breeds. For this reason, we genotyped Drežnica goats with an Illumina goat chip containing ~50.000 SNP markers (Tosser-Klopp *et al.*, 2014) to analyse the parameters of genetic diversity, population structure, and possible occurrence of crossbreeding with other breeds.

Materials & methods

We selected and sampled 133 representative animals of Drežnica goat and acquired six goat breeds from neighbouring countries Austria and Italy (Pogorevc *et al.*, 2021). Goats were genotyped with the Illumina Goat SNP50 BeadChip (Tosser-Klopp *et al.*, 2014). Additionally, we included the genome-wide genotypes of 16 goat breeds from the Alpine area available in the DRYAD repository. After forming the dataset

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_192, © N. Pogorevc *et al.* 2022

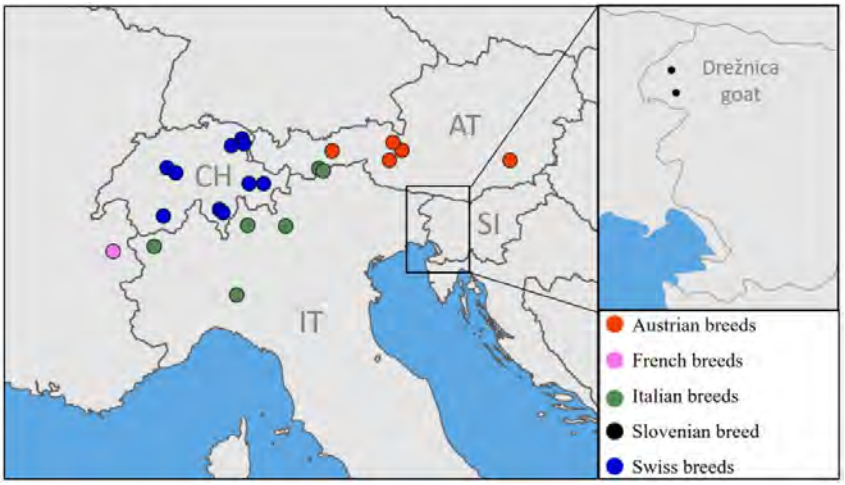


Figure 1. Geographic locations of 23 alpine goat breeds included in our datasets.

(Figure 1, Table 1), we performed quality control excluding SNPs with unknown chromosomal positions, inconsistent information according to relatives, a call rate <95%, and a minor allele frequency <0.025. We imputed missing genotypes for 48,297 autosomal SNPs that passed quality control (Browning *et al.*, 2007). For optimization, genetic diversity estimation, and genetic distance calculation described below, we used 5,658 short haplotypes consisting of four adjacent SNPs, each less than 50 kb apart.

Table 1. Number of private alleles and allelic richness of 23 alpine goat breeds.

Goat breed	Country	Sample size	Number of private alleles	Mean allelic richness	Publication
Camosciata Alpine	France	33	140	8.37	Oget *et al.*, 2019
Valdostana	Italy	17	142	8.14	Colli *et al.*, 2018
Saanen	Switzerland	38	107	6.96	Burren *et al.*, 2016
Valais	Switzerland	31	84	6.62	Burren *et al.*, 2016
Nera Verzasca	Switzerland	37	130	8.12	Burren *et al.*, 2016
Tessin grey	Switzerland	28	98	8.35	Burren *et al.*, 2016
Toggenburg	Switzerland	22	50	6.15	Burren *et al.*, 2016
Appenzell	Switzerland	19	43	6.19	Burren *et al.*, 2016
Booted	Switzerland	19	43	7.09	Burren *et al.*, 2016
Peacock	Switzerland	31	92	7.91	Burren *et al.*, 2016
Grisons striped	Switzerland	38	139	7.62	Burren *et al.*, 2016
Chamois Coloured	Switzerland	50	67	7.86	Burren *et al.*, 2016
Orobica	Italy	20	105	7.46	Colli *et al.*, 2018
Adamello Blond	Italy	20	225	9.55	Colli *et al.*, 2018
Camosciata Alpine	Italy	48	50	7.91	Colli *et al.*, 2018
Valpassiria	Italy	20	191	9.73	Colli *et al.*, 2018
Passeier	Italy	19	150	8.95	Pogorevc *et al.*, 2021
Blobe goat	Austria	28	189	9.05	Pogorevc *et al.*, 2021
Chamois Coloured	Austria	22	68	8.39	Pogorevc *et al.*, 2021
Pinzgau	Austria	22	158	8.54	Pogorevc *et al.*, 2021
Tauern Pied	Austria	24	129	7.03	Pogorevc *et al.*, 2021
Drežnica goat	Slovenia	50	301	7.86	Pogorevc *et al.*, 2021
Styrian Pied	Austria	27	382	9.94	Pogorevc *et al.*, 2021

To obtain objective results we optimized the dataset by excluding goats with admixture signatures from other breeds. We used a multivariate outlier test (Filzmoser *et al.*, 2005) and excluded outliers that had weaker genetic relationships with individuals from the same breed, stronger genetic relationships with individuals from other breeds, and a larger proportion of foreign alleles. Based on identity by descent we estimated unified additive relationships (UAR) and excluded closely related goats with UAR>0.25 (Yang *et al.*, 2010). After both optimization steps, we estimated genetic diversity parameters by calculating the number of private alleles and mean allelic richness. Genetic relationships between the breeds were calculated with Nei's distances (Nei *et al.*, 1983), which were used for the construction of a phylogenetic neighbour net with the program SplitsTree4 (Huson and Bryant, 2006). To determine the population structure, we used single-SNP alleles and analysed them with the programs SmartPCA (Price *et al.*, 2009) and Admixture (Alexander *et al.*, 2009).

Results

With allele count parameters based on haplotypes, we analysed genetic diversity within the breeds (Table 1). Drežnica goat had 301 private alleles, which is the second highest number. In terms of allelic richness, the Drežnica goat was positioned in the middle of breeds from the Alpine environment with a mean value of 7.86. The neighbour net tree of Nei's distances placed the distinct branch of Drežnica goat close to geographically neighbouring breeds from Austria and Italy (Figure 2B). These results were consistent with principal component analysis (PCA), which distinguished three major groups (Figure 2A). The cosmopolitan breeds from four countries (Camosciata Alpine and Chamois Coloured) formed one cluster, Austrian, Italian and Swiss breeds formed the second and the third cluster was the Drežnica goat. Austrian breeds were closest

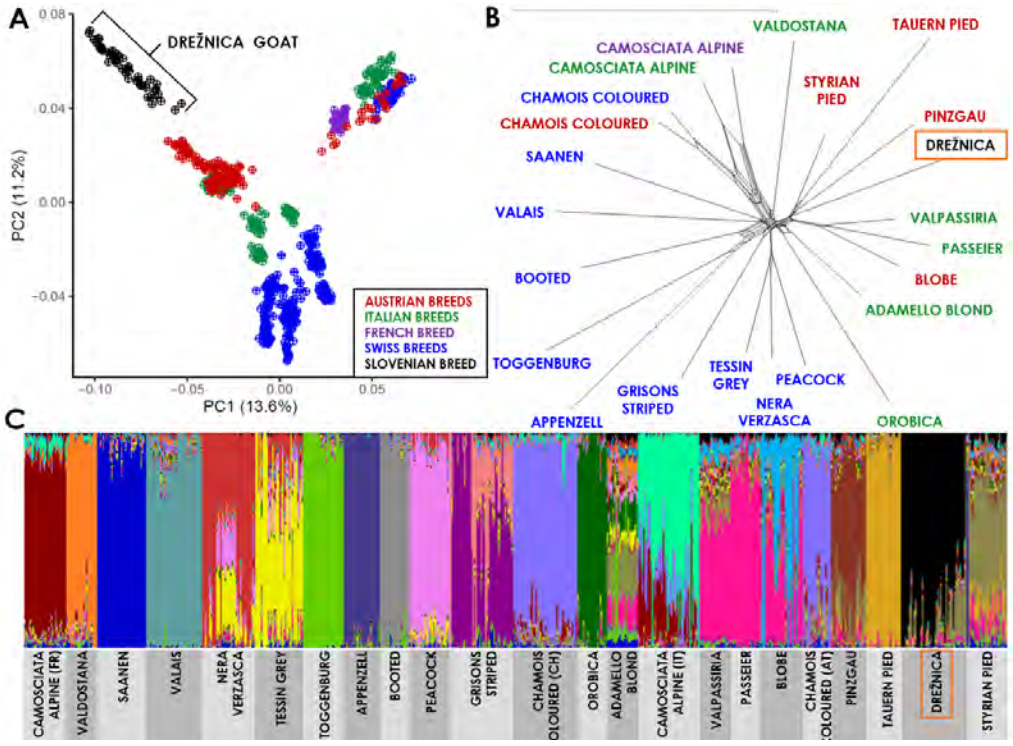


Figure 2. Genetic relationships of alpine breeds based on PCA analysis (A), Nei's distances (B), and their population structures of admixture analysis (C).

to the Drežnica goat, followed by Italian and then Swiss breeds. The lowest cross-validation error of the admixture analysis determined the optimal number of 21 hypothetical populations for our dataset (Figure 2C). There were 11 breeds that formed homogeneous clusters, while some individuals or whole populations of other breeds were slightly admixed, resulting in rather heterogeneous clusters. Drežnica goats showed mainly homogeneous blocks with traces of admixture in few individuals. With admixture plot, we obtained more detailed population structures, further supporting previous results observed in PCA and neighbour net graphs.

Discussion

Genomic characterization of endangered local breeds is essential for their efficient breeding and conservation programs. To achieve this, we compared the Drežnica goat with genotypes of 22 goat breeds from the Alpine environment. From the high number of private alleles and average allelic richness, we can conclude that the Drežnica goat retained distinct genetic signature. With the neighbour-joining tree, we demonstrated that this high diversity is not primarily due to admixture but to the distinct genetic origin of the Drežnica goat with its own branch. This supported the PCA analysis, where the Drežnica goat formed its own cluster clearly separated from other alpine breeds. An admixture analysis revealed that several Drežnica goats contained some admixture signatures, else the breed had a uniform population structure without clear separation of both production types. Since the breed has experienced bottlenecks in the past and the population size is low today, these diversity results are encouraging for future breeding and conservation programmes. Our study demonstrating distinct genetic identity therefore places the Drežnica goat on the world map of goat breeds.

References

Alexander D.H. *et al.* (2009). Genom Res 19:1655–64. https://doi.org/10.1101/gr.094052.109

Browning S.R. and Browning B. L. (2007) Am J Hum Genet 81(5):1084–1097. https://doi.org/10.1086/521987.

Burren A. *et al.* (2016) Anim Genet 47(6):727–739. https://doi.org/10.1111/age.12476

Colli L. *et al.* (2018) Genet Sel Evol 50(1):1–20. https://doi.org/10.1186/s12711-018-0422-x

Filzmoser P. *et al.* (2004) Comput Geosci 31(5):579–587. https://doi.org/10.1016/j.cageo.2004.11.013

Huson D.H. and Bryant D. (2006) Mol Biol Evol 23(2):254–67. https://doi.org/10.1093/molbev/msj030

Marsoner T. *et al.* (2018) Ecol Indic 94:55–63. https://doi.org/10.1016/j.ecolind.2017.06.046

Nei M. *et al.* (1983). J Mol Evol 19:153–170. https://doi.org/10.1007/BF02300753

Nicoloso L. *et al.* (2015) Genet Sel Evol 47:1–10. https://doi.org/10.1186/s12711-015-0140-6

Oget C. *et al.* (2019) Anim Genet 50(1):54–63. https://doi.org/10.1111/age.12752

Pogorevc N. *et al.* (2021) BMC Genom 22 (546). https://doi.org/10.1186/s12864-021-07802-z

Price A.L. *et al.* (2006) Nat Genet 38(8):904–9. https://doi.org/10.1038/ng1847

Tosser-Klopp G. *et al.* (2014) PLoS One 9(1). https://doi.org/10.1371/journal.pone.0152632

Yang J. *et al.* (2010) Nat Genet 42(7):565–9. https://doi.org/10.1038/ng.608

193. Genetic architecture of a composite beef cattle population

E. Hay[1*], S. Toghiani[2], A.J. Roberts[1], T. Paim[3], L.A. Kuehn[4] and H.D. Blackburn[5]

[1]USDA Agricultural Research Service, Fort Keogh Livestock and Range Research Laboratory, Miles City, MT 59301, USA; [2]USDA Agricultural Research Service, Beltsville Agricultural Research Center, Beltsville, MD, 20705, USA; [3]Instituto Federal de Educação, Ciência e Tecnologia Goiano, Campus Rio Verde, Rio Verde, Goias, Brazil, 76200-000; [4]USDA, Agricultural Research Service, US Meat Animal Research Center, Clay Center, 68933, USA; [5]National Animal Germplasm Program, USDA, Fort Collins, CO, 80521, USA; elhamidi.hay@usda.gov

Abstract

In this study, genomic information was used to evaluate the genetic composition and characteristics of a composite beef cattle population. This composite population consisted of 50% Red Angus, 25% Charolais, 25% Tarentaise. The results showed an increase in the proportion of Tarentaise to approximately 57% while Charolais decreased to approximately 5%, and Red Angus decreased to 38%. Evaluating runs of homozygosity showed the formation of new haplotypes in the composite population around the sixth generation. The study showed that progenitor combinations are not stable over generations and that either direct or natural selection plays a role in modifying the proportions. Furthermore, the results showed that Tarentaise contributed useful attributes to the composite in a cool semi-arid environment and suggests a re-exploration of this breed's role may be warranted.

Introduction

Composite breeds are traditionally developed by combining two or more breeds that have complementary traits. In 1979, the USDA Agricultural Research Service's station in Miles City, Montana, US developed a composite population using *Bos taurus* breeds with distinct production characteristics. For this cold semi-arid environment, a three-breed composite (50% Red Angus, 25% Charolais, 25% Tarentaise) was formed to evaluate its performance in a limited nutrient production system. In brief, the CGC was started with breeding 300 Red Angus dams to either 14 Charolais sires or 12 Tarentaise sires using artificial insemination. The F1 individuals were produced between years 1980 to 1984. Crossing of F1 individuals began in 1981 and F_2 individuals were generated from 1982 until 1988. The F3 generation was produced between 1984 and 1990. With this mating plan the CGC effective population size was computed to be 95.7. Subsequent generations have been produced since 1987.

The objective of this study was to evaluate the genetic composition of CGC, if changes have occurred, what parental populations were favored/disfavored and given the small population size can it adequately be a research model for composite beef populations.

Materials & methods

The dataset consisted of a total of 248 animals, where 10 were founder crosses to form CGC, 69 were CGC, 61 Red Angus, 79 Charolais and 29 Tarentaise. The ten early founder crosses were born in 1980 and were 50% Charolais and 50% Red Angus or 50% Tarentaise and 50% Red Angus. The CGC animals were born between 1986 and 2017. All animals were genotyped with High-density SNP panel (777,962 SNP, BovineHD Beadchip, Illumina, San Diego, CA, USA). Markers with minor allele frequency lower than 1% were removed. In addition, markers with a call rate lower than 95% or not physically mapped to the bovine genome assembly UMD3.1 were discarded from the analyses. Markers on sex chromosomes were also removed, the final number of SNP genotypes used was 203,297.

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_193, © E. Hay *et al.* 2022

To evaluate the genetic relationship between the pure breeds and CGC, principal component and ADMIXTURE (Alexander *et al.* 2009) analyses were conducted. Additionally, runs of homozygosity (ROH) analysis was conducted using PLINK (Purcell *et al.* 2007). The parameters were set to be a minimum run length equal to 1 Mb with a minimum of SNP markers calculated using the equation in Lencz *et al.* (2007) and allowing runs to contain up to two heterozygotes and five missing genotypes with a maximum gap equal to 50 kb and a minimal SNP density of 1 SNP per 50 kb.

Results

The principal components analysis showed population divergence with CGC clustered intermediately among the progenitor breeds (Figure 1). Furthermore, the progenitor breed proportions of CGC animals were estimated by ADMIXTURE and evaluated according to the generation number. The results showed decreased proportions of Charolais and Red Angus and an increase in Tarentaise in subsequent generations of CGC. As shown in Figure 2, there was an increase in the proportion of Tarentaise to approximately 57% while Charolais decreased to approximately 5%, and Red Angus decreased to 38%.

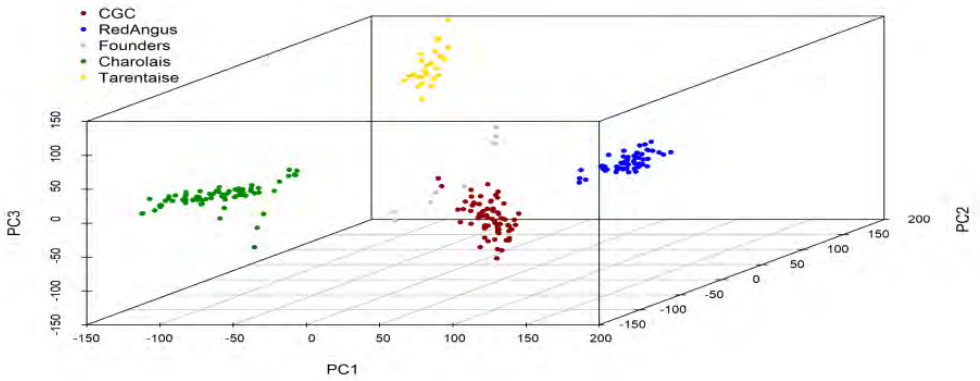


Figure 1. PCA plot for CGC founders, CGC, Red Angus, Charolais and Tarentaise.

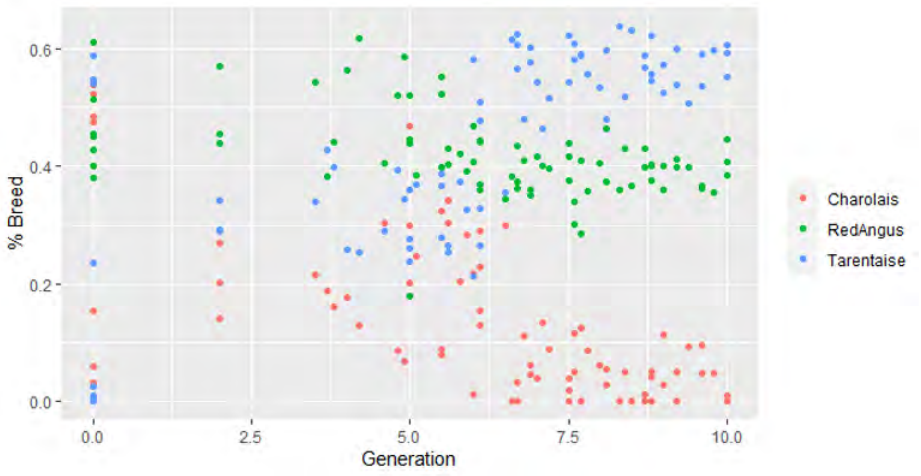


Figure 2. Genomic breed composition of CGC animals according to their equivalent generation number using Admixture analysis.

The ROH analysis showed Red Angus to have the highest number of ROH per animal and the longest total length of ROH (Table 1). Additionally, runs of homozygosity were classified according to the expected number of prior generations tracing back to a common ancestor. Figure 3 shows the change in the average ROH length according to the equivalent generation number of animals. Around the sixth-generation formation of longer ROH was observed.

Table 1. Summary description of the number, individual and total length of ROH segments (in Mb) per animal.

Parameters[1]	Breeds	Mean	SD	Min	Max
ROH_n	*CGC*	40.76	10.86	2	65
	Charolais	45.21	9.80	19	78
	Red Angus	92.26	12.09	70	127
	Tarentaise	70.72	10.99	50	98
ROH_L	*CGC*	3.33	3.60	1	55.41
	Charolais	3.70	4.40	1	100.54
	Red Angus	3.79	3.99	1	85.66
	Tarentaise	2.97	3.22	1	52.87
ROH_T_L	*CGC*	141.10	56.93	4.55	262.47
	Charolais	172.46	88.49	37.99	525.16
	Red Angus	349.92	81.52	226.98	738.47
	Tarentaise	209.27	44.21	115.42	304.64

[1] ROH_n = number of individual ROH segments per animal, ROH_L = length of individual ROH segments in Mb, ROH_{TL} = total length of ROH segments in Mb per animal.

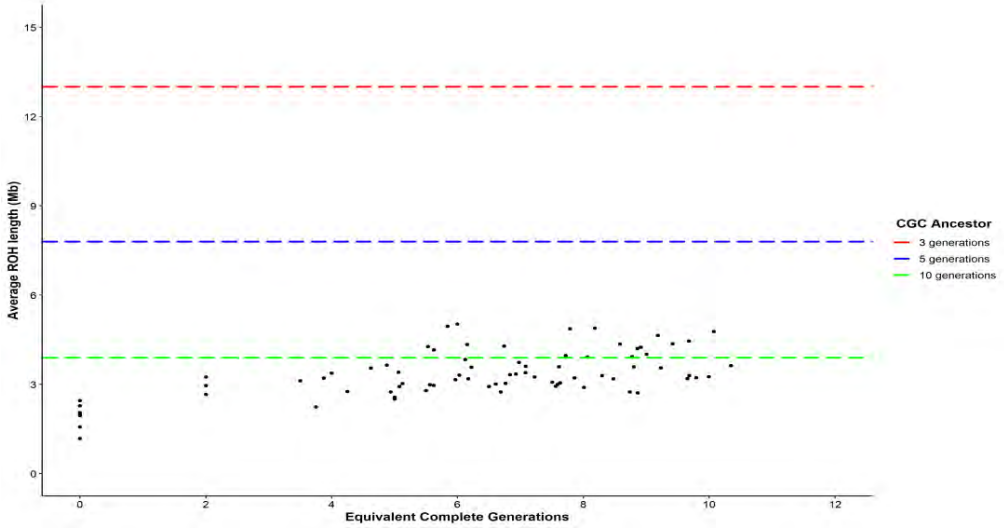


Figure 3. Runs of homozygosity observed in CGC cattle according to the equivalent generation number. Dashed lines indicate the length threshold for ROH that relates to a common ancestor at 3, 5, and 10 prior generations (red, blue, and green, respectively).

Discussion

Our results suggest several points about composite population development and potential use. As with other studies (Blackburn *et al.* 2017; Paim *et al.* 2020a; Paim *et al.* 2020b) with advancing generations the composite population develops its own unique cluster when compared to progenitor breeds (Figure 1). The formation of CGC cluster (generation 6) appears to be in concordance with previous work where a Brangus cluster occurred between generation 5 and 6 (Paim *et al.*, 2020a).

Evaluating runs of homozygosity showed new islands in CGC formed around generation 6 (Figure 3). This could be interpreted as when new breed formation occurred. The ROH based genomic inbreeding also showed a sharp increase to 10% in generation 6 and is different than the UN Food and Agriculture Organization (FAO) Guidelines for *in vivo* conservation of animal genetic resources (FAO 2013) which suggest three generations of inter se mating are required to establish a new composite breed after the initial crosses have been made.

The CGC population was constructed for increased cow productivity with the ability to perform in the northern Great Plains and has not been selected for meat quality, feedlot performance, nor increased mature size. Rather, the emphasis on selection tended to focus upon cow fertility and calf weaning weight. Given these selection objectives it is not surprising that the composition of the progenitor breeds shifted with time for CGC. Specifically, the proportion of Tarentaise shifted from a target of 25% to an average 57% in the last two generations, Red Angus decreased from a target of 50% to approximately 38%, and Charolais decreased from a target of 25% to approximately 4% (Figure 2).

References

Alexander, D. H., J. Novembre and K. Lange, 2009 Fast model-based estimation of ancestry in unrelated individuals. Genome research 19**:** 1655-1664.

Blackburn, H., B. Krehbiel, S. Ericsson, C. Wilson, A. Caetano *et al.*, 2017 A fine structure genetic analysis evaluating ecoregional adaptability of a Bos taurus breed (Hereford). PloS one 12**:** e0176474.

Lencz, T., C. Lambert, P. DeRosse, K. E. Burdick, T. V. Morgan *et al.*, 2007 Runs of homozygosity reveal highly penetrant recessive loci in schizophrenia. Proceedings of the National Academy of Sciences 104**:** 19942-19947.

Paim, T. d. P., E. Hay, C. Wilson, M. Thomas, L. Kuehn *et al.*, 2020a Dynamics of genomic architecture during composite breed development in cattle. Animal genetics 51**:** 224-234.

Paim, T. d. P., E. H. A. Hay, C. Wilson, M. G. Thomas, L. A. Kuehn *et al.*, 2020b Genomic breed composition of selection signatures in Brangus beef cattle. Frontiers in genetics 11**:** 710.

Purcell, S., B. Neale, K. Todd-Brown, L. Thomas, M. A. Ferreira *et al.*, 2007 PLINK: a tool set for whole-genome association and population-based linkage analyses. The American journal of human genetics 81**:** 559-575.

194. Variation in patterns of recombination result in genetic variation in intrachromosomal shuffling in the domestic pig

C. Brekke[1*], S.E. Johnston[2], A.B. Gjuvsland[3] and P. Berg[1]

[1]Norwegian University of Life Sciences, Oluf Thesens vei 6, 1433 Ås, Norway; [2]University of Edinburgh, EH9 3FL, Edinburgh, United Kingdom; [3]Norsvin SA, Storhamargata 44, 2317 Hamar, Norway; cathrine.brekke@nmbu.no

Abstract

Meiotic recombination leads to shuffling of loci located on the same chromosome. The amount of intrachromosomal shuffling from one generation to the next is affected by the number of crossovers, location of crossovers and crossover interference. In the domestic pig, *Sus Scrofa*, genome wide recombination rates are higher in females than in males. However, in this study we find that the genome wide intrachromosomal shuffling between pairs of loci is higher in males than in females due to difference in distribution of crossovers along the chromosome in the sexes. We show that this pattern is consistent in four of five different pig breeds and that there is a genetic component to the variation in genetic shuffling.

Introduction

An important part of meiosis is the exchange of genetic material between homologues chromosomes through meiotic recombination, which results in gametes with novel haplotypes. This event breaks down linkage disequilibrium and lead to haplotypic diversity that can be exploited in selection. Recombination also has a vital role in the proper alignment and segregation of homologues chromosomes. (Sherman *et al.*, 1991; Koehler *et al.*, 1996; Hassold *et al.*, 1995; Fledel-Alon *et al.*, 2009). Recombination can however also break up beneficial linkage previously built up by selection (Charlesworth and Barton, 1996), and there is evidence for increased mutation rates in recombination hotspots (Halldorsson *et al.*, 2019; Arbeithuber *et al.*, 2015). Recombination rates vary between taxa and species, and even within and between closely related populations (reviewed in Ritz *et al.*, 2017; and Stapley *et al.*, 2017). Most species show some level of heterochiasmy (Burt *et al.*, 1991) and in some species there is also a substantial sexual dimorphism in the distribution of crossovers along the genome (Sakamoto *et al.*, 2000; Lien *et al.*, 2011; Tortereau *et al.*, 2012; Johnston *et al.*, 2016). In breeding this variation is of interest because it affects the production of novel allelic combinations. Veller *et al.* (2019) suggests shuffling of maternal and paternal alleles from one generation to the next as an alternative measure that pics up both the number and distribution of crossovers. The aim of this study was to compare our previous results of variation in recombination rates within and between five domestic pig breeds with measures of intrachromosomal genetic shuffling and investigate whether there is variation between sex and breeds.

Materials & methods

Data. This study focused on five purebred commercial breeding populations with pedigree and genotype data: two sow breeds, Landrace (LR) and Large White (LW); and three boar breeds, Duroc (DU), Synthetic (SY) and Pietrain (PI). The genotype data and filtering are described in detail in Brekke *et al.* (2022).

Linkage mapping and crossover detection. Detailed description of linkage mapping, gamete phasing, crossover detection and fine scale recombination mapping can be found in Brekke *et al.* (2022).

Genetic shuffling within individuals. Genetic shuffling was calculated as the probability that a randomly chosen pair of loci was shuffled during gamete production following the method suggested by Veller *et al.*

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_194, © C. Brekke *et al.* 2022

(2019). We defined the parameter $\bar{r}$ as the probability of two alleles on a chromosome being shuffled due to recombination, i.e. excluding the part due to independent assortment in Equation 4 in Veller *et al.* (2019):

$$\mathbb{E}[\bar{r}] = \sum_{k=1}^{n} 2p_k(1-p_k)\,L_k^2$$

where k is the chromosome number (1-18), p is the proportion of grandpaternal alleles, $1\text{-}p$ is the proportion of grandmaternal alleles and L is the length of the chromosome as a fraction of the total length of the genome. For each phased gamete transmitted from a mother or father to an offspring, shuffling was calculated following Equation 1 and assigned as an observation to the parent, resulting in multiple observations for the phenotype 'genetic shuffling', hereafter referred to as $\bar{r}$, for each parent (hereafter focal individual, or FID).

Genetic variation. We estimated variance components for individual $\bar{r}$ with a repeatability model using the restricted maximum likelihood (REML) method and average information (AI) algorithm in DMU v 6 (Madsen *et al.*, 2014). The following model was used:

$$\bar{r} = id1 + id2 + e$$

where id1 is the random additive genetic effect of the FID, id2 is the random effect of the FID permanent environment (i.e. individual and/or environmental effects affecting all gametes from an FID) and e is the residual effect. The narrow-sense heritability (h^2) was defined as the proportion of phenotypic variance explained by the additive genetic effect and was estimated separately for each breed and sex.

Results

Mean $\bar{r}$ is significantly higher in males than in females in the LR (P=3.12e^{-105}), DU (P=2.51e^{-85}), PI (P=3.70e^{-72}) and SY (P=2.38e^{-53}) breed. In the LW line $\bar{r}$ is only slightly, but significantly (P=1.40e^{-21}) higher in females. Means and distribution are plotted in Figure 1. Intrachromosomal shuffling is a heritable trait in females in all breeds, but only in the LR breed in males. These results are presented in Table 1.

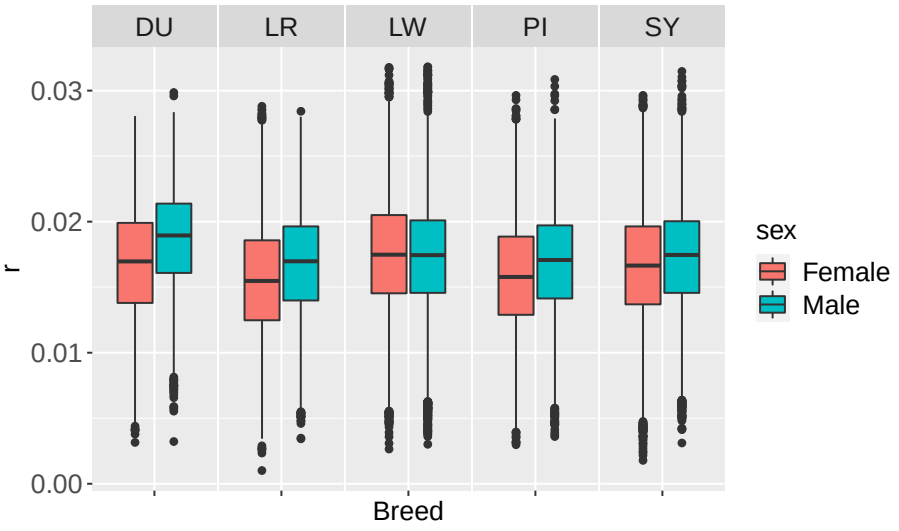


Figure 1. Sex difference in mean and distribution of shuffling.

Table 1. Results from variance component estimation of $\bar{r}$.

Breed	Sex	Mean(SD)	h^2(SE)	No obs[1]	No ind[2]
LR	M	0.0167(0.0039)	0.03(0.01)	11,805	155
LR	F	0.0155(0.0043)	0.13(0.02)	11,805	1,960
DU	M	0.0186(0.0038)	0.03(0.02)	4,090	89
DU	F	0.0168(0.0043)	0.09(0.02)	4,090	661
LW	M	0.0172(0.0039)	0.04(0.01)	41,237	273
LW	F	0.0175(0.0042)	0.15(0.01)	41,237	4,704
PI	M	0.0168(0.0040)	0.03(0.01)	12,159	196
PI	F	0.0158(0.0042)	0.09(0.01)	12,159	1,355
SY	M	0.0172(0.0039)	0.03(0.01)	25,705	224
SY	F	0.0166(0.0043)	0.05(0.01)	25,705	2,635

[1] Total number of observations (gametes).

[2] Total number of unique males or females with repeated observations.

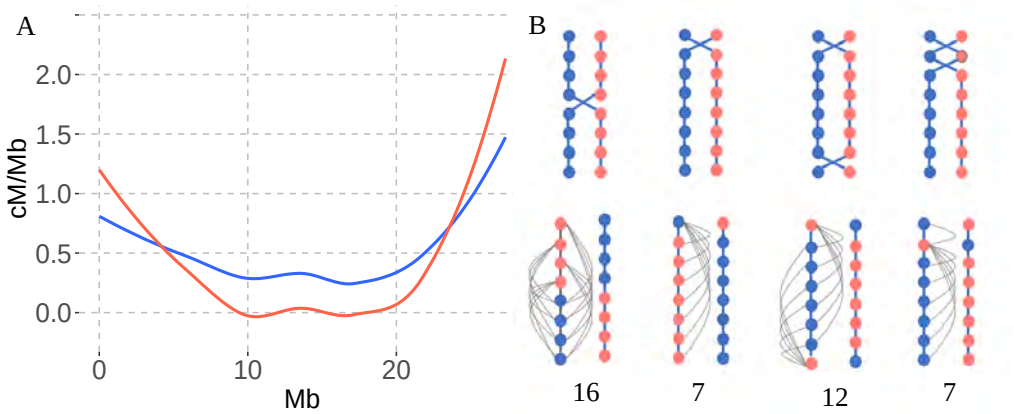


Figure 2. (A) Variation in recombination rates along chromosome 1 in the PI breed for males in blue and females in red plotted with the loess method in the geom_smooth function in ggplot2 (Wickham, 2007). (B) Example of crossover positions and the resulting number of pairs of loci shuffled (modified from Veller *et al.* (2019)).

Discussion

Our results show that the probability of shuffling between two loci on the same chromosomes is low relative to the shuffling of alleles at different chromosomes (average of 0.5 due to Mendelian segregation), but also that the intra-chromosomal shuffling differs between breeds and sexes and is a heritable trait in the domestic pig. Genome wide recombination rates in these pig breeds are higher in females than in males, and rates tend to be elevated in the telomeric regions in both sexes (Brekke *et al.* 2022). In some chromosomes this pattern is more extreme in females, i.e. recombination rates are higher than males in the telomeric regions, but lower than in males closer to the centromere (e.g. as in Figure 2a). This could explain why $\bar{r}$ is lower in females despite higher genome wide recombination rates. Figure 2b illustrates why a central crossover leads to more shuffling and why the position of the crossover may have a higher impact on the probability of shuffling between two loci than the number of crossovers. It is however puzzling that one of the breeds show the opposite sex difference in $\bar{r}$ (Figure 1). This breed (LW) is the breed with the highest genome-wide recombination rate (Brekke *et al.* 2022). More evenly spread crossovers lead to more shuffling (Figure 2b).

Differences in genetic shuffling can thus also be caused by differences in crossover interference between the breeds, explaining part of the difference in $\bar{r}$. Our results show that even if overall levels of sex-differences in recombination is the same in closely related populations, the shuffling might be different, potentially because of rapidly evolving hotspot usage (Paigen and Petkov, 2010; Weng *et al.*, 2014, 2019). It is not clear, however, how the difference between the sex is maintained in a population from generation to generation as each offspring receives a paternal and maternal gamete. Even if a different number of unique sires and dams mated in each generation, the number of maternal and paternal gametes in each generation is always exactly the same. A next step could be to look at differences in recombination and genetic reshuffling between X and Y paternal gametes. The population level shuffling might be more influenced by variation in $\bar{r}$ within the sex, and in males in particular in pigs as the selection pressure is higher. In conclusion this study shows that variation in crossover distribution affects the production of novel haplotypes from one generation to the next and that there is variation in the shuffling caused by recombination between breeds, sex and individuals in the domestic pig.

References

Arbeithuber, B. *et al.* (2015) PNAS, 112(7), pp. 2109–2114. https://doi.org/10.1073/pnas.1416622112.

Burt, A., Bell, G. and Harvey, P.H. (1991), *J. evol. Biol.*, pp. 259-277, https://doi.org/10.1046/j.1420-9101.1991.4020259.x

Brekke *et al.* 2022. Recombination rates vary between breeds, sex and individuals in the domestic pig and is associated with *RNF212*, *SYCP2*, *PRDM7, MEI1* and *MSH4*. [Manuscript submitted for publication]. Norwegian University of Life Sciences.

Charlesworth, B. and Barton, N.H. (1996), Gen. Res., 67(1), pp. 27–41. https://doi.org/10.1017/s0016672300033450.

Fledel-Alon, A. *et al.* (2009) *PLoS Gen.*, 5(9). https://doi.org/10.1371/journal.pgen.1000658.

Halldorsson, B. v. *et al.* (2019), *Science*, 363(6425). https://doi.org/10.1126/science.aau1043.

Hassold, T. *et al.* (1995), *Am. J. Hum. Genet*. 57(4): 867–874. PMCID: PMC1801507

Johnston, S.E. *et al.* (2016), *Gen.*, 203(1), pp. 583–598. https://doi.org/10.1534/genetics.115.185553.

Koehler, K.E. *et al.* (1996) Hum. Mol. Gen. 5.1495-1504. https://doi.org/10.1093/hmg/5.Supplement_1.1495

Lien, S. *et al.* (2011) *BMC Genomics*, 12. https://doi.org/10.1186/1471-2164-12-615.

Madsen, P., Jensen, J. *et al.* (2014). *Proceedings of the 10th world congress of genetics applied to livestock production* (pp. 18-22).

Paigen, K. and Petkov, P. (2010) *Nat. Rev. Gen.*, pp. 221–233. https://doi.org/10.1038/nrg2712.

Team, R. Core.(2013) 'A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing.' *URL* http://www.R-project.org

Ritz, K.R., Noor, M.A.F. and Singh, N.D. (2017) *Trends in Genetics*. Elsevier Ltd, pp. 364–374. https://doi.org/10.1016/j.tig.2017.03.003.

Sakamoto, T. *et al.* (2000) Genetics155:1331–1345, https://doi.org/10.1093/genetics/155.3.1331.

Sherman, S.L. *et al.* (1991) *Am. J. Hum. Genet*. 49(3):608-620, PMCID: PMC1683138.

Stapley, J. *et al.* (2017) *Phil. Trans. Roy. Soc. B.*, 372(1736). https://doi.org/10.1098/rstb.2016.0455.

Tortereau, F. *et al.* (2012) *BMC Genomics*, 13(1). https://doi.org/10.1186/1471-2164-13-586.

Veller, C., Kleckner, N. and Nowak, M.A. (2019) *Proceedings of the National Academy of Sciences*, 116(5), pp. 1659–1668. https://doi.org/10.1073/PNAS.1817482116.

Weng, Z.Q. *et al.* (2014) *Gen. Sel. Evo.*, 46(1). https://doi.org/10.1186/1297-9686-46-34.

Weng, Z. *et al.* (2019) *J. An. Sci. Bio.*, 10(1). https://doi.org/10.1186/s40104-019-0332-y.

Wickham, H., & Wickham, M. H. (2007). The ggplot package.

195. Identification of runs of homozygosity islands and functional analysis for Caqueteño Creole cattle

A.M. Toro-Ospina[1*], J.A.II.V. Silva[2,4], G.P. Schettini[3], V.H. Vallejo[1] and E.G. Ortiz[1]

[1]Universidad de la Amazonia, Sede Porvenir Calle 17 Diagonal 17 con Carrera 3F, 180001, Florencia Caqueta, Colombia; [2]School of Veterinary Medicine and Animal Science, Unesp, 18618-307 Botucatu, SP, Brazil; [3]Virginia Polytechnic Institute and State University 24061, Blacksburg, Virginia, USA; [4]National Council for Scientific and Technological Development, CNPq, Brasília, 71605-001, Brazil; toroospina92@gmail.com

Abstract

The study aimed to identify genomic regions of ROH hotspots/islands and genes of interest in Colombian Creole cattle. The Caqueteño Creole population used belongs to the University of Amazonia, Caquetá, Colombia. The dataset consisted of the genotypes of 127 animals genotyped with the Bovine HD 777,000 SNPs chip. The ROH islands were measured by the PLINK 1.9 program. Since genomic regions with the highest frequency of ROH occurrence are called hotspots. The present study identified ROH hotspots in BTA3, BTA5, BTA6, BTA8, BTA16, and BTA20 chromosomes. The identified genes associated with fertility, reproduction, muscular development, environmental resistance, and immunity traits suggest a natural selection effect on Caqueta creole breed over the years. Notwithstanding, more research needs to be carried out on the Caqueteño Creole breed to identify the specialized production traits of the population.

Introduction

The diversity and genetic adaptation of *Bos taurus* and *Bos indicus* subspecies in the American continent provided changes in their genomes, forming new breeds. Colombia is, in Latin America, one of the countries with more creole breeds adapted to humid tropics performing rusticity and good reproductive features (Bejarano *et al.*, 2012). Originated in the Caquetá department (Colombia), the Caqueteño creole is a breed adapted to tropical conditions. Nowadays, this breed has a low specimen number, and it has been a poorly studied, which becomes it essential to preserve this valued genetic resource for native Colombian breeds. With the advancement of genomics applied to cattle, it was possible to carry out genomic studies with high-density SNP chips among several breeds. Also, this tool made it feasible to perform studies in native breeds without any knowledge *a priori* about the population structure or information at the genome level (McKay *et al.*, 2008). Runs of homozygosity (ROH) have been used extensively in livestock to characterize their population structure and demographic history (Mastrangelo *et al.*, 2018). Furthermore, the ROH analysis can assist in identifying regions associated with economically important traits. Homozygous regions are desirables since they can contribute to breed conservation and recovery. Therefore, the study aimed to identify genomic regions of ROH hotspots/islands and genes of interest in Colombian Creole cattle.

Materials & methods

The Caqueteño Creole population assessed in this study comprises a little herd from three farms in the Caquetá department (Colombia), and a Creole population belonged to the University of Amazonia, Colombia. A total of 127 animals genotyped with the HD Bovine 777,000 SNPs chip (Illumina, San Diego, CA) composed the present study dataset. Through a quality control carried out by the PLINK 1.9 program, the genotypes were filtered with MAF<0.01, a call rate for SNPs <0.90, and a call rate for samples <0.98.

ROH hotspot. The ROH islands were calculated via PLINK 1.9 program (Purcell *et al.*, 2007) using the command *--homozyg-group--*. After, ROH hotspot/island segments with a higher frequency in shared regions were identified in the population. Also, the linkage disequilibrium average was estimated. Genes and orthologs that were settled within the standard limits were pointed out through the *BiomaRt* package in the R-environment.

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_195, © A.M. Toro-Ospina *et al.* 2022

Results

The genomic regions with the highest frequency of ROH occurrence are called hotspots. In this sense, we identified six genomic regions located on the BTA3 (29615176-98903642), BTA5 (62276336 -87155753), BTA6 (38632996-116998306), BTA8 (65897661-105645859), BTA16 (43074568-454221344) and BTA22 (674568-45422124) with one ROH hotspot in each chromosome. In addition, on the BTA 20, we got four ROH hotspots (20114020-63708551, 38347984-40037991, 12658291-51430813, 10769304-63708551) (Figure 1).

Discussion

The genomic regions with the highest frequency of ROH islands identified in the Caqueteño Creole population (Figure 1) suggest a relationship with genes associated with important economic traits. Identified in BTA3, the PHTF1 gene was associated with the fertility index in Holstein cattle (Mesbah-Uddin *et al.*, 2021). The ROH region in the BTA6, the NUP54 gene was associated with body fat deposition, and muscle development (Nolte *et al.*, 2019). According to these genes, it is possible to suggest the Creole Caqueteño has some important features which are supported by their phenotypic performance, precocity, and fertility rates (Celis Parra, Chinguad Taramuel and Garcia Bustos, 2020).

The TUT7 gene located in BTA8 was responsible to regulate the response of the immune cells, meanwhile, the PEX14 and CORT genes located in BTA16 were related to environmental adaptation and stress response (Singh *et al.*, 2020). These traits have valuable importance for cattle breeds to resist environmental changes. Some identified genes are in concordance with Colombian breed traits, such as rusticity and adaptability to tropical regions. The SPEF2 and PRLR putative genes, highlighted in BTA20, were associated with an important effect on cattle coat length, related to SLICK phenotype, as well as being linked to cattle thermal regulation in tropical climates (Huson *et al.*, 2014). Single nucleotide variants have the potential to assist, through genomic selection, to obtain shorter coat animals, which will lead to better thermotolerance and, consequently, higher production levels in hot and humid climates.

The identified genes associated with fertility, reproduction, muscular development, climate resistance, and immunity traits, lead us to suggest that it could be present in the Caqueta creole breed provided by the natural selection over the years. It should be noted that it is the first study that includes genomic information to understand the selection and history of the Caqueteño Creole population. Thus, these results suggest potential genes associated with immunity, adaptation, and fertility traits that could be manifested in the

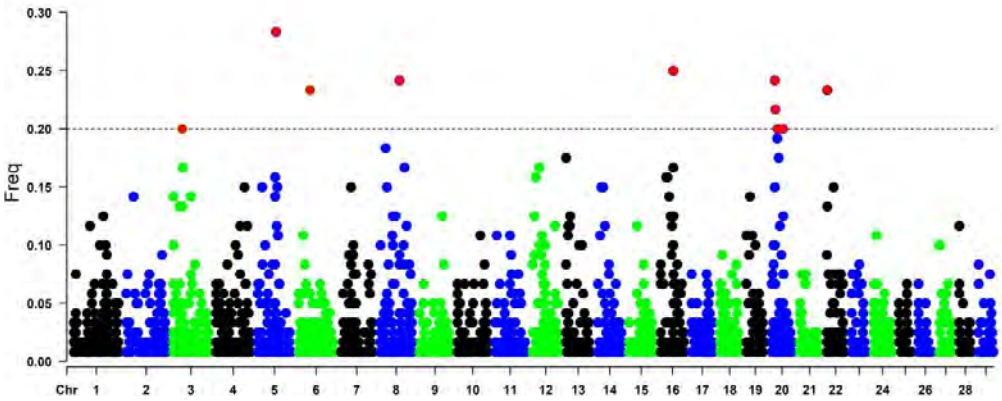


Figure 1. Manhattan plot of the distribution of ROH islands in population Creole Caqueteño cattle.

population. However, more research needs to be carried out on the Creole Caqueteño breed and native breeds in tropical countries, since livestock faces daily challenges, such as those associated with global climate changes that affect animal production.

References

Bejarano G.D., Pedraza L.A., Rocha J.F., Martínez S.R. (2012). Corpoica. pp. 97-107. http://dx.doi.org/10.21930/rcta.vol13_num1_art:246.

Celis Parra, G.A., Chinguad Taramuel, L.G. & Garcia Bustos, J.J. (2020). Trop Anim Health Prod 52, 3241–3250. https://doi.org/10.1007/s11250-020-02350-2.

Huson H.J., Kim E. S., Godfrey R.W., Olson T.A., McClure M.C., *et al.* (2014). Frontiers in genetics, 5, 101. https://doi.org/10.3389/fgene.2014.00101.

Mastrangelo S., Sardina M.T., Tolone M., Di Gerlando R., Sutera A.M, *et al.* (2018). Animal. 12(12):2480-2488. https://doi.org/10.1017/S1751731118000629.

McKay, S.D., Schnabel R.D., Murdoch B.M., Matukumalli L.K., Aerts J. (2008). BMC genetics; 9:37. https://doi.org/10.1186/1471-2156-9-37.

Mesbah-Uddin M., Guldbrandtsen B., Capitan A., Lund M.S., Boichard D., *et al.* (2021). Journal of Dairy Science. https://doi.org/10.3168/jds.2021-20655.

Nolte W., Weikard R., Brunner R.M., Albrecht E., Hammon H.M, *et al.* (2019). Frontiers in genetics, 10, 1130. https://doi.org/10.3389/fgene.2019.01130.

Purcell S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A. R., *et al.* (2007). The American Journal of Human Genetics, 81(3), 559–575. https://doi.org/10.1086/519795.

Singh AK, Upadhyay RC, Chandra G, Kumar S, Malakar D, *et al.* (2020). Cell Stress Chaperones. 25(2):327-344. https://doi.org/10.1007/s12192-020-01076-2.

196. Candidate genes related to signatures of selection for body weight and harvest length in American mink

D.N. Do*, K. Karimi, G. Hu and Y. Miar

Department of Animal Science and Aquaculture, Dalhousie University, Truro, NS B2N 5E3, Canada; duy.do@dal.ca

Abstract

Mink is a major animal used in the fur industry worldwide, and growth is an economically important trait for mink farmers. Mink farmers select animals with better growth traits, such as longer body length (pelt size), but little is known about the genetic mechanisms underlying growth traits in mink. This study examined the signatures of selection for body weight (BW) and body length (BL) in mink using whole-genome sequence data. We identified several genes, such as *PHF3* and *EYS* that may influence variation in BW and BL in overlapping extremes of fixation index and nucleotide diversity results. Enrichment of positional candidate genes indicated that genes involved in lipid metabolism, hormone secretion, and IL-17 signaling pathways might be related to the selection of these traits. This study provides a better understanding of the genetic background of BW and BL variations in mink.

Introduction

Growth is an economically important trait for farm animals, and better knowledge of animal growth is necessary for optimized management and feeding practices. Artificial selection leads to better growth performance in farm animal species. Exploring the genetic mechanisms underlying domestication and phenotypic variation using signatures of selection has been performed in various species using different statistics (reviewed in Saravanan *et al.*, 2020). Mink is a major animal used in the fur industry worldwide. Body weight (BW) and length (BL) are correlated traits and important for mink farmers as they are related to the pelt price. Generally, mink farmers select their animals with better reproductive performance, growth (longer BL), and fur quality. This selection has increased the body length of mink and led to better reproductive performance. However, little is known about the genetic mechanisms underlying the selection of growth traits in mink. The first study for selection signatures in American mink was performed for Aleutian disease by Karimi *et al.* (2021a), which detected 63 genes associated with the mink's response to the disease infection. Identified genetic markers involving selection signature of BL and BW could be used for genetic or genomic selection; therefore, this study examined the potential genes or markers for the signature of selection of the traits in American mink raised in Canada.

Materials & methods

Animals and samplings. The studied animals were raised under standard farming conditions in two farms (the Canadian Center for Fur Animal Research (CCFAR) at Dalhousie Faculty of Agriculture (Truro, NS, Canada) and Millbank Fur Farm (Rockwood, ON, Canada)). Details of 100 animals used for the whole genome sequencing in this study has been given in Karimi *et al.* (2021b). Mink were fed *ad libitum* with the commercial feed and their BW and BL were measured at the harvesting time in December 2018. Tongue samples were collected after humanly euthanization for DNA isolation.

Whole genome sequencing and variant calling. Genomic DNA was extracted from tongue tissues using DNeasy Blood and Tissue Kit (Qiagen, Hilden, Germany) following the manufacture protocol. DNA samples were then sequenced (100 bp pair-end reads) using the BGISEQ-500 platform at Beijing Genomics Institute (BGI, Guangdong, China). After sequencing, SOAPnuke software was used to remove adapters and low-quality reads. Then, high-quality reads were aligned against the last American

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_196, © D.N. Do *et al.* 2022

mink reference genome (https://www.ncbi.nlm.nih.gov/assembly/GCF_020171115.1/) using Burrows-Wheeler Aligner (BWA) 0.7.17 with default options. The aligned files were converted to binary alignment map (BAM) format and sorted using SAMtools 1.11. Potential PCR duplicates were removed using the MarkDuplicates command tool of Picard (http://broadinstitute.github.io/picard/). The BAM files were then indexed by SAMtools software. Variant calling was performed with two pipelines of (1) 'mpileup' module of SAMtools/BCFtools and (2) Haplotype caller using Genome Analysis Toolkit 4.1.9.0 (GATK) pipeline (https://gatk.broadinstitute.org/). To increase the accuracy of SNP calling, the overlapping SNPs between two conducted SNP calling pipelines were extracted. Then the final SNPs obtained by GATK pipeline were filtered using the following criteria: phred-scaled quality score <80.0, quality by depth <2.0, phred scaled *P*>60.0, mapping quality <40.0, strand odds ratio >4.0, mapping quality rank sum test <-12.5, read position rank-sum test <-8.0, minor allele frequency (MAF) >0.05, call rate <0.9, and Hardy-Weinberg Equilibrium with *P*<0.000001. Moreover, only biallelic and autosomal SNPs were kept for further analyses. After quality control, a total number of 9,758,965 SNPs from 100 animals was remained for signatures of selection analyses.

Detection of signature selection. The top 20 mink (10 males and 10 females each) with the highest and lowest values of BW and BL were selected for the signature of selection analyses. Fixation index (F_{ST}) and nucleotide diversity ($\theta\pi$) were used to detect selection signatures for BW and BL. Both $\theta\pi$ and F_{ST} values were calculated using vcftools 0.1.16 in a window size of 100 kb with a step size of 25 kb. To normalize the F_{ST} values, Z-transformation was performed using the scale function in the R program, and genomic windows harboring the top 5% Z-transformed F_{ST} values ($Z(F_{ST})$) were identified. The $\theta\pi$ ratios were computed as $\theta\pi$ high BW or BL divided by $\theta\pi$ low BW or BL. The $\theta\pi$ ratios were then log-transformed. The overlapped windows harboring the top 5% values from both methods were considered as significant windows for the traits. The most recent American mink genome assembly and bedtools 2.30.0 software were used for gene annotation. To identify the biological pathways enriched in the candidate gene lists, gene enrichment analysis was performed using clusterProfiler R package for Gene ontology and Kyoto Encyclopedia of Genes and Genomes (KEGG) terms. Overrepresented annotated genes were identified by Fisher's exact test, considering the *P*<0.05 as the threshold.

Results & discussion

The largest numbers of significant SNPs for both traits, regardless of methods, were located on chromosome 1 (Figure 1a and b). Notably, chromosome 1 was the longest chromosome in the mink genome. A total of 1,058 and 1,691 overlapping windows between the top 5% of F_{ST} and $\theta\pi$ was detected for BW and BL, respectively (Figure 1c). A similar ratio (around 20%) of overlapping windows between F_{ST} and $\theta\pi$ was also reported by Salek *et al.* (2020) when they analyzed the signature of selection for pony and light horse types. A total of 411 windows were involved in the selection for both BW and BL (Figure 1c), which might be due to the fact that the two traits had a high genetic correlation of 0.78 (Do *et al.*, 2021); therefore, sharing common genetic mechanisms.

The $Z(F_{ST})$ of these overlapping windows ranked from 1.92 to 10.24 for BW and 1.79 to 17.36 for BL, respectively. The windows with $Z(F_{ST})$ >7 with the positional candidate genes for BW were listed in Table 1. A novel gene *LOC122901584* was the only gene located in the window with the highest $Z(F_{ST})$ value at 172.73 MB on chromosome 2. Some other positional candidate genes for BW included *PHF3, EYS, TRNAK-CUU, TFAP2D,* and *TFAP2B* on Chromosome 1, *CCDC7* on Chromosome 12, or *USP3, FBXL22,* and *HERC1* on chromosome 13. For BL, the most significant region was centered at 105.88 Mb on chromosome 1, where *PHF3, EYS,* and *TRNAK-CUU* were considered as the positional candidate genes. Multiple windows with $Z(F_{ST})$ >7 and candidate genes were found around the regions of 105.88 Mb on chromosome 1. Another important window region for BL was at 74.88 Mb on chromosome 10 since several windows with $Z(F_{ST})$ >7

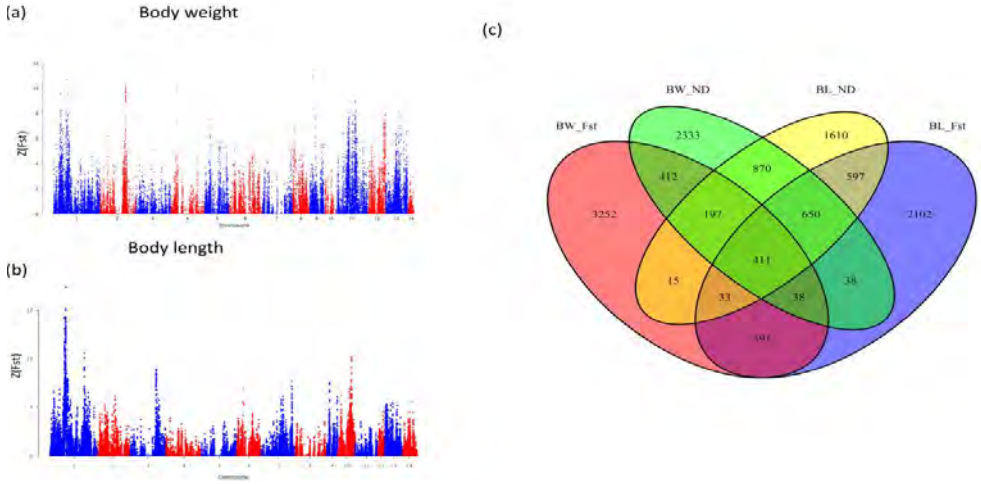


Figure 1. Manhattan plot of $Z(F_{ST})$ values for (a) BW and (b) BL, and (c) Venn diagram of overlapping windows. BW_Fst, BL_Fst, BW_ND and BL_ND were the top windows for BW and BL based on the $Z(F_{ST})$ and θπ ratios, respectively.

Table 1. The top windows based on $Z(F_{ST})$ for body weight.

Chromosome	Start (Mb)	End (Mb)	$Z(F_{ST})$	Log(θπ ratios)	Genes
2	172.73	172.83	10.24	1.71	*LOC122901584*
1	105.88	105.98	7.99	1.16	*PHF3, EYS,* and *TRNAK-CUU*
1	96.93	97.03	7.90	0.46	*TFAP2D* and *TFAP2B*
12	116.93	117.03	7.90	0.51	*CCDC7*
1	46.75	46.85	7.83	0.40	*LOC122905089*
13	105.25	105.35	7.80	1.38	*USP3, FBXL22,* and *HERC1*
2	170.35	170.45	7.67	4.31	*LOC122899239*
1	89.6	89.7	7.44	0.80	*PTK7, SRF,* and *CUL9*
12	107.3	107.4	7.15	0.39	*LOC122892061*
12	117.08	117.18	7.12	0.47	*LOC122891753*
1	89.55	89.65	7.02	0.88	*KLHDC3, CUL7, MRPL2, KLC4,* and *DNPH1*

were located in this region. *RASSF5, EIF2D, IKBKE, LAX1,* and *ZC3H11A* were potential candidate genes for BL in this region.

The functional enrichment analyses showed ubiquitin-like protein transferase activity, motor activity, and epidermal growth factor receptor binding were the three most significant GO terms enriched for BW. The protein phosphatase regulator activity, Rac GTPase binding, and rRNA methyltransferase activity were the three most significant GO terms enriched for BL. A total of 27 and 17 KEGGs pathways were significantly ($P<0.05$) enriched for candidate genes of BW and BL, respectively. Sphingolipid and IL-17 signaling pathways were among the two of the three most significant KEGG pathways enriched for both BW and BL. Some other significant enriched pathways included Parathyroid hormone synthesis, secretion and action, Wnt signaling pathway or Cytokine-cytokine receptor interaction for BW and ABC transporters, Thyroid hormone signaling pathway, and Glycerophospholipid metabolism for BL, respectively.

Although more than hundreds of genes were annotated as potential candidate genes for these traits, many genes such as *LOC122901584*, *LOC122905089,* and *LOC122899239* (Table 1) were novel genes with no known functions in mink. The functions of these genes were inferred based on the information from human studies. *PHF3* and *EYS* might involve the selection signatures for both BW and BL as they were located in the window associated with both traits. *PHF3* might function as a transcription factor, while *EYS* was known for its roles in retinal degeneration. There are no functional studies of these genes in mink so far. The functional enrichment indicated that BW candidate genes are involved in growth factor binding, which is one of the important functions of growth development. The involvement of BW candidate genes was also supported by the results of KEGG pathways enrichment. The enriched pathways, including Sphingolipid signaling pathway or Parathyroid hormone synthesis, secretion, and action pathway, were involved in lipid metabolism or hormone secretion – the important processes for regulating body weight. Interestingly, the IL-17 signaling pathway was enriched for both candidate genes of BW and BL. The IL-17 genes family is primarily involved in the host defense mechanisms against infection of bacteria, fungi, and helminth (Chung *et al.*, 2021); therefore, the enrichment results of the IL-17 signaling pathway might reflect the selection of higher BW or longer BL might be related to the health status of animals.

Acknowledgements

The authors gratefully acknowledge the financial support from the Natural Sciences and Engineering Research Council (NSERC) of Canada, Canada Mink Breeders Association, Nova Scotia Mink Breeders Association, and Mink Veterinary Consulting Research and Services. We also thank the CCFAR and Millbank Fur Farm staff for collecting and providing the data.

References

Chung S.-H., Ye X.-Q. and Iwakura Y. (2021) International immunology 33(12): 723-729.

Do D.N., Hu G., Salek Ardestani S. and Miar Y. (2021) Journal of Animal Science 99(3): 1-7.

Karimi K., Farid A.H., Myles S. and Miar Y. (2021a) Scientific reports 11(1): 1-13.

Karimi K., Ngoc Do D., Sargolzaei M. and Miar Y. (2021b) Genes 12(2): 258.

Salek Ardestani S., Aminafshar M., Zandi Baghche Maryam M.B., Banabazi M.H., Sargolzaei M. *et al.* (2020) Genome 63(8): 387-396.

Saravanan K., Panigrahi M., Kumar H., Bhushan B., Dutt T. *et al.* (2020) Livestock Science 241(104257).

197. Maintenance of genetic diversity in subdivided populations using different genomic coancestry matrices

E. Morales-González[1*], B. Villanueva[1], M.Á. Toro[2] and J. Fernández[1]

[1]INIA-CSIC, Ctra. de La Coruña, km 7.5, 28040 Madrid, Spain; [2]UPM, Ciudad Universitaria s/n, 28040 Madrid, Spain; morales.elisabeth@inia.es

Abstract

The main objective in conservation programs of subdivided populations is to maximize the global genetic diversity while restricting the levels of inbreeding within subpopulations. This can be achieved using the Optimal Contribution method that originally relied on pedigree-based coancestries. However, marker-based coancestries could lead to a more efficient management although different marker-based measures can have different effects on the diversity maintained. In this simulation study, we evaluated the potential benefit of using different marker-based genomic coancestry matrices in terms of global genetic diversity and its distribution within- and between-subpopulations. Our results show that using the genomic matrix describing the excess in the observed number of alleles shared by two individuals relative to the expected number under Hardy–Weinberg equilibrium could be the best option for managing subdivided populations as it maintains more global diversity and leads to less inbreeding within subpopulations.

Introduction

Most populations of endangered species or breeds within species are subdivided into different separated breeding groups. This fragmentation implies a reduced census in each subpopulation that leads to strong genetic drift and, thus, to fast increases in inbreeding and decreases in local genetic variability.

The consensus method to maintain genetic variability is the Optimal Contribution (OC) method that was extended by Fernández *et al.* (2008) to optimally manage subdivided populations. Their method determines the optimal number of offspring (contribution) of each individual to the next generation to maximize the global genetic diversity while balancing the relative levels of coancestry between and within subpopulations (and therefore the levels of inbreeding within subpopulations) and optimizes migration between subpopulations. As originally proposed, the method of Fernández *et al.* (2008) used pedigree-based coancestries. However, in undivided populations, it has been shown that the use of marker-based genomic coancestries leads to a more efficient maintenance of genetic diversity (e.g. Gómez-Romano *et al.* 2013) although different genomic coancestry matrices may lead to different levels of genetic variability (Meuwissen *et al.* 2020; Morales-González *et al.* 2021).

The objective of this study was to evaluate, via stochastic simulations, the global levels of genetic diversity maintained, its distribution within- and between-subpopulations and the migration flow between subpopulations when using different genomic coancestry matrices in the OC method for managing subdivided populations.

Materials & methods

The genome was composed of 20 chromosomes of 1 Morgan each. Two types of biallelic loci (500,000 SNPs and 500,000 'unobserved loci' per chromosome) were simulated. SNP loci were used for computing the genomic coancestry matrices involved in the management and the unobserved loci were used for calculating the different parameters evaluated. The conservation program started from a base population (t=0) made up of 100 individuals, which was created in three steps. Firstly, a population at mutation-drift equilibrium was generated by simulating 10,000 discrete generations of random mating for a population

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_197, © E. Morales-González *et al.* 2022

of n=100 as described in Morales-González *et al.* (2021). Secondly, the population was expanded in order to have enough individuals for sampling 100 replicates (n=25,600). Then, five panmictic generations were performed with a small group of individuals (10 ♂ and 10 ♀) isolated from the rest of the population (40 ♂ and 40 ♀). Afterwards, the large group was randomly divided into four groups of 10 ♂ and 10 ♀ each. Thus, at the beginning of the management (t=0), one of the subpopulations was genetically differentiated and more inbred than the others. At t=0, the average number of SNPs and unobserved loci still segregating was 48,519 and 48,376, respectively. Management was performed during 10 discrete generations following the OC method of Fernández *et al.* (2008) that maximizes the global genetic diversity but balancing the relative importance of within-subpopulation coancestry by a factor λ. Values of λ tested were 1 and 5. A restriction on the maximum number of migrants between subpopulations (M=5) was also included. The method was implemented using the pedigree-based coancestry matrix (Θ_P) and two matrices based on genomic information (i.e. on SNPs segregating at t=0): the matrix describing the excess in the observed number of alleles shared by two individuals relative to the expected number under Hardy-Weinberg equilibrium (Li and Horvitz 1953), Θ_L, and the genomic relationship matrix obtained using method 2 described in VanRaden (2008), Θ_V. Base population frequencies were used to compute both matrices. Mating was at random within subpopulations.

Management scenarios were compared in terms of the genetic variability retained in the global population and its distribution between and within subpopulations. The amount of genetic variability retained in the global population was measured as the expected heterozygosity (H_T). The genetic variability within subpopulations was measured as the average expected heterozygosity across subpopulations (H_S) and the genetic variability between subpopulations was calculated as $D = H_T - H_S$ (also referred to as the average distance between subpopulations). Given that matings within subpopulations were random, the average molecular inbreeding for subpopulation i was $F_{Si} = 1 - H_{Si}$, where H_{Si} is the expected heterozygosity for that subpopulation. Additionally, scenarios were compared in terms of the particular flow of offspring from and to subpopulation 1 (i.e. the number of migrants involving this particular subpopulation). Results presented for all parameters are averages across the 100 replicates.

Results

Table 1 shows a decrease in H_T across the 10 generations when the management was based on Θ_P or Θ_V for both values of λ, while H_T increased when using Θ_L for λ=1. This increase in H_T with Θ_L was accompanied by a higher reduction in H_S and a higher increase in D. However, for λ=5 (i.e. giving a higher weight to within-subpopulations variability), management using Θ_L was able to keep higher levels of diversity both in the global population (H_T) and within subpopulations (H_S) than when using Θ_P or Θ_V.

As expected, inbreeding was practically the same for subpopulations 2 through 5 and it was lower for λ=5 than for λ=1 (Table 2). For λ=1, the lowest levels of F_{S1} and $F_{S2\text{-}5}$ were obtained when using Θ_P and the highest levels were obtained when using Θ_L (the matrix giving the lowest H_S).

For λ=1, on average, each subpopulation sent 1 individual and received 1 individual from other subpopulation across generations (results not shown). However, for λ=5 (Figure 1), whatever the matrix used in OC (Θ_P, Θ_L or Θ_V), subpopulation 1 sent ~1 migrant and received 3-4 migrants at t=0 and, as the generations go by, the number of migrants received decreased (1 migrant after 4-5 generations). This is a reflection of λ=5 implying a higher pressure to increase the originally lower levels of diversity in subpopulation 1 compared to other subpopulations. In fact, due to the differential migration between subpopulations, Fs_1 decreased from t=0 to t=10 until it reached a value close to $Fs_{2\text{-}5}$ (Table 2).

Table 1. Expected heterozygosity in the global population (H_T), average expected heterozygosity in subpopulations (H_S) and distance between subpopulations (D) across generations (t) when contributions were optimized using pedigree-based (Θ_P), Li and Horvitz (Θ_L) and VanRaden (Θ_V) coancestry matrices.[1]

	t	Θ_P			Θ_L			Θ_V		
		H_T	H_S	D	H_T	H_S	D	H_T	HS	D
λ=1	0	0.188	0.180	0.008	0.188	0.180	0.008	0.188	0.180	0.008
	5	0.186	0.173	0.013	0.190	0.167	0.024	0.186	0.172	0.013
	10	0.184	0.169	0.015	0.191	0.162	0.030	0.183	0.167	0.016
λ=5	0	0.188	0.180	0.008	0.188	0.180	0.008	0.188	0.180	0.008
	5	0.186	0.176	0.010	0.189	0.178	0.011	0.185	0.176	0.010
	10	0.183	0.172	0.011	0.189	0.177	0.012	0.183	0.172	0.010

[1] Standard errors for H_T and H_S ranged from 6.58×10^{-5} to 4.87×10^{-4}.

Table 2. Average inbreeding in subpopulation 1 (F_{S1}) and in subpopulations 2 to 5 ($F_{S2\text{-}5}$) across generations (t) when contributions were optimized using pedigree-based (Θ_P), Li and Horvitz (Θ_L) and VanRaden (Θ_V) coancestry matrices.[1]

	t	Θ_P		Θ_L		Θ_V	
		F_{S1}	$F_{S2\text{-}5}$	F_{S1}	$F_{S2\text{-}5}$	F_{S1}	$F_{S2\text{-}5}$
λ=1	0	0.834	0.817	0.834	0.817	0.834	0.817
	5	0.834	0.826	0.841	0.833	0.835	0.826
	10	0.835	0.831	0.843	0.839	0.837	0.832
λ=5	0	0.834	0.817	0.834	0.817	0.834	0.817
	5	0.825	0.824	0.824	0.822	0.826	0.824
	10	0.828	0.828	0.824	0.823	0.828	0.828

[1] Standard errors ranged from 1.45×10^{-4} to 5.16×10^{-4}.

Discussion

The highest global genetic diversity maintained was obtained when using Θ_L. This result was also observed when dealing with undivided populations (Meuwissen *et al.* 2020; Morales-González *et al.* 2021). However, when giving the same weight to between- and within-subpopulation coancestry (λ=1), the higher global diversity obtained with Θ_L was at the cost of an increase in the genetic distance between subpopulations (D) and an increase in local inbreeding (lower H_S). This agrees with theory that predicts that the maximum global genetic diversity is achieved when the population is composed by many isolated inbred subpopulations (i.e. low local genetic diversity). However, when giving more weight to the within-subpopulation coancestry (λ=5), the use of Θ_L in OC not only led to higher global genetic diversity, but also to lower local inbreeding than the use of Θ_P or Θ_V. As expected, for all matrices lower levels of inbreeding were achieved for λ=5 than for λ=1, mainly due to the increase of migrants sent to subpopulation 1. This shows the ability of the methodology to detect the differences in genetic diversity between subpopulations and correct them by optimising the contributions and the migration flow. Allowing a higher number of migrants would probably lead to low inbreeding values in fewer generations. However, this is not a realistic scenario due to the cost and risk of moving animals between subpopulations.

In conclusion, it seems that Θ_L could be the best option for managing subdivided populations as it maintains more global diversity and leads to less inbreeding.

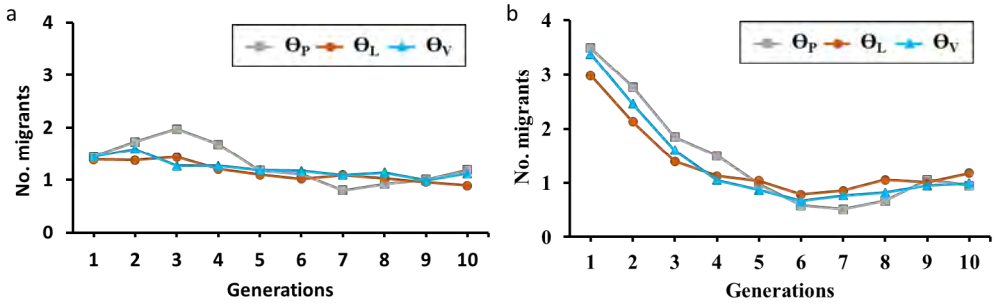


Figure 1. Number of individuals sent (a) and received (b) by subpopulation 1 across generations when contributions were optimized using pedigree-based (Θ_P), Li and Horvitz (Θ_L) and VanRaden (Θ_V) coancestry matrices for λ=5.

Acknowledgements

This research was funded by MCIN/AEI /10.13039/501100011033 (Project PID2020-114426GB-C2).

References

Fernández, J., Toro, M.A., and Caballero, A. (2008) Genetics 179(1):683-692. https://doi.org/10.1534/genetics.107.083816

Gómez-Romano, F., Villanueva, B., de Cara, M.A.R., and Fernández, J. (2013) Genet. Sel. Evol. 45:38. https://doi.org/10.1186/1297-9686-45-38

Li, C.C., and Horvitz, D.G. (1953) Am. J. Hum. Genet. 5(2):107–117.

Meuwissen, T.H.E., Sonesson, A.K., Gebregiwergis, G., and Woolliams, J.A. (2020) Front. Genet. 11:880. https://doi.org/10.3389/fgene.2020.00880

Morales-González, E., Fernández, J., Pong-Wong, R., Toro, M.Á., and Villanueva, B. (2021) Genes 12(5):673. https://doi.org/10.3390/genes12050673

VanRaden, P. (2008) J. Dairy Sci. 91(11):4414–4423. https://doi.org/10.3168/jds.2007-0980

198. Long-term changes in genetic mean and genic variance and underlying allele frequencies in breeding programmes

G.A. de Oliveira Junior[1], M. Spehar[2], F.S. Schenkel[1], C.F. Baes[1] and G. Gorjanc[2*]

[1]Centre for Genetic Improvement of Livestock, Department of Animal Biosciences, University of Guelph, Guelph, ON, Canada; [2]The Roslin Institute, University of Edinburgh, Edinburgh, United Kingdom; gregor.gorjanc@roslin.ed.ac.uk

Abstract

The infinitesimal model assumes that polygenic traits are influenced by (infinitely) many quantitative trait loci with (infinitely) small effects and that selection on such traits has minimal impact on the frequency of alleles at individual loci, hence minimal impact on genetic variance. Here we evaluated trends in genetic mean and genic variance for milk yield over 25 years in 9,403 proven and genotyped bulls from the Canadian Dairy Network. The results showed significant change in genetic mean and minor change in genic variance, supporting the infinitesimal model. A detailed analysis of SNP marker allele frequencies indicated sizeable changes at individual loci. We grouped the SNP markers by their initial allele frequency and direction of change, to indicate how these groups of makers contribute to changes in mean and variance in the analysed period of 25 years and predictions for the next 75 years showing sizeable changes in mean and variance.

Introduction

Animal breeding programmes focus on additive genetic values that represent the sum of allele substitution effects over all quantitative trait loci for the trait of interest. Much of the theory and estimation of additive genetic values is based on the infinitesimal model. The infinitesimal model assumes that polygenic traits are influenced by (infinitely) many quantitative trait loci with (infinitely) small effects (Fisher, 1918). Consequently, this model implies that selection on polygenic traits has (infinitely) small impact on allele frequency changes at individual loci (Fisher, 1918). Despite the (infinitely) small impact on allele frequency changes, selection can have a significant cumulative change across all loci, as manifested by breeding programmes.

Since the infinitesimal model predicts (infinitely) small allele frequency changes under selection, this implies (infinitely) small short-term changes in genetic variance. This prediction is in line with observations that response to selection in most breeding programmes are not levelling off (e.g. Hill, 2016). There are reports that recent introduction of genomic selection has sped up the rate of change in genetic mean and variance (Jannink, 2010; Hidalgo *et al.*, 2020; Wientjes *et al.*, 2021), which could also be due to the Bulmer effect, negative linkage-disequilibrium component of genetic variance (Bulmer, 1971; Lara *et al.*, 2021).

Abundant genome-wide marker data is enabling rapid genetic improvement via genomic selection, but also detailed dissection of underlying genome changes that drive the rapid genetic improvement. Compared to phenotypic selection, genomic selection is increasing the risk of losing favourable alleles (Jannink, 2010; Wientjes *et al.*, 2021), particularly for the rare alleles or those alleles that are in unfavourable linkage-disequilibrium with genomic selection markers. A systematic evaluation of the change in past and predicted future allele frequencies and their relation to genetic mean and variance is lacking and the aim of this contribution.

Materials & methods

Data. We used data from the routine milk yield genetic evaluation of the Canadian Dairy Network – Lactanet (Guelph, ON). The data comprised 9,403 proven bulls, born between 1989 and 2014. We retrieved

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_198, © G.A. de Oliveira Junior *et al.* 2022

de-regressed estimated breeding values and associated reliabilities for the bulls. We also retrieved their single nucleotide polymorphism (SNP) marker genotypes. We retained 40,448 SNP markers from the autosomes having minor allele frequency above 0.05.

Estimation of marker effects. We estimated allele substitution effects of the SNP markers using the Bayesian ridge-regression model $y = Xb + M\alpha + e$, where y is a vector of de-regressed estimated breeding values for milk yield scaled to mean zero and unit variance, b is a vector of fixed effects (the intercept), α is a vector of allele substitution effects, e is a vector of residuals, X is an incidence matrix linking y with b, and M is a matrix of SNP marker genotypes. The model assumptions were $\boldsymbol{\alpha} \sim N(\mathbf{0}, \boldsymbol{I}\sigma_\alpha^2)$ and $\boldsymbol{e} \sim N(\mathbf{0}, \boldsymbol{W}\sigma_e^2)$, where W is a diagonal weight matrix with values equal to (1-Reliability$_i$)/Reliability$_i$ (for the i-th bull) adjusting heterogeneous residual variance in the de-regressed estimated breeding values.

We inferred the model parameters using the Gibbs sampler implemented in the JWAS package with default prior specifications (Cheng *et al.*, 2018). We obtained 100,000 samples from posterior distribution of allele substitution effects and saved every 100th sample. We also obtained posterior samples of breeding values (a) for the i-th sample using $a^i = M\alpha^i$.

We further obtained samples from posterior distribution of the mean of breeding values $\mu_{a,k}^i = E(\boldsymbol{a}_k^i)$ and variance of breeding values $\sigma_{a,k}^{2,i} = Var(\boldsymbol{a}_k^i)$ for each (k-th) year of birth. Summarising the $\mu_{a,k}^i$ and $\sigma_{a,k}^{2,i}$ samples and plotting the summaries against the year gives time trends in genetic mean and variance. Lara *et al.* (2021) showed that changes in genetic variance are driven by changes in genic variance (function of allele frequencies and allele substitution effects at the causal loci; $\sigma_{a,l}^2 = 2p_l(1-p_l)\alpha_l^2$ for the l-th locus and $\sum_{l=1}^{n_l} \sigma_{a,l}^2$ for all n_l loci) and in linkage-disequilibrium covariance (function of correlation between causal locus genotypes and their allele substitution effects). Here we focus only on genic variance. Preliminary analysis showed minor changes in genic variance in the period between 1989 and 2014, despite significant changes in genetic mean, corroborating the infinitesimal model. To further analyse these results, we analysed changes in allele frequencies of the SNP markers.

Estimation of allele frequency time trends. We calculated allele frequency of the SNP markers for each year of birth (1989 to 2014) and estimated their rate of change with beta regression (*betareg* R package; Cribari-Neto and Zeileis, 2010), independently for each SNP marker. We have then predicted change in allele frequency from the fitted beta regression model for each SNP marker for 100 years $\hat{p}_{l,k}$ for the k-th year), including the observed period (1989 to 2014).

To further analyse change in genetic mean and no apparent change in genic variance we have grouped the SNP markers according to the initial allele frequency (in year 1989) (above or below 0.5) and change in allele frequency (increasing or decreasing). The rational for this grouping is that following the $2p_l$ (1 – p_l) expression we expect that 'InitialAFAbove0.5IncreasingAF' markers will show decrease in genic variance, 'InitialAFBelow0.5IncreasingAF' markers will show increase in genic variance until frequency of 0.5, 'InitialAFAbove0.5DecreasingAF' markers will show increase in genic variance until frequency of 0.5, and 'InitialAFBelow0.5DecreasingAF' markers will show decrease in genic variance. For each of the groups we plotted predicted allele frequencies and showed their contribution to predicted genetic mean ($\hat{\mu}_{a,k} = \sum_{l=1}^{n_l} 2\hat{p}_{l,k}\hat{\alpha}_l$) and genic variance ($\left(\sum_{l=1}^{n_l} \hat{\sigma}_{a,l}^2\right)_k = \sum_{l=1}^{n_l} 2\hat{p}_{l,k}(1-\hat{p}_{l,k})\hat{\alpha}_l^2$), with $\hat{\alpha}_l$ being mean of the 1000 samples α_l^i.

Results

Figure 1 shows the predicted changes in allele frequency for the SNP markers over 100 years. The change of allele frequencies in each of the four groups of course match the grouping based on the estimated rate of

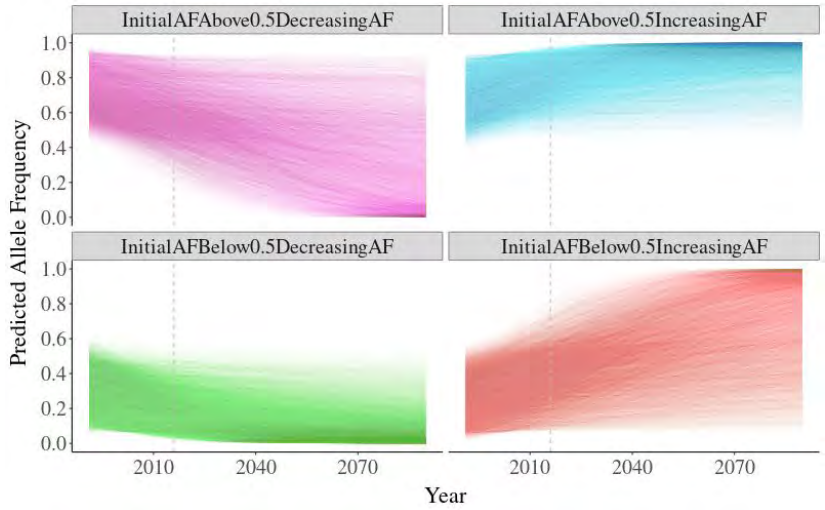


Figure 1. Predicted allele frequencies over 100 years for the four groups of SNP markers.

change, but there is considerable variation within each group. Some SNP markers are predicted to fixate, but many allele frequency changes are small.

Figure 2 shows the predicted genetic mean over 100 years and contribution of the four groups of SNP markers to the genetic mean. Genetic mean increased over the initial period and is predicted to increase in the future, above 6 genetic standard deviations of genetic gain. The largest contribution is from the 'InitialAFAbove0.5IncreasingAF' group, followed by the 'InitialAFBelow0.5IncreasingAF' group, with a considerably smaller contribution of the other two groups of SNP markers.

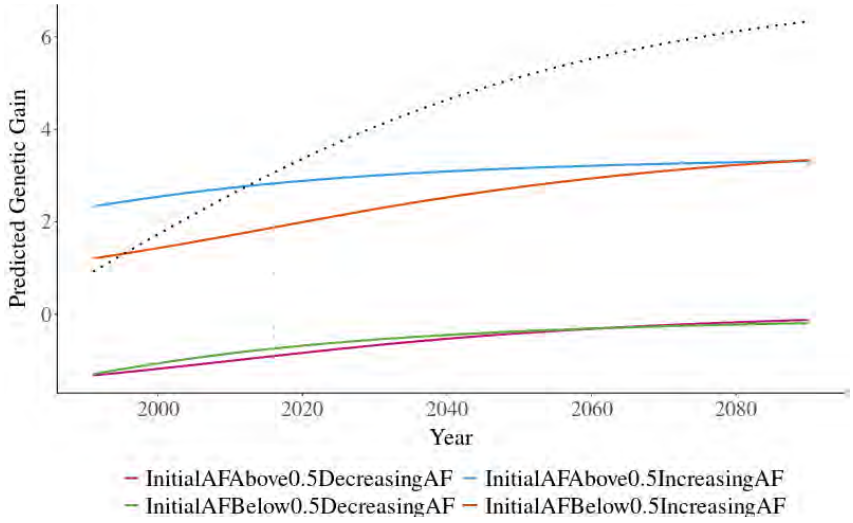


Figure 2. Predicted genetic mean over 100 years and contribution of the four groups of SNP markers to the genetic mean.

Figure 3 shows the predicted genic variance over 100 years and contribution of the four groups of SNP markers to the genic variance. Genic variance was almost constant during the initial period but is predicted to decrease for about a third in future. Contribution of the four groups varied over time, with relatively similar contributions in the initial period and the largest contribution of the 'InitialAFBelow0.5IncreasingAF' and 'InitialAFAbove0.5DecreasingAF' groups for most of the period, though these contributions decreased after some point. This result is in line with the expectation according to the $2p_i(1 - p_i)$ expression.

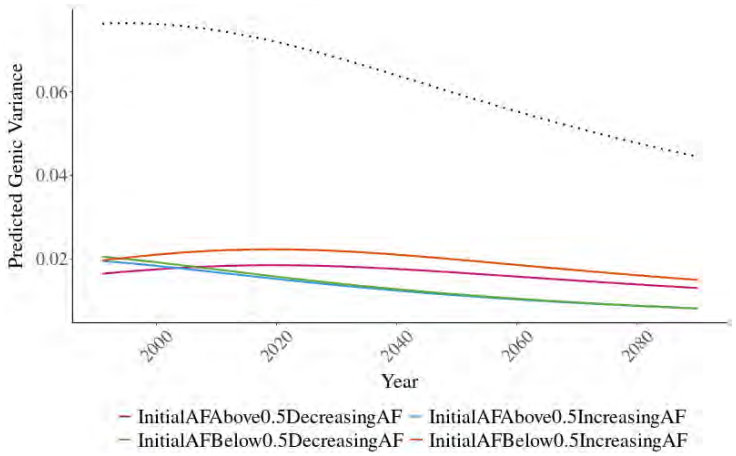


Figure 3. Predicted genic variance over 100 years and contribution of the four groups of SNP markers to the genic variance.

Discussion

We analysed change in genetic mean and variance over 25 years as well as underlying changes in SNP marker allele frequencies and predicted the trends in the same quantities for further 75 years. Results show that allele frequencies of the SNP markers changed sizeably during the analysed period of 25 years, and are predicted to change even more in the future, though arguably over a long period. The results suggests that there is considerable dynamics at individual loci that is not seen at the cumulative level. The results also give insights into the assumptions of the infinitesimal model and its predictions for changes in genic variance. The results also clearly showed different contribution of the four groups of SNP markers, both to genetic mean and genic variance. All these predictions assume that estimated rate of allele frequency changes in the initial period will hold in the future, that allele substitution effects will not change, hence that there are no non-additive genetic effects and that we have stable linkage-disequilibrium between the SNP markers and causal loci, and that there will be no mutations. These are clearly strong assumptions. The presented analysis nevertheless enables in-depth study of genetic changes in breeding populations.

References

Bulmer (1971) Am Nat 105:201–211. https://doi.org/10.1086/282718

Cheng *et al.* (2018) 11th WCGALP, Auckland, New Zealand.

Cribari-Neto and Zeileis (2010) J Stat Soft 34:1-24. https://doi.org/10.18637/jss.v034.i02

Fisher (1918) Proc Roy Soc 52:399–433. https://doi.org/10.1017/S0080456800012163

Hidalgo et. al. (2020) J Anim Sci 98:1–2. https://doi.org/10.1093/jas/skaa032

Hill (2016) Genetics 202:877–881. https://doi.org/10.1093/genetics/202.3.NP

Jannink (2010) Genet Sel Evol 42:35. https://doi.org/10.1186/1297-9686-42-35

Lara *et al.* (2021) Heredity 128:21–32. https://doi.org/10.1038/s41437-021-00485-y

Wientjes *et al.* (2021) bioRxiv 2021.03.16.435664. https://doi.org/10.1101/2021.03.16.435664

199. The long-term effects of genomic selection: allele frequency changes and fixation of loci

Y.C.J. Wientjes[1*], P. Bijma[1], J. van den Heuvel[1], B.J. Zwaan[1], Z.G. Vitezica[2] and M.P.L. Calus[1]

[1]Wageningen University & Research, P.O. Box 338, 6700 AH Wageningen, the Netherlands; [2]INRAE, UMR 1388 GenPhySE, 31326 Castanet-Tolosan, France; yvonne.wientjes@wur.nl

Abstract

Allele frequencies change due to drift and selection. We investigated those changes for different selection methods. We simulated a livestock population with 50 generations of selection for an additive trait controlled by 2,000 segregating loci at the start of selection. Our results show that the average absolute change in allele frequency with genomic selection was slightly higher than with pedigree selection, and considerably higher than with mass selection. Genomic and pedigree selection both resulted in fixation of roughly three times more loci than mass selection, and roughly five times more loci became fixed for the unfavourable allele. While this was mainly a result of increased genetic drift, genomic selection also lost additional favourable alleles due to hitchhiking. This suggests that genomic selection can limit long-term genetic gain.

Introduction

Despite the sometimes strong selection pressure in animal breeding programs, genetic variation and rates of genetic gain have been stable for many generations in several animal populations (Beniwal *et al.*, 1992; Havenstein *et al.*, 2003). Therefore, these pedigree-based selection methods have proven to be sustainable so far.

Implementation of genomic selection has increased the rates of genetic gain in the short term. Moreover, it has likely accelerated the changes in allele frequencies across generations in certain regions on the genome (Heidaritabar *et al.*, 2014; Doekes *et al.*, 2018) and has increased the risk of losing rare favourable alleles (Jannink, 2010; Liu *et al.*, 2014). However, the long-term consequences of genomic selection are still largely unknown.

The current availability of genomic information has also opened up the possibility to investigate the changes in allele frequency as a result of selection, which is essential for understanding the long-term effects of selection. The vast majority of causal loci are, however, unknown. Therefore, our aim is to investigate and quantify the allele frequency changes at causal loci and how this is affected by selection methods using simulations.

Materials & methods

We simulated a livestock population under 50 generations of selection. First, we simulated a historical population in QMSim software (Sargolzaei and Schenkel, 2009) that showed a linkage disequilibrium pattern and allele frequency distribution comparable to livestock populations. Then we used our own developed Fortran program to generate the next 50 generations of selection. In every generation, the best 100 females and 100 males were selected and randomly mated using a mating ratio of 1:1 and a litter size of 10 (5 females and 5 males). Each scenario was replicated 20 times. For more details, see Wientjes *et al.* (2021).

Genome and phenotypes. The simulated genome had 10 chromosomes of 100 cM each. At the end of the historical population, 20,000 segregating loci with a uniform allele frequency distribution were selected as marker, and 2,000 segregating loci with a U-shaped allele frequency distribution were selected as causal

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_199, © Y.C.J. Wientjes *et al.* 2022

loci with an effect on the simulated trait. Moreover, 4,000 non-segregating loci were randomly selected as locations for causal mutations, which started to affect the trait when the locus was segregating after a mutation took place. The number of new mutations for each individual was sampled from a Poisson distribution, with an average of 0.6, which resulted in a mutational variance of ~$0.001\sigma_e^2$ per generation.

Phenotypes were simulated based on an additive model. For all causal loci and mutations, an additive effect was sampled from $N(0,1)$. The additive effects and genotypic information of causal loci were used to calculate the true breeding value for each individual. Finally, a non-genetic residual was added to the true breeding value, to obtain a heritability of 0.4.

Selection process. We compared five methods to select the parents of the next generation:

1. RANDOM, which randomly selected the animals.
2. MASS, which selected the animals with the best phenotypes.
3. PBLUP_OP, which selected the animals with the best pedigree based estimated breeding value, estimated including own performance information.
4. GBLUP_NoOP, which selected the animals with the best genomic estimated breeding value (GEBV), estimated excluding own performance information.
5. GBLUP_OP, which selected the animals with the best GEBV, estimated including own performance information.

With PBLUP, the last eight generations of pedigree and last three generations of phenotypes were used. With GBLUP, the last three generations of genotypes and phenotypes were used. Breeding values were estimated with the MTG2 software (Lee and van der Werf, 2016), using a model including a fixed mean, a random additive genetic effect and a residual.

Allele frequency changes and fixation of loci. We compared the average change in allele frequency and the number of loci that became fixed during the 50 generations of selection, requiring a minimal change in allele frequency of 0.2 to disregard loci that initially were already close to fixation. For the loci that became fixed, we investigated the average additive effect and starting allele frequency and whether the favourable alleles was fixed or lost.

To disentangle the impact of drift and selection, additional simulations were run with random selection using a population with the same effective population size (N_e) as in the selection scenarios. For each scenario, the N_e was estimated using the pedigree kinship coefficient based on the off-diagonal elements of the pedigree relationship matrix. Then we measured the effect of selection as the additional number of loci that became fixed in the selection scenarios compared to the drift scenario.

Results

All scenarios with selection resulted in an improvement of the phenotype over the 50 generations. This improvement was largest for GBLUP_OP and lowest for GBLUP_NoOP (Table 1). Both GBLUP methods lost ~82% of the genetic variance, which considerably reduced the rate of genetic gain across generations. The loss in genetic variance was considerably less with MASS (~60%). Therefore, the rate of genetic gain after 50 generations was highest for MASS. The average change in allele frequency was 2-3 times larger with selection than with RANDOM, and largest for the GBLUP scenarios, followed by PBLUP_OP and finally MASS. Compared to RANDOM (N_e=222), selection reduced N_e most for PBLUP_OP (46), followed by the GBLUP scenarios (74-82), and finally MASS (136).

Table 1. Change in average phenotype, genetic variance and allele frequency across 50 generations of selection[1].

	Phenotypic change[2]		Proportion of genetic variance lost		Average change in allele frequency		Effective population size	
RANDOM	0.02	(0.11)	0.03	(0.03)	0.023	(0.000)	222	(0.8)
MASS	27.31	(0.44)	0.60	(0.01)	0.054	(0.001)	136	(1.8)
PBLUP_OP	27.02	(0.46)	0.79	(0.01)	0.062	(0.000)	46	(0.9)
GBLUP_NoOP	25.25	(0.45)	0.82	(0.01)	0.063	(0.001)	74	(1.4)
GBLUP_OP	29.37	(0.42)	0.83	(0.00)	0.065	(0.001)	82	(1.5)

[1] Results are shown as averages across replicates with standard errors across replicates between brackets.

[2] Expressed in base generation additive genetic standard deviations.

With RANDOM, very few loci became fixed, and there was an equal chance of fixing or losing the favourable allele (Table 2). Selection resulted in fixation of a considerable number of loci, being similar for PBLUP_OP, GBLUP_NoOP and GBLUP_OP, which was ~2.7 times higher than with MASS. With selection, more loci became fixed for the favourable allele than for the unfavourable allele, which was more emphasized with MASS (~89%) than with the other selection scenarios (~80%).

With selection, the average additive effect of the loci where the favourable allele became fixed was larger than for an average locus (average α across all loci was 0.041), and was largest with MASS. The average starting allele frequency of those loci was lower with selection, and lowest for GBLUP_OP, GBLUP_NoOP and PBLUP_OP. In contrast, the loci where the unfavourable allele became fixed had on average a lower additive effect than an average locus (0.024 vs 0.041). Moreover, the average starting allele frequency at those loci was largest for GBLUP_OP, GBLUP_NoOP, PBLUP_OP, followed by MASS and lowest for RANDOM.

For all selection methods, more loci became fixed for the favourable allele due to selection than due to drift, while more loci became fixed for the unfavourable allele due to drift. Both GBLUP methods and PBLUP_OP fixed a similar number of loci, but the impact of selection on the fixation of loci was higher for the GBLUP methods. Moreover, with both GBLUP methods, an additional number of loci became fixed for the unfavourable allele as a result of selection.

Table 2. Characteristics of loci that became fixed over 50 generations of selection[1].

			Favourable allele			Unfavourable allele		
	Total nr.[2]	Prop. fav.[3]	Avg. α[4]	Start freq.[5]	Sel.[6]	Avg. α[4]	Start freq.[5]	Sel.[6]
Random	2 (0.4)	0.46 (0.09)	0.035	0.74	0.00	0.046	0.25	0.00
MASS	133 (5.3)	0.89 (0.01)	0.061	0.59	0.18	0.021	0.34	0.02
PBLUP_OP	358 (4.6)	0.80 (0.01)	0.053	0.52	0.25	0.025	0.39	0.01
GBLUP_NoOP	356 (5.7)	0.79 (0.00)	0.052	0.52	0.30	0.027	0.40	0.07
GBLUP_OP	357 (5.3)	0.82 (0.00)	0.054	0.51	0.31	0.024	0.40	0.06

[1] Results are shown as averages across replicates with standard errors of the mean between brackets.

[2] Considering loci with a minimum change in allele frequency of 0.2.

[3] Proportion of loci fixed for the favourable allele.

[4] Average absolute statistical additive effect (α) in generation 0 in genetic standard deviation units.

[5] Frequency of the favourable allele in generation 0.

[6] Proportion of loci segregating in generation 0 that became fixed due to selection.

Discussion

The loss in genetic variance was considerably larger with the BLUP methods than with MASS, and MASS had the highest rate of genetic gain in the last generations. Therefore, for the longer term, MASS is expected to outperform GBLUP_OP in terms of cumulative genetic gain. The BLUP methods also showed a faster change in allele frequency and a greater rate of fixation of favourable alleles than MASS, which can be explained by the higher accuracy of the BLUP methods (Liu *et al.*, 2014).

The N_e was lowest for PBLUP_OP, which is a result of more family selection with PBLUP compared to GBLUP (Liu *et al.*, 2014; Wientjes *et al.*, 2021). So even though the average change in allele frequency and total number of loci becoming fixed was similar for GBLUP and PBLUP_OP, this was more due to drift and less due to selection with PBLUP_OP (21% due to drift and 25% selection) than with GBLUP (15% due to drift and 31% due to selection).

With the two GBLUP scenarios, an additional number of unfavourable alleles became fixed over and above the effect of drift. This was surprising, since selection was expected to prevent loss of favourable alleles on top of the effect of drift. An additional loss of favourable alleles due to selection did not happen with MASS and PBLUP, indicating that GBLUP resulted in greater fixation of unfavourable alleles due to genetic hitchhiking (Smith and Haigh, 1974; Barton, 2000).

Overall, GBLUP resulted in slightly larger and faster changes in allele frequencies of causal loci than PBLUP, and much larger and faster changes than MASS. As a consequence, GBLUP lost more favourable alleles. Maintaining rare favourable alleles is especially important for long-term genetic gain (Jannink, 2010), suggesting that GBLUP can limit the long-term genetic gain and should be accompanied by a management of genetic diversity.

Acknowledgements

This research is supported by the Netherlands Organisation of Scientific Research (NWO; VENI grant 16774). The use of the HPC cluster has been made possible by CAT-AgroFood (Shared Research Facilities Wageningen UR).

References

Barton N.H. (2000) Phil. Trans. R. Soc. B 355(1403): 1553-1562. https://doi.org/10.1098/rstb.2000.0716

Beniwal B.K., Hastings I.M., Thompson R., and Hill W.G. (1992) Heredity 69(4): 361-371. https://doi.org/10.1038/hdy.1992.136

Doekes H.P., Veerkamp R.F., Bijma P., Hiemstra S.J., and Windig J.J. (2018) Genet. Sel. Evol. 50(1): 15. https://doi.org/10.1186/s12711-018-0385-y

Havenstein G.B., Ferket P.R., and Qureshi M.A. (2003) Poult. Sci. 82(10): 1500-1508. https://doi.org/10.1093/ps/82.10.1500

Heidaritabar M., Vereijken A., Muir W.M., Meuwissen T., Cheng H. *et al.* (2014) Heredity 113(6): 503-513. https://doi.org/10.1038/hdy.2014.55

Jannink J.-L. (2010) Genet. Sel. Evol. 42(1): 35. https://doi.org/10.1186/1297-9686-42-35

Lee S.H., and van der Werf J.H.J. (2016) Bioinformatics 32(9): 1420-1422. https://doi.org/10.1093/bioinformatics/btw012

Liu H., Sørensen A.C., Meuwissen T.H.E., and Berg P. (2014) Genet. Sel. Evol. 46(1): 8. https://doi.org/10.1186/1297-9686-46-8

Sargolzaei M., and Schenkel F.S. (2009) Bioinformatics 25(5): 680-681. https://doi.org/10.1093 /bioinformatics/btp045

Smith J.M., and Haigh J. (1974) Genet. Res. 23(1): 23-35. https://doi.org/10.1017/S0016672300014634

Wientjes Y.C.J., Bijma P., Calus M.P.L., Zwaan B.J., Vitezica Z.G. *et al.* (2021) BioRxiv: 2021.2003.2016.435664. https://doi.org/10.1101/2021.03.16.435664

200. The effect of recombination rate on genomic selection in simulation

M. Johnsson[1*], A.M. Johansson[1], T. Klingström[1] and E. Rius-Vilarrasa[2*]

[1]Department of Animal Breeding and Genetics, Swedish University of Agricultural Sciences, Box 7023, 750 07 Uppsala, Sweden; [2]Växa Sverige, P.O. Box 30204, 104 25 Stockholm, Sweden; martin.johnsson@slu.se

Abstract

Genome features such as recombination rate vary between different regions of the genome and affect the distribution of genetic variation, thus potentially affecting the response of the genome to selection and the accuracy of genomic selection. In this paper, we build a quantitative genetic simulation with variable recombination rate based on real linkage maps. It confirms that higher recombination rate causes a somewhat higher long-term genetic gain due to preservation of additive genetic variance, and that high recombination rate lowers the accuracy of genomic selection. While the benefit happens in the long term, the decreased accuracy is immediate.

Introduction

In this paper, we build a quantitative genetic simulation with variable recombination rate in order to explore the effect of recombination rate on genomic selection in simulation. Features of the genome such as recombination rate, mutation rate and historical natural selection vary along the genome, and shape the distribution of genetic variation as they affect the supply of standing variation (Corbett-Detig *et al.* 2015, Harpak *et al.* 2016) and the extent of selective interference between variants (Hill and Robertson 1966).

Simulations are widely used in animal genetics to test the performance of analysis methods, understand the response to selection in breeding program designs, but also for mechanistic modelling of genetic variation. These simulations make many simplifying assumptions, including, often, abstracting away genome features and treating the genome as a more or less uniform grid of markers. Some of these assumptions may be entirely warranted, relying on the generality of quantitative genetic theory, but others may not.

In our research on the genome dynamics of livestock breeding, we hypothesise that there is additional knowledge about genetic architecture, and potentially genomic selection accuracy, to be gained by modelling the relationship between variation and features of the genome in more detail. This paper presents some of our pilot work on the effect of recombination rate on selection, using published physical and genetic maps of the cattle and chicken genomes.

Materials & methods

We used AlphaSimR (Gaynor, Gorjanc and Hickey 2021) to build a simulation of a closed breeding population with different karyotypes and recombination rate variation based on real linkage maps from cattle and chickens.

Simulation of population and traits. The founder population was generated using the coalescent simulator MaCS within AlphaSimR, using the 'GENERIC' population history, which models a large historical population size decreasing during domestication, and a final effective population size of 1000. We simulated one quantitative trait with either 100, 1000 or 10,000 causative variants drawn at random, proportioned to chromosomes based on their physical length. Additive genetic coefficients were drawn from a standard normal distribution. Dominance degrees were drawn from a normal distribution with mean 0.2 and variance 0.1, similar to values estimated by Bennewitz and Meuwissen (2010). Normally distributed noise was added to the phenotypes to achieve a heritability of 0.4.

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_200, © M. Johnsson *et al.* 2022

Chromosomes and linkage maps. Chromosomes were either based on the cattle karyotype, the chicken karyotype or were 20 chromosomes of equal size. Table 1 describes the different genomes simulated, with number of chromosomes, physical and genetic sizes. We only included autosomes that were included in the linkage map, meaning that the X and Z chromosome are missing, respectively, as well as some chicken microchromosomes. We used linkage maps from Ma *et al.* (2015) for cattle and Elferink *et al.* (2010) for the chicken. In order to update the physical coordinates, we used the UCSC LiftOver service to sequentially lift the physical position to ARS-UCD1.2 for cattle and GRCg6a for the chicken. We filtered the lifted maps to remove markers that caused disagreements in marker order, leaving a total of 58,010 in cattle and 12,907 in the chicken. We adjusted the genetic maps from coalescent simulation, which have uniform recombination rate, to correspond to the real linkage maps before breeding. We converted the initial genetic position of each site to a physical position, and linearly interpolated the new genetic position on the real linkage map based on that physical position. Linkage maps and code are available at https://github.com/mrtnj/lifted_recombination_maps.

Breeding structure. Each generation, 50 males (10%) and 100 females (20%) were selected to be the parents of next generation, and mated to produce 1000 offspring, half of each sex. The first ten generations of each replicate used phenotypic selection, and for the next ten generations, we either ran another ten generations of phenotypic selection or ten generations of genomic selection. In order to measure the decline in genomic accuracy, training was performed once using data from generation 6-9. The genomic selection model was SNP-BLUP based on 50,000 SNP markers randomly chosen but equally proportioned to chromosomes based on physical length. We evaluated accuracy as the correlation between true and the estimated breeding value. Each simulation case was run for 50 replicates. The simulation code is available at https://github.com/mrtnj/wcgalp_recombination.

Simulation with linkage equilibrium. For comparison, we created a simulation where the causative variants had the same starting frequencies, and distributions of additive and dominance coefficients, but where the loci were made to be in linkage equilibrium independently sampling the genotypes from a binomial distribution each generation. The linkage equilibrium simulation also used the same heritability, population size and a sex-averaged selection intensity as the AlphaSimR simulation.

Results

High recombination rate led to a lower loss of additive genetic variance on average (Figure 1). This effect was noticeable when there were many causative variants (i.e. not in the case with 100 loci), at later generations, but the difference was small compared to the variation between replicates. This was accompanied by somewhat higher genetic mean. Considering the case with 10,000 loci, in generation 20, the genetic mean was 4.4% higher for 1 M chromosomes, 6.4% higher for 2 M chromosomes, 3.6% higher for the cattle genome and 4.0% higher for the chicken genome compared to the shortest genome (0.5 M chromosomes).

Table 1. Genomes simulated.

Case	Physical length (Mbp)	Genetic length (M)	Chromosomes
0.5 M chromosomes	2,000	10	20
1 M chromosomes	2,000	20	20
2 M chromosomes	2,000	40	20
Cattle genome	2,489	24	29
Chicken genome	941	27	27

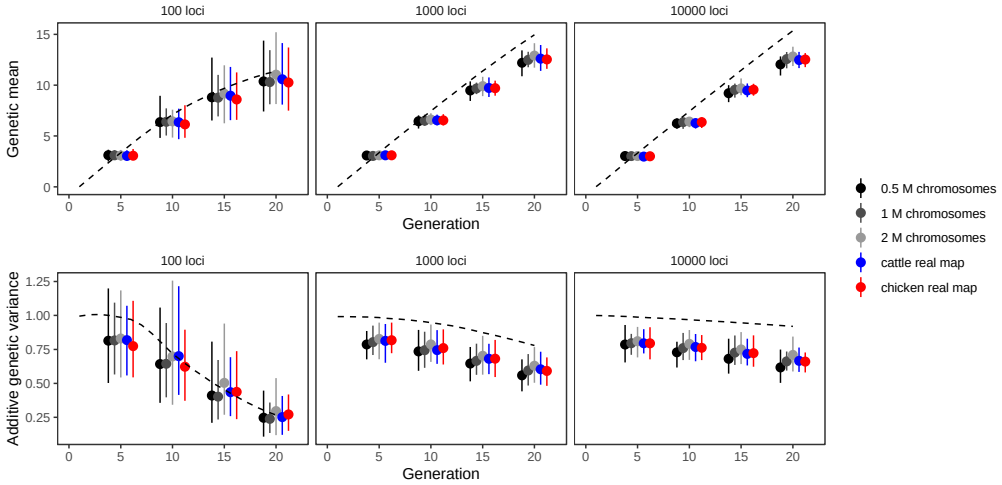


Figure 1. Changes in additive genetic variance with different recombination rates. The dashed line is the average from a simulation where causative variants are in linkage equilibrium.

On the other hand, higher recombination rate caused a decrease in genomic prediction accuracy (Figure 2). Considering the case with 10,000 loci, the first generation of genomic selection, average accuracies after one generation were 19% lower for 1 M chromosomes, 36% lower for 2 M chromosomes, 25% lower for the cattle genome, and 22% for the chicken genome compared to the most accurate 0.5 M chromosome case. When the genomes used the real cattle or chicken karyotype but had uniform recombination rate, the reduction was somewhat greater than with real linkage maps (27% for cattle and 26% for the chicken).

Discussion

The slight increase in genetic gain due to higher recombination is consistent with results from Battagin *et al.* (2011), who found that higher recombination rate led to higher genetic gain due to a smaller reduction in additive genetic variance, though the effect was small. The quantitative results are not comparable, as they used a different breeding structure, and selection on true breeding value. With genomic selection, the benefit from high recombination rate would be offset by a decreased genomic selection accuracy; while the improvement due to higher recombination happens in the long term and is small, the decrease in accuracy is immediate.

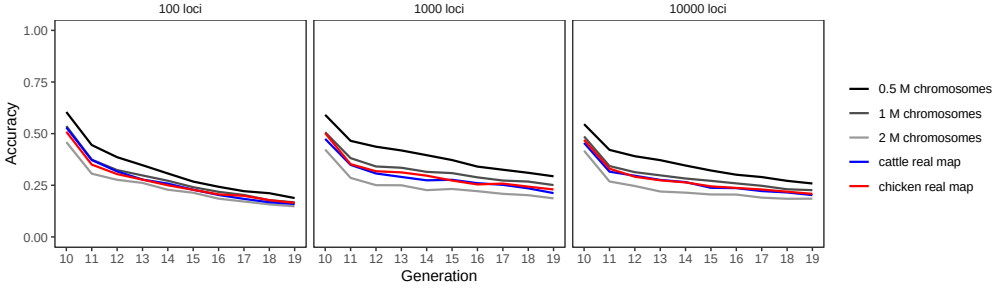


Figure 2. Decline in genomic selection accuracy over generations increasingly removed from a constant reference population, with different recombination rates.

As there is substantial variation within the genome in recombination rate, different regions of the genome should display local variation in its contribution to genomic selection accuracy, and to a lesser extent to long term gain. In the linkage maps used here, the 25 and 75% quantiles of recombination rate in between-marker intervals were 0.060 and 1.6 cM/Mbp for the cattle linkage map and 0.0 and 4.1 cM/Mbp for chicken linkage map.

Except for the real linkage maps, most aspects of these simulations were highly simplified. These are some avenues for development: the simulations used a generic population history, with a domestication bottleneck and large ancestral population. Work is ongoing to infer population history, and synthesise that with published estimates. As the real linkage maps were only introduced after the coalescent simulation of the founder population, the regional variation in linkage disequilibrium was only built up during the simulated breeding. Also, the real recombination rate landscape is more punctuated, with short 'hotspots' of a few kilobasepairs. Work is ongoing to infer fine-scale recombination rate and use it to improve the realism of the founder simulation. Finally, we used a generic breeding structure of a small closed population. The flexibility of AlphaSimR as well as recent developments to communicate and visualise breeding structures (Simianer *et al.* 2021) will be helpful for improving this aspect.

Acknowledgements

The research was supported by grant Dnr. 2020-01637 from Formas – a Swedish Research Council for sustainable development.

References

Battagin, M., Gorjanc, G., Faux, A. M., Johnston, S. E., & Hickey, J. M. (2016) Genet Sel Evol 48(1) https://doi.org/10.1186/s12711-016-0221-1

Bennewitz, J., and Meuwissen, T. H. E. (2010) J Anim Breed Genet 127(3), 171-179 https://doi.org/10.1111/j.1439-0388.2009.00847.x

Corbett-Detig, R. B., Hartl, D. L., and Sackton, T. B. (2015) PLoS Biol 13(4):e1002112. https://doi.org/10.1371/journal.pbio.1002112

Elferink, M. G., van As, P., Veenendaal, T., Crooijmans, R. P., and Groenen, M. A. (2010) BMC Genet 11(1) https://doi.org/10.1186/1471-2156-11-11

Gaynor, R. C., Gorjanc, G., and Hickey, J. M. (2021) G3 11(2):jkaa017 https://doi.org/10.1093/g3journal/jkaa017

Harpak, A., Bhaskar, A., and Pritchard, J. K. (2016) PLoS Genet, 12(12):e1006489. https://doi.org/10.1371/journal.pgen.1006489

Hill, W. G., and Robertson, A. (1966) Genetics Res 8(3):269-294 https://doi.org/10.1017/S0016672300010156

Ma, L., O'Connell, J. R., VanRaden, P. M., Shen, B., Padhi, A., *et al.* (2015). PLoS Genet 11(11): e1005387 https://doi.org/10.1371/journal.pgen.1005387

Simianer, H., Büttgen, L., Ganesan, A., Ha, N. T., and Pook, T. (2021) J Anim Breed Genet 138(2), 137-150 https://doi.org/10.1111/jbg.12534

201. Fine-scale study of meiotic recombination in cattle

C. Oget-Ebrad[1*], G.C.M. Moreira[1], L. Karim[2], W. Coppieters[2], C. Charlier[1], M. Georges[1] and T. Druet[1]

[1]Unit of Animal Genomics, GIGA-R and Faculty of Veterinary Medicine, University of Liège (B34), 4000 Liège, Belgium; [2]Genomics Platform, GIGA, University of Liège (B34), 4000 Liège, Belgium; claire.oget-ebrad@inrae.fr

Abstract

We herein characterized fine-scale meiotic recombination rate (RR) variation in a whole-genome sequenced cattle pedigree including 264 individuals. With a family-based approach, we identified 6,644 cross-over (CO) events (median length: 3.9 kb) whereas with a population-based method, we observed the presence of 12,776 recombination hotspots (median length: 1.5 kb). Although these hotspots represented only 0.97% of the genome, 38.3% of the CO overlapped with these hotspots (40.6% in males, 34.8% in females), suggesting that they might still be in use in the current population. Subsequently, we determined that RR was higher in open chromatin regions, in CpG islands, and in predicted PRDM9 motifs. Conversely, lower RR were estimated close to genes except in the transcription start sites. Finally, higher RR were estimated around H3K4me3 histone mark in testis, further supporting the implication of the PRDM9 zinc finger protein in meiotic recombination in cattle.

Introduction

Meiotic recombination is an important biological process as it ensures the proper segregation of homologues during meiosis and it shapes the patterns of linkage disequilibrium (LD) by creating new combinations of alleles (see for instance Coop and Przeworski, 2007, Baudat *et al.*, 2010, and Peñalba and Wolf, 2020, for reviews). Two approaches can be used to study meiotic recombination from genotypes (Peñalba and Wolf, 2020). First, family-based approaches allow to identify sex-specific cross-over (CO) events and provide current recombination rates (RR). Second, population (or LD)-based approaches provide historical sex-averaged RR. These latter maps represent relative estimates as they are influenced by other parameters such as past effective population size or selection intensity. Two mechanisms directing recombination are currently known (Baudat *et al.*, 2010). The first one is PRDM9-independent, where recombination is concentrated at TSS and CpG islands, and where we observe stable patterns of recombination across generations, as seen in yeasts, plants, and vertebrate species such as birds. In contrast, in most mammals such as in human, recombination is PRDM9-directed, that is, concentrated at the positions where PRDM9 protein is binding, away from genes, and where we observe a fast evolution of PRDM9 binding-motifs and a divergence between close species (Baudat *et al.*, 2010). Here, we take advantage of a unique sequenced cattle pedigree to characterize the patterns of RR in the bovine genome and to study the relationship between recombination and genomic features.

Materials & methods

Sequence data. Whole-genome sequence data was available for a small pedigree containing 264 Holstein-Friesian individuals. Based on filtering and processing steps described in Oget-Ebrad *et al.* (2021), we selected 8.4 million SNPs for further analysis.

CO events identification. We used the family-based approach implemented in LINKPHASE3 (Druet and Georges, 2015) to identify CO events in the sequence data. We removed double CO occurring within 1 Mb, and multiple identical CO within a family to remove putative errors. CO intervals were then refined using ShapeIT4.1 (Delaneau *et al.*, 2019).

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_201, © C. Oget-Ebrad *et al.* 2022

LD-based RR and hotspots. We used LDhelmet (Chan *et al.*, 2012) to estimate population-based RR and sequenceLDhot (Fearnhead, 2006) to identify recombination hotspots.

PRDM9-related features. We predicted the PRDM9 binding-motifs for each allele of the BTA1 paralog previously characterized for the Holstein breed (Zhou *et al.*, 2018), using the approach from Persikov and Singh (2014). We scanned the genome for potential binding locations of each of these predicted motifs with FIMO (Grant *et al.*, 2011).

Other genomic features. We used available public database to locate genomic features such as the transcription start sites (TSS) (Goszczynski *et al.*, 2021), and CpG islands (https://hgdownload.soe.ucsc.edu/goldenPath/bosTau9/database/). Functional annotation was available from experiments performed in testis, including an Assay for Transposase-Accessible Chromatin using sequencing (ATAC-seq), and epigenetic data for histone marks (H3K27me3, H3K4me3, H3K4me1, H3K27ac) and for one DNA-binding protein (CTCF).

Results

Broad-scale variation of RR. We identified 6,644 CO events (3,895 for males and 2,749 for females; median length: 3.9 kb) with the family-based approach and characterized 12,776 LD-based hotspot regions (median length: 1.5 kb; RR=4.1/kb in the hotspots vs 0.18/kb as whole-genome average), the latter representing 0.97% of the genome. Thirty-eight percent of the CO overlapped with LD-based hotspots (40.6% in males, 34.8% in females) and more than 50% were within 5 kb from the hotspots (Figure 1). This proportion was highly significant compared to random sets of CO events ($P<0.001$; 2.4% overlap on average with 1000 simulations). We also observed that both CO and hotspots density were higher at telomeres.

Fine-scale variation of RR. In our population, we observed the five alleles of the BTA1 *PRDM9* paralog previously described by Zhou *et al.* (2018) at following frequencies: 12.3% for allele 1, 51.1% for alleles 2 and 3, 23.1% for allele 4, and 13.4% for allele 5. A total of 149,620 PRDM9 predicted motifs were identified in the genome. Although we observed increased RR around these motifs (Figure 2a), only 11.3% of the LD-based hotspots contained at least one motif. We observed reduced RR close to TSS (Figure 2b) and increased

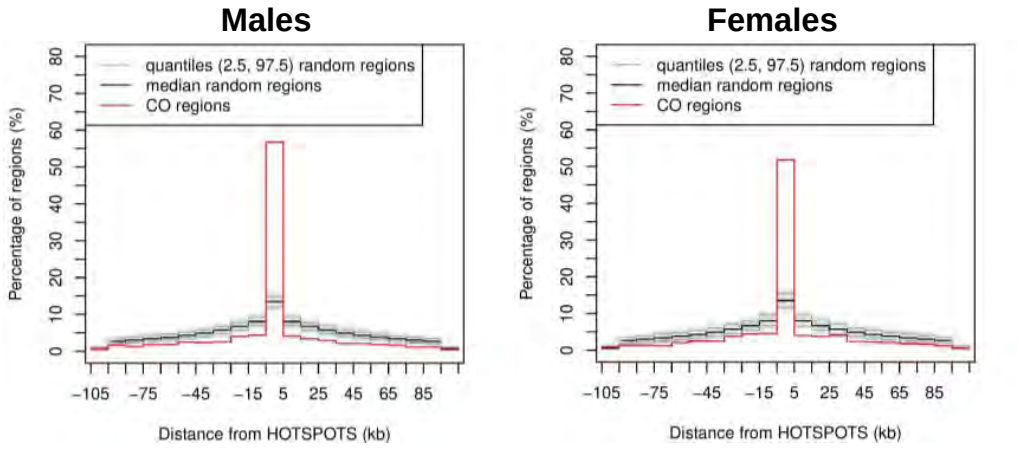


Figure 1. Percentage of CO events (red) identified with the family-based approach in males and females as a function of the distance (in kb) from the closest hotspot from the population-based approach, compared to 1000 simulated random sets of CO matching in number and size (grey for quantiles, black for median).

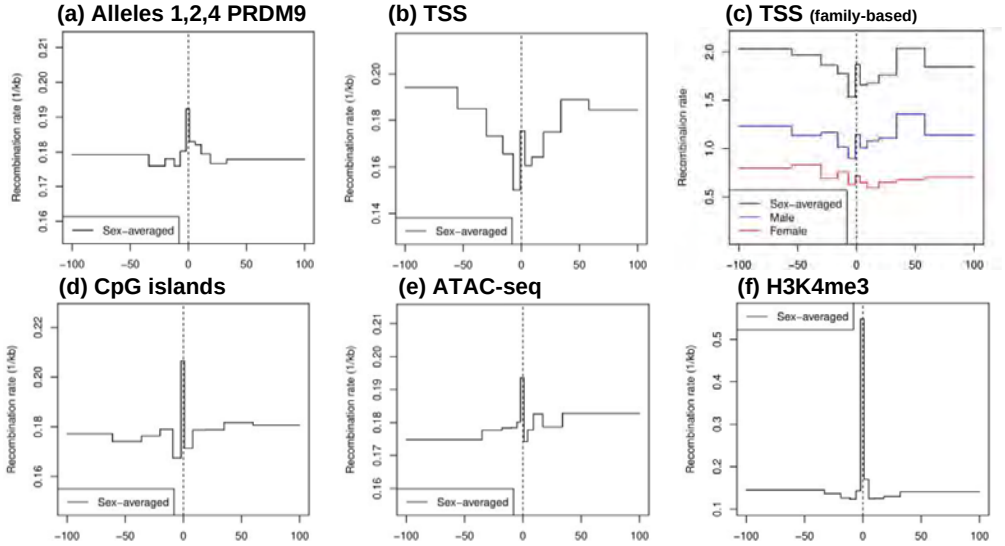


Figure 2. Recombination rates (RR) as a function of the distance (in kb) from the closest (a) BTA1 PRDM9 predicted binding-motif for alleles 1, 2 & 4, (b-c) transcription start site (TSS), (d) CpG island, (e) ATAC-seq peak, and (f) H3K4me3 signal. RR were estimates from the population-based approach, except in (c).

RR in CpG islands (Figure 2d). RR were higher in accessible DNA regions (ATAC-seq peaks, Figure 2e) and in CTCF signals. Methylation marks from active promoters (H3K4me1 and H3K27ac) presented lower RR. We observed a strong increase of RR in the H3K4me3 methylation mark (Figure 2f). This was specific to assays performed in testis and not observed with ATAC-seq data from other tissues (e.g. liver). Some of these associations were also observed with the set of identified CO (Figure 2c), with no striking difference between males and females, although the patterns were less clear than with the population-based approach.

Discussion

In cattle, association between PRDM9 and individual RR and CO events location (at relatively broad scale) has been previously reported, as in human and mice. Meiotic recombination has been indeed previously studied in cattle with sparser marker maps (Ma *et al.*, 2015). Based on whole-genome sequence data, our findings further support the role of PRDM9 in this species. First, we observed the presence of recombination hotspots (median length: 1.5 kb) in cattle, as in human and mice (Baudat *et al.*, 2010). These recombination hotspots represent 0.97% of the genome (1.7-1.9% in human (Kong *et al.*, 2010)). They were significantly overlapping with identified CO, suggesting that the historical hotspots are still in use in the present population. Then, we observed increased RR around BTA1 PRDM9 binding-motifs, as well as a small proportion of hotspots (11.3%) containing at least one PRDM9 binding-motif. This relatively low proportion (10.1% expected with random location) might result from inaccuracies in hotspots identification or in binding-motif predictions. The hotspots might also be associated with other *PRDM9* alleles that we did not identify, for instance alleles associated with the *PRDM9* paralogue on BTAX or alleles that are no longer present in the population. Overall, when multiple *PRDM9* alleles contribute to recombination hotspots, motifs enrichment analysis might fail to identify specific *PRDM9* alleles. Finally, a strong increase of RR in the H3K4me3 methylation mark was observed. This mark is known to be deposited by PRDM9 after binding DNA and is mandatory for the initiation of the recombination process (Diagouraga *et al.*, 2018), further supporting the implication of PRDM9 in meiotic recombination in cattle.

More generally, we found that RR increased in CpG islands and in open chromatin regions as seen in human (Halldorsson *et al.*, 2019), and decreased in regions associated to active promoters. CO would thus be kept away from genes and motifs should evolve rapidly; the recombination landscape could thus differ among breeds or between closely related species.

References

Baudat F., Buard J., Grey C., Fledel-Alon A., Ober C., *et al.* (2010) Science 327(5967):836-840. https://doi.org/10.1126/science.1183439

Chan A.H., Jenkins P.A., and Song Y.S. (2012) PLoS Genet. 8(12):e1003090. https://doi.org/10.1371/journal.pgen.1003090

Coop G. and Przeworski M. (2007) Nat. Rev. Genet. 8(1):23-34. https://doi.org/10.1038/nrg1947

Delaneau O., Zagury J.-F., Robinson M.R., Marchini J.L., and Dermitzakis E.T. (2019) Nat Commun 10(1):5436. https://doi.org/10.1038/s41467-019-13225-y

Diagouraga B., Clément J.A.J., Duret L., Kadlec J., de Massy B., *et al.* (2018) Molecular Cell 69(5):853-865.e6. https://doi.org/10.1016/j.molcel.2018.01.033

Druet T. and Georges M. (2015) Bioinformatics 31(10):1677-1679. https://doi.org/10.1093/bioinformatics/btu859

Fearnhead P. (2006) Bioinformatics 22(24):3061-3066. https://doi.org/10.1093/bioinformatics/btl540

Goszczynski D.E., Halstead M.M., Islas-Trejo A.D., Zhou H., and Ross P.J. (2021) Genome Res. 31(4):732-744. https://doi.org/10.1101/gr.267336.120

Grant C.E., Bailey T.L., and Noble W.S. (2011) Bioinformatics 27(7):1017-1018. https://doi.org/10.1093/bioinformatics/btr064

Halldorsson B.V., Palsson G., Stefansson O.A., Jonsson H., Hardarson M.T., *et al.* (2019) Science 363(6425). https://doi.org/10.1126/science.aau1043

Kong A., Thorleifsson G., Gudbjartsson D.F., Masson G., Sigurdsson A., *et al.* (2010) Nature 467(7319):1099-1103. https://doi.org/10.1038/nature09525

Ma L., O'Connell J.R., VanRaden P.M., Shen B., Padhi A., *et al.* (2015) PLoS Genet. 11(11):e1005387. https://doi.org/10.1371/journal.pgen.1005387

Oget-Ebrad C., Kadri N.K., Moreira G.C.M., Karim L., Coppieters W., *et al.* (2021) bioRxiv:2021.10.27.466052. https://doi.org/10.1101/2021.10.27.466052

Peñalba J.V. and Wolf J.B.W. (2020) Nat. Rev. Genet. 21(1):476-492. https://doi.org/10.1038/s41576-020-0240-1

Persikov A.V. and Singh M. (2014) Nucleic Acids Res. 42(1):97-108. https://doi.org/10.1093/nar/gkt890

Zhou Y., Shen B., Jiang J., Padhi A., Park K.-E., *et al.* (2018) DNA Res. 25(2):183-194. https://doi.org/10.1093/dnares/dsx048

202. Sex differences in recombination maps are associated with differential hotspot usage in sheep

H. Vassilief[1], M. Id Bella[1], D. Hazard[1], F. Tortereau[1], T. Faraut[1], S.E. Johnston[2] and B. Servin[1*]

[1]INRAE, GenPhySe, 24 chemin de Borde-Rouge, 31320 Castanet-Tolosan, France; [2]University of Edinburgh, Institute of Evolutionary Biology, University of Edinburgh, Ashworth Laboratories, Charlotte Auerbach Road, Edinburgh EH9 3FL, United Kingdom; bertrand.servin@inrae.fr

Abstract

Recombination is a fundamental biological process for the reproduction and evolution of species. Recombination phenotypes have been shown to exhibit large inter-individual variation with a significant genetic determinism. Here we make use of large genotyping datasets in the Sheep to: (1) study the distribution of recombinations along the genome (recombination maps); and (2) evaluate its inter-individual variation using a phenotype termed *hotspot usage* (HSU). We precisely estimated sex specific recombination maps and found that sex differences in recombination rates are concentrated in 16% of the genome, mostly at chromosome extremities. Individual variation in HSU is dominated by a large difference between sexes: males are found to preferentially use recombination hotspots contrary to females. This difference is most pronounced in regions with large sex differences in recombination rate. This suggests that sex difference in recombination maps in Sheep could be due to different crossover determination processes in male and female meioses.

Introduction

Recombination is a fundamental biological process that exhibits variation among species, populations and individuals. Two main recombination phenotypes have been studied at the individual level: recombination rate and recombination localisation. Individual recombination rates can be measured by counting the number of crossovers per meiosis (using pedigree or cytogenetic approaches). However, individual variation in crossover localisation is more difficult to measure as it requires data from a large number of meioses per individual and/or a very precise localisation of individual crossovers. One of the individual phenotypes that can be measured to evaluate variation in recombination maps is *hotspot usage (HSU)*. It is the propension for an individual to recombine in small genomic regions of high average historical recombination rate called hotspots. In this study we used a new dataset of precisely localised crossovers in sheep families to estimate HSU in multiple individuals and investigate its variation between the sexes.

Materials & methods

Datasets. This study exploits three datasets. First, two previously published datasets of medium density genotypes (~50K SNPs) in large pedigrees in Lacaune (Petit *et al.* 2017) and Soay (a primitive breed of domestic sheep; Johnston *et al.* 2016) sheep populations to build sex-specific recombination maps. Second, a new dataset of small nuclear families in a Romane population genotyped for a high-density SNP array (HD; ~600K SNPs) and used to study HSU variation. Datasets were phased and crossover positions identified using the software yapp (https://yapp-doc.netlify.app/).

Sex specific recombination maps. Recombination maps were estimated for each sex using poisson regression on crossover (CO) counts observed in non-overlapping 1 Mb intervals along the genome using the glm procedure in R. Sex-specific recombination maps were estimated considering each population by sex as a group in a Poisson-Log-Normal Model framework using the PLNmodels R package (Chiquet, Mariadassou, and Robin 2021). An extension of the model allowing for a mixture of poisson distributions assigned each 1 Mb interval to a specific mixture component and identified intervals with clear differences in recombination rates between sexes.

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_202, © H. Vassilief *et al.* 2022

Hotspot usage estimation. Hotspot usage was estimated in individuals from the HD Romane dataset and hotspots identified in Petit *et al.* (2017) using the procedure described in Coop *et al.* (2008). Briefly, this method estimates HSU in a collection of crossovers as the proportion that use a hotspot, adjusted for the size of the crossover localisation interval and the local hotspot density. First the probability that each crossover overlaps a hotspot by chance is estimated by randomly shifting its position locally. This allows to account for the size of the localisation interval (crossovers with large intervals have a higher probability of overlap by chance) and the local recombination rate (crossovers located in regions of high hotspot density have higher probability of overlap by chance). Then given a set of crossovers (for an individual, a sex …), the HSU is estimated by maximizing the likelihood that depends on these overlap probabilities (see Coop *et al.* (2008) for details).

Results

Table 1 summarises the crossover data for the three datasets. The two medium density datasets identify a large number of crossovers and can be used to precisely estimate average recombination rates. The HD dataset resolves crossovers in intervals roughly 10 times smaller and allows estimation of HSU.

Sexes differ in their recombination maps in few genomic intervals. Statistical analysis of crossover counts in the two MD datasets reveals that genomic intervals can be separated into 2 components in terms of recombination rate variation between groups. The majority component I (84% of the genome, green on Figure 1) exhibits essentially the same recombination rate in males and females. The minority component II (16% of the genome, purple on Figure 1) corresponds to intervals with large between sex variation, most of them located at the extremities. In both classes, the Soay and Lacaune populations have essentially the same recombination rates in males and females, despite their large genetic distance.

Hotspot usage exhibits strong sex differences and within sex variation. Pooling all crossovers from the Romane dataset, the HSU estimate is 59% with 95% confidence interval [57-61]. Males had significantly higher HSU (72%, CI [70-75]) than females (29%, CI [24-34]) (LRT=207.67 (1 df), P=4.5×10^{-47}). Figure 2 left shows that there is a large variation in individual estimates of HSU but mostly due to the large difference between sexes. However, a few outliers can be identified (two females with HSU>75 and 3 males with HSU<60) and the likelihood ratio tests on individual variation within each sex are significant (for females LRT=57.34 (27 df) P=5.8×10^{-4}, for males LRT=52.36 (28 df) P=3.5×10^{-3}).

Sex difference in hotspot usage and recombination rate are associated. Estimating hotspot usage in component II vs component I intervals reveals that females have an extremely low hotspot usage in intervals of class II while males have the same high hotspot usage in both components (Figure 2 right).

Table 1. Crossover data used in this study.

Population	Sex	# Parents	# Meioses	# CO	CO localization size (kb)[1]
Lacaune	Female	286	483	13,789	715 (249-4,168)
Lacaune	Male	542	8,991	325,487	687 (202-6,311)
Soay	Female	1,001	4,471	123,852	1,204 (266-9,614)
Soay	Male	631	4,157	144,237	829 (160-8,129)
Romane	Female	28	59	2,665	82.3 (18-2,605)
Romane	Male	29	166	6,433	91.7 (17-4,012)

[1] Median size (lower – upper 5% quantiles)

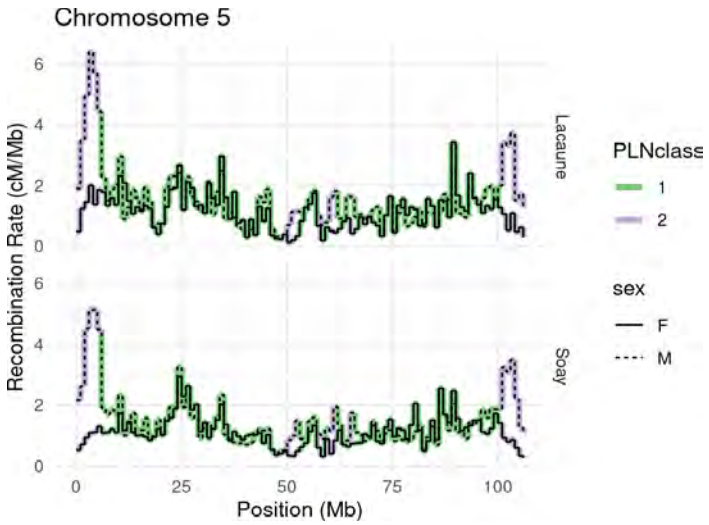


Figure 1. Example of recombination maps along chromosome 5 in two sheep populations and two sexes. Genomic intervals are coloured according to their mixture component (PLNclass).

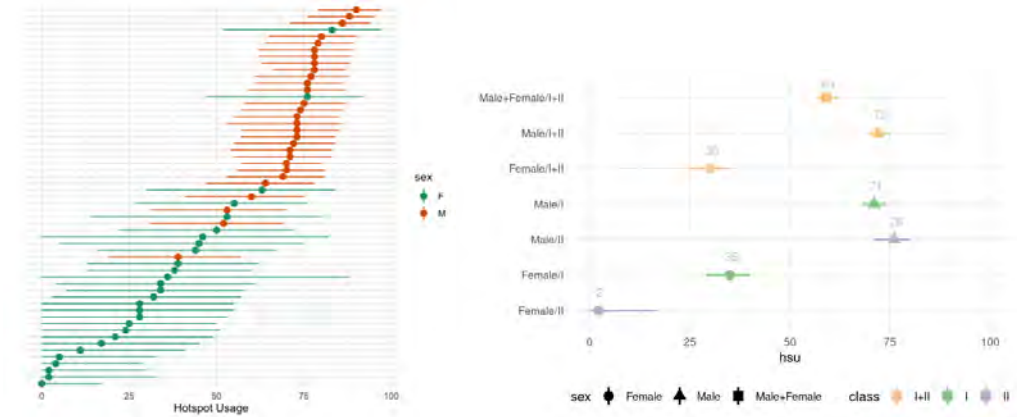


Figure 2. Hotspot Usage in the Romane population. Left panel: individual estimates of HSU Right panel: HSU in different genomic interval components and sex.

Discussion

We have established precise recombination maps in males and females in two highly differentiated populations of sheep. While sex differences in the recombination landscape are large, within sex between population variation is mostly negligible at the megabase scale so maps established here can be generally useful in many sheep populations. However, our results suggest that genetic analysis of segregation in pedigrees or genomic selection methods exploiting gametic variance, e.g. Bijma *et al.* (2020), should take account of between sex variation in recombination.

The sex difference in recombination rate between males and females in sheep is characterized by females recombining less than males due to a lower recombination rate in a small part of the genome mostly located at chromosome extremities. This is consistent with previous results in many mammals including cattle

(Ma *et al.* 2015), mouse (Peterson and Payseur 2021) and human (Broman *et al.* 1998) but different from red deer (Johnston *et al.* 2017) or pig (Tortereau *et al.* 2012) where females rather than males tend to have elevated recombination rates at chromosome ends.

Analysis of inter-individual variation in recombination localisation (HSU) show that the sex-difference in the recombination landscape in sheep is associated with a lower overall hotspot usage in females, particularly in the genomic intervals that exhibit large sex-differences in recombination rates. These differences could be due to differential use of crossover resolution pathways as was shown recently in the mouse (Brick *et al.* 2018; Powers *et al.* 2020).

References

Bijma P., Wientjes Y.C.J., Calus M.P.L., 2020. *Genetics* 214(1):91-107. https://doi.org/10.1534/genetics.119.302643

Brick, K., Thibault-Sennett, S., Smagulova, F., Lam, K.-W. G., Pu, Y., *et al.* 2018. *Nature* 561 (7723): 338–42. https://doi.org/10.1038/s41586-018-0492-5.

Broman K.W., Murray, J.C., Sheffield V.C., White, R.L., Weber J.L., 1998. *AJHG* 63(3):861 http://doi.org/10.1086/302011

Chiquet, J., Mariadassou, M. & Robin, S. 2021. *Frontiers in Ecology and Evolution* 9. https://doi.org/10.3389/fevo.2021.588292.

Coop, G., Wen, X., Ober, C., Pritchard, J. K. & Przeworski, M. 2008. *Science* 319 (5868): 1395–98. https://doi.org/10.1126/science.1151851.

Johnston, S. E., Bérénos, C., Slate, J. & Pemberton, J. M. 2016. *Genetics* 203 (1): 583–98. https://doi.org/10.1534/genetics.115.185553

Johnston S.E., Huisman J., Ellis P.A., Pemberton J.M., *G3* 7(8):2859-2870 http://doi.org/10.1534/g3.117.044198

Ma, L., O'Connell J.R., VanRaden P.M., Shen B. Padhi, A. *et al.* 2015. *PLoS Genetics* 11 (11):e1005387 https://doi.org/10.1371/journal.pgen.1005387

Peterson A. L., Payseur B. A., 2021. *Genetics* 217 (1):iyaa019 https://doi.org/10.1093/genetics/iyaa019

Petit, M., Astruc, J.-M., Sarry, J., Drouilhet, L., Fabre, S., *et al.* 2017. *Genetics* 207 (10): 767–84. https://doi.org/10.1534/genetics.117.300123.

Powers, N. R., Dumont, B. L., Emori, C., Lawal, R. A., Brunton, C., *et al.* 2020. *Science Advances* 6 (43): eabb6606. https://doi.org/10.1126/sciadv.abb6606

Tortereau F., Servin B., Frantz L., Megens H.J., Milan D., *et al.* 2012. *BMC Genomics* 13(1):586 https://doi.org/10.1186/1471-2164-13-586

203. Is it possible to differentiate meat products of a local breed from those of its sister breed based on genotypes?

H. Wilmot[1,2*], G. Glorieux[3], X. Hubin[3] and N. Gengler[2]

[1]National Fund for Scientific Research (F.R.S.-FNRS), Rue d'Egmont, 5, 1000 Brussels, Belgium; [2]Gembloux Agro-Bio Tech, University of Liège, Passage des Déportés, 2, 5030 Gembloux, Belgium; [3]Walloon Association of Breeders, Rue des Champs Elysées, 4, 5590 Ciney, Belgium; helene.wilmot@uliege.be

Abstract

To maintain endangered breeds and preserve their intrinsic diversity, it is often advocated to develop derived labelled products and the related certification process. However, this certification process is not always easily implemented if the endangered breed is closely related to another one. The breed traceability is only possible if the genetic diversity between those breeds is high enough. In this study, it was determined if meat of Dual-Purpose Blue (DPB) animals can be differentiated from meat of its sister breed, the Beef Belgian Blue (BBB), by the application of a genomic test. The results showed that the meat of DPB was completely distinguished from the meat of BBB, and this, with a probability of one for all meat samples. It therefore seemed possible to differentiate the DPB derived products from those of BBB with a high accuracy, meaning the implementation in routine of a certification process seemed possible.

Introduction

The Dual-Purpose Blue (DPB) breed can be considered as the old type of the previously called Mid and Upper Belgium Breed, from which the well-known Beef Belgian Blue (BBB) also derived. Both branches (DPB and BBB) separated during the seventies (Mota *et al.*, 2017). However, from this point, the BBB gained popularity while the number of DPB decreased to the point it is now considered an endangered breed. The BlueSter project (BlueSter, 2021) aimed to maintain this breed by, among other measures, *in situ* conservation and the development of local products. The consumer confidence can be ensured through a certification process using a genomic breed assignment tool for breed derived-products (e.g. Judge *et al.*, 2017). However, implementing this kind of certification is not always easy if the endangered breed is closely related to another one, as it is in our case. The objective of this study was therefore to determine if it was possible to differentiate meat products of the endangered DPB breed from those of its mainstream sister breed, the BBB, by the use of a genomic breed assignment test previously developed.

Materials & methods

All computations were run on R v.4.1.2 (R Core Team, Vienna, Austria, 2021) and visualized on R Studio (R Studio Team, Boston, USA, 2020).

Breeds. The two main breeds of interest in this study were the DPB, a transboundary dual-purpose breed of Southern Belgium and the North-East of France, and the BBB, a beef breed mostly located in Belgium. Whereas all BBB animals are homozygous for a partially recessive allele 'mh' resulting in muscular hypertrophy, a disparity is found in the DPB population where the wild type of the allele '+' is still existing (Mota *et al.*, 2017). In other words, depending on the selection objective of the breeder, DPB animals can be 'mh/mh' (in farms with a more meat-orientated breeding goal, mostly found in Belgium), 'mh/+' or '+/+' (in farms with a more dairy-orientated breeding goal, as most French breeders), and two or three genotypes can be coexisting in the same farm.

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_203, © H. Wilmot *et al.* 2022

Datasets. Two different datasets were used in this study: the first one was used for the principal component analysis, aiming to determine if the breeds of interest can be differentiated, and the second one was used by the breed assignment model. For the principal component analysis, 60 DPB and 60 BBB, considered as the reference population of their respective breed, were studied. Both Belgian (n=52) and French animals (n=8) were represented in the DPB population. For the breed assignment model, a total of 32 samples of meat pieces, coming from three breeds, were genotyped with the EuroG MD v2 (Illumina, San Diego, CA, USA). The three breeds included were the BBB (n=8), the DPB (n=16) and the Holstein (HOL, n=8). The BBB animals were born in four different farms but all fattened at the same place. The DPB animals came from two different breeders in a balanced manner (n=8 for each breeder). One of the breeders was located in Belgium and only provided 'mh/mh' animals to be slaughtered. The other one was located in France and raised both 'mh/+' and '+/+' animals. It was thus considered that the DPB breed was well represented as animals from the three genotypes were sampled. The HOL animals were also raised by different breeders and were considered as a 'control' for the study.

Principal component analysis. The PCA was based on a total of 17,667 SNPs as in a previous study (Wilmot *et al.*, 2022) and on the first dataset. Missing values per individual were imputed following a procedure of the missMDA v1.18 R package (Josse and Husson, 2012), related to PCA, and described in Wilmot *et al.* (2022). The FactomineR v.2.4 R package (Lê *et al.*, 2008) was used to perform the PCA on the matrix of correlations.

Breed assignment model. The breed assignment model used in this study to determine the breed of origin of meat samples for each of the 32 slaughtered animals was previously described in Wilmot *et al.* (2021) as the second best model to be used for breed assignment. Because the second best model used less SNPs than the first one (2,005 vs 7,153), it was preferred for routine analyses. In brief, the chosen model was based on 2,005 SNPs and used the Nearest Shrunken Centroids method to classify each individual. Twelve breeds can be detected with this model of which the BBB, the DPB and the HOL. Each meat sample was assigned to the breed for which it had the highest probability to belong to.

Results

Principal component analysis. The first two dimensions of the PCA are illustrated in Figure 1. There was a gap between animals from both breeds, even if one BBB animal was closely related to the DPB cluster. Therefore, it seemed possible to differentiate DPB from BBB. Moreover, as expected because of the different selection objectives and the three muscular genotypes existing in the DPB breed, there was a greater diversity in the DPB population compared to BBB animals i.e. DPB animals were more widespread than BBB animals in Figure 1.

Breed assignment of meat samples. A model based on 2,005 SNPs and on the Nearest Shrunken Centroids method was used to classify all the meat samples in the breed they were supposed to belong to. From the 2,005 SNPs, 167 were used to differentiate DPB from the other breeds existing in the reference population and 173 were used to differentiate BBB from the other reference breeds. An overlap of seven SNPs was observed between both SNP panels. The confusion matrix of the breed assignment test is available in Table 1. All meat samples were correctly attributed to their breed of origin, and this, with the maximum probability.

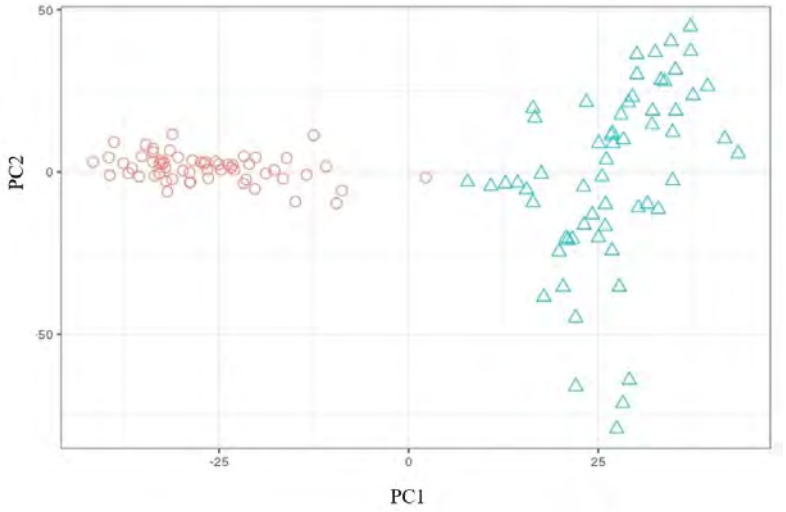


Figure 1. First two dimensions of the principal component analysis based on genotypes of reference Dual-Purpose Blue and Beef Belgian Blue. O = Beef Belgian Blue; Δ = Dual-Purpose Blue.

Table 1. Confusion matrix of meat samples based on the breed assignment test.

Predicted breed	Breed of origin		
	Beef Belgian Blue	Dual-Purpose Blue	Holstein
Beef Belgian Blue	8	0	0
Dual-Purpose Blue	0	16	0
Holstein	0	0	8

Discussion

The possibility to distinguish local meat products of an endangered breed, the DPB, from a mainstream sister breed, the BBB, based on genotypes, was studied. The breed assignment test allowed to perfectly differentiate the DPB meat samples from the BBB meat samples, with a probability of one for all meat samples, even those of DPB animals carrying the 'mh/mh' genotype, highlighting the high accuracy of the test.

This high accuracy can be partially explained by the fact that the SNPs used by the model were selected to differentiate each breed in the reference population from the others. Therefore, the panel comprised SNPs dedicated to differentiate DPB animals from BBB animals. It is then expected that significant phenotypic and genetic differences have arisen between both breeds since their divergence in the seventies, even for 'mh/mh' DPB animals. It has for example been suggested that the dairy performances were not necessarily inconsistent with the 'mh/mh' genotype as the more dairy DPB cows ('+/+') were producing on average 5,000 litres whereas the meat producing-DPB cows ('mh/mh') were still able to produce 4,000 litres (BlueSter, 2021). Moreover, it has been reported that, even when DPB animals carry the 'mh/mh' genotype, caesarean sections were more frequent in BBB than in DPB. It seemed that DPB breeders selected more for decreased birth weight and decreased calving difficulties, even within the 'mh/mh' animals, than BBB breeders (Mota *et al.*, 2017).

The fact that the breed assignment test can make the difference between genotypes of these two very closely related breeds, that diverged only from each other less than 50 years ago, is not trivial. It means that the genetic diversity between DPB and BBB was sufficiently high to differentiate these breeds. As a counter-example, the Angus and Red Angus cattle could not be effectively assigned to their respective breed in the study of Kuehn *et al.* (2011), even if both breeds started to separate much earlier (Márquez *et al.*, 2010) than DPB and BBB breeds. Besides, whereas the same three colour patterns (black-pied, blue-pied and white) are co-existing both in DPB and BBB, Angus and Red Angus clearly segregate from this point of view. However, the method chosen by Kuehn *et al.* (2011) was based on a regression of allelic frequencies and their objective was slightly different: determining breed composition instead of assigning animals to a breed.

Finally, the results showed that it would be possible to use the studied breed assignment test in routine as one step of a certification process of local products, especially if a SNP chip containing all the selected SNPs is used.

Acknowledgements

Support of the Walloon Government and the INTERREG program are acknowledged. H. Wilmot, as a research fellow, acknowledges the support of the FNRS (Brussels, Belgium).

References

BlueSter. (2021). BlueSter. Available at: https://www.projet-bluester.eu/

Josse J., and Husson F. (2012). J. Soc. Fr. Stat., 153(2), 77–99.

Judge M. M., Kelleher M. M., Kearney J. F., Sleator R. D., and Berry D. P. (2017). Anim., 11(6), 938–947. https://doi.org/10.1017/S1751731116002457

Kuehn L. A., Keele J. W., Bennett G. L., McDaneld T. G., Smith T. P. L., *et al.* (2011). J. Anim. Sci., 89(6), 1742–1750. https://doi.org/10.2527/jas.2010-3530

Lê S., Josse F., & Husson F. (2008). J. Stat. Softw., 25(1), 1–18. https://doi.org/10.18637/jss.v025.i01

Márquez G. C., Speidel S. E., Enns R. M., and Garrick D. J. (2010). J. Animal Sci., 88(1), 59–68. https://doi.org/10.2527/jas.2008-1292

Mota R. R., Mayeres P., Bastin C., Glorieux G., Bertozzi C., *et al.* (2017). J. Anim. Sci., 95(10), 4288–4299. https://doi.org/10.2527/jas2017.1748

R Core Team. (2021). R: A language and environment for statistical computing. Available at: http://www.r-project.org/

R Studio Team. (2020). RStudio: Integrated Development for R. RStudio. Available at: http://www.rstudio.com/

Wilmot H., Bormann J., Soyeurt H., Hubin X., Glorieux G., *et al.* (2022). J. Anim. Breed. Genet., 139(1), 40–61. https://doi.org/10.1111/jbg.12643

Part 14

Genetic diversity and inbreeding

204. Assessing existing admixture in indigenous veld ecotypes and non-descript goat populations of Southern Africa

C. Visser* and S.F. Lashmar

University of Pretoria, Private Bag X20, 0028, Hatfield, Pretoria, South Africa; carina.visser@up.ac.za

Abstract

Indigenous goat populations play an important role in food security and livelihoods of resource-poor communities in Southern Africa. These adapted animal resources could contribute significantly to future food production strategies, but limited research has been performed on their genomic architecture. A total of 192 indigenous goats, comprising of three recognized ecotypes (n=92), as well as five non-descript, communal populations (n=100) were genotyped using the Illumina® Goat SNP50 version 2 single nucleotide polymorphism (SNP) genotyping panel. Genome-level genetic relatedness analyses indicated that population differentiation was driven by both genetic distinctness and geographic isolation. Some of the communal individuals clustered with previously recognized ecotypes, while others indicated unique and distinct ancestry. Elevated inbreeding levels could probably be ascribed to artificial selection in commercial ecotypes, and small communal population sizes. A conservation through utilization approach could benefit the genetic diversity of these populations.

Introduction

Approximately 70% of Southern African goats are kept in non-commercial production systems and contribute to the nutritional and financial security of resource-poor households and communities (Monau *et al.*, 2020). Goats farmed in smallholder and communal farming systems are typically kept in small herds of less than 20 animals and are often confronted with environmental stressors and nutritional inadequacies, with little to no management inputs or supplementation (Chokoe *et al.*, 2020; Monau *et al.*, 2020). The endurance of these goats within different agroecological zones, and their subsequent development into uniquely adapted breeds or ecotypes, is believed to be a result of natural (rather than artificial) selection pressures. As a result of their exposure to these pressures, indigenous goats have acquired anatomical and physiological adaptive mechanisms to tolerate heat and drought, resist parasites (external and internal) and reduce their maintenance requirements.

Indigenous goat ecotypes and non-descript populations represent an important genomic reservoir for local chevon production and have served as a foundation for the development of commercial breeds, largely due to their adaptive traits (Mdlala *et al.*, 2016). The Indigenous Veld Goat (IVG; https://www.indigenousveldgoats.co.za/) society was recently established to improve pedigree and performance recording efforts for recognized ecotypes that are distinguished based on their morphological features (e.g. Nguni, Cape Lob Ear, Cape Speckled). Communally farmed goat populations are, however, for the most part non-descript with an increased number of community-based breeding programs (CBBP) being established to facilitate improved productivity and profitability (Mtshali *et al.*, 2021). The relevance of the breed concept is increasingly being brought into question, specifically for locally adapted goat populations, and there is a tendency towards advocating conservation through utilization of existing admixtures that were historically introduced (e.g. do Prado Paim *et al.*, 2019).

In order to capitalize on their adaptive traits for improved sustainability of future goat production strategies, it is necessary to profile the existing population structure and elucidate between-population relationships of local goat genetic resources. The aim of this research was to report on the genetic diversity and population structure of recognized indigenous goat ecotypes and non-descript populations occurring in Southern Africa.

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_204, © C. Visser and S.F. Lashmar 2022

Materials & methods

Ethical approval was granted by the Ethics Committee of the Faculty of Natural and Agricultural Sciences, University of Pretoria, for the use of the genotypes in this study (Ethics number: NAS313/2021). Single nucleotide polymorphism (SNP) data was available for 192 indigenous southern African goats comprised of three recognized ecotypes namely the Cape Speckled (n=30), Nguni (n=20) and Xhosa Lob Ear (n=42), as well as five non-descript, communal populations (Cedara, Jozini, Mboza, Msinga and OSCA; n=20 each). Within the Cape Speckled and Xhosa Lob Ear populations eight and ten animals, respectively, originated from Namibia (NM); the remaining samples originated from South Africa (SA), with the communal populations all originating from different regions in the KwaZulu-Natal (KZN) province. Genotyping was conducted using the Illumina® Goat SNP50 version 2 genotyping panel (59,727 SNPs), and data was subjected to quality control procedures in PLINK software (Purcell *et al.*, 2007). Mapped, autosomal SNPs were removed based on low call rate (<90%), low minor allele frequency (MAF<5%) and violation of Hardy-Weinberg Equilibrium (HWE; $P<0.001$). Linkage disequilibrium-based filtering was additionally applied (LD; $r^2>0.2$), where after a final set of 26,792 independent SNPs were retained.

Genome-level genetic relatedness between individual goats was estimated by means of principal component analysis (PCA) derived eigenvectors calculated using GCTA software (Yang *et al.*, 2011). A model-based clustering algorithm, applied using default ADMIXTURE 1.23 software (Alexander *et al.*, 2009) settings, was used to determine the most likely number of ancestral contributors (K-value) for each population. The K-value with the lowest cross-validation (CV) error was chosen as the most probable one, and population structure bar plots were visualized using Genesis version 0.2.3 software (Buchmann and Hazelhurst, 2014). The inbreeding levels of the various populations were quantified using PLINK's SNP-based F-statistic (F_{SNP}).

Results & discussion

The within and between population differences observed from the PCA analysis was seemingly driven by two factors namely genetic distinctness and geographic isolation (Figure 1). The three recognized ecotypes i.e. the Nguni (Mbuzi), Speckled and Xhosa types formed separate clusters, albeit dispersed. The Cedara and Msinga populations, which were previously described as non-descript, clustered with the Nguni type suggesting that these animals might belong to this ecotype. Interestingly the Jozini, Mboza and OSCA populations formed a distinct cluster from the recognized ecotypes, indicating possible association with a different ecotype. The sampling locations supported geographic isolation as a contributor to genetic distinctness as the Jozini and Mboza populations are located in northern parts of the KwaZulu-Natal (KZN) province, whereas the OSCA population is more centrally located and the Cedara and Msinga populations are found in the south of the province. An example of within-ecotype genetic isolation was observed for the Speckled goats; the Namibian samples subclustered, indicating a greater genetic distance from South African samples. The dispersed patterns of clustering are consistent with Chokoe *et al.* (2020) that genetically characterized communal, indigenous goat populations in other provinces of South Africa.

When ADMIXTURE's model-based algorithm was applied, a K-value of three (CV error=0.618) separated the recognized Speckled and Xhosa Lob Ear ecotypes from the remaining goats, which displayed varying proportions of a third predominant ancestor. The most likely number of inferred ancestral populations (K=6; CV error=0.613) revealed unique ancestry for the Namibian Speckled goats (mean proportion shared=0.900±0.086), different from that of SA Speckled goats. This observation was in concordance with an African-wide trend identified by Colli *et al.* (2018) that goat genepools are more associated with geographical areas than single populations. Similarly, the OSCA population that is geographically separated from populations in the north (Jozini and Mboza) and south (Cedara and Msinga) of KZN, largely derived its genetic composition from a completely unique ancestor (mean proportion shared=0.740±0.239). The Jozini and Mboza populations, which are in close proximity to one another, shared 0.487 and 0.821 proportion

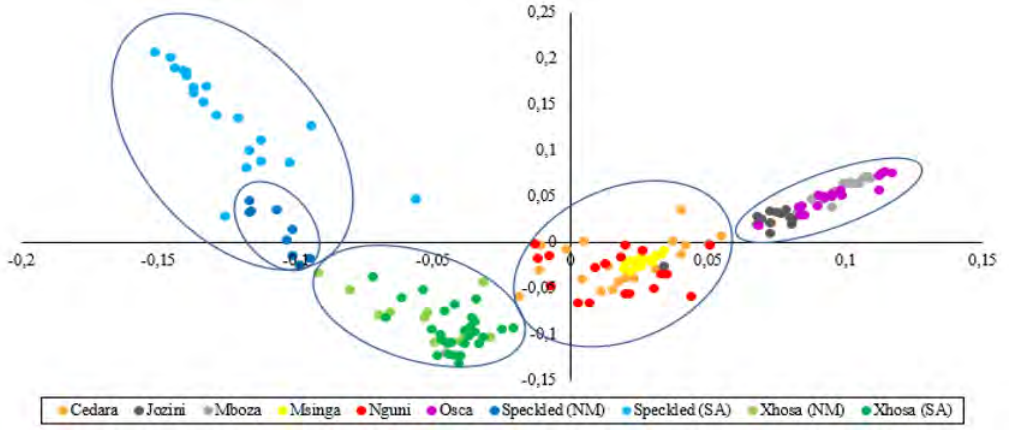


Figure 1. Principal component analysis (PCA) of genetic relatedness between indigenous South African goat populations. NM=Namibian; SA=South African.

units of a distinct ancestral population. One predefined Mboza animal, shared 90% of its ancestry with the Xhosa Lob Ear population, and this may have resulted from the tendency in the communal sector to purchase live animals and resell them in other areas (Gwaze *et al.*, 2008).

The SNP-based inbreeding coefficients (F_{SNP}) ranged from -0.030 (Nguni) to 0.146 (Speckled). In the case of the Speckled goat, a higher level of inbreeding may have resulted from its introduction to breed improvement, and hence pedigree- and performance-based selection (Indigenous Veld Goats, 2018). Higher levels of inbreeding in communal populations e.g. the Mboza (F_{SNP}=0.140) and OSCA (F_{SNP}=0.079) may be due to small census and effective population sizes, and may pose a risk to the existing genetic diversity.

Conclusions

Results from the present study indicated genome-level relatedness of communal goats previously described as non-descript with recognized goat ecotypes. The PCA and model-based clustering analyses additionally

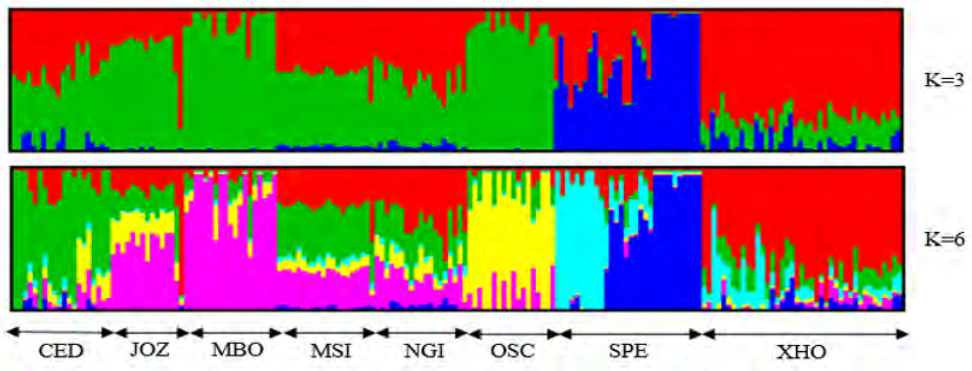


Figure 2. Model-based clustering of indigenous southern African goat populations. CED=Cedara, JOZ=Jozini, MBO=Mboza, MSI=Msinga, NGI=Nguni, OSC=OSCA, SPE=Speckled, XHO=Xhosa.

indicated unique and distinct ancestry, not common to existing, recognized ecotypes which may convey the presence of previously unrecognized ecotypes. It is recommended that existing admixture be conserved for future utilization, as some ancestral genomic segments may contribute to the expression of adaptive traits.

References

Alexander, D.H., Novembre, J. and Lange, K. (2009) Fast model-based estimation of ancestry in unrelated individuals. Genome Res. 19(9):1655-1664. http://www.genome.org/cgi/doi/10.1101/gr.094052.109

Buchmann, R. and Hazelhurst, S. (2014) Genesis manual. Johannesburg: University of the Witwatersrand. Available at: http://www.bioinf.wits.ac.za/software/genesis.old/.

Chokoe, T.C., Mdladla-Hadebe, K., Muchadeyi, F., Dzomba, E., Matelele, T., *et al.* (2020) Genetic Diversity of South African Indigenous Goat Population from Four Provinces Using Genome-Wide SNP Data. Sustainability 12(24):10361. https://doi.org/10.3390/su122410361

Colli, L., Milanesi, M., Talenti, A., Bertolini, F., Chen, M., *et al.* (2018) Genome-wide SNP profiling of worldwide goat populations reveals strong partitioning of diversity and highlights post-domestication migration routes. Genet. Sel. Evol. 50(1):1-20. https://doi.org/10.1186/s12711-018-0422-x

do Prado Paim, T., Faria, D.A., McManus, C., Lanari, M.R., Esquivel, L.C., *et al.* (2019) New world goat populations are a genetically diverse reservoir for future use. Sci. Rep. 9(1):1-12. https://doi.org/10.1038/s41598-019-38812-3

Gwaze, F.R., Chimonyo, M. and Dzama, K. (2009) Communal goat production in Southern Africa: a review. Trop. Anim. Health Prod. 41(7):1157-1168. https://doi.org/10.1007/s11250-008-9296-1

Indigenous Veld Goats (IGV) 2018. Available at: https://www.indigenousveldgoats.co.za/.

Mdladla, K., Dzomba, E.F., Huson, H.J. and Muchadeyi, F.C. (2016) Population genomic structure and linkage disequilibrium analysis of South African goat breeds using genome-wide SNP data. Anim. Genet. 47(4):471-482. https://doi.org/10.1111/age.12442

Monau, P., Raphaka, K., Zvinorova-Chimboza, P. and Gondwe, T. (2020) Sustainable utilization of indigenous goats in Southern Africa. Diversity 12(1):20. https://doi.org/10.3390/d12010020

Mtshali, T.F., Mapholi, O.N., Ncube, K.T., Dzomba, E.F., Matelele, T.C., *et al.* (2021) Goat farmers production objectives and trait preferences in the North West province of South Africa: An approach to identify selection criteria for community-based breeding program. Int. J. Livest. Prod. 12(2):64-75. https://doi.org/10.5897/IJLP2020.0743

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A.R., *et al.* (2007) PLINK: a tool set for whole-genome association and population-based linkage analysis. Am. J. Hum. Genet. 81:559-575. https://doi.org/10.1086/519795

Yang, J., Lee, S.H., Goddard, M.E. and Visscher, P.M. (2011) GCTA: a tool for genome-wide complex trait analysis. Am. J. Hum. Genet. 88(1):76-82. https://doi.org/10.1016/j.ajhg.2010.11.011

205. Estimation of sex chromosome inbreeding depression on milk production in cattle

I. Curik[1*], H. Vostra-Vydrova[2], M. Shihabi[1], J. Sölkner[3] and L. Vostry[2]

[1]University of Zagreb – Faculty of Agriculture, Svetošimunska 25, 10000 Zagreb, Croatia; [2]Czech University of Life Science Prague, Kamycka 129, 16500 Prague, Czech Republic; [3]University of Natural Resources & Life Sciences Vienna, Gregor-Mendel-Strasse 33, 1180 Vienna, Austria; icurik@agr.hr

Abstract

Inbreeding depression in dairy production is an extensively studied phenomenon. However, we are not aware of a single study that has estimated inbreeding depression concerning sex-chromosome. Using Bayesian inference, we analyzed inbreeding depression in milk yield (4,749 Czech Holstein cows genotyped for >40,000 SNPs with milk yield in the first lactation) in terms of the combined effect of inbreeding (per 1% increase) obtained from autosomes and sex-chromosome. We used three models, each based on different estimated inbreeding coefficients (pedigree, ROH>4Mb, or ROH>8Mb). Significant inbreeding depression estimates for autosomes were observed in all three models (medians; bF_{PED}=-13.2 kg, $bF_{ROH>4Mb}$=-30.0, and $bF_{ROH>8Mb}$=-34.4 kg), while significant inbreeding depression was observed only for pedigree sex-chromosome estimates (medians; $bF_{PED\text{-}sex}$=-7.6 kg, $bF_{ROH>4Mb\text{-}sex}$=-0.1 kg, and $bF_{ROH>8Mb\text{-}sex}$=3.9 kg). Low genomic correlations between inbreeding estimated from autosomes and sex-chromosome argue for the need to estimate inbreeding depression from sex-chromosome.

Introduction

Inbreeding depression is defined as a reduction in the population mean for quantitative traits such as reproduction, fecundity, survival, and other traits (Leroy 2014; Doekes *et al.*, 2021). In genetically small populations, inbreeding and consequently inbreeding depression cannot be avoided, especially in populations subject to strong artificial selection, as is the case with dairy cattle populations. Inbreeding depression estimated from pedigree or genomic data (Curik *et al.*, 2017) is a well-documented phenomenon in dairy cattle (Miglior *et al.*, 1995; Pryce *et al.*, 2014; Doekes *et al.*, 2019). However, we are not aware of a single study that estimated inbreeding depression in relation to sex chromosome, either from inbreeding estimated from pedigree or inbreeding estimated from genomic data.

The main objective of this study was to estimate the joint effect of inbreeding (per 1% increase) of sex and autosomal chromosomes on milk yield in the Czech Holstein population.

Materials & methods

Animals, phenotypes, pedigree, and genotypes. Milk yield data were analysed for a total of 4,749 first lactation cows that calved during 2018-2021 (first lactation only). The total pedigree included 54,196 individuals and is of good quality with a mean complete generation equivalent of 12.3 (range 8.6 to 14.2). Most cows (4,478) were genotyped with the Illumina BovineSNP50 BeadChip (versions v3_A1), while 205 and 67 individuals were genotyped with the CRV-developed 60 k Illumina Panel (versions v2) and the Illumina EuroGMD Chip (EuroGenomics), respectively. Our analysis was based only on SNP with known positions on autosomes and sex chromosomes. Quality control was performed using PLINK v1.9 software (Chang *et al.*, 2015), and we removed all SNPs with a call rate <90% and SNPs that deviated from Hardy-Weinberg equilibrium with $P<10^{-7}$. After removing individuals with a call rate <95%, we estimated ROH-based inbreeding coefficients for 4,749 cows with different SNP numbers. Milk data, pedigrees, and genotypes were kindly provided by the Holstein Cattle Breeders Association of The Czech Republic.

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_205, © I. Curik *et al.* 2022

Calculation of relationship matrices and inbreeding coefficients. The pedigree-based additive relationship matrices, autosome A (Tier, 1990) and sex-chromosome A-sex (Fernando and Grossman, 1990), were estimated using our code. In contrast, the related inbreeding coefficient F_{PED} and $F_{PED\text{-}sex}$ were calculated from the diagonal of the relationship matrices (diagonal minus one). The genomic relationship matrices for autosomes (G) and sex-chromosome (G_X) were calculated using the first model proposed by VanRaden (2008). All ROH-based inbreeding coefficients (FROH>4Mb, FROH>8Mb, FROH>4Mb-sex, and FROH>8Mb-sex) were calculated using the approach described in McQuillan *et al.* (2008) and Curik *et al.* (2014). ROHs were estimated for each individual with the R package 'detectRUNS' (Biscarini *et al.*, 2019) using the consecutive method with the following parameters: maxgap = 2 Mb, minSNP = 15, one heterozygote and one missing allowed.

Estimation of inbreeding depression. Inbreeding depression was estimated by the following model:

$$y_{Milk} = HY_i + Month_j + b_1 Age + b_2 F_X + b_3 F_X\text{-sex} + cow_X + cow_{X\text{-}Sex} + e, \quad (1)$$

where HY is a fixed effect of Herd by Year of birth; Month is a fixed effect of birth month; b_1Age is a regression coefficient for cow age; b_2F_X is a regression coefficient for autosomal inbreeding coefficient (X is F_{PED}, $F_{ROH>4Mb}$ or $F_{ROH>8Mb}$); b_3F_X-sex is a regression coefficient for the sex-chromosome inbreeding coefficient (X is $F_{PED\text{-}sex}$, $F_{ROH>4Mb\text{-}sex}$ or $F_{ROH>8Mb\text{-}sex}$; cow_X is a random effect of cow for autosomes (X is autosomal relationship matrix, additive A or genomic G); $cow_{X\text{-}sex}$ is a random effect of cow for sex-chromosome (X is sex relationship matrix, additive A-sex or genomic G-sex); and e is a random residual effect. Our estimation followed the Bayesian inference approach using gibbs2f90 software (Misztal, 2003) with two parallel Gibbs sampling chains with 250,000 iterations, as described in Farkas *et al.* (2007). The convergence of the Gibbs sampling chains was analysed using the ARIMA procedures of the SAS software and additionally examined using the Z-criterion of the Geweke analysis (Geweke, 1992). The first 50,000 iterations (burn-in period) were shifted and only every 150th iteration (tinning) was sampled to reduce dependence between successive iterations. Descriptive statistics and graphical representations of posterior distributions were obtained using the tidyverse package collection (Wickham *et al.*, 2019) from Rstudio (R Core Team, 2019; RStudio Team, 2020).

Results & discussion

Descriptive statistics for milk yield and various inbreeding estimates are provided in Table 1. In contrast to the ROH-based estimates, where much higher estimates were observed for the sex-chromosome, lower estimated inbreeding was observed for the pedigree sex-chromosome. Variation in sex chromosomal inbreeding was always more variable than variation in autosomal inbreeding (Table 1).

The correlation of 0.534 was observed between pedigree inbreeding coefficients (F_{PED} and $F_{PED\text{-}sex}$), while very low correlations were observed between ROH-based inbreeding coefficients, 0.08 between $F_{ROH>4Mb}$ and $F_{ROH>4Mb\text{-}sex}$ and 0.05 between $F_{ROH>8Mb}$ and $F_{ROH>8Mb\text{-}sex}$. Such low correlations suggest that the effects of genomic inbreeding estimated from autosomes do not account for the deleterious load that may be present on the sex chromosome.

The distributions of estimates of inbreeding depression obtained by three models, each based on differently defined inbreeding coefficients, using the Bayesian inference approach are shown in Figure 1.

Table 1. Descriptive statistics of the milk production, inbreeding coefficients and pedigree quality.

Variable	Median	Mean	Range	SD[1]
Milk yield (kg)	9,902	9,925	4,852-15,315	1,768
F_{PED} (%)	6.1	6.3	0.9-23.1	1.8
$F_{ROH>4Mb}$ (%)	10.5	10.7	0.2-28.7	3.0
$F_{ROH>8Mb}$ (%)	6.8	7.1	0.0-25.5	2.7
$F_{PED-sex}$ (%)	4.2	5.2	0.2-29.8	3.5
$F_{ROH>4Mb-sex}$ (%)	20.0	22.0	0.0-78.6	10.7
$F_{ROH>8Mb-sex}$ (%)	6.3	9.2	0.0-74.2	11.8

[1] SD represents standard deviation.

[2] CGE is complete generation equivalent.

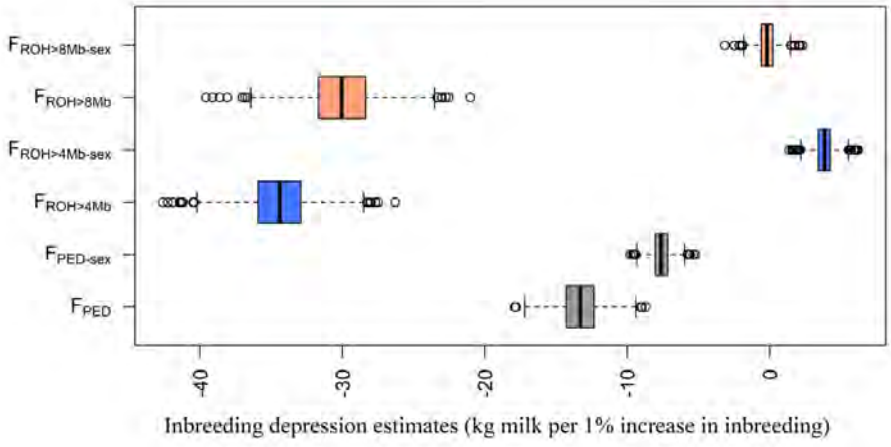


Figure 1. Box-plot distribution of estimates of inbreeding depression (Bayesian inference) resulting from three models, each based on differently defined inbreeding coefficients: (1) pedigree (F_{PED} and $F_{PED-sex}$); (2) ROH>4Mb ($F_{ROH>4Mb}$ and $F_{ROH>4Mb-sex}$); and (3) ROH>8 Mb ($F_{ROH>8Mb}$ and $F_{ROH>8Mb-sex}$).

According to Figure 1, significant inbreeding depression estimates for autosomes were observed in all three models (medians; bF_{PED}=-13.2 kg, $bF_{ROH>4Mb}$=-30.0, and $bF_{ROH>8Mb}$=-34.4 kg), while significant inbreeding depression was observed only for pedigree sex-chromosome estimates (medians; $bF_{PED-sex}$=-7.6 kg, $bF_{ROH>4Mb-sex}$=-0.1 kg, and $bF_{ROH>8Mb-sex}$=3.9 kg).

We were unable to explain the difference between pedigree and genomic estimates of inbreeding depression for the X chromosome. At the same time, little is known about the effects of inbreeding on X chromosome inactivation. Our knowledge of the interplay of inbreeding, selection, and accumulation of deleterious load on the X chromosome in animal populations is also insufficient. However, to our knowledge, this is the first estimate of inbreeding depression with respect to the X chromosome in livestock populations, and further research is needed to properly assess the potential negative consequences of X chromosome inbreeding in dairy cows.

Acknowledgements

This study was supported by the ANAGRAMS-IP-2018-01-8708, ALGSEQ18-IP-2018-01-7317, Project No. QK1810253 and Project No. LTAUSA19117.

References

Biscarini F., Cozzi P., Gaspa G., and Marras G. (2019) detectRUNS: DetectRuns of Homozygosity and Runs of Heterozygosity in Diploid Genomes. R package version 0.9.6. Available at: https://CRAN.R-project.org/package=detectRUNS

Chang C.C., Chow C.C., Tellier L.C., Vattikuti S., *et al.* (2015) Gigascience 4(1): s13742-015. https://doi.org/10.1186/s13742-015-0047-8

Curik I., Ferenčaković M., and Sölkner J. (2014) Livestock Science 166: 26-34. https://doi.org/10.1016/j.livsci.2014.05.034

Curik I., Ferenčaković M., and Sölkner J. (2017) Revista Brasileira de Zootecnia 46: 773-782. https://doi.org/10.1590/S1806-92902017000900010

Doekes H.P., Veerkamp R.F., Bijma P., de Jong G., *et al.* (2019) Genetics Selection Evolution 51(1): 1-16. https://doi.org/10.1186/s12711-019-0497-z

Doekes H.P., Bijma P., and Windig J.J. (2021) Genes 12(6): 926. https://doi.org/10.3390/genes12060926

Farkas J., Curik I., Csató L., Csörnyei, Z., *et al.* (2007) Livestock Science 112(1-2): 109-114. https://doi.org/10.1016/j.livsci.2007.01.160

Fernando R.L., and Grossman M. (1990) Theoretical and applied genetics 80(1): 75-80. https://doi.org/10.1007/BF00224018

Geweke J. (1992). Bayesian statistics 4: 641-649.

Leroy G. (2014) Animal genetics 45(5): 618-628. https://doi.org/10.1111/age.12178

McQuillan R., Leutenegger A., Abdel-Rahman R., Franklin C., *et al.* (2008) American Journal of Human Genetics 83(3): 359–372. https://doi.org/10.1016/j.ajhg.2008.08.007

Miglior F., Burnside E.B., and Kennedy B.W. (1995) Journal of Dairy Science 78(5): 1174-1180. https://doi.org/10.3168/jds.S0022-0302(95)76735-2

Misztal I., (2003) BLUPF90 Manual.

Available at http://nce.ads.uga.edu/~ignacy/newprograms.html

Pryce J.E., Haile-Mariam M., Goddard M.E., and Hayes B.J. (2014) Genetics Selection Evolution 46(1): 1-14. https://doi.org/10.1186/s12711-014-0071-7

RStudio Team (2020) RStudio: Integrated Development for R. RStudio, PBC, Boston, MA. Available at:http://www.rstudio.com/

R Core Team (2019) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Available at: https://www.R-project.org/

Tier B. (1990) Genetics Selection Evolution 22: 419–430. https://doi.org/10.1186/1297-9686-22-4-419

VanRaden P.M. (2008) Journal of dairy science 91(11): 4414-4423. https://doi.org/10.3168/jds.2007-0980

Wickham H., Averick M., Bryan J., Chang W. *et al.* (2019) Journal of Open Source Software 4(43):1686. https://doi.org/10.21105/joss.01686

206. Assessing genomic diversity and heterozygosity-rich regions of transboundary Nguni cattle of South Africa and eSwatini

E. van Marle-Köster[1*], S.F. Lashmar[1], M. Okpeku[2], F.C. Muchadeyi[3], N.O. Mapholi[4] and C. Visser[1]

[1]University of Pretoria, Department of Animal Science, Hatfield, 0028, South Africa; [2]University of KwaZulu Natal, Discipline of Genetics, School of Life Sciences, Westville, 3629, South Africa; [3]Agricultural Research Council, Biotechnology Platform, Onderstepoort, 0110, South Africa; [4]University of South Africa, Department of Agriculture and Animal Health, Florida, 1709, South Africa; este.vanmarle-koster@up.ac.za

Abstract

The genetic integrity of southern African Nguni ecotypes is under potential threat due to both indiscriminate crossbreeding and uncontrolled inbreeding. This study aimed to assess the transboundary genetic diversity and heterozygosity-rich regions of 192 South African (half stud, half research herds) and 96 eSwatini Nguni cattle. A set of 40,930 common single nucleotide polymorphisms (SNPs) were used to study their population structure and runs of heterozygosity (ROHet) profiles. Principal component analysis (PCA) and population structure (K=5) analyses illustrated genomic distinctiveness between South African (SA) and eSwatini (ES) populations, greater admixture for SA research (shared ancestry=0.631±0.353) compared to SA stud (shared ancestry=0.741±0.123), and three subpopulations for ES. The largest proportion (0.650) of the 935 ROHet segments fell within the 0.5-1 Mb length category. Overall, results illustrated that genetic distinctiveness resulted from geographic isolation and production strategies. Although no impending threat to genetic diversity was observed, further loss should be monitored to prevent possible ecotype endangerment.

Introduction

The Nguni cattle breed is classified as an indigenous Sanga breed (*Bos taurus africanus*) and a transboundary animal genetic resource that manifests unique adaptation abilities across distinct agroecological zones. Nguni cattle are known for their relatively low maintenance requirements (Musemwa *et al.*, 2010), longevity and high calving rates (Matjuda *et al.*, 2014), heat tolerance (Katiyatiya *et al.*, 2017) as well as resistance to ticks and tick-borne diseases (Mapholi *et al.*, 2014), which deem them adaptable to the diverse production environments characteristic of southern Africa. Accompanying the settlement of several ethnic groups in different geographic and climatic regions of southern Africa, distinct Sanga cattle ecotypes, phenotypically distinguishable by coat colour variations, developed over time (Bester *et al.*, 2003) which include the Shangaan, Pedi and Nkone (van Marle-Köster *et al.*, 2021).

In the early 1930's, South African (SA) Nguni cattle populations were classified as non-descript and an official breed society was only established in 1986. The SA Nguni is most popular in the regions with higher temperature and humidity in both commercial and smallholder production systems but has a relatively short history of animal recording and objective selection (van Marle-Köster *et al.*, 2021). In eSwatini (ES), Nguni are found in all six ecological regions of the country where they play a significant role in the economy of rural communities. The mandate of the four existing conservation and breeding stations changed in 1975 from conservation alone to a combined focus on conservation and improvement for beef cattle production. The improvement program consists of inter-crossing of the six different indigenous lines for multiplication purposes and up-grading of selected populations with exotic breeds that include the Angus and Brahman (Food and Agriculture Organization, 2004).

Uncontrolled breeding and the approved influx of exotic genetic material for crossbreeding and upgrading of indigenous populations is considered a threat to adapted local breeds, such as the Nguni. This study

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_206, © E. van Marle-Köster *et al.* 2022

aimed to assess the genetic diversity and heterozygosity-rich regions that exists both across countries (ES: eSwatini; SA: South Africa) and within-country (SA), between purebred stud animals (SA-S) and research herds (SA-R).

Materials & methods

Ethical approval was granted by the Ethics Committee of the Faculty of Natural and Agricultural Sciences, University of Pretoria, for the use of external data (EC-180000097) and written consent was provided by the University of eSwatini's UNESWA Research Board. Subsets of 96 SA purebred stud animals (SA-S), 96 SA animals from research stations (SA-R) and 96 eSwatini cattle from three governmental breeding stations (ES) were studied. Deoxyribonucleic acid (DNA) was extracted from hair (ES, and SA-S) and/or blood samples (SA-R). Post-quality control, 40,930 common, autosomal SNPs were retained after the exclusion of non-autosomal, unmapped and duplicated SNPs. Using PLINK software (Purcell *et al.*, 2007) exclusion of samples and SNPs was based on sample call rate (<90%), SNP call rate (<95%), low minor allele frequency (MAF<1%) and Hardy-Weinberg Equilibrium (HWE; P<0.001). Liberal linkage disequilibrium (LD) based filtering was additionally employed (r^2>0.5) before genetic relatedness and population structure analyses.

To assess the genetic relatedness, a genomic relationship matrix (GRM) was constructed using GCTA software (Yang *et al.*, 2011) and used in principal component analysis (PCA) to estimate eigenvectors. The cross-validation (CV) procedure was used to identify the most likely number of ancestral populations (K) in ADMIXTURE software (Alexander *et al.*, 2009). The K-value producing the lowest CV error was considered ideal and visualized with GENESIS software (Buchmann and Hazelhurst, 2014). Heterozygosity-rich segments were identified using the R package *detectRUNS* (Biscarini *et al.*, 2019) by executing the consecutive-SNP-based detection method (Marras *et al.*, 2015). A minimum of 10 SNPs and length of 10 kb was required to constitute a segment. The number of opposing (homozygous) and missing genotypes allowed were zero and two, respectively. In contrast to what is suggested by Meyermans *et al.* (2020) for ROH (i.e. no MAF filtering), MAF filtering was applied prior to runs of heterozygosity (ROHet) analyses to prevent underestimation of segment lengths as low-MAF SNPs tend to have homozygous genotypes, which may partition longer ROHet segments (i.e. more but smaller ROHet are identified). No LD filtering was applied prior to ROHet analysis.

Results & discussion

The PCA analysis (Figure 1a) indicated a clear separation of ES and SA populations, with more dispersion observed within the ES population than the overall SA population, and is consistent with a longer history of methodical within-breed selection for SA overall (since 1986) compared to a longer history of indiscriminate crossbreeding with exotic breeds for ES (since 1975). The tighter cluster observed for SA-S than SA-R supports more intense exposure to artificial selection and the utilization of high-impact animals across herds. Population structure results (Figure 1b) supported the clustering patterns depicted in the PCA. The ideal number of ancestral populations (K) estimated, with the lowest cross-validation error (CV=0.499), for the merged Nguni data set was five. The genomic composition of the SA Nguni population was largely derived from two ancestral populations, with SA-S being more uniform (mean proportion shared=0.741) and SA-R displaying a higher degree of admixture (mean±standard deviation proportion shared=0.631±0.353). The three distinct ancestral contributors observed for ES (shared genetics=0.404 for yellow, 0.249 for pink and 0.286 for green) supported the existence of three separate breeding lines (consistent with sampling), however, the degree of admixture illustrated increased crossbreeding across lines and individual ecotypes could be at risk of genetic erosion.

Runs of heterozygosity (ROHet) are genomic regions where diversity might be beneficial (e.g. adaptive traits) and may indicate balancing selection events (Biscarini *et al.*, 2020). The sum totals of ROHet

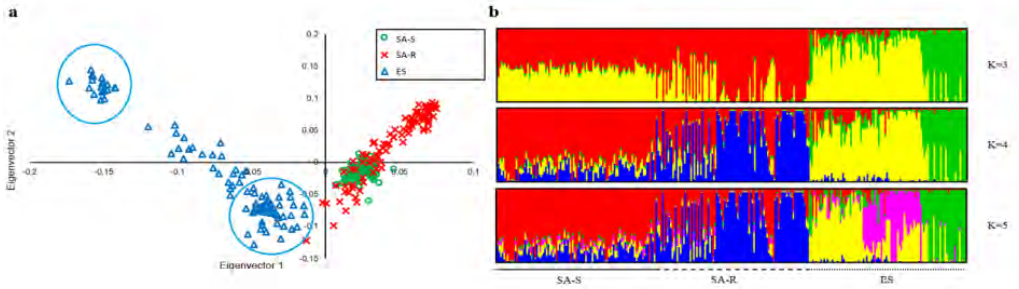


Figure 1. Genetic structure of southern African Nguni cattle, according to principal component analysis (a), and model-based clustering (b).

segments were 343, 309 and 283 for the ES, SA-R and SA-S populations, respectively, which was within the range of per-population numbers observed in other studies (e.g. 2 for Creolo from Guadalupe to 1,702 for Montana Tropical composite; Mulim *et al.*, 2021). The mean ROHet segment length was 0.63 megabase pairs (Mb) across the three populations. Albeit complex to compare (due to differing criteria for defining ROHet segments), the mean length was similar to a mean ROHet length of 0.70 Mb reported by Biscarini *et al.* (2020) for semi-feral Maremanna cattle. Similar to the afore-mentioned authors, and depicted in Figure 2, the 0.5-1 Mb length category ranked the highest in terms of mean proportion of ROHet segments (0.650 across populations). The most prevalent consensus ROHet segment was identified only in the SA-R population and encompassed 26 *Ensembl* (https://www.ensembl.org/index.html) protein-coding genes assembled to the ARS-UCD 1.2 cattle assembly. The 0.35 Mb ROHet region on BTA18 (base pair position range: 34,817,149-35,163,513), present in 19.5% of the SA-R animals, includes genes that may play a role in cilium biogenesis (e.g. *ATP6V0D1*), cellular response to starvation (e.g. *RIPOR1*) and phagocytosis (e.g. *ELMO3*). Identifying regions of conserved genomic diversity may be advantageous in characterizing adaptive traits and the identification of shared ROHet segments in regions containing, for example, immune response related genes may facilitate better elucidation of disease resistance, which is one of the proxy-indicators of adaptive ability (Mulim *et al.*, 2021). These results should, however, be interpreted with caution and a larger data set, in terms of sample size and marker density (or whole-genome sequencing), is expected to provide improved resolution.

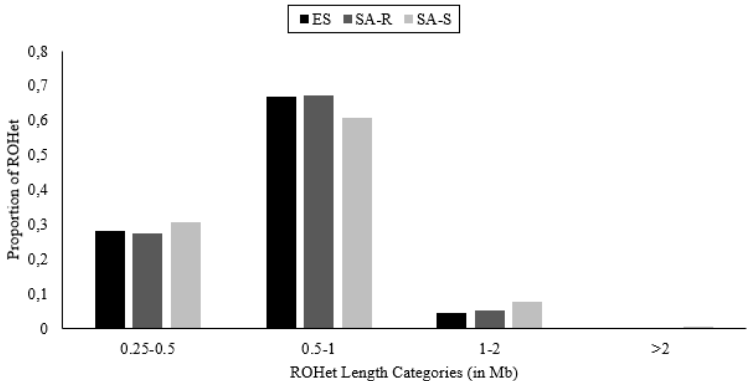


Figure 2. The proportion of heterozygosity-rich segments (ROHet), within different length categories, identified for eSwatini and South African Nguni cattle.

Conclusions

The genomic information proved useful to provide insight into the genomic diversity and heterozygosity-rich regions among the three Nguni populations. The SA and ES populations could be distinguished as separate clusters and require further investigation for potential ecotypes. Breeding strategies for ES populations should be monitored to prevent genetic erosion of subpopulations. Higher resolution genomic profiling (e.g. with whole-genome sequencing information) of a broader spectrum of Nguni populations, in both countries, will provide a more comprehensive picture of the genes included in consensus ROHet region(s). These regions of conserved genome-level variability may pinpoint genes that will assist in better describing the Nguni's adaptive mechanisms (e.g. disease resistance).

References

Bester, J., Matjuda, L.E., Rust, J.M., and Fourie, H.J. (2003) Nguni: a case study. In: Vilakati *et al.* (eds) 2003. Community-based management of animal genetic resources. FAO, Rome, Italy, pp. 45-68.

Biscarini, F., Cozzi, P., Gaspa, G., and Marras, G. (2019), detectRUNS: Detect runs of homozygosity and runs of heterozygosity in diploid genomes. Available at: https://github.com/bioinformatics-ptp/detectRUNS/tree/master/detectRUNS.

Biscarini, F., Mastrangelo, S., Catillo, G., Senczuk, G., and Ciampolini, R. (2020) Anim. 10(12):2285. https://doi.org/10.3390/ani10122285

Buchmann, R., and Hazelhurst, S. (2014), Genesis manual. Available at: http://www.bioinf.wits.ac.za/software/genesis/Genesis.pdf.

Food and Agriculture Organization (FAO), Country report on state of animal genetic resources in Swaziland: A contribution to the first report on the state of the world's animal genetic resources 2004. Available at: http://www.fao.org/3/a1250e/annexes/CountryReports/Swaziland.pdf.

Katiyatiya, C.L.F., Bradley, G., and Muchenje, V. (2017) J. Therm. Biol. 69:85-94. https://doi.org/10.1016/j.jtherbio.2017.06.009

Mapholi, N.O., Marufu, M.C., Maiwashe, A., Banga, C.B., Muchenje, V., *et al.* (2014) Ticks and Tick-Borne Dis. 5:475-483. https://doi.org/10.1016/j.ttbdis.2014.04.006

Marras, G., Gaspa, G., Sorbolini, S., Dimauro, C., Ajmone-Marsan, P., *et al.* (2015) Anim. Genet. 46(2):110-121. https://doi.org/10.1111/age.12259

Matjuda, L.E., MacNeil, M.D., Maiwashe, A., Leesburg, V.R. and Malatje, M. (2014) S. Afr. J. Anim. Sci. 44:161-172. https://doi.org/10.4314/sajas.v44i2.9

Meyermans, R., Gorssen, W., Buys, N., and Janssens, S. (2020) Genomics 21(1):1-14. https://doi.org/10.1186/s12864-020-6463-x

Mulim, H.A., Brito, L.F., Pinto, L.F.B., Ferraz, J.B.S., Grigoletto, L., *et al.* (2021) Res. Square. https://doi.org/10.21203/rs.3.rs-785546/v1

Musemwa, L., Mushunje, A., Chimonyo, M., and Mapiye, C. (2010) J. Sustain. Dev. Africa 12(5):209-226. ISSN: 1520-5509.

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A.R., *et al.* (2007) Am. J. Hum. Genet. 81:559-575. https://doi.org/10.1086/519795

Van Marle-Köster, V., Visser, C., Sealy, J., and Frantz, L. (2021) Sustain. 13(8):4388. https://doi.org/10.3390/su13084388

Yang, J., Lee, S.H., Goddard, M.E., and Visscher, P.M. (2011) Am. J. Hum. Genet. 88:76-82. https://doi.org/10.1016/j.ajhg.2010.11.011

207. Infographics for management of inbreeding and kinship in small populations

M.A. Schoon[1,2*], M. Neuteboom[1,2], S.J. Hiemstra[1,2], C. Danchin-Burge[3], J. Fernández[4] and J.J. Windig[1,2]

[1]Wageningen Livestock Research, P.O. Box 338, 6700 AH Wageningen, the Netherlands; [2]Centre for Genetic Resources the Netherlands, P.O. Box 338, 6700 AH Wageningen, the Netherlands; [3]Institute d'Elevage, 149 Rue Bercy, 75595 Paris 12, France; [4]Departamento de Mejora Genética Animal, INIA, Ctra. de La Coruña, km 7.5, 28040 Madrid, Spain; mira.schoon@wur.nl

Abstract

We developed infographics explaining inbreeding and genetic management for breeders, breeding organisations and the general public. Infographics are illustrations combined with text to obtain clear communication and understanding of concepts. The infographics consisted of three parts: (1) what is inbreeding and what are the consequences; (2) how to manage inbreeding for animal breeders and owners; (3) how to manage inbreeding for breeding organizations. The infographics can be downloaded for free and are available in several languages.

Introduction

Genetic diversity is vital for animal populations to respond to selection and adapt to changing environments. Many breeds, however, have become small in number, with the associated risk of loss of genetic diversity and a parallel increase in inbreeding. This implies potential detrimental consequences such as inbreeding depression and expression of genetic defects. Many breeds, especially in Northern Europe, are managed by small holders and hobby breeders. In these cases, their breeding organisations are often run by volunteers without a professional education in animal breeding. Even though other rare breeds, especially in Southern Europe, are kept at professional farms, inbreeding and kinship are concepts that are difficult to fully grasp. Consequently, there is a great need for information for non-professionals and professionals on how to manage genetic diversity, inbreeding and relationships.

Genetic management is riddled with pitfalls and misunderstanding. For example, many breeders do not realise that inbreeding is not heritable and that a ban on breeding with inbred animals is, therefore, unnecessary. At the same time it can be a challenge to explain why breeders should not all go for a superior sire in terms of the breed standard, top production or championships. These superior sires tend to be highly related which is a threat for the breed in the long term. On the other hand, an explanation on the use of the mean kinship (MK) of an animal with all other animals in the breed can be an eye-opener, leading to better management and maintenance of genetic diversity in the long term.

The theory of what inbreeding is and its relation to genetic diversity, the occurrence of genetic defects and inbreeding depression is well developed in population genetics (Falconer and Mackay 1996, Frankham *et al.* 2010, Lynch and Walsh 1998). Moreover, guidelines have been developed on how to manage populations (e.g. FAO 2013). This, however, is a vast body of material directed towards professional researchers and breeders. A comprehensive summary targeted to a broad audience is needed to reach practitioners and policy makers. To do so, we have developed some infographics that explain principles and practices with colourful graphics that are attractive to read.

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_207, © M.A. Schoon *et al.* 2022

Development of the infographics

Infographics are visuals (such as illustrations) combined with text to obtain clear communication on complicated matters. The infographics were developed in cooperation with a professional company specialised in visual communication. The aim of infographics is to inspire the viewer and reader to gain new insights and draw practical conclusions. In order to do so, good infographics take into account how the human brain works to distinguish and to group objects guided by their colour and by the background. The general reading direction in Western culture from left to right and top to bottom is important as well for the design of infographics.

The set of infographics were developed in a process that combined the expertise of the population geneticists with the designers of the agency. First, to get the designers acquainted with the problem, explanations such as those used by the population geneticists over the years for training students and breeding organisations were presented and discussed. Next, all relevant information on the topic was analysed and filtered to keep only the essential concepts. Next, initial sketches were made, texts added, discussed and adapted and, finally, the design was elaborated in full colour.

The first set of infographics was produced in Dutch and presented in an on-line series of seminars for an audience of animal breeders, breeding organisations and other people interested in breeding and inbreeding. Next, the infographics were translated into English, in consultation with experts from different countries. Currently versions in other languages are in preparation.

Structure of the infographics

The information was split into three infographics (Figure 1) in order to avoid too dense information to better communicate the most important messages towards three different target groups. The first infographic describes the basics of inbreeding and is targeted to anyone interested in inbreeding. The second infographic is targeted to animal owners and breeders and shows how to deal with inbreeding at the level of individual selection and mating. The third infographic is targeted to breeding organisations and population managers and visualizes how to deal with inbreeding and relationship at the population level.

Infographic 1: basics of inbreeding. The basic rule that an animal receives half of its DNA from the mother and the other half from the father is the starting point of this infographic. The definition of inbreeding is given and the fact that inbreeding leads to more homozygosity and less diversity is illustrated as well. Furthermore, the problem of recessive genetic defects is treated and why inbreeding can lead to their expression explained. Finally, examples of traits affected by inbreeding depression and common genetic defects are illustrated.

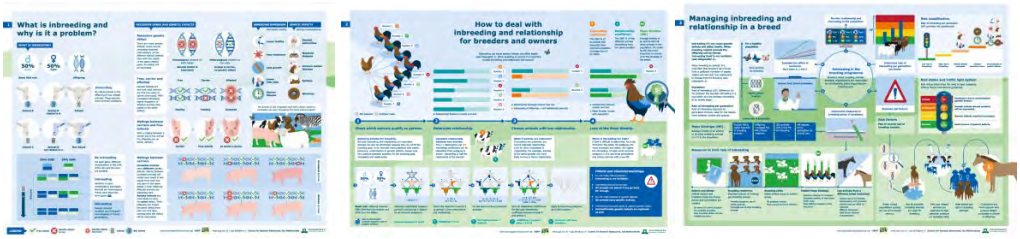


Figure 1. Thumbnails of infographics in English.

Infographic 2: managing inbreeding for breeders and owners. How to choose mates is the main question breeders and owners are faced with, when it comes to managing inbreeding. A flow chart is presented to illustrate the steps to take. It starts with a hen and seven roosters from which its mate can be chosen. Some mates may be unsuitable, because of other reasons besides inbreeding (such as health and breeding goal), and these are discarded in step 1. In step 2 relationship coefficients are determined. At this point, to facilitate the comprehension, the calculation of inbreeding and relationship coefficients is illustrated in the lower part of the infographic. In step 3 mates with a high relationship coefficient (i.e. between close relatives) are discarded. And in the 4th and final step an extra factor is introduced which makes the animal with a low MK to be favoured and chosen. The infographic ends by pointing out four common pitfalls or misunderstandings around inbreeding and providing the correct interpretation.

Infographic 3: managing inbreeding at population level. The information for population managers and breeding organisation is based on the cycle used in animal breeding education (Oldenbroek and De Waay, 2015) with the following steps: (1) Monitor the population by determining average inbreeding and relationship coefficients. (2) Determine the inbreeding rate (ΔF) per generation. (3) Analyse risk factors for the population. (4) Take measures and set up rules to counteract risk factors. (5) Set up communication. (6) Evaluate results. The inbreeding rate is the main indicator tool here, which is illustrated with a traffic light in vibrant colours for each of the categories (Table 1). The term effective population size is avoided because this generally gives a lot of confusion, requiring lengthy explanation to make general public aware of the fact that census size is only one, and often a minor, factor determining effective population size. Instead, risk factors for high inbreeding rates are illustrated. Possible actions to decrease inbreeding rates, such as the use of MKs and a breeding circle, are illustrated as well, with special emphasis on MKs as this is one of the most effective tools.

Discussion

Frequently, there is a rather incomplete understanding of inbreeding. This can lead, on the one hand, to attributing everything that is wrong to inbreeding and, on the other hand, to a denial of any detrimental effect of inbreeding. Consequently, this has led to heated debates and even splitting of breeding organisations of already small populations. The infographics are designed to increase the knowledge of breeders and breeding organisations. The use of social media and other actions may enhance their distribution even further. Evaluation of the effect of infographics on the practices of breeders and breeding organisations will be a next step.

Infographics may be useful for other concepts in animal breeding that are used widely but not always well understood by the general public. Examples can be breeding against diseases or genomic selection. The development of infographics also forces the designers and scientists to evaluate what are the core concepts needed for a proper understanding of the issue, and as such can be a useful exercise for teachers and students as well.

Table 1. Risk assessment based on inbreeding rate.

ΔF per generation	Assessment	Risk	Colour
>1%	Critical	Extinction due to accumulation of genetic defects	Red
0.5-1%	Endangered	Genetic defects almost certainly will be expressed	Orange
0.25-0.5%	Vulnerable	Genetic defects might be expressed	Yellow
<0.25%	Normal	Small chance of genetic defects	Green

Conclusions

The infographics clearly fill in a gap. In the Netherlands, they were readily picked up and links soon appeared after the introduction on websites of several breeding organisations. Three video presentations were made containing elements of the infographics, the first one about the basics of inbreeding has (12/12/2021) over 500 views on YouTube (https://youtu.be/OSzCsI4UOjs). This shows the need for this information and that the developed infographics turned out to be really helpful to get the message across.

The infographics can be downloaded from www.fokkenmetverstand.nl (Dutch) and the ERFP website https://www.animalgeneticresources.net/index.php/publications/ (English). At the time of writing, other languages are in preparation, and links will be made available on the ERFP website.

Acknowledgements

Development of the infographics was financially supported by the Dutch Ministry of agriculture, nature and food quality (projects KB-34-013-002 and WOT-03-003-056) and the European Regional Focal Point for Animal Genetic Resources (ERFP). We also would like to thank Eline de Kool for translating most parts from Dutch to English and Marleen Hijlkema for initial discussions on the set up of the infographics.

References

Falconer D.S., and Mackay T.F.C. (1996). Introduction to quantitative genetics. Longman Group, Harlow, UK

FAO (2013). *In vivo* conservation of animal genetic resources. FAO Animal Production and Health Guidelines. No. 14. Rome

Frankham R., Ballou J.D., and Briscoe D.A. (2010) Introduction to conservation genetics. Cambridge University Press, Cambridge, UK.

Lynch M., and Walsh B. (1998) Genetics and analysis of quantitative traits. Sinauer, Sunderland.

Oldenbroek K., and de Waaij L. (2015) Textbook Animal Breeding and Genetics for BSc students. Centre for Genetic Resources the Netherlands and Animal Breeding and Genomics Centre, Wageningen, Groen Kennisnet: https://wiki.groenkennisnet.nl/display/TAB/

208. Y-chromosomal haplogroups from wild and domestic goats reveal ancient migrations and recent introgressions

J.A. Lenstra[1*], VarGoats Consortium[2], I.J. Nijman[1,3], B.D. Rosen[4], P. Bardou[5], T. Faraut[5], T. Cumer[6], K.G. Daly[7], Z. Zheng[8], Y. Cai[8], H. Asadollahpour[8], B. Çınar Kul[9], W.-Y. Zhang[10], G. E[10], A. Ayin[11], M. Bakhtin[12], V.A. Bâlteanu[13], D. Barfield[14], H. Baird[15], B. Berger[16], T. Blichfeldt[17], G. Boink[18], S.R.A. Bugiwati[19], Z. Cai[20], S. Carolan[21], E. Clark[22], V. Cubric-Curik[23], M.I.A. Dagong[19], T. Dorji[24], L. Drew[25], J. Guo[26], J. Hallsson[27], S. Horvat[28], J. Kantanen[29], F. Kawaguchi[11], P. Kazymbet[12], N. Khayatzadeh[16], N. Kim[30], M. Kumar Shah[31], Y. Liao[32], A. Martínez[33], J.S. Masangkay[34], M. Masaoka[11], R. Mazza[35], J. McEwan[15], M. Milanesi[36], F. Md. Omar[37], Y. Nomura[11], N.-A. Ouchene-Khelifi[38], F. Pereira[39], G. Sahana[20], S. Sasazaki[11], A. Da Silva[40], M. Simčič[28], J. Sölkner[16], A. Sutherland[41], J. Tigchelaar[18], H. Zhang[23], Econogene Consortium[42], P. Ajmone-Marsan[43,44], D.G. Bradley[7], L. Colli[43,45], C. Drögemüller[46], C. Lei[8], H. Mannen[11], F. Pompanon[6], G. Tosser-Klopp[5] and Y. Jiang[8]

[1]Utrecht University, the Netherlands; [2]http://www.goatgenome.org/vargoats.html; [3]Univ. Medical Center Utrecht, Utrecht University, the Netherlands; [4]Agriculture Research Service, Beltsville, MD, USA; [5]GenPhySE, Univ. Toulouse, INRA, INPT, ENVT, Castanet Tolosan, France; [6]Université Grenoble Alpes, Université Savoie Mont Blanc, CNRS, LECA, Grenoble, France; [7]Trinity College Dublin, Dublin, Ireland; [8]College of Animal Science & Technology, Northwest A&F Univ., Yangling, China; [9]Faculty of Veterinary Medicine, Ankara Univ., Turkey; [10]Southwest University, Chongqing, China; [11]Kobe Univ., Kobe Japan; [12]Astana Medical Univ., Astana, Kazakhstan; [13]Inst. of Life Sciences, Univ. Agricultural Sciences and Veterinary Medicine of Cluj-Napoca, Cluj-Napoca, Romania; [14]Pleystole, Queensland, Australia; [15]AgResearch, Invermay Agricultural Centre, Mosgiel, New Zealand; [16]Univ. Natural Resources and Life Sciences Vienna (BOKU); [17]Norwegian Association of Sheep and Goat Breeders, Aas, Norway; [18]Stichting Zeldzame Huisdierrassen, Wageningen, the Netherlands; [19]Hasanuddin Univ., Makassar, Indonesia; [20]Aarhus Univ., Tjele, Denmark; [21]Old Irish Goat Society, Mulranny, Ireland; [22]Univ. Edinburgh, Edinburgh, United Kingdom; [23]Univ. Zagreb, Zagreb, Croatia; [24]International Centre for Integrated Mountain Development, Kathmandu, Nepal; [25]Finch Hatton, Queensland, Australia; [26]Sichuan Agricultural Univ., Chengdu, China; [27]Agricultural Univ. Iceland, Reykjavik, Iceland; [28]Biotechnical Faculty, Univ. Ljubljana, Ljubljana, Slovenia; [29]Natural Resources Inst. Finland (Luke), Helsinki, Finland; [30]Korea Research Institute of Bioscience and Biotechnology (KRIBB), Daejeon, South Korea; [31]National Animal Nutrition Research Centre, Lalitpur, Nepal; [32]Guangxi Key Laboratory of Livestock Genetic Improvement, Guangxi, China; [33]Universidad de Córdoba, Córdoba, Spain; [34]Univ. of the Philippines, Los Banos, Philippines; [35]Laboratorio Genetica e Servizi, Agrotis srl, Cremona, Italy; [36]Univ. degli studi della Tuscia, Viterbo, Italy; [37]Bangladesh Agricultural Univ., Mymensingh, Bangladesh [38]Science Veterinary Inst., University of Blida, Blida, Algeria; [39]IDENTIFICA Genetic Testing Maia & Centre for Functional Ecology, Porto, Portugal; [40]E2LIM, Univ. Limoges, Limoges, France; [41]New Zealand Arapawa Goat Association, New Zealand; [42]https://cordis.europa.eu/project/id/QLK5-CT-2001-02461; [43]Univ. Cattolica del S. Cuore di Piacenza and BioDNA Biodiversity and Ancient DNA Res. Centre, Piacenza, Italy; [44]UCSC PRONUTRIGEN Nutrigenomics Res. Centre, Piacenza, Italy; [45]UCSC BioDNA Biodiversity and Ancient DNA Res. Centre, Piacenza, Italy; [46]Univ. Bern, Bern, Switzerland; j.a.lenstra@uu.nl

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_208, © J.A. Lenstra *et al.* 2022

Abstract

By its paternal transmission, Y-chromosomal haplotypes are sensitive markers of population history and male-mediated introgression. We used whole-genome sequences (WGSs) of 386 domestic goats from 75 modern breeds and 7 wild goat species generated by the VarGoats goat genome project. Phylogenetic analyses indicated five domestic haplogroups Y1AA, Y1AB, Y1B, Y2A and Y2B. Haplogroup distributions for 180 domestic breeds indicate ancient paternal population bottlenecks during the migration into northern Europe, southern Asia and Africa. Sharing of haplogroups reveals male-mediated introgressions: from Asia into Madagascar and, more recently, into the South-African Boer goat; then from this breed into other southeastern African goats; and from Europe into native Korean and Ugandan goats. This study illustrates the power of the Y-chromosomal variation for the reconstructing the history of domestic species with a wide geographic range.

Introduction

The genetic diversity of goats has been reported to show a weak phylogeography of the mtDNA (Luikart *et al.*, 2001), but a strong geographic partitioning of the autosomes (Amills *et al.*, 2017; Colli *et al.*, 2018; Zheng *et al.*, 2020). This is in sharp contrast to similar studies of sheep (Kijas *et al.*, 2012; Ciani *et al.*, 2020), but has been confirmed by studies of ancient DNA (Daly *et al.*, 2022). However, the goat Y-chromosome as a marker for the paternal lineages has not yet been studied at a worldwide scale. For many species, Y-chromosomal variants have proved to be highly informative markers for paternal origin, which generally show a much stronger phylogeographic differentiation than mitochondrial or autosomal variants (Jobling and Tyler-Smith, 2017).

Previous analyses of Y-chromosomal diversity in European and Asian goats on the basis of low number of SNPs defined four haplotypes Y1A, Y1B, Y2A and Y2B, which have a strong geographic differentiation (Lenstra and the Econogene Consortium, 2005; Tabata *et al.*, 2018). Here, we use WGS data from the VarGoats project for a broad sample of European, Asian and African goats in order to define the major haplogroups, their phylogeny and the relationships with haplotypes from Iranian and Anatolian bezoar. After combining with data from other sources, we characterize the phylogeographic patterns across the three continents. We present evidence for population bottlenecks during the ancient migration of domestic goat and for more recent introgressions within and between continents.

Materials & methods

We extracted SNPs from four contigs in the male-specific part of the Y-chromosome, which after quality control were used for constructing Neighbour-Joining (NJ) trees and a Median-Joining Network of bezoars and domestic goats. We identified Y-chromosomal haplogroups for 2,214 goats on the basis of: (1) the VarGoats WGS data; (2) dideoxy sequencing of DNA segments covering SNPs that are diagnostic for the haplogroups; (3) published data for French, Swiss, Turkish and South-Asian goat breeds; (4) KASP genotyping of diagnostic SNPs; and (5) SNP alleles extracted from WGS data.

Results

Neighbour-Joining phylogenetic trees (not shown) and a Median-Joining Network (Figure 1) on the basis of the VarGoats data gave completely consistent patterns of clustering and show five domestic haplogroups that correspond to the previously found haplotypes: Y1AA, Y1AB (both representing a distinct cluster of Y1A haplotypes), Y1B, Y2A and Y2B. All domestic haplotypes differ from the Iranian or Anatolian bezoar haplotypes, which also differ from each other. The bezoar haplotypes are associated with the domestic Y1AA (CaY1AA) or Y2B (CaY2B) clusters, are linked to the Y1 or Y2 roots (CaY1 and CaY2) or are between the Y1 and Y2 haplogroups (Y0, in the NJ tree close to the root).

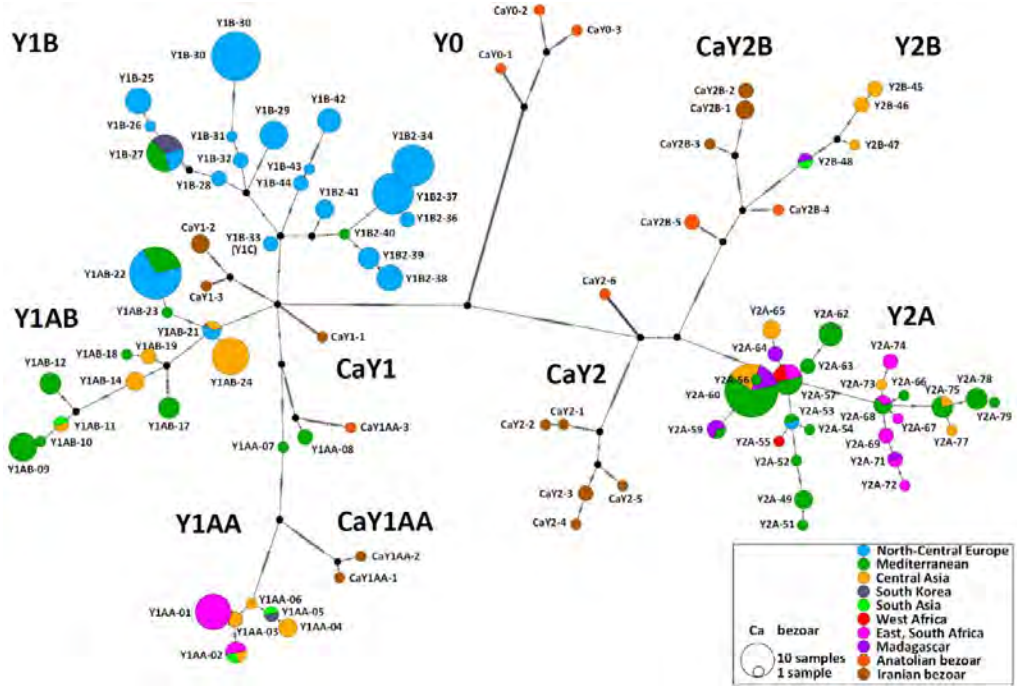


Figure 1. Median-Joining Network of 286 male domestic goats and 27 bezoars.

The network further shows a remarkable geographic differentiation:

1. Haplogroup Y1B is predominant in central and northern Europe, but is also found in South Korea and Uganda.
2. Y2A is almost absent in central and northern Europe, but is the predominant haplogroup in Spain, Anatolia and Africa. The British Anglo-Nubian (not shown) carries Y2A and Y1AA because of its Asian and African ancestry. Remarkably, most Y2A haplotypes on Madagascar are more related to Asian than to continental-African Y2A haplotypes.
3. Y2B does not occur in Europe, continental Africa or west Asia, but is a major haplotype in east and southeast Asia. It was also found in a single goat from Madagascar.
4. Y1AA is sporadic in Europe, but is a major lineage in Asia. Two Y1AA haplotypes in East and South Africa are closely related to Asian haplotypes. This reflects the Asian ancestry of the popular Boer breed, which has been crossed into other southeastern African goats.

Discussion

We propose that the phylogeographic pattern is the result of ancient population bottlenecks and recent introgressions (Figure 2). The Y-chromosomal link between Madagascar and Asia is in line with the genomic composition of the Malagasy zebu (Magnier *et al.*, 2022) and the Austronesian origin of the Malagasy language.

We conclude that the Y-chromosomal variation of goats reveals bottlenecks, expansions and introgressions, illustrating the power of Y-chromosomal markers for inferring the genetic origin of mammalian populations.

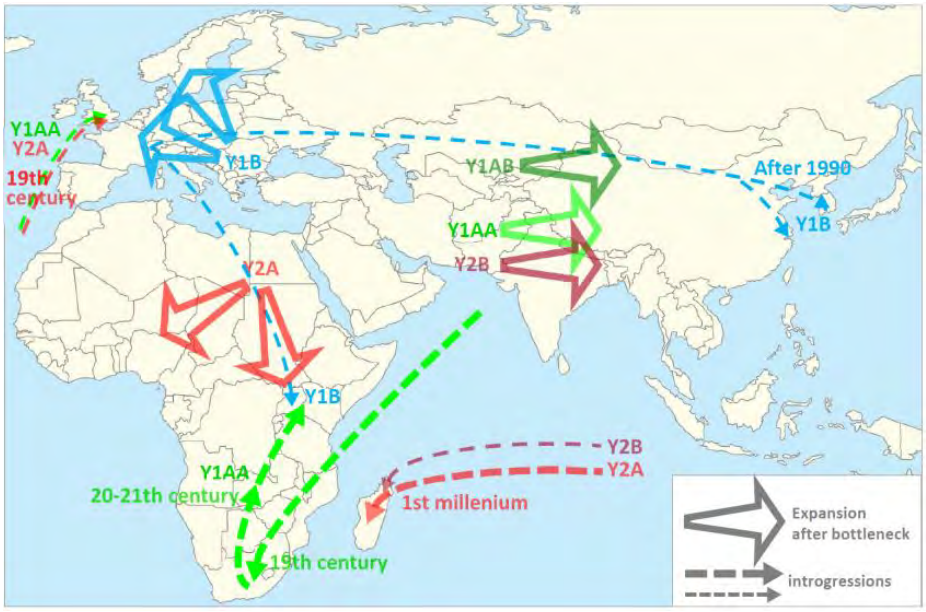


Figure 2. Bottlenecks, expansions and migrations inferred from the distribution of Y-chromosomal haplotypes.

References

Amills, M., Capote, J. and Tosser-Klopp, G. (2017) Anim. Genet., 48(6):631–644. https://doi.org/10.1111/age.12598

Ciani, E., Mastrangelo, S., Da Silva, A., Marroni, F., Ferenčaković, M. *et al.*. (2020) Genet. Selection Evol., 52(1):1–14. https://doi.org/10.1186/s12711-020-00545-7

Colli, L., Milanesi, M., Talenti, A., Bertolini, F., Chen, M., *et al.* (2018) Genet. Selection Evol., 50(1):1–20. https://doi.org/10.1186/s12711-018-0422-x

Daly, K.G., Mattiangeli, V., Hare, A.J., Davoudi, H., Doost, *et al.* (2021) Proc. Natl. Acad. Sci. USA 118(25):e2100901118. https://doi.org/10.1073/pnas.2100901118

Jobling, M.A. and Tyler-Smith, C. (2017) Nature Rev. Genet., 18(8):485–497. https://doi.org/10.1038/nrg.2017.36

Kijas, J.W., Lenstra, J.A., Hayes, B., Boitard, S., Neto, L.R., *et al.* (2012) PLoS Biology, 10(2). https://doi.org/10.1371/journal.pbio.1001258

Lenstra, J.A. and the Econogene Consortium. (2005) Proc. Of The Role of Biotechnology for the Characterization of Crop, Forestry, Animal and Fishery Genetic Resources. Turin, Italy.

Luikart, G., Gielly, L., Excoffier, L., Vigne, J., Bouvet, J. *et al.* (2001) Proc. Natl. Acad. Sci. USA 98(10):5927–5932. https://doi.org/10.1073ypnas.091591198

Magnier, J., Druet, T., Naves, M., Ouvrard, M., Raoul, S. *et al.* (2022) Genes|Genomces| Genetics, doi.org/10.1101/2021.10.08.463737

Tabata, R., Kawaguchi, F., Sasazaki, S., Yamamoto, Y., Bakhtin, M. *et al.* (2018) Anim. Sci. J., 90(3:17–322. https://doi.org/10.1111/asj.13144

Zheng, Z., Wang, X., Li, M., Li, Y., Yang, Z., Wang *et al.* (2020) Sci. Adv., 6(21). https://doi.org/10.1126/sciadv.aaz5216

209. Characterization and effect of inbreeding on growth and reproductive traits in Nellore Cattle

L.F. Benfica[1*], R.D. do Bem[1], E.R Carrara[2], L.F. Brito[3], D.A. Silva[1], R. Carvalheiro[1], J.N.S.G. Cyrillo[4] and M.E.Z. Mercadante[4]

[1]School of Agricultural and Veterinarian Sciences, São Paulo State University (Unesp), 14884-900, Jaboticabal-SP, Brazil; [2]Federal University of Viçosa (UFV), Viçosa-MG, Brazil; [3]Purdue University, West Lafayette, IN 47907, USA; [4]Beef Cattle Research Center, Institute of Animal Science, 14174-000, Sertãozinho-SP, Brazil; lorena.benfica@yahoo.com.br

Abstract

The objective of this study was to evaluate the effect of inbreeding, through pedigree and genomic-based inbreeding coefficients, on the selection (postweaning) weight (WSel), scrotal circumference (SC), and days to calving (DC) and to evaluate the effect of recent and ancient inbreeding in these traits in Nellore cattle. A total of 1,446 animals genotyped was used. Individual inbreeding coefficients were estimated based on pedigree (FPED), runs of homozygosity (ROH, FROH) and diagonal of the H matrix (pedigree-genomic relationship matrix) (FH). For each 1% increase in FPED, FROH and FH a decrease of 1.29, 1.27 and 1.31 kg in WSel and an increase of 1.48, 1.35 and 0.55 days in the DC were observed, respectively. However, there were no clear patterns of the effect of inbreeding in the SC. Especially for DC, ancient inbreeding was less harmful compared to recent inbreeding. In conclusion, inbreeding negatively affects postweaning weight and female fertility trait in Nellore cattle.

Introduction

Mating between related animals increases the proportion of homozygous genes associated with inbreeding, which can lead to inbreeding depression (Doekes *et al.*, 2021). Inbreeding coefficient (F) is commonly used to describe the proportion of autozygosity in an individual's genome and can be estimated from pedigree information or genomic data (e.g. Ferenčaković *et al.*, 2013).

The objective of this study was to assess the estimated inbreeding levels in an experimental herd of Nellore cattle, by using three methods. In addition, the effect of inbreeding on the selection (postweaning) weight (WSel), scrotal circumference (SC) and days to calving (DC) and the effect of recent and ancient inbreeding was evaluated.

Materials & methods

All management procedures followed animal welfare guidelines and were conducted in accordance with the State Law No. 11,977 of the State of São Paulo, Brazil. The database refers to Nellore cattle from a selection experiment established in 1980 at the Institute of Animal Science (IZ), Beef Cattle Research Center, Sertãozinho, SP, Brazil (Benfica *et al.*, 2020).

Genomic data of 1,446 animals born from 1983 to 2019 were available, with 757 animals being genotyped with GGPHD 75K (GeneSeek® Genomic Profiler Indicus) and 689 with HD (High-Density Bovine BeadChip, Illumina Inc., San Diego, CA) panels. Genotypes obtained with the lower density panel were imputed to the high definition (HD) panel using the FImpute v.3 software (Sargolzaei et. al., 2014). Quality control (QC) was performed by excluding SNPs with no defined position or located on the sex chromosomes, with a call rate lower than 0.85, and the SNPs that deviate from Hardy-Weinberg expectation ($P<1\times10^{-5}$). After QC a total of 612,154 SNPs were available. Pedigree file included 3,075 animals (378 bulls and 1,510 dams) and 8 generations. The traits considered were the selection (postweaning) weight adjusted to 378 days of age for

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_209, © L.F. Benfica *et al.* 2022

males and postweaning weight adjusted to 550 days of age for females (WSel), the scrotal circumference at yearling (SC), and the repeated days to calving considering all breeding seasons in which the female participated (DC). Days to calving were obtained for all females that calved as the difference between the calving date and the date of the beginning of that breeding season. Cows that did not calve were penalized by adding 21 to the largest days to calving value within the breeding group (Ferreira Júnior *et al.*, 2018).

Three methods were used to calculate the inbreeding coefficient: Pedigree-based inbreeding coefficient (FPED) was calculated using the BLUPF90 family programs (Misztal *et al.*, 2009). Plink 2.0 software (Purcell *et al.*, 2007) was used to identify the ROH. Following criteria were adopted to define a ROH: a minimum physical length of 1 Mb; a maximum gap of 500 kb between two successive SNP; a minimum of 50 consecutive homozygous SNP and a minimum density of one SNP per 100 kb. After identification, the ROH were classified into five length classes: (1) 1-2 Mb (FROH1); (2) 2-4 Mb (FROH2); (3) 4-8 Mb (FROH3); (4) 8-16 Mb (FROH4); and (5) >16 Mb (FROH5). These length classes indicate inbreeding resulting from ancient to most recent ancestors. A total inbreeding coefficient based on all ROH classes (FROH) was also estimated. Inbreeding coefficient based on pedigree-genomic relationship matrix (FH) was obtained by diagonal of the matrix H minus 1 (Legarra *et al.*, 2020) using preGSf90 (Misztal *et al.*, 2009).

The effect of inbreeding on phenotypes was estimated using BLUPF90 family of programs (Misztal *et al.*, 2009) including inbreeding coefficient as a linear covariate in single-trait analysis. The model for WSel was previously described by Benfica *et al.* (2020), and the models for SC and DC were described by Ferreira Junior *et al.* (2018).

Results

Descriptive statistics for the traits evaluated are in Table 1. Pearson correlations between all estimated inbreeding coefficients are presented in Figure 1A. The correlation coefficients between pedigree-based inbreeding and genomic-based inbreeding were moderately high, of 0.62 (FPED × FROH) and 0.58 (FPED × FH). The highest correlation was observed between FROH and FH (0.77). Classical estimates of inbreeding (FPED, FROH, FH) and FROH classes were positively and moderately correlated, however inbreeding based on smaller ROH segments were weakly correlated with inbreeding based on larger ROH segments.

The variation of the three inbreeding estimates, FPED, FROH and FH, over the years of selection for postweaning weight, and more recently for postweaning weight and feed efficiency (Benfica *et al.*, 2020) were shown in Figure 1B. FPED showed an increasing tendency over the years from 1982, during the first decade of the selection experiment. A greater number of inbreeding events was captured by ROH that kept mean values higher than the other inbreeding measures.

The regression coefficients of inbreeding effect are shown in Table 2. In general, inbreeding depression was observed for WSel and DC, but not for SC.

Table 1. Number of animals (and records) (N), mean, standard deviation (SD), minimum and maximum values for selection (postweaning) weight (WSel), scrotal circumference (SC) and days to calving (DC).

Trait	N	Mean	SD	Min	Max
WSel, kg	1,446	354	53.0	136	512
SC, cm	982	23.1	2.72	14.5	32.5
DC, day	452 (1,604)	344	33.6	263	417

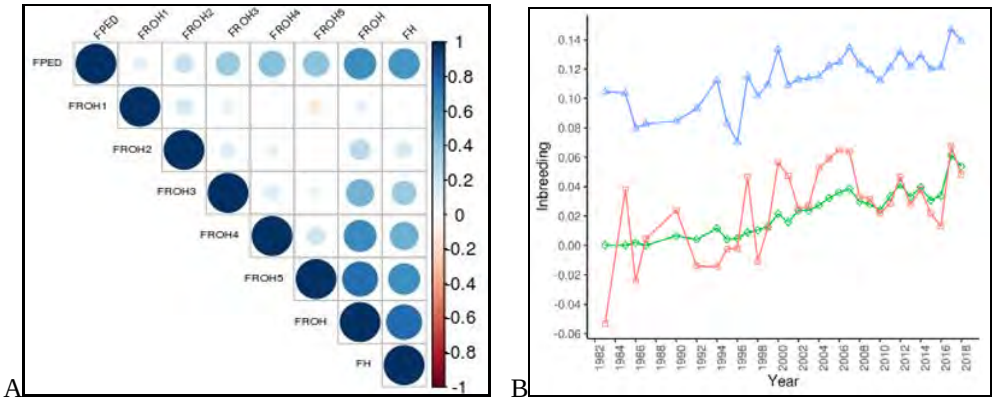


Figure 1. (A) Pearson's correlations between inbreeding measures; (B) Mean of FPED (green), FROH (blue) and FH (red) by year of birth.

Table 2. Estimates of inbreeding depression for selection (postweaning) weight (WSel), scrotal circumference (SC) and days to calving (DC), expressed as the change in phenotype per 1% of increasing in inbreeding.

Inbreeding method[1]	Trait					
	WSel, kg		SC, cm		DC, day	
	Estimate	SE	Estimate	SE	Estimate	SE
FPED	-1.29	0.85	0.09	0.07	1.48	0.84
FROH1 (0-2 Mb)	0.19	2.96	-0.39	0.24	-4.28	3.07
FROH2 (2-4 Mb)	-4.16	1.83	-0.11	0.14	-2.59	2.01
FROH3 (4-8 Mb)	-0.42	1.03	-0.003	0.08	-2.02	1.12
FROH4 (8-16 Mb)	-1.48	0.72	-0.01	0.05	1.66	0.89
FROH5 (>16 Mb)	-1.09	0.47	-0.02	0.04	2.19	0.51
FROH (total)	-1.27	0.37	-0.03	0.03	1.35	0.42
FH	-1.31	0.30	-0.001	0.02	0.55	0.33

[1] FPED: pedigree-based inbreeding; FROH: inbreeding based on runs of homozygosity; FGRM: inbreeding based on genomic relationship matrix.

Discussion

In the present study, as well as in previous studies (Doekes *et al.*, 2019; Makanjuola *et al.*, 2020), FPED was moderately to highly correlated with FROH. The higher correlation between FPED and FROH5 may be a direct reflection of pedigree depth, as ROH reflect past and recent relatedness and FPED estimates are based on pedigree records, which may not extend back many generations (Ferenčakovic *et al.*, 2013).

The inbreeding age can be derived from the length of ROH, with shorter lengths reflecting more ancient inbreeding, as it has already gone through the process of breakdown and recombination. Theoretically, ancient inbreeding should be less harmful, as a decrease in the frequency of deleterious alleles over time is expected as a result of natural selection (Doekes *et al.*, 2019; Makanjuola *et al.*, 2020). For WSel trait, there was no clear difference between effect of long ROH (recent inbreeding) and short ROH (ancient inbreeding). For DC, the results are in accordance with theory and ancient inbreeding even caused a favorable result in the trait, reducing the number of days to calving. Overall, both long and short ROH appear to contribute to inbreeding depression.

Inbreeding influenced the reproductive performance of females. Increasing 1% in FPED, FROH and FH increased 1.47, 1.35 and 0.55 days in DC. Martikainen *et al.* (2017) reported a 0.4 day increase in interval between inseminations for 1% increase in FROH of dairy cows. Overall, the increase in inbreeding had an unfavorable effect on productive and fertility traits, except for SC. Recent inbreeding is more harmful for female fertility trait than ancient inbreeding. Hence, the management of inbreeding should be highlighted in breeding programs for Nellore cattle.

Acknowledgements

This study was supported by National Council for Scientific and Technological Development (CNPq, 434681/2018-0), by São Paulo Research Foundation (FAPESP, 2017/10630-2), and by Coordination for the Improvement of Higher Education Personnel, Brazil (CAPES, Finance Code 001).

References

Benfica L.F.; Sakamoto L.S., Magalhães A.F.B., de Oliveira M.H.V., de Albuquerque L.G. *et al.* (2020) J. Anim. Sci. 98(11):1-9. https://doi.org/10.1093/jas/skaa350

Doekes H.P., Bijma P., Windig J.J. (2021) Genes. 12(6):926. https://doi.org/10.3390/genes12060926

Doekes H.P., Veerkamp R.F., Bijma P., de Jong G., Hiemstra S.J. *et al.* (2019) Genet. Sel. Evol. 51:54. https://doi.org/10.1186/s12711-019-0497-z

Doublet A.C., Croiseau P., Fritz S., Michenet A., Hozé C. *et al.* (2019) Genet Sel Evol. 51:52. https://doi.org/10.1186/s12711-019-0495-1

Ferenčaković M., Hamzic E., Solberg T.R., Klemetsdal G., Curik I. *et al* (2013) J. Anim. Breed. Genet. 130:286-293. https://doi.org/ 10.1111/jbg.12012

Ferreira Júnior, R.J., Bonilha, S.F.M., Monteiro, F.M., Cyrillo, J.N.S.G., Branco, R.H. *et al.* (2018) J. Anim. Sci. 96:4035-4044. https://doi.org/ 10.1093/jas/sky276

Legarra A., Aguilar I., and Colleau J.J. (2020) J. Dairy Sci. 103:3363–3367. https://doi.org/10.3168/jds.2019-17750

Makanjuola B.O., Maltecca C., Miglior F., Schenkel F.S., and Baes C.F. (2020) BMC Genomics 21:605. https://doi.org/10.1186/s12864-020-07031-w

Martikainen K., Tyrisevä A.M., Matilainen K., Pösö J., and Uimari P. (2017) J. Anim. Breed. Genet. 134:383–392. https://doi.org/ 10.1111/jbg.12285

Misztal I., Legarra A., and Aguilar I. (2009) J. Dairy Sci. 92:4648–4655. https://doi.org/10.3168/jds.2009-2064

Purcell S., Neale B., Todd-Brown K., Thomas L., Ferreira M.A.R. *et al.* (2007) Am J Hum Genet. 81:559–75. https://doi.org/10.1086/519795

Sargolzaei, M., Chesnais, J.P., and Schenkel, F.S. (2014) BMC Genomics 15:478. https://doi.org/10.1186/1471-2164-15-478

210. Genomic diversity of Czech and Slovak native goat breeds

L. Vostry[1*], H. Vostra-Vydrova[1], J. Rychtarova[2], B. Hofmanova[1], N. Moravcikova[3], R. Kasarda[3], M. Margetin[3], I. Drzaic[4], V. Cubric-Curik[4], J. Sölkner[5] and I. Curik[4]

[1]Czech University of Life Sciences, Kamycka 129, 16500 Prague, Czech Republic; [2]Institute of Animal Science, Pratelstvi 815, Prague, Czech Republic; [3]Slovak University of Agriculture in Nitra, Tr. A. Hlinku 2, 94976 Nitra, Slovakia; [4]University of Zagreb Faculty of Agriculture, Svetošimunska c. 25, 10000 Zagreb, Croatia; [5]University of Natural Resources & Life Sciences Vienna, Gregor-Mendel-Strasse 33, 1180 Vienna, Austria; vostry@af.czu.cz

Abstract

Due to that Czech indigenous goat breeds, together with Slovak indigenous goat breeds, are closed population, face this population loss of genetic diversity, because these populations are closed. The aim of this study was to perform a genome-wide diversity characterization using Illumina GoatSNP50 BeadChip of indigenous Czech Braun Shorthair and White Shorthair goat breeds together with international Alpine goat from Czech Republic and indigenous White Shorthair goat breed from Slovak Republic. For comparison was in this analysis included indigenous alpine goat breed from Austria and Switzerland. To determine key conservation risk parameters, we estimated inbreeding (F_{ROH}) and current and historical effective population size (Ne). The observed estimates for Czech and Slovak indigenous breed show similar values as other alpine goat breeds. In addition, by comparing genetic differences by Neighbour-net, no differences were recorded between the Czech-Slovak WSG. The study provides data and information utilizable in the management of conservation programs planned to reduce inbreeding and to minimize loss of genetic variability.

Introduction

The main aim of the conservation of the livestock population is maintaining genetic diversity. This is mainly because about one-fifth of livestock breeds became extinct during the second half of the twentieth century and a large part of these breeds are endangered (Hodges 2006). Much of the lost genetic variability was mostly unknown and undocumented (Bruford *et al.*, 2003). For these reasons, it is necessary to study and documented genetic diversity by local breeds, which can be used in future breeding with commercial breeds, especially when future breeding goals differ from those of today (Woelders *et al.* 2006).

White Shorthair goat (WSG) and Brown Shorthair goat (BSG) are local breeds of domestic dairy goats from the Czech and Slovak Republic. The WSG was created by crossing the native Czechoslovak landrace goats crossed with Saanen goats from Switzerland and Germany. The BSG was created by crossing the Czech native breed with non-uniform colour with German Brown bucks in the period between 1900 and 1930, mainly in the Czech-German borderlands.

The aim of this study was to analyse the genetic diversity and conservation status of Czech and Slovak native goat breed.

Materials & methods

Data collection and genotyping. Our dataset, contained samples of two Czech indigenous breeds Brown Shorthair (CZ_BSG) and White Shorthair (CZ_WSG), one international breed Alpine goat (CZ_A) collected in the Czech Republic and one Slovak indigenous breed White Shorthair (SK_WSG) collected in the Slovak Republic, we genotyped 121 individuals. In addition to our original data, the dataset of Austria and Switzerland was used. The five breeds from Austria (Blobe – AT_BLB, Chamois Coloured – AT_CHA,

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_210, © L. Vostry *et al.* 2022

Pinzgau – AT_PNZ, Styrian Pied – AT_STP, Tauern Pied – AT_TAP) and ten Switzerland were: (Appenzell – CH_APP, Grisons striped – CH_GST, Chamois Coloured – CH_CHA, Nera Verzasca – CH_NVR, Peacock – CH_PEA, Saanen – CH_SAA, Booted – CH_BOT, Tessin grey – CH_TSG, Toggenburg – CH_TGB, Valais – CH_VAL) In these breeds a historical admixture with Czech-Slovakia indigenous breeds is expected. The animals were genotyped with the Illumina goat SNP50 BeadChip. Data cleaning was performed using software PLINK 1.9 (Chang *et al.*, 2015) to select only common autosomal SNP markers with known chromosomal positions. Then, only animals and common autosomal SNP markers with a call rate higher than 95% across all breeds were retained in the database.

Runs of homozygosity inbreeding. Runs of homozygosity (ROH) segments were detected using detectRuns (Biscarini *et al.* 2018). The ROH segments were defined with a specific minimum length (>4 Mbp, >8 Mbp) reflecting different generations of ancestors in the past. The different proportions of missing calls within homozygous regions were allowed depending on the ROH length class (>4 Mbp 1 call, >8 Mbp 2 calls). No heterozygous call was allowed for all ROH segments.

Population structure and gene flow. Inference of the phylogenetic relationship between 20 goat breeds was represented as the Neighbour-net network inferred from pairwise Nei's genetic distances (Nei, 1972). Nei's distances were calculated using the R package StAMPP R (Pembleton *et al.*, 2013). A NeighborNet network was created and represented using SplitsTree4 software.

Results

Genotype data of Alpine goat breed with a substantial number of genotypes available were selected to compare the genetic diversity of Czech and Slovak native goat breeds.

The relatedness between analysed breeds was evaluate by Neighbour-net (Figure 1) based on pairwise Nei's genetic distances. As expected, the results showed that CZ_WSG and SK_WSG create together separated cluster within the main reticulation. The CZ_BSG create separate cluster on separate reticulation. The CZ_A created cluster together with AT_CHA and CH_GST.

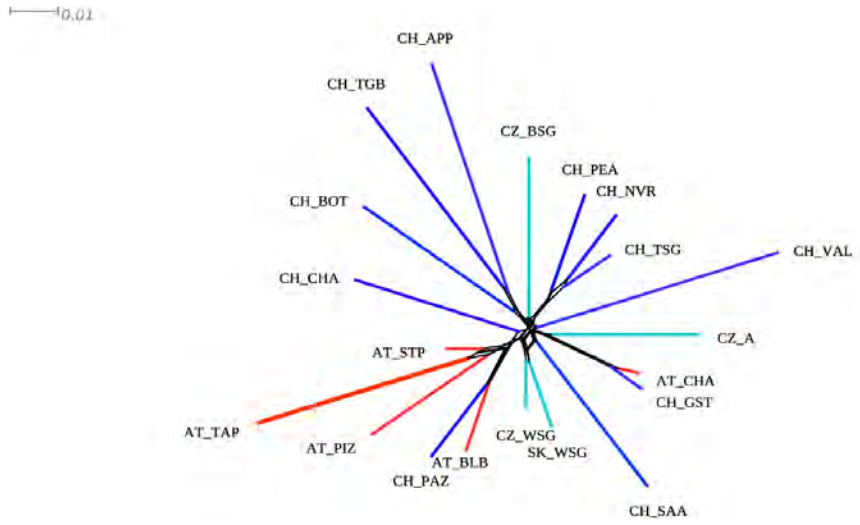


Figure 1. Neighbour-Net graph of Nei's distances between breeds. Each color represent one country: Austria, Czech and Slovak Republic, Swizerland.

The estimation inbreeding coefficient based on ROH>4 and >8 Mb and effective population size across Czech and Slovak native goat and representative Apline goat breed from each cluster are presented in Table 1. While Czech breeds have a higher value of current and historical inbreeding in comparison with other breeds. For Slovak goat breeds was estimated two times the lower value of both inbreeding with comparison to the same breeds keep in the Czech Republic. The highest value of inbreeding from Czech breeds based on different length of ROH was estimated in BSG and the lowest value in WSG. The overall highest value of inbreeding was observed in the AT_TAB breed. The highest value of inbreeding based on a short length of ROH ($F_{ROH>4Mb}$) points to more frequent mating of related individuals in the history than in current breeding.

The estimates of current effective population size indicate that analysed populations are not at risk of loss of genetic diversity. The lowest value of the current population size of Czech and Slovak native goats was found in breed SK_WSG (N_e=127) and the highest in breed CZ_WSG (Ne=267). A comparison of the current Ne and Ne 10 generations ago shows that most breeds have decreased Ne. On the contrary, in breeds CZ_BSG there was an increase in the effective size of the population.

Discussion

The study of genetic diversity in Czech and Slovak native goat breeds confirm that breeds are affected by the loss of genetic diversity, especially CZ_BSG. The estimated genomic inbreeding coefficient (FROH) of Czech native breeds, show a higher value than the inbreeding coefficients estimate based on pedigree data (CZ_WSG – 0.02 and CZ_BSG – 0.05, Vostra-Vydrova *et al.*, 2020). Contrariwise, the current effective population size estimated in this analysis shows a smaller value than Ne estimated on pedigree data (311-CZ_WSG and 87 – CZ_BSG). Thy higher N_e value of CZ_BSG compared with CZ_WSG and SK_BSG are affected by a smaller number of individuals of analyzed the breed. Increased Ne by CZ_BSG is affected by expected improving other breeds during 10 generation ago.

Our study points to tendency clustered CZ_WSG and SK_WSG between Austria and alpine goat. The Neighbour-net indicated the close genetic distance between CZ_WSG and SK_WSG breed and higher distance of CZ_BGB and Czech and Slovak WSG. This higher similarity between WSG depends on the

Table 1. Inbreeding coefficient, effective population size and number of animals for Czech and Slovak native breeds and selected Alpine breeds.

	Inbreeding Coefficient				Effective population size		Number of animals
	$F_{ROH>4Mb}$		$F_{ROH>8Mb}$		Current	10 generations ago	
	Mean	SD[1]	Mean	SD[1]			
CZ_A	0.047	0.025	0.030	0.023	153	166	26
AT_BLB	0.019	0.031	0.013	0.028	66	3,093	34
CZ_WSG	0.045	0.069	0.037	0.062	267	411	36
CZ_BSG	0.074	0.053	0.050	0.050	166	93	34
AT_PIZ	0.067	0.064	0.054	0.058	141	87	28
CH_BOT	0.078	0.043	0.057	0.040	126	55	23
SK_WSG	0.021	0.043	0.016	0.041	124	1,555	24
AT_TAP	0.092	0.031	0.063	0.029	143	40	28
CH_NVR	0.045	0.045	0.034	0.041	90	575	42
CH_SAA	0.060	0.050	0.033	0.046	119	201	64

[1] Standard deviation.

historical development of these breeds. Until 1993, breeds CZ_WSG and SK_WSG were considered as one breed. After the division of Czechoslovakia were these breeds split into two populations. The close relatedness between CZ_WSG, SK_WSG and CH_SAA presented by Neighbour-net confirms the expectations that CH_SAA was used for improvement of the WSG in the past. The study provides data and information utilizable in the management of conservation programs planned to reduce inbreeding and to minimize loss of genetic variability.

Acknowledgements

This study was funded by Ministry of Agriculture of the Czech Republic (grant number QK1910156), by Slovak Research and Development Agency (grant numbers APVV-20-0161 and APVV-17-0060) and by ANAGRAMS-IP-2018-01-8708 project, funded by Croatian Science Foundation.

References

Biscarini F., Cozzi P., Gaspa G., Marras G. (2018). detectRUNS: R package version 0.9.5. Available at https://CRAN.R-project.org/package=detectRUNS.

Bruford M.W., Bradley D.G., Luikart G. (2003) Nature Reviews Genetics. https://doi.org/10.1038/nrg1203.

Hodges J. (2006) Poultry Science 85, 200-209. https://doi.org/10.1093/ps/85.2.200

Chang C.C., Chow C.C., Tellier L.C.A.M., Vattikuti S., Purcell S.M. *et al.*, (2015) GigaScience 4, 1. https://doi.org/10.1186/s13742-015-0047-8

Nei M. (1972) American Naturalist, 106, 283-292. http://dx.doi.org/10.1086/282771

Pembleton L.W., Cogan N.O.I. and Forster J.W. (2013) Molecular Ecology Resources 13, 946-952. https://doi.org/10.1111/1755-0998.12129

Vostra-Vydrova H., Hofmanova B., Moravcikova N., Rychtarova J., Kasarda R., *et al.*, (2020) Livestoc Science 240, 104163. https://doi.org/10.1016/j.livsci.2020.104163

Woelders H., Zuidberg C.A., Hiemstra S.J. (2006) Poultry Science 85, 216-22. https://doi.org/10.1093/ps/85.2.216

211. Population structure of U.S. Holsteins allows for a snapshot of allele frequency changes and family specific SNPs

T.J. Lawlor[1*], Y. Steyn[2], S. Tsuruta[2], Y. Masuda[2], D.A.L. Lourenco[2] and I. Misztal[2]

[1]Holstein Association USA Inc., 1 Holstein Pl, Brattleboro, VT 05301, USA; [2]University of Georgia, 425 River Road, Athens, GA 30605, USA; tlawlor@holstein.com

Abstract

Genetic change occurs in the U.S. Holstein population through the heavy use of specific bulls for a relatively short period of time. By focusing on a specific time-period, we were able to group a high percentage of the descendants of several prominent bulls into five different clusters. The average F_{st} across clusters was 0.03. Comparison between clusters revealed a heterogeneous mixture of allele frequency changes with varying degrees of magnitude and direction. Non-parallel responses between families suggests alternative goals and/or non-additive gene action. SNP effects for the trait stature were estimated independently for the five clusters and used to predict additive breeding values (BV). Correlations of within-cluster BVs with BVs based upon all animals combined varied from 0.70 to 0.88. By stratifying a population into subpopulations, family specific SNPs can be identified and used to increase or maintain genetic diversity.

Introduction

A top dairy bull can have tens of thousands of daughters and an even larger number of granddaughters. These family members, with similar degrees of relationship, appear at different time periods. For example, the first group of descendants of a young prominent bull will be his daughters. They have an additive genetic relationship of 0.5 with their sire. After another year or two, the sons of this prominent bull will start producing granddaughters, and there will be a mixture of both daughters and granddaughters in our population. And after another couple of years, the additive genetic relationship will drop to 0.25 as almost all the living descendants will be granddaughters. As we move through time, the average relationship of this prominent bull with our active breeding population will decrease as his descendants become more distant.

K-mean clustering results in animals with similar relationships being grouped together. By focusing on a specific time-period, the descendants of certain bulls are clustered together. That is the progeny of the most prominent young sires are clustered together, descendants of the slightly older sires with breeding age sons are grouped into a separate cluster, and so on down to the final cluster of animals with the lowest average relationship amongst the current breeding population. When looking at GWAS results of prominent Holstein sires, Ma *et al.* (2012) stated:

> Artificial insemination and genetic selection are major factors contributing to population stratification in dairy cattle. Artificial insemination has been widely used and this has increased the likelihood of the presence of related individuals in randomly selected samples and the presence of large half-sib families; both of which contribute to population stratification.

Materials & methods

A total of 20,099 genotyped Holstein animals were used as target population of breeding animals born between 2010 and 2012. They included 3,902 males with at least 25 progeny and 16,197 females with a classification score. K-means clustering with a built-in R package (kmeans) was applied on the genomic relationship matrix (G) to separate the population into five separate clusters. The pedigrees of the target animals were traced back for 10 generations. The frequency of each SNP was calculated for each generation. Changes in allele frequency (AF) is defined as the difference in frequency of each SNP between

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_211, © T.J. Lawlor *et al.* 2022

two successive generations. The F_{st} across clusters, a measure of the variation in allele frequencies among subpopulations, was calculated as in Bonhomme *et al.* (2010).

Expected inbreeding within or across cluster was calculated based on pedigree information assuming non-zero inbreeding for unknown parents (Aguilar and Misztal, 2008) with the INBUPGf90 package within the BLUPF90 software suite (Misztal *et al.*, 2014). The genetic correlations across clusters were estimated using stature as trait and the method by Duenk *et al.* (2020). This correlated the additive breeding value of a cluster using the SNP effects of the same cluster, or when using the SNP effects of another cluster. Female animals of each cluster were used to estimate SNP effects with the POSTGSF90 package (Misztal *et al.*, 2014) and applied on the males of each cluster. Unlike the method by Duenk *et al.* (2020), the correlations were adjusted using the method by Calo *et al.* (1973). Assuming a heritability of 0.45 in all clusters, this adjustment factor is 1/0.45.

Results & discussion

Successful population stratification was confirmed in several ways. PCA analysis validated the clustering of target animals into five different but overlapping subpopulations. Inbreeding analysis showed higher within cluster relationships and lower relationships between clusters. Average inbreeding by cluster was 0.22, 0.20, 0.18, 0.17 and 0.10; with average between cluster inbreeding of 0.11. A multi-modal distribution of inbreeding values within the first three clusters indicated the inclusion of daughters and granddaughters of the prominent bulls. Average F_{st} value across all cluster was 0.03. Indicating allele frequency differences between the clusters. The age of the prominent bull for each cluster was associated with its average inbreeding Table 1.

All clusters combined. Alternative procedures for identifying the top SNPs associated with selection were investigated. Grouping all target animals together increases the likelihood of identifying those SNPs with a similar and consistent allelic change across all clusters. The method of Rowan *et al.*, 2020, which looks for the SNPs that are highly associated with changes over time was investigated. Most of the top 100 SNPs identified by this method could be described as changing in a parallel way across all clusters. Interpretation of these allele frequency changes are consistent with a polygenic shift.

Separate clusters. Clustering similarly related descendants of a unique bull into individual groups provides a snapshot of the genotypic makeup of different families at a specific point in time. With alternative sources of genetic drift and linkage disequilibrium, each cluster may exhibit its own unique trajectory of AF changes. Three procedures were utilized to determine the most significant family specific SNPs. Two procedures involved comparing within-cluster AF changes and then looking for SNPs with the largest between cluster contrast. The 100 SNPs exhibiting the greatest range or the greatest variance between clusters were selected. The third method involved selecting the 100 SNPs with the highest $F_{st.}$

Table 1. Association of the average inbreeding value of the animals within the cluster with the age of the most prominent ancestor for that cluster.

Cluster description[1]	Most prominent ancestor[2]	Birthdate of bull	Average inbreeding of the cluster
A. Highest	Planet	March 2003	0.22
B.	Goldwyn	January 2000	0.20
C.	Shottle	July 1999	0.18
D.	O Man	March 1998	0.17
E. Lowest	Many	---	0.10

[1] Clusters are ranked by the average additive relationship of its members.

[2] Most prominent ancestor was identified as the bull with the largest numbers of descendants within a cluster.

The starting allele frequency for each cluster was similar with all animals tracing back to a common group of ancestors. The SNPs with the highest $F_{st.}$ would usually have one cluster exhibit a sharp change in its trajectory in the final generation. This is an indicator of genetic drift and a strong association between a cluster and a specific family. The genotype of the prominent sire causes a dramatic shift in the allele frequency.

Two procedures aimed at contrasting AF changes between clusters identified rapid directional changes, dramatic reversals in direction, and a mixture of parallel and non-parallel responses across clusters. Replicate Frequency Spectrum (Barghi *et al.* 2019) indicated that individual SNPs that showed a large directional change in AF for a specific cluster were replicated in the other clusters approximately 60% of the time. Heterogeneous changes in AF across clusters indicates genotypic redundancy whereby a variety of different genotype combinations provide different genetic solutions to the selection goals in a specific time-period.

Although we can only speculate on the reasons that may be involved in these non-parallel changes across clusters, two possible explanations would be differences in selection goals in different families and/or non-additive gene actions. The most prominent sires of families 1 and 2 are Planet and Goldwyn who are known to differ dramatically for fat and protein yield. The trajectory of the DGAT gene which has a significant genetic effect on fat and protein changes accordingly Figure 1.

Several known candidate genes identified by Ma *et al.* 2019 as having large AF in a very similar population were investigated. We also observed AF changes of a similar magnitude for the same identical alleles. However, their trajectories are not always the same for all families. For example, *ERBB4* (Chromosome 2) does show a shared and parallel change in AF across all families. Whereas the AF for *SPATA6* (chromosome 3) and *USP13* (chromosome 1) changed in a non-parallel way in the different families.

Population stratification caused by prominent sires allows for the investigation of alleles with strong LD (Barton, 2017). The fate of individual alleles is highly contingent on the allelic makeup at other loci. SNPs

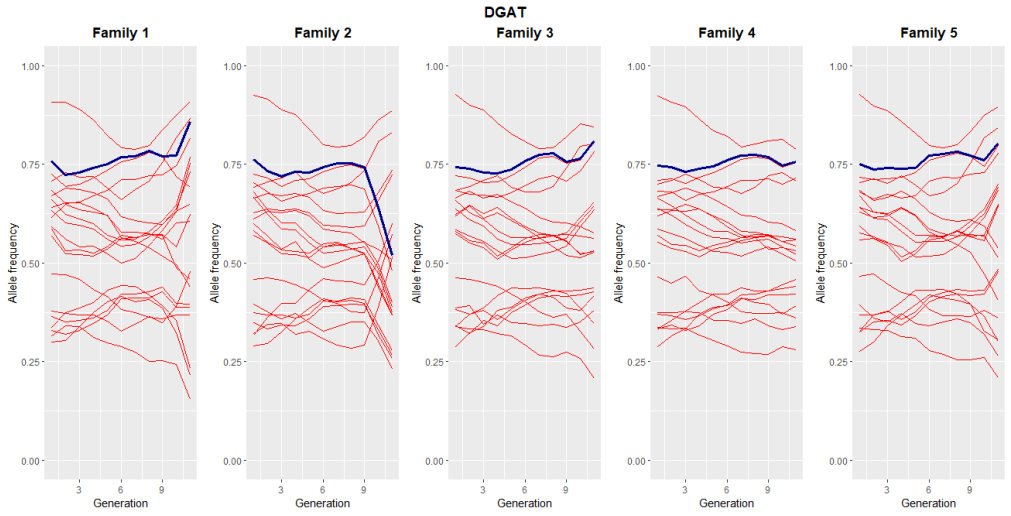


Figure 1. The allele frequency of the SNP that showed the *DGAT* gene (blue) and the surrounding 20 SNP markers (red) per generation within each family.

with an AF trajectory that includes a reversal in direction (or flip) are good candidates for the investigation of epistasis. Using the criteria of an overall reversal in direction of 0.1, we observed 11% or approximately 6,300 SNPs changing direction. In comparison, fixation of alleles was infrequent across the whole population (3 alleles). Number of alleles that became fixed within each family were 38, 22, 22, 59, and 40, respectively.

Conclusions

This study identified different subpopulations within the Holstein breed for a specific time-period. These clusters are associated with unique and identifiable families. AF changes between clusters involved rapid directional changes, dramatic reversals in direction, and a mixture of parallel and non-parallel responses. Heterogeneity in AF changes across families is an indicator of genotypic redundancy and a source of genetic variation.

References

Aguilar, I., and Misztal, I. (2008) J Dairy Sci 91(4):1669-1672. https://doi.org/10.3168/jds.2007-0575

Barghi, N, Tobler, R., Nolte, V., Jaksic, A.M., Mallard, F., *et al.* (2019) PLoS Biol. 17(2):e3000128. https://doi.org/10.1371/journal.pbio.3000128

Barton,N.H. (2017) Heredity (118):96–109. https://doi.org/10.1038/hdy.2016.109

Bonhomme, M., Chevalet, C., Servin, B., Boitard, S., Abdallah, J. *et al.* (2010) Genetics 186(1):241-262 https://doi.org/10.1534/genetics.110.117275

Calo, L.L., McDowell, R.E., VanVleck, L.D., and Miller, P.D. (1973) J Anim Sci. 37(3):676-682. https://doi.org/10.2527/jas1973.373676x

Duenk, P., Bijma, P., Calus, M.P.L, Wientjes, Y.C.J., and Van der Werf, J.H.J (2020) G3 10(2):783-795 https://doi.org/10.1534/g3.119.400663

Ma, L., Wiggans, G.R., Wang, S., Sonstegard, T.S., Yang, J., *et al.* (2012) BMC Genom. 20(1):128 https://doi.org/10.1186/1471-2164-13-536

Ma, L., Sonstegard, T.S., Cole, J.B, Sonstegard, T.S., VanTassell, C.P., Wiggans, G.R., *et al.* (2019) BMC Genom. 20(1):128 https://doi.org/10.1186/s12864-019-5459-x

Misztal, I., Tsuruta, S., Lourenco, D.A.L., Aguilar, I., Legarra, A. *et al.* (2014) Manual for BLUPF90 family of programs. Available at: http://nce.ads.uga.edu/wiki/lib/exe/fetch.php?media=blupf90_all2.pdf

212. Genome-wide evaluation of runs of homozygosity in Indian Sahiwal cattle

S.K. Illa[1], S. Mumtaz[2], S. Mukherjee[3], S. Nath[3] and A. Mukherjee[3*]

[1]Livestock Research Station, Garividi, Andhra Pradesh, 535101, India; [2]State Animal Husbandry, Kolkata, West Bengal, 700106, India; [3]ICAR-National Dairy Research Institute, Karnal, Haryana, 132001, India; writetoanupama@gmail.com

Abstract

Genome wide scan of ROH provide information on inbreeding, population history and genes under selection. The present study aimed: (1) to validate PLINK input settings and characterization of ROH patterns; (2) to assess the ROH islands to identify the major genes and gene hubs; and (3) to compare different inbreeding measures in the selected population in Sahiwal cattle. The study revealed optimum thresholds as gap (250 KB), density (120 KB/SNP), window-snp (10) and window-threshold (0.05) and based on this ROH patterns were assessed. The ROH islands revealed hub genes i.e. *ACAD11, RFX4*and *BANP* controlling various traits. The values of mean F_{PED}, F_{ROH}, F_{HOM}, F_{GRM} and $F_{GRM0.5}$ were 0.009, 0.091, -0.001, -0.104 and -0.009 respectively and the correlations ranged from -0.03 to 0.90among inbreeding measures. The ROH estimates helps in understanding the evolutionary mechanism and planning sustainable breeding programme in Sahiwal breed.

Introduction

The Sahiwal cattle was distributed around the globe especially the tropical countries and is known for their milk attributes (Ilatsia *et al.*, 2011). The Sahiwal herd was established at National Dairy Research Institute (NDRI)in 1955 and from 2010 genetic improvement and propagation of superior germplasm under All India Coordinated Research Projects (AICRP)was initiated. The selective breeding was followed since then in the herd and optimum level of inbreeding based on pedigree (<10%) was maintained. But due to incomplete and lack of ancestor information for few animals pedigree based method may not be accurate and it is important to estimate autozygosity at genomic level for genetic improvement in the herd. Thus, monitoring of genetic diversity and genomic inbreeding is a key for sustainable breeding programme in Sahiwal cattle.

ROH are the contiguous homozygous stretches of genotypes transmitted to offspring from parents through the common ancestor (Broman *et al.*, 1999; Gibson *et al.*, 2006). The ROH estimates elucidates about the population structure, demography history of individuals (Purfield *et al.*, 2012) estimate level of autozygosity and differentiate identical by state with descent (Curik *et al.*, 2014). Although many studies have been reported (Mukherjee *et al.* 2018; Martikainen *et al.*, 2020; Peripolli *et al.*, 2020),inconsistent choice of parameter settings makes difficult to compare the values across the studies(Meyermans *et al.*, 2020). Thus, present study was undertaken to develop guideline for uniform, reliable ROH parameter and its validation on SNP data prior to analysis. The study aimed: (1) to validate PLINK input settings and characterization of ROH patterns; (2) to assess the ROH islands to identify the major genes and gene hubs; (3) to compare different inbreeding measures in the selected population in Sahiwal cattle.

Materials & methods

The study was conducted on 193 Sahiwal cows from NDRI herd and pedigree based inbreeding coefficient was estimated. The equivalent complete generations (ECG) ranged from 1 to 5 and average ECG of individuals was 2.29±0.92, the inbreeding coefficients and the average relatedness were in the range of 0 and 0.06, 0.0002 and 0.032, respectively. The maximum 13 generations was found in the study. The

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_212, © S.K. Illa *et al.* 2022

Sahiwal cows were genotyped with a medium-density INDUSCHIP2, a customized Illumina Bovine SNP chip developed by National Dairy Development Board (NDDB), India (Annual Report, 2018) consisted of 52,363 SNPs with the approval by Institute Animal Ethics committee (NDRI approval 43-IAEC-18-8). The quality control of genotypic data with flag --geno 0.1 and --mind0.1 was performed by PLINK v1.9 (Purcell *et al.*, 2007) software. A total of 739 SNPs were removed; no individuals were excluded and remaining 49,697 SNPs were used for downstream analysis. The sliding window approach was used to validate the various parameters settings i.e. density, maximum gap (Kb), minimum SNP per window, minimum length (Kb),window size (SNP) and threshold on detection of ROH. Each parameter is validated by keeping other parameters as default and the genomic coverage percent was calculated for each validation. The genomic coverage percent is expressed as the per cent of largest ROH identified to the total length of genome (autosome) that was utilized in the validation of ROH analysis. Finally, the parameters with greater genome coverage percent were utilized in the estimation of ROH segments. The ROH segments were grouped into five classes based on the length: 0-2 Mb, 2-4 Mb, 4-8 Mb, 8-16 Mb and >16 Mb. The --homozyg-group function in PLINK was used to identify ROH islands shared among at least 25% of individuals. Candidate genes within ROH islands were ascertained using Ensembl Biomart tool (Haider *et al.*, 2009). Significant GO terms and KEGG pathways were obtained by DAVID (Huang*et al.,* 2009). The genomic inbreeding estimates F_{ROH}(based on ROH), F_{HOM} (ratio of observed to expected homozygous genotypes), F_{GRM}($F_{GRM} = 1/n \sum_{i=1}^{n} \frac{(x_i - 2p_i)^2}{h_i} - 1$), $F_{GRM0.5}$ (estimated from diagonal of GRM with allele frequencies as 0.5) (Bjelland *et al.*, 2013; Doekes *et al.*, 2019) and F_{PED} (pedigree based) were estimated and compared.

Results

In this study, observed better genome coverage at the density of 120 kb/SNP, maximum gap of 250 kb, window SNP of 10 with 0.05 threshold were identified (Table 1). The greater portion of genome consisted of medium size segments (2-4 Mb) with corresponding length as 37.70% whereas; with shorter and longer ROH segments was 0.19 and 5.68% respectively (Table 2). The genome coverage (%) due to longer ROH segments was higher than other segments.

Table 1. Validation of parameters for ROH analysis using 50K SNP array.

Parameter	Value	Genomic coverage %	Number	Average length
Maximum gap	250	3.64	8.49	9.68
Density	120	7.477	14.25	12.05
SNP	50	4.01	11.77	7.82
KB	30	3.45	7.34	10.62
--homozyg-window -snp	10	3.63	7.86	10.52
--homozyg-window -threshold	0.05	3.45	7.34	10.62

Table 2. Summary of runs of homozygosity, number and length (Mb) by ROH length class.

Class	n ROH	Percent	Cumulative (%)	Mean length	SD length	Total length	Genome coverage%
ROH_{0-2}	334	5.65	5.65	1,417.01	284.63	473,282.06	0.19
ROH_{2-4}	1,893	32.04	37.69	3,054.41	489.64	5,782,005.62	2.30
ROH_{4-8}	1,873	31.70	69.39	5,614.43	1,127.15	10,515,843.09	4.19
ROH_{8-16}	1,207	20.43	89.82	11,086.93	2,229.55	13,381,924.50	5.33
$ROH_{>16}$	600	10.18	100	23,776.74	11,679.50	14,266,045.53	5.68

The mean F_{PED}, F_{ROH}, F_{HOM}, F_{GRM} and $F_{GRM0.5}$ were 0.009, 0.091, -0.001, -0.104 and -0.009, respectively with coefficient of variation from35.73 to 65.01%. The high correlations found for F_{PED} with F_{HOM} (0.90); medium of F_{PED} with F_{GRM} (-0.43) and with $F_{GRM0.5}$ (0.21) (Table 3).

The most frequent ROH islands were located on BTA5 and BTA7containing 113 genes related with milk production, fertility, body conformation, stress tolerance and host immunity. The effect of various density values on genomic coverage revealed that genomic coverage spiked up significantly between 40 and 70 Kb/SNP of SNP density and in agreement to Meyermans *et al.* (2020). The maximum genome coverage achieved at 120 Kb/SNP in our study. It was challenging to obtain maximum genome coverage with higher gap threshold. The shorter segments (<1 Mb) couldn't be detected with default gap settings (1000 Kb). A gap of 250 Kb setting was found to be optimum for ROH analysis in medium dense array.

Discussion

The genomic coverage percent was considered as a best indicator for the validation procedure. Different parameters such as density (Kb/SNP), maximum gap (Kb), window size and threshold were validated. The maximum gap of 250 Kb parameter was better thresholds and confirms Meyermans *et al.* (2020) to skip pruning genotypes for LD and MAF as it would extremely affect ROH analysis with negligible gain in ROH detection.

In this study, the inbreeding was due to the presence of longer ROH segments that are contributed by recent ancestors. The recombination events were unable to break the segments that might result the longer segments (Kirin *et al.*, 2010). The similar results were reported in tropical cross bred cattle and mithun (Mukherjee *et al.*, 2018; Peripolli *et al.*, 2020). The low average F_{PED} and F_{ROH} values observed contrary to these Dixit *et al.* (2020) and Peripolli *et al.*(2020) reported higher estimates. The study revealed common ROH islands on chromosome 5 and 7(40% similarity) against earlier report of 92, 90 and 93%, respectively in Gyr, Piedmontese and other indicine cattle (Karimi, 2013; Marras *et al.*, 2015; Peripolli *et al.*, 2018). The *ACAD11* (Acyl-CoA dehydrogenase family member 11) has significant effects on average daily gain, dry matter intake and residual feed intake traits in cows. *RFX4* (Regulatory factor X4), located on the chromosome 5 is important and supposed to affect 21 different traits in dairy cattle (Hu *et al.*, 2019). *BANP* (BTG3 associated nuclear protein) gene is related with somatic cell score (*VT:1000072, LPT:1000072, CMO:0000793*) of dairy cows (Jiang *et al.*, 2019).This study provided the insights of the ROH role in detecting the selection foot prints and genomic inbreeding in the Sahiwal population. This is probably the first study to compare genomic inbreeding with pedigree based estimate with characterization of

Table 3. Inbreeding based on Pedigree(F_{PED})on ROH lengths(F_{ROH}), F_{HOM} and F_{GRM}.

F_x	Mean	Median	Minimum	Maximum	CV (%)	N
F_{PED}	0.009	0.000	0.000	0.062	12.19	187
$F_{ROH0-2\ Mb}$	0.0012	0.001	0.0003	0.003	55.50	151
$F_{ROH2-4\ Mb}$	0.011	0.011	0.001	0.024	35.73	192
$F_{ROH4-8\ Mb}$	0.023	0.023	0.001	0.048	40.59	183
$F_{ROH8-16\ Mb}$	0.029	0.029	0.003	0.079	49.57	180
$F_{ROH>16\ Mb}$	0.036	0.033	0.006	0.148	65.01	156
F_{ROH}	0.091	0.094	0.000	0.242	49.33	192
F_{HOM}	-0.001	0.029	-0.307	0.170	84.84	192
F_{GRM}	-0.104	-0.016	-0.289	0.892	2.09	192
$F_{GRM0.5}$	-0.009	-0.014	-0.116	0.155	5.79	192

runs of homozygosity in Sahiwal cattle. The understanding of demographic details and genetic control of performance traits will help in formulating breeding strategy for genetic improvement in Sahiwal cattle.

References

2018-19, *AnnualReport of NDDB*. *December*, 2–2. https://www.nddb.org/sites/default/files/NDDB_AR_2018_19_Eng.pdf

Bjelland D.W., Weigel K.A., Vukasinovic N., and Nkrumah J.D. (2013)J. Dairy Sci. 96(7):4697–4706. https://doi.org/10.3168/jds.2012-6435

Broman K.W., and Weber J.L. (1999)Am. J. Hum. Genet.65(6):1493–1500. https://doi.org/10.1086/302661

Curik I., Ferenčaković M., and Sölkner J. (2014)Livest. Sci.166(1):26–34. https://doi.org/10.1016/j.livsci.2014.05.034Dixit S.P., Singh S., Ganguly I., Kumar Bhatia A., Sharma A.,*et al.*(2020) Front. Genet. 11:92. https://doi.org/10.3389/fgene.2020.00092

Doekes H.P., Veerkamp R.F., Bijma P., de Jong G., andHiemstra S.J. (2019) Genet. Sel. Evol. 51:54. https://doi.org/10.1186/s12711-019-0497-z

Gibson J., Morton N. E., and Collins A. (2006)Hum. Mol. Genet. 15(5):789–795. https://doi.org/10.1093/hmg/ddi493

Haider S., Ballester B., Smedley D., Zhang J., Rice P.,*et al.*(2009)Nucleic Acids Res. 37(suppl_2): 23-27. https://doi.org/10.1093/nar/gkp265

Huang D.W., Sherman B.T., and Lempicki R.A. (2009) Nat. Protoc. 4:44-57. https://doi.org/10.1038/nprot.2008.211

Hu Z.L., ParkC.A.,and ReecyJ.M.(2019)Nucleic Acids Res.https://doi.org/10.1093/nar/gky1084

Ilatsia E.D., Roessler R., Kahi A.K., and Valle Zárate A. (2011) Animal Genetic Resources 49:65–74. https://doi.org/10.1017/s2078633611000336

Jiang J., Ma L., PrakapenkaD., VanRadenP.M., ColeJ.B., *et al.*(2019) Front. Genet.10:412. https://doi.org/10.3389/fgene.2019.00412

Karimi Z. (2013).*European Master in Animal Breeding and Genetics Runs*, *July*.

Kirin M., McQuillan R., Franklin C.S., Campbell H., Mckeigue P.M., *et al.* (2010). *PLoS ONE*, *5*(11), 1–7. https://doi.org/10.1371/journal.pone.0013996

Martikainen K., Koivula M., and Uimari P. (2020)Sci. Rep. 10(1):3804. https://doi.org/10.1038/s41598-020-60830-9

Meyermans R., Gorssen W., Buys N., and Janssens S. (2020)BMC Genom.21(1):94. https://doi.org/10.1186/s12864-020-6463-x

Mukherjee A., Mukherjee S., Dhakal R.,Longkumer I., Haque N., *et al.*(2018)Sci. Rep. 8(1):10316. https://doi.org/10.1038/s41598-018-28718-x

Peripolli E., Stafuzza N.B., Amorim S.T., de LemosM.V.A., Grigoletto L.,*et al.*(2020). J. Anim. Breed. Genet. 137(2):155–165. https://doi.org/10.1111/jbg.12428

Peripolli E., Stafuzza N.B., Munari D.P., Lima A.L.F., Irgang R.,*et al.*(2018)BMC Genom.19(1):34. https://doi.org/10.1186/s12864-017-4365-3

Purcell S., Neale B., Todd-Brown K., Thomas L., Ferreira M.A.R.,*et al.*(2007)Am. J. Hum. Genet.81(3):559–575. https://doi.org/10.1086/519795.

Purfield D.C., Berry D.P., McParland S., and Bradley D.G. (2012)BMC Genet.13:70. https://doi.org/10.1186/1471-2156-13-70

213. Following the hidden footsteps of past human migrations via the goat genome

A. Manunza[1*], J. Ramirez Diaz[1,2], T. Bobbo[1,3], P. Cozzi[1], B. Lazzari[1], F. Biscarini[1], S. Biffani[1], A.M. Johansson[4], L. Grøva[5] and A. Stella[1]

[1]Institute of Agricultural Biology and Biotechnology, National Research Council, via Edoardo Bassini, 15, 20133 Milan, Italy; [2]Università Cattolica del Sacro Cuore, Via Emilia Parmense 84, 29122 Piacenza, Italy; [3]University of Milan, Via Celoria 2, Milan, Italy; [4]Department of Animal Breeding and Genetics, Swedish University of Agricultural Sciences, Box 7023, 75007 Uppsala, Sweden; [5]Norwegian Institute of Bioeconomy Research, Gunnars vei 6, 6630 Tingvoll, Norway; arianna.manunza@ibba.cnr.it

Abstract

The aim of this study was to assess the genetic diversity and the degree of introgression of foreign alleles into autochthonous goat genomes following putative land and maritime routes of human migrations. Genotypic data from sixteen populations were analysed for diversity statistics and admixture. Results show genetic fingerprints of old gene flow that could be related to past human expansions for explorations, raids or demographic reasons. In addition, this study could add missing pieces to indirectly recreate historically and still uncertain human land and maritime routes.

Introduction

Archaeological and historical sources have driven the reconstruction of human migrations. Recently, cutting-edge genomics and bioinformatics analysis has contributed to this field. Since their domestication, small ruminants constituted an irreplaceable source of food and ecosystem services for humans and their small size made them ideal for carrying them by land and sea, following human migratory routes (Amills *et al.*, 2017). A deeper analysis of literature (Colli *et al.* 2018, Cardoso *et al.*, 2018), investigating the genetic diversity in several breeds from the five continents, led us to catch subtle but intriguing similarities among geographically distant populations. In this study, we aimed to infer signs of gene flow following putative routes of human migrations by analysing the genomes of native modern goat populations. In particular, we focused on two prolonged migration events that occurred between the 500 BCE and the 900 CE: the Celtic expansion from Continental Europe and the Viking diaspora from Scandinavia (Viking Age, VA).

Materials & methods

Our dataset includes data from Stella *et al.* (2018): ARR = Traditional Aran, BLB = Bilberry and OIG = Old Irish (Ireland), BAG = Bagot and OEG = Old English (UK), DNK = Danish Landrace (Denmark), FIN = Finnish landrace (Finland), ICL = Icelandic landrace (Iceland), NLD = Dutch Landrace (the Netherlands), NRW =Norwegian Landrace (from Tanzania); from Berg *et al.* (2020) (SKO = Skorpa coastal goat, SEL = Sejle coastal goat) (Norway); from Cortellari *et al.* (2021) (ORO = Orobica, NVE = Nera Verzasca and VLS = Vallesana (northern Italy); from Smarter project, https://www.smarterproject.eu/) SWE = Swedish landrace (Sweden). Genotype data were from Illumina GoatSNP50 BeadChip (Tosser-Klopp *et al.*, 2014). Data were remapped against ARS1.2 assembly. Quality control was performed using PLINK v1.9 (Chang *et al.* 2015). We used Arlequin v.3.5 (Excoffier *et al.*, 2010) to calculate the pairwise Fst values and the SNPrelate R package (Zheng *et al.*, 2012) to perform the Principal Component Analysis (PCA). We inferred the estimation of individual ancestries using ADMIXTURE v.1.3 with K ranging from 2 to 30 (Alexander *et al.*, 2009). To make a comparative dataset considering the maternal contributions, we selected from GenBank the HV1 control regions (392 bp) of 63 domestic goats (Scandinavia, British Isles and northern Italy). A Maximum-Likelihood (ML) phylogeny analyses using Tamura-Nei and 1000 bootstrap was carried out with MEGA v.11 (Tamura *et al.*, 2021) and the resulting tree was drawn using iTOL v.6.4.3 program (Letunic and Bork, 2021).

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_213, © A. Manunza *et al.* 2022

Results

F-statistics highlighted a surprisingly low genetic diversity that characterizes some populations, considering a European continental context. Fst ranged from 0.071 to 0.621. The NVE breed shows the lowest Fst values in all pairwise comparisons and a close similarity with all populations from Scandinavia (above all DNK, 0.096) and secondarily with the British Isles (0.143-0.193). The ICL breed differentiates in a well-defined cluster (ADMIXTURE and PCA analysis). It presents Fst values a little lower of 0.3 with the Scandinavian breeds, definitely a lower value in comparison with British breeds (0.6) and slightly higher with Italian breeds (0.4). The breeds from the British Isles are quite similar to each other but show a lower genetic distance with DNK and NRW (0.159-0.208) followed by the Italian populations (0.184-0.217). The cross-validation method in the ADMIXTURE (Figure 1) analysis indicates K=19 as the most probable number of ancestries. Figure 1A shows a significant level of admixture in most of breeds, except for ICL, ARR, SKO and NLD. A small introgression level is detected among some populations from Scandinavia, British Isles and Italy. The genetic clustering analysis by PCA (Figure 2) groups the populations in three clusters: Icelandic breed is closer to Scandinavian breeds; British Iles forms a unique cluster with Italian breed near to NLD, whereas DNK is a bit far from the rest making a single cluster Also, we can see a certain level of overlapping between NVE with BLB and among SKO, SEL and SWE (dotted circle in the Figure 2).

The control region tree depicts an interesting link across populations. Individuals are coloured depending on their geographical origin and the groups of interest are highlighted in different colours. Six well-established groups are identified, whose nodes are supported by a bootstrap value around 50% or higher. mtDNA analysis confirms a certain degree of past introgression between NVE and DNK, Norwegian

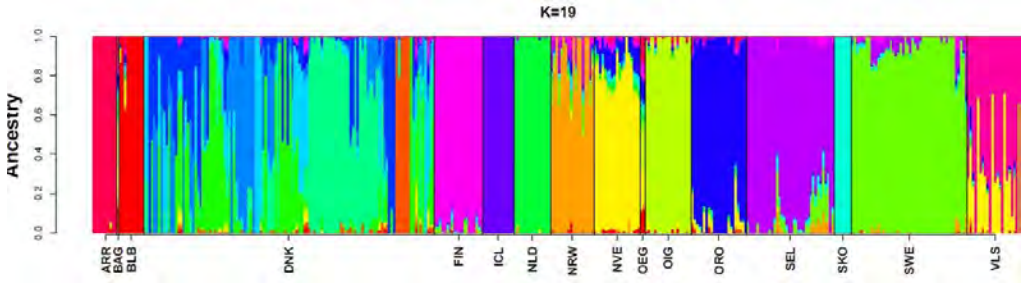


Figure 1. Individual ancestries inferred by ADMXTURE analysis: different colours refer to different ancestries.

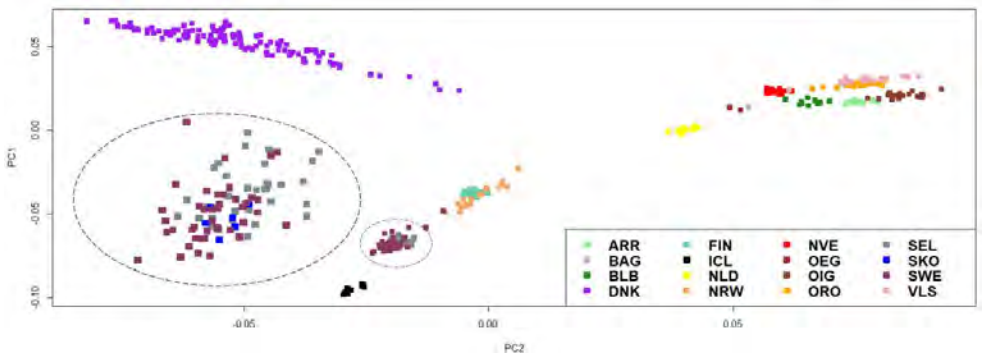


Figure 2. PCA analysis: the first two PCs explain the 10.3% of variance.

with British Iles individuals (with historical samples in the same branches) and Icelandic with NVE. The ancient samples from the British Isles are also clustering together. H= historical and M=modern prefix ID of samples as in Cassidy *et al.* (2017).

Discussion

All our findings suggest that crossing and introgression events among these native breeds occurred in the past, possibly following human migrations. The BLB breed for example, seems to have contributed to the genetic composition of both Italian and Scandinavian breeds, during two different and temporally separated contacts or through a slow but continued genetic exchange. We can speculate that the foreign fingerprint found in these populations is traceable back to the Celtic migration from central Europe towards Ireland and Britain and later with the Viking diaspora, spanning from the 500 BCE to the 900 CE. A previous study on the historical gene flow based on mtDNA of British Isles breeds (Cassidy *et al.*, 2017) found that ancient samples from the islands of Scotland clustered together with modern Irish goats and a Norwegian landrace goat. Another interesting grouping contains a modern British goat from central-east Ireland alongside Welsh, English and Icelandic goats. Both results suggest a genetic link with native goats from Scandinavian countries and settlements. In our mtDNA analyses, we included more individuals from Scandinavian countries, Iceland and northern Italy. It is noteworthy that a deeper inspection of the inferred ancestries and the goat haplotypes distributions reveals additional similarity between Scandinavian breeds and the rest of breeds. This suggests a bidirectional introgression and that a common situation has emerged in the genetic composition of modern Scandinavian and British Isles human populations. In fact, two studies aiming to explore the genomic history of the Celtic insular origin and the VA (Byrne *et al.*, 2018; Margaryan *et al.*, 2020), showed that the genetic composition of modern individuals from Scandinavia and Ireland exhibited different degrees of admixture that mirror the most famous expansion routes (Norwegian-like haplotypes into peoples from British Isles and Iceland, Danish-like haplotypes into Britain). In addition, several individuals with large amounts of South European ancestry in Denmark and southwest Sweden were retrieved during the VA, supporting our findings. Considering these results in the context of historical recorded raids and human demographic movements, we can postulate that the genome

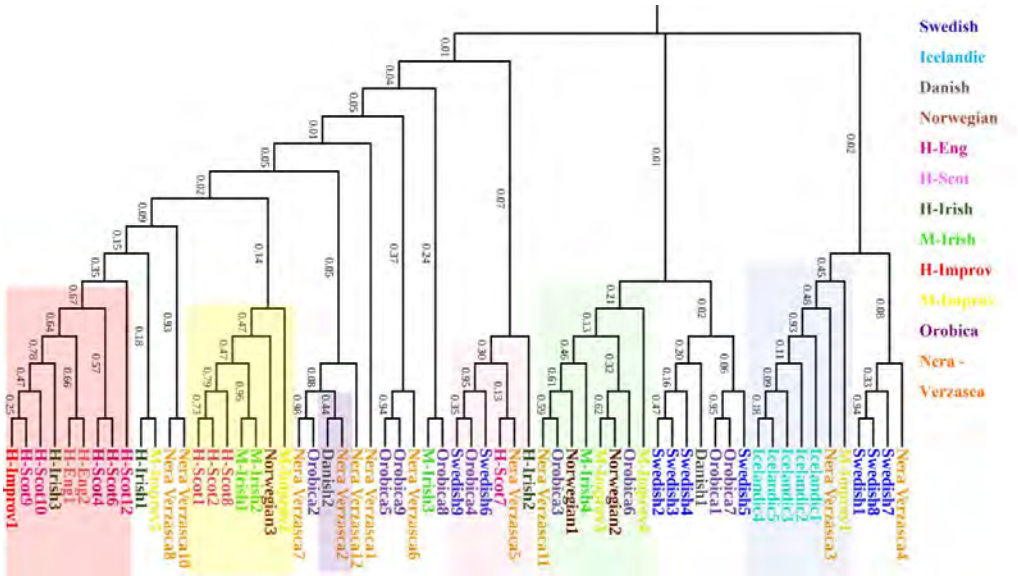


Figure 3. Phylogeographic analysis of 63 control regions of native goat breeds.

of modern indigenous goats could retain clues of old foreign genetic contribution as a consequence of an extensive trade and transport of goats in their history.

Acknowledgements

This project was funded by the European Union's Horizon 2020 research and innovation program under grant agreement No 772787 (SMARTER).

References

Alexander D.H., Novembre J., Lange K. (2009) Genome Res. 19:1655–1664. https://doi.org/10.1101/gr.094052.109.

Amills M., Capote J., Tosser-Klopp. (2017) Anim. Genet. 48(6):631-644. https://doi.org/10.1111/age.12598.

Berg P., Groeneveld L.F., Brekke C., Våge D.I., Sørheim K.M. *et al.* (2020) Acta Agric. Scand. 69:47-52. https://doi.org/10.1080/09064702.2020.1729852.

Byrne R.P., G., *et al.* (2018) PLoS Genet. 14(1): e1007152. https://doi.org/10.1371/journal.pgen.1007152.

Cardoso T.F., *et al.* (2018) Genet. Sel. Evol.50, 56. https://doi.org/10.1186/s12711-018-0425-7.

Cassidy L.M., *et al.* (2017) Biol. Lett.13(3): 20160876.https://doi.org/10.1098/rsbl.2016.0876.

Chang C. C., Chow C. C., Tellier LCAM, Vattikuti S., Purcell S.M., *et al.* (2015) GigaScience 4(1): s13742-015-0047-8. https://doi.org/10.1186/s13742-015-0047-8.

Colli L., Milanesi M., Talenti A., Bertolini F., Chen M. *et al.* (2018) Genet Sel Evol 50, 58 https://doi.org/10.1186/s12711-018-0422-x

Cortellari M., Barbato M., Talenti A., Bionda A, Carta A., *et al.* (2021) Sci. Rep. 11, 19042. https://doi.org/10.1038/s41598-021-89900-2.

Excoffier L., Lischer H.E.L. (2010) Mol. Ecol. Resour. 10: 564–567.

Letunic I., Bork P. (2021) Nucleic Acids Res. 49(W1):W293–W296. https://doi.org/10.1093/nar/gkab301.

Margaryan A., *et al.* (2020) Nature 585, 390–396. https://doi.org/10.1038/s41586-020-2688-8.

Stella A., *et al.* (2018) Genet. Sel. Evol. 50, 61. https://doi.org/10.1186/s12711-018-0427-5.

Tamura K., Stecher G., Kumar S. (2021) Mol. Biol. Evol. 38(7):3022–3027. https://doi.org/10.1093/molbev/msab120.

Tosser-Klopp G., Bardou P., Bouchez O., Cabau C., Crooijmans R. *et al.* (2014) PLoS One 9(3): e86227. https://doi.org/10.1371/journal.pone.0086227.

Zheng, X., Levine D., Shen J., Gogarten S.M., Laurie C., *et al.* (2012) Bioinformatics 28(24):3326–3328. https://doi.org/10.1093/bioinformatics/bts606.

214. Are autozygosity estimates from imputed genotypes reliable?

M. Ferenčaković[1], I. Curik[1], J.F. Garcia[2], J. Sölkner3 and R. Carvalheiro[4,5*]

[1]Department of Animal Science, Faculty of Agriculture, University of Zagreb, Svetošimunska 25, 10000 Zagreb, Croatia; [2]School of Veterinary Medicine, Sao Paulo State University (UNESP), R. Clóvis Pestana 793, 16050-680, Araçatuba, SP, Brazil; [3]Department of Sustainable Agricultural Systems, Division of Livestock Sciences, University of Natural Resources and Life Sciences, Gregor Mendel Str. 33, 1180 Vienna, Austria; [4]School of Agricultural and Veterinarian Sciences, Sao Paulo State University (UNESP), Via de Acesso Prof. Paulo Donato Castelane s/n, 14884-900, Jaboticabal, SP, Brazil; [5]National Council for Science and Technological Development, 71605-001, Brasilia, DF, Brazil; rcar@fcav.unesp.br

Abstract

Estimates of the inbreeding coefficient based on runs of homozygosity (F_{ROH}) are widely used in population genetics and livestock studies. The accuracy of F_{ROH} estimates is determined in part by the density of the genotyping chip. Compared with high density (HD) chips, low density (LD) chips tend to overestimate F_{ROH} in shorter homozygous segments. It is unclear whether imputation of genotypes from LD to HD would reduce or eliminate this overestimation. Our goal was to determine whether F_{ROH} calculations based on imputed LD genotypes are accurate. To this end, F_{ROH} estimates were determined for Nelore cows using HD or LD imputed genotypes (F_{ROH_HD} and F_{ROH_LDimp}). Our results showed that LD imputed genotypes reduced but did not eliminate differences between estimates. Differences were observed in both directions, with under- and overestimates of F_{ROH}. This suggests that imputation errors can either disrupt or generate homozygous segments.

Introduction

Runs of homozygosity (ROH), which are continuous segments of homozygous genotypes, have a variety of applications in population genetics and livestock breeding (McQuillan *et al.*, 2008; Curik *et al.*, 2014). Because they do not rely on 'complete' and correct pedigree information, are not based on expected values, and account for recent and remote inbreeding depending on segment size, inbreeding coefficient estimates based on ROH (F_{ROH}) may be expected to estimate autozygosity more precisely than traditional inbreeding coefficient estimates derived from pedigree data (Keller *et al.*, 2011; Peripolli *et al.*, 2016).

The density of the SNP chip influences the quality and precision of F_{ROH} estimates, as low density (LD) chips tend to overestimate F_{ROH} in shorter ROH segments compared to high density (HD) chips (Ferenčaković *et al.*, 2013; Hillestad *et al.,* 2017; Ceballos *et al.,* 2018). To reduce genotyping costs, it is common practice in animal breeding to genotype a portion of the individuals with an LD chip and impute their missing HD genotypes using information from a reference population genotyped with an HD chip (Khatkar *et al.*, 2012). Although this strategy has been shown to be cost-effective in providing accurate genomic predictions, among other applications, it is unknown whether F_{ROH} estimates from imputed genotypes are reliable. The goal of this study was to compare F_{ROH} estimates based on HD and LD imputed genotypes using Nelore (Bos indicus) cattle genotypic information.

Materials & methods

F_{ROH} estimates were first calculated for 1,221 Nelore cows genotyped with the Illumina® BovineHD chip (HD). These cows were born in six different herds between 1993 and 2008 and were part of the DeltaGen® breeding program's genomic selection reference population (Carvalheiro *et al.,* 2014). Only SNPs from autosomes and with call-rates greater than 0.95 were considered, totaling 660,297 SNPs. The proportion of the autosomal genome covered by ROH segments was calculated as F_{ROH} (McQuillan *et al.,* 2008).

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_214, © M. Ferenčaković *et al.* 2022

ROH were identified using the SNP & Variation Suite v.8.9.0 (Golden Helix, Bozeman, MT, USA) under the following conditions: 15 or more consecutive homozygous SNPs with gaps of no more than 1000 kb between SNPs, at a density of at least 1 SNP every 100 kb. Due to the possibility of genotyping errors, a limited number of heterozygous and missing SNPs were permitted within ROH segments, with the number varying according to the length of the ROH segment (Ferenčaković *et al.*, 2013).

To assess the quality of autozygosity estimates from imputed genotypes, F_{ROH} was recalculated assuming that cows were genotyped with LD chips and their genotypes were imputed on HD. To do this, partial information from the HD chip was used to simulate the possibility that cows were genotyped with the following LD chips: Illumina® BovineLD (7k), Illumina® BovineSNP50 v2 (50k) and two other simulated/customized LD chips (21k and 52k) tested in the imputation study conducted by Carvalheiro *et al.* (2014). The number of SNPs from each LD chip corresponded to 6,848 (7k), 39,642 (50k), 21,570 (21k) and 52,590 (52k). The imputation process from LD to HD was performed using FImpute v.2.2 software (Sargolzaei *et al.,* 2014) and HD genotypes from 994 influential Nelore bulls as reference population.

Finally, for each cow, differences in F_{ROH} estimates obtained with the HD (F_{ROH_HD}) and imputed LD (F_{ROH_LDimp}) genotypes were computed, taking into account overall ROH segment lengths (>1 Mb) and specific ROH segment lengths (1-2 Mb, 2-4 Mb, 4-8 Mb, 8-16 Mb, and >16 Mb). F_{ROH_HD} was also compared to FROH from 50k that was not imputed to HD (F_{ROH_50k}).

Results & discussion

For the studied Nelore cattle population, the mean, median, minimum, and maximum values of overall (>1 Mb) F_{ROH_HD} were 0.0516, 0.0497, 0.0001, and 0.2380, respectively. Zavarez *et al.* (2015) present a more detailed description of this population's ROH pattern. The mean value of F_{ROH_50k} was 0.2843, indicating that autozygosity was overestimated when compared to F_{ROH_HD}. Ferenčaković *et al.* (2013) discovered a similar pattern in taurine breed cattle. This overestimation occurred primarily at smaller (1-2 Mb) ROH (Table 1), indicating that the 50k chip lacks sufficient coverage in these small segments to detect heterozygous SNPs, which the HD chip detects. The greatest autozygosity estimate difference was found in a cow with F_{ROH_HD} and F_{ROH_50k} values of 0.0268 and 0.3347, respectively.

Imputed genotypes provided F_{ROH} estimates that were comparable to HD genotypes on average. In comparison to F_{ROH_HD}, the 7k imputed genotypes had a slight overestimation of F_{ROH}, with a mean F_{ROH} of 0.0640. The remaining LD imputed genotypes had mean F_{ROH} values of 0.0534 (21k), 0.0546 (50k), and 0.0526 (52k), all of which were very close to the mean value of F_{ROH_HD} (0.0516). Aside from having comparable average F_{ROH} values, some cows had differences in F_{ROH_HD} and F_{ROH_LDimp} estimates. The minimum and maximum differences shown in Table 1 indicate that the discrepancies occurred in both directions, implying that the imputed LD chips underestimated or overestimated the autozygosity of some cows when compared to F_{ROH_HD}. Increasing the SNP chip density from 7k to 21k, 50k, or 52k reduced the percentage of cows with a discrepancy between F_{ROH_HD} and F_{ROH_LDimp} estimates but had almost no effect on the range between minimal and maximal F_{ROH} estimate differences (Table 1).

It is critical to note that the imputation accuracies were high, with average values of 0.9268 (7k), 0.9826 (21k), 0.9761 (50k), and 0.9909 (52k). A cow's minimum imputation accuracy was 0.8077 (7k), 0.9186 (21k), 0.9002 (50k), and 0.9486. (52k). There was no correlation found between imputation accuracy and differences between F_{ROH_HD} and F_{ROH_LDimp} estimates, indicating that imputation errors could either break or create ROH segments (data not shown). We were unable to predict or diagnose individuals with differing F_{ROH_LDimp} compared to F_{ROH_HD} because no pattern of differences was observed. Additional analyses will be carried out in order to determine the source of these discrepancies. In conclusion, our findings show

Table 1. Summary statistics of the differences[1] of autozygosity estimates obtained with the HD chip (F_{ROH_HD}) and lower density chips[2] (F_{ROH_LD}), for different ROH segment lengths.

	Statistic[3]	50k	$7k_{imp}$	$21k_{imp}$	$50k_{imp}$	$52k_{imp}$
$dF_{ROH\ >1Mb}$	mean	0.2327	0.0123	0.0018	0.0029	0.0010
	min	0.0415	-0.1217	-0.1258	-0.1241	-0.1244
	max	0.3079	0.1171	0.0954	0.0972	0.0928
	%dif>0.01	100.00	57.25	20.15	19.82	20.07
	correlation	0.6694	0.6786	0.7773	0.7746	0.7806
$dF_{ROH\ 1-2Mb}$	mean	0.1263	0.0003	-0.0011	-0.0007	-0.0016
	min	0.0368	-0.0675	-0.0676	-0.0654	-0.0671
	max	0.1632	0.0191	0.0199	0.0211	0.0209
	%dif>0.01	100.00	8.11	8.27	7.94	7.86
	correlation	0.1888	0.1947	0.3382	0.3134	0.3904
$dF_{ROH\ 2-4Mb}$	Mean	0.0454	0.0019	-0.0010	-0.0004	-0.0011
	Min	0.0043	-0.0419	-0.0396	-0.0399	-0.0396
	Max	0.0734	0.0206	0.0197	0.0193	0.0187
	%dif>0.01	99.59	11.14	6.96	6.80	6.31
	correlation	0.4014	0.2809	0.5175	0.4728	0.5312
$dF_{ROH\ 4-8Mb}$	mean	0.0070	0.0031	0.0002	0.0003	0.0002
	min	-0.0306	-0.0371	-0.0500	-0.0483	-0.0508
	max	0.0383	0.0314	0.0276	0.0254	0.0275
	%dif>0.01	29.07	18.84	8.19	7.45	7.21
	correlation	0.6671	0.4141	0.6478	0.6279	0.6705
$dF_{ROH\ 8-16Mb}$	mean	0.0514	0.0035	0.0014	0.0013	0.0013
	min	-0.0002	-0.0530	-0.0542	-0.0517	-0.0516
	max	0.0987	0.0390	0.0339	0.0359	0.0339
	%dif>0.01	97.13	19.82	8.76	10.07	8.85
	correlation	0.4858	0.5485	0.7294	0.6978	0.7435
$dF_{ROH\ >16Mb}$	mean	0.0026	0.0036	0.0023	0.0024	0.0022
	min	-0.0182	-0.0133	-0.0121	-0.0105	-0.0120
	max	0.0606	0.1155	0.1035	0.1072	0.1039
	%dif>0.01	8.60	12.78	7.29	7.21	6.55
	correlation	0.8713	0.7842	0.8360	0.8374	0.8413

[1] $dF_{ROH} = F_{ROH_LD} - F_{ROH_HD}$.

[2] 50k (not imputed); and 7k ($7k_{imp}$), 21k ($21k_{imp}$), 50k ($50k_{imp}$) and 52k ($52k_{imp}$) imputed to HD.

[3] min: minimum; max: maximum; %dif>0.01: percentage of abs ($F_{ROH_LD} - F_{ROH_HD}$)>0.01, corr: correlation between F_{ROH} estimates.

that F_{ROH} estimates from imputed genotypes are, on average, comparable to F_{ROH} estimates based on HD chip. Nonetheless, some animals had very different autozygosity estimates when comparing F_{ROH_HD} and F_{ROH_LDimp}. Despite having the potential to influence selection and mating decisions, it is unclear to what extent these differences influence the results of other studies, such as those related to genomic inbreeding depression (Stoffel *et al.*, 2021) or the detection of signals of positive selection (Cheruiyot *et al.*, 2021). Further research into this subject is advised. While the reasons for the differences in F_{ROH} estimates remain unknown, higher density LD chips (e.g. ≥21k) should be preferred over lower density LD chips (e.g. 7k) to obtain more reliable F_{ROH} estimates based on imputed genotypes if HD genotyping of all individuals is not possible.

This study was supported by the ANAGRAMS-IP-2018-01-8708.

References

Browning, B. & Browning, S. (2009) Am J Hum Genet 84(2):210-223. https://doi.org/10.1016/j.ajhg.2009.01.005

Carvalheiro, R., Boison, S.A., Neves, H.H.R., Sargolzaei, M., *et al.* (2014) Genet Sel Evol 46:69. https://doi.org/10.1186/s12711-014-0069-1

Ceballos, F., Joshi, P., Clark, D., *et al.* (2018) Nat Rev Genet 19, 220–234. https://doi.org/10.1038/nrg.2017.109

Cheruiyot, E.K., Haile-Mariam, M., Cocks, B.G., *et al.* (2021) Sci Rep 11:16619. https://doi.org/10.1038/s41598-021-95816-8

Curik, I., Ferenčaković, M & Sölkner, J. (2014) Livest Sci 166:26-34. https://doi.org/10.1016/j.livsci.2014.05.034

Ferenčaković, M., Curik, I & Sölkner, J. (2013) Genet Sel Evol 45:42. https://doi.org/10.1186/1297-9686-45-42

Hillestad, B., Woolliams, J. A., Boison, S. A., Grove, H., Meuwissen, T., *et al.* (2017) Acta Agric. Scand. A Anim. Sci., 67(3-4):107-116. https://doi.org/10.1080/09064702.2018.1501088

Keller, M., Visscher, P & Goddard, M. (2011) Genetics 189(1):237–249. https://doi.org/10.1534/genetics.111.130922

Khatkar, M.S., Moser, G., Hayes, B.J. & Raadsma, H.W. (2012). BMC Genomics 13:538. https://doi.org/10.1186/1471-2164-13-538

McQuillan, R., Leutenegger, A., Abdel-Rahman, R., Franklin, C., *et al.* (2008) Am J Hum Genet, 83(3):359–372. https://doi.org/10.1016/j.ajhg.2008.08.007

Peripolli, E., Munari, D.P., Silva, M.V.G.B., Lima, A.L.F. *et al.* (2016) Anim Genet 48:255–271. https://doi.org/10.1111/age.12526

Sargolzaei, M., Chesnais, J.P. & Schenkel, F.S. (2014) BMC Genomics 15:478. https://doi.org/10.1186/1471-2164-15-478

Stoffel, M.A., Johnston, S.E., Pilkington, J.G. *et al.* (2021). Nat Commun 12: 2972. https://doi.org/10.1038/s41467-021-23222-9

Zavarez, L.B., Utsunomiya, Y.T., Carmo, A.S., Neves, H.H.R., *et al.* (2015) Front Genet 6:5. https://doi.org/10.3389/fgene.2015.00005

215. Genetic diversity and population structure for resistance to mastitis in Braunvieh cattle

M.Z. Trujano-Chavez[1*], R. Sánchez-Ramos[2], P. Pérez-Rodríguez[3] and A. Ruíz-Flores[1]

[1]Posgrado en Producción Animal, Universidad Autónoma Chapingo, Carretera Federal México-Texcoco Km 38.5, 56227 Texcoco, México; [2]Recursos Genéticos y Productividad, Colegio de Postgraduados, Carretera Federal México-Texcoco Km 36.5, 56230 Texcoco, México; [3]Socio Economía Estadística e Informática-Estadística, Colegio de Postgraduados, Carretera Federal México-Texcoco Km 36.5, 56230 Texcoco, México; zulealizee@gmail.com

Abstract

Resistance to mastitis is a difficult trait to improve due to its multifactorial causes. The objective was to characterize a Mexican Braunvieh cattle population for genetic resistance to mastitis. We used 22 SNP markers for 19 candidate genes in 150 animals. Average heterozygosity was 0.468±0.034, a value higher than those reported for other European breeds. The inbreeding coefficient was low, 0.057±0.066, similar to those reported for tropical breeds. The genotypic profiles for the candidate loci per K-means group were obtained, as well as the group's distribution through the graphics of the principal component analysis. The genotypic profiles showed high genetic diversity among groups. There are highly heterozygous groups, a favourable indicator of the presence of genetic diversity. The results of this study constitute evidence of the genetic potential of the Mexican Braunvieh population to improve resistance to mastitis.

Introduction

Resistance to mastitis is a complex trait that is multifactorial in nature and, thus, difficult to improve. According to Haugaard *et al.* (2013), the genomic prediction ability for clinical mastitis is low, 0.19, compared to that reported by Brunes *et al.* (2021) for traits like dry matter intake (0.52±0.04). The conventional selection scheme for resistance has not been effective against mastitis due to the large number of candidate genes related to several non-immune factors that influence it (Meier *et al.*, 2020), and the low heritability reported for the resistance to mastitis (Kašná *et al.*, 2018). In addition to the control of environmental factors to prevent it at the farm level, genomic methods have been used as an alternative solution to the problem of increasing resistance to mastitis. The aim is to obtain animals that are immune to the main mastitis-causing pathogens and have the ability that allow them to adapt to the environmental factors that cause mastitis (Suleiman *et al.*, 2018).

According to Ali *et al.* (2019), manipulation of genetic variation is key to the development and use of any genetic improvement program for dairy and beef cattle breeds. Genetic diversity has been studied in the original Braunvieh and derivative breeds aiming to disentangle the genetic basis of their adaptation to diverse environments and increasing productive performance. The objective of this study was to determine the genetic diversity and structure of a Braunvieh population in terms of mastitis resistance.

Materials & methods

Source of information, quality control and candidate markers. Genomic information was obtained using hair samples from 150 animals born between 2001 and 2016 on five farms of the Mexican Association of Braunvieh Purebred Breeders. The samples were genotyped at GeneSeek (Lincoln, NE, USA, http://genomics.neogene.com) using the Genomic Profile Bovine LD chip with 50,000 SNP markers. The missing genotypes were imputed using the observed allele frequencies. Additionally, a 0.05 threshold for the minor allele frequency was considered. The candidate genes associated with resistance to mastitis in *Bos taurus* were identified (197), according to literature reports. The position ranked by base pairs were obtained from

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_215, © M.Z. Trujano-Chavez *et al.* 2022

the Gene Library (Gene, 2021). All the analyses were carried out using the R software, version 4.0.4 (R Core Team, 2021). With this information, the intragenic loci were searched (±25 kb), 22 markers associated with resistance to mastitis were found and the 10% most informative markers were determined using the Shannon index (Sherwin, 2010).

Population genetic structure and multivariate analyses. Expected (He) and observed (Ho) heterozygosity, and Hardy-Weinberg equilibrium were estimated using the adegenet package (Jombart and Ahmed, 2011). The inbreeding coefficients by individual with respect to the population (F_{IT}) were calculated with adegenet and a histogram including a nonparametric estimate of the density function was obtained (Jombart and Ahmed, 2011). The differences in the genotypic profile of individuals were identified with the K-means algorithm included in the stats and factoextra packages in R (Kassambara and Mundt, 2020). The optimum number of groups to be formed was obtained with Silhouette and Elbow of factoextra. The principal component analysis (PCA) was performed using adegenet package (Jombart and Ahmed, 2011) using the sample covariance matrix obtained with the candidate SNP. The groups of K-means were used to identify divergent genotypes among groups of animals and to characterize genetic diversity based on the PCA.

Results

Candidate markers and population genetic structure. Table 1 shows the markers obtained after filtering with the Shannon Index. Their candidate genes, chromosome, and the number assigned to visualize the results in the graphs are shown. For resistance to mastitis, 19 genes with 22 loci were obtained. All loci were in Hardy-Weinberg Equilibrium.

The He and Ho were 0.495±0.006 and 0.468±0.034, respectively. The *P*-value with the Bonferroni correction showed non-significant difference between He and Ho. The genomic coefficient of inbreeding was 0.057±0.066. Figure 1 shows the distribution of the animals according to their individual F_{IT}.

Clustering and principal component analysis. The optimum number of groups to equilibrate the silhouette width and the means of squared errors for K-means converged at 11 groups, Figure 2. Silhouette width and the average proportion of variability among groups relative to the total variability (Elbow method) were 0.08 and 34.9%, respectively.

Discussion

In México, studies using genomic information from large animal populations are scarce. Usually, the main restriction is the availability of economic resources for animal genotyping. (Mrode *et al.*, 2019). Despite having used only 150 animals in our study, the sample size was enough to detect significant differences with an effective size of 0.4 (*P*=0.05). The power of 0.87 for the Xi^2 test obtained in the present study is adequate to obtain conclusive, reliable results (Khalilzadeh and Tasci, 2017).The estimates for Ho obtained in our study for mastitis (average 0.468±0.034) are above those obtained in breeds derived from the original Braunvieh, but below those reported for some *Bos taurus* breeds, with adaptation to the tropics. In Colombia, Creole cattle (Ocampo *et al.*, 2021) reported Ho of 0.66. This Ho estimate are quite different from those reported by (Mocarelli *et al.*, 2020), who found an average of 0.35±0.167 for Ho in populations related to Braunvieh. This suggests that the Mexican Braunvieh population studied has increased its genetic diversity. The overall average for F_{IT} was 0.057±0.066 for resistance to mastitis. This value is similar to those found (0.07-0.09) for the African tropical breeds King *et al.* (2021) and Creole Colombian cattle (Ocampo *et al.*, 2021). In contrast, the F_{IT} for Gyr (0.013±0.106), and Guernsey (0.02±0.217) (Strucken *et al.*, 2021) was 30% lower than that found in our study. The distribution of animals by their F_{IT} value can be observed in Figure 1. In general, animals fall in the range of 0 to 0.3 F_{IT}. The average F_{IT} is lower than those reported

Table 1. Single nucleotide polymorphisms (SNP) associated with resistance to mastitis in Braunvieh Cattle.

Gene[1]	Chromosome	SNP marker name	Locus number
ARHGAP10	17	ARS-BFGL-NGS-113821	1
		BovineHD1700002890	2
BDH2	6	BovineHD0600006001	3
CAPG	11	Hapmap54495-rs29018810	4
CHD5	16	BovineHD1600013051	5
CST6	29	BovineHD2900013196	6
ELMO1	4	BovineHD0400016236	7
FBL	18	ARS-BFGL-NGS-113564	8
FOCAD	8	BovineHD0800007122	9
IMMP2L	4	BovineHD0400015680	10
LOC510112	10	BovineHD1000009188	11
		BovineHD1000009179	12
MBL2	26	BovineHD2600001441	13
		BovineHD2600001446	14
MYO1E	10	BovineHD1000015285	15
PON1	4	ARS-BFGL-NGS-48351	16
PPP3CA	6	BovineHD0600006555	17
ST7	4	BovineHD0400014185	18
SULF2	13	BovineHD1300022072	19
TBC1D8	11	BovineHD1100002208	20
TBCK	6	BovineHD0600005226	21
ZNFX1	13	BovineHD1300022388	22

[1] Candidate genes reported by Chen *et al.* (2015), Cai *et al.* (2018), Kurz *et al.* (2019), and Moretti *et al.* (2021).

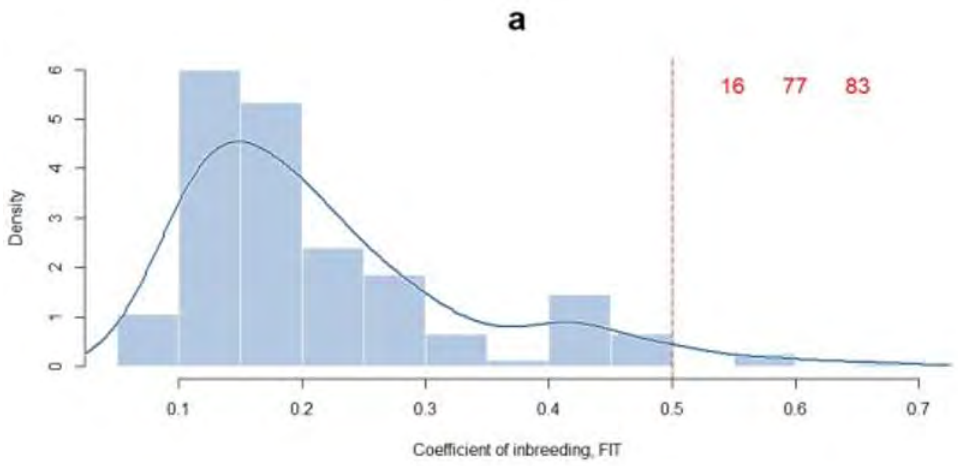


Figure 1. Histogram and density plot (cosine Kernel Function) of the coefficient of inbreeding (F_{IT}) for resistance to mastitis in a Braunvieh cattle population. bw= Bandwidth. Animals with a F_{IT} greater than 0.5 are shown in red.

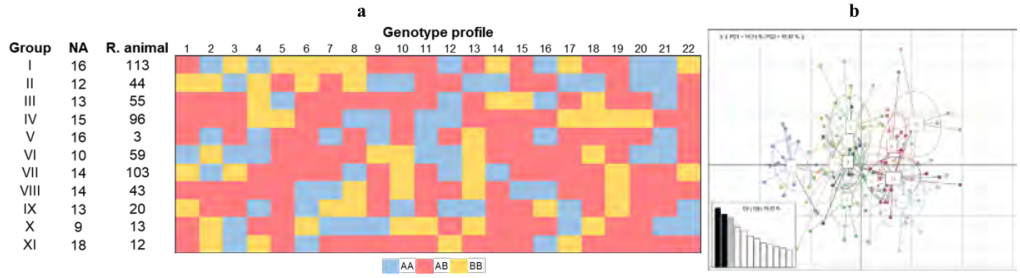


Figure 2. Clustering results (a) for 150 individuals and graphic representation of the principal components analysis (b) for resistance to mastitis in a Braunvieh cattle population. NA = Number of animals in the group. R. animal = representative animal (the closest to the centroid). PC1/2 = percentage of variability explained by dimension one (x-axis) and two (y-axis). Note: the names of each locus are indicated in Table 1.

by Signer-Hasler *et al.* (2004) for *Bos taurus* breeds (0.071), Brown Swiss (0.071), Braunvieh (0.059), and Original Braunvieh (0.023) (Moscarelli *et al.*, 2020).

The K-means algorithm and PCA are good statistical tools to visualize groups of animals with different genotypic profiles. Figure 2 shows the high genetic diversity of the Braunvieh population, since AB genotypes are abundant, especially in group XI (a). PCA chart (b) suggests that there are clear differences between the groups formed by K-means.

The results of our research give evidence of the genetic potential of the Mexican Braunvieh population for the improvement of mastitis-related traits in the tropic.

References

Ali, I., Suhail, S.M., and Shafiq, M. (2019) Reprod. Domest. Anim. 54:1026-1033. https://doi.org/10.1111/rda.13458

Brunes, L.C., Baldi, F., Lopes, F.B., Narciso, M.G., Lobo, R.B. *et al.* (2021) Animal 15(2):100085. https://doi.org/10.1016/j.animal.2020.100085

Gene. (2021) Bethesda (MD): National Library of Medicine (US), National Center for Biotechnology Information. https://www.ncbi.nlm.nih.gov/gene/ (accessed on 15 June 2021).

Haugaard, K., Tusell, L., Perez, P., Gianola, D., Whist, A. C., *et al.* (2013) J. Dairy Sci. 96(6):3986–3993. https://doi.org/10.3168/jds.2012-6133

Jombart, T. and Ahmed, I. (2011) Bioinformatics (27):3070–3071.

Kašná, E., Zavadilova, L., and Stipkova, M. (2018) Czech J. Anim. Sci. 63(11):443–451. https://doi.org/10.17221/105/2018-cjasKassambara, A. and Mundt, F. (2020) https://CRAN.R-project.org/package=factoextra (accessed on 16 November 2021).

Khalilzadeh, J. and Tasci, A.D.A (2017) Tour. Manag. 62:89-96.

King, F.J.M., Banga, C.B. and Visser, C. (2021) Trop. Anim. Health Prod. 53:117. https://doi.org/10.1007/s11250-021-02562-0

Meier, S., Arends, D., Korkuć, P., Neumann, G.B., and Brockmann, G.A. (2020) J. Dairy Sci. 103(11):10289-10298. https://doi.org/10.3168/jds.2020-18209

Moscarelli, A., Sardina, M.T., Cassandro, M., Ciani, E., Pilla, F. *et al.* (2020) Anim. Genet. 52:21-31. https://doi.org/10.1111/age.13019

Mrode, R., Ojango, J. M. K., Okeyo, A. M., and Mwacharo, J. M. (2019) Front. Genet. 9:694. https://doi.org/10.3389/fgene.2018.00694

Ocampo, R.J., Martínez, J.F. and Martínez, R. (2021) Trop. Anim. Health Prod. 53:122. https://doi.org/10.1007/s11250-021-02563-z

R Core Team. (2021) https://www.R-project.org/ (accessed on 16 November 2021).

Sherwin, W.B. (2010) Entropy 12:1765-1798.

Strucken, E.M., Gebrehiwot, N.Z., Swaminathan, M., Joshi, S., Al Kalaldeh, M. *et al.* (2021) Genet. Sel. Evo. 53:47. https://doi.org/10.1186/s12711-021-00640-3

Suleiman, T.S., Karimuribo E.D., and Mdegela, R.H. (2018) Trop. Anim. Health Prod. 50:259–266. https://doi.org/10.1007/s11250-017-1424-3

216. Evaluation of the Istrian cattle population structure by pedigree analysis

A. Ivanković[1*], J. Ramljak[1], G. Šubara[2], M. Pećina[1], E. Šuran[2], Z. Ivkić[3] and M. Konjačić[1]

[1]Faculty of Agriculture University of Zagreb, Svetošimunska 25, 10000 Zagreb, Croatia; [2]Agency for Rural Development of Istria, Tugomila Ujčića 1, 52000 Pazin, Croatia; [3]Croatian Agency for Agriculture and Food, Vinkovačka cesta 63c, 31000 Osijek, Croatia; aivankovic@agr.hr

Abstract

Local cattle breeds are an important part of genetic diversity, but due to globalization and industrialization of production, they are often marginalized. Istrian cattle is a local breed that almost became extinct three decades ago. Active measures have been taken to protect the Istrian cattle and the population has increased, gaining a new economic importance. The objective of this study was to analyze the status and trends of Istrian cattle using pedigree information. The population of Istrian cattle has stabilized trends in the number of animals and herds, although it requires constant monitoring. The levels of relatedness and inbreeding have plateaued (at 4.00 and 3.14%, respectively) and bottleneck in the population has not been observed. The mean number of equivalent generations is relatively low (2.99) and further systematic monitoring is important.

Introduction

Genetic diversity is one of nine control variables in the planetary boundaries framework (Steffen *et al.*, 2015). The genetic diversity of cattle, the most important domestic species in terms of global production of food of animal origin, is gradually being depleted (FAO, 2015). Of the total number of breeds of mammalian species used for agriculture worldwide, cattle breeds account for 22.90% (FAO, 2019). Although the number of extinct cattle breeds is relatively low (12.8%), a large number of breeds are endangered (21.1%) and only 12.3% of cattle breeds are in 'not endangered' status. For most cattle breeds, their endangered status is unknown (53.8%) (FAO, 2019).

Istrian cattle is one of the three Croatian local cattle breeds. The primary breeding area of Istrian cattle is the Istrian peninsula and the neighboring island, and in last ten years has been extended to karst area (Lika, Dalmatia). It belongs to the Podolian group of cattle breeds and is very similar to the Italian Podolian and Maremmana breeds (Maretto *et al.*, 2012). Until the early twenties of the 20th century, cattle production in Croatia was based on local cattle breeds. Due to the modernization of agricultural production, the use of mechanization and the introduction of allochthonous breeds, local breeds were marginalized, and their survival was in danger. In the nineties of the 20th century, the program of protection of Istrian cattle was launched and the initial population included 103 cows and eight bulls (Ramljak *et al.*, 2011). According to the Croatian Agency for Agriculture and Food (2021), the current population of Istrian cattle is 1,118 adult animals (65 bulls and 1,053 cows) housed in 174 breeding farms. Previous studies on Istrian cattle included phenotype analysis and genetic analysis (Ivanković *et al.*, 2007; Ramljak *et al.*, 2011; Maretto *et al.*, 2012), but pedigree analysis was not performed. Pedigree information is a useful source for conservation programs to maintain genetic variability and minimize inbreeding (Santana *et al.*, 2016), and to analyse the breeding status of population (Jarnecka *et al.*, 2021). Therefore, the objective of this study was to describe the population structure and trends of the Istrian cattle population through pedigree analysis.

Materials & methods

The database of Istrian cattle includes 7,399 animals born between 1971 and 2020, belonging to 174 herds located mainly on the Istrian peninsula. For each animal in the database, information such as the

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_216, © A. Ivanković *et al.* 2022

identification number, date of birth, sex, sire, dam, and herd were recorded. In this study, active (alive) animals of both sexes were selected as reference population (REF 01). Demographic and population parameters are presented for the last two decades, because Herdbook was closed in 2000. Pedigree analysis was performed using ENDOG v4.8 software (Gutiérrez and Goyache, 2005). Pedigree completeness was determined using the mean number of complete generations, mean number of maximum generations, and mean number of equivalent generations (Boichard *et al.*, 1997). Parameters related to inbreeding (mean inbreeding coefficient (*F*), mean average relatedness coefficient (*AR*), effective population size (*Ne*)), probability of gene origin (effective number of founders (f_e), effective number of ancestors (f_a), effective number of founder herds (f_h)), and generation interval (*GI*) were assessed. The effective population size was estimated via an individual increase in inbreeding per generation (Gutiérrez *et al.*, 2008). The average relatedness coefficients and individual inbreeding coefficients were calculated based on Gutiérrez and Goyache (2005). The generation intervals were calculated for the four paths: sire-son, sir-daughter, dam-son and dam-daughter.

Results

Demographic trends in the Istrian cattle population are shown in Figure 1. The number of cows has been constant1y increasing per year. According to Figure 1, the number of herds and number of new registered animals (offspring's included in the Herdbook per year) have increased from 2000 to 2011 and have only been roughly stable in the last decade.

An analysis of all pedigree entries revealed 6,756 animals for which both parents were known. The active living population consisted of 2,207 active animals and the effective population size is 91.92 (Table 1). The number of founder animals was 235 and the mean number of equivalent complete generations was 2.99.

The ratio between effective number of founders and effective number of ancestors is 1.03, The average generation interval was 5.93±3.65 years and ranged from 4.77±3.07 (sire – son) to 6.69±3.84 (dam – son). The average inbreeding coefficient in the REF 01 population was 3.14%, ranging from a minimum of 1.41% (2002) to a maximum of 3.19% (2018) (Figure 2). The *AR* in the population was 4.00%, ranging from 3.31% (2008) to 4.27% (2000).

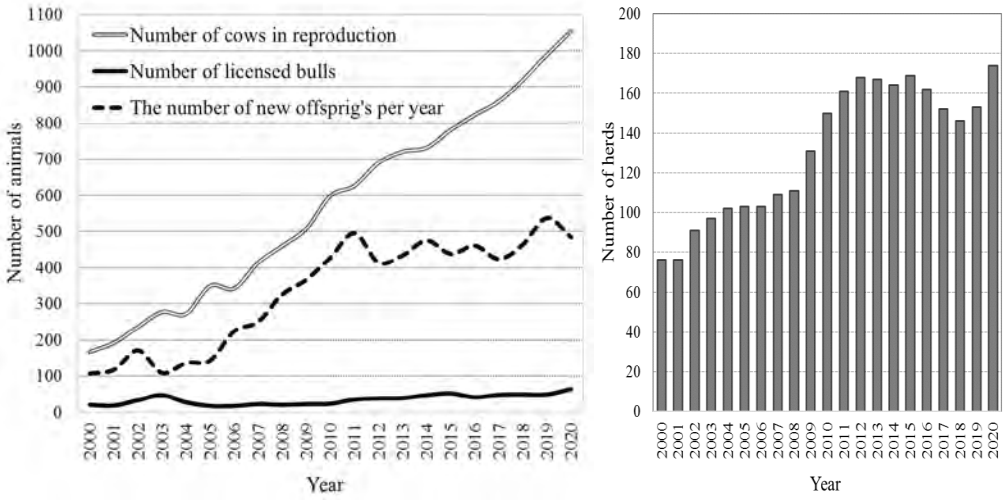


Figure 1. Population trends in the Istrian cattle population in the last two decades.

Table 1. Population size and genetic diversity indices of the Istrian cattle.

Item	Total
Original dataset (TP; total population)	7,399
Reference population 1 (REF 01; live animals)	2,207
Mean inbreeding coefficient (*F*)	3.14
Mean Average Relatedness coefficient (*AR*)	4.00
Effective population size (N_e)	91.92
Mean number of complete generations (*MCG*)	1.79
Mean number of equivalent generations (*MEG*)	2.99
Mean number of maximum generations (*MMG*)	5.81
Number of founders contributing to reference population	235
Effective number of founders for the reference population (f_e)	36
Effective number of ancestors for the reference population (f_a)	35
Effective number of founder herds (f_h)	17.10
Effective number of founders/effective number of ancestors (f_e/f_a)	1.03

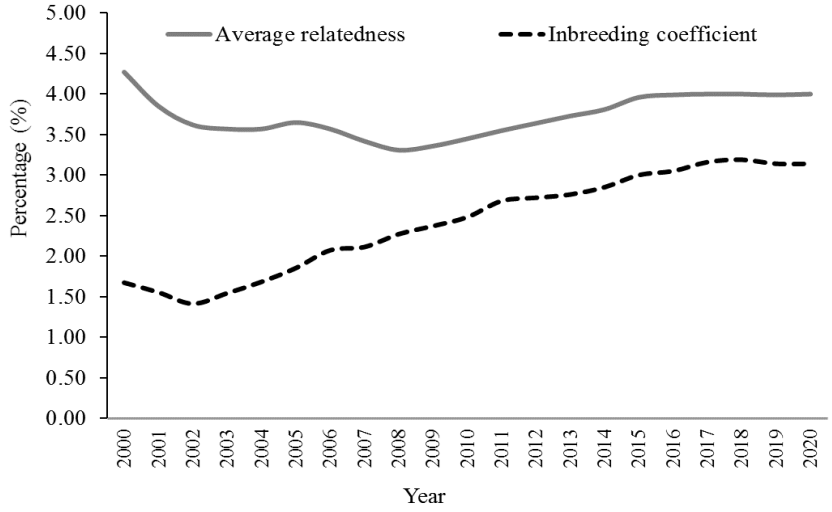


Figure 2. Level of inbreeding coefficient and average relatedness (in %) in the Istrian cattle from 2000 to 2020.

Discussion

The Istrian cattle population is subject to a systematic protection program, and an economic revitalization program has been implemented in the last decade. Considering the number of animals, the population trend is positive, as well as the increase in the average size of the herd from 2.5 adult animals (in 2000) to 6.4 (in 2020). The sex ratio in the population (cows/bulls) increased from 8.0 (in 2000) to 16.2 (in 2020). In the last ten years, the growth of the inbreeding level has slowed down, while in the last five years it has stagnated around 3.14%. The level of inbreeding observed in Istrian cattle is not alarming, but requires constant monitoring. We need to be careful in monitoring the level of inbreeding because Boichard *et al.* (1997) pointed out that the *F* trend has some limitations, especially when pedigree completeness is low. It should consider the fact that pedigree completeness is still low in the Istrian cattle population (*MCG*=1.79). The relationships between N_e, f_e and f_a provide information about the occurrence of bottleneck in a population.

For example, a f_e/f_a ratio close to 1 suggests that bottlenecks have not occurred in the population (Boichard *et al.*, 1997). A high f_e/f_a ratio indicates disproportionate use of breeding animals (Stephens and Splan, 2013). The f_e/f_a ratio in Istrian cattle is currently 1.03, indicating that there is no obvious bottleneck in the population. We assume that this is a consequence of the higher presence of natural mating compared to artificial insemination in practice.

Funding

This research has been fully supported by the Croatian Science Foundation (Genetic, Economic and Social Interactions of Local Breed Conservation Programs, GGD LocBreed), grant number IP-2020-02-4860.

References

Boichard D., Maignel L., and Verrier, E. (1997) Genet. Sel. Evol. 29(5):5-23. https://doi.org/10.1186/1297-9686-29-1-5

Croatian Agency for Agriculture and Food (CAAF), Cattle Breeding - Annual Report 2021. Available at: https://www.hapih.hr/wp-content/uploads/2021/05/

Caballero A. and Toro M.A. (2000) Genet. Res. 75(3):331-343. https://doi.org/10.1017/S0016672399004449

FAO (2015) The second report on the state of the world's Animal genetic resources for food and agriculture. Food and Agriculture Organisations, Rome, Italy.

Food and Agriculture Organisations (FAO), Status and Trends of Animal Genetic Resources 2019. Available at: http://www.fao.org/3/my867en/my867en.pdf

Gutiérrez J.P., and Goyache F. (2005) J. Anim. Breed. Genet. 122(3):172-176. https://doi.org/10.1111/j.1439-0388.2005.00512.x

Gutiérrez J.P., Cervantes I., Molina A., *et al.* (2008) Genet. Sel. Evol. 40: 359-378.
https://gsejournal.biomedcentral.com/track/pdf/10.1186/1297-9686-40-4-359.pdf

Ivanković A., Kelava N., Konjačić M., *et al.* (2007) Stočarstvo 61:161-173.

Jarnecka O., Bauer E.A., and Jagusiak W. (2021) Animal 15(6):100238. https://doi.org/10.1016/j.animal.2021.100238

Maretto F., Ramljak J., Sbarra F., *et al.* (2012) Livest. Sci. 150(1-3):256-264. https://doi.org/10.1016/j.livsci.2012.09.011

Ramljak J., Ivanković A., Veit-Kensch C.E., *et al.* (2011) J. Anim. Breed. Genet. 128(1):73-84. https://doi.org/10.1111/j.1439-0388.2010.00905.x

Santana M.L., Pereira R.J., Bignardi A.B., *et al.* (2016) Livest. Sci. 187:6-15. https://doi.org/10.1016/j.livsci.2016.02.002

Steffen W., Richardson K., Rockström J., *et al.* (2015) Science 347(6223):1259855. https://doi.org/10.1126/science.1259855

Stephens T.D., and Splan R.K. (2013) Animal Genetic Resources 52:31-38. https://doi.org/10.1017/S2078633613000052

217. Population dynamics of a long-term selection experiment in chickens divergently selected for high or low body weight

S.J. Harrison[1*], P.B. Siegel[2], C.F. Honaker[2] and R.M. Lewis[3]

[1]The University of Zambia, P.O. Box 32379, 10101 Lusaka, Zambia; [2]Virginia Polytechnic Institute and State University, 2170 Litton Reaves Hall, Blacksburg, VA 24061, USA; [3]University of Nebraska-Lincoln, P.O Box 830908, 68583-0908, Lincoln, NE, USA; sylvia.harrison@unza.zm

Abstract

The population dynamics of two lines of chickens from a long-term selection experiment was assessed. White Plymouth Rock chickens were selected for high or low 8-week body weight for 59 generations. The objective was to determine whether the two lines maintained similar population structures over the selection horizon. A complete pedigree of 30,943 birds was available: 102 founders, 16,292 low weight select (LWS), and 14,549 high weight select (HWS). Average inbreeding rates per generation were less than 1% in both lines. At generation 59, average relatedness coefficients were 0.53 (SD 0.001) for LWS, and 0.66 (SD 0.001) for HWS. Effective population size, whether based on inbreeding or coancestry, was 39 in LWS and 33 in HWS. Effective number of founders, ancestors, and genome equivalents were 17, 12.3, and 2.5 in LWS; correspondingly, in HWS they were 15, 8.4, and 1.9. Since the two lines had similar population structures, selection responses can be reliably compared.

Introduction

Population dynamics refers to the way in which the size and age structure of populations change over time, and the characterization of that change in mathematical terms. Such change is brought about by selection, as individuals that differ in viability and fertility contribute differently to the next generation. A long-term selection experiment in chickens based on high and low body weight (BW) at 8-wk of age provides a clear illustration of response to artificial selection (Márquez *et al.*, 2010). In a closed population, such as this one, inbreeding is inevitable, which effects the variance of gene frequency from one generation to the next (Konig *et al.*, 2010). There also are impacts on the structure of the population, manifest in its effective population size and contributions of founders, ancestors, and genetic drift. Gutiérrez *et al.* (2003) reported that genetic variability and evolution can be explained by a well-documented pedigree. The objective of this study was to assess population dynamics and genetic diversity in the high and low BW lines now after 59 years of selection, using pedigree data.

Materials & methods

Data and husbandry. Data were from 59 discrete generations. The two selection lines were established from a cross of 7 inbred lines of White Plymouth Rock chickens. Fifteen roosters and 87 hens from this cross were the founders of both lines. Their heavier offspring were selected as parents of the high weight select (HWS) line while their lighter offspring were selected as parents of the low weight select (LWS) line. From that 1,957 hatch, 8 sires and 48 dams were selected to establish the parental generation for each line. A complete pedigree, including sex, generation, and selection line, was constructed for each bird. The data analysed included 102 founders, 16,292 LWS and 14,549 HWS birds.

Inbreeding and average relatedness. The analyses were conducted using ENDOG v4.8 (Gutiérrez and Goyache, 2005). Individual inbreeding coefficients (F_t), and changes in inbreeding over time, were computed to deduce trends and infer the remaining heterozygosity. The inbreeding rate statistic was calculated as:

$$\Delta F_t = (F_t - F_{t-1}) / (1 - F_{t-1}) \quad (1)$$

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_217, © S.J. Harrison *et al.* 2022

where ΔF_t was the change in mean inbreeding between successive generations, and F_t was the mean inbreeding coefficient in generation *t*. Coancestry coefficients (f_{ij}) between each pair of individuals, *i* and *j*, were obtained and used to calculate Wright's F-statistics (F_{IS}, F_{ST}; Wright, 1978). The average relatedness coefficient (*AR*) was computed as explained by Dunner *et al.* (1998) accounting for both inbreeding and coancestry.

Effective population size. The effective population size describes the effective number of breeding individuals responsible for observed inbreeding rates if the population was randomly mated. It was calculated as:

$$N_e = 1 / (2 \times \Delta F_\mu) \quad (2)$$

where ΔF_μ was the average:

$$\Delta F_i = 1 - (\sqrt[t-1]{(1 - F_i)}) \quad (3)$$

of the birds in generation 59, and *t* was the number of equivalent complete generations (Gutiérrez *et al.*, 2009). Additionally, such realized effective population size was computed in a similar way but based on coancestry (N_{ec}; Cervantes *et al.*, 2011).

Effective number of founders, ancestors and genome equivalents. The effective number of founders (f_e) was computed to deduce the number of individuals that would be expected to produce the observed genetic diversity in a population if all the individuals had contributed equally to the population, as:

$$f_e = 1/\Sigma_{(i = 1,f)} q^2_i \quad (4)$$

where q_i was the proportion of genes originating with the i^{th} founder, and *f* was the total number of founders. The effective number of ancestors (f_a), was computed to account for unbalanced use of potential parents resulting in losses in genetic diversity producing bottlenecks (Boichard *et al.*, 1997), as:

$$f_a = 1 / \Sigma_{(j = 1,a)} p^2_j \quad (5)$$

where p_j was the marginal genetic contribution of the j^{th} ancestor, and *a* was the total number of ancestors. Founder genome equivalents (f_g) were obtained to quantify the number of equally contributing founders that would lead to the same level of genetic diversity as seen in the population under study assuming no random losses of founder alleles (Lacy, 1989).

Results

Inbreeding and average relatedness. Mean inbreeding coefficients per generation are shown in Figure 1 as the regression of $\log(1 - F_t)$ on *t* (R^2>0.98). The slope for LWS of -0.58±0.19% and for HWS of -0.68±0.32% approximated the -*ΔF* for each line. Maximum and average inbreeding rates per generation were similar in both lines. The population-level F_{ST} was 0.1202, suggesting that 12% of the total variation in the population was due to line differences. That finding was consistent with the pairwise F_{ST} between lines of 0.1215, implying that the selection lines had diverged substantially genetically. The maximum inbreeding coefficient and average relatedness in each line at generation 59 is summarized in Table 1.

Table 1. Summary statistics for genetic diversity measures at generation 59.

Parameter	Low weight select	High weight select
Number	193	185
Maximum F_i	0.64	0.63
Mean *AR* (SD)	0.53 (0.001)	0.66 (0.001)
N_e	39.3	33.2
N_{ec}	38.6	32.9
f_e	17	15
f_a	12.3	8.4
f_g	2.5	1.9

[1] F_i is individual inbreeding coefficient; *AR* is average relatedness coefficient; N_e is realized effective population size based on inbreeding; N_{ec} is realized effective population size based on coancestry; f_e is effective number of founders; f_a is effective number of ancestors; f_g is founder genome equivalents.

Effective population size. The realized effective population sizes at generation 59, whether based on inbreeding or coancestry, were similar in both lines at about 39 in the LWS and 33 in the HWS (Table 1). This coincided with their similar rates of inbreeding.

Effective number of founders, ancestors and genome equivalents. The effective number of founders, ancestors, and genome equivalents at generation 59 are shown in Table 1. Their values were similar for both lines. In Figure 2, the cumulative marginal contributions of ancestors to genetic diversity in each line at generation is 59 plotted. Fewer than 33 ancestors contributed to a line. In the LWS, five males and one female ancestor explained 50% of the genetic variation at generation 59. In the HWS, only three male ancestors explained that same amount of variation.

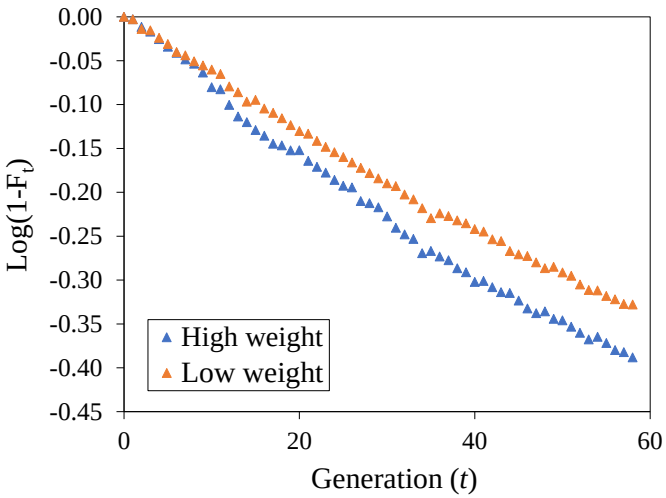


Figure 1. Plot of log(1-F_t) against *t* by line across 59 generations, where F_t was the mean inbreeding coefficient at generation *t*.

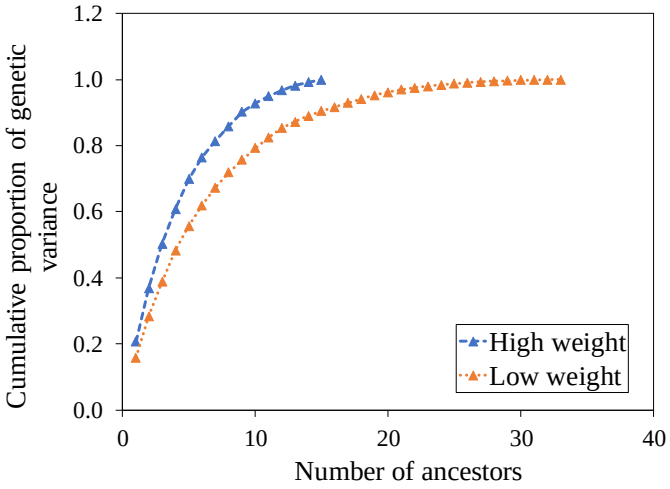


Figure 2. Cumulative marginal contributions of ancestors to genetic diversity in each line at generation 59.

Discussion

Inbreeding. A gradual increase in inbreeding coefficients in the LWS and HWS lines was observed. Because inbreeding increases the frequency of homozygous genotypes in a closed population, expression of deleterious recessive alleles can lead to inbreeding depression, which may negatively affect fertility and survivability. This population was propagated by short generation intervals, which is the case with chickens. The moderate levels of inbreeding identified were inevitable, with individuals became more closely related with increasing generations. However, the risk of inbreeding depression was less drastic in this population because the founders were entirely unrelated, and selection and mating decisions consciously avoided accumulating inbreeding. With on average less than 1% increase in inbreeding per generation, there was relatively low risk of inbreeding depression or loss of a line.

Effective population size. The N_e observed was similar in both lines and this similarity can be attributed to the relatively constant rate of inbreeding across generations. At generation 59, the N_e was about 60 and 68% of the actual population size (number of selected parents) in the LWS and HWS lines, respectively. Franklin and Frankham (1998) reported a critical N_e of 500 to maintain adequate genetic variation in natural populations. However, that study did not indicate whether that threshold was specific to a particular species or breed, or due to biological differences such as reproductive rate. Smaller livestock populations (e.g. N_e>100) were considered viable to maintain long-term selection responses (Goddard, 1992; Weigel, 2001).

Effective number of founders, ancestors and genome equivalents. The f_e measures the expected founder contributions to descendants across generations. However, the f_a is more befitting for a population that has undergone bottlenecks, which risks a potential loss of alleles. In generation 59 of this study, the f_a [12 (LWS); 8 (HWS)] was relatively low compared to the actual number of founders (102), taking note that an ancestor many not necessarily be a founder. The cumulative marginal contribution of the ancestors showed that few ancestors were most influential in genetic contribution to descendants, as was the case with founders. Similar percentages were observed with regards the number of equally contributing founders that would lead to the same level of genetic diversity.

Based on various statistics that characterize the dynamics of a population, the LWS and HWS lines were similar in their genetic architecture. Robust comparisons of the performance of these two lines over the entire selection profile therefore can be made with confidence.

References

Boichard, D., Maignel L., and Verrier E. (1997) Genet. Sel. Evol. 29:5-23. https://doi.org/10.1186/1297-9686-29-1-5

Caballero A., and Toro M. A. (2000) Genet. Res. 75:331-343. https://doi.org/10.1017/s0016672399004449

Cervantes I., Coyache F., Molina A., Valera M., and Gutiérrez J.P. (2011) J. Anim. Breed. 128(1):56-63.https://doi.org/10.1111/j.1439-0388.2010.00881.x

Dunner, S., Checa, M.L., Gutierrez, J.P., Martin, J.P., and Cañon, J. (1998) Genet. Sel. Evol. 30:397-405. https://doi.org/10.1186/1297-9686-30-4-397.

Franklin J.R., and Frankham R. (1998) Anim. Conserv. 1:69-73. https://doi.org/10.1111/j.1469-1795.1998.tb00228.x

Gutiérrez, J.P., Altarriba J., Diaz C., Quintanilla R., Caňõn J., *et al.* (2003) Genet. Sel. Evol. 35:1-21. https://doi.org/10.1186/1297-9686-35-1-43

Gutiérrez J.P., and Goyache F. (2005) J. Anim. Breed. Genet. 122(3):172-176. https://doi.org/10.1111/j.1439-0388.2005.00512.x

Konig S., Tsehay F., Sitzenstock F., Von Borstel U.U., Schmutz M., *et al.* (2010). Poult. Sci. 89:658–667. https://doi.org/10.3382/ps.2009-00543

Lacy R.C. (1989) Zoo Biol. 8:111-123. https://doi.org/10.1002/zoo.1430080203

Márquez G.C., Siegel P.B., and Lewis R.M. (2010) Poult. Sci. 89:2580–2588. https://doi.org/10.3382/ps.2010-01034

Goddard, M.E. (1992) J. Dairy Sci. 75:2902-2911. https://doi.org/10.3168/jds.S0022-0302(92)78052-7

Weigel, K.A. (2001) J. Dairy Sci. 84(E. Suppl.):E177-184. https://doi.org/10.3168/jds.S0022-0302(01)70213-5

Wright, S. (1978) Evolution and genetics of populations. Vol 4. Variability within and among populations. University of Chicago Press, Chicago, USA.

218. Genetic management meets genomics

J.A. Woolliams[1*] and T.H.E. Meuwissen[2]

[1]The Roslin Institute, Royal (Dick) School of Veterinary Studies, University of Edinburgh, Easter Bush Campus, Midlothian EH25 9RG, United Kingdom; [2]Norwegian University of Life Sciences, P.O. Box 5003 NMBU, 1432 Ås, Norway; john.woolliams@roslin.ed.ac.uk

Abstract

Genomics opens the opportunity for observing multiple genetic perspectives but genetic management requires decisions that resolve them towards desired outcomes. Management using alternative genomic relationship matrices (G) may create a covariance between drift and initial allele frequency and this creates a distinction between natural measures of inbreeding based on drift (F_{drift}) and homozygosity (F_{hom}). These measures coincide when inbreeding management is based on identity by descent (IBD), with more gain relative to either F_{drift} or F_{hom} when G was based on IBD compared to drift or homozygosity. The covariance implies that management based on drift or homozygosity drives rare alleles respectively towards extinction or intermediate frequencies either of which may be undesired. Management by IBD allows frequency changes to be as expected with random drift, separating out the management of diversity from the selective advantages arising from artificial or natural selection.

Introduction

Genomics opens a window on the diversity that delivers gain in breeding schemes, and so offers opportunities for re-imagining them. The key to sustaining more than short-term gain is the diversity itself, both for traits underlying productivity and those underlying robustness, encompassing health and welfare in livestock and companion animals. Therefore management of diversity, and understanding how genomics may play a role, is important in re-imagining. A crucial perspective is that managed populations in their current form have value, e.g. productivity in livestock, temperament in companion animals, feral traits in zoo populations; consequently many attributes are to be maintained alongside desired improvements. Therefore the current population is a reference point framing the objectives for managing diversity.

In the absence of genomics the ideas of diversity were built upon IBD referring back to a base population, a reference point (so well-suited to the management needs) where all alleles are considered distinct irrespective of their identity by state (IBS). In this inbreeding framework, core concerns in managing diversity coincide: genetic drift from the reference, which may promote lethal recessives, and loss of heterozygosity, in turn associated with loss of genic variance and increased risks of fixing deleterious and losing desired alleles. These are all predictable from the classical F, and controlling the rate of inbreeding (ΔF) provides a means to managing diversity. Control is achieved with optimum contributions (OC; e.g. Meuwissen, 1997) which, in one form, maximises rates of gain (ΔG) for a given ΔF, using the NRM (A).

The problem with this framework is that the classical theory refers to neutral and selection-free loci, and genomics has demonstrated that while there may be many neutral loci, these are likely linked to loci affecting selection. Routine prediction of EBVs using a G based on dense chips of primarily *a priori* neutral SNPs compounds this problem as these SNPs do not remain neutral. Furthermore, the genomic data for these SNPs allows direct observation of heterozygosity and drift from their frequencies in a reference population. Therefore the simple substitution of G for A in OC, as in BLUP, offers a viable solution (Sonesson *et al.* 2012). However Beukelaer *et al.* (2017) hint that with this solution the coincidence of drift and loss of heterozygosity may disappear. Furthermore, while there is one A, there are many possible G, raising the possibility that what may be true for one G may not be true for another. To resolve these issues extensive

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_218, © J.A. Woolliams and T.H.E. Meuwissen 2022

simulation studies were conducted and reported by Meuwissen *et al.* (2020) and what follows is a summary of their findings with special attention for the discrepancies between drift and homozygosity inbreeding in genomic diversity management.

Materials & methods

Genomes and SNPs. The genome was 10 simulated chromosomes, each 1M, and developed over 400 generations with N_e=100, and a mutation rate of 10^{-8}/base pair/generation to achieve drift-mutation-linkage disequilibrium (LD) balance. This resulted in 33k SNPs in a base population with allele frequencies (p_{0k} for SNP k) distributed comparably to whole genome sequence (WGS). Four separate panels of 7k SNPs were randomly sampled: M, used for predicting EBVs and (in most results) also constructing G to manage diversity; Q, used as QTLs with a Normal distribution of substitution effects, with variance independent of p_{0k}, and combined with environmental variance to give a trait for selection with breeding values g and h^2=0.4; N, used as truly neutral loci to assess outcomes; and D, used (for some results) in conjunction with M to separate genomic data for predicting EBVs and managing diversity.

Predicting EBV. A ridge regression model on scaled allele counts was used to predict EBV

$$y = 1m + Xb + e \quad (1)$$

where y is a vector of phenotypes, m a fitted mean, 1 a vector of 1's, b a vector of coefficients for panel M markers, assumed MVN(0,σ^2I), X their design matrix, and e a vector of residuals. X contained the counts of a random chosen alternative allele at SNP k scaled by $[2p_{0k}(1-p_{0k})]^{1/2}$. The use of p_{0k} was maintained across generations. The scaling leads to models analogous to Model 2 of Van Raden (2008). EBVs were then predicted as $g^{hat}=Xb^{hat}$.

Diversity management. OC was simulated with a number of different relationship matrices G. Genomics vastly expands possible G, and can be classified by their underlying target. A: drift, exemplified by G_{VR1} and G_{VR2}, based on methods 1 and 2 of Van Raden (2008), but with frequencies remaining 'centred' on p_{0k}; and $G_{i(p)}$ based on transforming a change in allele frequency from p_{0k} to a scale of total intensity applied, whether or not random (Liu *et al.*, 2008). B: homozygosity, exemplified by $G_{0.5}$, constructed as a Van Raden matrix with all frequencies 'centred' on 0.5. C: IBD, exemplified by A, and G_{FG} using genomics to aggregate IBD across the genome by linkage analysis (Fernando and Grossman, 1989). A hybrid of B and C using ROH is also feasible, G_{ROH}. All were considered by Meuwissen *et al.* (2020), including details of their calculation, but only a subset will be described in these results.

Simulations and summaries. OC schemes were run for 19 generations after a generation of random selection in a dioecious population with ΔF=0.005 (N_e=100), although exact values cannot be guaranteed when using e.g. A or $G_{0.5}$. In one scenario the EBVs in each generation were randomly generated. Each form was replicated 100 times and averaged. In each replicate ΔG was calculated, together with 2 natural measures of inbreeding coefficient following Wright: if p_{tk} is the allele frequency of locus k at time t, H_{tk} = $2p_{tk}(1-p_{tk})$, and $\delta p_{tk} = (p_{tk}-p_{0k})$, these were $F_{hom} = 1- \Sigma_{loci} H_{tk}/[H_{0k}N_{loci}]$ and $F_{drift} = \Sigma_{loci} \delta p_{tk}^2/[p_{0k}(1-p_{0k})N_{loci}]$. Corresponding rates were calculated using -1x the slope of the regression of $\log_e(1-F)$ on t.

Results

With these definitions of F_{hom} and F_{drift}, then their difference is twice the covariance of the allele frequency change δp_{tk} for locus k and its initial deviation from ½, standardising both terms by $[p_{0,k}(1-p_{0,k})]^{1/2}$, which is the genic s.d. for individual allele frequency in fully inbred individuals, i.e. $F_{hom}-F_{drift}=2\text{cov}(\delta p_{tk}/[p_{0,k}(1-p_{0,k})]^{1/2},(p_{0k}-½)/[p_{0,k}(1-p_{0,k})]^{1/2})$. This is the covariance of Beukelaer *et al.* (2016). Compared to classical drift

Table 1. Observed ΔF measured by drift or homozygosity and the difference between total F in SNP panels M and N after 20 generations of OC selection.

		EBV predictors: Panel M			Neutral loci: Panel N		
Scheme[1]	Type[2]	ΔF_{hom} [3]	ΔF_{drift} [3]	F_{hom}-F_{drift} [3]	ΔF_{hom} [3]	ΔF_{drift} [3]	F_{hom}-F_{drift} [3]
G_{VR2}(M,M)	A	0.0146	0.0050	0.147	0.0103	0.0068	0.054
G_{VR2}(M,D)	A	0.0100	0.0069	0.048	0.0101	0.0068	0.050
G_{VR2}(~,M)	A	0.0109	0.0050	0.093	0.0085	0.0059	0.041
G_{VR1}(M,M)	A	0.0096	0.0056	0.063	0.0080	0.0069	0.021
$G_{0.5}$(M,M)	B	0.0008	0.0213	-0.348	0.0073	0.0176	-0.170
G_{FG}(M,M)	C	0.0044	0.0049	-0.009	0.0043	0.0049	-0.010
A(M,~)	C	0.0072	0.0083	-0.016	0.0070	0.0084	-0.021

[1] G_{VR2}(M,D) denotes SNP panel M used to predict EBVs, and SNP panel D used to construct G_{VR2} for use in OC. Others follow by analogy, with ~ indicating no panel required either due to random EBVs or use of A.

[2] Target for controlling diversity (see text): A, drift; B, homozygosity; C, IBD.

[3] Standard errors $<2.5\times10^{-5}$ for ΔF_{hom} and ΔF_{drift} and $<2.2\times10^{-4}$ for F_{hom}-F_{drift}.

where F_{hom}=F_{drift}, a covariance >0 indicates minor alleles are more likely to be lost, and <0 indicates they are more likely to be retained but drift towards 0.5 and further from p_{0k} (especially in WGS data).

Distinction between F_{hom} and F_{drift}. Table 1 shows that with genomic management of diversity the classical coincidence of F_{hom} and F_{drift} is no longer automatic. Diversity management based upon VR1 or VR2 resulted in $F_{hom}>F_{drift}$, from promoting the loss of minor alleles and reducing the larger magnitude drift towards fixation through penalties based on δp_{tk}^2. $F_{hom}>F_{drift}$ was also observed in the unmanaged neutral panel N, but less strongly, as these are driven by LD with the managed panel M. This phenomenon is not due to directional selection as $F_{hom}>F_{drift}$ was observed with random selection (G_{VR2}(~,M)), and it is not due to the use of the same marker panel for predicting EBV and diversity management as $F_{hom}>F_{drift}$ was observed with G_{VR2}(M,D). The converse, $F_{drift}>F_{hom}$, was observed when diversity management was based on homozygosity, with larger magnitude changes in allele frequency permitted to reduce the loss of alleles. As observed with $G_{0.5}$(M,M), a low ΔF_{hom} in the managed panel is not necessarily replicated in an unmanaged panel of neutral loci. However when the diversity management was tracking IBD the difference between F_{hom} and F_{drift} was much smaller than for all other schemes. For G_{FG}(M,M) most of the observed difference was explained by the non-random mating arising from the 2 sexes; F_{hom} is defined for HWE whereas in the simulations F_{hom} was as observed in the simulated population.

Gain in relation to F_{hom} and F_{drift}. Figure 1 shows gain achieved as a function of F_{hom} or F_{drift} over the 20 generations for 5 schemes in Table 1, with the target F=0.1 highlighted. Maximum gain was achieved using IBD for management when assessed per unit of either F_{drift} or F_{hom}. The two G_{VR} schemes gave very similar trajectories of gain, slightly lower than $G_{0.5}$(M,M) per unit of F_{hom} but greater than $G_{0.5}$(M,M) per unit of F_{drift}.

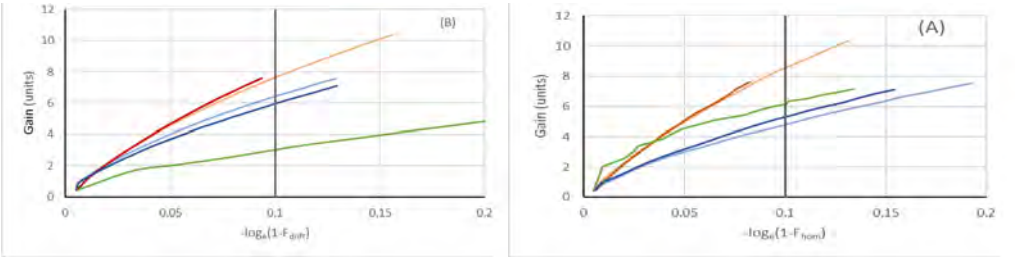


Figure 1. The trajectory of total gain as a function of different measures of inbreeding: (A) related to F_{hom}; (B) related to F_{drift}. Schemes shown are: G_{FG}(M,M) dark red; A(M,~) light red; $G_{0.5}$(M,M) green; G_{VR1}(M,M) dark blue, and G_{VR2}(M,M) light blue.

Discussion

The near coincidence of F_{hom} and F_{drift}, showing at most small covariance between MAF and drift is justified from theory as classical inbreeding can be imagined as arising from a set of 2N alleles at each locus for the N individuals in the base. Hence, all alleles at all loci used for management have frequency $(2N)^{-1}$ and the covariance will be 0, and equivalence follows. In turn, this offers an explanation as to why gain is greater when G was based on IBD, as the selective advantages of the QTL and EBV markers will be influenced through LD by the covariance in promoting changes not beneficial to the trait: either by resisting desired movement towards intermediate frequencies when using G_{VR1} or G_{VR2}, or resisting desired movement away from intermediate frequencies when using $G_{0.5}$. The effects observed are potentiated by $|p_0$-½| and suggests that using G_{VR1} rather than G_{VR2} may be preferable for generating gain as it weights G towards intermediate p_0 rather than equally irrespective of MAF, but the results for ΔG do not support this despite the reduced magnitude of F_{hom}-F_{drift}.

For breeding and generating directional gain the most attractive option was using IBD for managing diversity. Note that A does not offer control of inbreeding, as the true ΔF will always exceed the target to an unknown degree; G_{FG} overcomes this but computation remains challenging. Objectives may solely concern protecting diversity in the valued reference population and $G_{0.5}$ may appear attractive, but purging will be hindered, IBD increased, and epistatic variance promoted. In conclusion, the pre-genomic idea of IBD management allows allele frequency changes to be as expected under random drift, separating out the management of diversity from the selective advantages arising from artificial or natural selection.

References

Beukelaer de H., Badke Y., Fack V., and de Meyer G. (2017) Genetics 206:1127-1138. https://doc.org/10.1534/genetics.116.194449

Fernando R.L, Grossman M. (1989) Genet. Sel. Evol. 21:467-477. https://doc.org/10.1186/1297-9686-21-4-467

Liu A.Y., Woolliams J.A. (2010) Genet. Res. (Camb.) 92:157-166. https://doc.org/10.1017/S0016672310000145

Meuwissen, T.H.E. (1997) J. Animal Science 75(4):934-940. https://doc.org/10.2527/1997.754934x

Meuwissen T.H.E, Sonesson A.K, Gebregiwergis G, and Woolliams J.A. (2020) Front. Genet. 11:880. https://doc.org/10.3389/fgene.2020.00880

Sonesson, A.K., Woolliams, J.A., and Meuwissen T.H.E. (2012) Genet. Sel. Evol. 44:42. https://doc.org/10.3389/fgene.2020.00880

219. Genetic characterization of the indigenous Sanga cattle of Namibia

D.A. Januarie[1,2*], E.D. Cason[2] and F.W.C. Neser[2]

[1]*Ministry of Agriculture, Water and Land Reform, Private Bag 13184, Windhoek, Namibia;*
[2]*Department of Animal Sciences, P.O. Box 339, University of Free State, Bloemfontein, 9300, South Africa;*
deidre.januarie@gmail.com

Abstract

The primary objective of this study was to determine the levels of genetic variability and inbreeding within the different ecotypes found north and south of the Veterinary Cordon Fence (VCF). Hair samples from the Kavango- (n=124), Caprivi- (n=135), Ovambo- (n=269), Kunene- (n=118), Herero (n=17) Sanga ecotypes as well as Nguni (n=148) cattle were collected and analysed using a 150K Single Nucleotide Polymorphism (SNP) chip. Marker and individual quality control, and linkage disequilibrium (LD) filtering were performed on a total of 120,700 markers and 837 animals. Observed (Ho) and expected (He) heterozygosities, and inbreeding coefficients (*F*) were calculated. Except for the Herero ecotypes, low positive *Fs* for the other ecotypes were observed. Expected heterozygosity values for Kavango = 0.325±0.0001, Caprivi = 0.316±0.00006, Nguni = 0.322±0.0002, Kunene = 0.331±0.0004, Ovambo = 0.321±0.0001 and Herero = 0.346±0.00004 were observed. PCA scatter plots and admixture (*K=9*) showed four (4) cluster groups, despite some admixture being reflected. *Fst* coefficients between pairs of populations were calculated and low *Fst* coefficients were however observed when compared to PCA plots.

Introduction

Cattle are usually grouped into two major categories namely *Bos Taurus* (taurine) and *Bos Indicus* (indicine) (Meyer, 1984; Groeneveld *et al.* 2010; Horsburgh *et al.,* 2013; Mwai *et al.* 2015; Madilindi *et al.,* 2020). Indigenous cattle such as the Nguni, Pedi, Tuli, Afrikaner, is collectively known as Sanga cattle (Meyer 1984; Strydom, 2008; Els *et al.* 2000). Sanga cattle are a combination of African taurine and African indicine (Makina *et al.*, 2016). Namibia does not have its own suitable name for the indigenous cattle and as such are also referred to as Sanga cattle (Els *et al.* 2000).

Most of the Namibian indigenous Sanga cattle are found in the Northern Communal Areas (NCA) of Namibia. According to Els (2001), the indigenous Sanga cattle are divided into different ecotypes namely Kunene, Ovambo, Kavango and Caprivi. Due to the presence of Foot and Mouth Disease (FMD) and Contagious Bovine pleurophneumonia (CBPP) or Lungsickness in the NCA, the North-South movement of all cloven-hoofed animals in Namibia are controlled by the Veterinary Cordon Fence (VCF). Thus creating isolated cattle populations of the same breed. Hanotte *et al.* (2010) is of the opinion that livestock that experienced selection pressure due to environmental conditions such as drought, display a greater genetic diversity across habitats when compared to a 'neutral' genomic background.

Very little information is available on the phenotypic and genetic levels of diversity of indigenous livestock in Africa (Rege, 1999; Mwai *et al.,* 2015) and in Namibia only Notier *et al.* (2002) looked at genetic diversity levels in indigenous Sanga cattle using 10 microsatellite markers. Recently the government of Namibia embarked on a genetic diversity study using SNPs (150K) aimed at characterizing the indigenous Sanga cattle of Namibia to further support the conservation of the vulnerable indigenous animals.

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_219, © D.A. Januarie *et al.* 2022

Materials & methods

Sample collection. A total of 837 hair samples were collected from unrelated indigenous Sanga/Nguni cattle ecotypes in all fourteen (14) regions in Namibia. The ecotypes included, Kavango (n=124), Caprivi (n=135), Nguni (n=148), Kunene (n=118) and Ovambo (n=269). An additional ecotype, Herero Sanga (n=17), was also collected. Most of the indigenous Sanga/Nguni ecotypes (Kunene, Ovambo, Kavango and Caprivi) occur in the NCAs, north of the VCF, while the Nguni and Herero ecotypes were collected south of the VCF. All the hair samples from the NCA stood a quarantine period of three (3) months and were baked in an oven for 30 minutes at 72 °C to sterilize the samples from FMD and CBPP as per requirements import permit.

Analysis of hair samples. All hair samples were analyzed at GeneSeek/Neogene (Nebraska, USA) and the Agricultural Research Council (ARC) in South Africa using a 150K Single Nucleotide Polymorphism (SNP) chip (Igenity and Illumina). Sample analysis was performed as described by King *et al.* (2021). Briefly, PLINK (1.07) (Purcell, 2010) software was used to perform statistical analysis. SNP quality control was performed where SNPs with missing genotypes in more than 5% individuals or less than 95% (--geno 0.05), with a minor allele frequency (MAF) of less than 10% (--maf 0.01) and that deviated from the Hardy-Weinberg equilibrium (--hwe 0.001) were excluded from the analysis. Observed (Ho) and expected (He) heterozygosities as well as inbreeding coefficients were calculated; highly inbred and crossbred animals were also removed. A total of twenty six (26) animals were excluded after quality control. Principle Component Analysis (PCA) scatter plots were done using R-Studio, an integrated development environment (IDE) for the programming language R (R Core Team, 2021). Admixture (1.3) (Alexander *et al.* 2020) (*K=9*) was used to determine the genetic relationship between and among populations. *Fst* coefficients were calculated using PLINK (1.9) (Purcell and Chang, 2019).

Results

Table 1 shows that low positive inbreeding coefficients *(F)* were observed for all the ecotypes, except for the Herero Sanga (exhibiting more heterozygous animals than inbred). Expected heterozygosities (He) varied from 0.316±0.00006 to 0.346±0.00004; Caprivi ecotype being the least heterozygous while the Herero Sanga cattle were the most heterozygous. The Kavango, Nguni, Kunene and Ovambo ecotypes were very similar in heterozygosity. Studies done by Makina *et al.* (2016), Madilindi *et al.* (2020) and King *et al.* (2021) showed Sanga breeds showed similar trends as seen with this data.

Four cluster groups were observed when using PCA scatter plots. The Kunene and Ovambo ecotypes clustered as one group, while the Kavango and Caprivi ecotypes clustered as another group. Nguni and Herero Sanga also cluster separately from the previous groups. Admixture (Figure 2) confirms similar results as seen with the PCA scatter plot (Figure 1). In a study by Madilindi *et al.* (2020) on three Nguni populations it was found that they did cluster independently even though there was some admixture.

Table 1. Inbreeding Coefficients, Observed (Ho) and Expected (He) Heterozygosities of the Indigenous Sanga of Namibia.

Ecotype	n	F	Ho ± std	He ± std
Kavango	124	0.0097	0.322±0.0141	0.325±0.0001
Caprivi	135	0.0113	0.313±0.0118	0.316±0.00006
Nguni	148	0.0110	0.318±0.0167	0.322±0.0002
Kaoko (Kunene)	118	0.0002	0.331±0.0158	0.331±0.0004
Ovambo	269	0.0045	0.319±0.0163	0.321±0.0001
Herero	17	-0.0515	0.364±0.0144	0.346±0.00004

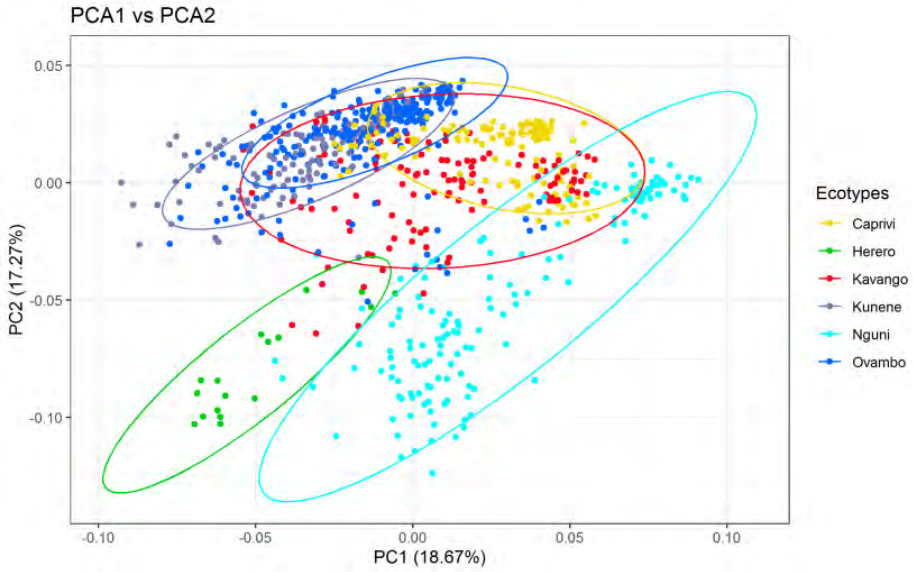


Figure 1. Cluster groups of the Indigenous Sanga cattle of Namibia.

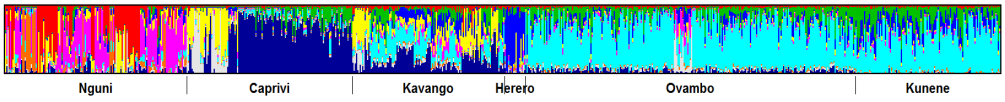


Figure 2. Admixture of Indigenous Sanga cattle in Namibia (*K*=9).

Low *Fst* coefficients showed a low level of differentiation (Table 2). This could possibly be ascribed to that all these Sanga ecotypes belong to one breed that has common ancestory (Maiorano *et al.*,2018; Bhuiyan *et al.*, 2021).

Discussion

The natural environment in which these ecotypes exist clearly plays a role in the genetic diversity of the animals. The Nguni and Herero Sanga cattle are found south of the VCF, the FMD free area where no restrictions on animal movement exist. The remaining ecotypes (except for those at the government research stations) are in the NCA. Caprivi ecotype is in Zambezi region, where animal movement to other regions are restricted (due to FMD). In the NCA, the cattle have adapted to different climates, from dry harsh north-

Table 2. Fst coefficients between the six Sanga ecotypes in Namibia.

	Kavango	Caprivi	Nguni	Kunene	Ovambo	Herero
Kavango						
Caprivi	0.0065					
Nguni	0.0097	0.0133				
Kunene	0.0081	0.0139	0.0165			
Ovambo	0.0058	0.0093	0.0142	0.0041		
Herero	0.0351	0.0513	0.0386	0.0347	0.0421	

western Namibia to wet north-eastern Namibia. In the NCA no direct selection for genetic improvement is practiced, thus natural selection occurs mostly. This is because subsistence farming is practiced, and no clear breeding plans and objectives exist. The result is that populations subjected to similar environments cluster closer together. Hanotte *et al.* (2010), Madilindi *et al.* (2020) and King *et al.* (2021) confirm these results in terms of adaptability to natural environment and selection due to the environment in which the animals exist. However, more genetic exchange also occurs between these regions.

References

Alexander, D. H., Lange, S. S. S. and Novembre, J. (2020) Admixture (1.3)

Bhuiyan, M. S. A., Lee, S., Hossain, S. M.J., Deb, G.K, Afroz, M.F *et al.* (2021) Animals 11: 2381. https://doi.org/10.3390/ani11082381

Els, J. F. (2001) Field Guide for the identification of the indigenous Livestock of Namibia. SADC/UNDP/FAO project.

Els, J. F., Bester, F. V., Jessen, P. T. and Reed, E. R. (2000) The Sanga Cattle. Manual for Agricultural Extensionists and Farmers.

Groeneveld, L.F., Lenstra, J.A., Eding, H., Toro, M.A., Scherf, B., *et al.* (2010) Anim. Genet. 41(Supple.1): 6-31.

Hanotte, O., Dessie, T., and Kemp, S. (2010). *Science*. 328: 1640-1641.

Horsburgh, K. A., Prost, S., Gosling, A., Stanton, J-A., Rand, C. *et al.* (2013). PLoS ONE. 8 (8): e71956. https://doi.org/10.1371/journal.pone.0071956

King, F. J. M., Banga, C. B and Visser C. (2021). Trop. Anim. Health Prod. 53:117. https://doi.org/10.1007/s11250-021-02562-0

Purcell, S. 2010. PLINK (1.07)

Purcell, S. & Chang, C. 2019. PLINK (1.9) www.cog-genomics.org/plink/1.9/

Madilindi, M.A., Banga, C. B., Bhebhe, E., Sanarana, Y. P., Nxumalo, K. S., *et al.* (2020) Trop Anim Health Prod. 52:753-762. https://doi.org/10.1007/s11250-019-02066-y

Makina, S. O., Whitacare, L. K., Decker, J.E., Taylor, J. F., MacNeil, M. D., Scholtz, *et al.* (2016) Genet. Sel. Evol. 48: 88

Maiorano, A.M., Lourenco, D.L., Tsuruta, S., Ospina, A.M.T., Stafuzza, N.B., *et al.* (2018) PLoS ONE 13(8):e0200694. https://doi.org/10.1371/journal.pone.0200694

Meyer, E. H. H. (1984) Proc. of the 2nd World Congress on Sheep and Beef Cattle Breeding. Pretoria, South Africa.

Mwai, O., Hanotte, O., Kwon, Y. and Cho S. (2015) Asian-Australas. J. Anim. 28(7): 911-921. http://dx.doi.org/10.5713/ajas.15.0002R

Nortier, C. L., Els, J. F., Kotze, A. and Van der Bank, F. H. (2002) Proc. 7th WCGALP, Montpellier, France.

R Core Team (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Rege, J.E.O. 1999. Animal Genetic Resource Information Bulletin No. 25: 1-25

Strydom, P. E. (2008) Meat Sci. 80: 86-93. https://doi.org/10.1016/j.meatsci.2008.04.017

220. Hedging genetic diversity using a similarity matrix based on Mendelian sampling

A.A. Musa* and N. Reinsch

Research Institute for Farm Animal Biology (FBN), Institute of Genetics and Biometry, Wilhelm-Stahl-Allee 2, 18196 Dummerstorf, Germany; musa@fbn-dummerstorf.de

Abstract

A major limitation of genomic selection criteria is the trade-off between short- and long-term genetic gains. A similarity matrix based on Mendelian sampling was recently proposed as an additional criterion to optimise this trade-off while hedging genetic diversity. However, its effect on genetic gain and diversity in breeding programmes under recurrent selection has not been investigated. Here the effects of the similarity matrix in conjunction with breeding value and an index combining breeding value and Mendelian sampling variance on genetic gain and diversity are investigated. We found no apparent effects of including the similarity matrix on genetic gains over generations. However, compared to the conventional breeding value and index selection, its inclusion preserved up to 6 and 19% of genetic diversity, respectively. This result is due to the similarity matrix's diversifying effect, which enables the selection of an optimal set of genetically diverse parents that maximises genetic gain in the next generations.

Introduction

Selecting parents of future generations based on their genomic breeding values (BV) maximise short-term genetic gains. Continuous selection on BV, on the other hand, accelerates the loss of genetic diversity, jeopardising long-term genetic gains. An index combining BV and Mendelian sampling variance (MSV) was recently proposed to optimise the trade-off between short and long term gains (Bijma *et al.*, 2020). MSV quantifies the expected variability among full-sibs and is vital if offspring variability is of interest. Under high selection intensity, the index tends to choose similar parents with high MSV potential, resulting in unintended losses of genetic diversity.

To hedge genetic diversity, a similarity matrix based on the Mendelian sampling values of parents was developed, in which the diagonal elements represent the similarity of a parent to itself, which equals its MSV (Musa, 2021; Reinsch and Musa, 2021). A high similarity between parents indicates that they share many heterozygous quantitative trait loci (QTL), with large effects on a trait in the same linkage phase. Similar to how the covariance and correlation matrices of asset prices are used for portfolio diversification in finance (Markowitz, 1952), the similarity matrix can be combined with selection criteria to obtain an optimal set of genetically diverse parents with high expected returns (Musa, 2021; Reinsch and Musa, 2021). However, its effect on genetic gain and diversity in a breeding programme under recurrent selection is yet to be investigated. Using a simulated cattle breeding programme, we investigate the potential of the similarity matrix combined with BV and an index comprising BV and MSV for increasing genetic gain and hedging genetic diversity in future generations. The results are compared to those obtained using BV and the index.

Materials & methods

Simulation of a cattle breeding programme. A base population of 1000 animals with 500 cows was simulated with a predefined parameter set for cattle using a coalescent process in AlphaSimR (Gaynor *et al.*, 2020). An average genetic map for both sexes with an architecture of 10 chromosomes with an average length of ~100 cM was assumed. For each chromosome, 1,200 segregating sites were assigned, from which 200 were QTL affecting a trait. The QTL effects are assumed to be additive, with a mean of

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_220, © A.A. Musa and N. Reinsch 2022

zero and a variance of 100. The heritability of the trait was assumed to be 0.25. Then we performed 20 generations of selection using various criteria. Each generation included the following steps: (1) calculation of selection criterion; (2) selection of the top 1% of bulls and 50% of cows as parents; (3) allotting mates with a maximum allocation of cows to a bull of 20%; and (4) mating to form a new generation of animals with four offspring per mate-pair, while maintaining a constant population size of 1000. Genetic gain and diversity of true breeding values were derived using 1000 animals. The entire simulation was repeated 100 times. The simulations and subsequent statistical analyses were performed in R software (R Core Team, 2020).

Selection criteria – BV. Let $\mathbf{z}_i'$ be a row vector of QTL genotype indicators (coded 1, 0, and -1 for respective genotypes *AA*, *AB* or *BA*, and *BB* at each locus) and m a vector of population additive QTL effects, then the BV of animal i is computed as $bv_i = \mathbf{z}_i'\mathbf{m}$ and for mate-pair ij as the mean BVs of animals i and j.

Selection criteria – Index. The index I_i of animal i is calculated according to Bijma *et al.* (2020) as $I_i = bv_i + \sqrt{2} \cdot x_p \cdot \sigma_{g_i}$ and for mate-pair ij as mean the mean indices of animals i and j. Here, x_p is the standardised truncation point belonging to the selected proportion p, and σ_{g_i} is the square root of MSV for animal i. Then MSV of additive genetic values of gametes produced by animal i is obtained using and $\sigma^2_{g_i} = \sum \left(\boldsymbol{m}_i^c\right)' \boldsymbol{R}^c \boldsymbol{m}_i^c$, where and $\boldsymbol{m}_i^c$ is the parent-specific vector of additive QTL effects for parent i on chromosome c and R^c is the expected within-family linkage disequilibrium matrix of QTL on the same chromosome (Bonk *et al.*, 2016; Musa, 2021). The element m_{ik}^c of the parent-specific vector $\boldsymbol{m}_i^c$ is given by the formula $m_{ik}^c = \delta_{ik} m_k^c$, where δ_{ik} and m_k^c are the phase indicator and QTL effect at locus k. If the standard allele is observed on the first haplotype, the phase indicator δ_{ik} is 1; otherwise, it is -1. For each homozygous genotype, the indicator is set to zero $\delta_{ik} = 0$. The element ρ_{kl}^c of R^c is calculated as $\rho_{kl} = \exp(-2d_{kl}^c)/4$ (Haldane, 1919), where d_{kl}^c is the genetic distance (Morgan) between QTL k and l on chromosome c.

Computation and application of similarity matrix. Let $\boldsymbol{m}_i^c$, $\boldsymbol{m}_j^c$, $\boldsymbol{m}_u^c$, and $\boldsymbol{m}_v^c$ be the parent-specific vector of QTL effects for parents i, j, u, and v, then the similarity $s_{ij,uv}$ between the Mendelian sampling values of zygotes produced by parent pairs ij and uv is given for a single trait as $s_{ij,uv} = \sum_c \left(\left| \left(\boldsymbol{m}_i^c\right)' \boldsymbol{R}^c \boldsymbol{m}_u^c \right| + \left| \left(\boldsymbol{m}_j^c\right)' \boldsymbol{R}^c \boldsymbol{m}_v^c \right| \right)$ (Musa, 2021). A similarity matrix S can be derived by assembling $s_{ij,uv}$. The diagonal and off-diagonal elements of matrix S represent MSVs and similarities between Mendelian sampling values of zygotes produced by mate pairs, respectively. Finally, matrix S was standardised to obtain a correlation matrix K within the range of $0 \leq k_{ij,uv} \leq 1$ (Musa, 2021).

The matrix K was then used in the same way as in the portfolio diversification approach in finance (Markowitz, 1952), in which the similarities between mate pairs were minimised for a given level of expected returns to obtain a portfolio that contains the optimal set of mate pairs. The expected genetic return of a portfolio was defined as $\Psi = n'p$, where vector p is the vector of BV or index for mate pairs ij and n is a vector of the relative weights allotted to mate-pairs ij. The optimisation problem can be stated as:

$$\min_n \boldsymbol{n'Kn}$$

Subject to $\Psi = \boldsymbol{n'p} \geq 0$

$$\boldsymbol{n'1} = q$$

$$n_{ij} \in \{0,1\}, _{ij} \in \{1, \ldots, N\}$$

where l is a vector of ones, q is the number of mate pairs to be selected, n_{ij} is a binary decision showing whether mate pair ij is selected (n_{ij}=1) or not (n_{ij}=0), and N is the total number of all possible mate pairs. The minimum and maximum allocations were set to 0 and 1, respectively, for unbiased comparison with other criteria. This optimisation was solved using *gurobi* (Gurobi Optimisation LLC, 2021) to obtain vector n at varying levels of expected return Ψ to derive an efficient frontier (EF). The EF offers several portfolios containing a set of optimal allocations of mate pairs with the lowest possible risk of losing genetic diversity (low similarities) for each level of expected returns.

Results & discussion

Figure 1A shows a subset of a similarity matrix for mate pairs, with the red boxes separating one bull from the other. Due to differences in heterozygosity and QTL linkage, some families have larger MSVs and thus produce more variable offspring than those with smaller MSVs, as indicated by the diagonal elements. Large MSVs are typically caused by heterozygous QTL that have large effects on a trait, whereas small effects cause small MSVs (Musa, 2021). Consequently, small-effect QTL may be lost by only selecting mate pairs with high MSV potentials. Mate pairs with large MSVs are more similar than those with small MSVs (see off-diagonals) because they share many heterozygous QTL with large effects on a trait in the same linkage phase. Hence, the similarity matrix is of practical importance for hedging and visualising genetic diversity. The standardised similarity matrix (Figure 1B) contains the same information as Figure 1A, except that the diagonals are one, and the off diagonals better highlight small similarities between mate pairs (Musa, 2021; Reinsch and Musa, 2021).

Figure 1C shows the EF, which provides several portfolios (blue and red points) containing an optimal set of genetically diverse mate pairs at a given expected return. Additionally, the EF illustrates the trade-off between expected returns and diversity (similarity), with expected returns increasing as the similarity between mate pairs and, consequently, among offspring increases. The choice of a portfolio depends on the breeders' goals and objectives. Nonetheless, the EF enables breeders to make the most informed decisions regarding the genetic diversity they are willing to sacrifice for greater genetic gain (Musa, 2021; Reinsch and Musa, 2021). Here we chose the portfolio in red with the highest similarity and expected returns.

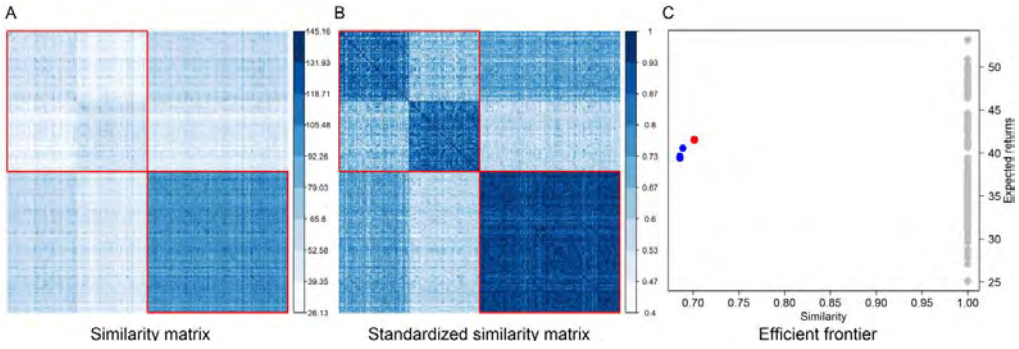


Figure 1. Similarity matrix (A), standardised similarity matrix (B) and efficient frontier (C) of matings. The red boxes on the similarity matrices (A and B) separate matings between bulls. The grey points represent the expected returns of each mate-pair. The blue and red coloured points (efficient portfolios) are called the efficient frontier (C), which provide a set of diverse mate pairs with low haplotype similarities at a given expected return level. The portfolio in red is the chosen portfolio in this study.

The genetic gains over generations (Figure 2A) were nearly identical for BV and index selection, with minor differences occurring only in later generations. There were no apparent effects of including the similarity matrix on genetic gains. Genetic diversity, however, was better preserved (10-55%) by index selection compared to selection solely on BV (Figure 2B). This effect was amplified by 3-19% when the index was combined with a similarity matrix (INDEX_SIM). In the case of BV selection, including the similarity matrix (BV_SIM) also had a beneficial effect, albeit to a much lesser extent (1-6%). These results are due to the similarity matrix' diversifying effect, allowing for the selection of a genetically diverse set of parents that maximises genetic gain while hedging genetic diversity in subsequent generations.

The similarity matrix can be included as an additional GS criterion to optimise the trade-off between short and long term genetic gains while hedging genetic diversity to sustain longer-term genetic gain.

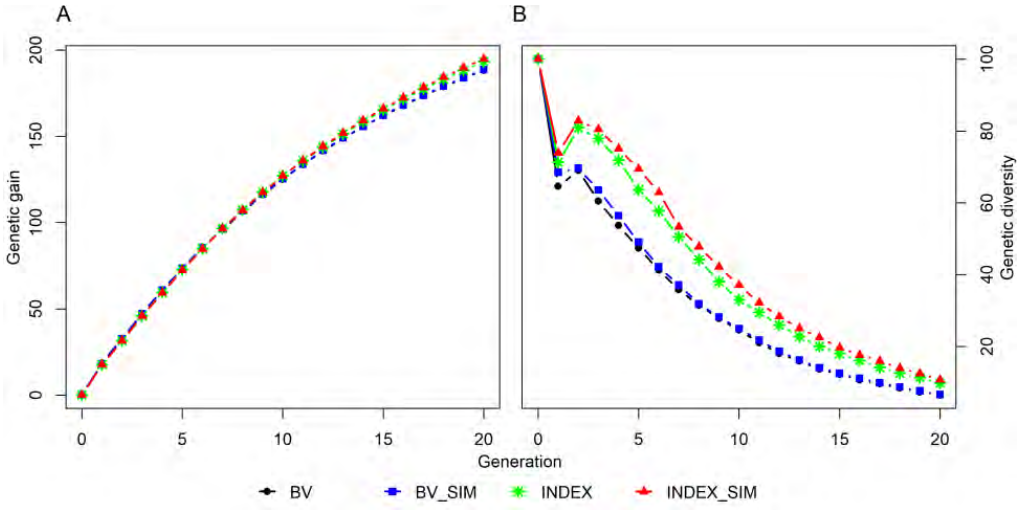


Figure 2. Genetic gain (A) and diversity (B) of true breeding values over 20 generations.

References

Bijma, P., Wientjes, Y.C.J., and Calus, M.P.L. (2020) Genetics 214(1):91-107. https://doi.org/10.1534/genetics.119.302643

Bonk, S., Reichelt, M., Teuscher, F., Segelke, D., and Reinsch, N. (2016) Genet. Sel. Evol., 48(1):1-11. https://doi.org/10.1186/s12711-016-0214-0

Gaynor, C.R., Gorjanc, J., and Hickey, J.M. (2020) G3 Genes, Genomes, Genet. 11(2):1-5. https://doi.org/10.1093/g3journal/jkaa017

Gurobi Optimization LLC (2021) Available at https://www.gurobi.com

Haldane, J.B.S. (1919) J. Genet., 8:299-309.

Markowitz, H. (1952) J. Finance 7(1):77-91.

Musa, A.A. (2021) Dissertation. Available at https://macau.uni-kiel.de/receive/macau_mods_00002304.

R Core Team (2020) Available at https://www.r-project.org

Reinsch, N., and Musa, A.A. (2021). Züchtungskunde, 93(3):201-209. https://rb.gy/2u6jcb.

221. Genetic basis of sexually divergent early life survival in red deer (*Cervus elaphus*)

L. Peters[1,2*], J.M. Pemberton[1] and S.E. Johnston[1]

[1]Institute of Evolutionary Biology, School of Biological Sciences, University of Edinburgh, Edinburgh, EH9 3FL, United Kingdom; [2]Laboratoire de Genetique Cellulaire, INRAE, 31320 Auzeville-Tolosane, France; lucy.peters@inrae.fr

Abstract

Males and females largely share the same genome, but the optimal value of a trait can differ between the sexes. If these traits share a genetic basis between the sexes, intra-locus sexual conflict can prevent them from reaching their optima. Here, we explore the genetic basis of sex-specific early life survival in wild red deer, a species with extensive sexual dimorphism, using a genomic dataset of >38,000 genotyped SNPs for 3,067 individuals. We conducted a variance component analysis using a Bayesian mixed model approach. Males were ~14% more likely to die before the age of two. Early life survival was significantly heritable (h^2=~0.1) and influenced by maternal effects. No SNPs were significantly associated with sex-specific survival to the second year and a polygenic genetic architecture is likely. Overall, this study gives valuable insight into the genetic architecture of sexually-dimorphic traits and its contribution to maintaining genetic variance.

Introduction

In species with separate sexes, males and females often follow very different life-history strategies. In many species, females invest more heavily in parental care compared to males, e.g. through gestation and food provisioning, whereas males invest more heavily in maximising their reproductive success (Clutton-Brock, 1991; Schärer *et al.*, 2012). This can lead to diverging fitness optima in the two sexes (where fitness is defined as total lifetime breeding success or its components such as survival and fecundity) and results in males and females differing substantially in a range of physical and behavioural traits. As a result, the genetic architecture of such traits can be sex-specific and play an important role in maintaining genetic polymorphisms, which has been documented for some major effect loci (Barson *et al.*, 2015; Rostant *et al.*, 2015). However, in most cases the genetic architecture of sexually divergent traits and thus their effect on genetic variance is unknown.

Red deer show extensive sexual dimorphism in morphology. Due to intense male competition, traits such as body size and fighting ability are important in stags, leading to significantly higher birth weight and faster growth rates in male calves (Clutton-Brock, 1985; Kruuk *et al.*, 1999). However, male calves also have lower fat storage compared to female calves, and consequently often much lower first winter survival (Clutton-Brock, 1985; Gauzere *et al.*, 2020). As a result, male and female red deer have different fitness optima for a number of traits, including those affecting survival. Here, we use long-term individual data on survival and a genomic dataset from red deer on the Isle of Rum, Scotland, to investigate the genetic basis of sexually divergent early life survival and its role in maintaining genetic variance.

Materials & methods

Study system and genomic data-set. The red deer study population is situated in the North Block of the Isle of Rum, Scotland (57°02'N, 6°20'W) and has been subject to individual monitoring since 1971. Deer calves are marked with ear tags shortly after birth to enable recording of detailed life histories of individuals. DNA is routinely extracted from neonatal ear punches. DNA samples were genotyped on the Cervine Illumina BeadChip (Brauning *et al.*, 2015) using an Illumina genotyping platform and Illumina GenomeStudio software (Illumina Inc., San Diego, CA, USA). A total of 3,067 pedigreed individuals

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_221, © L. Peters *et al.* 2022

genotyped at 39,587 polymorphic SNP loci were available for analysis. In addition, a linkage map specific to the Rum population is available, with 38,083 SNPs assigned to linkage groups corresponding to the 33 deer autosomes and X chromosome (Johnston *et al.*, 2017).

Imputation of missing genotypes. Missing genotypes in the genomic dataset were imputed using AlphaImpute version 1.98 (Hickey *et al.*, 2012) to prepare the dataset for subsequent analyses. The following parameters were used: threshold of non-missing SNPs for an animal to be included in high density group = 50%; threshold for allowed disagreement between cores of two surrogate parents = 0.5; five phasing iterations were run, the size of cores including tails being 100, 200, 300, 400 and 500 and the respective core sizes 50, 150, 250, 350 and 450.

Bayesian mixed model analysis of early life survival. To identify sexually antagonistic SNPs that affect early life survival, a Bayesian general linear mixed model approach was adopted using the R package *BGLR* version 1.0.8 (Pérez and De Los Campos, 2014). Survival to the second year of life was used as a binomial response variable, coded 1 if the animal survived and 0 if it died. All models employed the following formula:

$$y = Xb + Z1\ a + Zr\ ur + e \quad (1)$$

where y is a vector of survival to the second year, X is an incidence matrix relating individual measures to the vector of fixed effects b; Z1 and Zr are incidence matrices relating individual measures to additive genetic and other random effects respectively; a is a vector of SNP effects or the genomic co-variance matrix; ur is a vector of additional random effects; and e is a vector of residual effects. Models were run both with sex as a fixed effect and as separate models for each sex. Models were selected based on their genomic prediction accuracy i.e. the ability of an individual's genotypes to predict its phenotype. The genomic prediction accuracy was calculated as the area under the curve (AUC) of a Receiver Operating Characteristic (ROC) curve. AUC calculation and significance testing was performed in R using the *pROC* package version 1.17.0.1 (Robin *et al.*, 2011). The genetic priors tested were the BayesA model, which assumes all markers have non-zero effects and uses a scaled-t distribution, and the BayesB model, which implements a mixture of a distribution with highest density at 0 and a scaled-t distribution (Meuwissen *et al.*, 2001). The additional random effects tested included birth year, second winter year, and mother ID. For the non-genetic random effects, a Bayesian Ridge Regression (BRR) model was chosen, which uses a Gaussian prior. Additionally, Bayesian Reproducing Kernel Hilbert Spaces Regressions (RKHS) models using a genomic co-variance matrix was run to determine the heritability of early life survival; a full model with both sexes and two sex specific models were run. All models were run with the default parameters except for the number of iterations, which was set to 120,000 with a burn-in of 20,000; the thinning interval was kept the default value of 5.

Results

After model selection the best model was a BayesB model with additional random effects to account for the influence of common environment (birth year) and maternal effects (mother ID). Males had on average a ~14% lower probability of survival to the second year compared to females (effect = -0.1407, 95% confidence interval (CI) = [-0.1998; -0.0807]). The heritability of early life survival was substantial (0.10) as estimated by the RKHS model, which included sex as fixed effect and the environmental and maternal random effects. The maternal effect on survival explained ~14% of total phenotypic variance, which was slightly larger than the additive genetic effect. The environmental effect of birth year had larger uncertainty associated with its estimate. Sex specific variance components suggested maternal effects have a stronger effect on male than on female survival (Table 1). None of the SNP effect sizes in either the model containing

Table 1. Variance components of the RKHS *BGLR* model for early life survival.

Variable	Variance (95% CI)	Proportion Vtotal[1](95% CI)
Additive genetic	0.3184 (0.1410; 0.5697)	0.1024 (0.0496; 0.1664)
Mother ID	0.4037 (0.2357; 0.6201)	0.1320 (0.0746; 0.2020)
Birth year	1.3810 (0.7087; 2.4529)	0.4350 (0.2983; 0.5821)
Additive genetic female	0.3487 (0.1390; 0.6910)	0.1446 (0.0649; 0.2525)
Mother ID female	0.3198 (0.1297; 0.6195)	0.1340 (0.0571; 0.2381)
Birth year female	0.7078 (0.3466; 1.3113)	0.2916 (0.1757; 0.4332)
Additive genetic male	0.3077 (0.1136; 0.6567)	0.1128 (0.0464; 0.2126)
Mother ID male	0.5055 (0.2255; 0.9057)	0.1867 (0.0910; 0.3003)
Birth year male	0.8818 (0.4438; 1.6066)	0.3208 (0.2024; 0.4649)

[1] Total phenotypic variance.

sex as fixed effect or in the two sex specific models had 95% CIs that did not contain zero, and therefore no SNP could clearly be identified as having an effect on survival (Figure 1).

Discussion

Survival to the second year in red deer was significantly heritable, with additive genetic variance explaining about 10% of the total phenotypic variance. This was similar to previous findings for this population that showed first and second winter survival separately each had a heritability of about 5% (Gauzere *et al.*, 2020). However, no individual loci with significant effects on post-natal early life survival were identified. Therefore, although early life survival clearly has a genetic component, the trait is likely polygenic with many loci of small effect rather than a small number of loci explaining most of the genetic variance. This kind of genetic architecture is typical for fitness-related traits, where additive genetic variance is thought to be maintained by high mutational variability due to large number of loci influencing survival (Houle *et al.*, 1996; Merilä and Sheldon, 1999). Thus, genetic variance can be maintained even in the face of strong selection if a trait is highly polygenic (Barton and Keightley, 2002). There was a strong sex effect on early life survival probability, with female offspring ~14% more likely to survive to their second year compared to their male counterparts, confirming earlier findings on sex differences in early life survival in the red deer (Clutton-Brock, 1985). Additionally, maternal effects were also found to play a significant role in offspring survival, explaining ~14% of the overall phenotypic variance in the model including both sexes and this effect was stronger for males in the sex specific models than in females (~19% compared to ~13%). This is in line with earlier findings that mothers invest more heavily in their male offspring (Clutton-Brock, 1991).

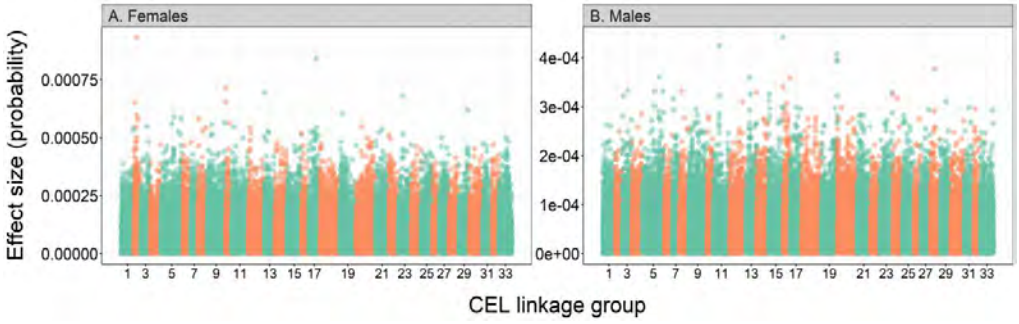


Figure 1. SNP effect sizes from sex-specific BGLR models for females (panel A) and males (panel B). Effect sizes are given as absolute change on a probability scale.

Maternal effects can weaken the link between an individual's genotype and their phenotype, thus adding to the conservation of genetic variance by moderating the influence of selection (Yamamichi and Hoso, 2017). A polygenic architecture together with maternal effects likely contributes to the maintenance of genetic variance for early life survival in this population.

References

Barson N.J., Aykanat T., Hindar K. *et al.* (2015) Nature 528(7582):405–408. http://doi.org/10.1038/nature16062

Barton N.H., Keightley P.D. (2002) Nat. Revi. Genet. 3(1):11–21. https://doi.org/10.1038/nrg700

Brauning R., Fisher P.J., McCulloch A.F., *et al.* (2015) bioRxiv.

Clutton-Brock T.H. (1985) SciAm 252(2):86–92. https://doi.org/10.1038/scientificamerican0285-86

Clutton-Brock TH (1991) The Evolution of Parental Care. Princeton University Press Princeton, US

Gauzere J., Pemberton J.M., Morris S., Morris A., Kruuk L.E., *et al.* (2020) Evolution 74(7):1378–1391. https://doi.org/10.1111/evo.14000

Hickey J.M., Kinghorn B.P., Tier B., Van Der Werf J.H. and Cleveland M.A. (2012) Genet. 44(1):1–11. https://doi.org/10.1186/1297-9686-44-9

Houle D., Morikawa B. and Lynch M. (1996) Genetics 143(3):1467–1483. https://doi.org/10.1093/genetics/143.3.1467

Johnston S.E., Huisman J., Ellis P.A. and Pemberton J.M. (2017) G3 7(8):2859–2870. https://doi.org/10.1534/g3.117.044198

Kruuk L.E., Clutton-Brock T.H., Rose K.E. and Guinness F.E. (1999) Proc. Royal Soc. B 266(1429):1655–1661. https://doi.org/10.1098/rspb.1999.0828

Merilä J. and Sheldon B.C. (1999) Heredity 83(2):103–109. https://doi.org/10.1046/j.1365-2540.1999.00585.x

Meuwissen T.H., Hayes B.J. and Goddard M.E. (2001) Genetics 157(4):1819–1829. https://doi.org/10.1093/genetics/157.4.1819

Pérez P. and De Los Campos G. (2014) Genetics 198(2):483–495. https://doi.org/10.1534/genetics.114.164442

Robin X., Turck N., Hainard A., *et al.* (2011) BMC Bioinform. 12(1):1–8. https://doi.org/10.1186/1471-2105-12-77

Rostant W.G., Kay C., Wedell N. and Hosken D.J. (2015) BMC Bio. 13(1):34. https://doi.org/10.1186/s12915-015-0143-3

Schärer L., Rowe L. and Arnqvist G. (2012) Trends Ecol. Evol. 27(5):260–264. https://doi.org/10.1016/J.TREE.2011.12.006

Yamamichi, M. and Hoso, M. (2017) Evolution 71(2):449-457. https://www.jstor.org/stable/48575963

222. Changes in genomic inbreeding over half a century in Swedish Red and Swedish Holstein cattle

S. Eriksson*, E. Strandberg and A.M. Johansson

Department of Animal Breeding and Genetics, Swedish University of Agricultural Sciences, P.O. Box 7023, 75007 Uppsala, Sweden; susanne.eriksson@slu.se

Abstract

Over the last half century, new artificial insemination technology, methods for genetic evaluation, and breeding material from other countries have been introduced in Swedish cattle breeding. In this study, we aimed at investigating changes in homozygosity and genomic inbreeding over time in commercial dairy breeds. Genotype data from 50K and 150K SNP arrays for in total 5,636 Swedish Red and 3,115 Swedish Holstein bulls born 1956-2015 were analysed. The highest levels of both average homozygosity and inbreeding coefficient based on runs of homozygosity were seen in the earliest period of the data, although rapidly decreasing until the 1980s. After a period of more stable estimates, the inbreeding coefficients increased in both breeds after the introduction of genomic selection around 2010. Principal component analysis of relatedness showed a gradual transition over time for both breeds, but most pronouncedly for Swedish Holstein that was much influenced by foreign Holstein bulls.

Introduction

Commercial dairy cattle breeds are intensively selected to suit current or predicted future market needs. At different time periods, new techniques and breeding regimes have been implemented for Swedish dairy cattle, such as the use of frozen semen for artificial insemination in the 1950s, and the introduction of the Interbull genetic evaluation of bulls in 1994 and the joint Nordic breeding values in 2005. The method best linear unbiased prediction (BLUP) was implemented for genetic evaluation in 1984, and in more recent years, genomic selection has been implemented as an even more efficient selection tool (Bengtsson *et al.*, 2020). Along with these improvements, the need to monitor genetic diversity has increased (Forutan *et al.*, 2018).

Today, the Swedish Red (SR) is a synthetic breed, largely influenced by Finnish, Norwegian and Danish red breeds, but that also shows genetic similarities to some old native Swedish cattle breeds (Upadhyay *et al.* 2019). The Swedish Holstein (SH) started as a strain of Friesian cattle (SLB), but the influence of Holstein increased from the 1960s until an almost complete Holsteinization of the breed. With the aim to study consequences of such developments, we estimated and compared different measures of genetic diversity, and inbreeding in Swedish commercial dairy bulls during different time periods since the 1950s.

Materials & methods

In total 5,636 SR and 3,115 SH bulls born from 1956 to 2015 were genotyped (Table 1). Most bulls (3,032 SH and 5,545 RDC) were genotyped using the 50K Illumina SNP array (v1 or v2) and their genotypes were obtained from the Nordic Cattle Genetic Evaluation (NAV). In addition, 83 old HOL and 90 old SR bulls were genotyped using the 150K Genomic Profiler SNP array from GeneSeek®. In total 39,161 autosomal SNPs overlapping between the different arrays, with a non-zero minor allele frequency, were kept for further analyses.

Calculation of expected and observed level of homozygosity (Ho_{exp} and Ho_{obs}), runs of homozygosity (ROH), ROH-based inbreeding coefficient F_{ROH}, inbreeding coefficient based on deviation of observed from expected homozygosity (based on allele frequencies of bulls born before 1980) (F_{dev}), and principal

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_222, © S. Eriksson *et al.* 2022

Table 1. Number of genotyped bulls born in different years for Swedish Red (SR) and Swedish Holstein (SH).

	Birth year								
Breed	<1965	1965-1979	1980-1984	1985-1989	1990-1994	1995-1999	2000-2004	2005-2009	2010-2015
SR	33	54	42	266	425	383	351	452	3627
SH	25	33	19	60	90	146	177	592	1973

component analysis (PCA) was done using PLINK v1.9 (Purcell *et al.*, 2007). Ho_{obs} was defined as the proportion of homozygous SNPs and the deviation from expected homozygosity was estimated using the --het command, using the allele frequencies of bulls born before 1980. For ROH definition, a window size of 40 SNPs within 1000 kb, minimum SNP density of 75 (kb/SNP), up to 500 kb allowed gap between SNPs, 0.08 as proportion overlapping windows required to be homozygotic, 1 missing and 0 heterozygotic SNPs allowed per window, were used. F_{ROH} for each individual was calculated as the aggregate length of ROH segments in kbase pairs divided by the total length of the genome covered by SNPs, and the resulting value was adjusted by dividing with the F_{ROH} (0.9197) that a simulated completely inbred individual would get in the analysis.

A random subset of 395 SR and 236 SH bulls per birth decade starting from 1990 and 2000, respectively, were used in the PCA, to avoid bias due to the large number of genotyped individuals born in later years.

Results

Higher average homozygosity was estimated for SR than SH bulls in all studied time periods, but the difference was very small in later years (Figure 1). The inbreeding coefficient based on deviation from expected homozygosity (F_{dev}) was positive in the earliest years in both breeds, then negative after 1980, but increasing towards 0 from 1990s in SH (Figure 1). The F_{ROH} were similarly showing a higher level of inbreeding before the 1980s, that rapidly declined until the 1990s, stayed rather stable until 2009 and then started to increase again in both breeds in the last time period (Figure 1). This measure of inbreeding was higher for SR than SH before the 1980s, but from the 1990s the F_{ROH} was slightly higher in the SH than in the SR bulls.

The PCA-plot for each breed showed a gradual but distinct pattern related to the birth time periods of the bulls, from until the mid-1960s to 2015 (Figure 2). This pattern in relatedness within breed was mainly explained by the first principal component, and was more marked for the SH than for the SR bulls.

Discussion

The reduction in generation interval as a consequence of the introduction of genomic selection has been suggested to be the main cause of increased inbreeding rates per year in dairy cattle in recent years. In contrast, the use of pedigree-based BLUP would theoretically give a higher inbreeding rate per generation (Forutan *et al.*, 2018). In the Swedish commercial populations, a more rapid increase in F_{ROH} was as expected seen after 2009, following the introduction of genomic selection. It is noticeable that the highest levels of F_{ROH} were seen in the oldest bulls when breeding was a local business, but declined rapidly when the international exchange of genetic material increased. In practice, the old Swedish Friesian cattle (SLB) was crossbred with, and gradually replaced by, American Holstein (Bett *et al.*, 2010), and this transition is visible in the PCA plot.

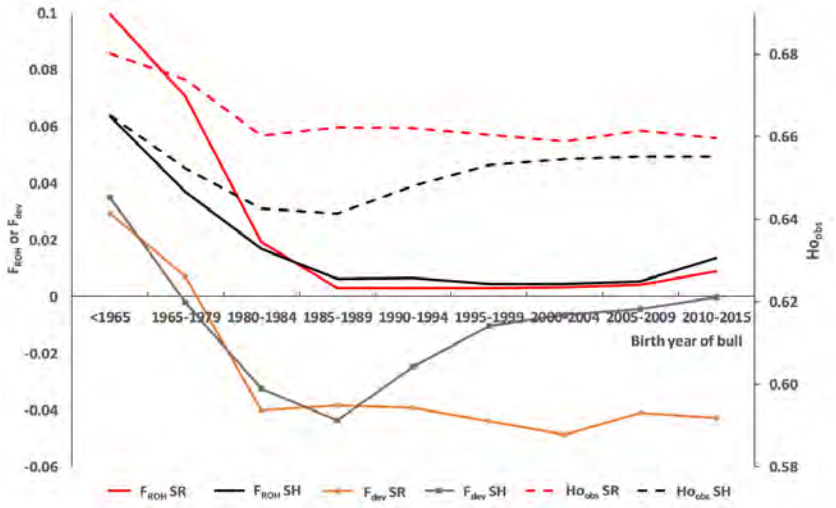


Figure 1. Average inbreeding coefficient based on runs of homozygosity (F_{ROH}), and on deviation from expected homozygosity (F_{dev}), and average observed homozygosity (Ho_{obs}) for Swedish Red (SR) and Swedish Holstein (SH) bulls born in different time periods.

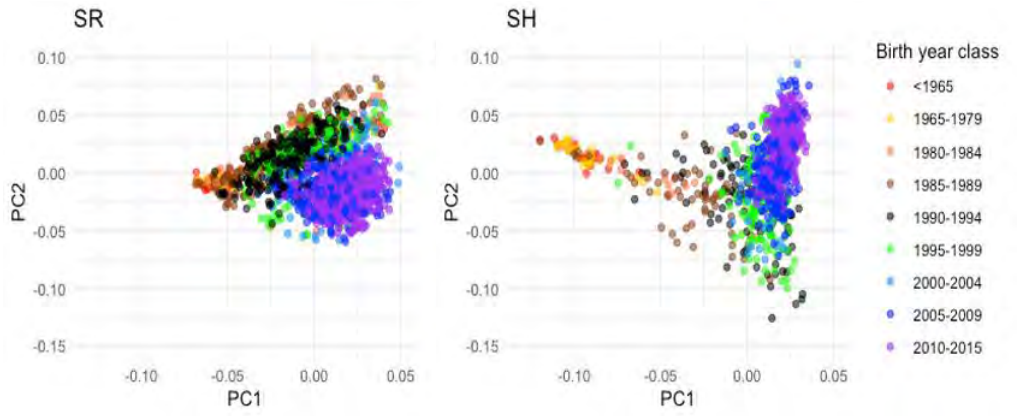


Figure 2. Principal component (PC) analysis of genetic relatedness between bulls born in different years, for Swedish Red (SR) and Swedish Holstein (SH). PC1 and PC2 explained 2.1 and 1.2% of the variation for SR, and 2.9 and 1.1% for SH, respectively.

At about the same time that BLUP was introduced in Swedish cattle breeding, the decrease in average F_{ROH} stopped, and the trend became more stable. This could also be due to a reduced effect of crossbreeding, especially in the SH in which the proportion of Holstein at that time was large. The continued transition also in later years shown mainly in the second principal component in the PCA-plot would thus likely be due to intensive selection for favourable genotypes.

Most studies of inbreeding trends that go far back in time are based on pedigree information. The low pedigree completeness in historical data, such as illustrated by Makanjuola *et al.* (2020), may cause under-estimation of inbreeding levels in the oldest animals, which can be avoided by using genomic information. Due to recombination, old inbreeding is seen as shorter ROH than recent inbreeding, and

genomic measures such as F_{ROH} do not accumulate inbreeding over time to the extent that homozygosity and pedigree-based measures do (Meuwissen *et al.*, 2020). The current levels of F_{ROH} in both SR and SH appear to be considerably lower than in the Canadian Holstein population studied by Makanjuola *et al.* (2020), where it increased from around 0.1 in 1990 to around 0.17-0.19 in 2018. Martikainen *et al.* (2017) found F_{ROH} ranging from 0.04 to 0.09 in Finnish Red cows born 2002-2014, which is more comparable, but still higher than our results. However, their estimate of Ho_{obs} was 0.63, which was lower than our values. The actual level notwithstanding, our results show that the inbreeding level was considerably higher half a century ago than at present, but that the inbreeding level in both SR and SH is currently increasing, which should be considered in breeding strategies.

Acknowledgements

We thank NAV and VikingGenetics (VG) for providing genotype data and VG for old bull sperm samples, with a special thanks to Hans Stålhammar who helped to select the samples. The study was funded by the Swedish research council Formas.

References

Bengtsson C., Stålhammar H., Strandberg E., Eriksson S., Fikse W.F. (2020) J. Dairy Sci. 103(7): 6383-6391. https://doi.org/10.3168/jds.2019-17963

Bett R.C., Johansson K., Zonabend E., Malmfors B., Ojango J. *et al.* (2010). Proc. of the 9th WCGALP, Leipzig, Germany.

Forutan M., Ansari Mahyari S., Baes C., Meltzer N., Schramm Schenkel F., *et al.* (2018) BMC Genomics 19, 98. https://doi.org/10.1186/s12864-018-4453-z

Makanjuola B.O., Miglior F., Abdalla E.A., Maltecca C., Schenkel F.S. *et al.* (2020) J. Dairy Sci. 103(6):5183-5199. https://doi.org/10.3168/jds.2019-18013

Martikainen K., Sironen A., Uimari P. (2017) J. Dairy Sci. 101(12):11097-11107. https://doi.org/10.3168/jds.2018-14805

Meuwissen T.H.E., Sonesson A.K, Gebregiwergis G., and Woolliams J.A. (2020) Front. Genet. 11:880. https://doi.org/10.3389/fgene.2020.00880

Purcell S., Neale B., Todd-Brown K., Thomas L., Ferreira M.A.R., *et al.* (2007) Am. J. Hum. Genet., 81(3): 559-575. https://doi.org/10.1086/519795

Upadhyay M., Eriksson S., Mikko S. Strandberg E., Stålhammar H, *et al.* (2019) Genet. Sel. Evol. 51, 56. https://doi.org/10.1186/s12711-019-0496-0

223. Runs of homozygosity in Nguni and Bonsmara cattle populations

P. Soma[1*], B.B. Kooverjee[1] and M.M Scholtz[1,2]

[1]Agricultural Research Council, Animal Production Campus, Private Bag X2, Irene, 0062, South Africa; [2]Department of Animal, Wildlife and Grassland Sciences, University of the Free State, Bloemfontein, South Africa; pranisha@arc.agric.za

Abstract

Runs of homozygosity (ROH) are used in livestock research for identification of homozygous genomic regions and are valuable to characterize inbreeding. The aim of this study was to use BovineSNP150K data to investigate inbreeding levels and occurrence of ROH in indigenous Bonsmara and Nguni cattle populations. 506 genotypes were included for analysis. Quality control was performed using PLINK. ROH were computed using the R package detectRUNS. A total of 17,411 ROH in total with mean number of ROH per individual per breed ranging from 203 to 397 were obtained. Both breeds displayed low ROH based inbreeding (F_{ROH}) of 0.003±0.002 for Nguni and 0.012±0.005 for the Bonsmara population. Within the reported ROH, genes relating to heat tolerance and immunity were detected in both breeds, while those relating to growth was found only in the Bonsmara population. These breeds provide valuable insights for understanding the genetic basis of adaptive traits.

Introduction

Runs of homozygosity (ROH), an indicator of genomic autozygosity, are defined as long and continuous stretches of homozygous genotypes (Ku *et al.*, 2011). ROH results from parents transmitting identical haplotypes to their offspring. This autozygosity may be present in inbred as well as non-inbred populations due to several population occurrences such as inbreeding, genetic drift as well as natural and artificial selection. Over the past two decades, there has been a rapid development of high-throughput genomic analysis methods, including next-generation sequencing (NGS) and genotype-based microarrays. Both methods allow for effective identification of ROH. However, single nucleotide polymorphisms (SNP) microarrays remain most popular for determining ROH due to its lower cost and less-time consuming analysis, compared to NGS. SNP arrays in cattle have been successfully used to estimate inbreeding (Zhang *et al.*, 2013). In addition, ROH was found to be effective in distinguishing non-autozygotic identical by state (IBS) segments from autozygotic identical by descent (IBD). The frequency and length of ROH are important parameters in determining the causative forces of genomic change. Therefore, the identification of ROH may assist in revealing the population structure of a population. The Nguni, an indigenous South African (SA) breed has many beneficial traits such as parasite resistance, heat and drought tolerance. Bonsmara, a composite breed performs well in beef production systems due to its high fertility, calm temperament and easy handling (van Marle-Köster *et al.,* 2021). Thus, the aim of this study was to determine ROH in indigenous Nguni and Bonsmara cattle populations.

Materials & methods

Ethical approval. Ethical approval was obtained from the Agricultural Research Council, Animal Production Ethics Committee (APIEC21/06).

SNP data and quality control. Raw SNP data using GeneSeek Genomic Profiler 150K was obtained for two SA cattle breeds. The quality control parameters involved removing SNPs with less than 95% call rate, SNPs less than 2% minor allele frequencies (MAF), samples with more than 10% missing genotypes and SNPs that deviated from Hardy Weinberg Equilibrium using PLINK (*P*£0.001) (Purcell *et al.*, 2007). Closely related individuals with a PI_HAT value greater than 0.25 were removed. Following quality control,

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_223, © P. Soma *et al.* 2022

27 individuals and 40,224 SNPs were removed from the Nguni population, while 32 individuals and 28,409 SNPs removed from the Bonsmara population.

ROH analysis and Gene annotation. ROH were determined using the R package detectRUNS (Biscarini *et al.*, 2018). The minimum length that constituted the ROH was set to 1 Mb. The following parameters were used to define the ROH: (1) one missing SNP and up to one possible heterozygous genotype was allowed in the ROH; (2) the minimum number of SNPs that constituted the ROH was set to 50; (3) the minimum SNP density per ROH was set to one SNP every 100 Kb; and (4) the maximum gap between consecutive homozygous SNPs was 250 Kb. The inbreeding coefficient (F_{ROH}) was estimated based on ROH for each animal and averaged per breed. F_{ROH} was calculated within detectRUNS using the following formula:

$$F_{ROH} = L_{ROH} / L_{AUTO}$$

where L_{ROH} is the total length of ROH on autosomes and L_{AUTO} is the total length of the autosomes covered by SNPs (Biscarini *et al.*, 2018). Analyses of common ROH revealed SNPs most common in both populations by using the tableRUNS function with the threshold set at 0.2 found in the detectRUN R package. Results indicated the top ROH island positions and number of SNPs. Genes within these islands were identified using Ensembl BioMart (Kasprzyk, 2011). The enrichment analysis was performed using PANTHER with Bonferroni correction ($P<0.05$) to group genes based on function or processes unique to each population.

Results

In Table 1, the number of animals and SNPs remaining after quality control along with the inbreeding coefficient is shown.

The genes that are unique to the two breeds are shown in Table 2 and the enrichment analysis of gene ontology terms for biological processes and molecular function are shown in Table 3.

Discussion

ROH provides useful information with regard to genetic relatedness among individuals, assisting in managing the inbreeding rate in a population, revealing possible deleterious variants in the genome. The calculation of the inbreeding coefficient (F_{ROH}) from ROH was reported to be more accurate for estimating autozygosity and detecting historic and recent inbreeding effects, rather than estimating inbreeding from pedigree data (Alemu *et al.*, 2021). F_{ROH} for the Nguni was lower than the Bonsmara. This may be attributed to Nguni being an indigenous breed with higher genetic diversity than Bonsmara cattle (van Marle-Köster *et al.,* 2021). Other studies reported higher levels of inbreeding in cattle (Dixit *et al.*, 2020). Findings in this study are consistent to that reported by Soma *et al.*, 2021, whereby low levels of inbreeding were detected using microsatellite data. Genes relating to heat tolerance (*DNAJCB5, DNAJB1*) and immunity (*IRF1, STING1*), were detected in both populations, while growth (*GHR*) related genes were identified in the Bonsmara. DNAJC18 gene, plays a role in heat stress, was detected only in Nguni. The gene ontology term chaperone-mediated protein folding (GO: 0061077) and unfolded protein binding (GO:0051787) present in the Bonsmara population, supports the presence of heat shock proteins (*HSPH1, HSPB2, HSPA9*). While the

Table 1. Mean and Standard deviation of ROH based inbreeding (F_{ROH}) of Nguni and Bonsmara cattle.

Breed	Number of Animals	Number of SNPs	Mean ± SD F_{ROH}
Nguni	231	100,952	0.003±0.002
Bonsmara	252	110,566	0.012±0.005

Table 2. Genes identified within the common runs of homozygosity in Nguni and Bonsmara populations.

Chr	nSNP	from	to	Genes
Nguni				
14	22	23939005	24014579	*DNAJC5B*
20	99	5395165	5826896	*GHR*
5	82	48094474	48494942	*HSP90B1*
5	80	48555140	49104631	*DNAJB7*
6	6	39367782	39391155	*CABS1*
7	75	434978	833790	*HSPA4, HSPA9,*
7	95	50712879	54632239	*DNAJC18, STING1*
Bonsmara				
11	23	56825445	57184891	*HSPA5*
11	22	57948246	58576022	*DNAJC27, DNAJC5G*
12	62	27807391	29372191	*DNAJC3, HSPH1*
14	36	41828105	42950017	*DNAJC5B*
15	21	56842162	57523769	*DNAJB13, DNAJC24, HSPA8, HSPB2*
20	84	5591162	5826896	*HSPB3, MROH2B*
25	58	1127441	2164566	*CAHSP1, DNAJA3*
5	61	1,12E+08	1,13E+08	*DNAJB7*
7	20	434978	599836	*HSPA4, HSPA9*
7	50	23112707	23327925	*DNAJB1, IRF1*

Table 3. Significant gene ontologies identified in the Nguni and Bonsmara population (Bonferonni correction $P<0.05$).

Gene ontology ID	Breed	Reference genes[1]	Observed genes	Expected	Over/under[2]	Fold	*P*-value
GO biological process							
GO:0051085	Bon	32	5	0,06	+	82 ,84	6 ,00E-05
GO:0051084	Bon	35	5	0,07	+	75 ,74	9 ,02E-05
GO:0006458	Bon	37	5	0,07	+	71 ,65	1 ,16E-04
GO:0061077	Bon	69	6	0,13	+	46 ,10	4 ,38E-05
GO:0010862	Ngu	48	4	0,09	+	44 ,18	2 ,14E-02
GO:0060393	Ngu	57	4	0,11	+	37 ,21	4 ,07E-02
GO:0006457	Bon	184	9	0,35	+	25 ,93	6 ,20E-07
GO molecular function							
GO:0051787	Bon	24	3	0,05	+	66 ,27	4 ,52E-02
GO:0051082	Bon	98	5	0,18	+	27 ,05	3 ,65E-03

[1] *Bos taurus* genome assembly used as reference.

[2] Over representation (+), Under representation (-); Fold enrichment (Fold); Bonsmara (Bon); Nguni (Ngu).

enriched term regulation of pathway-restricted SMAD protein phosphorylation (GO:0051082) identified in the Nguni supports the presence of GHR, a growth hormone receptor targeted during phosphorylation. Results of this study may assist in obtaining new insights into the genetic characteristics of these two cattle breeds. Findings may further contribute to future beef cattle breeding even in uncontrolled low input commercial, emerging and communal production systems.

Acknowledgements

The support of the Gauteng Department of Agriculture & Rural Development is acknowledged under project ID P02000181.

References

Alemu, S.W., Kadri, N.K., Harland, C., Faux, P., Charlier, C., *et al.* (2021) Heredity 126(3):410-423. https://doi.org/10.1038/s41437-020-00383-9

Biscarini F, Cozzi P, Gaspa G and Marras G. (2018) CRAN (The Comprehensive R Archive Network) https://cran.r-project.org/web/packages/detectRUNS/index.html

Dixit, S.P., Singh, S., Ganguly, I., Bhatia, A.K., Sharma, A., *et al.* (2020) Frontiers in genetics 11:92. https://doi.org/10.3389/fgene.2020.00092

Kasprzyk, A. (2011). Database. https://doi.org/10.1093/database/bar049

Ku C.S., Naidoo N., Teo S.M. and Pawitan Y. (2011) Hum. Genet. 129: 1. https://doi.org/10.1007/s00439-010-0920-6

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A., *et al.* (2007) The American journal of human genetics 81(3):559-575. https://doi.org/10.1086/519795

Soma, P., Kooverjee, B.B., Swart, H., & Yika, B.A. (2021) Proc. of the 13th Annual Agriculture Research Symposium, Gauteng, South Africa.

Van Marle-Köster E, Visser C, Sealy J, Frantz L. (2021) Sustainability 13(8):4388. https://doi.org/10.3390/su13084388

Zhang H., Li J., Zhao F., Ren H., Xu L., *et al.* (2013) PLoS ONE 8.e66569. https://doi.org/10.1371/journal.pone.0066569

224. Estimation of old and new inbreeding and their effects on growth traits using pedigree and genomic information

P. Sumreddee[1], A. Ling[2], S. Toghiani[3], E.H. Hay[3], E. Hartono, S.E. Aggrey[4] and R. Rekaya[2*]

[1]The Bureau of Biotechnology in Livestock Production, Department of Livestock Development, Pratumtani, 12000, Thailand; [2]Department of Animal and Dairy Science, The University of Georgia, Athens, GA 30602, USA; [3]USDA Agricultural Research Service, Miles City, MT 59301, USA; [4]Department of Poultry Science, The University of Georgia, Athens, GA 30602, USA; rrekaya@uga.edu

Abstract

Inbreeding caused by the mating of animals related through a recent ancestor is expected to have more pronounced effects on traits than old inbreeding. Estimating these effects requires a clear definition of new (recent) and old (ancient) inbreeding. Current methods to classify inbreeding are largely based on heuristic criteria. The subjectivity of the thresholds used to distinguish between ancient and recent inbreeding has led to inconsistency in results. To remedy this situation, we developed a new method to classify inbreeding based on a search algorithm driven by the hypothesis that new inbreeding tends to have a more pronounced effect on traits. Four growth traits were used to quantify the effects of old and new inbreeding in a Hereford herd. Thresholds to classify inbreeding varied across traits and sources of information. Using pedigree information, inbreeding generated in the last 10 to 11 generations was considered recent. When genomic information was used, thresholds ranged between 4 to 7 generations. In general, the proposed method seemed to have similar performance compared to existing approaches using ROH segments and to be slightly better using pedigree information.

Introduction

Although genealogy information has been traditionally used to estimate individual level of inbreeding (F), the pedigree-based measure of inbreeding (F_{ped}) is highly influenced by the quality of the pedigree and its completeness (Cassell *et al.*, 2003). The availability of high-density single nucleotide polymorphisms (SNPs) provides an alternative to quantify the genome-wide homozygosity through runs of homozygosity (ROH) segments. Several studies showed that inbreeding coefficients calculated based on ROH (F_{ROH}) are more accurate, and they are more powerful in assessing the effects of inbreeding (Keller *et al.* 2011). The distribution of ROH segment length could be a valuable resource to distinguish between new and old inbreeding. Long ROH segments are likely produced by recent common ancestors (recent inbreeding), whereas shorter segments are generated many generations in the past (older inbreeding). Knowing the distribution of ROH segments will provide a useful tool to understand the purging of deleterious alleles and to assess inbreeding depression. Several approaches based on tracing the pedigree back to a pre-specified number of generations (threshold) to identify recent inbreeding were proposed (Doekes *et al.* 2019). Unfortunately, the threshold is arbitrarily set, and it is likely to be population or even breed dependent. Additionally, multiple studies have recently attempted to discriminate between recent and ancient inbreeding using genomic autozygous segments. Several approaches have been proposed to categorize ROH segment length into different age-classes based on arbitrary thresholds. These approaches to discriminate between recent and old inbreeding have led to inconsistent results. It was hypothesized that recent inbreeding has a greater impact on traits compared to inbreeding due to distant common ancestors. Ancient inbreeding allows selection to have enough time to purge deleterious recessive mutations. The ability to identify inbreeding classes associated with a greater impact on phenotypes could be useful not only for quantifying its impact on phenotypes, but also for better herd management. The objectives of this study are to: (1) develop a new method to distinguish between recent and old inbreeding using pedigree and ROH information; and (2) compare the performance of the proposed approach with existing methods.

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_224, © P. Sumreddee *et al.* 2022

Materials & methods

Data used in this study was collected from purebred Line 1 Hereford cattle herd maintained at USDA-ARS, Fort Keogh Livestock and Range Research Laboratory, Miles City, MT. Although the population was historically selected for increased yearling weight, since 2011, the selection was primarily focused on improving calving ease. A complete pedigree comprising 10,478 animals was used. Pedigree analyses of the Line 1 population showed relatively complete, deep genealogical records with an average of 17.15 equivalent complete generations for all animals. A subset of 797 animals born between 1953 and 2016 was genotyped with different density panels (3k to 50k markers). Imputed genotypes for 30,220 SNPs on 29 autosomal chromosomes for 785 animals, generated as described in Sumreddee *et al.* (2019), were used.

Grid search approach to discriminate between old and recent inbreeding. We developed a grid search algorithm to determine the appropriate threshold for discriminating between new and old inbreeding based on the basic assumption that recent autozygosity has more harmful effects than old inbreeding (Sumreddee *et al.*, 2019). Thus, the threshold separating old and recent inbreeding can be identified by comparing their effects on traits. By varying the threshold to cluster ROH into short and long segment classes, the optimal cut-off-point (threshold) can be inferred. Standardized coefficients for new and old inbreeding based on varying thresholds (ROH length or number of generations) were included as covariates in the model for analyzing inbreeding depression. The magnitudes of the estimated inbreeding effects associated with partial inbreeding were then compared, and a threshold for recent inbreeding was determined. The threshold was defined as the minimum number of ancestral generations (t) or the longest length m (in Mb) used to classify ROH segments that resulted in a larger negative effect of recent inbreeding compared to its old counterpart. The resulting new and old inbreeding are denoted hereafter as $F_{new\,t}$ and $F_{old\,t}$ using pedigree information and F_{long_m} and F_{short_m} using ROH segments. The proposed method was compared to several existing approaches used to partition inbreeding into new and old classes.

Results

Figure 1 presents the correlations between different estimates of new and old inbreeding. In general, F_{ped} has moderate to high correlation (r=0.63 to 0.91) with old age-class inbreeding (e.g. $F_{old\,t}$, $F_{anc\,Kal}$). However, it was weakly correlated (r=0.24 to 0.41) with recent inbreeding (e.g. $F_{new\,t}$, $F_{new\,Kal}$). Different patterns were observed between F_{ROH} and partial inbreeding coefficients derived based on different definitions of ROH segment length classifications. F_{ROH} was in general highly correlated with both short and long ROH-based inbreeding coefficients (r=0.69 to 0.98). However, high correlations were observed only between F_{ROH} and old inbreeding derived based on pedigree (r=0.47 to 0.62). Correlation between F_{ROH} and pedigree based new inbreeding ranged between 0.15 to 0.24. The higher correlations of both F_{ped} and F_{ROH} with pedigree based old inbreeding seem to suggest that the majority of autozygosity in the used pedigree is ancient. Moderate negative correlations (r=-0.45 to -0.55) were observed between pairs of new and old inbreeding based on pedigree using our proposed thresholds (10 and 11 generations), while almost no linear relationships were observed between pairs of new and old inbreeding using a 5 generations literature threshold (r=-0.18) and between Kalinowski's new and ancestral inbreeding (r=0.09).

Inbreeding depression based on new and old pedigree-based inbreeding. The effects of new and old inbreeding based on pedigree information were present in Table 1. Using the proposed method, the threshold for discriminating between new and old inbreeding was 10 (for BW, WW and YW) and 11 (for ADG) generations. An additional 1% of new inbreeding resulted in a loss of 1.78 kg and 8 gram/day in YW and ADG, respectively. Old inbreeding also showed a significant negative effect of 5.5 gram/day on ADG but was not significant for YW. All growth traits consistently showed a larger negative effect of new inbreeding compared to its old counterpart. The proposed method yielded consistent results with those using the ancestral inbreeding approach ($F_{new\,Kal}$ and F_{anc_Kal}) based on Kalinowski *et al.* (2000) in terms

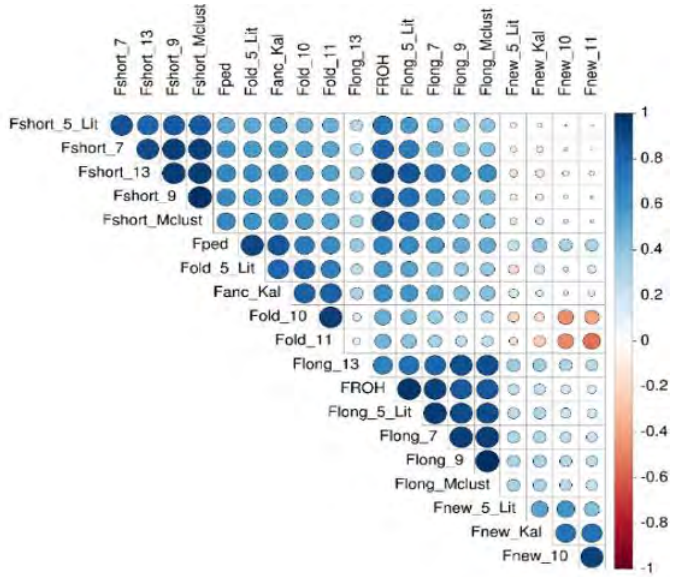


Figure 1. Correlations between different estimates of inbreeding coefficients with hierarchical clustering order. Blue and red circles represent positive and negative correlations, respectively. Color intensity and the size of the circle are proportional to the correlation coefficients. F_{ped} = pedigree-based inbreeding from the full pedigree; F_{old_t}, F_{new_t} = old and new inbreeding based on tracing the pedigree back to t ancestral generations; $F_{old\ 5\ Lit}$ and $F_{new\ 5\ Lit}$ = old and new inbreeding based on 5 generations for recent generations according to the literature; F_{anc_Kal}, $F_{new\ Kal}$ = ancestral and new inbreeding based on Kalinoski *et al.* (2000); F_{ROH} = genome-wide ROH-based inbreeding; F_{short_Mclust} and F_{long_Mclust} = inbreeding due to short and long ROH based on model-based clustering method; $F_{short_5_Lit}$ and $F_{long_5_Lit}$ = inbreeding due to short and long ROH based on 5 Mb as defined from the literature; F_{short_m} and F_{long_m} = inbreeding due to short and long ROH determined using proposed method with *m* Mb threshold.

Table 1. Estimates of the regression coefficients (SE)[1] for genome-wide and partial inbreeding coefficients based on pedigree using the proposed and existing methods.

Measure of inbreeding[2]		BW[3]	WW	YW	ADG
Genome-wide	F_{ped}	-0.018 (0.065)	-0.137 (0.312)	-1.170 (0.518)	-0.0065 (0.0016)
	F_{ROH}	-0.043 (0.040)	-0.168 (0.211)	-0.890 (0.364)	-0.0050 (0.0013)
Existing	$F_{old\ 5\ Lit}$	-0.017 (0.075)	-0.233 (0.361)	-1.436 (0.592)	-0.0070 (0.0017)
	$F_{new\ 5\ Lit}$	-0.020 (0.105)	0.106 (0.553)	-0.405 (0.972)	-0.0043 (0.0038)
	F_{anc_Kal}	-0.038 (0.112)	0.468 (0.550)	-0.265 (0.973)	-0.0059 (0.0032)
	$F_{new\ Kal}$	-0.152 (0.401)	-3.701 (2.059)	-7.282 (3.632)	-0.0276 (0.0129)
Proposed	$F_{old\ 10}$	0.0007 (0.079)	0.257 (0.383)	-0.693 (0.640)	-
	$F_{new\ 10}$	-0.038 (0.081)	-0.637 (0.421)	-1.779 (0.707)	-
	$F_{old\ 11}$	-	-	-	-0.0055 (0.0018)
	$F_{new\ 11}$	-	-	-	-0.0080 (0.0020)

[1] Significant effects are indicated by a bold font. The estimates of inbreeding effects from the proposed method are shown only for specific traits due to a trait-specific threshold to define new inbreeding.

[2] $F_{old\ 5\ Lit}$ and $F_{new\ 5\ Lit}$ = old and new inbreeding based on 5 generations for recent generations according to the literature; F_{anc_Kal} and $F_{new\ Kal}$ = ancestral and new inbreeding based on Kalinowski *et al.* (2000); $F_{old\ t}$ and $F_{new\ t}$ = old and new inbreeding based on tracing the pedigree back to t ancestral generations determined using the proposed method.

[3] BW = birth weight (kg); WW = weaning weight (kg); YW = yearling weight (kg); ADG = average daily gain (kg/d)

of inbreeding depression, especially for YW and ADG. Kalinowski's new inbreeding (F_{anc_new}) was more harmful than ancestral inbreeding (F_{anc_Kal}) for most of the growth traits; however, the significant effects were found only for YW and ADG. Using the literature-defined threshold of 5 generations to separate new and old inbreeding only the latter ($F_{old\ 5\ Lit}$) showed a significant unfavorable effect on both YW and ADG. Although a greater harmful impact would be expected for a very recent inbreeding (5 generations), this small threshold value favored older inbreeding to capture the majority of the genome-wide inbreeding. In fact, about 93% of the total inbreeding (F_{ped}) was captured by old inbreeding when the threshold was fixed at 5 generations. Consequently, magnitudes of effects of $F_{old\ 5\ Lit}$ and F_{ped} on YW and ADG were similar. In this respect, although the old inbreeding is likely to have a small deleterious effect, the accumulation of a relatively larger number of old genetic variants could dominate the overall inbreeding depression. Likewise, a smaller effect of new inbreeding obtained using 5 recent generations ($F_{new\ 5\ Lit}$) could be due to its minimal contribution (about 7%) to the total inbreeding. These results are concerning regarding the potential bias in the estimation of inbreeding depression using an arbitrary threshold.

Inbreeding depression based on recent and ancient ROH-based inbreeding. Partitioning of ROH segments into long and short classes offers the opportunity to assess the unfavorable effects of recent (F_{long}) and ancient (F_{short}) inbreeding. Using the proposed method, both inbreeding coefficients had negative effects on growth performances; however, the effects are statistically significant only for YW and ADG. Inbreeding based on ROH segments longer than 7 Mb (F_{long_7}) estimated using the proposed method resulted in a significant decrease in ADG (4.4 gram/day per 1% increase in inbreeding). Overall, the differences in the effects of new and old inbreeding based on the length of ROH segments were less evident compared to those observed using pedigree information. The proposed method seemed to have similar performance compared to existing approaches using ROH segments and to be slightly better using pedigree information.

References

Cassell B.G., Adamec V. and Pearson R.E. (2003) J. Dairy Sci. 86(9): 2977-2983. https://doi.org/10.3168/jds.S0022-0302(03)73895-8

Doekes H.P., Veerkamp R.F., Bijma P., De Jong G., Hiemstra S.J. *et al.* (2019) Genetics Selection Evolution 51(1): 54. https://doi.org/10.1186/s12711-019-0497-z

Kalinowski S.T., Hedrick P.W. and Miller P.S. (2000) Conservation Biology 14(5): 1375-1384. https://doi.org/10.1046/j.1523-1739.2000.98209.x

Keller M.C., Visscher P.M. and Goddard M.E. (2011) Genetics 189(1): 237-249. https://doi.org/10.1534/genetics.111.130922

Sumreddee P., Toghiani S., Hay E.H., Roberts A., Agrrey S.E. *et al.* (2018) J. Anim. Sci. 97(1): 1-18. https://doi.org/10.1093/jas/sky385

225. Genome wide characterisation of runs of homozygosity on the main indigenous goat breed of Botswana

P.I. Monau[1*], K. Raphaka[2] and S.J. Nsoso[1]

[1]Botswana University of Agriculture and Natural Resources, Faculty of Animal and Veterinary Sciences, Department of Animal Science, Private Bag 0027, Gaborone, Botswana; [2]National Agricultural Research & Development Institute, Private Bag 00177, Gaborone, Botswana; pmonau@buan.ac.bw

Abstract

Runs of homozygosity (ROH) provides knowledge on the demographic history and management of livestock. The aim of this study was to characterise the distribution of runs of homozygosity (ROH) in Tswana goat genome. The Illumina Goat SNP50 Bead Chip was used to genotype 192 Tswana goats. Quality control and detection of ROH were analysed using Plink software. A total of 2,486 ROH were detected in Tswana goat population and the average number of ROH per animal was 13.08±8.71. The mean length was 7.18±7.89 and ranged from 3.02 Mb to 28.06 Mb. The majority of ROH segments for Tswana goats were shorter and accounted for approximately 77% in contrast to 10% of segments longer than 16 Mb. The highest chromosomal inbreeding levels were found on chromosome 6 with about 7.8%. The findings form a base for future goat genomics research and goat breeding programs in Botswana.

Introduction

Tswana goat is the predominant ancient genetic resource valuable for sustainable production and nutrition of poor-resource farmers in Botswana. It constitutes about 70% of the 1.3 million goat's population kept extensively under traditional management systems (Statistics Botswana, 2019). The traditional sector consists of approximately 96% of goat production (Statistics Botswana, 2019). This breed has been exposed to natural selection over time, enforced by environmental changes, breeding conditions and different husbandry practices, resulting in a wide range of adaptability such as low feed intake, heat, and disease resistance. Selection of certain characters over time often leads to reduced genetic diversity and autozygosity. Autozygosity generally occurs when individuals possess identical haplotypes from common ancestors resulting in long incessant homozygous segment known as runs of homozygosity (ROH) (Nosrati *et al.*, 2021). The length and frequency of ROH play a major role in shaping the genomic change of livestock (Dixit *et al.*, 2020). The short ROH may arise because of ancient consanguinity from genetic drift and founder effect whilst longer ROH may reveal recent common ancestor (McQuillan *et al.*, 2008). The distribution and size of ROH provide insights on individual autozygosity, evolution history and demography, diseases, and traits architecture, ultimately assisting on guiding breeding strategies and sustainable genetic improvement.

Molecular genetics has facilitated in scanning the genome for ROH in different livestock species using high-density markers. The extent of ROH across different goat genome is however, limited particularly for indigenous goats in the tropics including the indigenous Tswana goat. It is necessary to assess autozygosity of indigenous goats to implement relevant breed improvement and design breed conservation programs to cope with current and future environmental changes. The aim of this study was therefore, to characterise genome wide patterns of runs of homozygosity of indigenous Tswana goats of Botswana.

Materials & methods

Animals, genotyping and quality control. Animal sampling was conducted in conformity with the guidelines of Botswana University of Agriculture and Natural Resources ethical committee (BUANEC). A total of 96 goat hair samples were randomly collected from traditional farms in all different agro-ecological regions of Botswana and the other 96 from Agricultural Research Stations. The hair samples were

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_225, © P.I. Monau *et al.* 2022

despatched to Animal Genetic Laboratory in France, Labogena DNA platform (Domaine de Vilvert, CS 80009, 78353 Jouy en Josas cedex), for DNA extraction and genotyping with the Illumina Goat SNP50 Bead Chip. All quality control procedures were carried out using PLINK v1.07 software (Purcell *et al.*, 2007). Only autosomal SNPs were included in the analysis and markers that did not map to any chromosome were discarded. SNPs with minor allele frequency (MAF) greater than 0.05, call rate and genotyping rate greater than 95%, Hardy-Weinberg equilibrium $>1\times10^{-4}$ and individuals with missing genotyping rate of more than 15% were excluded from the analysis. The ultimate dataset retained for down streaming analyses included 190 animals and 47,040 markers.

Detection of runs of homozygosity. Runs of homozygosity (ROH) were detected using PLINK v1.07 software (Purcell *et al.*, 2007). To define the ROH, a sliding window of 50 SNPs was used, and the following parameters were used: the minimal number of SNPs in ROH was set to 20; the maximal gap between the adjacent SNPs was set to 2 Mb; the minimum SNP density per ROH was set to 1 SNP every 100 kb; one heterozygote genotype allowed, and 5 missing genotypes allowed. To characterize the ROH distribution, the following parameters were estimated; the average ROH per individual, the number of ROH, the mean total length of ROH, and the genomic inbreeding coefficient (F_{ROH}). The genomic inbreeding coefficient based on ROHs ($F_{ROH)}$ was computed following the method proposed by McQuillan *et al.* (2008). The detected ROHs were specified based on length to four categories: 2-4 Mb, 4-8 Mb, 8-16 Mb, and >16 Mb recognised as $ROH_{2\text{-}4}$ Mb, $ROH_{4\text{-}8}$ Mb, $ROH_{8\text{-}16}$ Mb, and $ROH_{>16}$ Mb, respectively. All ROHs per individual were summed and averaged to calculate the mean sum of ROH length.

Results

A total of 2,486 ROH were detected across the autosome of 190 Tswana goats and the average number of ROH per animal was 13.08±8.71, with values ranging from 2 to 59. The descriptive statistics for runs of homozygosity of indigenous Tswana goats are summarised in Table 1. The mean length was 7.18±7.89 and ranged from 3.02 Mb to 28.06 Mb. The longest segment of 66 Mb constituting of 1,258 SNPs was found on chromosome 2 (CHI2). The majority of ROH segments for Tswana goats were shorter (2-8 Mb) and covered relatively small part of the genome. These segments accounted for approximately 77% in length in contrast to 10% of ROH segments longer than 16 Mb that covered a higher proportion of the genome (Table 1). The genomic inbreeding coefficient (F_{ROH}) was 0.005.

The relationship among animals regarding the number of ROH and the length of the genome covered by them is portrayed in Figure 1. Most of the individuals clustered close to the beginning of coordinates because each one carried 2 to 20 ROH with a total length less than 300 Mb. The highly inbred animal displayed a ROH genome coverage of 315.31 Mb of the total autosomal genome with 51 number of ROH found within a range of 2-8 Mb and a mean length of 5.34 whilst the least inbred animal exhibited a ROH genome coverage of 8.05 Mb and a mean length of 4.0 with all ROHs found within a range of 2- 8 Mb.

Table 1. Descriptive statistics for runs of homozygosity for indigenous Tswana goat.

Class length (Mb)	No of ROH[1]	Percentage	Mean length	Standard deviation	Genome coverage
$ROH_{2\text{-}4}$	1,147	46	3.02	0.48	0.14
$ROH_{4\text{-}8}$	763	30	5.54	1.15	1.69
$ROH_{8\text{-}16}$	353	14	11.0	2.23	1.55
$ROH_{>16}$	223	10	28.06	11.11	2.51

[1] Number of runs of homozygosity.

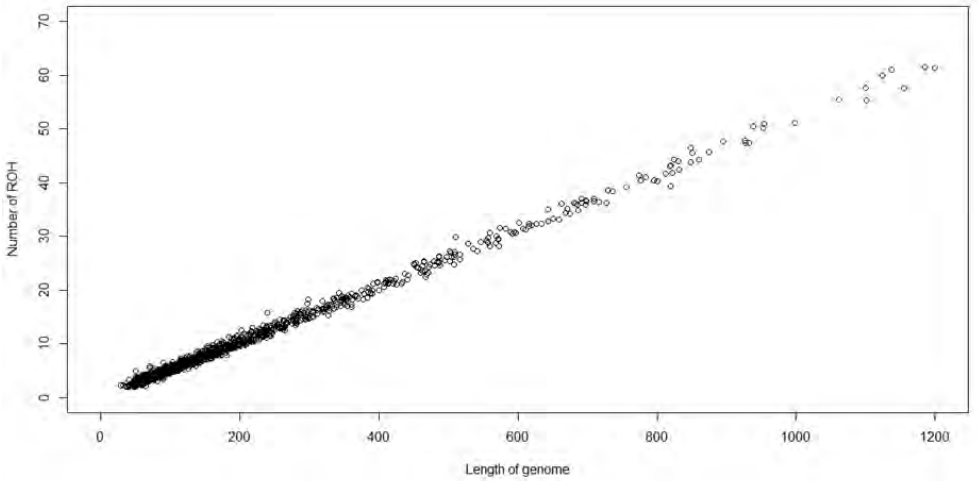


Figure 1. Number of runs of homozygosity per individual Tswana goat and the length of the genome covered by ROH (Mb).

The frequency of ROH per chromosomes are presented in Figure 2. ROH segments were identified on all 29 autosomes and the percentages of ROH declined as the number of chromosomes increases. Major fraction of ROH was observed on chromosome 6 with about 7.8%.

Discussion

Genomic inbreeding based on runs of homozygosity (ROH) provides an important information due to non-existent pedigree information of goats kept under traditional management system. The abundance of short ROH measuring up to 8 Mb segments in Tswana goat population demonstrate that the breed has been formed by small founding populations. The substantial low contribution of short ROH to the genome could indicate differences in selection events in ancestral populations. The long ROH that covered most of the genome though in small percentiles reflect low possibility of recombination and recent inbreeding of

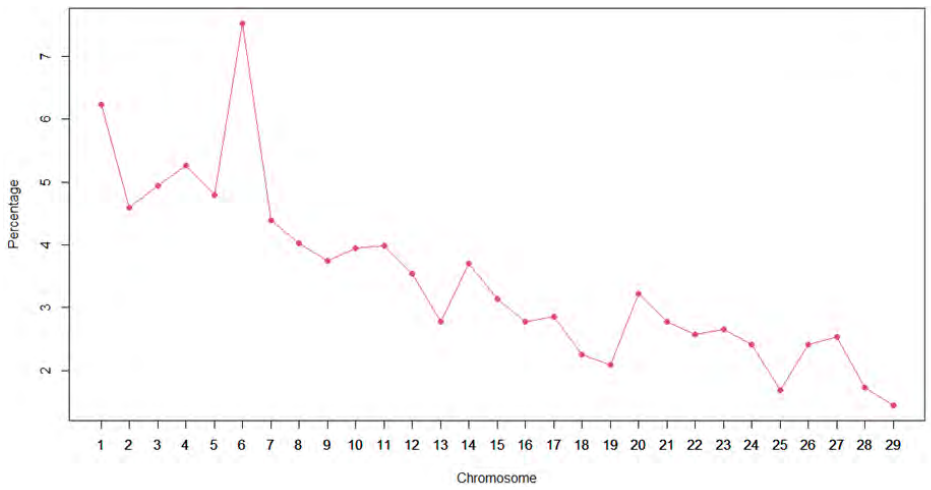


Figure 2. Percentage of SNPs in a ROH for each chromosome of Tswana goats.

3-5 generations ago (Onzima *et al.*, 2017). Recent inbreeding is more destructive than ancient inbreeding because selection lessens the frequency of harmful alleles over generations (Zhao *et al.*, 2021). The pattern of ROH coverage observed in this study is in line with other studies of goats (Brito *et al.*, 2017; Onzima *et al.*, 2018; Manstrangelo *et al.*, 2021). The low genomic inbreeding coefficients provides the opportunity for better management and designing appropriate conservation programs to preserve genetic diversity to cope with current and future predicted effects of climate change. The findings of the study will advance our understanding on the adaptation and functional traits of Tswana goats.

Conclusions

Runs of homozygosity are randomly distributed across Tswana goat genome. The population is not highly affected by recent inbreeding and the ancient haplotypes may have a major positive impact on the functional traits of Tswana goats. The findings provide insight for future goat genomics research, goat breeding program, and sustainable utilization of the breed.

References

Brito, L.F., Kijas, J.W., Ventura, R.V., Sargolzaei, M., Porto-Neto, L.R., Cánovas, A. *et al.* (2017) BMC Genom. 18(1): 1-20. https://doi.org/10.1186/s12864-017-3610-0

Dixit, S. P., Singh, S., Ganguly, I., Bhatia, A. K., Sharma, A. *et al.* (2020). Front genet. 11, 92. https://doi.org/10.3389/fgene.2020.00092

Mastrangelo, S., Di Gerlando, R., Sardina, M.T., Sutera, A.M., Moscarelli, A. *et al.* (2021) Animals, 11(6):1510. https://doi.org/10.3390/ani11061510

McQuillan, R., Leutenegger, A.L., Abdel-Rahman, R., Franklin, C.S., Pericic, M., Barac-Lauc, L. *et al.* (2008) Am. J. Hum. Genet. 83(3):359-372. https://doi.org/10.1016/j.ajhg.2008.08.007

Nosrati, M., Nanaei, H.A., Javanmard, A. and Esmailizadeh, A. (2021) Genomics, 113(3):1407-1415. https://doi.org/10.1016/j.ygeno.2021.03.005

Onzima, R.B., Upadhyay, M.R., Doekes, H.P., Brito, L., Bosse, M. *et al.* (2018) Front genet. 9:318. https://doi.org/10.3389/fgene.2018.00318

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A.et al. (2007) Am. J. Hum. Genet. 81(3):559-575. https://dx.doi.org/10.1086%2F519795

Statistics Botswana, 2019. Annual agricultural survey report of 2013. Gaborone, Botswana. Available at: www.statsbots.org.bw

Zhao, G., Liu, Y., Niu, Q., Zheng, X., Zhang, T. *et al.* (2021) BMC Genom. 22(1):1-12. https://doi.org/10.1186/s12864-021-07992-6

226. Candidate genes for disease, reproduction and meat quality traits in Portuguese native breeds

M.V.G.B. Silva[1*], L.L. Verardo[2], M.A. Machado[1], J.C.C. Panetto[1], I. Carolino[3] and N. Carolino[3]

[1]Brazilian Agricultural Research Corporation (EMBRAPA), Rua Eugênio do Nascimento, 610, 36038-330, Juiz de Fora – MG, Brazil; [2]Universidade Federal dos Vales do Jequitinhonha e Mucuri, Campus JK, 39100-000, Diamantina, Brazil; [3]Instituto Nacional de Investigação Agrária e Veterinária, Fonte Boa, 2005-048, Vale de Santarém, Portugal; marcos.vb.silva@embrapa.br

Abstract

Native Portuguese cattle breeds are a 'biodiversity deposit' resulting from genetic and environmental effects accumulated over the years. We assessed two cattle breeds representative of Portuguese native cattle (Barrosã – BAR and Brava de Lide – BRA) aiming the evaluation of ROH across populations, identification of ROH islands and functional analysis of the identified genes. The pattern of ROH differed across breeds mainly from short to median range. Based on ROH islands analysis, three regions were observed to be shared by more than 35% of the individuals in the two breeds. Besides, gene networks highlighted biological associations based on genes found on ROH islands with biological processes related with reproductive traits (*RBM19*) and immune systems (*DTX1*) in BAR breed, and immune systems and adipogeneses processes (*ZBTB16*) in BRA breed.

Introduction

Native Portuguese cattle breeds are a 'biodiversity deposit' resulting from genetic and environmental effects accumulated over the years, namely, geographical, edapho-climatic, land, historical, cultural, political-economic, among others. In the middle of the 20th century, with the implementation of the first Herd Books and the reorientation of production, mainly for beef, the current native bovine populations became properly organized. Since the 1990s, according to the size of the herd and the priorities defined by breeders for each breed, they have been included in conservation or genetic improvement programs. Except for the Brava de Lide breed, whose selection is oriented towards behavioural traits, the other breeds have been selected essentially for maternal and reproductive traits.

History of the population, breeding system, and pattern of geographic subdivision may be reflected in runs of homozygosity (ROH) patterns throughout the genome. ROH are continuous homozygous segments of DNA (Gibson *et al.*, 2006), and its calculation is essential to identify islands. The islands of homozygosity shared among individuals, may be helpful in identify possible results of selection events (Zhang *et al.*, 2015). Moreover, genes identified in those regions can be functionally analysed through gene networks (Verardo *et al.*, 2021) aiming to a better understand of their biological role in each population. Thus, we assessed two cattle breeds representative of Portuguese native cattle, with different breeding goals (Barrosã – BAR and Brava de Lide – BRA) aiming the evaluation of ROH across populations, identification of ROH islands and functional analysis of the identified genes.

Materials & methods

The cattle breeds of Portugal used in this study are representative of the two trunks of cattle that were formed in the westernmost region of the Iberian Peninsula and classified by Miranda do Vale (1949), according to their ethnic characteristics and geographical distribution: Mauritanian or *Bos taurus mauritanicus* (Barrosã – BAR) and Iberian or *Bos taurus ibericus* (Brava de Lide – BRA). Animals were genotyped using the commercial genotyping array of Affymetrix (60k). SNP genotypes from sexual chromosomes were excluded from the analysis. Also, genotypes were excluded when the call rate <0.95 and minor allele frequency <0.01.

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_226, © M.V.G.B. Silva *et al.* 2022

In addition, samples showing overall call rates <0.90 were excluded from further analysis. A total of 96 and 94 animals were analysed after quality control with 34,997 and 33,482 SNPs in the genotypes of BAR and BRA, respectively.

The ROH approach using the software PLINK v1.07 (Purcell *et al.*, 2007) for each population with the following parameters were used: minimum window length of 120 SNPs, maximum gap size between two SNPs of 1000 kb, minimum ROH length of 1000 kb, minimum number of potential marked SNPs of 50, one heterozygote allowed per window, maximum of five missing calls per window, sliding window length of 50 SNPs and proportion of overlapping windows that must be a homozygous >0.05. To identify ROH islands throughout the genome, the most observed homozygous segments shared by individuals in each breed were considered as an indication. The '--homozyg-group' function implemented in PLINK was used to assess ROH islands shared among individuals. The GenBank annotation based on the ARS-UCD1.2 assembly of the bovine genome was used to identify genes in ROH regions.

Gene networks highlighting biological processes among the gene sets identified for each breed were generated using the ClueGO plugin for Cytoscape (Bindea *et al.*, 2009). This procedure allowed the creation of gene networks highlighting biological roles, and the comparison of gene clusters by visualizing their functional differences or similarities.

Results

The pattern of ROH differed across breeds mainly from short to median range (Figure 1A). BAR and BRA breeds showed similar percentage of long-range ROH (>31 Mb; <5%), while BAR showed higher percentages of short range ROH (<5 Mb; >50%). In addition, based on ROH islands analysis, three regions were observed to be shared by more than 35% of the individuals in the two breeds (Table 1). Gene networks highlighting biological associations were generated based on genes found on ROH islands (Figure 1B). Several biological processes were highlighted, including Regulation of T cell differentiation and positive regulation of embryonic development.

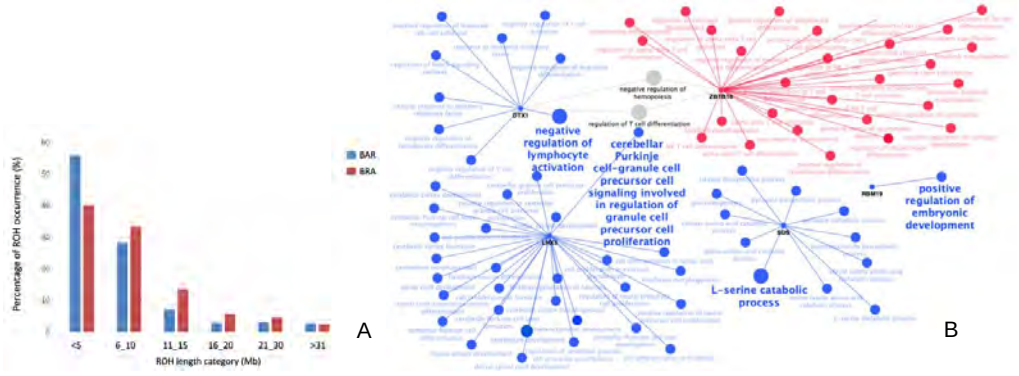


Figure 1. (A) Percentage of runs of homozygosity (ROH) per length categories in two Portuguese Iberian cattle breeds. The sum of ROH was calculated per animal within each ROH length category. BAR: Barrosã; BRA: Brava de Lide. (B) Gene networks highlighting biological associations of genes found on ROH islands in two Portuguese Iberian cattle breeds. Blue and red nodes are biological processes related with genes identified in BAR and BRA breeds, respectively. Grey nodes are biological processes shared between both breeds genes.

Table 1. Chromosomal position, length in base pairs (bp), number of markers and gene content of runs of consensus homozygosity (ROH) islands most shared (ROH freq.) identified in two Portuguese Iberian breeds.

Breed[1]	Chr.	Begin	End	Length (bp)	Nº SNP	ROH freq. (%)	Annotated genes
BAR	17	60,517,694	61,094,324	576,630	11	35.42	*PLBD2, SDS, SDSL, DTX1, LHX5, LOC104974667, LOC101907408, TRNAG-CCC, RBM19, LOC100847943*
BRA	15	24,406,959	24,891,500	484,541	7	43.62	*ZBTB16, HTR3A, HTR3B, LOC511161, LOC101907641, LOC782610, TRNAG-CCC, LOC112441707*
	15	24,166,187	24,231,362	65,175	3	43.62	*TMPRSS5*

[1] BAR: Barrosã; BRA: Brava de Lide.

Discussion

We first aimed to evaluate the run of homozygosity in two Portuguese breeds, Barrosã and Brava de Lide. A run of homozygosity is the probability that all consecutive markers on a pair of homologous chromosome segments, in the same or different individual(s), display identical alleles (Hayes *et al.*, 2003). The extent and frequency of ROH may provide information about the ancestry of an individual and its population. Moreover, inbreeding may be inferred from the presence of long ROH, with longer segments indicating recent inbreeding within a population (Kirin *et al.*, 2010). We observed that both breeds showed small and similar occurrence of longest ROH found in this study (<5%). This indicates a proper matting scheme in those breeds to prevent higher inbreeding. Besides, BAR showed higher prevalence of small ROH indicating more genetic diversity.

Moreover, both breeds showed ROH islands. It is important to note that with few sampled genotypes, it is quite likely that some ROH are artifacts. This is particularly important when pointing to specific candidate regions/genes as these might be misleading. In this study, genes and their biological processes were analysed aiming to understand their roles in each breed. In BAR, we identified an island in chromosome 17 with genes associated, for example, with positive regulation of embryonic development (*RBM19*) and negative regulation of lymphocyte activation (*DTX1*) in the biological process network. *RBM19* encoding for RNA binding motif protein 19, may be involved in regulating ribosome biogenesis and was previously reported to be essential for preimplantation during mouse embryonic development (Zhang *et al.*, 2008). *DTX1* encode for deltex E3 ubiquitin ligase 1 protein being reported to play a regulatory role in immune responses in teleost fish (Zheng *et al.*, 2021). Thus, these two genes may play a role on reproductive traits (*RBM19*) and immune systems (*DTX1*) in BAR breed.

Also, ROH islands were observed on chromosome 15 of BRA breed. In those islands, based on biological process network, we observed a gene (*ZBTB16*) related with regulation of T cell differentiation which is also shared with the previous discussed gene of BAR breed (*DTX1*). The *ZBTB16* gene encodes for the zinc finger and BTB domain containing 16 protein and have been related with innate and innate-like lymphoid lineage development in mouse (Mao *et al.*, 2017) and defense response to gram-positive bacteria in goat (Rout and Verma, 2021). Also, its overexpression is suggested to promote white adipogenesis and induce brown-like adipocyte formation in bovine white intramuscular preadipocytes (Wei *et al.*, 2018). It suggests a role of this gene in immune systems and adipogeneses processes of BRA breed.

In this study, short ROH were observed in both Portuguese Iberian breeds BAR and BRA, suggesting a successful matting scheme, also highlighted by a lower occurrence of long ROH. Genes located in ROH islands were evaluated and explored throughout their biological processes, which provided genetic insights about the breeds. Genes were associated with important traits (e.g. *RBM19* with reproductive traits and

DTX1 with immune systems in BAR breed; and *ZBTB16* with immune systems and adipogeneses processes in BRA breed).

Acknowledgements

The authors would like to thank the *Associação dos Criadores de Bovinos de Raça Barrosã* and *Associação Portuguesa de Criadores de Toiros de Lide*, and the *Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq)/Instituto Nacional de Ciência e Tecnologia – Ciência Animal (INCT-CA)* for their financial support.

References

Bindea G., Mlecnik B., Hackl H., Charoentong P., Tosolini M., *et al.* (2009). Bioinformatics, 25(8), 1091-1093. https://doi.org/10.1093/bioinformatics/btp101

Do Vale M. (1949) Gado Bissulco – Suínos, Bovinos, Arietinos, Caprinos. Livraria Sá da Costa, Lisboa, PT.

Gibson J., Newton E.M. and Collins A. (2006). Human Molecular Genetics, 15, 789–95. https://doi.org/10.1093/hmg/ddi493

Hayes B.J., Visscher P.M., McPartlan H.C. and Goddard M.E. (2003). Genome Res. 13, 635-643. http://www.genome.org/cgi/doi/10.1101/gr.387103

Kirin M., McQuillan R., Franklin C. S., Campbell H., McKeigue P. M., *et al.* (2010). PLoS One 5:e13996. https://doi.org/10.1371/journal.pone.0013996

Mao A.P., Ishizuka I.E., Kasal D.N., Mandal M., and Bendelac A. (2017). Nature communications, 8(1), 1-14. https://doi.org/10.1038/s41467-017-00882-0

Purcell S., Neale B., Todd-Brown K., Thomas L., Ferreira M.A., *et al.* (2007). The American journal of human genetics, 81(3), 559-575. https://doi.org/10.1086/519795

Rout P.K., and Verma M. (2021). Scientific reports, 11(1), 1-16. https://doi.org/10.1038/s41598-021-85094-9

Verardo L.L., e Silva F.F., Machado M.A., do Carmo Panetto J.C., Faza D.R.D.L.R., *et al.* (2021). Frontiers in genetics, 12: 702822. https://doi.org/10.3389/fgene.2021.702822

Wei S., Zhang M., Zheng Y., and Yan, P. (2018). Cellular Physiology and Biochemistry, 48(6), 2528-2538. https://doi.org/10.1159/000492697

Zhang J., Tomasini A.J., and Mayer, A.N. (2008). BMC developmental biology, 8(1), 1-14. https://doi.org/10.1186/1471-213X-8-115

Zhang Q., Guldbrandtsen B., Bosse M., Lund M.S., and Sahana, G. (2015). BMC genomics, 16(1), 1-16. https://doi.org/10.1186/s12864-015-1715-x

Zheng W., Chu Q., Yang L., Sun L., and Xu, T. (2021). PLoS pathogens, 17(3), e1009438. https://doi.org/10.1371/journal.ppat.1009438

227. Runs of homozygosity harbour genes associated with muscle growth in Pietrain but not with fertility in Large White pigs

S.E. Palma-Vera[1*], L. Derežanin[1,2], J. Fickel[2,3], H. Reyer[4], N. Reinsch[5], H. Henne[6] and J. Schön[1,7]

[1]Institute of Reproductive Biology, Research Institute for Farm Animal Biology (FBN), Wilhelm-Stahl-Allee 2, 18196 Dummerstorf, Germany; [2]Department of Evolutionary Genetics, Research Institute for Zoo and Wildlife Research (IZW), Alfred-Kowalke-Straße 17, 10315 Berlin, Germany; [3]University of Potsdam, Institute for Biochemistry and Biology, Karl-Liebknecht-Str. 24-25, 14476 Potsdam, Germany; [4]Institute of Genome Biology, Research Institute for Farm Animal Biology (FBN), Wilhelm-Stahl-Allee 2, 18196 Dummerstorf, Germany; [5]Institute of Genetics and Biometry, Research Institute for Farm Animal Biology (FBN), Wilhelm-Stahl-Allee 2, 18196 Dummerstorf, Germany; [6]German Federal Hybrid Breeding Programme (BHZP GmbH), An der Wassermühle 8, 21368 Dahlenburg, Germany; [7]Department of Reproduction Biology, Research Institute for Zoo and Wildlife Research (IZW), Alfred-Kowalke-Straße 17, 10315 Berlin, Germany; palma-vera@fbn-dummerstorf.de

Abstract

Runs of Homozygosity (ROHs) are extensively found in the genomes of farm animals. The genes contained in these regions are important for the selected traits, having reached fixation through selective breeding, but also as a result of inbreeding. Two essential goals of breeding programs are to increase reproductive fitness and muscle growth, both of which are largely determined by ROHs and the genes therein. In this study, the genomes of pigs selected primarily for reproductive (Large White (LW), n=10) or growth (Pietrain (PI), n=10) traits were sequenced, and ROHs were detected. In both breeds, approximately one third of the genome was found to be homozygous. Although breed-specific genes overlapping ROHs were not significantly enriched for biological functions related to reproduction or growth, five genes were associated to muscle biology. Furthermore, these genes were affected by non-synonymous mutations, which may result from selection.

Introduction

Though maintaining genetic diversity is a priority of any breeding program, regions of homozygosity are prevalent in the genomes of farm animals, resulting from selection and inbreeding (Maltecca *et al.*, 2020). These regions are mostly studied in the form of Runs of Homozygosity (ROHs), relying on single nucleotide variants (SNVs) for their detection (Ceballos *et al.*, 2018).

High reproductive performance and high lean body weight are two important goals in pig breeding. Under commercial conditions, it is common to cross highly prolific females with highly muscular males, in order to obtain a large number of piglets that will meet market standards. This is known as terminal crossing and it often involves Large White (LW) females and Pietrain (PI) males. Therefore, these two breeds offer a convenient model to explore fertility and growth's genetic background.

This study aimed to identify regions of homozygosity in ROHs and detect genes associated with reproduction and growth by sequencing the genomes of pigs highly selected for maternal (LW) and paternal traits (PI). Results from these analyses provide insights into the genetic diversity of these two important commercial pig breeds.

Materials & methods

Sample collection and whole genome sequencing. Genomic DNA was extracted from blood (10 animals per breed) using QIAamp DNA Mini Kit (Qiagen, Hilden, Germany). Blood samples were collected

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_227, © S.E. Palma-Vera *et al.* 2022

in accordance to the health and hygiene monitoring at the German Federal Hybrid Breeding Programme (BHZP) (for more details see (Dauben *et al.*, 2021)). Sequencing libraries were prepared with the Illumina® TruSeq® Nano DNA Library Prep kit. Sequencing was done with Illumina® NovaSeq6000 generating paired end reads of 151 bp length.

Read processing and SNV calling. Adapter removal and quality trimming were performed with FASTP (Chen *et al.*, 2018). Read quality was evaluated with FastQC (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/). Reads were aligned to the pig genome build Sscrofa11.1 using the Burrow-Wheeler Aligner software in MEM mode (Li and Durbin, 2010).

Alignment processing and SNV calling followed GATK best practices (4.0.6.0) (Depristo *et al.*, 2011; McKenna *et al.*, 2010; Poplin *et al.*, 2017; Van der Auwera *et al.*, 2013). Only bi-allelic SNVs were retained. Sites were filtered by removing 5-10% of the poorest-scoring variants according to the following annotations: QualByDepth, FisherStrand, StrandOddsRatio, RMSMappingQuality, MappingQualityRankSumTest and ReadPosRankSumTest. SNVs were genotyped as missing if the depth of coverage was either too low (<4 reads), too high (3 standard deviations higher than the sample mean depth), or if the genotype quality was too low (<20 in Phred scale). Variants missing in at least one sample were removed. Annotations were done using SnpEff v4.3t (Cingolani *et al.*, 2012).

Detection of runs of homozygosity. ROHs were estimated using the ROH extension (Narasimhan *et al.*, 2016) in SAMtools/BCFtools v1.5 (Li *et al.*, 2009). For this, allele frequencies at each SNV site and a constant recombination rate of 0.7 cM/Mb were provided. Genes overlapping ROHs were detected, transformed into human gene identifiers and characterized based on Gene Ontology Biological Process (GOBP) annotations using enrichr (Xie *et al.*, 2021).

Data handling and visualization. Genomic intervals (genes, SNVs and ROHs) were overlapped using GenomicRanges (Lawrence *et al.*, 2013). Data processing and visualizations were done using R (R Core Team, 2020) and the tidyverse package (Wickham *et al.*, 2019).

Results

Average genome coverage was ~50× and a total ~16M SNVs were detected. Levels of homozygosity were similar in both breeds, with approximately 35% of the genome covered by ROHs (Figure 1). Genomic homozygosity was largely (~23%) composed by medium size ROHs between 100 Kb and 5 Mb. Small ROHs less than 100 Kb spanned only 5% of the genome in both breeds. In Pietrain, a slightly higher genomic fraction was covered by large ROHs (>5 Mb).

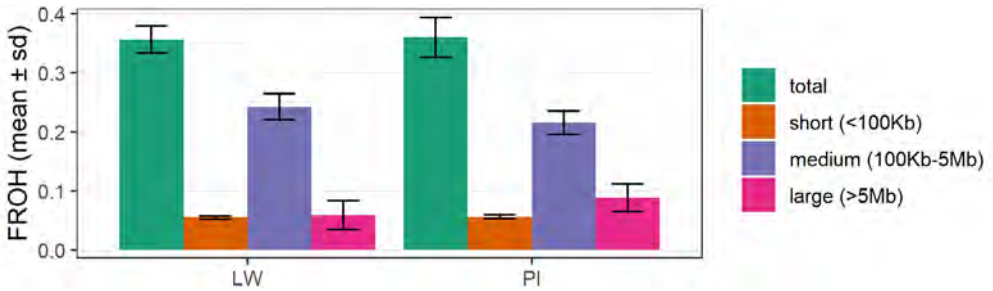


Figure 1. Mean genomic fraction covered by ROHs (FROH). For each breed the total FROH is shown and next decomposed into subcategories according to ROH size.

For each animal, genes overlapping ROHs were detected and breed-specific gene lists were created: LW-private with 216 genes and PI-private with 393 genes. Reproductive functions were not represented by LW-private genes, but significant enrichments (FDR<0.05) were found for 'ephrin receptor signaling pathway' (GO:0048013), 'positive regulation of cytoskeleton organization' (GO:0051495) and 'axon guidance' (GO:0007411). PI-private genes did not produce significant enrichments. However, for 14 genes, GOBPs included functions related to muscle growth. Five of these genes were affected by at least one missense or stop-gained homozygous SNV (Table 1).

Discussion

Although ROHs have been studied in LW and PI, this was so far done via SNP-arrays (Gorssen *et al.*, 2020; Shi *et al.*, 2020; Zhan *et al.*, 2020) or low-mid coverage WGS (Bosse *et al.*, 2012; Nosková *et al.*, 2021). In this study, samples were sequenced at an average coverage of ~50×, which allows to obtain a dense SNV set and to genotype heterozygous variants with high accuracy. These two attributes are critical to get a detailed view of the ROH landscape.

Private genes overlapping ROHs in LW were not involved in reproductive processes. Interestingly, the significantly enriched term 'ephrin receptor signalling pathway' (GO:0048013) has been reported to mediate FSH signalling in granulosa cells (Buensuceso and Deroo, 2013), which could be important for ovarian follicle development and ovulation. On the other hand, muscle function associated genes were found in the PI-private gene list, carrying mutations predicted to disrupt translation and altering the amino acid protein sequence. Since neither of these mutations have been reported within Pietrain ROHs, it is possible that they were missed by previous studies due to a lower SNV density than the present study.

This study provides new insights into the footprints of selection and inbreeding in the genomes of commercially important pig breeds. The genes spanned by ROHs may have important functions for the selected traits, but they may also represent the consequences of inbreeding.

Funding

This study was funded by the Leibniz Collaborative Excellence programme (K52/2017, SOS-FERT).

Table 1. Genes involved in muscle function affected by missense or stop-gained mutations.

Gene Symbol	No of SNVs	Chromosome:Position (Effect)
TNNI2	2	2:1252872 (stop_gained&splice_region_variant)
		2:1252872 (missense_variant)
TNNT3	3	2:1316619 (stop_gained)
		2:1316689 (missense_variant)
		2:1316747 (missense_variant)
AKAP13	1	7:88217238 (missense_variant)
ANK2	4	8:109320026 (missense_variant)
		8:109320032 (missense_variant)
		8:109371239 (missense_variant)
		8:109436996 (missense_variant)
ANKRD2	2	14:108948808 (missense_variant)
		14:108948871 (missense_variant)

References

Bosse M., Megens H.-J.J., Madsen O., Paudel Y., Frantz L.A.F., *et al.* (2012) PLoS Genet. 8(11):e1003100. https://doi.org/10.1371/journal.pgen.1003100

Buensuceso A. V, and Deroo B.J. (2013) Biol. Reprod. 88(1):1-12,25. https://doi.org/10.1095/biolreprod.112.100123

Ceballos F.C., Joshi P.K., Clark D.W., Ramsay M., and Wilson J.F. (2018) Nat. Rev. Genet. 19(4):220–234. https://doi.org/10.1038/nrg.2017.109

Chen S., Zhou Y., Chen Y., and Gu J. (2018) Bioinformatics 34(17):i884–i890. https://doi.org/10.1093/bioinformatics/bty560

Cingolani P., Platts A., le Wang L., Coon M., Nguyen T., *et al.* (2012) Fly 6(2):80–92. https://doi.org/10.4161/fly.19695

Dauben C.M., Pröll-Cornelissen M.J., Heuβ E.M., Appel A.K., Henne H., *et al.* (2021) BMC Genomics 22(1):1–15. https://doi.org/10.1186/s12864-021-07997-1

Depristo M.A., Banks E., Poplin R., Garimella K. V., Maguire J.R., *et al.* (2011) Nat. Genet. 43(5):491–501. https://doi.org/10.1038/ng.806

Gorssen W., Meyermans R., Buys N., and Janssens S. (2020) Anim. Genet. 51(1):32–42. https://doi.org/10.1111/age.12888

Lawrence M., Huber W., Pagès H., Aboyoun P., Carlson M., *et al.* (2013) PLoS Comput. Biol. 9(8):e1003118. https://doi.org/10.1371/journal.pcbi.1003118

Li H., and Durbin R. (2010) Bioinformatics 26(5):589–595. https://doi.org/10.1093/bioinformatics/btp698

Li H., Handsaker B., Wysoker A., Fennell T., Ruan J., *et al.* (2009) Bioinformatics 25(16):2078–2079. https://doi.org/10.1093/bioinformatics/btp352

Maltecca C., Tiezzi F., Cole J.B., and Baes C. (2020) J. Dairy Sci. 103(6):5302–5313. https://doi.org/10.3168/jds.2019-17846

McKenna A., Hanna M., Banks E., Sivachenko A., Cibulskis K., *et al.* (2010) Genome Res 20(9):1297–1303. https://doi.org/10.1101/gr.107524.110

Narasimhan V., Danecek P., Scally A., Xue Y., Tyler-Smith C., *et al.* (2016) Bioinformatics 32(11):1749–1751. https://doi.org/10.1093/bioinformatics/btw044

Nosková A., Bhati M., Kadri N.K., Crysnanto D., Neuenschwander S., *et al.* (2021) BMC Genomics 22(1):290. https://doi.org/10.1186/s12864-021-07610-5

Poplin R., Ruano-Rubio V., DePristo M.A., Fennell T.J., Carneiro M.O., *et al.* (2017) BioRxiv10.1101/201178. https://doi.org/10.1101/201178

R Core Team (2020) *R: A Language and Environment for Statistical Computing* R Foundation for Statistical Computing. https://www.r-project.org/

Shi L., Wang L., Liu J., Deng T., Yan H., *et al.* (2020) J. Anim. Sci. Biotechnol. 11(1):. https://doi.org/10.1186/s40104-020-00447-0

Van der Auwera G.A., Carneiro M.O., Hartl C., Poplin R., del Angel G., *et al.* (2013) Curr. Protoc. Bioinforma.*SUPL.43*. https://doi.org/10.1002/0471250953.bi1110s43

Wickham H., Averick M., Bryan J., Chang W., McGowan L.D., *et al.* (2019) J. Open Source Softw. 4(43):1686. https://doi.org/10.21105/joss.01686

Xie Z., Bailey A., Kuleshov M. V, Clarke D.J.B., Evangelista J.E., *et al.* (2021) Curr. Protoc. 1(3):e90. https://doi.org/10.1002/cpz1.90

Zhan H., Zhang S., Zhang K., Peng X., Xie S., *et al.* (2020) Genes (Basel). 11(5):577. https://doi.org/10.3390/genes11050577

228. Analysis of runs of homozygosity of cattle living in different climate zones

L. Falchi[1*], S. Mastrangelo[2], A. Cesarani[1], G. Senczuk[3], B. Portolano[2], F. Pilla[3] and N.P.P. Macciotta[1]

[1]Dipartimento di Agraria, University of Sassari, 07100 Sassari, Italy; [2]Dipartimento di Scienze Agrarie, Alimentari e Forestali, University of Palermo, 90128 Palermo, Italy; [3]Dipartimento di Agricoltura, Ambiente e Alimenti, University of Molise, 86100 Campobasso, Italy; lfalchi1@uniss.it

Abstract

The genome of the current cattle breeds has been shaped by domestication, adaptation, and by artificial selection. In this study, genomic data of 3,177 animals from 205 different Taurine and Indicine breeds spread worldwide were used to investigate the Runs of Homozygosity (ROH) pattern among three different climatic zones (North, Temperate and Tropics). Our findings confirmed that *Bos taurus* presents the lowest number of ROH per animal, *Bos indicus* the largest, and hybrids (*B. taurus* × *B. indicus*) have intermediate values, respectively. ROH tend to be longer in the temperate areas and shortest in the tropics.

Introduction

Modern cattle breeds originated from the extinct wild Auroch (*Bos primigenius*), that has been domesticated about 8.500 years ago in the Fertile Crescent (Magee and Bradley, 2006). The first hypothesis is that *Bos taurus* derived from *B. primigenius*, and then *Bos indicus* would have been derived from *B. taurus* during early pastoral movements to India, at 4000 BC approximately (Magee and Bradley, 2006). The second hypothesis, which is largely accepted, is that modern cattle arose from two different domestications of aurochs: *B. indicus* would derive from a southern Asian subspecies of Aurochs, the *B. primigenius namadicus* (Bradley *et al.*, 1998). In any case, the modern breeds originated from a complex interaction between natural and artificial selection, and adaptation to environment (Gautier *et al.*, 2010). *B. indicus* is spread in Eurasia and eastern Africa, whereas *B. taurus* is diffused mainly in Europe, North and West Africa, and Middle East (Bruford *et al.*, 2003).

The genomes of these two subspecies and of their crosses (hybrids) can provide interesting insights about their different evolution histories. Among genomic metrics able to highlight these signals, Runs of Homozygosity (ROH) are increasing in popularity. ROH are continuous homozygous chromosomal segments (Gibson *et al.*, 2006) whose length is used as a temporal indicator of the inbreeding events: ROH due to recent inbreeding tend to be longer, whereas ROH due to remote inbreeding are shorter ones. Because homozygous genotypes could be due also to fixation of favorable alleles, ROH occurrence can be the result of both artificial and natural selection (Macciotta *et al.*, 2021).

The aim of this study was to analyze the distribution of ROH in different types of cattle (*B. taurus*, *B. indicus*, and hybrids) raised worldwide, and to assess whether the ROH composition would differ in different climatic regions.

Materials & methods

In this study, genotypes of 3,177 cattle (2,689 *B. taurus*, 330 *B. indicus*, and 158 Hybrid) belonging to 205 different breeds were available from a merge of SNP data conducted in previously published work (Mastrangelo *et al.*, 2020). After edits, 23,313 SNPs mapping on the 29 autosomes were retained for the analysis. The average number of animals per breed was 16±7 (range from 3 to 30). For each breed, the latitude and longitude of country of origin were retrieved from online databases. Breeds were then grouped

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_228, © L. Falchi *et al.* 2022

into three zones according to the latitude of their countries of origin: (1) North, latitude≥45°; (2) Temperate, 45°<latitude>23.26°; (3) Tropics, latitude≤23.26° (including also negative latitudes).

Runs of homozygosity. Consecutive ROH were computed using the '*detectruns*' R package (Biscarini *et al.*, 2018). The following parameters were adopted to define a ROH: at least 15 homozygote SNPs in a row covering a minimum of 2 million base pairs, with no heterozygote or missing allowed. Number of ROH per animal (nROH) and average (meanMb) ROH length were computed. A ROH that started and finished at exactly the same positions within chromosome was defined as unique (Cesarani *et al.*, 2018; Macciotta *et al.*, 2021).

Statistical analysis. The effects of Type (subspecies) and Zone on nROH were tested using the following generalized linear model with a lognormal distribution:

$$nROH = \mu + TYPE + ZONE + e \quad (1)$$

where:

μ is the overall mean; TYPE is the fixed effect of the type (i.e. *B. taurus*, *B. indicus*, and hybrids); ZONE is the fixed effect of the zone (i.e. Tropics, Temperate, and North); e is the residual term.

ROH length was analyzed with the following model:

$$y = \mu + TYPE + CHROM + ZONE + ANIM + e \quad (2)$$

where:

CHROM is the fixed effect of chromosome (29 levels); ANIM is the random effect of the animal, added to model the repeated measurements within animal (3,177 levels); and e is the random residual. The two random effects were assumed to be normally distributed with parameters $N(0,I\sigma_a^2)$ and $N(0,I\sigma_e^2)$, where I is an identity matrix and σ_a^2 and σ_e^2 the variances associated with the animal and the residual random effects, respectively. Both models (1 and 2) were run also including ZONE×TYPE, i.e. the interaction between Zone and Type, removing North because only *B. taurus* animals were present in that area. Mixed models were computed using SAS PROC GLMMIX (SAS Inc. 2012).

Results

A total of 136,922 ROHs were found in 3,177 animals. Removing the duplicated regions, 94,925 unique ROH were found: 1,453 ROH were highlighted at least one time in all the 3 zones, whereas only 649 were found in all the three considered Types. Table 1 shows the four unique ROH that have been found to be shared by more than 70 animals.

Table 1. ROH shared by more than 70 animals.

BTA	Region (Mb)	N°	Animals per zone[1]			Animals per type		
			Temperate	North	Tropics	Taurus	Indicus	Hybrid
12	63.40-69.68	94	51	38	5	86	6	2
16	1.03-4.77	91	52	32	7	80	10	1
27	0.43-6.04	86	45	34	7	81	4	1
15	1.03-6.30	77	40	32	5	62	9	6

[1] Nord = Latitude≥45°; Temperate = 45°<Latitude>23.26°; Tropics = Latitude≤23.26°.

Table 2 shows the results of the mixed model analysis. ZONE affected both nROH and meanMB. Animals farmed in the Tropics showed the highest nROH and the lowest meanMb, whereas lowest nROH and the highest meanMb were found for the Temperate zone, respectively. All the three types differed significantly for nROH, with *B. indicus* having the largest value and *B. taurus* the smallest, respectively. In the model for meanMb CHROM was highly significant. Repeatability for meanMb was 0.09 and 0.13 for the model without and with interaction, respectively.

LS means of the interaction between Zone and Types are reported in Figures 1a and 1b.

The *indicus* showed the highest nROH ($P<0.001$) in both climate regions whereas differences between *taurus* and *hybrids* ($P<0.001$) can be found only in the temperate zone.

Discussion

An effect of the farming zone and, to a lesser extent, of the subspecies on ROH features of the investigated sample of cattle, has been detected. Tropics exhibited the largest number of ROH per animal and the

Table 2. Least squares means of number of ROH and average ROH length per animal.

	Number of ROH per animal	Avereage ROH length
ZONE[2]	($P<0.0001$)	($P<0.001$)
North (Latitude≥45°)	55.34 ± 1.03^{B}	3.56 ± 1.03^{B}
Temperate (45°<Latitude>23.26°;)	41.38 ± 1.03^{C}	3.79 ± 1.03^{A}
Tropics (Latitude≤23.26°)	61.24 ± 1.03^{A}	3.28 ± 1.03^{B}
TYPE	($P<0.0001$)	($P<0.0001$)
Hybrid	40.07 ± 1.05^{B}	3.42 ± 1.06^{B}
Indicus	122.30 ± 1.04^{A}	3.03 ± 1.04^{B}
Taurus	28.61 ± 1.02^{C}	4.28 ± 1.02^{A}

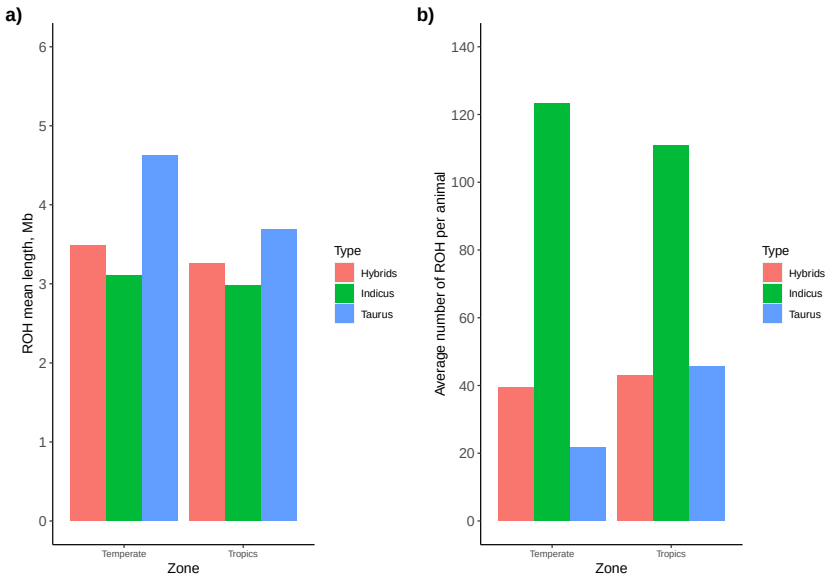


Figure 1. Least squares means of the Zone×Type interaction for the average number of ROH per animal (a) and the average ROH length (b).

smallest average ROH length. Effects of geographical area on the number and length of ROH have been recently reported, albeit in a more limited geographical context, for goats (Cortellari *et al.*, 2021) and sheep (Cesarani *et al.*, 2022). Geographical clines in ROH distributions have been reported also for human populations (Nothangel *et al.*, 2010). A study on mammalian species has evidenced a relationship between ROH length and latitude of farming area (Bruniche Olsen *et al.*, 2018). The higher nROH observed in *B. indicus*, especially due to a larger occurrence of short ROHs (<4 Mb), could be at least partially ascribed to ascertainment bias in genotyping of this subspecies. However, this issue has been at least partially addressed by the criteria here adopted for ROH detection. A larger ROH length was observed for *B. taurus* only in temperate climate (Figure 1b). Usually longer ROH are interpreted as evidence of a stronger selection (Purfield *et al.*, 2012). Thus, results here obtained seems to highlight an effect of selection in favor of *B. taurus* mainly in temperate areas. Finally, an interesting result is represented by the sharing of same unique ROHs across zones and, especially, subspecies.

References

Biscarini F., Cozzi P., Gaspa G., and Marras G. (2018) detectRUNS: Detect runs of homozygosity and runs of heterozygosity in diploid genomes.

Bradley D.G., Loftus R.T., Cunningham P., and MacHugh D.E. (1998) Evol. Anthropol. 6(3):79-86.

Bruford M., Bradley D., and Luikart G. (2003) Nat. Rev. Genet. 4:900-910. https://doi.org/10.1038/nrg1203.

Brüniche-Olsen A., Kellner K.F., Anderson C.J., and DeWoody J.A. (2018) Conserv. Genet. 19:1295-1307. https://doi.org/10.1007/s10592-018-1099-y.

Cesarani A., Sorbolini S., Criscione A., Bordonaro S., Pulina G., *et al.* (2018) Anim. Genet. 49(5):371-383. https://doi.org/10.1111/age.12697.

Cesarani A., Gaspa G., Correddu F., Dimauro C., and Macciotta, N.P. (2022) J. Anim. Breed. Genet. 00:1-15. https://doi.org/10.1111/jbg.12666.

Cortellari M., Barbato M., Talenti A., Bionda A., Carta A., *et al.* (2021) Sci. Rep. 11(10986). https://doi.org/10.1038/s41598-021-89900-2.

Gautier M., Laloë D., and Moazami-Goudarzi K. (2010) PLoS ONE 5(9):e13038. https://doi.org/10.1371/journal.pone.0013038.

Gibson J., Morton N.E., and Collins A. (2006) Hum. Mol. Genet. 15(5):789-795. https://doi.org/10.1093/hmg/ddi493.

Macciotta N.P., Colli L., Cesarani A., Ajmone-Marsan P., Low W.Y., *et al.* (2021) Genet. Sel. Evol. 53(20):1-21. https://doi.org/10.1186/s12711-021-00616-3.

Magee D., and Bradley D. (2006) Genetics And Origins Of Domestic Cattle. In: Zeder M., Bradley D., Emshwiller E. and Smith B. (eds) Documenting Domestication: New Genetic and Archaeological Paradigms. University of California Press, Berkeley, CA, pp. 317-328. https://doi.org/10.1525/9780520932425-025.

Mastrangelo S., Tolone M., Ben Jemaa S., Sottile G., Di Gerlando R., *et al.* (2020) Sci. Rep. 10:14522. https://doi.org/10.1038/s41598-020-71375-2.

Nothnagel M., Lu T.T., Kayser M., and Krawczak M. (2010) Hum. Mol. Genet. 19, 2927-2935. https://doi.org/10.1093/hmg/ddq198.

Purfield D.C., Berry D.P., McParland S. Bradley D.G. (2012) BMC Genet. 13(70). https://doi.org/10.1186/1471-2156-13-70.

229. The genomic inbreeding trend over the last 25 years in three Italian pig breeds

G. Schiavo[1*], S. Bovo[1], A. Ribani[1], S. Tinarelli[2], M. Cappelloni[2], M. Gallo[2] and L. Fontanesi[1]

[1]Department of Agricultural and Food Sciences, University of Bologna, Viale G. Fanin, 40127 Bologna, Italy; [2]Associazione Nazionale Allevatori Suini (ANAS), Via Nizza 53, 00198 Rome, Italy; giuseppina.schiavo2@unibo.it

Abstract

Inbreeding is an important parameter that should be monitored in livestock populations under selection. The traditional inbreeding coefficient (F_{PED}) is based on pedigree data. Runs of homozygosity (ROH) can be used to define one of the most accurate genomic inbreeding parameters, defined as the proportion of the autosomal genome covered with ROH (F_{ROH}). The aim of this work was to evaluate how 25 years of selection have shaped the inbreeding level of the three heavy pig breeds raised in Italy: Italian Large White, Italian Landrace and Italian Duroc. For all breeds, the average F_{ROH} increased over time, as expected in populations under selection. What was interesting is that the rate of inbreeding was higher for F_{PED} than for F_{ROH}. This might mean that the rate of inbreeding could be overestimated by using only pedigree information. F_{ROH} can complement traditional F_{PED} coefficient to monitor inbreeding in these populations.

Introduction

In livestock populations under selection, inbreeding increases as an inevitable consequence of the selection program. Therefore, inbreeding should be monitored to avoid the negative effects, such as inbreeding depression. The traditional inbreeding coefficient (F_{PED}) is calculated using pedigree data (Falconer and Mackay, 1996). This way to calculate inbreeding has several limitations. For example, it needs complete and correct pedigree registrations, it does not account for the true relatedness of the founder animals of the base population, and it does not consider the stochasticity of the recombination events over generations.

High throughput genotyping data can be used to calculate several genomic inbreeding parameters. One of these parameters is based on the identification and then of the sum over the whole animal genome of runs of homozygosity (ROH). ROH are long and continuous stretches of homozygous chromosome regions in which all adjacent single nucleotide polymorphisms (SNPs) are homozygous. The proportion of the autosomal genome covered by ROH can be used to define F_{ROH}, that is considered a good estimation of the true level of inbreeding in an individual animal, which can be averaged over all animals of a population (Ferenčaković *et al.* 2013).

Italian Large White (ILW), Italian Duroc (ID) and Italian Landrace (IL) are the main breeds raised in Italy for the production of heavy pigs. Animals from these breeds are slaughtered at about 160 kg of live weight when they are at least 9 months old and are raised for the production of Protected Designation of Origin (PDO) products. Their breeding programs started in the second half of 1980s.

Genomic inbreeding based on ROH and other parameters has been preliminarily evaluated in several Italian pig breeds by Schiavo *et al.* (2020). ROH have been also analysed in many other European pig breeds, with useful results for the inference of the genetic history and genetic events of these autochthonous pig genetic resources (Schiavo *et al.* 2021).

In this study, to have a general overview of the inbreeding trend in the Italian heavy pig breeds, we evaluated their average inbreeding level (calculated using pedigree and ROH data) over the last 25 years.

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_229, © G. Schiavo *et al.* 2022

Materials & methods

Animals and data. A total of 6,982 pigs of the ILW, ID and IL breeds, born between 1995 and 2019, registered in their breed herd books managed by the Italian Pig Breeders Association (ANAS), were included in this study. The average pedigree minimum depth for these pigs was 5.28± 1.64 for ILW, 4.80±1.627 for ID and 3.43±2.44 for IL. The same animals were genotyped with the GeneSeek® GGP Porcine HD Genomic Profiler SNP chip for a total of 68,516 SNPs. Quality check and filtering of the genotyping data was performed with PLINK 1.9 (Chang *et al.* 2015), with the following parameters: Hardy Weinberg Equilibrium *P*-value>0.001; call rate >0.9. No filter was applied for minor allele frequency. The final dataset included 3,540 ILW, 1,201 ID and 1,969 IL pigs with 51,930 SNPs.

Inbreeding coefficients. Two methods for estimating F_{PED} were used: F_{PEDW}, according to Wright's coefficient with the algorithm proposed by Meuwissen and Luo (1992), and F_{PEDR}, with the recursive method proposed by VanRaden (1992). This last method takes into consideration individuals with unknown relatives, which by definition have a F_{PEDW} coefficient equal to 0, and recursively updates their inbreeding coefficient with the average coefficient of their year of birth. Both methods are implemented in Inbupgf90 (Aguilar and Mizstal, 2008). ROH were computed using PLINK 1.9 with the following parameters: at least 15 consecutive homozygous SNPs; ROH length of at least 1 Mb; no heterozygous SNPs allowed; minimum density of SNPs in a genome window of 1 SNP every 100 kb. F_{ROH} was defined as the sum of ROH of an individual over the autosome size. Different F_{ROH} parameters were estimated according to the minimum length of ROH considered. Namely F_{ROH1} was defined with ROH with minimum length of 1 Mb, F_{ROH4} with a minimum length of 4 Mb, and F_{ROH16} with ROH longer than 16 Mb.

Comparison between coefficients. Pairwise Pearson's correlations between the three inbreeding coefficients were calculated. The correlation values between all methods are reported in Table 1. For the three methods and for each breed, the average inbreeding per year was calculated. Breed-year combinations with more than 15 animals were considered for the analyses. Inbreeding rates of the coefficients from the

Table 1. Pearson's correlations between the different inbreeding coefficients for the three breeds.

		F_{PEDW}	F_{PEDR}	F_{ROH1}	F_{ROH4}	F_{ROH16}
Italian Large White	F_{PEDW}	-	0.982	0.526	0.531	0.515
	F_{PEDR}	0.982	-	0.505	0.511	0.501
	F_{ROH1}	0.526	0.505	-	0.987	0.903
	F_{ROH4}	0.531	0.511	0.987	-	0.926
	F_{ROH16}	0.515	0.501	0.903	0.926	-
		F_{PEDW}	F_{PEDR}	F_{ROH1}	F_{ROH4}	F_{ROH16}
Italian Duroc	F_{PEDW}	-	0.991	0.543	0.560	0.535
	F_{PEDR}	0.991	-	0.528	0.545	0.526
	F_{ROH1}	0.543	0.528	-	0.986	0.908
	F_{ROH4}	0.560	0.545	0.986	-	0.927
	F_{ROH16}	0.535	0.526	0.908	0.927	-
		F_{PEDW}	F_{PEDR}	F_{ROH1}	F_{ROH4}	F_{ROH16}
Italian Landrace	F_{PEDW}	-	0.940	0.558	0.577	0.564
	F_{PEDR}	0.940	-	0.528	0.545	0.533
	F_{ROH1}	0.558	0.528	-	0.988	0.899
	F_{ROH4}	0.577	0.545	0.988	-	0.919
	F_{ROH16}	0.564	0.533	0.899	0.919	-

three different methods were defined as the opposite slope of the regressed values of ln(1-F) on the years of birth, where F was the average value for the related year, as described for example in Doekes *et al.* (2018).

Results

The average inbreeding value (calculated considering all animals included in this dataset) was higher in ID with respect to the other two breeds. In particular, the average F_{PEDW} was 0.054±0.07 in ID, 0.052±0.05 in ILW and 0.033±0.06 in IL. Similar results were obtained with F_{PEDW}, for which the average value was 0.058±0.07 for ID, 0.062±0.05 for ILW and 0.045±0.06 for IL. The average F_{ROH1} was 0.241±0.05 in ID, 0.152±0.05 in IL and 0.139±0.45 in ILW. The trends of F_{ROH1} and F_{PEDW} coefficients are shown in Figure 1. When comparing F_{PEDR} and F_{PEDW}, along with a very high correlation coefficient (R>0.94), also the trends resulted to be very similar. F_{ROH1} was highly correlated with F_{ROH4} (R>0.98), while F_{ROH16} had a R>0.93 with the other two F_{ROH} coefficients. The Pearson's correlations between a genomic-based and a pedigree-based measures ranged from a minimum of 0.5 in ILW to a maximum of 0.58 in IL. All genomic-based values were better correlated with F_{PEDW} than with F_{PEDR}. All correlations are reported in Table 1. The comparison between the slopes of the regressed ln(1-F) on years of birth, with F representing the averaged inbreeding coefficient of a certain year, indicated that the inbreeding rate for ILW and ID was lower with F_{ROH1} (with

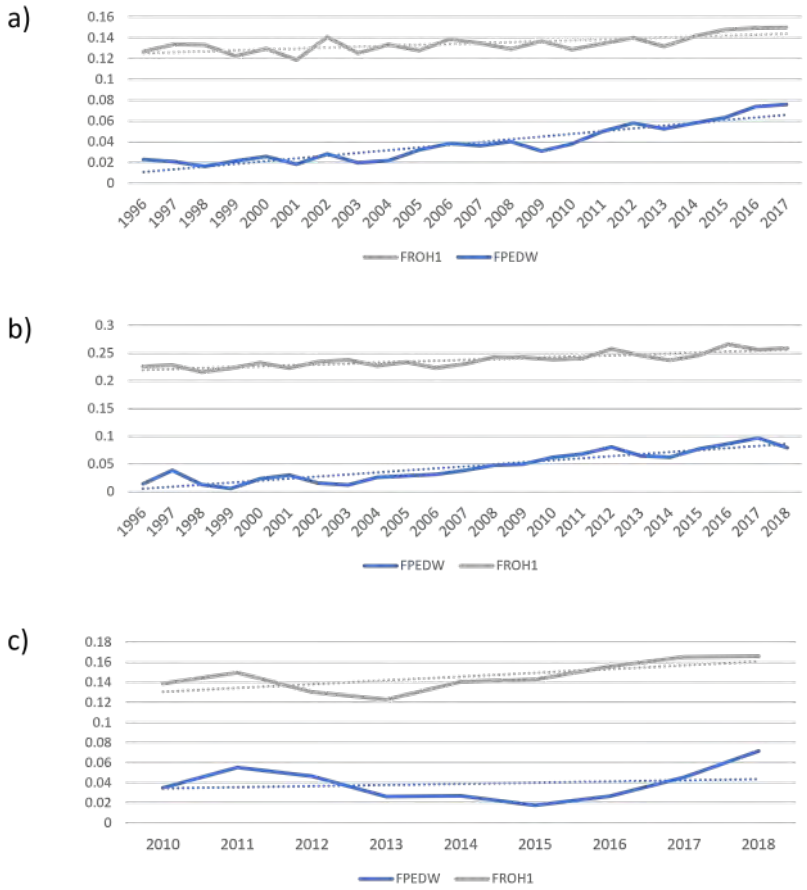


Figure 1. Trends of the F_{PEDW} and F_{ROH1} coefficients over years for the three breeds: (a) Italian Large White; (b) Italian Duroc and (c) Italian Landrace.

an opposite slope of the regression of 0.0002 for ILW and 0.0009 for ID) than with F_{PEDW} (opposite slope of the regression coefficient of 0.00086 for ILW, 0.00169 for ID) and F_{PEDR} (0.00088 ILW and 0.00159 for ID).

Discussion

All breeds had a relatively low level of inbreeding for F_{PEDR}, F_{PEDW}, F_{ROH1}, F_{ROH4}, reflecting a good control of consanguinity during the last decades of the breeding programmes. The low F_{ROH16} indicates scarce occurrence of recent inbreeding events; its relative low correlation with other methods reveals that this parameter is not the proper indicator of inbreeding for these pig breeds. The correlation between F_{PEDR} and F_{PEDW} was high, thus the two methods are comparable for these breeds, probably due to the high proportion of reliable records in the pedigree. The correlation between F_{ROH1} and F_{ROH4} was very high, so the two methods give similar information about the inbreeding level of the animals. The F_{PED} trend over the last 25 years tended to increase for all methods, as expected, but the slope of this trend resulted much more relevant for F_{PED} with respect to the F_{ROH}. This might indicate that there could be an actual overestimation of the inbreeding level for the whole population when using only pedigree information. This means that, in future breeding plans, there could be less limitations in the choice of mating pairs. These results suggest that it could be useful to integrate genomic information aside of traditional inbreeding coefficients in breeding plans defined with optimum contribution selection strategy when the population has a relatively high average inbreeding level. This would make it possible to better control the inbreeding increase over generations.

References

Aguilar I., and Misztal I. (2008). J. Dairy Sci. 91(4):1669-1672. https://doi.org/10.3168/jds.2007-0575

Chang C.C., Chow C.C., Tellier L.C., Vattikuti S., Purcell S.M. *et al.* (2015). Gigascience 4(1):s13742-015. https://doi.org/10.1186/s13742-015-0047-8

Doekes H.P., Veerkamp R.F., Bijma P., Hiemstra S.J., and Windig, J.J. (2018). Genet. Sel. Evol. 50, 15. https://doi.org/10.1186/s12711-018-0385-y

Falconer D.S. and Mackay T.F.C. (1996). Introduction into quantitative genetics. Essex: Prentice Hall.

Ferenčaković M., Hamzić E., Gredler B., Solberg T.R., Klemetsdal G., Curik I. *et al.* (2013). J. Anim. Breed. Genet. 130(4):286-293. https://doi.org/10.1111/jbg.12012

Granado-Tajada I., Rodríguez-Ramilo S.T., Legarra A., and Ugarte E. (2020). J. Dairy Sci. 103(6):5215-5226. https://doi.org/10.3168/jds.2019-17743

Meuwissen T., and Luo Z. (1992) Genet. Sel. Evol. 24:305. https://doi.org/10.1186/1297-9686-24-4-305

Schiavo G., Bovo S., Bertolini F., Tinarelli S., Dall'Olio S. *et al.* (2020). Animal 14(5):910-920. https://doi.org/10.1017/S175173111900332X

Schiavo G., Bovo S., Muñoz M., Ribani A., Fontanesi, L. et al (2021). Anim. Genet. 52(2): 155-170. https://doi.org/10.1111/age.13045

VanRaden, P.M. (1992). J. Dairy Sci. 75(11):3136-3144. https://doi.org/10.3168/jds.S0022-0302(92)78077-1

230. Inbreeding and coancestry trends in Nordic Holstein

S. Tenhunen[2*], J.R. Thomasen[2], L.P. Sørensen[2], G.P. Aamand[3], P. Berg[1,4] and M. Kargo[1]

[1]Aarhus University, Centre for Quantitative Genetics and Genomics, Blichers Alle 20, 8830 Tjele, Denmark; [2]VikingGenetics, Ebeltoftvej 16, 8960 Randers SØ, Denmark; [3]Norwegian University of Life Sciences, NMBU, Universitetstunet 3, 1433 Ås, Norway; [4]Nordic Cattle Genetic Evaluations, Skejby, 8200 Aarhus N, Denmark; saten@vikinggenetics.com

Abstract

Increased rates of inbreeding have been observed in several Holstein populations across the world since genomic selection (GS) was implemented into the breeding programs. The objective of this study was to test whether this is also the case in the Nordic Holstein population using pedigree data. Rates of inbreeding were estimated before, during and after the GS implementation. We could not detect any increase in inbreeding or coancestry rates in Nordic Holstein. Effective population size increased over the studied time periods from 68 to 104, despite fewer sires being used. This might be because more of sires of sons have been used to produce AI bulls compared to the time before implementation of GS.

Introduction

Inbreeding rates have been higher in several Holstein populations around the world after genomic selection (GS) was implemented, compared to before GS (Doekes *et al.*, 2018a; Doublet *et al.*, 2019; Forutan *et al.*, 2018). In theory, GS should decrease loss of genetic diversity (Meuwissen *et al.*, 200; Meuwissen *et al.*, 2016), but in practice the opposite has been observed in several breeds and populations. Loss of genetic diversity due to increasing rate of inbreeding in dairy cattle populations is concerning because it may lead to inbreeding depression, which affects many dairy cattle traits negatively e.g. milk yield, fertility, and udder health traits (Doekes *et al.*, 2019; Makanjuola *et al.*, 2020a), and loss of genetic variation. The Food and Agriculture Organization (FAO) recommends that the increase in inbreeding per generation should not exceed 1% to avoid negative effects of inbreeding depression (FAO, 2013).

Within the Nordic (Denmark, Finland and Sweden) Holstein population, inbreeding trends have not been investigated in recently. The analyses of Sørensen *et al.* (2005) were carried out prior to the introduction of GS in 2011. The Nordic countries have had a common Holstein breeding program since 2010 when the cooperative VikingGenetics (VG) was founded. For this reason, there is a demand for updated research on Nordic Holstein (NH), due to the changes in the breeding program after introduction of GS, to investigate how this has affected the population structure. Thus, the objective of this study is to analyse inbreeding and coancestry trends in the NH population based on pedigree data.

Materials & methods

Pedigree data was supplied by Nordic Cattle Genetic Evaluations. EVA software (Berg *et al.*, 2006) was used to calculate inbreeding coefficient, expected inbreeding coefficient under random mating, alpha (deviation from random mating), coancestry (C), pedigree completeness index (PCI), average generation coefficient (Gencoef), and generation interval (L) (NordGen, 2022). Nordic Holstein females expected to contribute to the next generation of the Holstein breeding population were selected as the base population. Selected females were candidates for VG' Holstein breeding program. They were from milk recording herds and born from January 2016 to November 2018. Only animals with a Nordic Total Merit value >8, Holstein breed percentage >97% and PCI>0.8 were accepted into the base population for pedigree tracing. In total, pedigrees were traced for 372,955 females which reflect approximately 1/3 of the NH female population

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_230, © S. Tenhunen *et al.* 2022

born in 2016-2018. Pedigrees of the selected base population animals were traced back to the known founders, on average 12.4 generations.

Rates of inbreeding were calculated from the regression of the natural logarithm ln(1-F) on year of birth. Yearly rates of change in inbreeding (ΔF) and coancestry (ΔC) were calculated as $\Delta F/\Delta C = 1 - e^{\beta 1}$ (Pérez-Enciso, 1995), where β1 is the regression slope. Effective population size, Ne, (Falconer and Mackay, 1996) was calculated from ΔF as Ne= 1/2ΔF×L, where L is the generation interval. Expected future effective population size (Exp. Ne) was calculated from ΔC as Ne = 1/2ΔC×L. Animals were allocated to three different groups depending on birth year: before implementation of GS 2007-2010, during the implementation 2011-2014, and after implementation of GS 2015-2018. Calculations of effective population size were done separately for each group.

Results & discussion

The generation interval has decreased in NH after implementation of GS. Before GS, the generation interval was on average 4.9 years, and it has decreased to 3.2 years on average after the implementation (Table 1). Inbreeding and coancestry trends (Figure 1) have been stable in NH across time periods, and we were not able to detect any increase in inbreeding in contrast to what has been seen in some other Holstein populations (Doekes *et al.*, 2018b; Forutan *et al.*, 2018; Makanjuola *et al.*, 2020b). On the contrary, ΔF in NH has remained unchanged (0.0016) from before to after implementation of GS. This was also seen for the ΔC (0.0018). Over the time periods number of starting AI sires has decreased, but in the last time period a wider range of sires of sons has been used compared to earlier periods (Figure 1). These observations agree with the theory of decreasing loss of genetic diversity with genomic selection. Similar results for Nordic animals were seen in the inbreeding survey by World Holstein Friesian Federation (Feddersen, 2020).

The increase in effective population size (Ne) from 68 to 104 across the time periods supports the theory, because even though rates of changes have been the same over the time periods, generation interval has decreased which leads to larger Ne. Expected future effective population size similarly increased over the time periods, but is smaller than the Ne, which might indicate that even though animals are less inbred, they are still related to each other. This might be caused by more intense use of related bulls from same imported sires. This will cause a decrease in inbreeding, but coancestry will increase or remain at the same level. After implementation of GS, sires have been used more equally meaning that all bulls will have offspring in the breeding population even though they might not be used as much in the Holstein production population. GS has made selection of AI sires more reliable, and daughter proven sires has become less competitive; therefore, bulls can be circulated faster in the breeding program. In this study focus was more on the inbreeding trend behind NH females. Studying inbreeding trends behind Nordic AI sires will be beneficial because other studies show that increase in inbreeding is higher for the bulls (Doekes *et al.*, 2018a; Doublet *et al.*, 2019).

Table 1. Yearly rates of change of inbreeding (ΔF) and coancestry (ΔC), generation interval (L), effective population size (Ne) and expected future effective population size (Exp. Ne) in each group of animals before, during and after genomic selection was implemented in NH.

	ΔF	ΔC	L	Ne	Exp. Ne
2007-2010	0.0016	0.0018	4.86	68	59
2011-2014	0.0017	0.0014	4.08	71	88
2015-2018	0.0016	0.0018	3.17	104	87

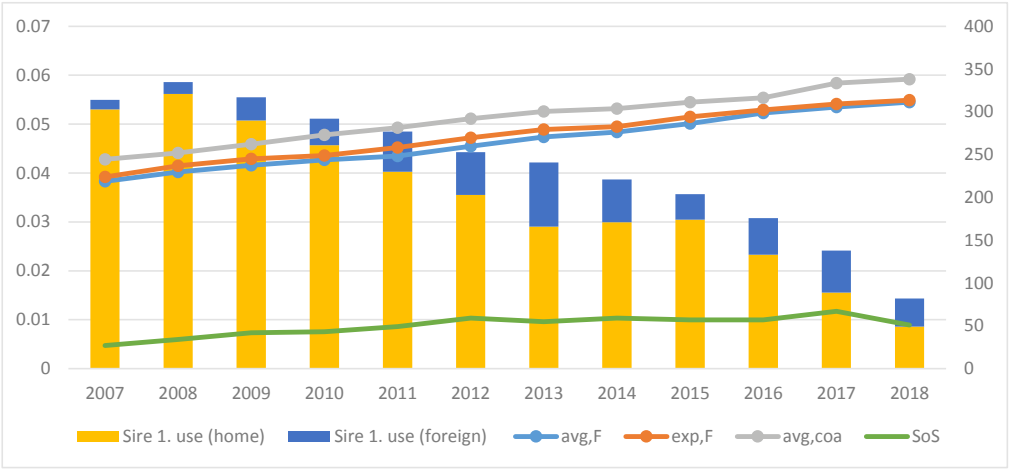


Figure 1. Number of starting Nordic and foreign AI bulls, average inbreeding, expected inbreeding, average coancestry and number of sires of sons (SoS) between years 2007-2018 in NH population.

Considering the effective population size, there has been improvement since the last study on NH from 2005 (Sørensen *et al.*, 2005), where the historical Ne was 49-69 and expected future effective population size was 43-51 across the time periods. Even though Ne for NH has increased, there is reason to monitor the development and improve the breeding scheme to gain higher Ne. Populations with Ne between 50-100 are facing risk of inbreeding depression, and even higher Ne might be needed for long-term (FAO, 2013; Frankham *et al.*, 2014). After the implementation of GS, NH is above the short-term threshold but in the long-term the breed might be in danger of facing issues related to inbreeding depression, when the expected future effective population size is on the threshold for short-term survivability. Therefore, methods for controlling future increase in inbreeding more efficiently need to be developed.

Breeding programs today are based on genomic selection and there has been recommendations that methods to control inbreeding should then be based on genomic information (Sonesson *et al.*, 2012). There has been debate about the best way to manage inbreeding in livestock populations (Meuwissen *et al.*, 2020; Henryon *et al.*, 2019a; Henryon *et al.*, 2019b) and more research on the topic is required. In the future research, inbreeding and coancestry trends should be based on genomic information to investigate whether similar results are seen when studying only genomic tested animals in the breeding population for NH.

References

Berg, P., Nielsen, J., and Sørensen, M. K. (2006). Proc. of the 8th WCGALP, Belo Horizonte, Minas Gerais, Brazil, 13-18 (pp. 27-09). Instituto Prociência.

Doekes, H.P., Veerkamp, R.F., Bijma, P., Hiemstra, S.J., and Windig, J.J. (2018a). Genet. Sel. Evol. 50(1):1-16. https://doi.org/10.1186/s12711-018-0385-y

Doekes, H.P., Veerkamp, R.F., Hiemstra, S.J., Bijma, P., van der Beek, S., and Windig, J.J. (2018b, February). Proc. of the 11th WCGALP, Auckland, New Zeland (pp. 11-16).

Doekes, H.P., Veerkamp, R.F., Bijma, P., de Jong, G., Hiemstra, S.J., and Windig, J.J. (2019). Genet. Sel. Evol. 51(1):1-16. https://doi.org/10.1186/s12711-019-0497-z

Doublet, A.C., Croiseau, P., Fritz, S., Michenet, A., Hozé, C., Danchin-Burge, C., and Restoux, G. (2019). Genet. Sel. Evol. 51(1):1-13. https://doi.org/10.1186/s12711-019-0495-1

Falconer, D.S., and Mackay, T.F.C. (1996). Introduction to quantitative genetics, Longman Group Limited, United Kingdom, 192, 65-71.

FAO. 2013. *In vivo* conservation of animal genetic resources. FAO Animal Production and Health Guidelines. No. 14. Rome.

Feddersen, E. (2020). World Inbreeding Trend in Holsteins. Available at: http://www.whff.info/documents/WHFF-Council_WorldInbreedingTrend.pdf

Forutan, M., Mahyari, S.A., Baes, C., Melzer, N., Schenkel, F. S., and Sargolzaei, M. (2018). BMC Genomics, 19(1):1-12. https://doi.org/10.1186/s12864-018-4453-z

Frankham, R., Bradshaw, C.J., and Brook, B.W. (2014). Biol. Conserv. 170:56-63. https://doi.org/10.1016/j.biocon.2013.12.036

Henryon, M., Liu, H., Berg, P., Su, G., Nielsen, H.M., Gebregiwergis, G.T., and Sørensen, A.C. (2019a). Genet. Sel. Evol. 51(1):1-12. https://doi.org/10.1186/s12711-019-0475-5

Henryon, M., Berg, P., Liu, H., Su, G., Ostersen, T., and Sørensen, A.C. (2019b). In Proc. of the AAABG 23rd Conference (Vol. 27, pp. 222-5), Armidale, Australia

Makanjuola, B.O., Maltecca, C., Miglior, F., Schenkel, F.S., and Baes, C.F. (2020a). BMC Genomics, 21(1):1-15. https://doi.org/10.1186/s12864-020-07031-w

Makanjuola, B.O., Miglior, F., Abdalla, E.A., Maltecca, C., Schenkel, F.S., and Baes, C.F. (2020b). J. Dairy. Sci. 103(6):5183-5199. https://doi.org/10.3168/jds.2019-18013

Meuwissen, T.H.E., Hayes, B.J., and Goddard, M.E. (2001). Genetics 157(4):1819-1829. https://doi.org/10.1534/genetics.116.189795

Meuwissen, T.H.E., Hayes, B., and Goddard, M. (2016). Anim. Front. 6(1):6-14. https://doi.org/10.2527/af.2016-0002

Meuwissen, T.H., Sonesson, A.K., Gebregiwergis, G., and Woolliams, J.A. (2020). Front. Genet. 11:880. https://doi.org/10.3389/fgene.2020.00880

NordGen (2022), EVA program. Available at https://www.nordgen.org/en/our-work/nordgen-farm-animals/eva-program/

Pérez-Enciso, M. (1995). J. Anim. Breed. Genet. 112(1-6):327-332. https://doi.org/10.1111/j.1439-0388.1995.tb00574.x

Sonesson, A.K., Woolliams, J.A., and Meuwissen, T.H. (2012). Genet. Sel. Evol. 44(1):1-10. https://doi.org/10.1186/1297-9686-44-27

Sørensen, A.C., Sørensen, M.K., and Berg, P. (2005). J. Dairy. Sci. 88(5):1865-1872. https://doi.org/10.3168/jds.S0022-0302(05)72861-7

231. Effect of genomic selection on genetic variability in chromosomal level in the Finnish Ayrshire

K. Sarviaho*, P. Uimari and K. Martikainen

Department of agricultural sciences, University of Helsinki, P.O. Box 28, 00014, Helsinki, Finland; katri.sarviaho@helsinki.fi

Abstract

This research aimed to study the effect of genomic selection (GS) on median length of runs of homozygosity (ROH) and average genomic inbreeding estimated from the ROHs in the Finnish Ayrshire. The transition from traditional phenotype-based breeding value prediction to genomic-based estimation was assumed to have taken place during 2012-2014. The median length of the ROHs and the average genomic inbreeding varied by chromosomes. In addition, the median length of ROHs and the average genomic inbreeding were generally higher in bulls than in cows. After the introduction of GS, ROHs became longer in the majority of chromosomes. However, the average genomic inbreeding did not follow the trend of ROH length. Those chromosomes, where the median length of ROHs and the average genomic inbreeding increased, included quantitative trait loci associated with a range of traits under selection in the Finnish Ayrshire.

Introduction

Nordic Red dairy population (RDC) constitutes of Finnish Ayrshire cattle (FAY), Swedish Red cattle, and Danish Red cattle (NAV 2021a). They have a joint genetic evaluation by Nordic Cattle Evaluation (NAV) and share a balanced breeding goal: Nordic Total Merit (NTM) (NAV 2021a). Traits included in NTM for RDC are yield, growth, fertility, calving, udder health, general health, claw health, feet and legs, udder, milkability, temperament, longevity, youngstock survival, and saved feed (NAV 2021a). Application of genomic selection (GS) in dairy cattle breeding began in the 2010's (NAV 2021b). Despite its benefits, for example, faster genetic progress, GS could increase average inbreeding, decrease genetic variability, and cause inbreeding depression (e.g. Makanjuola *et al.* 2020). Inbreeding can be measured from runs of homozygosity (ROH) (Purfield *et al.*, 2012). Both selection for and selection against the traits could increase the median length of ROHs and result in increased genomic inbreeding (F_{ROH}) (Kim *et al.*, 2015). This research aimed to study the effect of GS on ROH length and average genomic inbreeding on chromosomal level in FAY (NAV 2021a).

Materials & methods

The genotype data was received from NAV, and the pedigree and breed composition data were obtained from the Finnish Animal Breeding Association. Data included genotypes of 46,914 SNPs (single nucleotide polymorphism) from 24,658 animals that were at least 50% Finnish Ayrshire (FAY). There were 3,228 bulls and 3,856 cows born in 1980-2011, while 1,897 bulls and 15,677 cows were born in 2015-2020. Bulls were genotyped with Illumina BovineSNP50 v1.0 and v2.0 Beadchip, and the majority of the cows were genotyped with Illumina BovineLD v.1 Beadchip, and the genotypes were imputed to 50K with FImpute (Sargolzaei *et al.*, 2014). The genotype data included SNP variants in chromosomes *Bos Taurus* autosomes (BTA) and did not include any missing genotypes.

Traditional phenotype-based breeding value estimation was assumed between 1980-2011, while estimation mainly based on genomic information was assumed from 2015. ROHs were identified using PLINK version 1.90b6.20 (Chang *et al.* 2015, Purcell and Chang 2021) following a method by Meyermans *et al.* (2020). Briefly, no pruning for minor allele frequency, deviations from Hardy-Weinberg equilibrium, or linkage disequilibrium was done. The sliding window was set to 63 SNPs, minimum number of SNPs in ROH

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_231, © K. Sarviaho *et al.* 2022

to 63, gap size to 500 Kbp, minimum density to 1 SNP/64 Kbp, and threshold of 0.05. One heterozygous genotype per window was allowed. Estimates of F_{ROH} were calculated as the number of SNPs in ROHs per total number of SNP-variants (Pryce *et al.*, 2014).

Results & discussion

Both the median length of ROHs (Figure 1) and the average F_{ROH} (Figure 2) varied by gender and by chromosomes. In general, ROHs were longer, and the average F_{ROH} was higher in bulls than cows, as was expected, since bulls tend to be highly selected. Additionally, we found the median length of ROHs increasing in the majority of BTAs after the introduction of GS (Figure 1). The median length of ROHs in bulls born between 1980-2011 and between 2015-2020 was 5.44 Mbp and 5.61 Mbp, respectively, and in cows between 1980-2011 and between 2015-2020, 5.36 Mbp and 5.52 Mbp, respectively. However, the average F_{ROH} did not follow the trend of the ROH length (Figure 2). Indeed, the average F_{ROH} in bulls born between 1980-2011 and between 2015-2020 was 0.067 (SD 0.023) and 0.066 (SD 0.023), respectively. Likewise, the average F_{ROH} in cows born between 1980-2011 and between 2015-2020 was 0.065 (SD 0.022) and 0.063 (SD 0.022), respectively. GS and optimization of genetic contributions simultaneously could lengthen the ROHs while the rate of inbreeding is restricted (e.g. Clark *et al.* 2013, Kim *et al.* 2015).

A range of QTLs associated with traits with the biggest weights in NTM, fertility, and yield have been identified in RDC and FAY, which could explain the differences in ROH length found in this study. For example, QTLs associated with fertility traits have been identified in RDC and FAY in BTA01, BTA02, and BTA04 (Höglund *et al.* 2014, Höglund *et al.* 2015, Schulman *et al.* 2011). In addition, in RDC, QTLs associated with fertility have been identified in BTA03, BTA06, BTA07, BTA09, BTA11, BTA13, BTA15, and BTA24 (Höglund *et al.* 2014, Höglund *et al.* 2015). In our study, the median length of ROHs increased

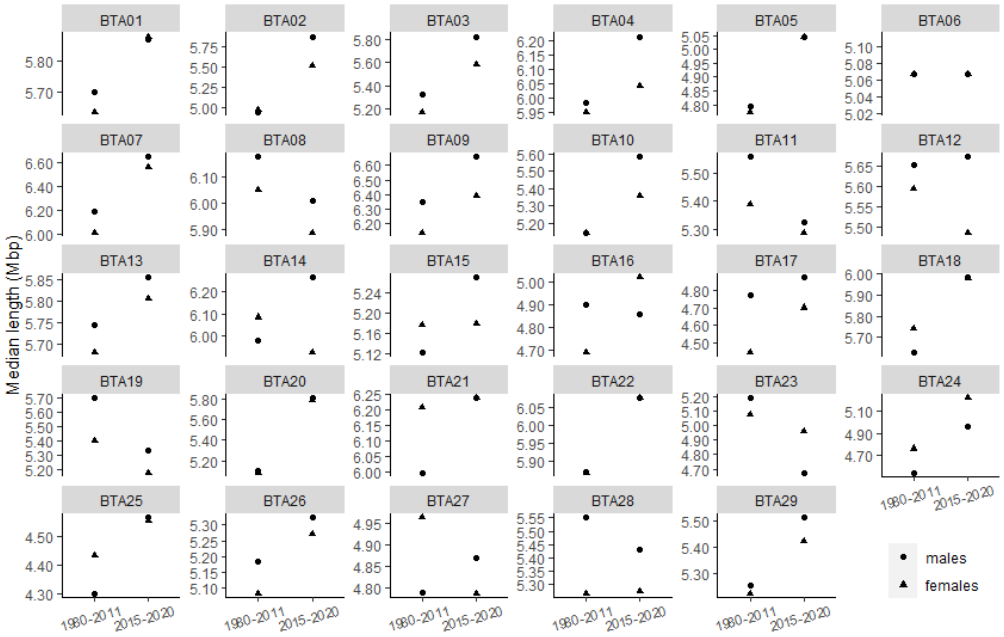


Figure 1. Median length. Median length of ROHs in chromosomes of *Bos Taurus* autosomes (BTA) before and after the introduction of genomic selection by gender.

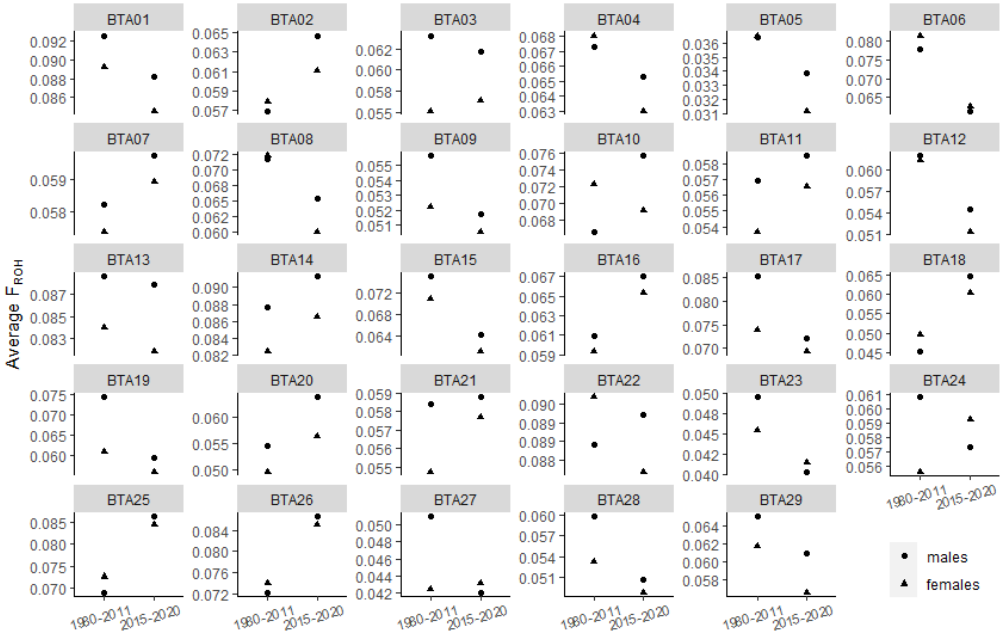


Figure 2. Average F_{ROH}. Average genomic inbreeding F_{ROH} in *Bos Taurus* autosomes (BTA) before and after the introduction of genomic selection by gender.

in all forementioned chromosomes after the introduction of GS. However, in BTA11, the median length decreased, and in BTA15, it increased only in bulls. In comparison, the average F_{ROH} generally decreased with few exceptions. Interestingly, in BTA06, the median length of ROHs was similar before and after the introduction of GS. Additionally, QTLs associated with fertility have been identified in FAY in BTA 20 (Schulman *et al.*, 2011). Both the median length of ROHs and the average F_{ROH} increased in BTA20 after the introduction of GS. In addition, QTLs associated with either fertility or yield traits have been identified in RDC in BTA05 and BTA20 (Höglund *et al.* 2015, Iso-Touru *et al.* 2016, Kadri *et al.* 2015). Again, the median length of ROHs increased in BTA05 after GS was introduced. However, the average F_{ROH} decreased. Finally, QTLs associated with either fertility or yield have been identified in FAY in BTA12 (Kadri *et al.*, 2014). In our study, the median length of ROHs increased only slightly in bulls and decreased in cows after the introduction of GS, while the average F_{ROH} decreased in both bulls and cows.

Conclusions

ROHs became longer after the introduction of genomic selection in majority of the *Bos Taurus* autosomes, where QTLs associated with fertility have been identified. However, no distinct trend in ROH length or average genomic inbreeding was observed in the chromosomes, including QTLs associated with yield traits. In addition, the average genomic inbreeding was similar before and after the introduction of genomic selection and did not follow the trends of ROH length. Further research on potential ROH-islands (e.g. Nothnagel *et al.*, 2010) is needed to assess the effect of genomic selection on genomic variability in the Finnish Ayrshire breed.

Acknowledgements

We are grateful for Suomen Naudanjalostussäätiö for funding our research. The authors thank NAV (Nordic Cattle Genetic Evaluation, Aarhus, Denmark) and Faba (The Finnish Animal Breeding Association, Hollola, Finland) for providing the genotype and pedigree data used in this study.

References

Chang, C. C., Chow C. C., Tellier, L. CAM., Vattikuti, S. *et al.* (2015) GigaScience 4 (1): 7. https://doi.org/10.1186/s13742-015-0047-8

Clark, S. A., Kinghorn, B. P., Hickey, J. M., and van der Werf, J. H. J. (2013) Genet. Sel. Evol. 45:44. http://dx.doi.org/10.1186/1297-9686-45-44

Höglund, J. K., Sahana, G., Guldbrandtsen, B., and Lund, M. S. (2014) BMC Genet. 15:8. https://doi.org/10.1186/1471-2156-15-8

Höglund, J. K., Buitenhuis, B., Guldbrandtsen, B., Lund, M. S., and Sahana, G. (2015) BMC Genet. 16:110.

Iso-Touru, T., Sahan, G., Guldbrandtsen, B., Lund, M. S., and Vilkki, J. (2016) BMC Genet. 17:55. https://doi.org/10.1186/s12863-016-0363-8

Kadri, N. K., Sahana, G., Charlier, C., Iso-Touru, T. *et al.* (2014) PLoS Genetics 10: e1004049. https://doi.org/10.1371/journal.pgen.1004049

Kadri, N. K., Guldbrandtsen, B., Lund, M. S., and Sahana, G. (2015) J. Dairy Sci. 98: 9015-9025. https://doi.org/10.3168/jds.2015-9599

Kim, E. S., Sonstegard, T. S., Van Tassell, C. P., Wiggans, G., and Rothschild, M. F. (2015) PLoS ONE 10: e0129967. https://doi.org/10.1371/journal.pone.0129967

Makanjuola, B. O., Maltecca, C., Miglior, F., Schenkel, F. S., and Baes, C. (2020) BMC Genomics 21:605. https://doi.org/10.1186/s12864-020-07031-w

Meyermans, R., Gorssen, W., Buys, N., and Janssens, S. (2020) BMC Genomics 21: 94. https://doi.org/10.1186/s12864-020-6463-x

NAV. (2021a) NAV routine genetic evaluation of Dairy Cattle – data and genetic models. Available at: https://nordicebv.info/wp-content/uploads/2021/10/NAV-routine-genetic-evaluation_EDITYSS-08102021.pdf.

NAV. (2021b) Changes in the official calculation of breeding values in NAV for dairy breeds. Available at: https://nordicebv.info/wp-content/uploads/2020/05/Changes-in-the-official-calculation-of-breeding-values-in-NAV-for-dairy-breeds-1999-2020.pdf.

Nothnagel, M., Lu, T. T., Kayser, M., and Krawczak, M. (2010) Hum. Mol. Genet. 19: 2927-2935. https://doi.org/10.1093/hmg/ddq198

Pryce, J. E., Haile-Mariam M., Goddard, M. E., and Hayes, B. J. (2014) Genet. Sel. Evol. 46: 71. https://doi.org/10.1186/s12711-014-0071-7

Purcell, S. M. and Chang, C.C. (2021) Available at: https://www.cog-genomics.org/plink/1.9/.

Purfield, D. C., Berry, D. P., McParland, S., and Bradley, D. (2012) BMC Genet. https://doi.org/10.1186/1471-2156-13-70

Sargolzaei, M., Chesnais, J. P., and Schenkel, F. S. (2014) BMC Genomics 15: 478. https://doi.org/10.1186/1471-2164-15-478

Schulman, N. F., Sahana, G., Iso-Touru, T., McKay, S. D. *et al.* (2011) Anim. Genet. 42: 263-269. https://doi.org/10.1111/j.1365-2052.2010.02149.x

232. Genomic diversity of two indigenous Croatian goat breeds

I. Drzaic[1*], V. Orehovački[2], L. Vostrý[3], I. Curik[1], N. Mikulec[1] and V. Cubric-Curik[1]

[1]University of Zagreb Faculty of Agriculture, Svetošimunska c. 25, 10000 Zagreb, Croatia; [2]Ministry of Agriculture, Department of the Livestock Gene Bank, Poljana Križevačka 185, 48260 Križevci, Croatia, [3]Czech University of Life Sciences Faculty of Agrobiology, Food, and Natural Resources, Kamycka 129, 16500 Prague, Czech Republic; ikovac@agr.hr

Abstract

This study used genome-wide SNP data to characterise the Istrian and Croatian Coloured goats genetically. Population structure and conservation status were analysed using 11 Istrian and 33 Croatian Coloured goats and 506 additional animals belonging to 19 European goat breeds. Neighbour-net and admixture results show that the two breeds are unique and retain a great proportion of the ancestral genetic structure. Despite low N_e for Istrian goats, both breeds showed low levels of inbreeding (F_{ROH}~0.05). Considering these results, the Istrian goat is at high risk of extinction and should be carefully monitored.

Introduction

The domestic goat (*Capra hircus*) is one of the first domesticated animals. A relatively small animal with modest requirements and good resistance to harsh environmental conditions, it migrated with humans and colonised Croatian territory around 7,000 years BP (Amills *et al.*, 2017). Since then, goat rearing has been a constant presence in Croatian livestock production, although there have been some difficult times. In 1954, the government banned goat keeping in open areas, which led to a drastic decline in goat populations in Croatia. The indigenous Croatian goats are raised in extensive systems with low production. The Croatian Coloured goat (GCH) has 2261 animals in selection, while the Istrian goat (GCI) has only 14 animals in selection (Croatian Agency for Agriculture and Food, 2021). Local breeds are part of the genetic and socio-cultural heritage of the local community.

For this reason, it is essential to preserve the genetic diversity of the remaining breeds, mostly captured in non-selected autochthonous breeds (Medugorac *et al.*, 2009). To date, there is little knowledge about the genetic structure of indigenous Croatian goat breeds. This study used genome-wide genotype data to analyse the genetic diversity and conservation status of two Croatian indigenous goat breeds, the Istrian goat and the Croatian Coloured goat.

Materials & methods

Sample collection and data genotyping. A total of 11 hair samples from GCI and 34 tissue samples from GCH were genotyped using the Illumina Goat 50K SNP Chip (Illumina INC., San Diego, CA). Quality control was performed using SVS ver 8.9.0 (SNP and Variation Suite, Golden Helix, USA). It included removal of all: (1) loci on the mtDNA, sex chromosome and without position; (2) genotypes with a call rate <90%; (3) individuals with a call rate <95%. We used 506 additional samples downloaded from the Dryad repository (https://doi.org/10.5061/dryad.v8g21pt) representing 19 European goat breeds (VAL – Valdostana, ORO – Orobica, BIO – Bionda dell'Adamello, VSS – Valpassiria, ASP – Aspromontana, ARG – Argentata, GAR – Garganica, DIT – Di Teramo, CCG – Ciociara Grigia, SAR – Sarda, SIT – Saanen, NIC– Nicastrese, CRS – Corse, SFR – Saanen, FSS – Fosses, PTV – Poitevine, PYR – Pyrenean, PVC – Provencale, ARR – Arran, CRP – Carpatian).

Population structure. Population structure was determined using a model-based clustering algorithm computed in ADMIXTURE ver. 1.3.0 software (Alexander and Lange, 2011) for 2-25 populations. Prior to

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_232, © I. Drzaic *et al.* 2022

calculation, data was pruned in PLINK 1.9 software (Chang *et al.*, 2015) using the indep-pairwise 50 5 0.03 command. Nei's distances (Nei, 1972) was calculated using the StAMPP package (Pembleton *et al.*, 2013) in R software (R Core Team, 2020) and used to generate the Neighbour-net graph using SplitsTree ver 4.14.2 software (Huson and Bryant, 2006)

Conservation status. Individual genomic inbreeding (F_{ROH}) was assessed using ROH data and then averaged by breed. ROH was determined using SVS software with parameters previously defined in Ferenčaković *et al.* (2013). Effective population size (N_e) analysis was calculated based on LD using GONE (Santiago *et al.*, 2020) and using the molecular co-ancestry method implemented in NeEstimator V2.1 (Do *et al.*, 2014.) with default parameters.

Results & discussion

After quality control, 10 GCI and 33 GCH with 54,216 genotypes remained. The merged Croatian and European dataset consisted of 548 goats and 44,867 genotypes, and after pruning, 19,530 SNP remained.

The genetic relationship between the analysed breeds is represented with the Neighbour-net (Figure 1). Consistent with PCA (result not shown), several clear clusters are revealed in which the breeds form sub-branches. The GCI was clustered with the SIT and SFR, while the GCH was clustered with the CRP goat and is very close to the BIO and VSS. The longest branch formed Ireland ARR goat due to island isolation, similar to the results of Colli *et al.* (2018).

The ADMIXTURE plots showed the results for K=3, K=8 and K=14 (Figure 2). First breed split occurred at K=3, and continued as the K value increased. After calculating CV errors, K=14 was identified as the best-fitting representation. K=14 confirmed clusters that diverged at lower K values and exhibited varying breed admixture degrees. The breeds ARR, DIT, FSS, GAR, ORO, PTV, PVC, PYR, SFR, SIT, SAR, VAL were assigned to separate clusters. Other breeds studied showed a high degree of admixture, as previously described in Nicoloso *et al.* (2015) and Colli *et al.* (2018). GCI showed a common genetic background

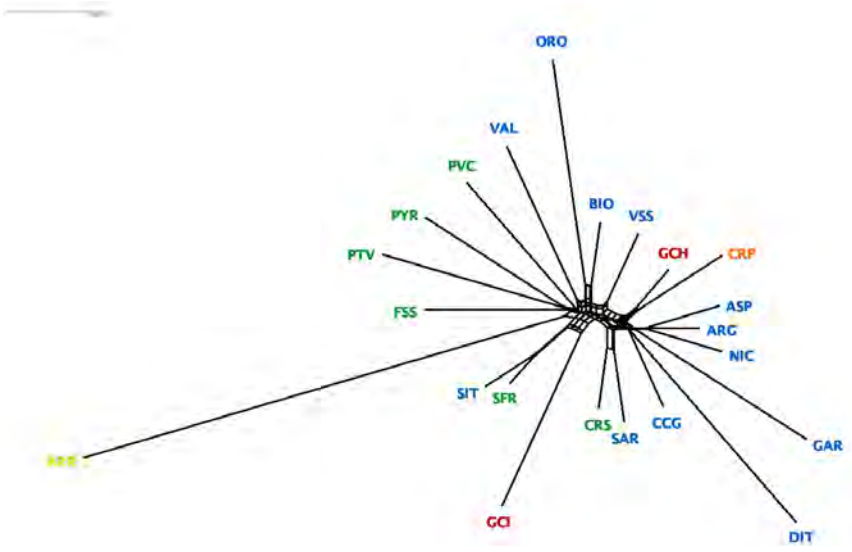


Figure 1. Neighbour-Net graph of Nei's distances between breeds. Each color represents one country: Croatia, Italy, France, Ireland, Romania.

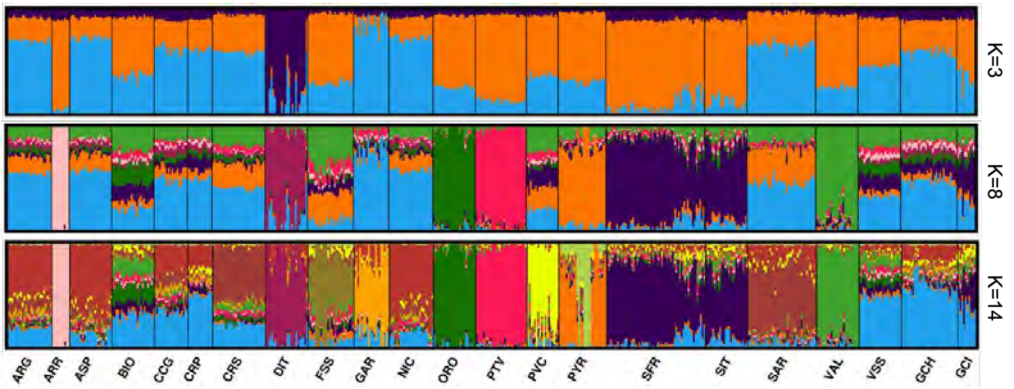


Figure 2. Population structure of 21 selected goat breeds using ADMIXTURE.

of mixed origin similar to BIO and VSS with minor components of ORO, PVC, and Saanen breeds. The admixture pattern of GCH was similar to ARG, ASP, CRP, and NIC with a minor component of SAR. The Croatian goat breeds retained a large part of the historical genetic structure.

The conservation status of Croatian goats was estimated using genomic inbreeding (F_{ROH}) and effective population size (Ne). The distribution of F_{ROH} and estimates of N_e are presented in Table 1. Effective population size was shown for generation 1 (NeEstimator, GONE) and generation 5 (GONE) to be comparable with other authors. GCI had a lower N_e value than the GCH, which was expected due to the smaller population size of the GCI. The results for the GCI should be viewed with caution because the sample size is small. We observed an initial pattern of decreasing N_e in both breeds, but then a small increase in N_e was observed in GCH, probably due to a bottleneck. In general, highly productive breeds (Angora, Boer, Sannen, Alpine) show lower effective population size than local breeds with lower production (Colli *et al.*, 2018). N_e should be at least 100 to ensure that an animal population is viable in the long term in terms of genetic diversity (Meuwissen, 2009), which means that the GCI is at high risk of extinction.

Considering the small N_e it is surprising that the GCI, together with CGH, has a very low genomic inbreeding of 0.06 and 0.04, respectively. The low inbreeding could be explained by high genetic admixture. Similar F_{ROH} values were obtained for native Italian ARG (Bertolini *et al.*, 2018; Cortellari *et al.*, 2021; Mastrangelo *et al.* 2021) SAR, BIO, VSS, NIC (Bertolini *et al.*, 2018, Cortellari *et al.*, 2021) and French CRS (Bertolini *et al.*, 2018) breeds. F_{ROH} values decreased with increasing minimum ROH length, suggesting lack of recent anthropogenic influence. In summary, Croatian goat breeds have great genetic diversity, with a large part of historical admixture resulting from an extensive production system. Decrease in recent inbreeding (F_{ROH}>16Mb) suggest no recent human selection. Caution should be taken when implementing

Table 1. Effective population size and inbreeding coefficients for Croatian goat breeds.

Population	Ne					
	NeEstimator	GONE		F_{ROH}		
		Generation 1	Generation 5	>2 Mb	>8 Mb	>16 Mb
CGI	5.40	27.21	∞	0.06	0.05	0.04
CGH	119.20	477.10	439.27	0.04	0.03	0.02

breeding programmes, especially for the GCI, as the small N_e could seriously affect the overall biological fitness of the breed. This study was supported by ANAGRAMS: IP-2018-01-8708, project No.QK1810253 and Sir je IN: KK.01.1.1.04.0058.

References

Alexander D.H. and Lange K. (2011) BMC Informatics 12, 246. https://doi.org/10.1186/1471-2105-12-246

Amills M., Capote J., Tosser-Klopp G. (2017) Animal Genetics 48, 631-644. http://doi.org/10.1111/age.12598

Bertolini F., Figueiredo Cardoso T., Marras G., Nicolazzi E., *et al.* (2018) Genet Sel Evol 50:59. https://doi.org/10.1186/s12711-018-0424-8

Brito L.F., Jafarikia M., Grossi D.A., Kijas J.W., *et al.* (2015) BMC Genetics 16:67. https://doi.org/10.1186/s12863-015-0220-1

Chang C.C., Chow C.C., Tellier L.C.A.M., Vattikuti S., *et al.*, (2015) GigaScience 4, 1. https://doi.org/10.1186/s13742-015-0047-8

Colli L., Milanesi M., Talenti A., Bertolini F., *et al.* (2018) Genet Sel Evol 50:58. https://doi.org/10.1186/s12711-018-0422-x

Cortellari M., Bionda A., Negro A., Frattini S., *et al.* (2021) Genet Sel Evol 53:92. https://doi.org/10.1186/s12711-021-00685-4

Croatian Agency for Agriculture and Food (2021) Sheep, Goats and Small Animals Breeding, Annual Report 2020. Croatian Agency for Agriculture and Food, Osijek, Croatia.

Do C., Waples R.S., Peel D., Macbeth G.M., *et al.* (2014) Molecular Ecology Resources 14, 209-214. https://doi.org/10.1111/1755-0998.12157

Ferenčaković M., Solkner J. and Curik I. (2013) GSE 45:42. http://www.gsejournal.org/content/45/1/42

Huson D.H. and Bryant D. (2006) Mol Biol Evol 23(2), 254-267. https://doi.org/10.1093/molbev/msj030

Mastrangelo S., Di. Gerlando R., Sardina M.T., Sutera A.M., *et al.* (2021) Animals 11, 1510. https://doi.org/10.3390/ani11061510

Medugorac I., Medugorac A., Russ I., Veit-Kensch C.E., *et al.*, (2009) Mol Ecol 18, 3394-3410. https://doi.org/10.1111/j.1365-294X.2009.04286.x

Meuwissen T. (2009) *Acta Agriculturae Scand. Section A - Anim. Sci.* 59(2), 71–79. https://doi.org/10.1080/09064700903118148

Nei M. (1972) Am Nat 106, 283-292. https://doi.org/10.1086/282771

Nicoloso L., Bomba L., Colli L., Negrini R., *et al.* (2015) Genet Sel Evol 47:62. https://doi.org/10.1186/s12711-015-0140-6

Pembleton L.W., Cogan N.O.I. and Forster J.W. (2013) Molecular Ecology Resources 13, 946-952. https://doi.org/10.1111/1755-0998.12129

R Core Team (2020). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Available at: https://www.R-project.org/

Santiago E., Novo I., Paradinas A.F., Saura M., *et al.* (2020) Mol Bio Evol 37(12), 3642-3653. https://doi.org/10.1093/molbev/msaa169

233. Consequences on the genome homozygosity of breeding program decisions to improve milk production in Manchega dairy sheep

A. Rubio-de Juan[1], C. Diaz[2*], M.J. Carabaño[2], M.R. Fernandez-Santos[3] and M. Ramón[1]

[1]Instituto Regional de Investigación y Desarrollo Agroalimentario y Forestal de Castilla-La Mancha (IRIAF), Av. del Vino s/n, 13300 Valdepeñas, Spain; [2]Instituto Nacional de Investigación y Tecnología Agraria y Alimentaria CSIC-INIA, Ctra. de la Coruña km 7.5, 28040 Madrid, Spain; [3]SaBio IREC (CSIC – UCLM – JCCM), Faculty of Pharmacy, UCLM, Albacete, Spain; cdiaz@inia.es

Abstract

Selection focused only on productive traits, which has been the case in small ruminant breeding programs for many years, has resulted in damage to fitness traits. This deterioration could be due, in part, to increased homozygosity in specific genomic regions linked to these traits. With this hypothesis, this study has made use of genomic information from HD chips to characterize ROH regions sheep to explore whether there has been an increase in homozygosity in genome regions associated with traits of economic interest for the breeding program. Results show an increase in homozygosity in regions associated with productive (milk, fat, and protein yields), reproductive (reproductive seasonality), or health traits (clinical mastitis), among others.

Introduction

Breeding programs have been widely used for many years to improve traits of economic interest in various production systems. In some species, gains achieved in traits under selection have been substantial, with yields doubling for many of the traits (Lu Cao *et al.*, 2020). However, due to the complex nature of the relationships between production and functional traits (fertility, health, resilience), which in many cases are antagonistic, the improvement of production levels has resulted in an impairment of functionality (Berry *et al.*, 2014). Possible explanations for this antagonism are the existence of pleiotropic antagonisms or hitchhiking effects and the accumulation of homozygosity in specific regions of the genome (Mastrangelo *et al.*, 2018).

The Manchega sheep breed has been under selection for more than 25 years. The main selection objective has been milk yield, with fat and protein contents and udder shape being included in recent years. The response to selection in milk has resulted in almost doubling the production of this breed, while fertility or udder shape has deteriorated.

Different information can be used to quantify the effects that selection for productive traits may have on other traits. Particularly, the development of genotyping platforms has allowed the availability of individual genome-level information, making it possible to monitor changes in the genome along generations to assess the impact of selection and correlated responses.

This work aimed to make use of genome information to analyze the distribution of homozygosity along the genome and infer if selection for productive traits may have increased homozygosity in regions of the genome associated with functional traits.

Materials & methods

Genomic information from 324 AI rams belonging to the Manchega sheep breeding program was used. This information was provided by AGRAMA breeders association and consisted of SNP data from the

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_233, © A. Rubio-de Juan *et al.* 2022

ovine Illumina HD (700K) bead chip. Rams were selected to maximize representativeness of animals in the population. Quality Control (QC) removed samples and variants with missing call rates exceeding 10%, and only autosomes were considered. A total of 540,733 markers and 322 rams remained after QC. Criteria followed to define ROH were the following (Macciotta *et al.* 2021, Signer-Hasler *et al.* 2019, Peripolli and Baldi 2016): only regions greater than 1 Mb and with at least 100 SNPs were considered; the scanning window could contain up to 5 missing calls and 1 heterozygous, must have had at least 50 SNPs per 50 kb on average and if two consecutive SNPs were more than 1000 Kb apart they could not be part of the same ROH. ROH were classified into three groups based on their length: small ranges (from 1 to 2.5 Mb), mid ranges (from 2.5 to 5 Mb), and long ranges (>5 Mb). Distribution of ROHs between animals belonging to different generations was studied to identify ROH regions that have changed throughout the development of the breeding program. The information on the sheep QTL database (https://www.animalgenome.org/cgi-bin/QTLdb/OA/index) was used to have a first approximation of the traits of interest to which these regions of homozygosity could be associated.

Results

Figure 1 shows the number of genotyped animals available in each generation and the coverage of ROH regions per chromosome. In general, the homozygosity observed in this breed was low in agreement with previous estimates carried out in this breed using pedigree information or 50K SNP chips (results provided by the breeders association). ROH regions classified as small range were the most abundant, with abundance decreasing as ROH length increases.

Some interesting regions were identified when examining differences in ROH distribution and abundance among generations. First, two regions of homozygosity conserved in all animals were found on chromosomes OAR2 and OAR10 (Figure 2). In these regions, QTLs associated with production traits such as milk yield, and fat, and protein contents, and morphological traits such as the presence of horns and the testes' weights (Hao Li *et al.*, 2020) have been described. Some genomic regions showing increasing homozygosity over generations were found. These regions contain QTLs associated with milk (OAR3 and OAR16), fat (OAR13, OAR16) and protein (OAR3) yields, milk casein percentage (OAR16), body weight (OAR6), fat density (OAR6), reproductive seasonality (OAR3), clinical mastitis and milk bacterial count (OAR16) (Mateescu and Thonney, 2010; García-Gámez *et al.*, 2012; Banos *et al.*, 2017; Dettori *et al.*, 2017; Hao Li *et al.*, 2020).

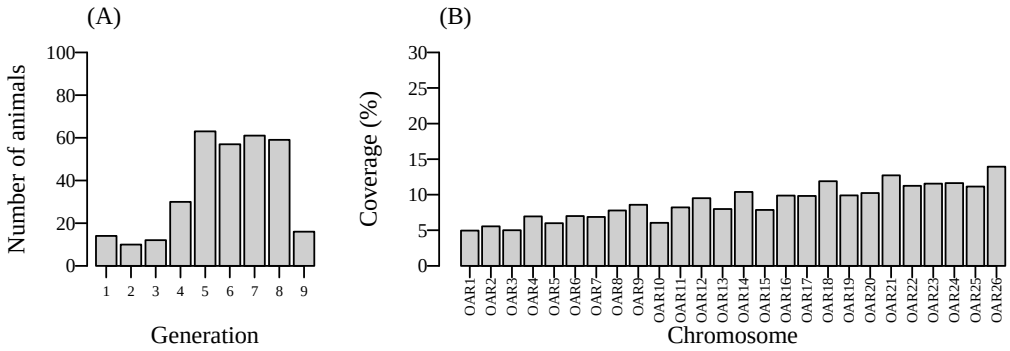


Figure 1. (A) Number of animals genotyped on each generation (from 1990 to 2016). (B) Coverage of ROH regions on the sheep genome.

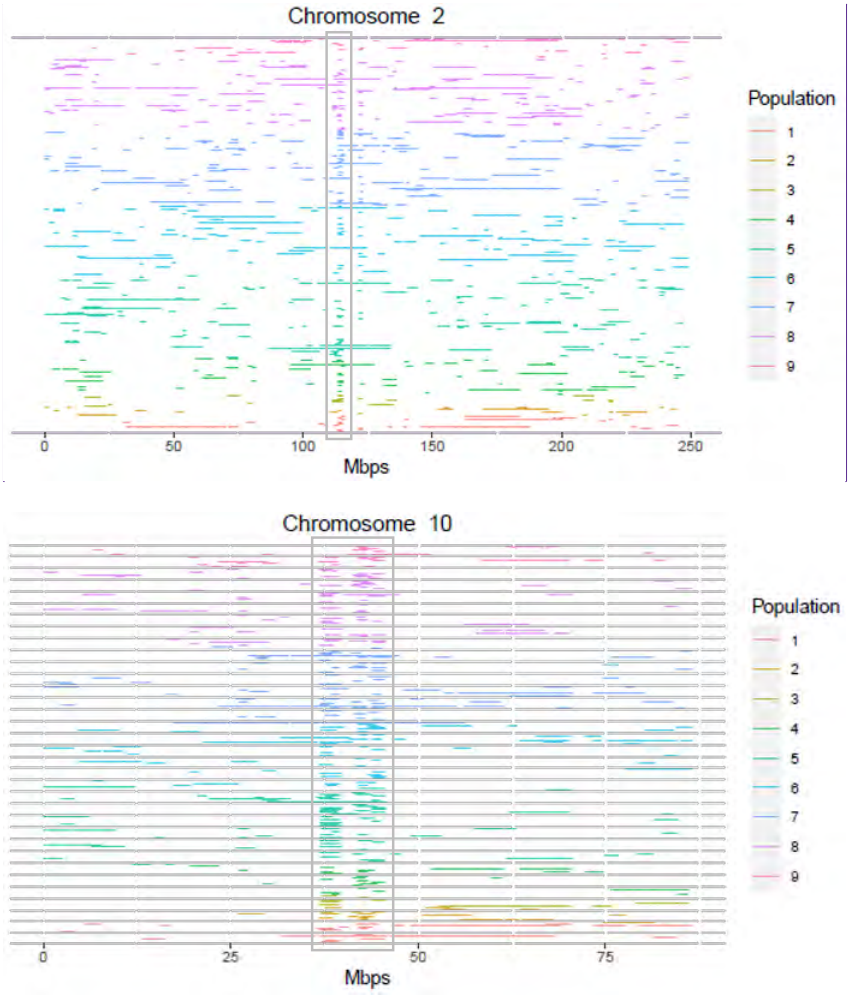


Figure 2. Distribution of ROH regions on chromosomes OAR2 and OAR10 for the 322 animals included in the study. In each graph, rows represent the distribution of ROHs along the chromosome for each ram. Rams belonging to different generations (1-9) are identified by a color scale, being distributed from the oldest (bottom of the graph) to the most recent (top of the graph). A rectangle was used to highlight regions of conserved homozygosity.

Discussion

Characterization of the regions of homozygosity using genomic information from HD SNP chips has allowed the identification of some regions with conserved homozygosity as well as other regions where increased homozygosity has been observed. Most of them were small ranges, something expected since long ROHs have been usually associated with recent inbreeding, and these are usually broken within a few generations as a consequence of recombination (Sumreddee *et al.*, 2021).

The two regions of conserved homozygosity identified on chromosomes OAR2 and OAR10 have been associated with production and morphological QTLs in sheep. More interestingly, some regions have been identified in which homozygosity has increased over generations. In these regions, QTLs associated with traits of productive interest such as fat and protein yields, milk casein percentage, traits such as body size

or body fat density, or fitness traits related to reproductive seasonality or udder health have been described. Thus, selection for increased milk production may have led to increased inbreeding in these regions that would determine the quality of milk production or the greater or lesser susceptibility of animals to mastitis.

In a following work, we will examine the effect that the homozygosity on these regions could have on the characters that have been described in these regions. Thus, we will collect productive, reproductive and health information on these animals and see if there is an effect of inbreeding depression on these traits.

Acknowledgements

This work has been funded by project IMPAR (SBPLY17/180501/000369). We also thank the breeders association AGRAMA for providing genotypic and phenotypic data.

References

Banos G., Bramis G., Bush S.J., Clark E.L., McCulloch M.E.B. *et al.* (2017). BMC genomics 18(1): 624. https://doi.org/10.1186/s12864-017-3982-1

Berry D.P., Wall E., Pryce J.E. (2014). Animal 8(1):105–121. https://doi.org/10.1017/S1751731114000743

Dettori M.L., Pazzola M., Paschino P., Amills M., Vacca G.M. (2018). Journal of dairy science 101(11): 9978-9986. https://doi.org/10.3168/jds.2018-14914

García-Gámez E, Gutiérrez-Gil B., Sahana G, Sánchez J-P., Bayón Y, *et al.* (2012). PLoS One 7(10): e47782. https://doi.org/10.1371/journal.pone.0047782

Li H., Wu X-L., Tait Jr R.G., Bauck S., Thomas D.L., et al (2020) Animal Genetics 51(4): 624-628. https://doi.org/10.1111/age.12956

Lu Cao, Huiming Liu, Han A. Mulder, Mark Henryon, et al (2020). Frontiers in Genetics. https://doi.org/10.3389/fgene.2020.00251

Macciotta Nicolo P. P., Colli L., Cesarani A., Ajmone-Marsan P., Low W.Y., *et al.* (2021) Genet Sel Evol 53: 20. https://doi.org/10.1186/s12711-021-00616-3

Mastrangelo S., Sardina M.T., Tolone M., Di Gerlando Rm et al (2018). Animal 12(12): 2480-2488. https://doi.org/10.1017/S1751731118000629

Mateescu R.G., Thonney M.L. (2010). Animal Genetics 41(5): 454-9. https://doi.org/10.1111/j.1365-2052.2010.02023.x

Peropolli E., Baldi F. (2016) Animal Genetics 48(3):255-271. https://doi.org/10.1111/age.12526

Signer-Hasler H., Burren A., Ammann P., Drögemüller C., Flury C. (2019) Animal Genetics 50 (4):512-525. https://doi.org/10.1111/age.12828

234. Heterosis can be explained by thousands of partially recessive, deleterious mutations with very small effects at sites that are conserved across species

M.E. Goddard

Faculty of Veterinary and Agricultural Science, University of Melbourne and Agriculture Victoria, Bundoora, Victoria, Australia; mike.goddard@agriculture.vic.gov.au

Abstract

A model is proposed to explain heterosis and inbreeding depression. The model assumes the Kacser and Burns model of gene action at thousands of sites which are conserved across species. This implies many polymorphisms with very small effects and partial dominance. This explains 3 surprising features of heterosis – little dominance variance is observed despite dominant gene action; heterosis occurs between populations that have diverged at high N_e; and heterosis is only slightly less than expected from estimates of inbreeding depression.

Introduction

Two fundamental techniques for genetic improvement of animals and plants have been selection and crossbreeding. Crossbreeding benefits from heterosis for which the most parsimonious explanation is directional dominance and differences between breeds in allele frequencies. Directional dominance also explains inbreeding depression. Typically, inbreeding depression and heterosis occur for traits related to fitness such as fertility and many production traits (Doekes *et al.* 2021), but we know very little about the loci responsible. That is, we do not know the number of loci, their effect sizes, the degree of dominance or their allele frequencies. For instance, genome wide association studies (GWAS) have found thousands of markers associated with quantitative traits but most show no significant dominance (Palmer *et al.* 2021).

Three features of inbreeding depression and heterosis seem surprising – there is little or no dominance variance; heterosis occurs even in crosses between subspecies such as *Bos indicus* and *Bos taurus*; and heterosis per degree of inbreeding is only a little less than inbreeding depression per degree of inbreeding (Goddard and Ahmed, 1982). Despite dominance causing inbreeding depression, dominance variance is hard to estimate and many estimates are not significantly different from zero, although this may not be so surprising when the number of loci is large and minor allele frequencies are low (Hill *et al.*, 2008). If populations diverge at moderate effective population size (N_e), selection should prevent deleterious alleles from drifting to high frequency and so there should be little heterosis when the populations are crossed. In reality heterosis is high in wide crosses such as *B. indicus* × *B. taurus* (Goddard and Ahmed 1982, Buning *et al.* 2019). If populations diverged entirely due to drift, the divergence at loci causing heterosis should be the same as at neutral loci. Therefore, we can estimate the inbreeding (%F) that has occurred since two populations diverged using random markers. If this inbreeding coefficient is multiplied by the inbreeding depression per 1%F observed in matings of close relatives we can predict the heterosis expected. Goddard and Ahmed (1982) did this and found the heterosis observed was only slightly less than expected.

The aim of this paper is to present a model that can explain these observations concerning heterosis and inbreeding depression.

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_234, © M.E. Goddard 2022

Results

Inbreeding depression is explained by variation at conserved sites. Sites in the genome that have been conserved across vertebrates have been identified. Xiang *et al.* (2021) found that homozygosity at conserved sites, but not at other sites, decreased mean trait values in cattle. Yengo *et al.* (2021) found the same result in human traits. Genomic sites conserved across vertebrates must have been subject to very long term, consistent selection. This selection would maintain the most favourable allele and all mutations would decrease fitness.

Kacser and Burns model. This conclusion can be combined with the model of Kacser and Burns (1981) in which a trait that is additively inherited (e.g. concentration of an enzyme =x) affects a quantitative trait or fitness (y) in a curvilinear manner such as y =kx/(1+x) (The Michaelis-Menten equation for enzyme reactions). In this case, y increases as x increases but approaches an asymptote at y=k. Selection will increase x and tend to maintain or conserve alleles that increase x. In this model, mutations which when homozygous reduce x to zero also reduce y to zero but the heterozygote shows dominance. The degree of dominance depends on the value of x in the wild type homozygote. The fitness of heterozygotes for a lethal mutation has been estimated to be approximately 5% below that of the wild type homozygote and this corresponds to a value of x in the wild type homozygote of 19 so that the y = 19k/20 in the homozygote and 9.5k/10.5 in the heterozygote.

Conserved sites are also enriched for additive effects on quantitative traits but they only explain part of the additive variance (e.g. 20% in Xiang *et al.*, 2021). At non-conserved sites that affect quantitative traits, the most favourable allele must vary depending on the background genotype and the environment. For instance, for a trait subject to stabilising selection, the value of a mutation depends on the optimum value of the trait and the current mean. The Kacser and Burns model does not seem to apply to these loci, at least not to their effect on fitness.

Additive and dominance effects. I will assume that the Kacser and Burns model applies only to conserved sites. For small effects on the underlying scale (a_x) the effects on the y scale are

$$a_y = ka_x/(1+x)^2$$

$$d_y = ka_x^2/(1+x)^3 = a_y^2\,(1+x)/k$$

$$d_y/a_y = a_x/(1+x)$$

where x = the mean of x.

The genetic variances and inbreeding depression (D) in y due to one locus are

$$V_a = 2pqa_y^2$$

$$D = 2pqd_y = 2pqa^2\,(1+x)/k = V_a\,(1+x)/k$$

$$V_d = (2pqd_y)^2$$

where p and q=1-p are allele frequencies.

These quantities depend on 3 parameters – the heterozygosity (2pq), the size of the additive effect on x relative to the mean of x ($a_x/(1+x)$) and k/(1+x) which converts from the scale of x to that of y.

Provided V_a is not correlated with (1+x)/k across loci, then $D = V_a (1+x)/k$ holds for the combined effect of all loci.

An example. For a locus that explains 0.0001 of the additive genetic variance in y and has a heterozygosity of 2pq = 0.2, then $k/(1+x) = 0.1\sigma_a$. Then for this locus, $D = 0.0001V_a (1+x)/k = 0.00890\sigma_a$. and d/a = 0.2. If we assume that conserved sites explain 20% of additive genetic variance then there are 2,000 loci contributing to inbreeding depression and combined $D = 1.78\ \sigma_a$. This is approximately the case for milk yield and other traits that show moderate heritability and inbreeding depression. The dominance variance $V_d = 0.0016\ V_a$ or close to zero.

Table 1 gives the results for a range of input assumptions. Reducing the effect size on the underlying x scale, reducing the number of loci, reducing the heterozygosity all reduce the inbreeding depression, while increasing the number of loci increases inbreeding depression.

The conclusion is that commonly observed values of additive and dominance variance and inbreeding depression can be explained by a model with thousands of loci with small effects and partial dominance, and moderate level of heterozygosity.

Slow inbreeding with selection. The inbreeding depression calculated above applies to rapid inbreeding, such as matings between sibs, where selection has little opportunity to act. However, populations of animals diverge more slowly and selection acts to prevent a rise in the frequency of deleterious mutations. For the small effect loci described above the selection coefficients may be $s < 0.001$. At effective population sizes $N_e < 10{,}000$, these loci will drift but at larger N_e selection will prevent their frequency rising greatly. Thus, the small effect mutations at conserved sites will experience drift in breeds with $N_e = 100$ but the wild type allele will be conserved in species with $N_e > 10{,}000$. During rapid inbreeding the inbreeding depression is proportional to heterozygosity (2pq). However, when 2 breeds or populations are crossed the maximum gain in fitness or a trait occurs when there are no homozygotes for the deleterious allele in the F1. This eliminates the loss of fitness caused by homozygosity for these deleterious, partially recessive alleles which is proportional to q^2 where q is the frequency of the deleterious allele. If there is no effective selection and allele frequencies drift randomly then the inbreeding depression and heterosis are equivalent. However, if selection is effective in keeping q small then q^2 is much smaller than 2pq and so heterosis is less than

Table 1. Inbreeding depression (D) in units of genetic standard deviation and dominance variance as functions of the number of conserved loci, the additive variance (Va) per locus, the size of variant effects on the x scale (a/(1+x)), and the heterozygosity (2pq).

No. conserved loci	Va per locus	a/(1+x)[1]	2pq	Inbreeding depression[2]	Vd
2000	0.0001	0.2	0.2	1.78	0.0016
2000	0.0001	0.2	0.1	1.26	0.0008
2000	0.0001	0.1	0.2	0.89	0.0004
200	0.001	0.2	0.2	0.56	0.0016
4000	0.0001	0.2	0.2	3.57	0.0032

[1] Units are the additive genetic standard deviation of y.

[2] Units of genetic standard deviation of y at F=1.

inbreeding depression. In other words, lethal recessive mutations contribute to inbreeding depression and even to heterosis if breeds diverge with very small N_e, but they do not contribute much to heterosis among breeds that diverge slowly at high N_e. Conversely, numerous partially dominant mutations of small effect can contribute to inbreeding depression and to heterosis.

Discussion

The model put forward here of thousands of partially recessive, deleterious mutations of very small effect at genomic sites that are conserved across species can explain 3 surprising features of heterosis, that is, that there is little dominance variance, heterosis occurs even between populations that have diverged at high N_e and heterosis per degree of inbreeding is only a little less than inbreeding depression per degree of inbreeding. I do not imply that this is the only model that can explain the observed facts but only that it is one possibility.

As the number of loci increases while the inbreeding depression is kept constant, the dominance variance declines because inbreeding depression depends on the dominance effect whereas the dominance variance depends on that effect squared (Robertson and Hill 1983). The low dominance variance explains why GWAS seldom find significant dominance effects (just as additive variance explains the power to find additive effects).

Although it is hard to cite published evidence, anecdotal evidence suggests that heterosis occurs even in crosses between species. Farmers in Bhutan deliberately cross Yaks and cattle or Mithun and cattle, and the value of mules is famous. These cases could be due to a completely different mechanism to that discussed here but if they are due to the same mechanism as discussed here it implies that species carry many slightly deleterious, partially recessive alleles segregating at high allele frequency.

References

Bunning H., Wall E, Chagunda M.G.G., Banos G., Simm G. (2019) *J. Anim. Sci.*, 97, 29-34, https://doi.org/10.1093/jas/sky406

Doekes, H.P.; Bijma, P.; Windig, J.J.(2021).*Genes*,*12*, 926. https://doi.org/10.3390/genes12060926

Goddard, M. E., and Ahmed A.M. (1982). *Proc. of the 2nd WCGALP*, vol. 8, pp. 377-382.

Hill, W.G., Goddard, M.E. and Visscher, P.M. 2008. *PLoS Genetics* 4: e1000008. https://doi.org/10.1371/journal.pgen.1000008

Hivert V., Sidorenko J., Rohart F., Goddard M.E., Yang J., Wray N.R., Yengo L., Visscher P.M.,(2021) *Amer. J. Human Genetics*, 108, 786-798, https://doi.org/10.1016/j.ajhg.2021.02.014.

Kacser, H. and Burns J.A (1981). *Genetics* 97: 639-666.

Palmer DS, Zhou W, Abbott L, Baya N, Churchhouse C, Seed C, *et al.* (2021)

bioRxiv 2021.08.15.456387; https://doi.org/10.1101/2021.08.15.456387

Robertson, A.,Hill W.G. (1983) *Proc. Royal Soc Series B-Biological Sciences* 219, 253-264

Yengo, L., Yang J., Keller M.C., Goddard M.E., Wray N.R., and Visscher P.M. (2021). *The Amer. J. Human Genetics* 108:1488-1501. https://doi.org/10.1016/j.ajhg.2021.06.005

Xiang, R., Breen E.J., Bolormaa S., Vander Jagt C.J., Chamberlain A.J., Macleod I.M. and Goddard M.E. (2021). C*ommunications Biology* 4,: 1-10. https://doi.org/10.1038/s42003-021-02874-9

235. Signatures of selection and runs of homozygosity in the genome of several fancy and meat rabbit breeds

M. Ballan[1*], S. Bovo[1], G. Schiavo[1], M. Schiavitto[2], R. Negrini[3] and L. Fontanesi[1]

[1]Division of Animal Sciences, Department of Agricultural and Food Sciences, University of Bologna, Viale Giuseppe Fanin 46, 40127 Bologna, Italy; [2]Associazione Nazionale, Coniglicoltori Italiani (ANCI), Contrada Giancola snc, 71030 Volturara Appula (FG), Italy; [3]Associazione Italiana Allevatori, Via G. Tomassetti 9, 00161 Roma, Italy; mohamad.ballan2@unibo.it

Abstract

The European rabbit (*Oryctolagus cuniculus*) is a multi-purpose species raised for meat, fiber, and fur production. It is also considered an important animal model and a companion animal, with many fancy breeds distinguished by several exterior traits. In this study, we analysed SNP array datasets from 16 fancy and meat rabbit breeds using F_{ST} analysis and runs of homozygosity (ROH) analysis. We performed F_{ST} comparisons using the windows-based approach. ROH islands were defined by evaluating the percentile-based threshold of homozygous SNPs frequency. We identified 309 genomic regions that contained putative signatures of selection. ROH islands were detected in a total of 11 autosomes, which covered about 10-20 Mb of the rabbit genome. The putative regions under selection contain genes involved in pigmentation processes and external traits (e.g. *ASIP*, *MC1R*, *TYR*, and *LIPH*). The present study contributes to clarify the genetic architecture and mechanisms that affect phenotypic differences between rabbit breeds.

Introduction

The European rabbit (*Oryctolagus cuniculus*), known simply as rabbit, is a multi-purpose livestock species raised for meat, fiber, and fur production. It is also considered an important animal model and a companion animal. Since its domestication, human-driven artificial directional selection has constituted many breeds and lines. Fancy breeds are mainly distinguished by several exterior traits (e.g. coat colour, hair structure, body size, ear size and position, and structure of the skull), and production breeds that have been constituted by considering performance and production traits. Therefore, this species is characterized by a vast amount of phenotypic diversity linked to economically relevant traits and morphological characteristics. Most of the genetic diversity within this species has been shaped by a combination of several genetic events (e.g. genetic drift, bottleneck, isolation, introgression, migration, crossbreeding) that contributed to define a lot of genetic resources, thus far mostly untapped and uncharacterized.

The rabbit genome has been sequenced and explored (Bertolini *et al.*, 2014; Carneiro *et al.*, 2014) and a high-throughput single nucleotide polymorphism (SNP) genotyping platform is available. This SNP chip has been used in genome wide association studies carried out in several rabbit populations to detect genomic regions affecting economically relevant traits, including production traits such as intramuscular fat and its composition (Sosa-Madrid *et al.*, 2020a, 2020b, Laghouaouta *et al.*, 2020) and reproductive traits, such as litter size (Sosa-Madrid *et al.*, 2020c) and the number of teats (Bovo *et al.*, 2021). Other studies have contributed to explain the variability of external traits, e.g. coat colour, that largely define breed-specific features in most traditional and fancy breeds (Demars *et al.*, 2018; Fontanesi *et al.*, 2006, 2010, 2014a, 2014b; Utzeri *et al.*, 2014).

In this study, we analysed SNP array datasets from a total of 16 fancy and meat rabbit breeds with the aim to identify signatures of selection using F_{ST} analysis. Furthermore, we identified runs of homozygosity (ROH) and we evaluated the occurrence of ROH islands in the genome of some of these breeds to identify putative signatures of selection that might be derived by different selection histories and structures of these rabbit genetic resources.

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_235, © M. Ballan 2022

Materials & methods

Animals and genotyping. Biological specimens (hair roots or buccal swaps) were sampled from a total of 712 rabbits from three meat breeds (Italian White, n. 256; Italian Spotted, n. 93; Italian Silver, n. 20) and 12 fancy breeds (Belgian Hare, n. 24; Burgundy Fawn, n. 6; Champagne d'Argent, n. 19; Checkered Giant, n. 79; Coloured Dwarf, n. 20; Dwarf Lop, n. 20; Ermine, n. 20; Giant Grey, n. 27; Giant White, n. 20; Rex, n. 19; Rhinelander, n. 28; and Thuringian, n. 9) registered to the herd book of the National Rabbit Breeders Association. One commercial meat breed (n. 52) was from another albino meat population. All animals included in this study were selected in order to avoid highly related individuals (no full- or half-sibs). DNA was extracted using the Wizard Genomic DNA Purification kit (Promega Corporation, Madison, WI, USA). Animals were then genotyped using the Affymetrix Axiom OrcunSNP Array (Affymetrix Inc., Santa Clara, CA, USA) which interrogates 199,692 DNA markers. Quality check of the data via the Axiom Analysis Suite and PLINK v.1.9 (Chang *et al.*, 2015) led to discard low-quality samples and DNA markers (call rate <0.90). After filtering, a total of 702 animals and 139,922 SNPs were used in the subsequent analyses.

Data analyses. F_{ST} analyses were performed based on the Hudson estimator as this index is independent of sample size (Bhatia *et al.* 2013). Signatures of selection were computed in 350 kb sliding genome windows, with a step size of 100 kb. In total, 7,951 genome windows were computed after having tested windows of variable sizes as described by Rubin *et al.* (2010). Windows with less than 3 SNPs were discarded from the analyses. Two approaches were used to identify signatures of selection. The first one compared one breed against all others whereas the second one grouped some breeds together according to common features (colour patterns, body size, and use/specialization). Following Rubin *et al.* (2010) putative signatures of selection were identified from genome windows at the extreme lower end of the distributions (99.98th percentile of the distribution). This led to the identification of 14 genome windows for each comparison. We used PLINK v.1.9 (Chang *et al.* 2015) software to call ROH considering the following parameters: (1) the minimum length that constituted the ROH was 1 Mb; (2) a minimum number of 15 consecutive homozygous SNPs to call ROH and a maximum of 2 heterozygous SNPs in the homozygous SNP-windows; (3) the maximum gap between consecutive SNPs was 1000 kb. ROH islands were defined by evaluating the percentile-based threshold of homozygous SNP frequency. First, the proportion of SNPs residing within a ROH was calculated for a given breed by counting the number of times a SNP appeared in a ROH within a given breed divided by the total number of genotyped rabbits of that breed. Then, a percentile threshold of the frequency of a SNP in ROH was calculated. The percentile-based threshold was defined considering the top 1% of SNPs observed in a ROH in each breed. Adjacent SNPs, having a frequency of ROH occurrences over or equal to the identified thresholds, and with a distance ≤1 Mbp between them, constituted ROH islands (Schiavo *et al.* 2021). Genomic regions that displayed signatures of selection from F_{ST} analysis and ROH islands were annotated with the Bedtools v.2.17. (https://bedtools.readthedocs.io/) by retrieving annotated protein-coding genes from the OryCun2.0 NCBI's GFF file (https://www.ncbi.nlm.nih.gov/).

Results

In the pairwise F_{ST} comparison, the averaged F_{ST} values ranged from 0.095 (Italian White vs all) to 0.238 (Ermine *vs* all). In total, 309 unique genome regions that contain signatures of selection were obtained from F_{ST} analysis (single breed approach and group of breeds approach). Several genome regions harbouring coat colour related genes already detected by other studies to affect pigmentation in rabbits or other species were identified. These regions included *TYR* (Burgundy-Fawn, Italian White, Italian Spotted), *OCA2* (Checkered Giant), *MC1R* (Rhinelander), *MITF* (Giant White), and *ASIP* (Giant Grey). Other signatures were highlighted in genes related to: (1) coat structure, such as *LIPH* (Rex); and (2) body size, such as (*COL2A1, LCORL, and GRK5)* in Dwarf Lop breed.

ROH islands were identified in a total of 11 autosomes, which covered about 10-20 Mb of the rabbit genome. Considering overlapping ROH islands between breeds, a total of 22 independent ROH islands were detected. Manhattan plots presenting ROH islands in the four breeds for which more than 50 animals were genotyped are shown in (Figure 1). Several ROH islands contain genes involved in body size, body length, pigmentation process, carcass traits, growth, and reproductive traits as previously determined in humans or in other livestock species such as (*NELL1, TRIB2, NR3C1, TRIB1*, *IFRD1, AOX1, ITGB8, and OCA2*) genes.

Discussion

In this study, different fancy and commercial rabbit breeds were investigated. The methodologies used to perform F_{ST} analysis were able to capture signatures of selection in many genomic regions harbouring genes that were previously identified to affect coat colour (*ASIP*, *TYR,* and *MC1R*), coat structure (*LIPH*), and body size (*LCORL/NCAPG*) in rabbits. Furthermore, signatures of selection were detected in genome windows containing *EDNRB*, *EDNRA*, and *MITF,* that were not reported thus far to affect coat colour in rabbits, despite the role that those genes play in pigmentation. Interestingly, the two independent types of analysis (ROH islands and F_{ST}) pointed out the *OCA2* genome region in Checkered Giant breed. This gene encodes for the homolog of the mouse p (pink-eyed dilution) that is believed to be a melanocyte-specific transporter with an essential role in normal pigmentation. Other studies could be useful including additional breeds and using other statistical approaches will be needed to expand the detection of signatures of selection in rabbit genome.

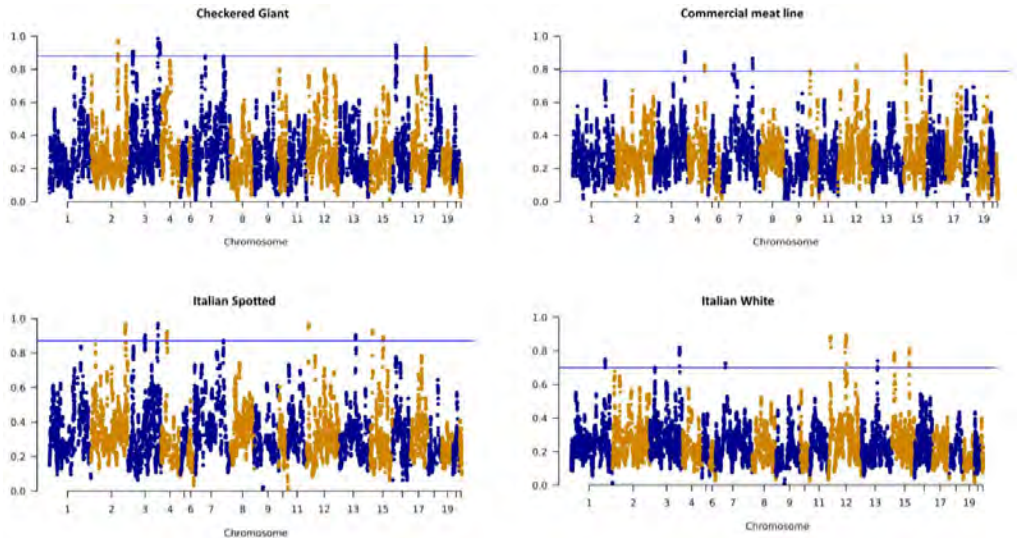


Figure 1. Manhattan plots showing ROH islands in Checkered Giant, commercial meat, Italian Spotted, and Italian White breeds. The blue line indicates the frequency corresponding to the top 1% of the most frequent SNPs in the population.

References

Bertolini F., Schiavo G., Scotti E., Ribani A., Martelli P.L., *et al.* (2014) Anim Genet 45:304-307. https://doi.org/10.1111/age.12121.

Bhatia G., Patterson N., Sankararaman S., Price A.L. (2013) Genome Res 23:1514-1521. https://doi.org/10.1101/gr.154831.113.

Bovo S., Schiavo G., Utzeri V.J., Ribani A., Schiavitto M., *et al.* (2021) Anim Genet 52:237-243. https://doi.org/10.1111/age.13036.

Carneiro M., Rubin C.J., Di Palma F., Albert F.W., Alföldi J., *et al.* (2014) Science 345:1074-1079. https://doi.org/10.1126/science.1253714.

Chang C.C., Chow C.C., Tellier L.C., Vattikuti S., Purcell S.M., *et al.* (2015) GigaScience 4.1:s13742-015. https://doi.org/10.1186/s13742-015-0047-8.

Demars J., Iannuccelli N., Utzeri V.J., Auvinet G., Riquet J., *et al.* (2018) Genes 9:E430. https://doi.org/10.3390/genes9090430.

Fontanesi L., Tazzoli M., Beretti F., Russo V. (2006) Anim Genet 37:489-493. https://doi.org/10.1111/j.1365-2052.2006.01494.x.

Fontanesi L., Forestier L., Allain D., Scotti E., Beretti F., *et al.* (2010) Genomics 95:166-175. https://doi.org/10.1016/j.ygeno.2009.11.003.

Fontanesi L., Scotti E., Allain D., Dall'Olio S. (2014a) Anim Genet 45:248-255. https://doi.org/10.1111/age.12104.

Fontanesi L., Vargiolu M., Scotti E., Latorre R., *et al.* (2014b) PLoS One 9:e93750. https://doi.org/10.1371/journal.pone.0093750.

Laghouaouta H., Sosa-Madrid B.S., Zubiri-Gaitán A., Hernández P., *et al.* (2020) Animals 10:2090. https://doi.org/10.3390/ani10112090.

Rubin C.J., Megens H.J., Martinez Barrio A., Maqbool K., Sayyab S., *et al.* (2012) Proc Natl Acad Sci U S A 109:19529-36. https://doi.org/10.1073/pnas.1217149109.

Sosa-Madrid B.S., Hernández P., Blasco A., Haley C.S., Fontanesi L., *et al.* (2020a) Anim Genet 51:58-69. https://doi.org/10.1111/age.12873.

Schiavo G., Bovo S., Muñoz M., Ribani A., Alves E., *et al.* (2021) *Animal Genetics*, *52*(2): 155-170.https://doi.org/10.1111/age.13045.

Sosa-Madrid B.S., Varona L., Blasco A., Hernández P., Casto-Rebollo C., *et al.* (2020b) Animal 14:2225-2235. https://doi.org/10.1017/S1751731120001263.

Sosa-Madrid B.S., Santacreu M.A., Blasco A., Fontanesi L., Pena R.N., *et al.* (2020c) J. Anim. Breed. Genet., early online https://doi.org/10.1111/jbg.12451. https://doi.org/10.1111/jbg.12451.

Utzeri V.J., Ribani A., Fontanesi L. (2014) Anim Genet 45:600-603. https://doi.org/10.1111/age.12171.

236. Investigating genetic redundancy as a source of genetic diversity and adaptability in the U.S. Holstein breed

Y. Steyn[1*], T.J. Lawlor[2], S. Tsuruta[1], Y. Masuda[1], A. Legarra[3], D.A.L. Lourenco[1] and I. Misztal[1]

[1]University of Georgia, 425 River Road, Athens, GA 30605, USA; [2]Holstein Association USA Inc., 1 Holstein PI, Brattleboro, VT 05301, USA; [3]INRA, UMR1388 GenPhySE, 31326 Castanet-Tolosam, France; yvette.steyn@uga.edu

Abstract

The continuous increase in inbreeding in the U.S. Holstein is a threat to genetic diversity. Loss of genetic diversity can hinder adaptation to changing environments and consumer demands. K-means clustering on the genomic relationship matrix was used to stratify 20,099 selected candidates into five clusters. These clusters were shown to be different through expected inbreeding within- and across cluster, the fixation index (F_{st}), and genetic correlations for stature between clusters. Expected inbreeding within cluster was mostly higher than across-cluster. The average F_{st} across cluster was 0.03. Genetic correlations ranged from <0.50 to >0.90. Families were formed by tracing pedigrees of each cluster back for 10 generations. Allele frequency changes over generations were evaluated. Polygenic shifts, selective sweeps, hitch-hiking, and epistasis were observed. The Replicate Frequency Spectrum was used to measure genetic redundancy. Results show that sub-populations within the breed have developed differently over time.

Introduction

The U.S. Holstein is generally viewed as a large, homogenous population known for its extensive use of only a few sires. This high level of relatedness among animals is a concern for genetic improvement and overall health of the population. Lack of variation can hinder the ability of populations to adapt to change (Markert *et al.*, 2010), which is a growing concern in the face of climate change and consumer preferences. While trait variation may reduce with selection, genetic variation can still exist. There are two main patterns in which allele frequency (AF) change occurs in a population, namely selective sweeps and polygenic shifts. Genetic hitchhiking can also increase the frequency of markers surrounding the gene of importance simply due to its proximity to the gene. Most traits of economic importance are highly polygenic. These alleles are more likely to undergo small polygenic shifts. Genetic redundancy is a phenomenon where an excess of beneficial variants exists to achieve the same genetic merit (Goldstein and Holsinger, 1992). Thus, different sub-populations may undergo heterogenous changes to converge to the same genotype or phenotype. The objectives of this study were to identify sub-populations within the breed and observe different AF changes over time.

Materials & methods

A total of 20,099 genotyped Holstein animals were used as selected candidates. This includes 3,902 males that were sires of animals born after 2010, with at least 25 progeny, and 16,197 females with measurements in 2013 or 2014. K-means clustering with a built-in R package was applied on the genomic relationship matrix (G) to separate the population into five separate clusters (C1 to C5). Alleles were coded as 0, 1, and 2.

Differences across cluster. Expected inbreeding within or across cluster was calculated based on pedigree information assuming non-zero inbreeding for unknown parents (Aguilar and Misztal, 2008) with the INBUPGf90 package within the BLUPF90 software suite (Misztal *et al.*, 2014). The F_{st} across clusters was calculated as in Bonhomme *et al.* (2010). The genetic correlations across clusters were estimated using stature as trait and an adjusted version of the method by Duenk *et al.* (2020). This correlated the additive breeding value of a cluster using the SNP effects of the same cluster, or when using the SNP effects of another

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_236, © Y. Steyn *et al.* 2022

cluster. Female animals of each cluster were used to estimate SNP effects with the POSTGSF90 package (Misztal *et al.*, 2014) and applied on the males of each cluster. A simple correlation will underestimate the genetic correlation since the accuracy of breeding values is not considered. Therefore, the correlations were multiplied by an adjustment factor using the method by Calo *et al.* (1973). The heritability was considered as a measure of accuracy. Assuming a heritability of 0.45 in all clusters, this adjustment factor is 1/0.45.

Allele frequency changes. Within each cluster, five families were formed (F1 to F5) by tracing pedigrees back for 10 generations (G10 being most recent and G0 earliest). The AF were calculated for each generation within family. Generations overlap and families are connected by shared ancestors, thus animals may appear in more than one generation within a family, or/and in more than 1 family. In G0, 82% of the animals appear in more than one family and 54% are common to all families. The number of shared ancestors decreases over time. By G6, the number of families in common to all is larger than the number of unique animals per family. The total number of unique animals per family is 4,355 (F1), 3,555 (F2), 3,574 (F3). 7,107 (F4), and 4,870 (F5). Specific SNP were identified for visual observation based on previous literature (Ma *et al.*, 2019), the variance of absolute change from G0 to G10 across families, the range between minimum change in a family and maximum change, and the F_{st}. Genetic redundancy was measured using the Replicate Frequency Spectrum (RFS) (Barghi *et al.* 2019). This looks at the number of times the 100 SNP with the greatest change in one family, change by more than 0.30 in each individual family. This reflects non-parallel changes over time.

Results

The inbreeding within each cluster averaged 0.22 for C1, 0.20 for C2, 0.18 for C3, 0.10 for C4, and 0.17 for C5. The average across-cluster inbreeding was 0.11 for C1, 0.11 for C2, 0.12 for C3, 0.10 for C4, and 0.11 for C5. The average F_{st} for markers across clusters was 0.03. The genetic correlations adjusted for heritability are presented in Table 1. The changes in AF over time within each family for the SNP that showed the greatest variance is presented in Figure 1. Table 2 presents the RFS. When the 100 SNP with the greatest change in F1 are observed in each family, 15 SNP change more than 0.30 only in F1, none only in one other family, and 14 in all families. Among the 100 SNP with the greatest change in F2, 15 change more than 0.30 only in F2, none only in one other family, and 16 in all families. Of the 100 SNP with the greatest change in F3, 12 change by more than 0.30 only in F3, none only in one other family, and 13 in all families. When the 100 SNP with the greatest change in F4 are chosen, 3 change by 0.30 only in F4, 3 only in F3, 4 only in F2, and 19 in all. Among the 100 SNP with the greatest change in F5, 18 change by more than 0.30 only in F5, none only in one other family, and 11 in all families. Change was not restricted to be in the same direction in all families.

Table 1. Genetic correlations across clusters, measured as the adjusted correlation between the additive breeding value of animals of each cluster when expressed in the genetic background of each cluster.

Genetic background	Cluster expressed				
	C1	C2	C3	C4	C5
C1	1.00	0.63	0.99	0.76	0.93
C2	0.77	1.00	0.62	0.59	0.51
C3	0.81	0.41	1.00	0.81	0.92
C4	0.80	0.74	0.93	1.00	0.90
C5	0.89	0.45	0.89	0.73	1.00

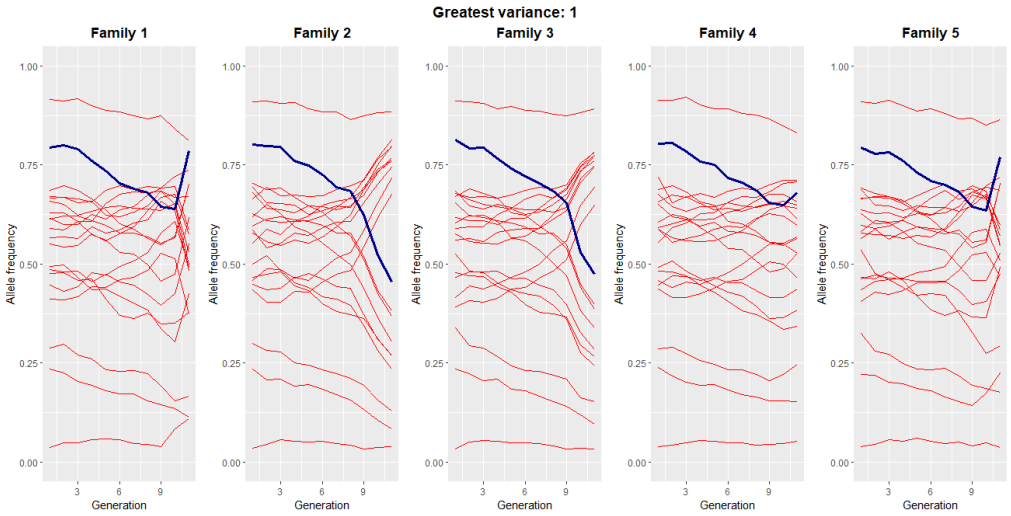


Figure 1. The allele frequency of the SNP that showed the greatest variance in change across families (blue) and the surrounding 20 SNP markers (red) per generation within each family.

Table 2. The Replicate Frequency Spectrum: The number of times the 100 SNP with the greatest change in a specific family (each row), change allele frequency (AF) by more than 0.30 in each family.

Family with top 100 SNP	Number of SNP with AF change >0.30				
	F1	F2	F3	F4	F4
F1	100	45	61	26	59
F2	57	100	68	25	51
F3	60	61	100	31	46
F4	68	67	78	55	60
F5	50	57	57	20	100

Discussion

The generally higher within-cluster inbreeding compared to across-cluster inbreeding, is an indication of genetic differences across clusters. Consistently lower inbreeding occurs for C4, regardless of whether mating was across- or within- cluster. This indicates that, although the clusters are separate from each other, C4 still contains enough variation to allow lower inbreeding. The average F_{st} of 0.03 is higher than what would be expected if no differences exist, and about half of that found between three different dairy breeds in a French study (Flori *et al.*, 2009).

There are two genetic correlations between families – one where the genotype of population 1 is expressed in the genetic background of population 2, and one where population 2 is expressed in the genetic background of population 1. These differ due to the AF in different populations. Although the genetic correlations between some clusters were above 0.85, correlations as low as 0.41 were observed, indicating that clusters show some genetic differences. In general, C2 has the lowest correlations with other clusters.

Some observed SNP showed non-parallel changes across families based on our criteria for selection, indicating genetic redundancy. Figure 1 shows a decrease in AF of the selected SNP (in blue) in F2 and F3

with no change in direction. Relatively small changes in AF are observed in F4, while directional changes in AF occur in F1 and F5. Observed SNP surrounding the selected markers (in red) show signs of potential partial selective sweeps and hitchhiking. Hitchhiking in opposite directions can be achieved when markers are in different phases. Directional changes could be due to a change in selection pressure, or epistasis with a different gene under selection.

The presence of genetic redundancy is further shown with the RFS. Even though many SNP change similarly in all families, there are family-specific SNP that change by more than 0.30 in only one family. While this study shows genetic redundancy, it does not determine whether specific changes were due to selection or drift. Gene-dropping computer simulations could be conducted to identify markers that have changed more than what could be expected under neutral inheritance. Such methods could be applied to identify genes under selection.

Conclusions

The U.S. Holstein breed is a complex mixture of family subgroups with different allelic frequencies and genotypic combinations. The study reveals genetic redundancy that has allowed sub-populations to change differently over time. Genetic redundancy has the ability to continue genetic improvement in the breed without drastically decreasing underlying genetic diversity. The different polygenic shifts in sub-groups avoids the fixation of alleles and the loss of genes that may allow the breed to adapt to climate change and consumer demand.

References

Aguilar, I., and Misztal, I. (2008) J Dairy Sci 91(4):1669-1672. https://doi.org/10.3168/jds.2007-0575

Barghi, N, Tobler, R., Nolte, V., Jaksic, A.M., Mallard, F., *et al.* (2019) PLoS Biol. 17(2):e3000128. https://doi.org/10.1371/journal.pbio.3000128

Bonhomme, M., Chevalet, C., Servin, B., Boitard, S., Abdallah, J. *et al.* (2010) Genetics 186(1):241-262 https://doi.org/10.1534/genetics.110.117275

Calo, L.L., McDowell, R.E., VanVleck, L.D., and Miller, P.D. (1973) J Anim Sci. 37(3):676-682. https://doi.org/10.2527/jas1973.373676x

Duenk, P., Bijma, P., Calus, M.P.L, Wientjes, Y.C.J., and Van der Werf, J.H.J (2020) G3 10(2):783-795 https://doi.org/10.1534/g3.119.400663

Goldstein, D.B., and Holsinger, K.E. (1992) Evolution 46(2):412-429 https://doi.org/10.1111/j.1558-5646.1992.tb02048.x

Flori, L., Fritz, S., Jaffrezic, F., Boussaha, M., Gut, I. *et al.* (2009) PloS One 4(8):e6595 https://doi.org/10.1371/journal.pone.0006595

Ma, L., Sonstegard, T.S., Cole, J.B, VanTassell, C.P., Wiggans, G.R., *et al.* (2019) BMC Genom. 20(1):128 https://doi.org/10.1186/s12864-019-5459-x

Markert, J.A., Champlin, D.M., Gutjahr-Gobell, R., Grear, J.S., Kuhn, A. *et al.* (2010) BMC Evol. Biol. 10(1):205 https://doi.org/10.1186/1471-2148-10-205

Misztal, I., Tsuruta, S., Lourenco, D.A.L., Aguilar, I., Legarra, A. *et al.* (2014) Manual for BLUPF90 family of programs. Available at: http://nce.ads.uga.edu/wiki/lib/exe/fetch.php?media=blupf90_all2.pdf

237. Inbreeding depression and heterosis for live-weight and carcass traits in purebred and crossbred sheep

P. Rafter[1*], N. McHugh[1], K. McDermott[2] and, D.P. Berry[1]

[1]*Animal & Grassland Research and Innovation Centre, Teagasc, Moorepark, Fermoy, Cork, Ireland;*
[2]*Sheep Ireland, Highfield House, Shinagh, Bandon, Cork, Ireland; pierce.rafter@teagasc.ie*

Abstract

Given the importance of inbreeding depression and heterosis, both should be accounted for in breeding and management strategies. The present study quantified the association between eight different pedigree- and genome-based inbreeding and outbreeding metrics with 40-day weight, weaning weight, carcass weight, carcass conformation, and carcass fat in a large multi-breed and crossbred population of sheep. Five of the inbreeding and outbreeding metrics tested were associated with at least one of the traits analysed. Of the five traits analysed, only 40-day weight was not associated with any of the inbreeding or outbreeding metrics used in the present study, although one of the inbreeding metrics for the dam was positively associated with 40-day weight of the lamb.

Introduction

Inbreeding depression is the reduction in progeny performance relative to the parental mean performance which occurs due to the close relationship between the sire and the dam (Falconer and Mackay, 1996). Heterosis (hybrid vigour) can be considered the opposite of inbreeding depression, in that it is an increase in progeny performance relative to the parental mean performance that occurs when the sire and dam originate from different populations (Falconer and Mackay, 1996). The purpose of the present study was to evaluate the association between different estimates of inbreeding and outbreeding with 40-day weight, weaning weight, carcass weight, carcass conformation, and carcass fat of lambs. Both pedigree-based and genome-based measures of inbreeding and outbreeding were considered.

Materials & methods

Genotype and phenotype data. Illumina OvineSNP50 single nucleotide polymorphism (SNP) array (54,241 SNPs) (Illumina Inc. San Diego, CA) genotypes were available for 31,171 sheep with a call rate of >95%. These 31,171 animals consisted of 21,325 crossbred sheep, 3,466 Texel, 2,539 Belclare, 1,511 Charollais, 1,234 Suffolk, and 1,096 Vendeen sheep. Single nucleotide polymorphisms on the Y or MT chromosome, SNPs with a call-rate <95%, SNPs with a GC score <0.55, and SNPs which deviated from Mendelian inheritance in >2% of verified parent-progeny pairs were all excluded. After edits, 47,349 SNPs remained; missing genotypes were imputed using the imputation software suite FindHap v3 (VanRaden *et al.,* 2013).

Five traits were investigated in the present study: 40-day weight (kg), weaning weight (kg), carcass weight (kg), carcass conformation, and carcass fat. Forty-day weight is the live-weight of the lamb measured at approximately 40 days of age while weaning weight is the live-weight of the lamb measured at weaning at approximately 100 days of age (McHugh *et al.*, 2018). Carcass fat cover was graded by processors on a scale of 1 [thin] to 5 [fat] scale. Carcass conformation was also assessed by processors on a 5-point scale where 1 represents poor conformation and 5 represents excellent conformation. Carcass records were retained for crossbred animals which were ≤300 days old at slaughter. After edits, carcass records were available for 5,872 animals while 40-day weight and weaning weight records were each available for 7,196 animals.

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_237, © P. Rafter *et al.* 2022

Measures of inbreeding and outbreeding. Four different estimates of inbreeding were used: F_{ped} which is the (traditional) pedigree estimate of inbreeding (Wright, 1922), F_{hom} which is a genomic estimate calculated from observed homozygosity (Kardos *et al.*, 2015), F_{ROH} which is the proportion of the genome in a run of homozygosity (ROH) (Kardos *et al.*, 2015), and homozygosity by loci which is a genomic metric that measures the deviation from expected heterozygosity across the genome (Aparicio *et al.*, 2006). Separately, four estimates of outbreeding were also calculated: F_{ROHet} which is the proportion of the genome in a run of heterozygosity (ROHet), observed heterozygosity which is the proportion of heterozygous SNPs in the genome, heterosis which is proportional to the sum of the product of parental breed fractions (VanRaden and Sanders, 2003), and recombination loss which is proportional to the sum of the square of parental breed fractions (VanRaden and Sanders, 2003). For the heterosis and recombination loss calculations, the breed fractions of the sire and the dam were calculated based on the pedigree information.

Runs of homozygosity and ROHet were called in each animal in the population using the R package 'detectRUNs' (Biscarini *et al.*, 2019) using a sliding window approach. The window size for each run was set to 15 SNPs; within each window a maximum of one contradictory genotype was allowed, i.e. for ROH, the contradictory genotype was a heterozygote and for ROHet, the contradictory genotype was a homozygote. The maximum gap between adjacent SNPs within a run was set at 250 kb and the minimum SNP density was one SNP per 100 kb; the minimum number of SNPs per run was five and the minimum length of a run was 1 kb.

Association analyses. The association between each inbreeding and outbreeding statistic with each carcass and live-weight trait was modelled separately for each trait across the entire population using a linear mixed model approach with the software suite ASREML (Gilmour *et al.*, 2015). Pedigree estimates for inbreeding or outbreeding statistics may be biased downwards relative to genomic estimates due to missing pedigree records or violations in the assumption that all founder animals are unrelated. For comparison purposes, the inbreeding and outbreeding statistics were each standardised to have a mean of 0 and a standard deviation of 1. Each of the linear mixed models were solved separately for the five different traits analysed and again separately for each of the eight inbreeding and outbreeding statistics so that overall, 40 different analyses were conducted. Forty-day weight and weaning weight were modelled with a maternal effects model as:

$$y = Xb + Za + Wm + e \quad (1)$$

where y was the vector of phenotypes. The fixed effects were denoted in the model by the matrix b which consisted of, the intercept treated as a continuous variable, the breed category of each animal (which was either crossbred, Texel, Belclare, Charollais, Suffolk, or Vendeen), the inbreeding or outbreeding statistic included in the model which was treated as a continuous variable, the sex of the animal treated as a categorical variable, the age of the animal treated as a continuous variable, an age-by-sex interaction variable, the age of the dam treated as a continuous variable, and the flock-date of each animal was included as a contemporary group in the model. The direct polygenic effect of each animal was denoted by a and was treated as a random effect which was assumed to have the distribution $N(0, \mathbf{A}\sigma^2_g)$, where A was the numerator relationship matrix and σ^2_g was the direct additive genetic variance. The maternal genetic effect, denoted by m, was also treated as a random effect and had the distribution $N(0, \mathbf{A}\sigma^2_m)$ where σ^2_m was the maternal genetic variance. The residual of the model was denoted by e and was treated as a random effect with the distribution $N(0, \mathbf{I}\sigma^2_e)$, where I was the identity matrix and σ^2_e was the residual variance. The matrices X, Z, and W were design matrices for the fixed effects, the direct polygenic random effects, and the maternal genetic effects, respectively. For the analysis of recombination loss and heterosis, purebred animals were not included in the analysis because they have no heterosis or recombination loss.

Each carcass trait was separately modelled using an animal model:

$$y = Xb + Za + e \quad (2)$$

The definitions for variables, y, Z, a, and e were the same as in Equation 1. The fixed effects, denoted by matrix b, were the inbreeding or outbreeding statistic included in the model which was treated as a continuous variable, an age-by-sex interaction variable, and the contemporary group for each animal, which was the flock-by-slaughter date of each animal.

Results & discussion

All full-siblings have, by definition, the same expected pedigree-based estimate of inbreeding because the calculation of the pedigree estimate is based on the pedigree relationship between the sire and the dam (Wright, 1922). Conversely, the genomic estimate of inbreeding between full-siblings can differ because the genomic estimate of inbreeding is calculated per individual based on the observed homozygosity of their genome. In the population used in the present study there were 5,969 full-sibling pairs; the average difference in F_{hom} between full-siblings was 0.0006 (standard deviation of 0.043) and the average difference in F_{ROH} between full-siblings was 0.0005 (standard deviation of 0.015). These results indicate that the genomic inbreeding estimates may have the capacity to detect a truer reflection of inbreeding.

Association analyses. The regression coefficient of each performance trait on each of the eight pedigree and genomic estimates of inbreeding and outbreeding are in Figure 1. Neither homozygosity by loci, heterosis, or F_{ROHet} were associated with any of the traits analysed; additionally, no inbreeding or outbreeding statistic of the lamb itself was associated with 40-day weight. However, while inbreeding or outbreeding statistics relating to the lamb were not associated with 40-day weight (Figure 1), one of the inbreeding

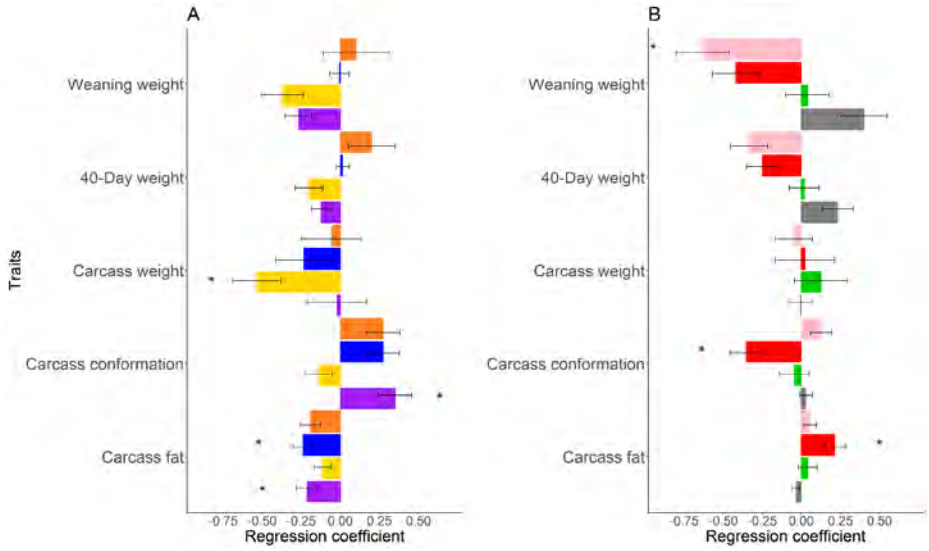


Figure 1. The regression coefficient of (A) inbreeding metrics – F_{hom} (purple), F_{ped} (yellow), F_{ROH} (blue), homozygosity by loci (HL) (orange), and (B) outbreeding metrics -F_{ROHet} (grey), heterosis (Het) (green), observed heterozygosity (OH) (red), and recombination loss (RL) (pink) on carcass fat, carcass conformation, carcass weight, 40-day weight, and weaning weight. The black error bars indicate one standard error. Bars annotated with stars indicate regression coefficients which were different from zero (P<0.001).

metrics (homozygosity by loci) for the dam was positively associated with 40-day weight in the lamb. For weaning weight the strongest association with dam inbreeding (or outbreeding) metrics was also a positive association with homozygosity by loci, although this association was not significant after adjusting for multiple testing. Pre-weaning weight in lambs has a maternal genetic component (Fitzmaurice *et al.*, 2020) which may account for the association between homozygosity of loci in the dam and day-40 weight in the lamb.

Both pedigree-based and genome-based estimates of inbreeding and outbreeding statistics were associated with at least one of the traits analysed. Of the five traits analysed only carcass conformation and carcass fat were both associated with inbreeding and outbreeding metrics (Figure 1). For carcass fat the inbreeding metrics had negative associations whereas the outbreeding metric had a positive association. Conversely, for carcass conformation the inbreeding metric had a positive association and the outbreeding metric had a negative association. These results are consistent with the reported positive genetic correlation between lower carcass fat and higher carcass conformation in sheep (Conington *et al.*, 2001).

Conclusions

Five of the eight estimates of inbreeding and outbreeding were associated with at least one of the live-weight or carcass traits analysed. Therefore, when practicable, pedigree-based and genome-based estimates of inbreeding and outbreeding should be used to account for inbreeding and crossbreeding in sheep populations.

References

Aparicio J.M., Ortego J., and Cordero P.J. (2006) Molecular ecology 15(14):4659-4665. https://doi.org/10.1111/j.1365-294X.2006.03111.x.

Biscarini F., Cozzi P., Gaspa G., and Marras G. (2019) detectRUNS. Available at: https://cran.r-project.org/web/packages/detectRUNS/

Conington J., Bishop S.C., Grundy B., Waterhouse A., and Simm G. (2001). BSAS, 73(3); 413-423. https://doi.org/10.1017/S1357729800058380.

Falconer D.S. and Mackay T.F.C. (1996) Quantitative genetics. Addison Wesley Longman Limited, Essex, UK.

Fitzmaurice S., Conungton J., Fetherstone N., Pabiou T., *et al.* Animal 14(5):899-909. https://doi.org/10.1017/S1751731119002908.

Gilmour A.R., Gogel B.J., Cullis B.R., Welham S.J., and Thompson R. (2015). ASREML user guide. Available at: www.vsni.co

Kardos M., Luikart G., and Allendorf F.W. (2015) Heredity 115:63-72. https://doi.org/10.1038/hdy.2015.17.

McHugh N., Pabiou T., McDermott K., Wall E., and Berry D.P. (2018) JAS, 96(6):2051-2059. https://doi.org/10.1093/jas/sky143.

VanRaden P.M., Null D.J., Sargolzaei M., Wiggans G.R., *et al.* (2013) JDS 96(1):668-678. https://doi.org/10.3168/jds.2012-5702.

VanRaden P.M. and Sanders A.H. (2003) JDS 86(3):1036-1044. https://doi.org/10.3168/jds.S0022-0302(03)73687-X.

Wright S. (1922) Am. nat. 56(645):330-338.

238. Phylogenetic analysis and nucleotide diversity of 69 cattle breeds including German Black Pied cattle using WGS

G.B. Neumann[1*], P. Korkuć[1], D. Arends[1], M.J. Wolf[2], K. May[2], S. König[2] and G.A. Brockmann[1]

[1]Humboldt University of Berlin, Albrecht Daniel Thaer- Institute for Agricultural and Horticultural Sciences, Invalidenstr. 42, 10115 Berlin, Germany; [2]Justus-Liebig-University of Gießen, Institute of Animal Breeding and Genetics, Ludwigstr. 21, 35390 Gießen, Germany; guilherme.neumann@hu-berlin.de

Abstract

Local populations such as the endangered German Black Pied cattle (DSN) are an important genetic resource for the future of livestock breeding. To provide more information on the genome composition of DSN, we performed phylogenetic analysis and calculated genome-wide nucleotide diversity using variants from whole-genome sequencing of 302 DSN, and 1,394 individuals from 68 taurine breeds. Phylogenetic analysis detected four main clusters that could be assigned to their respective geographical origin: Northern Europe, Central Europe, Jersey and Guernsey islands, and an area comprising Eastern Europe, Central Italy, and Asia. DSN showed close relationship to Dutch Friesian Red, Dutch Belted, Holstein, Eastern Flanders White Red, and Kholmogory. Nucleotide diversity was medium in DSN (0.001528) while it was highest in breeds from Asia and Eastern Europe (>0.0017). Nucleotide diversity was particularly high on BTA 23 between 19 and 37 Mb where the bovine major histocompatibility complex and additional immune genes are located.

Introduction

Local indigenous cattle breeds are a crucial source of genetic diversity for livestock breeding, which might be important for future local adaptations. According to the Food and Agriculture Organization of the United Nations (FAO), 84% of all local breeds in Europe are considered at risk (FAO, 2021). This is also the case for the dual-purpose German Black Pied cattle population (DSN, 'Deutsches Schwarzbuntes Niederungsrind'). Its initial breeding in the North Sea region of Germany and the Netherlands dates back to the 18th century. From there, animals were brought by emigrants to North America where they contributed to the breeding program responsible for the generation of the modern high-yielding dairy breed Holstein (Brade and Brade, 2013). Because of its lower milk yield, DSN has been almost entirely replaced by Holstein. Today, only about 2,500 DSN herd book cows remain. The current breeding goal for DSN aims to preserve DSN-typical and beneficial characteristics. For an effective conservation plan, genetic diversity measurements are needed to guarantee an optimised pool of genetic variants that provides sufficient adaptation capacity to changing environments and prevents inbreeding depression (Kristensen *et al.*, 2015). In this study, we compared the genetic relationship of 69 taurine cattle breeds including DSN using phylogenetic analysis and nucleotide diversity (π). Nucleotide diversity has been shown to be a good start point when evaluating the genetic diversity status of a given population (Kardos *et al.*, 2021). It is a region-wide metric used to quantify the degree of polymorphisms within a population.

Methods

Genomic data. Sequence variants were available from whole-genome sequencing data of 302 DSN animals (Neumann *et al.*, 2021). Further sequence variants of 1,394 animals from 68 other *Bos taurus* breeds and 1 Auroch (*Bos primigenius*) were obtained from the 1000 Bull Genomes Project – Run 9 (Hayes and Daetwyler, 2018). The sequence variants were detected based on the ARS-UCD1.2 reference genome. For high data quality, we required a sequencing coverage of at least 8-fold and at least 5 animals per breed. For breeds with >30 animals, 30 animals were randomly selected. For Holstein, 30 animals per country (Germany, Denmark, Netherlands, U.S.) and 30 labelled as 'Holstein Friesian', and not only 'Holstein'

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_238, © G.B. Neumann *et al.* 2022

in the 1000 Bull Genomes Project were randomly selected. The sequence variants from 302 sequenced DSN and from the 1000 Bull Genomes Project were merged using BCFtools v1.9. A total of 74,834,092 biallelic variants (68,486,380 SNPs and 6,347,712 indels) occurring in the tranche 99% (from the Variant Recalibration performed by 1000 Bulls Genome Project) with a call rate ≥90% were considered for the analysis.

Phylogenetic analysis. A genome-wide phylogenetic tree was built for all *Bos taurus* autosomes (BTA). The 74,834,092 sequence variants were first pruned removing all variants in high linkage (r^2>0.6) with PLINK v1.9 (Purcell *et al.* 2007) using a windows-size of 50 bp and a step-size of 5 bp. Using the resulting 21,426,106 variants, the Manhattan distance between animals was calculated and the UPGMA algorithm implemented in the biotite library in Python was used for clustering. The tree was rooted using Auroch as an outgroup. The phylogenetic tree was visualized using iTOL v6 (Letunic *et al.* 2019). Branches of the phylogenetic tree represent the position of the majority of animals from each breed.

Genomic diversity. Nucleotide diversity (π) was calculated per windows of 10 kb (π_{window}) for all BTAs using 74,834,092 sequence variants and the package scikit-allel in Python as:

$$\pi_{window} = \frac{\frac{n}{n-1}\sum_{ij} p_i p_j \pi_{ij}}{10\ kb} \quad (1)$$

Where p is the frequency of the respective i^{th} and j^{th} haplotypes, π^{th} is the number of nucleotide differences per site between the i^{th} and j^{th} haplotypes, and n is the number of haplotypes. The average nucleotide diversity per chromosome (π_{chrom}) was calculated as the mean over all windows of the respective chromosomes, and total nucleotide diversity (π_{total}) as the mean across all BTAs.

Results

Phylogenetic analysis. Phylogenetic analysis of 69 cattle identified four main clusters that can be assigned to geographical origins (Figure 1a): Northern Europe (green), Central Europe (violet), Jersey and Guernsey islands (red), and Eastern Europe, Central Italy, and Asia (blue). The cluster of Central Europe comprised breeds from Austria, Switzerland, Southern Germany and France. Those are mainly dual-purpose breeds kept in mountainous areas. However, the majority of the breeds including DSN and Holstein formed a cluster based on their origin in Northern European countries. Jersey and Guernsey cattle formed a separate cluster. The fourth cluster included breeds from Asian, Central Italy, and Eastern European countries. Most breeds showed clear clusters without or with very few misassignments of animals. This was not the case for the Modern Danish Red, Swedish Red, Norwegian Red, Ayrshire Finish, and Modern Angler which showed many misassignments. Regarding DSN, we observed that they clustered closest to Dutch Friesian Red, Dutch Belted, Holstein, Eastern Flanders White Red, and Kholmogory. Those are all breeds originating from Germany, the Netherlands, and Belgium. Kholmogory, a Russian breed, is descendent of animals imported from the Netherlands. Further breeds from the North Sea region, from territories of Germany, the Netherlands, Denmark, Belgium and Scandinavia showed also close relationship to DSN.

Genomic diversity. Total nucleotide diversity (π_{total}) for the 69 investigated breeds ranged between 0.001316 in Väneko and 0.002048 in Buryat (Figure 1a). π_{total} was higher in DSN (0.001541) than in Holstein (0.001505) and 28 other breeds (20 from Northern Europe), but lower than in other European and Asian breeds. The highest π_{total} values were detected in Asian and Eastern European breeds. The other European breeds showed similar π_{total} with a few lower exceptions (<0.0014): Ringamålako (0.001396), Jersey (0.001396), Guernsey (0.001385), Groningen White Headed (0.001345), Traditional Danish Red (0.001339), Scottish Highland (0.001328) and Väneko (0.001316). Regions with particular high nucleotide diversity (π_{window}) in DSN as well as in all other breeds were observed on BTA 4, 10, 12, 15, 29 (Figure 1b),

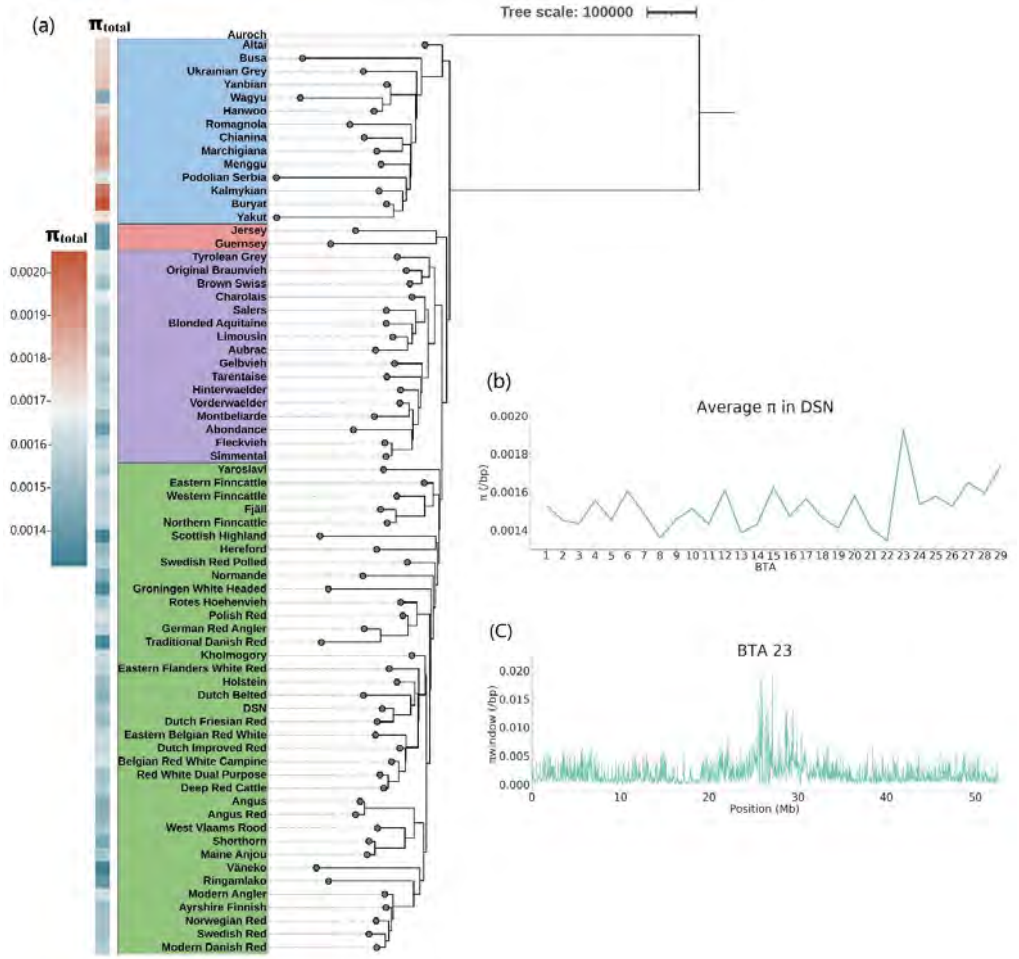


Figure 1: (a) Phylogenetic tree of 69 cattle breeds. Clusters are coloured by their geographical assignment: Central Italy, Eastern European and Asian countries (blue), Jersey and Guernsey (red), Central Europe (violet), Northern European countries (green). Individuals of each breed were collapsed to single clusters (grey spheres). Nucleotide diversity (π_{total}) is listed next to the leaves as gradient (blue=low, red=high). (b) Average nucleotide diversity in DSN per chromosome. (c) Nucleotide diversity in DSN per window on BTA 23.

and in particular on BTA23 between 19 and 37 Mb (Figure 1c), which includes the major histocompatibility complex (MHC) and the bovine leukocyte antigen (BoLA).

Discussion

DSN showed similar nucleotide diversity as the other breeds from the same cluster, indicating a good management of the endangered population. As expected, in DSN and all other breeds, BTA 23 contains a highly polymorphic region where the MHC is located. This high diversity around MHC is also seen in other species (Shiina *et al.*, 2017). Asian and Eastern European breeds showed higher nucleotide diversity (π_{total}) likely due to less intensive breeding programs, genetic bottlenecks, or founder effects in comparison to Central and Northern European breeds (Felius *et al.*, 2014). Further migration of these cattle to Southern, Central and finally Northern Europe are consistent with the clusters in our phylogenetic tree (Felius *et*

al., 2014). The lowest nucleotide diversity of Northern European Red cattle populations was evident in Traditional Danish Red and Groningen White Headed, which resulted from recent inbreeding in the small populations (Schmidtmann *et al.*, 2021). Recently, Schmidtmann *et al.* (2021) have found high levels of admixture between Modern Angler, Ayrshire Finnish, and Norwegian Red, which explains why our clustering between those breeds had many misassignments.

References

Brade, W. and Brade, E. (2013) Berichte über Landwirtschaft 91(2).

Food and Agriculture Organization of the United Nations (FAO). Risk status of livestock breeds. Available at: https://www.fao.org/sustainable-development-goals/indicators/252/en/

Felius, M., Beerling, M.L., Buchanan, D.S., Theunissen, B., *et al.* (2014) Diversity. 6(4):705-750. https://doi.org/10.3390/d6040705

Hayes, B.J., and Daetwyler, H.D. (2018) Annu. Rev. Anim. Biosci. 2019. 7:89–102. https://doi.org/10.1146/annurev-animal-020518

Kardos, M., Armstrong, E.E., Fitzpatrick, S.W., Hauser, S., Hedrick, P.W., *et al.* (2021) PNAS. 118(48). https://doi.org/10.1073/pnas.2104642118

Kristensen, T.N., Hoffmann, A.A., Pertoldi, C., Stronen, A. V. (2015) Front. Genet. 6(38). https://doi.org/10.3389/fgene.2015.00038

Letunic, I., Bork, P. (2019). Web Serv. issue Publ. 47. https://doi.org/10.1093/nar/gkz239

Neumann, G.B., Korkuć, P., Arends, D., Wolf, M.J., May, K., *et al.* (2021) BMC Genomics 22. https://doi.org/10.1186/S12864-021-08237-2

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A.R., *et al.* (2007). Am. J. Hum. Genet. 81, 559–575. https://doi.org/10.1086/519795

Schmidtmann, C., Schönherz, A., Guldbrandtsen, B., Marjanovic, J., Calus, M.,et al. (2021) GSE 53. https://doi.org/10.1186/s12711-021-00613-6

Shiina, T., Blancher, A., Inoko, H., Kulski, J.K., 2017. Immunology 150, 127–138. https://doi.org/10.1111/IMM.12624

239. Partitioning of autozygosity in different age-based classes in cattle populations with different demographic histories

T. Druet[1*], L. Flori[2] and M. Gautier[3]

[1]Unit of Animal Genomics, GIGA-R, University of Liège, 4000 Liège, Belgium; [2]SELMET, INRAE, CIRAD, L'Institut Agro, Université de Montpellier, 34988 Montpellier, France; [3]CBGP, INRAE, CIRAD, IRD, L'Institut Agro, Université de Montpellier, 34988 Montpellier, France; tom.druet@uliege.be

Abstract

Classification of homozygous-by-descent (HBD) segments or ROH in different age-related groups based on their length is useful for several applications. Here we compare the partitioning obtained with our newly developed multiple-HBD classes hidden Markov model (HMM) with an alternative approach consisting of length-based clustering of ROH. When the HMM was applied to simulated data sets, autozygosity was concentrated in the HBD-class associated with the ancestors causing inbreeding. This was however not observed for the ROH-classes corresponding to the expected length of ROH transmitted from these ancestors. On real cattle data, autozygosity was maximized with the HMM in the HBD-class matching the inferred period of reduced N_e. In summary, while ROH-based classifications must be interpreted cautiously, our newly developed multiple HBD-classes HMM seems promising to provide a better picture of the age-based partitioning of individual genomic inbreeding.

Introduction

Inbreeding is common in cattle populations as a consequence of intense selection of elite sires or due to their recent demographic history often characterized by declining effective population sizes or recent bottlenecks. Homozygous-by-descent (HBD) segments, genomic segments inherited twice and through different paths from a common ancestor, are present at high frequency in such inbred populations. HBD segments result in long stretches of homozygous genotypes referred to as run-of-homozygosity (ROH), the length of which being a function of the number of generations to the common ancestor. The distribution of HBD segment lengths is thus informative about the past demographic history of a population and is often summarized, in empirical studies, by grouping observed ROH in several discrete length classes; long/short ROH-classes being considered as resulting from recent/ancient inbreeding. Such classifications are valuable to understand past demographic events from the population, to monitor evolution of inbreeding levels, to infer the mating structure when pedigrees are unknown, to determine how deleterious variants are purged across generations, to compare the effects of recent and ancient inbreeding on phenotypes or even to estimate mutation rates. As an alternative approach, we recently developed an HMM with multiple-HBD classes to partition autozygosity in different age-related HBD classes (Druet and Gautier, 2017). We herein assess the accuracy and robustness of such partitioning of autozygosity based on simulated data sets and genotyping data available for three cattle populations with contrasted recent demographic histories, and compare these classifications with those obtained with ROH-based approaches.

Materials & methods

Classification approaches. The joint identification and classification of HBD segments was first realized with RZooRoH (Bertrand *et al.*, 2019) and a model with 12 HBD classes. In this multiple HBD-classes model, several nested layers of ancestors are successively modelled (Druet and Gautier, 2021). In each layer, the genome is described as a mosaic of HBD and non-HBD segments with an HMM. The length of HBD and non-HBD segments is exponentially distributed with a rate R_c specific to the layer c and related to the number of generations to the common ancestors (approximately $0.5 \times R_c$ generations). The frequency of HBD segments within layer c is defined by a parameter ρ_c, that can be interpreted as the inbreeding

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_239, © T. Druet *et al.* 2022

coefficient accumulated over the generations included in the layer. The non-HBD classes are modelled as a mixture of HBD classes from earlier generations and shorter non-HBD segments. Genotype probabilities in different states (HBD / non-HBD) are obtained from the allele frequencies and the genotyping error rates.

A ROH-based approach was additionally used. The clustering of ROH was carried out with GARLIC (Szpiech *et al.*, 2017) that automatically selects the optimal number of SNPs per window. GARLIC was also ran with 12 classes by selecting ROH boundaries matching HBD-classes. More precisely, the boundary between ROH class c-1 and c corresponded to the ROH-length for which the exponential distributions with rates R_{c-1} and R_c have equal probability.

Simulated data sets. To evaluate the classification approaches, we simulated 500 genomes as mosaics of HBD and non-HBD segments using different values of ρ, corresponding to the inbreeding coefficient *F*. Their length was exponentially distributed with a rate R_c. The genome consisted of 25 chromosomes 100 cM long, and individuals were genotyped for a total of 25,000 SNPs. We subsequently used Argon (Palamara, 2016) to simulate more realistic data under a Wright-Fisher model. In these simulations, bottlenecks occurred 16 or 64 generations ago and effective population size (N_e) dropped to 20 or 50.

Real cattle data sets. Identification and classification of HBD segments or ROH was applied to three cattle data sets with distinct recent demographic history (i.e. decline, severe bottleneck and expansion, respectively): (1) 145 Dutch Holstein (HOL) individuals (Alemu *et al.*, 2021); (2) 18 individuals from the feral cattle population of Amsterdam Island (TAF); and (3) 22 Zebus from Magadascar (ZMA) (Magnier *et al.*, 2021). After quality filtering, the samples were respectively genotyped for 37,675, 23,679 and 531,967 SNPs. Past N_e was first estimated with GONE (Santiago *et al.*, 2020).

Results

The expected length for a HBD segment associated with a common ancestor present G generations ago is L=100/(2G) cM, G is thus often estimated as 100/(2L) for a ROH of length L. However, we observed that classes with L>100/(2L) had larger contributions to autozygosity, more so when simulated inbreeding levels were high (Figure 1). Conversely, with the HMM, the HBD-class with a rate R_c equal to 2G had the main contribution to autozygosity. Similar patterns were observed for simulations with different values of R_c.

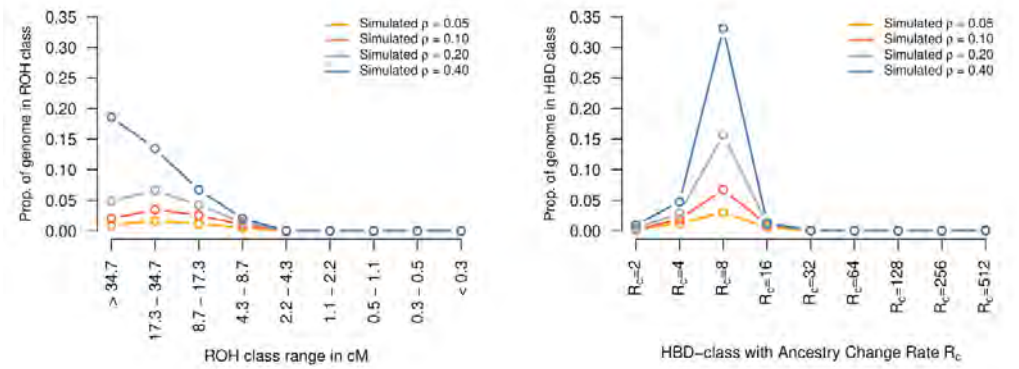


Figure 1. Classification of simulated HBD segments inherited from ancestors living 4 generations in the past (i.e. with an expected length L=12.5 cM) with an ROH approach and with the multiple HBD-classes HMM (expected R_c=8).

The behaviour of the two approaches was confirmed when populations that experienced a bottleneck were simulated under a Wright-Fisher model. With the HMM based approach, the HBD-class with a rate R_c equal to 2G (where G is the timing in generations of the bottleneck), and its neighbours, captured the HBD-segments associated to the bottleneck. On the other hand, with the ROH-based approach, ROH longer than 100/(2G) had the highest contributions to autozygosity, more so when the bottleneck was stronger and inbreeding levels were higher.

This trend was also confirmed when analysing real data on three cattle populations with very different recent histories (HOL, TAF and ZMA) inferred with GONE and which were found consistent with our prior expectations (Figure 2): (1) HOL presented a recent decline with current N_e being equal to 70; (2) TAF experienced a recent and severe bottleneck (17 to 25 generations ago, N_e dropped to 5) in agreement with the historically reported introduction in the Amsterdam island of six founder individuals in 1861 followed by a rapid expansion of the population (up to Ne=1,300); and (3) ZMA is in expansion (100 generations ago, N_e=100 up to a current Ne=3,000). The partitioning with the multiple HBD-classes model closely matches these demographic events (Figure 2). For instance, autozygosity is mainly associated with ancestors living around 16 generations in the past for TAF population, and around 100 generations in ZMA, just before the expansion. In HOL, inbreeding levels were associated to more recent ancestors. The inbreeding levels estimated with GARLIC were highly correlated with those from ZooRoH (0.963, 0.938 and 0.996 for the three populations) although values were lower, indicating that the smaller HBD segments were not captured. By decreasing the minimal ROH size to 20 SNPs with GARLIC, inbreeding levels were closer to those obtained with ZooRoH (although this is a rather small number of SNPs and the optimal values estimated by GARLIC were higher). As for the analyses on simulated data sets, the ROH-based approach associated most of autozygosity to classes with L>100/R_c (R_c is the rate of HBD-classes with the main contribution to autozygosity). Interestingly, for lower inbreeding levels and more recent ancestors, classifications from both approaches were more concordant.

Discussion

In ROH-based approaches, the number of generations to the common ancestor of a given ROH is often estimated as G=100/(2L), based on the expected length of a ROH inherited from an ancestor living G generations ago. ROH are often classified according to their length and G is then estimated independently for each class. In simulations with ancestors present G generations ago, we observed that classes with

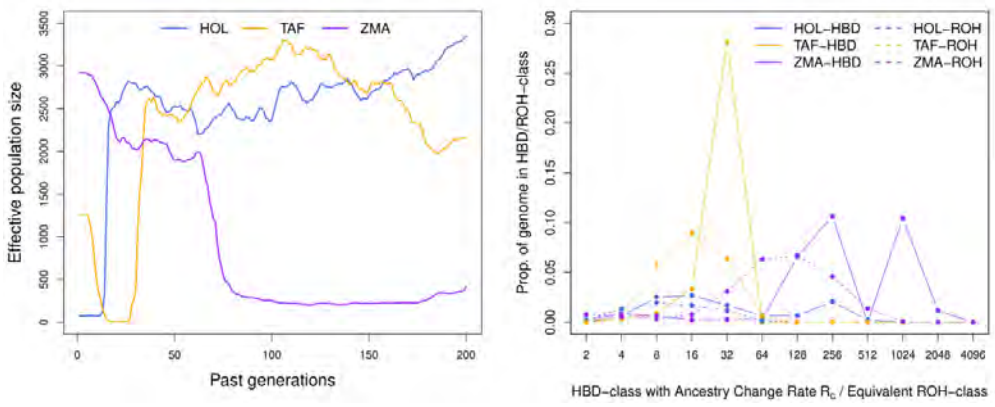


Figure 2. Past effective population size and partitioning in different HBD or ROH-classes for cattle populations with distinct demographic histories.

L>100/2G have larger contributions than classes with smaller ROH, leading to the incorrect interpretation that inbreeding might be associated with more recent ancestors. This result from the fact that ROH inherited from ancestors present G generations ago follow an exponential distribution and their length range spans generally over several classes, with only a subset of these ROH falling into the class centered around L. Such an approach thus fails to account for the possibility that ROH from different classes might come from the same group of founders. A better estimate may be derived by combining information from multiple ROH (from their average length), providing a single ancestor (or multiple ancestors living in the same generation) contributed to the individual inbreeding. The classification varied also as a function of the inbreeding level ρ, possibly because for higher ρ, the probability that two consecutive short and independent ROH appear as one unique long ROH increases, shifting the distribution towards classes with longer segments. Overall, ROH-based classifications should be interpreted very cautiously. By contrast, the HMM allows joint estimation of the parameters of a mixture of exponential distributions that would result in the observed distribution of HBD segment lengths. We are currently evaluating how precisely the age (in terms of generation to the common ancestor) of each HBD segments / ROH can be estimated with different approaches in more realistic demographic histories. For instance, if we want to understand the relationship between HBD class contributions and Ne, we should determine which past generations are captured by each HBD classes. We must also account for the fact that contributions from the most ancient classes might be swept by autozygosity associated with more recent ancestors. However, we showed that the parameters ρ did not require such corrections (Druet and Gautier, 2021).

References

Alemu S.W., Kadri N.K., Harland C., Faux P., Charlier C. *et al.* (2021) Heredity 126:410–423. https://doi.org/10.1038/s41437-020-00383-9

Bertrand A.R., Kadri N.K., Flori L. Gautier M., and Druet T. (2019) Methods Ecol Evol 10(6):860–866. https://doi.org/10.1111/2041-210X.13167

Druet T., and Gautier M. (2017) Mol Ecol 26:5820–5841. https://doi.org/10.1111/mec.14324

Druet T., and Gautier M. (2021) bioRxiv https://doi.org/10.1101/2021.05.25.445246

Palamara P. (2016) Bioinformatics 32:3032-34.https://doi.org/10.1093/bioinformatics/btw355

Magnier J., Druet T., Naves J., Ouvrard M., Raoul S., *et al.* (2021) bioRxiv https://doi.org/10.1101/2021.10.08.463737

Santiago E., Novo I., Pardiñas A.F., Saura M., Wang J., *et al.* (2020) Mol Biol Evol. 37(12):3642-3653. https://doi.org/10.1093/molbev/msaa169

Szpiech Z.A., Blant A., and Pemberton T.J. (2017) Bioinformatics 33(13):2059–2062. https://doi.org/10.1093/bioinformatics/btx102

Part 15

Genetic diversity and inbreeding: genebanking

240. New FAO guidelines for the management of animal genetic resources

P.J. Boettcher[1*], P. Ajmone-Marsan[2], R. Baumung[1], J. Boes[3], L. Colli[2], C. Ginja[4], M. Honkatukia[3], J. Kantanen[5], G. Leroy[1] and J.A. Lenstra[6]

[1]Food and Agriculture Organization of the United Nations, Viale delle Terme de Caracalla, 00153 Roma, Italy; [2]Università Cattolica del Sacro Cuore, Via Emilia Parmense 84, 29122 Piacenza, Italy; [3]Nordic Genetic Resource Center (NordGen), Växhusvägen 12, 23456 Alnarp, Sweden; [4]Universidade do Porto, Rua Padre Armando Quintas, 4485-661 Vairão, Portugal, [5]Natural Resources Institute Finland, Myllytie 1, 31600 Jokioinen, Finland; [6]Utrecht University, Postbus 80125, 3508 TC Utrecht, the Netherlands; paul.boettcher@fao.org

Abstract

Biotechnology offers powerful tools for the management of livestock genetic diversity. Genomics, bioinformatics, reproductive technologies and tools for cryopreservation of germplasm are particularly important. These biotechnologies have undergone notable advancements in recent years. Developing capacities and sharing expertise are two ways by which the Food and Agriculture Organization of the United Nations (FAO) supports its member nations. For these reasons, FAO has drafted two new guidelines, on: (1) genomic characterization; and (2) cryoconservation. The guidelines update previous versions from the early 2010s by providing information on exploitation of relevant recent biotechnological developments. They were prepared in collaboration with: (1) members of the International Society for Animal Genetics; and (2) partners of the IMAGE research consortium, including NordGen. The guidelines were written by teams of scientific experts and reviewed by FAO Member Nations. Draft versions are already available online and the finalized guidelines will be released during 2022.

Introduction

Animal genetic resources, and their diversity, are an essential part of the biological basis for world food security and contribute to the livelihoods of more than one billion people. Livestock genetic diversity is declining and the capacity to manage animal genetic resources is insufficient in many countries, especially in developing countries (FAO, 2007b and 2015). To establish an intergovernmental policy to address these issues, the Member Nations developed and adopted the Global Plan of Action for Animal Genetic Resources (GPA; FAO, 2007a). The main responsibility for implementing the GPA lies with countries, but international organizations, including FAO, are expected to play important roles, especially with cross-country collaboration. Among the actions and priorities included in the GPA is the development of 'technical standards and protocols for phenotypic and molecular characterization' and 'approaches and technical standards for conservation'. To this end, FAO had previously developed guidelines on molecular genetic characterization (FAO, 2011) and cryoconservation (FAO, 2012) of animal genetic resources, among others.

The last editions of guidelines on molecular genetic characterization and cryoconservation (i.e. gene-banking) were prepared a decade ago. Although much of the content of the two documents has remained perfectly valid, many technological advancements have occurred in the last 10 years, the greater application of which could enhance management of animal genetic resources. Therefore, the objective of this work was to update the FAO guidelines.

Materials & methods

The new guidelines were developed through substantial international cooperation.

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_240, © P.J. Boettcher *et al.* 2022

Genomic characterization. FAO has a long history of cooperation with the International Society of Animal Genetics (ISAG), including the ISAG/FAO Advisory Group on Animal Genetic Diversity, an ISAG standing committee (ISAG, 2021). Members of this Advisory Group served as authors and reviewers of the previous guidelines on molecular characterization (FAO, 2011). Five current and former members of the Advisory Group served as editors for the new guidelines (and co-authors of this paper). These editors formulated outlines for each chapter, identified and engaged with authors, and oversaw the preparation process. In some cases, the editors also served as authors.

Innovations in cryoconservation. From 2016 to 2020, FAO was a partner in the consortium of the European Union Horizon 2020 research project IMAGE (Innovative Management of Animal Genetic Resources; IMAGE, 2016). The objective of IMAGE was to 'enhance the use of genetic collections and to upgrade animal gene bank management', both in Europe and globally. One deliverable of IMAGE was the review of the existing FAO guidelines on cryoconservation (FAO, 2012) and recommendation of an outline and content for new guidelines. The Nordic Genetic Resource Center (NordGen) was the IMAGE partner responsible for this deliverable, although all partners contributed. NordGen staff were made co-editors of the new guidelines. Authors for each chapter consisted of at least one IMAGE member and (when possible) one author from a non-European country, to ensure global representation and relevance. Scientists from around the world provided case studies.

Preparation. The two guidelines were written over an approximately one-year period starting in the spring of 2020. The content was reviewed by guideline editors, authors, FAO staff, invited experts and public volunteers. Draft versions of the documents were presented to the 11th session of the Intergovernmental Technical Working Group on Animal Genetic Resources for Food and Agriculture (FAO, 2021a) for review by FAO Member Countries in May 2021. In addition, the draft guidelines on genomic characterization were reviewed by seven other members of the ISAG/FAO Advisory Group. In September 2021, the revised versions (FAO, 2021b,c) were presented to the 18th session of the Commission on Genetic Resources for Food and Agriculture, a governing body of FAO.

Results

As previously mentioned, the objective of this work was to update previous guidelines (FAO, 2011 and 2012). Regarding characterization, genomics greatly increases the precision of analyses of genetic variation, relative to earlier tools such as microsatellite markers, and have also opened previously non-existent opportunities for the study of functional variation and population histories, and the management of breeds. For cryoconservation, awareness has increased on the use of *in vitro* collections as sources of variation for *in situ* populations, rather than simply as an 'insurance policy' against breed extinction. The recognition of genomics and gene-banking as complementary tools is highlighted in the guidelines.

Genomic characterization. Table 1 provides the titles of each section of the guidelines on genomic characterization, the main topics of each section. Relative to the previous guidelines (FAO, 2011), the new version provides less information on the basics of genomic diversity studies and much more on genetic tools, methods and applications. Appendices include an extensive glossary, procedures for DNA (blood) sampling, questionnaires for collection of complementary data, description of popular software, and a summary of available DNA chips.

Innovations in cryoconservation. Table 2 has the titles of each section of the guidelines on cryoconservation. Case studies were included as text boxes to demonstrate the real application of the tools and approaches discussed. The guidelines also contain appendices that provide detailed instructions on cryopreservation of the most frequently used types of germplasm and tissues, as well as 'model' agreements for material acquisition and transfer, based on the tools developed by the European Regional Focal Point for Animal Genetic Resources.

Table 1. Title, main topics covered, and authors for each section of the FAO guidelines on genomic characterization animal genetic resources.

Title	Main topics
1. Introduction	Rationale for characterization; history and prospects
2. The basics of genomic diversity studies	Prerequisites and planning for optimal results; choice of DNA technology; complementary data to be collected; key processes undertaken in the field and laboratory; data analysis; translating and disseminating results; international coordination
3. Genomic tools and methods	SNP genotyping; candidate genes vs SNP chips; genotyping by sequencing; preparing a SNP data set; merging existing data sets; quality control; whole-genome sequencing; genotype imputation
4. Applications of genomics	Assessment of genomic variation within and between populations; assessing population structure; reconstructing population history; demographic modelling; mitochondrial DNA; sex chromosomes; selection signatures; genome-wide association studies; applying genomics for sustainable use and conservation of genetic resources
Conclusions and recommendations	Summary of key points from previous sections

Table 2. Titles, main topics covered for each section of the FAO guidelines on innovations in cryoconservation of animal genetic resources.

Title	Main topics
1. Building a gene banking strategy	Role of gene-banking in management of diversity; key issues and challenges; gene-banking objectives; governance; elements of a gene-banking strategy; utilization of stored material; rationalization of collections; multi-country gene-banking
2. Quality management	Quality management systems and benefits for gene banks; key processes; implementation; evaluation and assessment; continual improvement of gene-bank operations
3. Choice of biological material to be preserved	Semen; oocytes and embryos; gonads and gonadal tissues; germ, stem and somatic cells; matching material with national needs for utilization
4. Economics of gene banking	Costs and benefits of gene-banking; cost analysis challenges; cost-effectiveness analysis
5. Developing and using gene bank collections	Analysing genetic variation in a gene bank collection; pedigrees vs genomics; software for bank management; updating collections to account for changes over time
6. Collection and cryopreservation of germ plasm and tissues	Features of long-term cryopreservation; material collection, slow freezing vs vitrification; thawing and utilization; evaluation of quality and viability; semen; embryos and oocytes; gonads and gonadal tissue; germ and somatic cells
7. Sanitary issues	Material collection; testing for freedom of pathogens; processing; storage; distribution; national case studies
8. Databases and documentation	Critical issues for gene-bank data; data management plans; data types; data for internal vs external use; data security and protection; international information systems
9. Legal issues	Legal basis of gene banking; statutes and decision-making bodies; acquiring samples; regulating access; benefit-sharing
10. Capacity building and training	Academic instruction; training for gene-bank staff; public outreach and communication with gene-bank stakeholders

Discussion

Draft versions of the guidelines are already available online (FAO, 2021b,c). The finalized guidelines will be released in 2022. Also in 2022, FAO will organize a series of webinars to raise awareness of the guidelines and disseminate the material to scientists and other stakeholders in FAO Member Nations. Alas, development and dissemination of technical guidelines is merely an initial step in the process of using genomics and cryoconservation to improve the management of animal genetic resources. Ensuring that the knowledge in the guidelines has an impact requires increased political will, investment, and international cooperation.

References

FAO. 2007a. Global Plan of Action for Animal Genetic Resources and the Interlaken Declaration. FAO, Rome, Italy.

FAO. 2007b. The State of the World's Animal Genetic Resources for Food and Agriculture. Rischkowsky B. and Pilling D. (eds) FAO, Rome, Italy.

FAO 2011. Molecular genetic characterization of animal genetic resources. FAO, Rome, Italy.

FAO. 2012. Cryoconservation of animal genetic resources. FAO, Rome, Italy.

FAO. 2015. The Second Report on the State of the World's Animal Genetic Resources for Food and Agriculture. Scherf B.D. and Pilling D. (eds) FAO, Rome, Italy.

FAO, 2021a. 11th Session of the ITWG on Animal Genetic Resources for Food and Agriculture. Available at: www.fao.org/animal-genetics/events/events-detail/en/c/1369166.

FAO, 2021b. Draft practical guide on genomic characterization of animal genetic resources. Available at: https://www.fao.org/3/ng882en/ng882en.pdf.

FAO, 2021c. Draft practical guide on innovations in cryoconservation of animal genetic resources. Available at: https://www.fao.org/3/ng882en/ng882en.pdf.

IMAGE. 2016. IMAGE Innovative Management of Animal Genetic Resources. Available at: https://imageh2020.eu.

ISAG. 2022. Committees of the Society. Available at: https://www.isag.us/committees.asp.

241. Genebanks for animal genetic resources in Europe: quality management system as a tool to improve their operation

C. Danchin[1*], D. Bojkovski[2], D.Duclos[1], M. Neuteboom[3], M. Schoon[3], F. Tejerina Ampudia[4], M. Tixier-Boichard[5], and S.J. Hiemstra[3]

[1]Institut de l'Elevage, 149 rue de Bercy 75012 Paris, France; [2]Department of Animal science, Biotechnical Faculty, University of Ljubljana, Jamnikarjeva 101, 1000 Ljubljana, Slovenia; [3]Centre for Genetic Resources, the Netherlands (CGN), Wageningen University & Research, P.O. Box 338, 6700 AH Wageningen, the Netherlands; [4]Ministry of Agriculture, Fisheries and Food of Spain, C/ Almagro, 33, 28071 Madrid, Spain; [5]University Paris-Saclay, Inrae, AgroParisTech, GABI, 78350 Jouy-en-Josas, France; coralie.danchin@idele.fr

Abstract

In order to safeguard farm animal genetic diversity for current and future use, reinforcement of national genebank strategies is needed as an important complementary measure next to maintenance and improvement of live breeding populations. A Quality Management System (QMS) is an important tool to integrate the needs of various stakeholders into the genebank strategy, to mitigate risks associated with collections' management and to professionalize the genebank operations and protocols. A peer review system, where genebank staff visit and review other genebanks, was tested and appeared to be an effective tool to enhance the further development of the strategies of individual genebanks, their operations and impact. For instance the peer review was an opportunity to exchange about collection and freezing protocols, advises were given to better organize storage of genetic material and data management, or more general recommendation were done such as suggesting to develop a better communication strategy.

Introduction

The most remarkable advantage of cryoconservation of genetic material in genebanks is its capacity to store reproductive material without deterioration. Providing access to this stored material makes possible a large range of uses aimed at supporting *in vivo* conservation, as a backup in case of genetic or sanitary problems, for developing new lines/breeds, increasing the diversity in the living population, modifying and/or reorienting selection, or for research purposes. The relevance of genebanks is recognized by the FAO Global Plan of Action for Animal Genetic Resources, and by the United Nations in the 2030 Agenda, in relation to SDG 2 (Zero Hunger) and more specifically Indicator 2.5.1: 'number of animal genetic resources (AnGR) for food and agriculture secured in either medium or long term conservation facilities'.

In Europe, recent surveys showed that many countries have established national genebank collections of AnGR as a backup (see Leroy *et al.* 2019, Passemard *et al.* 2018) but still more work needs to be done. Meanwhile, EUGENA (https://eugena-erfp.net/en/), governed by the European Regional Focal Point (ERFP), was established to support *ex situ* conservation.

The development of genebanks in Europe will be overviewed in this article as well as how a QMS and a peer review systems will help further developing them.

State of development of genebanks in Europe

The analysis of SDG indicator 2.5.1 based on FAO online database (https://www.fao.org/dad-is/) shows that the cryopreservation status is known for two thirds of local breeds in Europe. Only 133 local breeds have sufficient material to allow reconstitution in case of extinction (3.5%). This proportion is coherent with findings of Leroy *et al.* (2020), who indicated that most breeds in European countries have either no or insufficient material in genebanks.

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_241, © C. Danchin *et al.* 2022

EUGENA is a network of nationally recognized genebanks for animal genetic resources. It aims to support the *ex situ* conservation and sustainable use of AnGR at European level. Exchange of knowledge and experiences as well as facilitating access to information about genebank collections in Europe are the main objectives of EUGENA. The participation in EUGENA is open to all institutions managing genebanks.

A specific ERFP *Ad Hoc* Action was launched to support the development of EUGENA, to identify candidate EUGENA genebanks, and improve the information about the genebanks in Europe. To do so, a survey was developed to collect information about genebanks in European countries, with questions related to the institution, its objective(s), the livestock species represented in the genebank, the type of biological material stored, and finally any link to communication tools such as a web page or social media channel.

The survey was sent via a large e-mailing to a wide array of stakeholders, including National Coordinators and experts within the ERFP network and the wider European animal breeding and reproduction sector in 34 European countries. A total of 103 answers were received from 26 European countries. Conservation was quoted as the main reason for establishing genebank collections by 58% of the respondents, followed by research (23%) and commercial purposes (18%). The majority of the responding genebanks are holding reproductive material (44 institutions) while 19 of them are specialized in genomic collections; 40 institutions are hosting both type of material. As already found in previous studies (Leroy *et al.* 2019, Passemard *et al.* 2018), cattle is the main species of interest (74 collections) and most collections are holding semen (83 genebanks). Beyond cattle, sheep, goat, pig, horse and chicken are regularly represented in the collections.

The combination of the results of the *Ad Hoc* Action and the study by Passemard *et al.*, 2018 allowed the ERFP network to build up an inventory of 126 different institutions operating as genebanks in the European region. Thus, collaboration and coordination between all these actors involved are needed to foster the *ex situ* conservation activities within EUGENA.

The EU H2020 GenResBridge project (grant # 817580) was a leverage that pushed the ERFP community to formalize an Animal Genetic Resources Strategy for Europe. This Strategy proposes a series of key recommendations with the objective to strengthen the EUGENA network. Recommendations include mapping of all *ex situ* conservation facilities and enlarging the EUGENA network, ensuring breed specific core collections, evaluating sanitary issues connected with stored material and its use, implementing QMS, initiating research on new cryopreservation technologies, and facilitating access to cryopreserved genetic material where possible. Finally, to secure continued funding of the *ex situ* conservation activities, the general public must be informed about the efforts, results and value of the genebanks.

Quality management system (QMS)

Genebanks must ensure high quality conservation by providing secure, long-term storage to guarantee future viability and biosecurity. Dedicated systems for quality management, be they formal or informal, are important tools that can assist genebanks managers to monitor their progress towards their targets and fulfil their mission.

A task performed within the H2020 IMAGE project (grant #677353) was a worldwide survey on QMS (Zomerdijk *et al.*, 2020). An online questionnaire was distributed through the FAO network of national coordinators. Ninety responses were received from 62 countries. It showed that 30% of genebanks have established a QMS, 17% have a formal certification (mainly ISO 9001:2015), 35% have formal cryoconservation goals, and 50% have identified the major risks to their collections. As far as collecting, processing, storage and documentation, 59% of genebanks use Material Transfer Agreements for (part of)

incoming samples, 88% follow specific Standard Operating Procedures for processing and freezing, 49% have a database for monitoring collections. The least compliance was for distribution of outgoing material since only 24% have formal procedures for distribution of the material.

FAO has developed Guidelines with the aim to give technical guidance on the design and establishment of animal genebanks (FAO, 2012), but without any specific reference to QMS. One important conclusion from the EU H2020 IMAGE project was that quality management principles should be better reflected in the FAO Guidelines. Therefore, in the updated version of the guidelines, a whole section is devoted to 'Quality management for improved organization and implementation' (FAO, 2021).

Another output of the H2020 IMAGE project was the design of a self-diagnostic tool to facilitate the uptake of quality management by genebanks which can be downloaded on the project website (https://www.imageh2020.eu/). The IMAGE survey showed that a significant number of European genebanks were interested in developing their QMS, up to the stage of being certified with reference to ISO 9001:2015 for instance. Yet, the development of a QMS is generally viewed as time-consuming and costly, so that few genebanks complete the exercise or hesitate to invest in it. The aim of the self-diagnostic tool is to help a genebank to assess its progress towards a full QMS, so that it can get a precise idea of where to focus its efforts.

Peer reviews to strengthen genebanks and their quality management

The H2020 GenResBridge project gave the opportunity to implement a 'peer review' assessment of genebanks. Peer reviews are based on the principle that genebank experts visit and review other genebanks, genebanks giving full transparency about the strategies, facilities and protocols, and having discussions about these. After each visit a report is written with observations and recommendations, in order to support further development of the genebank strategy and to improve the genebank management. Peer review visits of national livestock genebanks were organised in 2021, involving the French, Dutch and Slovenian National genebanks. After the visits, all reports were published on the ERFP website. Review reports are short and to the point, listing only major observations and associated recommendations.

The overall impression of the genebank review pilot was positive (Bojkovski *et al.*, 2021) and considered as a useful exercise, complementary to internal audits, both for the hosting genebank and its staff, as for the reviewers visiting the other genebank. The complete transparency of the participating genebanks resulted in excellent discussions about strengths and opportunities for improvements. Comparison of the three national livestock genebanks highlighted similarities as well as differences between the genebanks.

Sharing experiences and knowledge was beneficial for all three genebanks and resulted in new ideas and inspiration. These recommendations could be very practical, such as exchanges on collecting or freezing protocols, advises on how to better organize the gene bank, on how much and which material should be stored. QMS traceability documents were also compared. More general recommendations were also given; for instance it was suggested to a genebank to better develop its communication strategy; it was also stressed out the importance of drafting documents linked to the samples that clearly state the ownership of the material. The recommendations listed in the individual reports will help the genebanks to further develop their work and priorities.

The review team with different expertise and background consisted between one and three experts of the other two genebanks. The choice of three relatively different genebanks was very positive; two genebanks with long experiences but very different organizational structures versus one younger genebank; one genebank already with a certified QMS, the second in the process to obtain this and the third not having

this objective for the moment. A rotational system was suggested in case such a review system could be implemented on a larger group of genebanks.

Conclusions

The output of the IMAGE self-diagnostic tool and the peer reviews were both acknowledged of being a valuable experience for genebanks in Europe. A new ERFP *Ad Hoc* action is set to take place in 2022 that will combine both approaches. The expectation is that ten volunteer genebanks could use the tool and be assessed within this year. A cooperative and step by step implementation of a QMS approach is definitely a way forward to secure the gene pools that are conserved in the numerous genebanks from Europe. It is also hoped that this action will encourage genebanks to get involved in EUGENA.

References

Bojkovski D., Duclos D., Neuteboom M., Hiemstra S.J., Schoon M. (2021). Livestock Genebank Peer Reviews. Available at: https://www.animalgeneticresources.net/wp-content/uploads/2021/12/20211123_AnGR_GB_Report_PeerReview_GENERAL_photos.pdf

FAO. (2012) Cryoconservation of Animal Genetic Resources;Rome, Italy.

FAO, Draft practical guide on innovations in cryoconservation of Animal Genetic Resources 2021. Available at: https://www.fao.org/3/ng882en/ng882en.pdf

GenResBridge consortium (2021). Genetic Resources Strategy for Europe Available a: http://www.genresbridge.eu/fileadmin/templates/Genres/Uploads/Documents/GRS4E.pdf

Leroy, G.; Boettcher, P.; Besbes, B.; Danchin-Burge, C.; Baumung, R.; Hiemstra, S.J. (2019). Cryoconservation of Animal Genetic Resources in Europe and Two African Countries: A Gap Analysis. Diversity, 11, 240. https://doi.org/10.3390/d11120240

Passemard A.S, Hiemstra S.J., Tixier-Boichard M., and Danchin C.(2018) Proc. of the 10th WCGALP, Auckland, Canada.

Zomerdijk F., Hiemstra S.J., d'Arbaumont M., Tixier-Boichard M., and Boettcher P. (2020). Biopreservation and Biobanking.244-253.

242. Genetic characterization of endangered indigenous pigs of Angola

P. Sá[1], D. Santos[1,2], A. Leitão[1,2], J.M. Cordeiro[3], L.T. Gama[1,2] and A.J. Amaral[1,2*]

[1]CIISA – Centro de Investigação Interdisciplinar Em Sanidade Animal, Faculdade de Medicina Veterinária, Universidade de Lisboa, Avenida da Universidade Técnica, 1300-477 Lisbon, Portugal; [2]Laboratório Associado para Ciência Animal e Veterinária (AI4Animals); [3]Faculdade de Medicina Veterinária, Universidade José Eduardo dos Santos, Cuíto, Huambo, Angola; andreiaamaral@fmv.ulisboa.pt

Abstract

It is hypothesized that African pigs derive from Near Eastern pigs that were introduced by land through Egypt, spread to other locations in Africa establishing populations of African indigenous pigs, one of which is called the *Angola Extension*. Angola has been through dramatic social events that have led to a drastic reduction of the population of indigenous pigs, which is currently endangered. These pigs are well adapted to poor diets and harsh environments. Here we used whole-genome-sequencing to explore the genetic diversity of indigenous pigs from Angola and their relationships with worldwide breeds. Results show that these pigs harbour thousands of unique mutations and high levels of heterozygosity. The location of missense mutations suggest the occurrence of adaptation events related with metabolism and tissue remodelling.

Introduction

The history of pigs in sub-Saharan Africa has been poorly studied. The DAD-IS database (FAO) accounts for a large number of African indigenous breeds, most of them with an unknown risk status. Similarly as in Europe, it is hypothesized that African pigs derive from Near Eastern pigs that were introduced by land through Egypt (Blench and MacDonald, 2000). Later during the discovery age, it is known that the Portuguese introduced Iberian pigs along the seacoast in several trading posts. The use of linguist studies focused on the origin of the word pig allowed to identify four regions in which the word pig does not derive from the Portuguese word 'porco' and these have been classified as regions of the distribution of indigenous pigs in Africa. These are called the Ancient Berber, Sennar populations and two West populations, one called the West African Extension and the other the Angola Extension that ranges from Cameroon to Angola (Blench and MacDonald, 2000). Pigs in these regions are mostly bred free ranging with access to a diet poor in protein and are well adapted to the environmental conditions including endemic diseases. The levels of genetic diversity in several African pig populations assessed by microsatellites suggested that these harbour high levels of diversity (Ramírez *et al.*, 2009).

The region of the Angola Extension is still today a remote area where pigs are bred free ranging, whose animals have never been characterized regarding their genetics. In local dialect, pigs are called 'ongulu' or 'olongulu', a word not derived from Portuguese. The country has been through dramatic social events that have led to a drastic reduction of the population of indigenous pigs. The aim of the current study was to investigate the levels of genetic diversity in a population that remains in a remote region near the Angola extension and of its relatedness with worldwide breeds.

Materials & methods

Samples of ear tissue were collected from four local pigs (ANG) at the municipality of Bailundo in Angola near the Angola Extension region. Total DNA was extracted and sequencing libraries with paired 150bp reads expected to achieve 10-fold coverage were generated. Sequencing was performed by outsourcing at Novogene company, using an Illumina NovaSeq 6000 sequencer. Public datasets comprising data from 5 European domestic breeds (5 Iberian, 4 Duroc, 9 Large White, 5 Landrace, 4 Pietrain), and 13 samples from Meishan an Asian domestic breed, plus 13 Asian wild boars and 25 European wild boars (PRJEB9922,

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_242, © P. Sá *et al.* 2022

PRJEB1683, PRJNA320525 and PRJNA255085) were included for comparison. Prinseq-lite v.0.20.4 was used to select reads with length of 50-150nt and an average phred score quality higher than 20. After, reads were mapped to Sscrofa11.1 using BWA v.0.7.17 (BWA) (Li and Durbin, 2009). PCR Duplicates were removed using Picard v.2.23.4 (Broad Institute). Variant calling was performed using 'SAMtools Variant Caller' v.1.0.6. BCFtools v.1.10.2 (Li, 2011) was used to select SNPs with minimum read depth of 10 and a minimum number of 3 reads carrying the alternative allele. Variant Effect Predictor (VEP) (McLaren *et al.*, 2016) was used for SNP annotation. SNP summary statistics was estimated using populations module of STACKS (Catchen *et al.*, 2011). David 6.8 (Dennis *et al.*, 2003) was used to assign gene ontology (GO) terms and Revigo (Supek *et al.*, 2011) was used to generate a reduced representation of these.

Results

Raw data generated from ANG pigs, ranged from 95 million to 126 million reads across samples, corresponding to an average size of 108 million reads (Table 1). Read filtering, retained over 99.80% of the reads of which at least 99% were mapped against the reference genome. Read mapping depth ranged from 9.80 to 12.12, similarly to the 78 public datasets analysed. Variant calling identified more than eight million SNPs in autosomes for ANG pigs of which, ~900 thousand are unique to this population in comparison with European and Asian domestic and wild populations. The occurrence of SNPs was on average 1 SNP per 227 kbp. Most identified SNPs were located in intronic regions and have unknown functional significance. Among the exonic SNPs, approximately 21 thousand were annotated as missense, stop-and-start gain and loss.

In order to investigate the impact of missense SNPs in ANG pigs, we have selected the genes that harboured missense SNPs occurring in all the samples. This resulted in the identification of 1,536 genes, for which we were able to retrieve 50 GO terms, further summarized in 9 parental GO terms for biological processes.

GO terms grouped for tissue remodelling were the most frequent, comprising a total of 22 GO terms (Table 2).

Table 1. Summary and mapping statistics of ANG pigs' data and of public datasets.

Sample sets	N[1]	Raw reads	Paired reads	Mapping[1] (%)	Genome coverage[1] (%)	Read depth[2] (x)
ANG	4	108,695,832	108,549,926	99.35	97.91	10.75
PRJEB9922	37	210,848,029	152,596,663	99.21	97.14	11.27
PRJEB1683	37	111,879,123	103,604,188	99.94	80.13	9.29
PRJNA320525	1	211,300,791	192,172,164	97.06	98.11	14.97
PRJNA255085	3	179,282,739	164,017,544	97.90	97.63	12.55

[1] Number of samples.

[2] Values represent mean over samples.

Table 2. Parental GO terms obtained with Revigo's reduced representation of genes with missense SNPs in ANG pigs.

GO ID	Term	Number of child GO terms
GO:0048771	tissue remodelling	22
GO:0051122	hepoxilin biosynthetic process	7
GO:0086073	bundle of His cell-Purkinje myocyte adhesion involved in cell communication	5
GO:1903546	protein localization to photoreceptor outer segment	5
GO:0016139	glycoside catabolic process	4
GO:0060335	positive regulation of interferon-gamma-mediated signalling pathway	4
GO:0008037	cell recognition	1
GO:1904714	regulation of chaperone-mediated autophagy	1
GO:0010659	cardiac muscle cell apoptotic process	1

The observed heterozygosity (Ho) and expected heterozygosity (He) among ANG pigs was 0.78 and 0.41 respectively. The estimated FIS (inbreeding coefficient) was -0.13.

Discussion

In this study, we used whole genome sequencing for the identification of SNPs and characterization of genetic diversity of a local population of indigenous pigs from Angola. The data generated displayed a level of an average depth of 10x and a mapping rate of at least 97.5% similar to previous studies in other pig breeds (Li *et al.*, 2014) and which are within the range of the optimal value for SNP calling in pigs (Jiang *et al.*, 2019). SNP density was 1 SNP per 227 kbp which allows an high power for posterior identification of signatures of selection (Ma *et al.*, 2015). Approximately 900 thousand SNPs are unique to this African population in comparison with European and Asian domestic and wild populations, suggesting the existence of a rich genetic pool. Similarly to other studies in pigs (Groenen *et al.*, 2012; Li *et al.*, 2017) and in other species (Gao *et al.*, 2020; Weldenegodguad *et al.*, 2019) most SNPs were identified in non-coding regions. Among the exonic variants with functional impact these are mostly located in genes with functions related with tissue remodelling. This result is in agreement with the fact that body shape related traits have been a major focus of selection in pig breeds and for which we find 1,631 QTLs in the pig QTL database (Hu *et al.*, 2019). Signatures of selection related to body shape have also been previously reported (Li *et al.*, 2014). In fact, ANG pigs are of smaller stature and are able to survive being fed with leftovers mostly derived from corn and other vegetables, with no access to animal protein (Malungo, 2019). Finally, the levels of observed heterozygosity were higher in comparison with European and Asian pig breeds as previously described by Ramirez *et al.* (2009) in other populations of African pigs (Ramírez *et al.*, 2009). We observed a negative FIS value which indicates an excess of heterozygotes in relation to what was expected under random mating most likely due to the existence of a geographical dispersal of small groups (Parreira *et al.*, 2020) and the possibility of exchange between different regions (Pope, 1992). Thus reinforcing the need to sample in other regions of the country.

This study generated the first whole genome sequencing data of a pig population that is highly threatened. Our results have provided novel insights regarding the genetic value of these pigs. Their genomes harbour unique mutations. The fact that many of the SNPs are located in genes related with tissue remodelling suggests that this population may harbour unique signatures of selection that should be further investigated.

This work was supported by FCT- Fundação para a Ciência e Tecnologia IP (UIDB/00276/2020).

References

Blench R., and MacDonald K. (2000) The Origins and Development of African Livestock: Archaeology, Genetics, Linguistics and Ethnography. Routledge, Oxfordshire, UK.

Broad Institute, Picard Tools. Available at: http://broadinstitute.github.io/picard/.

Catchen, J. M., *et al.* (2011) G3 (Bethesda) 1(3):171–182. https://doi.org/10.1534/g3.111.000240

Dennis, G., *et al.* (2003) Genome Biol. 4(5):P3. https://doi.org/10.1186/gb-2003-4-9-r60

Gao, J. *et al.* (2020) Front. Genet., 11. https://doi.org/10.3389/FGENE.2020.00901/FULL

Groenen, M. A. M. *et al.* (2012) Nature 491(7424):393–398. https://doi.org/10.1038/nature11622

Hu, Z. L., Park, C. A., and Reecy, J. M. (2019) Nucleic Acids Res. 47(D1):D701–D710. https://doi.org/10.1093/NAR/GKY1084

Jiang, Yifan *et al.* (2019) BMC Bioinform., 20(1). https://doi.org/10.1186/S12859-019-3164-Z

Li, H. (2011) Bioinformatics 27(21):2987. https://doi.org/10.1093/BIOINFORMATICS/BTR509

Li, H. and Durbin, R. (2009) Bioinformatics 25(14):1754–1760. https://doi.org/10.1093/bioinformatics/btp324

Li, M. *et al.* (2014) Sci. Rep., 4:1–7. https://doi.org/10.1038/srep04678

Li, W.T. *et al.* (2017) Proc. Royal Soc. B., 284(1869). https://doi.org/10.1098/RSPB.2017.2437.

Ma, Y. *et al.* (2015) Heredity 115(5):426–436. https://doi.org/10.1038/HDY.2015.42
Malungo, O. Q. (2019) Caracterização da produção de suínos e avaliação dos sistemas de exploração no sector familiar no município de Chitembo, Província do Bié. José Eduardo dos Santos University, Huambo, Angola.
McLaren, W. *et al.* (2016) Genome Biol., 17(1):1–14. https://doi.org/10.1186/S13059-016-0974-4
Parreira, B. *et al.* (2020) Heredity 125(5):328–339. https://doi.org/10.1038/s41437-020-0345-5
Ramírez, O. *et al.* (2009) Mol. Biol. Evol., 26(9):2061–2072. https://doi.org/10.1093/MOLBEV/MSP118.
Pope, T. (1992) Evolution 46(4). https://doi.org/10.2307/2409760.
Supek, F., *et al.* (2011) PLoS One 6(7):e21800. https://doi.org/10.1371/journal.pone.0021800
Weldenegodguad, M. *et al.* (2019) Front. Genet. 10(1):728. https://doi.org/10.3389/FGENE.2018.00728/BIBTEX.

243. Genetic diversity and inbreeding in US Jersey population and germplasm collection

K. Srikanth[1*], M. Neupane[2], M.A. Jaafar[1], C.W. Wolfe[3], H.J. Huson[1], C.P. Van Tassell[2] and H.D. Blackburn[4]

[1]Department of Animal Science, Cornell University, 507 Tower Rd, Ithaca, NY 14853, USA; [2]Animal Genomics and Improvement Laboratory, Agricultural Research Service, USDA, 10300 Baltimore Ave, Beltsville, 20705-2350, MD, USA; [3]American Jersey Cattle Association, 6486 E. Main St, Reynoldsburg, OH 43068, USA; [4]National Animal Germplasm Program, USDA, 111 South Mason St, Fort Collins, CO 80521, USA; ks786@cornell.edu

Abstract

Genetic diversity is needed to improve health, production traits and for breed conservation. Germplasm collection conserves diversity and enables population reconstruction or corrective mating as needed. The objective of this study was to evaluate genetic diversity in US Jersey cattle and identify animals for germplasm preservation to capture population diversity. Genotype and pedigree information on 36,179 Jersey bulls were obtained from the Council on Dairy Cattle Breeding (CDCB), including the 782 Jersey bulls preserved by the USDA-ARS National Animal Germplasm Program (NAGP). Genetic diversity indices (*Ho*, *He*, *MAF*) were similar in both the groups. Admixture results suggested some differences in the genetic composition between the groups. The pedigree and genomic inbreeding ranged between 7.68 (F_{PED}) and 15.77 (F_{ROH}) for CDCB animals, and between 6.30 (F_{PED}) and 14.43 (F_{ROH}) in the NAGP collected animals. The average correlation between genomic and pedigree inbreeding was 0.63, while between genomic inbreeding estimates was 0.94.

Introduction

Genetic diversity in cattle is associated with improvement of economically important traits, management decisions, and breed conservation. The excessive accumulation of inbreeding or homozygosity leads to genetic diversity loss. Pedigree-based studies have suggested an increased rate of inbreeding and loss of diversity in Canadian Jersey and Holstein cattle (Stachowicz *et al.*, 2011). Melka *et al.* (2012) found lower levels of diversity in Jersey relative to Holstein, and Brown Swiss bull populations and suggested crossbreeding to recover genetic diversity, if diversity within the breeds continues to get worse. The study of genetic diversity promotes understanding of evolution, breed differentiation, and gene pool within the breeds (Doekes *et al.*, 2020 and Huson *et al.*, 2020; Melka *et al.,* 2012). Cryopreservation of recent and historic germplasm is essential to improve management and production of cattle breeds and maintain breed diversity. The USDA-ARS National Animal Germplasm Program (NAGP) collection helps to maintain and dissipate this diversity and reconstruct it in the future if needed (Blackburn *et al.*, 2019). The objective of this study was to evaluate genetic diversity in the US Jersey breed and to identify additional animals for NAGP germplasm collection that would capture and preserve this diversity in cyropreservation. It also aims to assess distant and recent rates of inbreeding in US Jersey cattle.

Materials & methods

Study population. A total of 36,179 genotyped Jersey bulls born between 1950 and 2020 were used in this study. Genotype and pedigree information on these animals were obtained from the national cooperator database maintained by the Council on Dairy Cattle Breeding (CDCB, https://www.uscdcb.com/). This population included 782 NAGP Jersey sires born between 1953 and 2016. Animals were genotyped using 45 different arrays and were imputed to 79,389 markers using Findhap ver 3 (VanRaden *et al.,* 2011). The SNPs used in this study are routinely used in US dairy genomic evaluations. Bulls missing pedigree and phenotype records were excluded, and genotypes for 36,370 (35,675 CDCB and 695 NAGP) animals were used in the study (Table 1).

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_243, © K. Srikanth *et al.* 2022

Table 1. Distribution of samples used in this study.

Decade	CDCB	NAGP	Total
1951-1980	57	29	86
1981-1990	254	110	364
1991-2000	1,352	243	1,595
2001-2010	2,388	212	2,600
2011-2020	28,892	101	28,993
2021-	2,732	0	2,732
Total	35,675	695	36,370

Genetic relationship and diversity estimates. Basic diversity indices such as observed (*Ho*), expected heterozygosity (*He*) and minor allele frequencies (MAF) were calculated using PLINK ver 1.9. Principal component analysis was performed using GCTA ver 1.94. Admixture composition was estimated using a maximum likelihood model implemented in ADMIXTURE ver 1.23 (Alexander *et al.*, 2009).

Inbreeding estimation. Pedigree inbreeding (F_{PED}) was estimated using an additive genomic relationship matrix A using only pedigree information to calculate probabilities that gene pairs are identical by descent. Genomic inbreeding was estimated by two approaches: (1) Genomic relationship matrix (F_{GRM}) G (VanRaden *et al.*, 2008). Both pedigree and F_{GRM} were calculated using inbreed1.f90 and inbreed2.f90 implemented in US dairy evaluations. (2) Runs of homozygosity (F_{ROH}) was determined using a sliding window approach following the criteria suggested by (Doekes *et al.* 2019). Inbreeding coefficient was calculated based on the method proposed by McQuillan *et al.* 2008. ROH length based inbreeding was estimated by classifying the identified ROH segments into different length classes: (1) >16 Mb; (2) 8-16 Mb; (3) 4-8 Mb; (4) 2-4 Mb; and (5) 1-2 Mb.

Results & discussion

Genetic diversity between bulls in the CDCB population and NAGP germplasm collection. The average observed heterozygosity, expected heterozygosity and minor allele frequencies in the two populations were similar (Table 2). Principal component analysis revealed considerable variation in the populations, particularly in recent decades (Figure 1). This result could reflect that most of the sires genotyped are from the recent decades (80% of the genotyped bulls were born from 2011 to 2020). ADMIXTURE results (K=2) suggested differences in composition between the animals in the NAGP collection and the wider population (CDCB), with lower levels of admixture in the NAGP collection (Figure 2a). Inclusion of purebred Holstein data in the analysis (K=3) suggested Holstein introgression for CDCB (8.9%) and NAGP (5.4%).

Table 2. Genetic diversity indices observed in the population and germplasm collection.

Group	*Ho*[1] ± Std	*He*[2] ± Std	MAF[3] ± Std
CDCB	0.37±0.12	0.38±0.13	0.28±0.14
NAGP	0.38±0.11	0.39±0.11	0.29±0.12

[1] Ho – observed heterozygosity.
[2] He – expected heterozygosity.
[3] MAF- minor allele frequency.

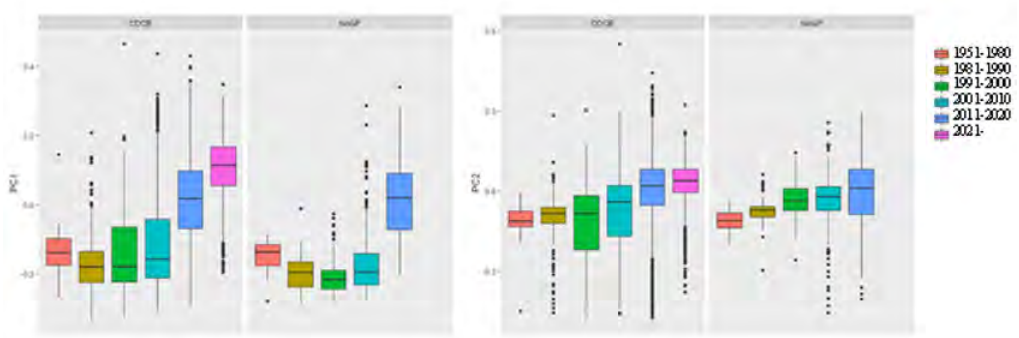


Figure 1. Box plot showing variation captured by Principal components (PC1 and PC2).

Genomic and pedigree inbreeding estimation. The average pedigree inbreeding was 6.3% in the NAGP and 7.68% in the CDCB populations. The genomic inbreeding estimates were 11.12%, 10.16% (F_{GRM}) and 15.77%, 14.43% (F_{ROH}) for CDCB and NAGP collections, respectively. ROH length-based inbreeding ranged between 3.8 and 15.2% for CDCB and 3.94 and 14.43% in NAGP animals (Table 3). The correlation between genomic and pedigree inbreeding ranged between 0.56 to 0.68 with genomic inbreeding estimates ranging between 0.87 to 0.97 (Figure 2b). The average pedigree and genomic inbreeding in NAGP bulls was 10.2 and 6.5 respectively (Table 3). This estimation was similar to previous reports in Jersey (Huson *et al.* 2020). The lack of pedigree completeness, and the random nature of recombination and segregation which pedigree-based measures do not capture, could be a contributing factor for low pedigree inbreeding estimates (Doekes *et al.* 2019). Inbreeding measures based on diagonal elements of the GRM do not distinguish between identity by state (IBS) and identity by descent (IBD) and could lead to inconsistent results (Pryce *et al.* 2014, Villanueva *et al.* 2021). Meuwissen *et al.* 2020, showed that in contrast to classical inbreeding theory, inbreeding measured in terms of drift (F_{DRIFT}) and homozygosity (F_{HOM}) can substantially vary, when inbreeding measures are based on GRM, derived using IBS (F_{GRM}), but not when derived using IBD (F_{ROH}). This could explain the differences in genomic inbreeding measures in this study and based on these reports, F_{ROH} might capture inbreeding more accurately than F_{GRM}. The information generated in this study can be used to improve Jersey diversity, health, and production. It will inform the US Jersey Association of the population status on diversity and inbreeding. Analyses to date suggest the *in situ* population and the NAGP collection have equivalent levels of genetic diversity, but further analysis will better identify gaps in the collection.

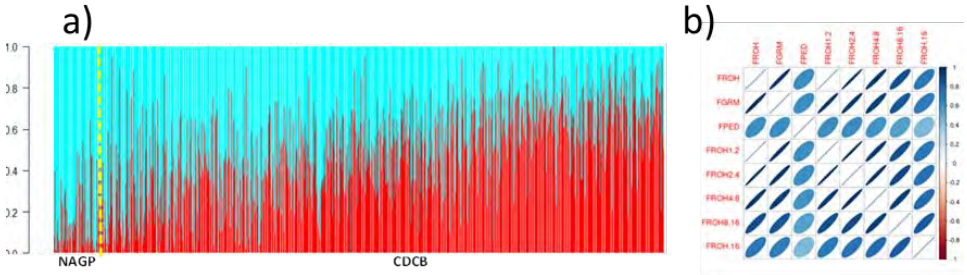


Figure 2. (a) Admixture analysis showing compositional difference between NAGP collection and CDCB population at K=2. (b) Pearson correlation between different inbreeding measures.

Table 3. Comparisons between different inbreeding measures.

Inbreed estimator	CDCB ± SD	NAGP ± SD
F_{PED}	7.68±2.95	6.30±3.16
F_{GRM}	11.12±4.12	10.16±4.15
F_{ROH}	15.77±3.63	14.43±3.73
$F_{ROH\ 1\text{-}2}$	15.23±3.41	14.64±3.02
$F_{ROH\ 2\text{-}4}$	12.56±3.70	12.38±3.83
$F_{ROH\ 4\text{-}8}$	9.34±3.66	9.44±3.73
$F_{ROH\ 8\text{-}16}$	7.16±3.33	7.06±3.41
$F_{ROH\ >16}$	3.80±2.47	3.94±2.59

References

Alexander D.H., Novembre J., and Lange K. (2009) Genome Research 19(9):1655-1664. https://doi.org/10.1101/gr.094052.109.

Blackburn H.D., Wilson C.S., and Krehbiel B. (2019) Diversity 11(12):244. https://doi.org/10.3390/d11120244.

Doekes H.P., Bijma, P., Veerkamp, R.F., de Jong, G., *et al.* (2020) Genetics Selection Evolution 52(1):1-18. https://doi.org/10.1186/s12711-020-00583-1.

Huson H.J., Sonstegard T.S., Godfrey J., Hambrook D., Wolfe C., *et al.* (2020) Frontiers in genetics 11:366. https://doi.org/10.3389/fgene.2020.00366.

McQuillan R., Leutenegger A.L., Abdel-Rahman R., Franklin C.S., Pericic, M., *et al.* (2008) The American Journal of Human Genetics 83(3):359-372. https://doi.org/10.1016/j.ajhg.2008.08.007

Melka M.G., and Schenkel F.S. (2012) BMC Research Notes 5(1): 1-8. https://doi.org/10.1186/1756-0500-5-161.

Meuwissen T. H., Sonesson A. K., Gebregiwergis, G., and Woolliams, J. A. (2020) Frontiers in genetics 11:880. https://doi.org/10.3389/fgene.2020.00880.

Pryce J. E., Haile-Mariam M., Goddard M. E., and Hayes, B. J. (2014). Genetics Selection Evolution, 46(1): 1-14. https://doi.org/10.1186/s12711-014-0071-7.

Stachowicz K., Sargolzaei M., Miglior F., and Schenkel F.S. (2011) Journal of dairy science 94(10):5160-5175. https://doi.org/10.3168/jds.2010-3308.

VanRaden PM., (2008) Journal of Dairy Science 91:4414–4423. https://doi.org/10.3168/jds.2007-0980.

Villanueva B., Fernández A., Saura, M., Caballero, A., Fernández, J., *et al.* (2021) Genetics Selection Evolution 53(1): 1-17.

244. Development and utilization of the United States gene bank collection

H.D. Blackburn[1*], C.S. Wilson[1] and C.D. Dechow[2]

[1]National Animal Germplasm Program United States Department of Agriculture, 1111 South Mason St., Fort Collins, CO 80521, USA; [2]Penn State University, Department of Animal Science, University Park, PA, USA; harvey.blackburn@usda.gov

Abstract

A large and substantial collection genetic resources, exceeding 59,000 animals and over 1.15 million samples, has been acquired by the national gene bank. Approximately 11,000 animals from the collection have been used for reconstituting genetic diversity and genomic studies. As an example, previous research showed *in situ* U.S. Holstein had only two different ancestral Y chromosome linages. It was determined two additional Y Chromosome linages were present in the gene bank and they were reintroduced to the *in situ* population. Reintroducing these two lost Y chromosomes from collection animals representing 1950's genetics illustrated how the collection can be used to address contracting genetic diversity, like limited Y chromosomes. The relatively high PTA measures for net merit and daughter lactation performance suggest gene bank collections may have broader uses than originally thought. Challenges still exist to increase gene bank collection use and to better quantify the diversity captured.

Introduction

For the last 23 years the national genebank has been charged with conserving livestock, aquatic and insect genetic resources important to agriculture (Blackburn, 2018; Blackburn *et al.* 2019). As a result, we have acquired 1.15 million samples from 59,640 animals, representing 191 breeds, 369 subpopulations within those breeds and 44 species. Approximately 50% of the breeds collected have exceeded 50% of the minimum collection targets recommended by FAO (2016). The collection is still being developed with new and existing breeds in the collection or their subpopulations. At this time, it is beneficial to explore how the collection has been used, by whom and some of the results obtained when using the collection.

Materials & methods

The collection has been developed by government, universities and private sector interactions. Program funding facilitates collection development, but the private sector has been highly engaged in donating germplasm samples from various populations. Universities have also contributed to collection development, including the donation of orphan collections developed by retired scientists. From its initiation the collection was designed to encompass all breeds and to serve multiple functions (e.g. research, corrective mating, reconstituting breeds). Therefore, the number of germplasm samples acquired for specific animals can exceed the quantity necessary to reconstitute a population in a worst-case scenario. The Animal-GRIN database contains information on collected samples, e.g. animal identification, phenotypes, genotypes, number and types of samples, geographic locations (https://agrin.ars.usda.gov/main_webpage_dev/ars?language=EN&record_source=US). Sample requests and releases are managed on-line through Animal-GRIN. Upon receipt, requests are reviewed to determine alignment with the program's mission.

For the data presented, Animal-GRIN was queried to determine animal birth year, species, and how the requested animals were to be used. Four sample use categories included: reintroduction (to reintroduce genetic variability or corrective mating), reproductive evaluation (samples used to evaluate sperm quality with live animal matings), genomic evaluation (samples genotyped and used to study genetic diversity or lethal alleles), and regeneration for live animals for research. An example of reintroduction and regeneration

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_244, © H.D. Blackburn *et al.* 2022

for live animal research is given for reintroducing lost Y chromosomes (LYC) into the Holstein breed (Dechow *et al.* 2020). In this example two of the collection's 1950's era bulls were mated to oocytes from current generation cows (ranking in the upper 70th percentile for net merit), resulting in calves from both sexes. The bull calves were raised and placed in a commercial stud and eight daughters remained in the Pennsylvania State University's herd for further study.

Results

To date samples for over 11,000 animals have exited the repository (Figure 1). Most species (cattle, sheep, goats, chickens, pigs, trout, oysters and catfish) in the collection have been used to varying degrees. From 2004 to 2020 a broad range of samples, based upon animal birth year, have been requested and used by stakeholders. There were 10 separate years where samples requested encompassed most birth years, including animals born before 1960. Most of the samples requested were used for molecular genetic studies (92%) by researchers or industry, with 5% being used for reintroduction of genetic variability (480 animals) into *in situ* populations; the remaining two categories accounted for 3% of animals exiting the collection.

Table 1 provides a snapshot of collection use and the results/impacts of those efforts. Industry, universities, and government have used the collection for various purposes. Table 1 also shows that some collection uses have made significant economic contributions to livestock sector. For example, animals from the collection were used in the initial development of genomic breeding values. Breed associations have also used the genotyped animals from the collection to determine if key ancestors were carriers of lethal mutations. Determining the carrier status of these ancestors has led to significant savings in genotyping offspring from sires shown not to carry the mutation.

Genomic predicted transmitting ability (PTA) for founder LYC bulls and their progeny were compared to the national average for net merit and it was determined that progeny from LYC bulls were near breed average (Figure 2) which is contrary to previously thought expectations. Genomic EBV for the LYC daughters averaged 908 kg less than contemporary herd mates; however, the LYC daughters produced 670 and 1,702 kg more milk in first and second parity, respectively (Jewell, 2021).

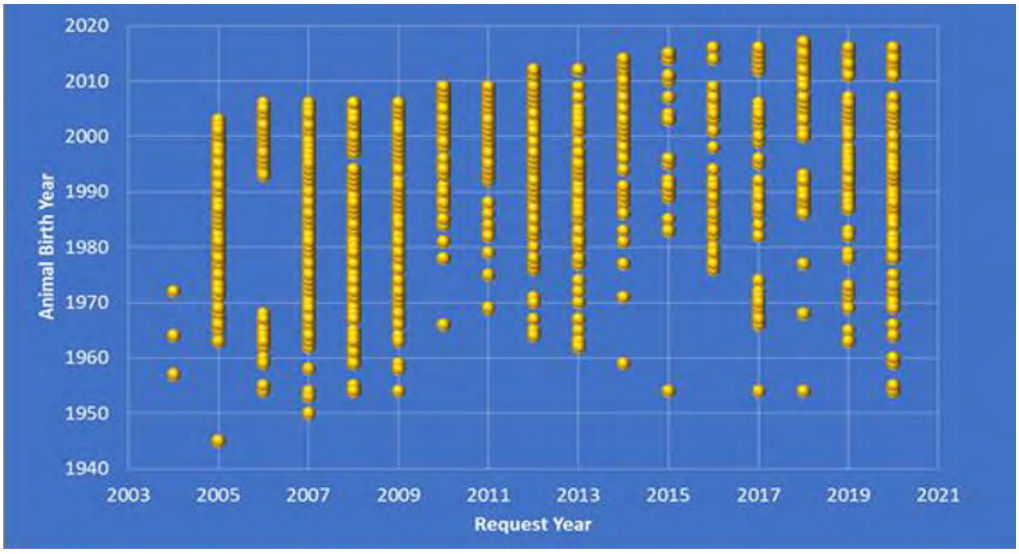


Figure 1. Samples from all species (n=11,000 animals) exiting the gene bank by year requested and animal birth year.

Table 1. Examples and impact of collection use.

Purpose	Breed	User	Results/impact
Develop genomic BV	Holstein Jersey	Research	$400 million/year revenue increase
Identify carriers of lethal genes	Angus, Brangus	Breed Assoc.	Saved breeders >$3 million in genotyping
Reintroduce lost Y chromosomes from 1950 era bulls	Holstein	University & industry	Increased genetic variability, Y chromosome research, progeny semen sold (Dechow *et al.*, 2020)
Reconstituted halothane/napole pig line	Crossbred	University	New research grants, increased university use
Adding genetic variability to rare breeds	Multiple[1]	Industry	Increased genetic variability for stakeholders
Change in allelic frequencies of imported genetics	Meishan	Gov't	Nagoya Protocol – Informed negotiation & policy positions (Blackburn *et al.*, 2014)

[1] Milking Shorthorn, Dexter, Large Black.

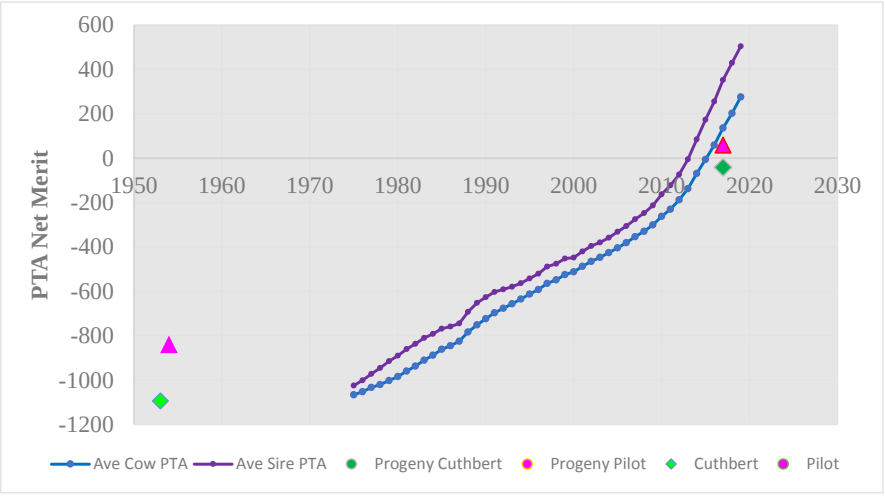


Figure 2. Comparison of 1950 era bulls (Cuthbert and Pilot), progeny of 1950 era bulls to the national average breeding value for net merit (Council of Dairy Cattle Breeding computed PTAs).

Discussion and conclusions

Results indicate that the collection has been widely used and the impact of use has been substantial economically and scientifically. Private breeders of major and minor breeds have used the collection to perform corrective mating for a variety of reasons, including redirecting breeding strategies and adding genetic variation into the *in situ* population.

Reintroduction of two LYC (Table 1) into the Holstein breed is an instructive example of collection use. Figure 2 compares founding sires, their progeny, and the national average PTA for net merit from 1953 to 2019. By mating the bulls to elite cows, the gap between current average bull breeding values and Y-chromosome progeny was substantially narrowed. Bull progeny (born in 2017) PTA were approximately equivalent to breed averages from 2012 to 2014. The lactation performance of LYC daughters was based on a relatively small sample, but substantially exceeded expectations. The LYC project underscores the need to further evaluate collection progeny to determine the extent that current genomic prediction methodology under-estimate phenotypic performance.

In the LYC project it was noted the two 1950 era sires PTA exceeded breed average for milk production by two standard deviations. Blackburn (2019) reported many breeds with animals in the collection exceed the mean BV when compared to other animals of their birth year by an equally large deviation. This finding suggests that while a particular animal may be chronologically older its genetics may be more useful than expected and when mated to above average females the gap between old and new genetics can be narrowed considerably. Generally, we have observed across species that animals in the collection that are outliers for BV may be above breed average for 4 to 6 future generations (8 to 30 years), depending upon selection pressure and generation interval within the breed. This observation has ramifications for how such individual animals might be utilized.

In summary, the collection has been widely used and made substantial contributions supporting its continued growth and development. Future challenges will be to better understand the genetic aspects of the collection, to use the collection effectively in corrective mating programs, and to encourage scientists and industry to increase their use of the collection for any variety of purposes.

References

Blackburn, H. D., Y. Plante, G. Rohrer, E. W. Welch, S. R. Paiva (2014) J. Anim. Sci. 92:1401-1411.

Blackburn, H. D. (2018) Ann. Rev. Animal Biosciences 6:69-82.

Blackburn, H. D., C. S. Wilson, B. Krehbeil (2019) Diversity 11(12), 144.

Dechow, C. D., W. S. Liu, L. W. Specht, and H. Blackburn. 2020. J. Dairy Sci. 103:4510-4516.

FAO (2012) Cryoconservation of animal genetic resources. FAO Animal Production and Health Guidelines No. 12. Rome.

Jewell, S. 2021. Exploring the Performance of Unique Genetic Lineages of Dairy Cattle. BS Honors Thesis. https://honors.libraries.psu.edu/files/final_submissions/7259.

245. Mapping genetic diversity in European gene banks: preliminary results on chickens for the validation of IMAGE001 array

M. Tixier-Boichard[1*], N. Peynot [1], D. Duclos[2], S. Weigend[3], L. Monteagudo[4], A. Martinez Martinez[5], J.V. Delgado[5], R. Gonzalez Prendes[6], R. Crooijmans[6] and G. Restoux[1]

[1]Université Paris-Saclay, INRAE, AgroParisTech, GABI, 78352 Jouy-en-Josas Cedex, France; [2]Institut de l'Elevage, 75595 Paris Cedex 12, France; [3]Friedrich-Loeffler-Institut, Institute of Farm Animal Genetics, 31535 Neustadt, Germany; [4]University of Zaragoza, Facultad de Veterinaria, 50013 Zaragoza, Spain; [5]University of Cordoba, Animal Breeding Consulting, S.L, 14071 Cordoba, Spain; [6]Wageningen University and Research, Animal Breeding and Genomics, 6708 PB Wageningen, the Netherlands; michele.tixier-boichard@inrae.fr

Abstract

Gene banks are a component of a national strategy for the preservation of genetic diversity. Gene bank managers need to have a global and comparable picture of the diversity in their collections in order to rationalize them. Facing a diversity of molecular tools is a difficulty. The IMAGE H2020 project aimed at developing a low cost 60k SNP array to facilitate the mapping of diversity in gene banks. The first test of this array was performed for chicken with 204 samples from 18 local breeds and nine experimental lines provided by Germany, France and Spain. The MAF across population was 0.34, showing that this tool is useful over a range of populations. The principal component analysis and the Neighbor-joining tree showed that local breeds did not cluster according to country and were generally homogenous. Comparison with on-farm populations remains to be done to assess the value of the gene bank collections.

Introduction

Genetic diversity is essential for the sustainability of animal breeding. Countries are investing in long term conservation of genetic diversity with the creation of gene banks. FAO promotes this strategy in the Global Plan of Action for Animal Genetic Resources. In Europe, the European Regional Focal Point is gathering national coordinators of animal genetic resources from all countries and includes a permanent working group on ex situ cryoconservation.

Until now, sampling strategies remain national and may vary according to countries, resulting in a variety of situations in terms of species and breeds represented. The ERFP is currently promoting the set up of a European network of Animal Gene Banks (EUGENA) to increase collaboration between national gene banks in Europe. An economic analysis of gene banking was recently performed in the frame of the IMAGE H2020 project (Grant # 677353). Results showed that the cost of preservation of a given number of breeds could be decreased by 20% by a better coordination across European countries (de Oliveira Silva *et al.*, 2018).

Assessing the value of gene bank collections for breeding programmes require data on within-breed diversity as well as between-breeds, since both levels are important for the management of farmed animal species. Whereas pedigree data have been used since a long time to characterize within-breed diversity, for *in situ* populations as well as *ex situ* collections, only molecular data make it possible to map the diversity at all levels within a species.

Mapping diversity has made significant progress thanks to the development of genotyping arrays in all farmed species, whereas whole genome sequencing can provide a deeper analysis of the status of collections. However, sequencing costs are still relatively high for gene bank managers, who generally lack the capacity to analyse such data. Thus, genotyping arrays are still a relevant tool to provide a global description of the

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_245, © M. Tixier-Boichard *et al.* 2022

diversity present in gene bank collections and to make comparisons between gene banks as well as between gene bank collections and on farm populations. However, several SNP arrays for chickens coexist, either private (around 60K) or public (around 580K), at a rather high cost (from 90 to 150 € per animal), with a low % of markers shared between the arrays. This is complicated for gene bank managers who need a global and comparable picture of the diversity in their collections. To facilitate a common approach across gene banks, the IMAGE H2020 project has developed a multi-species 60k array, with the aim to provide 10k markers for each of six species, i.e. cattle, sheep, goat, horse, pig and chicken, at a price of 20 US$ per sample maximum for both the array and the genotyping, thus excluding the cost of DNA extraction.

A subset of gene bank collections from different countries was genotyped to validate the IMAGE001 array. In this paper, we show how the first version of IMAGE001 can map diversity of chicken collections from gene banks in Germany, France and Spain.

Table 1. List of chicken populations included into the genotyping test for IMAGE001 array.

Type of population	Breeds	Samples	Country
Experimental lines	Congenic line_LB13 (MHC)	2	France
	Congenic line LB21 (MHC)	2	France
	Congenic line LB4 (MHC)	4	France
	Congenic line_LB19 (MHC)	3	France
	DPF-(low duration of fertility line)	4	France
	DPF+ (high duration of fertility line)	5	France
	DWNA (dwarf naked neck layer)	5	France
	Fat line(high abdominal fatness)	5	France
	Lean line Low abdominal fatness)	5	France
Local breeds	High quality Bresse chicken (LB99)	5	France
	Barbezieux	3	France
	Gasconne	3	France
	Gauloise_Doree	12	France
	Blue_Andalusian	10	Spain
	Deutsches Reichshuhn	10	Germany
	Krüper	10	Germany
	Deutsches Lachshuhn	10	Germany
	Langshan	10	Germany
	Ostfriesische Moewen	10	Germany
	Rheinlaender	10	Germany
	Sachsenhuhn	10	Germany
	Sundheimer	10	Germany
	Westfaelische Totleger	10	Germany
	Augsburger	11	Germany
	Bergische Schlotterkamm	11	Germany
	Deutsche Sperber	11	Germany
	Gallina_del_Sobrarbe	20	Spain

Materials & methods

Gene bank samples. A total of 211 chicken DNA samples were provided for 27 breeds from the gene banks of three countries (Table 1). Depending on the breed, the DNA had been extracted from frozen semen or from blood. Quality and concentration were determined with spectrophotometry and at least 1.5 µg of DNA was provided to the IMAGE partner at Wageningen University who was coordinating the test. Sampling represented only local breeds for Germany and Spain, with 10 to 20 animals per breed, whereas experimental lines were also provided by France, with a lower number of animals (2-5) per population (Table 1).

SNP selection. SNP selection was performed based on: (1) overlap with existing arrays with a high allele frequency across populations; (2) location in genes affecting phenotypic traits; (3) mtDNA; (4) Ancestral SNPs as compared to wild species of jungle fowl; (5) MHC region; (6) random genes located into QTL regions. A total of 9,306 SNPs were initially selected to be tested for the chicken, covering 32 autosomes (9,071 markers) and the sex chromosomes (235 markers).

Genotyping. Genotyping was performed on an Affymetrix platform by Eurofins® biotech company.

Analysis. Analysis was performed with *plink* software. The Minor Allele Frequency (MAF) was calculated per chromosome and per breed. Autosomal markers were used to determine the genetic structure with a Principal Component Analysis and a neighbour-joining tree.

Results

Genotyping. Altogether 7,743 autosomal markers passed the calling rate and were included in the analysis. Seven samples did not provide satisfactory results, leading to a total of 204 individual genotypes. The mean MAF per chromosome was above 0.30 for most of the autosomes, exceptions being chr 16 (0.03), chr 25 (0.27), chr 27 (0.20) chr 31 (0.02) and chr 33 (0.04). The overall MAF across all populations was 0.34. At population level, the median MAF was above 0.25 except for three congenic lines (LB4, LB13, LB19), which were less variable.

Principal component analysis. The first four principal components (PC) explained only 16% of the total variance. PC1 explained 7%, populations were scattered across a gradient from Langshan breed to congenic lines, Westfaelische Totleger and Ostfriesische_Moewen breeds. PC2 isolated the Gallina_del_Sobrarbe from other breeds, PC3 isolated the Deutsche Lachshuhn and PC4 isolated the Andalusian.

Neighbor-joining tree. The NJ tree (Figure 1) featured 4 main branches: branch A included only one Spanish breed (Blue Andalusian) and one German breed (Augsburger); branch B included three German breeds (Rheinlander, Westfaelische Totleger and Ostfriesische Moewen) and the four congenic lines which are all of White Leghorn origin; branch C included three German breeds (Krüper, Deutsche Sperber and Bergische Schlotterkamm); branch D included the remaining German breeds, most animals of the Gallina del Sobrarbe, the French local breeds and the experimental lines of a non White Leghorn origin. The divergent lines such as Fat/Lean lines and DPF+/DPF- clustered together.

Discussion

First of all, the high MAF value observed across all populations demonstrates the ability of the array to be used over a large range of populations. The total number of markers was quite reduced and another set of markers will be added to reach the target of 10k markers. Yet, a few chromosomes, (16, 31, 33) are likely to remain underrepresented in the array.

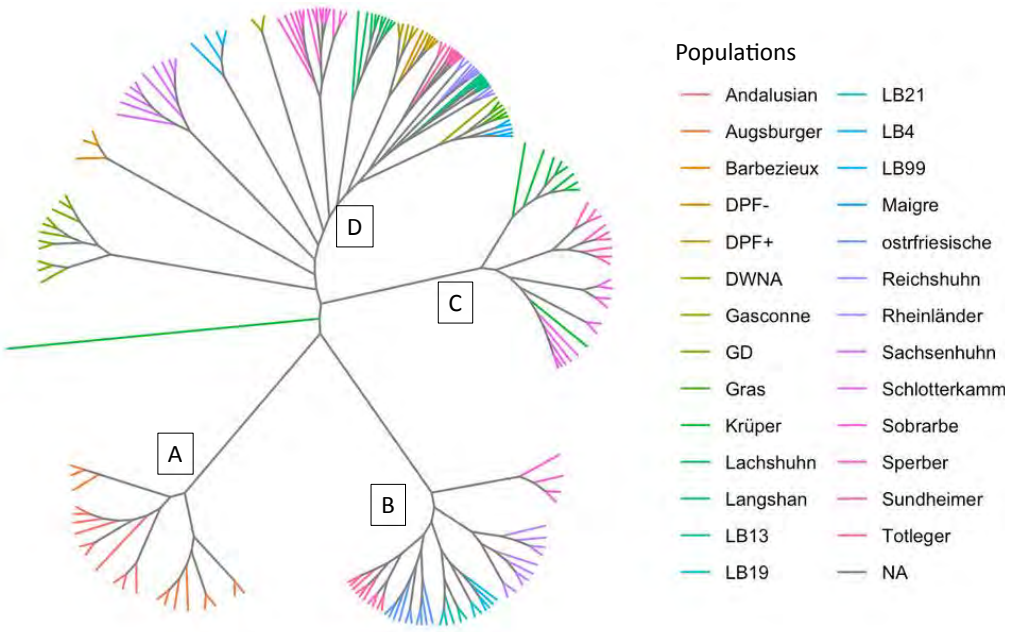


Figure 1. Neighbor-Joining tree of the populations. (A, B, C, D are discussed in the text).

The populations did not particularly cluster according to country, which shows the interest of sharing information across gene banks to assess their complementarity. This finding, however, may result from the unbalanced distribution of breeds according to country in the data set. There was only one branch made up of breeds from a single country (Germany) which is consistent with the fact that these breeds are typical Northwestern European breeds and quite different from Mediterranean breeds such as the White Leghorn. Enlarging the sampling to Northern Europe would make possible to check this hypothesis. More information is needed on the history of the two Spanish breeds to understand their separate clustering pattern in the NJ tree. Besides, the history of breeds from different countries which are located within the same branch of the NJ tree also needs to be further explored.

Each population appeared in a single branch of the NJ tree, with a few exceptions, such as the Gallina del Sobrarbe, for which most animals appeared mainly in branch D but a few were in branch B. Local chicken breeds managed by a breeders'club generally appear to be homogenous, unless the breed is distributed across farmers who do not exchange their animals, thus leading to genetic fragmentation.

In conclusion, using a single SNP array for mapping diversity in gene bank collections is feasible and the next step is to include data from on-farm populations in order to map the diversity within and outside gene banks.

Acknowledgements

This work was funded by H2020 IMAGE grant #677353.

References

De Oliveira Silva R., Ahmadi B.V, Hiemstra S.J., and Moran D. (2018) J. Anim. Breed. 136(1): 63 -73. https://doi.org/10.1111/jbg.12368

246. Conservation of an endangered pig breed using optimum contribution selection

H.M. Nielsen* and M. Kargo

Center for Quantitative Genetics and Genomics, Aarhus University, Blichers Alle 20, 8830 Tjele, Denmark, hannem.nielsen@qgg.au.dk

Abstract

The present study focuses on optimizing the breeding plan for an endangered Danish Landrace pigs. This original landrace pig breed (DL-1970) was until 1970 used for production of bacon, after which the breed was crossed with other landrace breeds. This study deals with the animals from DL-1970, which originates from the breed before crossing. Seven boars and 46 sows were available for selection. Optimum contribution selection (OCS) was applied to select boars and sows. Average relatedness between male candidates was 0.35, between female candidates 0.45 and between male and female candidates 0.39. Using OCS yielded lower inbreeding levels (0.141) compared to random mating (0.195). Due to the high level of relatedness between the animals and the very small population size it is recommended to use material from the gene bank to increase genetic variation in the future.

Introduction

The DL-1970 is a distinctive feature of Danish agricultural history, as the great export adventure with bacon in the first half of the last century was based on this pig breed. Around 1970, however, the breeding goal for Danish Landrace changed to having increased focus on growth, feed efficiency, fertility and meat quality (Sørensen and Nielsen, 2017). In the 1980s, crossbreeding was also performed between Danish Landrace pigs and landrace pigs from Norway and Finland. Today there are only a few individuals left of the original DL-1970 as it was in 1970 before crossing and change of the breeding goal. Therefore, the breed's existence is currently threatened. If a local breed is lost, a cultural value will be lost (Meuwissen *et al.*, 2009). DL-1970 have a large cultural value in Denmark, since this breed founded the Danish Bacon adventure and most pigs in the current production are founded on this bred. If the breed is to survive and increase in population size, it is important that there are optimal production methods for the breed as well as possible sales opportunities for the finishers. Last but not least, it is important to maintain the genetic variation as large as possible (Meuwissen, 2009). The present study focuses on optimizing the breeding plan for DL-1970 by use of Optimum contribution selection (OCS) for selection of sires and dams.

Materials & methods

Data. The data set included the identity (id) of the animal, id of the sire, id of the dam, sex of the animal and date of birth. There were 1179 individuals in the pedigree with the oldest animal born in 1983 and the youngest born in July 2019. Pedigree completeness (the proportion of known ancestors in each ascending generation, MacCluer *et al.*, 1983) was calculated by tracing the pedigree 8 generations back. Potential breeding candidates were seven boars and 46 sows born from January 2018 until July 2019. Matings were assumed to occur naturally. Thus, for the optimum contribution selection it was assumed that each boar could mate a maximum of 10 sows.

Optimum contribution selection. OCS maximizes the genetic level of the next generation of animals while at the same time restricting rate of future inbreeding by restricting the increase in average co-ancestry (Meuwissen, 1997). We used the EVA software (Berg *et al.*, 2006) to perform OCS. The function to maximize in OCS with overlapping generations was shown in Henryon *et al.* (2015). With full weight on average relationships and minimization of inbreeding, as applied in the current study, the function is:

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_246, © H.M. Nielsen and M. Kargo 2022

$$U_t(c) = \frac{\omega}{L^2}(c + Pv)'A(c + Pv) \quad (1)$$

where,

c is a vector with n (total number of animals in the pedigree) genetic contributions to the new cohort, ω is a penalty applied to the average relationship of the current population, L is the generation interval, P is a matrix of size n×k with contributions to each age class of animals in the current generation, k is the number of sex-age classes in the current generation, v is a vector with k expected relative contributions to future age classes, and A is an n×n matrix with additive genetic relationships.

Results

From 1980, pedigree completeness (Figure 1) was low, increasing to around 80% after 2000, when tracing the pedigree back 8 generations.

Figure 2 shows the average inbreeding of the progeny from year 1980 to 2019. The average inbreeding was high, up to 0.33 in year 2002. During the last 5 year period, the average inbreeding was around 0.15 to 0.20.

The average relatedness between male candidates were 0.35, between female candidates 0.45 and between male and female candidates 0.39. By applying OCS for selection of 20 matings, only 3 out of the 7 boars were selected (Table 1). Their contributions varied from 10 to 50%. Thus 2 sires had high contributions, which may be a problem in the future even with the use of OCS. Likewise, OCS selected 20 out of 46 available female candidates. It is clear from Table 1 that the boars selected are those with lowest relationship with the rest of the female populations.

Under random mating, expected average inbreeding was 0.195, whereas expected average inbreeding was 0.141 when applying OCS. In the last year (2017) observed with selection in the dataset, average inbreeding was 0.196. This means that by applying OCS it is possible to reduce average inbreeding in the progeny by around 0.05 compared to the previous year.

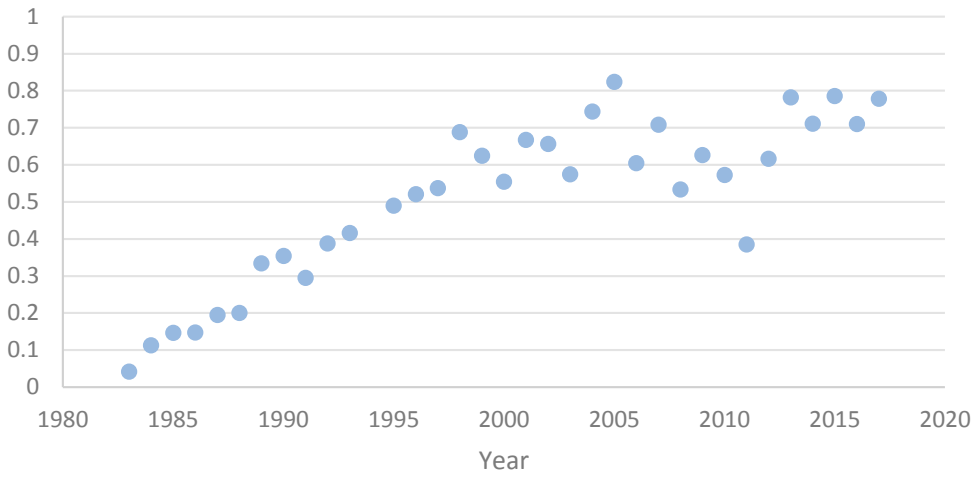


Figure 1. Pedigree completeness index (8 gen) from year 1983 to 2019.

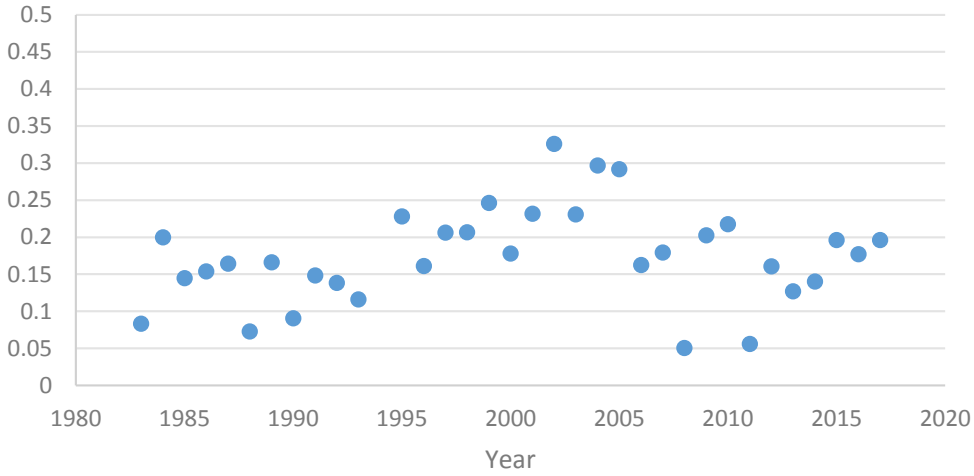


Figure 2. Average inbreeding of progeny over years.

Table 1. No of contributions (matings) for each of the sires and its relationship with the male and female population.

Boar	No of matings	Relationship with males	Relationship with females
1	0	0.405	0.386
2	10	0.424	0.260
3	2	0.439	0.281
4	0	0.484	0.447
5	0	0.507	0.548
6	8	0.424	0.266
7	0	0.530	0.543

Discussion

In the present study we used OCS to optimize the breeding plan for an endangered pig breed.

Using OCS yielded substantial lower inbreeding levels compared to random mating. However, the level of relatedness between the animals is very high, which means that the population is critically endangered.

In addition to using OCS, one option would be to use material from the gene bank (Hulsegge *et al.*, 2019), which contains both semen and embryos. This stock of sperm and embryos in the gene bank was established as a safeguard that can be used if the living population is critically endangered and needs to be able to increase the genetic variation. Freezing of semen and embryos from the breed began as early as in 1996. There is semen from a total of 11 boars produced in 1996, 2001 and in 2017. Seven of the 11 boars are from a previous control line, which has shown to be only distantly related to the current population of DL-1970 (Szekeres *et al.*, 2016). In addition, there are embryos from 16 sows and 13 boars. The living population of Danish Landrace pigs is at present at a size where use of the material in the gene bank is needed. Thus, the next step is to run OCS using semen and embryos from the gene bank.

References

Berg P., Nielsen J. and Sørensen, M.K. (2006).Proc. of the 8th WCGALP, Belo Horizonte, Brazil.

Henryon, M., Ostersen, T., Ask, B, Sørensen, A.C. and Berg, P. (2015). Genet. Sel. Evol. 47: 21. https://doi.org/10.1186/s12711-015-0107-7

Hulsegge, I., Calus, M., Hoving-Bolink, R., Lopes, M. Megens, H.-J *et al.* (2019). *Genet Sel Evol (2019) 51:60.* https://doi.org/10.1186/s12711-019-0502-6

MacCluer, J.W., Boyce, A.J., Dyke, B., Weitkamp, L.R., Pfennig *et al.* (1983). J. Heredity. 74: 394-399. https://doi.org/10.1093/oxfordjournals.jhered.a109824

Meuwissen, T.H.E. (1997). J. Anim. Sci. 75:934–940. https://doi.org/10.2527/1997.754934x

Meuwissen, T. (2009). Genetic management of small populations: A review. Acta Agric. Scand. Section A. 59:71-79. https://doi.org/10.1080/09064700903118148

Sørensen, L.H., and Nielsen, V.H. (2017). Danske husdyrgenetiske ressourcer. DCA report No. 100. 52 pp. (in Danish).

Szekeres, B., Schönherz, A, Nielsen, V.H. and Guldbrandtsen, B. (2016). Gamle danske husdyrracers genomer – racernes oprindelse og slægtskab. DCA report No. 082. 60 pp. (in Danish).

247. IDI: an index to assess the usefulness of cryopreserved collections for the management of animal genetic diversity

A. Jacques[1*], D. Duclos[2], C. Danchin-Burge[2], M.J. Mercat[3], M. Tixier-Boichard[1] and G. Restoux[1]

[1]Université Paris-Saclay, INRAE, AgroParisTech, GABI, 78350 Jouy-en-Josas, France; [2]Institut de l'Elevage, 149 rue de Bercy, 75595 Paris Cedex 12, France; [3]Institut du porc, La Motte au Vicomte, 35651 Le Rheu Cedex, France; alicia.jacques@inrae.fr

Abstract

Confronted with the erosion of genetic diversity in domestic animal populations, many countries such as France have created gene banks to limit this loss. However, it is crucial to evaluate the potential of these collections before using them for genetic diversity management. Thus, we propose a new Index of Diversity Impact, IDI, to evaluate and compare the diversity potential of collections. IDI is negative when the use of collections decreases the overall relatedness of the current population and positive when it increases it. IDI was used to evaluate the cryopreserved collections of the French National Cryobank of 14 breeds of three livestock species, pig, sheep and cattle. We obtained significantly negative IDIs for three breeds that differ in terms of collection strategy and population management. The IDI appears to be a relevant tool to help managers and breeders to quickly identify promising collections to be used in future breeding programmes.

Introduction

Genetic diversity is essential for the sustainability of livestock populations by allowing for continuous response to selection and for adaptation to new conditions. However, selection has led to an erosion of genetic diversity over time, which can limit further genetic gain. In addition, for small populations, such as local breeds, the effect of genetic drift is even more important and leads to accelerated depletion of genetic diversity within the breed.

To address this risk, many countries have invested in conserving genetic diversity through the creation of gene banks. These *ex situ* conservation entities have been storing genetic material for several years, including reproductive material such as semen. However, these genetic resources are still underused, even though they could have a potential interest for managing breed diversity. In their study, Leroy *et al.* (2011) showed that the use of frozen semen from cryobanked bulls can be a valid choice for introducing genetic diversity, provided that attention is paid to gene diffusion, or when there is a change in breeding goals. Two other studies showed that the use of cryopreserved semen from ancient bulls could introduce genetic diversity into a local breed under selection (Eynard *et al.*, 2018) but also in an international breed (Doekes *et al.*, 2018).

France is a country with a rich livestock biodiversity, with for instance 47, 57 and 12 French breeds officially recognized in 2015 for cattle, sheep and pigs, respectively, but a majority of local ones are threatened (Verrier *et al.*, 2015). French breeds greatly differ in terms of population size and management programme, with, for example, the Prim'Holstein, first French dairy breed under selection, having nearly 3,400,000 females born in the period 2016-2019 versus the local Tarentaise dairy breed, having only 17,266 females in the same period (Danchin-Burge, 2020). This raises the question of genetic erosion and motivated the creation of the French National Cryobank, FNC, in 1999, in order to preserve the reproductive material from French breeds of domestic animals. Currently, the FNC brings together reproductive resources of 21 species and nearly 8,000 donors with stored material over the last 20 years. However, until today, the use of these available resources to manage the genetic diversity of current populations is very scarce. Thus, this study aims at characterizing the collections of the FNC and evaluating their potential use for the management of breed diversity in future breeding programmes.

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_247, © A. Jacques *et al.* 2022

Materials & methods

Data for individuals in the national cryobank. Donor animal data were obtained from the FNC database (www.cryobanque.org). We focused on three species of livestock that are likely to use frozen semen in their management: cattle, sheep and pig. The national identifier (unique identification in a national pedigree database), the breed and the year of birth of donor animals were recorded.

Breeds, pedigree data and definition of cohorts. A total of 14 breeds were chosen to cover a large range of situations encountered in France, encompassing two pig breeds, the Pietrain (PIE) and the Cul Noir du Limousin (CNL), five sheep breeds, the Dairy Lacaune (LAC), the Manech Tête Noire (MTN), the Manech Tête Rousse (MTR), the Basco-Béarnaise (BAS) and the Mouton Vendéen (VEN), and seven cattle breeds, the Prim'Holstein (HOL), the Montbéliarde (MON), the Abondance (ABO), the Froment du Léon (FRO), the Tarentaise (TAR), the Saosnoise (SAO) and the Maraîchine (MAR). We defined contemporary male and female cohorts, M and F respectively, as individuals with at least one known parent, born between 2018 and 2020 for pigs, between 2017 and 2020 for sheep, and between 2015 and 2020 for cattle, according to the species' generation interval. These contemporary cohorts are assumed to be currently alive and available for breeding. A cohort composed of all the cryopreserved sires present in the FNC for each breed, C, was also defined. For the pig breed (PIE), 13 boars with semen cryopreserved by the breeding companies were added to our study. The pedigrees of all individuals in the three cohorts were extracted from the national database and limited to the 25 last generations. The pedigree quality of the contemporary male and female cohorts were evaluated with the overall equivalent generations as the weighted average of the generation equivalents (EqG) for each year using PEDIG (Boichard, 2002).

Computation of kinship and definition of an index of diversity impact. The calculations of the average kinship between the three cohorts were carried out using the *colleau* module of PEDIG (Colleau, 2002). For each breed, we calculated the average kinship between all the sires present in the cryobank (cohort C) and the contemporary female population (cohort F) ($\overline{\phi_{C-F}}$); and between the contemporary male and female cohorts ($\overline{\phi_{M-F}}$). The average kinship within cohort C between cryobank sires ($\overline{\phi_{C}}$) was also calculated. Then, we defined an Index of Diversity Impact (IDI) for cryobank individuals. This indicator is an estimation of the expected impact of the use of cryobank individuals in random mating with contemporary females on the level of diversity of the next generation under the assumption that only the males in the cryobank would be utilized. It is calculated following the Equation 1.

$$IDI = \left(\overline{\phi_{C-F}} - \overline{\phi_{M-F}}\right) \Big/ \overline{\phi_{M-F}} \quad (1)$$

It is expressed as a percentage of loss (negative) or gain (positive) in mean kinship relative to the current kinship of the population. The confidence intervals (CI) at 95% were defined from the individual IDI of each cryobank individual and applying Equation 2.

$$CI = \left[\overline{IDI} \pm 1,96 \frac{sd(IDI)}{\sqrt{n}}\right] \quad (2)$$

With *sd(IDI)* the standard deviation observed within the cryobank population of *n* donors.

All statistical analyses were conducted using R, version 4.0.2 (R Core Team, 2019).

Results

Composition of the cryobank's collections. For the three species, a total of 84 breeds are represented in the FNC, with 21 breeds for cattle, 15 for pigs and 48 for sheep. This corresponds to 977 bulls, 451 boars and 1,177 rams with frozen semen. The average number of donors per breed is 46.5 for cattle (min=1, max=210,

se=13.4), 30.1 for pig (min=10, max=76, se=4.4) and 24.5 for sheep (min=1, max=302, se=6.4). Donor birth years ranged from 1959 to 2013, 1979 to 2018 and 1974 to 2019 for cattle, pig and sheep respectively. This shows that material from ancient sires was retrieved by the FNC after 1999. The range of birth years of donors varied according to species and breed (Table 1).

Genetic diversity and IDI for the 14 breeds analysed with pedigree data. The number of donors in the cryobank for the 14 selected breeds ranged from 10 bulls in the SAO to 302 rams in the LAC. The size of the contemporary cohorts was very variable according to breed, with the number of females ranging from 534 cows for the FRO to 5,069,731 for the HOL.

The quality of the pedigrees ranged from 3.1 to 19.1 EqG on the female pathway and from 3.1 to 19.3 EqG on the male pathway (Table 1). The mean within-cryobank kinship values ranged from 0.8 to 24.3%. The average kinship values between cryobank sires and contemporary females (cryo-F) ranged from 3.1 to 24.9%, and the average kinship values between contemporary males and females (M-F) ranged from 1.9 to 36.1%. The IDI values varied between -31% for the CNL and 66.5% for the SAO. Three breeds exhibiting different situations have a significantly negative IDI: a local pig breed with quite ancient sires in the cryobank (CNL), a moderately selected sheep breed with a large number of males in the cryobank as compared to contemporary males (VEN), and a highly selected cattle breed (HOL) for which the cryobank favored donors with extreme breeding values and includes old material, thus, increasing the genetic distance with the current population. The 11 remaining breeds have a positive or null IDI.

Discussion

With more than 8,000 donor animals, the FNC needs to know the potential value of its collections for the management of genetic diversity, all the more that ancient sires are represented in the collections and that living populations can suffer from selection and drift. For sheep, the collections include donors born

Table 1. Table of genetic diversity measures for the 14 breeds studied by pedigree data.

	Cryobank sires		Contemporary cohort size[1]		No. of EqG [2]		Average kinship (in %)			Index of diversity impact (in %)	
Breed	No.	Birth year range	M	F	M	F	C	C-F	M-F	IDI	[CI][3]
SAO	10	1991;2011	3,084	3,089	3.2	3.3	00.8	03.1	01.9	66.5	[25.5;107.5]
FRO	12	1978;2006	536	534	3.1	3.1	10.1	12.8	11.3	13.5	[-08.2;35.2]
TAR	75	1994;2006	25,781	26,030	6.5	6.6	10.6	09.3	08.6	08.5	[03.1;14.0]
ABO	66	1995;2006	470	71,698	6.7	6.8	15.1	12.9	12.0	07.2	[01.1;13.4]
MON	142	1994;2013	3,658	1,201,388	9.4	9.5	12.1	10.8	10.8	00.0	[-03.0;03.0]
MAR	13	1989;2004	3,195	3,368	6.7	6.5	05.7	07.6	08.0	-04.7	[-28.5;19.2]
HOL	190	1993;2013	10,776	5,069,731	10.4	10.0	10.8	08.9	10.1	-12.6	[-14.7;-10.6]
PIE	27	2016;2017	49,073	48,503	19.3	19.1	12.8	12.7	12.4	02.3	[00.4;04.1]
CNL	32	1976;2013	1,243	1,367	15.5	15.5	24.3	24.9	36.1	-31.0	[-35.4;-26.6]
MTN	26	2007;2016	1,209	11,977	5.9	4.5	11.3	06.0	04.1	44.1	[34.1;54.1]
MTR	48	2008;2016	6,546	104,184	9.7	8.2	06.8	05.5	05.1	07.2	[-01.3;15.7]
LAC	302	1980;2018	19,206	708,459	13.3	10.8	05.7	03.5	03.4	03.3	[00.3;06.2]
BAS	32	2009;2016	2,418	32,547	8.4	6.1	10.1	06.8	06.6	03.2	[-04.9;11.3]
VEN	69	1989;2012	306	33,855	13.7	10.3	05.2	03.7	04.4	-16.3	[-20.8;-11.9]

[1] Number of individuals with at less one parent known in the contemporary cohorts.

[2] Number of Equivalents generations in the pedigree computed for the contemporary cohorts.

[3] Confidence intervals at 95%, computed from the individual IDI of each cryobank individual.

fairly continuously over time since 1985. For pigs, local breeds have ancient donors whereas breeds under selection have quite recent donors. For cattle, few donors are born after the 2010s following the switch to genomic selection, which considerably shortened the reproductive carrier of males and decreased the quantity of material available for long term storage.

For the breeds studied, the number of donors in the collection is not correlated with the average intra-breed relatedness, which rather depends on the quality of the pedigree of the individuals in collections and the breed management programmes. The IDI reflects the average impact of randomly mating sires in collections with contemporary females relative to expectations for the use of contemporary males only; and makes it possible to rank the usefulness of collections resulting from different sampling strategies. A positive or null IDI reveals a representative collection of the current population, either by the presence of the strongest contributors to the breed (12 AI bulls in collection out of the 13 used in the FRO), or by recent collections likely more related to current females (example of PIE or BAS). However, a positive index does not definitively mean that the use of cryobanked sires is not useful for genetic diversity management, but that on average there will be fewer individuals in the collection that are genetically distant from the current population. In these cases, large confidence intervals suggest that the use of some well-chosen cryobank sires could produce matings with a lower level of relatedness than with contemporary sires, even if the IDI is positive, as for FRO or MON.

In conclusion, the IDI summarizes the impact, on genetic diversity, of using cryopreserved individuals relative to the current level of genetic diversity of the population; it is thus a convenient tool for managers to rapidly screen collections. In our study, IDI values were obtained from pedigree data but could also be calculated from genomic data, especially to be applicable to local breeds without pedigrees. The next step is to analyse options for individual mating plans between cryobank sires and contemporary females.

References

Boichard D. (2002) Proc. of the 7th WCGALP, Montpellier, France.

Colleau J.J. (2002) Genet. Sel. Evol. 34:409–421. https://doi.org/10.1051/gse:2002015

Danchin-Burge C. Indicateurs de variabilité génétique – races bovines – Edition 2020. Available at: https://idele.fr/fileadmin/medias/Documents/2020_varume_BL.pdf

Doekes H.P., Veerkamp R.F., Bijma P., Hiemstra S.J., and Windig J. (2018) J. Dairy Sci. 101(11):10022–10033. https://doi.org/10.3168/jds.2018-15217

Eynard S.E., Windig J.J., Hulsegge I., Hiemstra S.J., and Calus M.P.L. (2018) J. Anim. Breed. Genet. 135(4):311–322. https://doi.org/10.1111/jbg.12333

Leroy G., Danchin-Burge C., and Verrier E. (2011) Genet. Sel. Evol. 43:36. https://doi.org/10.1186/1297-9686-43-36

R Core Team. R: A language and environment for statistical computing. Available at: https://www.R-project.org/

Verrier E., Audiot A., Bertrand C., Chapuis H., Charvolin E. *et al.* (2015) Anim. Genet. Resources 57:105–118. https://doi.org/10.1017/S2078633615000260

248. Conservation through utilization in German Red and White dual-purpose cattle

M. Slagboom[1*], V. Milkevych[1], H. Liu[1,2], J.R. Thomasen[3], M. Kargo[1] and C. Schmidtmann[4]

[1]Aarhus University, Centre for Quantitative Genetics and Genomics, Blichers Alle 20, 8830 Tjele, Denmark; [2]current affiliation: SEGES Cattle, Agro Food Park 15, 8200 Aarhus N, Denmark; [3]VikingGenetics, Ebeltoftvej 16, 8960 Randers SØ, Denmark; [4]Christian-Albrechts-University Kiel, Institute of Animal Breeding and Husbandry, Hermann-Rodewald-Straße 6, 24118 Kiel, Germany; margotslagboom@qgg.au.dk

Abstract

Smaller Red dairy cattle breeds such as German Red and White dual-purpose cattle (RDN) need to be improved genetically to be competitive. One way to achieve this, is by collaboration with a larger mainstream breed such as Red dairy cattle from Denmark, Finland and Sweden (RDC). The aim of this paper was to increase genetic gain in the RDN breeding program by breeding with RDC bulls. Four scenarios with different levels of collaboration between the two breeds were compared in terms of genetic gain, inbreeding, and breed proportions. The results indicate that for 10 years of selection, restricted selection of 20% RDC bulls allows for a double as high proportion of RDN bulls while achieving the same rate of genetic gain as in scenarios with unrestricted selection. However, for a longer time span, the proportion of RDN will likely reduce even more.

Introduction

Red dairy cattle breeds are kept all throughout Europe and can be divided in larger, mainstream breeds such as Red dairy cattle breed from Denmark, Finland and Sweden (RDC), and smaller local breeds such as the German Red and White dual-purpose cattle (RDN) (Schmidtmann, 2021). Conservation of local breeds that are numerically smaller may be achieved by increasing genetic gain in these breeds. Increasing the breed level may be possible by collaboration with a mainstream Red cattle breed such as RDC. Collaboration between these breeds means that RDN will have an increased accuracy of genomic selection due to an increase in the size of the reference population. In addition, there will be more selection candidates, which increases the selection intensity. The RDC breed may also profit from this collaboration, as RDN form a source of genetic diversity. The aim of this paper is to increase genetic gain in the RDN breeding program by breeding with RDC bulls, while keeping the genomic proportion of RDN in RDN bulls as large as possible.

Materials & methods

Four scenarios were simulated with a difference in the origin of bulls selected for breeding with RDN. All breeding bulls that were selected for breeding in the RDN breeding program originated from the RDN breeding program in scenario 'Separate'. In scenarios 'Restricted-20%' and 'Restricted-50%', 20 or 50% of the bulls selected for breeding to the cows in RDN could originate from RDC. All bulls selected for breeding in the RDN breeding program could originate from RDC in scenario 'Unrestricted'. Each scenario was replicated 20 times.

Simulation. A full genomic set-up was simulated in a newer version of stochastic simulation program ADAM (Pedersen *et al.*, 2009) that was updated for multi-breed simulations. Sequence data were available for 38 RDN and 146 RDC. These sequence data were reduced to the SNPs on the Illumina BovineSNP50v3 Beadchip (Illumina Inc., San Diego, USA). The data were split by chromosome and phased. The obtained haplotype data were used as input for the simulations. Simulated time was 30 years, in which the first 20 years were used to build up reference populations and the last 10 years were used to run a genomic

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_248, © M. Slagboom *et al.* 2022

selection scheme with varying numbers of RDC breeding bulls that were selected for breeding with RDN. In the progeny testing scheme (year 1-20), several 1-year-old bulls were genotyped and progeny tested 3 years later. At the end of the progeny testing scheme, the reference population of RDN consisted of 1,200 proven bulls and for RDC this number was 10,000. The genomic selection scheme (year 21-30) included genotyping young males and females every year, as well as multiple ovulation and embryo transfer. Breeding values were estimated using the DMU package (Madsen and Jensen, 2013).

Breeding goals. The simulated breeding goals (BG) included two traits: milk yield and udder health. Heritabilities were based on estimates found in literature of 0.48 and 0.04 (International Bull Evaluation Service, 2012, 2014), and the genetic correlation between milk yield and udder health was -0.32 (Norberg *et al.*, 2009). The BG weights for RDC were set to match the correlation of the milk production index of the Nordic Total Merit index (NTM) with the whole NTM (Sørensen *et al.*, 2018), with a correlation of 0.75 between milk yield and the BG, and a correlation of 0.38 between udder health and the BG. The BG for RDN was set to have more emphasis on health, with a correlation of 0.64 between milk yield and the BG, and a correlation of 0.53 between udder health and the BG.

Analysis. The outcome of the simulation was analysed in R statistical software (R Core Team, 2021), for year 26-30 of the simulation. Rate of inbreeding was calculated per generation, both for pedigree-based inbreeding and for IBD-based inbreeding. Rate of genetic gain was calculated per year and per trait in genetic standard deviation (σ_A) units. In addition, genetic gain in the aggregate genotype was calculated by using the BG weights and genetic gain per trait. An ANOVA test followed by Tukey's Honest Significant Differences test was used to find significant differences between scenarios. The percentage of bulls that originated from each breeding program and selected for breeding in the RDN breeding program was calculated per year for year 21-30. The breed proportion based on pedigree data of RDN of bulls selected for breeding in the RDN breeding program was calculated by use of the OptiSel package in R (Wellmann, 2021).

Results

Simulated scenarios showed large differences in rate of genetic gain and inbreeding in the RDN breeding program in years 26-30 (Table 1). Scenario 'Separate' caused the highest rate of inbreeding and the lowest rate of genetic gain. The other three scenarios caused the highest genetic gain in the aggregate genotype, without any significant differences between these scenarios. The rate of inbreeding was the lowest in scenario 'Restricted-20%'.

Table 1. Rate of inbreeding per generation and rate of genetic gain in the RDN breeding program per year for the different scenarios in year 26-30.

	Separate	Restricted-20%	Restricted-50%	Unrestricted
Rate of inbreeding (pedigree)	1.85%^{a}[1]	0.36%^{c}	1.06%^{b}	0.74%^{bc}
Rate of inbreeding (IBD)	2.69%^{a}	0.68%^{c}	1.49%^{b}	1.12%^{bc}
Aggregate genotype (in std. BG)	0.27^{b}	0.45^{a}	0.42^{a}	0.41^{a}
Milk production (in σ_A units)	0.27^{b}	0.43^{a}	0.43^{a}	0.40^{a}
Health (in σ_A units)	0.04^{a}	0.08^{a}	0.04^{a}	0.06^{a}

[1] Subscripts with different letters indicate significant differences between estimates in the same row.

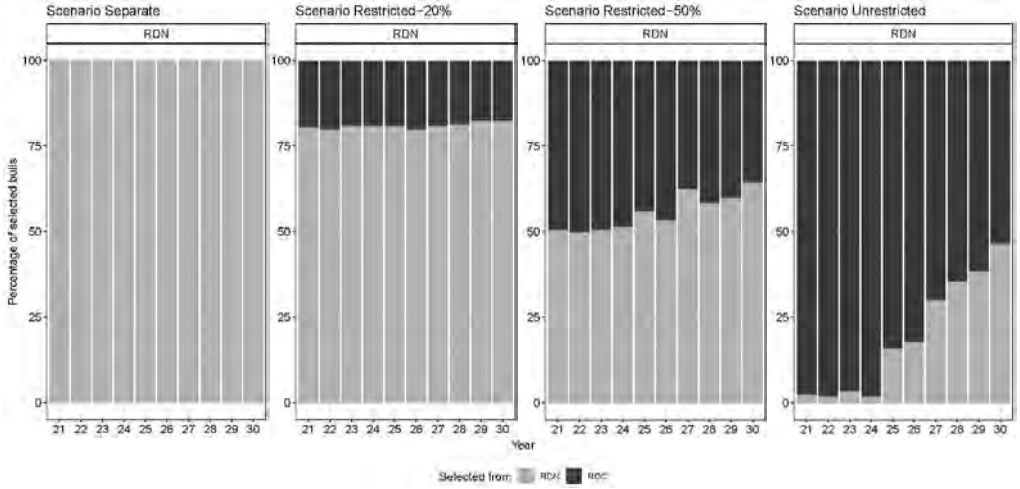


Figure 1. Percentage of selected bulls that are selected for breeding in the RDN breeding program from each breed, per scenario and per year.

The percentage of bulls selected for breeding in the RDN breeding program but originating from the RDC breeding program, or so called 'foreign' bulls was 0% in scenario 'Separate' (Figure 1). Approximately 20% of the selected bulls were foreign in scenario 'Restricted-20%'. The percentage of foreign bulls selected for breeding in the RDN breeding program in scenario 'Unrestricted' and 'Restricted-50%' changed over the years of the simulation. The breed proportion of RDN decreased over the years of the genomic selection scheme for all scenarios except 'Separate' (Table 2). The proportion of RDN genes in year 30 was twice as high in scenario 'Restricted-20%' compared to scenario 'Unrestricted'.

Table 2. Proportion of RDN genes in bulls born in the RDN breeding program and selected for breeding in the RDN breeding program, per year of the simulation.

Year	Separate	Restricted-20%	Restricted-50%	Unrestricted
21	1.00	1.00	1.00	1.00
22	1.00	1.00	1.00	1.00
23	1.00	0.57	0.56	0.50
24	1.00	0.55	0.54	0.75
25	1.00	0.42	0.29	0.27
26	1.00	0.38	0.28	0.26
27	1.00	0.34	0.24	0.18
28	1.00	0.30	0.23	0.16
29	1.00	0.26	0.19	0.12
30	1.00	0.24	0.17	0.12

Discussion

Different scenarios were simulated in this study with the aim of increasing the breed performance in the RDN breeding program by collaborating with the RDC breeding program. In the case of separate breeding programs, the rate of genetic gain in the RDN breeding program was significantly lower compared to the other scenarios. The rate of inbreeding would be too high compared to the rule of thumb of 1% per generation (Oldenbroek and Van der Waaij, 2015). When looking solely at genetic gain and inbreeding (Table 1), scenario 'Restricted-20%' was the best breeding strategy. This scenario caused equal rate of genetic gain in the aggregate genotype and on trait level compared to the other (Un)restricted scenarios, but with a lower rate of inbreeding. This rate of inbreeding was significantly lower in scenario 'Restricted-20%' compared to scenario 'Restricted-50%' because in 'Restricted-50%' some RDN bulls were selected twice due to the set-up of the simulation. In scenario 'Restricted-20%', at least 80% of the bulls that were selected for breeding in the RDN breeding program also originated from the RDN breeding program. However, even in this very restricted scenario, the proportion of RDN genes at the end of the genomic selection scheme dropped to 0.25 (Table 2). The question remains if this proportion is high enough to speak about conservation of the RDN, or that the breeding strategies studied in the current simulation will not lead to the conservation of the RDN genotype. An alternative breeding strategy might be to select bulls based on RDN breed proportions instead of in which breeding program a bull is born. In addition, optimum contribution selection can be used to strengthen the breeding program. This was however beyond the scope of this study.

References

International Bull Evaluation Service. 2012. National Genetic Evaluation Form Holstein Production traits. Available at: https://wiki.interbull.org/public/Nat_GE_Forms?action=AttachFile&do=get&target=DFSHOLpro.doc.

International Bull Evaluation Service. 2014. Interbull Routine Genetic Evaluation for Udder Health Traits, April 2014. Available at: http://www.interbull.org/web/static/mace_evaluations_archive/udder/uder-apr14.html.

Madsen P., and Jensen J.. 2013. A User's Guide to DMU. Version 6, release 5.2. Aarhus Univ. Foulum, Denmark.

Norberg E., Madsen P., and Pedersen J. 2009. A multi-trait genetic analysis of protein yield, udder health, and fertility in first lactation Danish Holstein, Danish Red, and Danish Jersey using an animal model. Acta Agric. Scand. Sect. a-Animal Sci. 59(4):197–203. https://dx.doi.org/10.1080/09064700903490588.

Oldenbroek K., and van der Waaij L. 2015. Textbook Animal Breeding and Genetics for BSc students. Centre for Genetic Resources and Animal Breeding and Genomics Group, Wageningen University and Research Centre, Wageningen, the Netherlands.

Pedersen L.D., Sørensen A.C., Henryon M., Ansari-Mahyari S., and Berg P. 2009. ADAM: A computer program to simulate selective breeding schemes for animals. Livest Sci 121(2-3): 343–344. https://dx.doi.org/10.1016/j.livsci.2008.06.028.

R Core Team. 2021. R Project for Statistical Computing. Available at: https://www.r-project.org/

Schmidtmann, C. 2021. Genomic characterization of Red dairy cattle in Northern Europe and the optimization of their breeding programs. Thesis. Christian-Albrechts-University Kiel.

Sørensen L.P., Pedersen J., Kargo M., Nielsen U.S., Fikse F. *et al.* 2018. Review of Nordic Total Merit Index Full Report November 2018. Available at: https://www.nordicebv.info/wp-content/uploads/2018/11/2018.11.06-NTM-2018-report-Full.pdf

Wellmann R. 2021. Package 'optiSel.' Available at: https://cran.r-project.org/web/packages/optiSel/index.html

249. Genetic rescue of small populations in the presence of deleterious variation

C. Reimer[1,2*], S. Weigend[1,2], J. Geibel[2,3] and T. Pook[2,3]

[1]Friedrich-Loeffler-Institut, Institute of Farm Animal Genetics, Höltystraße 10, 31535 Neustadt, Germany; [2]University of Goettingen, Center for Integrated Breeding Research, Carl-Sprengel-Weg 1, 37075 Göttingen, Germany; [3]University of Goettingen, Department of Animal Sciences, Animal Breeding and Genetics Group, Albrecht-Thaer-Weg 3, 37075 Göttingen, Germany; christian.reimer@fli.de

Abstract

Genetic rescue using external breeders or cryo-preserved semen is a reasonable strategy to restore diversity in small breeds. It is also known from natural populations that animals from large populations can convey substantial genetic load when they migrate into small, highly inbred populations, which may compromise the recipient population. We used stochastic simulation to evaluate different scenarios of genetic rescue of two conservation stocks of different sizes and found that introgression of males from larger stocks had a more detrimental effect on the viability of offspring than from smaller stocks, even though both strategies partially restored diversity. Using samples from gene-banks had a comparably negative effect on offspring viability, if the population was substantially larger at the time of cryo-conservation. Our results suggest that careful assessment of genetic load and its consequences is required whenever external animals should be transferred into endangered stocks. As larger breeds have higher genetic load due to less efficient purging, animals from small breeds seem beneficial for genetic rescue.

Introduction

The conservation of genetic diversity is of major concern in many livestock species as commercially less preferred breeds are often decreasing in size, leading to a high degree of inbreeding. It is believed, that such breeds will consequently suffer from inbreeding depression (Charlesworth and Willis, 2009) and might eventually become inviable (Soule, 1987). Deleterious variants seem to play an important role, even though the actual mechanisms are under debate (Charlesworth and Willis, 2009) and their distribution and frequency in livestock might be strongly dependent on population characteristics, demography and breeding practices (Bosse *et al.*, 2019). At the same time, strong recessive deleterious variations are subject to constant selective pressure called purging, which is even more efficient under high levels of inbreeding, leading to viable, though highly inbred, populations (Robinson *et al.*, 2018).

Conservation schemes based either on introgression of animals from related, but numerically larger populations (Kyriazis *et al.*, 2021) or insemination with cryo-preserved semen from gene banks (FAO, 2012) are considered as a reasonable strategy to restore genetic variability and counter inbreeding depression. Albeit, positive effects generally outweigh the negative (Frankham, 2015), outbreeding depression is a concern. Observations from an island grey wolf population, where a male from a much larger population entered a highly inbred insular stock, putatively introducing strongly deleterious variants, resulted in the extinction of the respective population (Kyriazis *et al.*, 2021). This implicates that the aforementioned strategies might not be without risk also in livestock. The aim of this study is to evaluate the effect of animal migration from larger populations and a cryo-conserve into smaller stocks in a livestock setting with specific focus on strong deleterious variations.

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_249, © C. Reimer *et al.* 2022

Materials & methods

We used stochastic simulation with the R-package MoBPS (Pook *et al.*, 2020) and started with an initial pig population of 1000 pigs with an equal sex ratio. Variants and maps were taken from Ensembl (Howe *et al.*, 2021). First, a quantitative trait with 1000 purely additive QTLs was simulated to mimic a performance trait like 6-month weight. Second, 1000 recessively lethal deleterious variants were randomly placed across the genome. Frequencies of the deleterious variants were sampled from a normal distribution with a mean of 1% and a standard deviation of 0.5%, while setting values below zero to zero. This results in a population with about 88% of all meiosis in the founder population leading to viable offspring.

The simulation was divided into three main stages. S1: Initial 15 generations in a breeding stock of equal sex ratio to generate a viable start population followed by S2: deriving four lines, Large (L), Medium (M), Endangered (E) and Critical (C) and breeding with constant to drastically decreasing size and a sex ratio of 1:2 for 20 generations. Animal numbers were cascaded down sequentially (Table 1). Mild selection for 6-month weight was applied to all stocks in S1 and S2.

In S3, genetic rescue of the 'Endangered' and 'Critical' stock was modelled in scenarios by mating always two males from other stocks or a cryo-conserve, taken from the stock to be rescued ten generations ago, with all females, with the exception of no mating between M and C. After genetic rescue, selection was switched to random mating with constant stock size and sex ratio of 1:2. Proportions of purged deleterious variants, average deleterious allele content per animal, the proportion of offspring being homozygous for at least one lethal allele, and inbreeding rates (calculated based on share of the genome in Identity-by-Descent) were tracked through all generations. All simulations were repeated 500 times. The interested reader is referred to https://github.com/tpook92/MoBPS/ for the R-Script.

Results

The initialisation period S1 in the first 15 generations lead to an average kinship of 3.4% and on average 14.5 deleterious variants per individual, and quickly fixed 60% of the deleterious loci. The decreased population size in S2 led to an average kinship in L, M, E, and C, of 11, 39, 73 and 93%, respectively. Nearly all deleterious variants (>95%) got purged from E, C and M, while L still contained ~12% of the initial deleterious mutations (Figure 1). During S2, in 33% of the simulations C, and in 5% E got extinct.

S3 clearly showed decreased proportions of viable offspring in the first generations after genetic rescue, and decreased inbreeding, when males were taken from other stocks or the cryo-reserve (Figure 2). Consistently across repetitions, males from the largest population or the cryo-reserve had a stronger negative impact than from smaller stocks. In 11 simulations, breeding males from E in C led to extinction of C, likewise in each internal pairing in C and E once each. Inbreeding increase flattened and the benefit was greatest when males were taken from independent stocks rather than the cryo-reserve.

Table 1. Simulated stock sizes in *S2*.

Stock	Generation			
	17	18-21	22-26	27-36
Large	1000	1000	1000	1000
Medium	800	400	200	100
Endangered	700	200	80	36
Critical	300	75	40	15

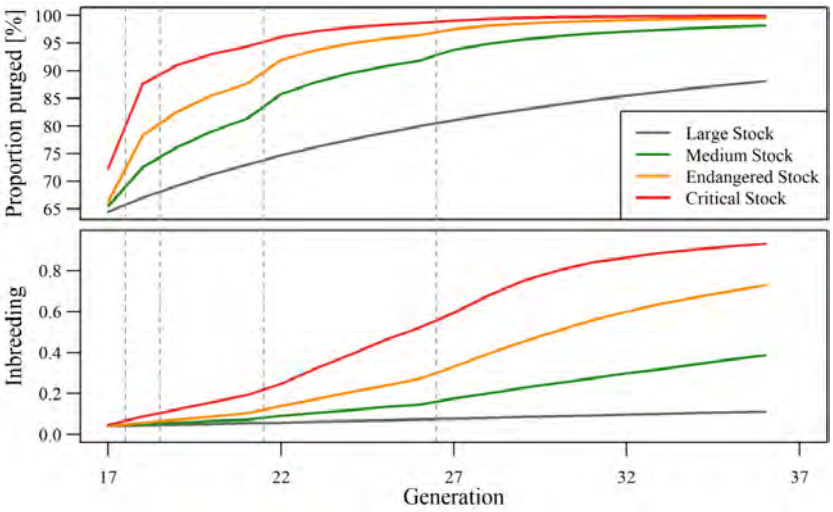


Figure 1. Top: Proportion of purged deleterious loci within a stock in S2, calculated on the basis of the founder generation. Bottom: Development of Inbreeding, calculated as average IBD per genotype.

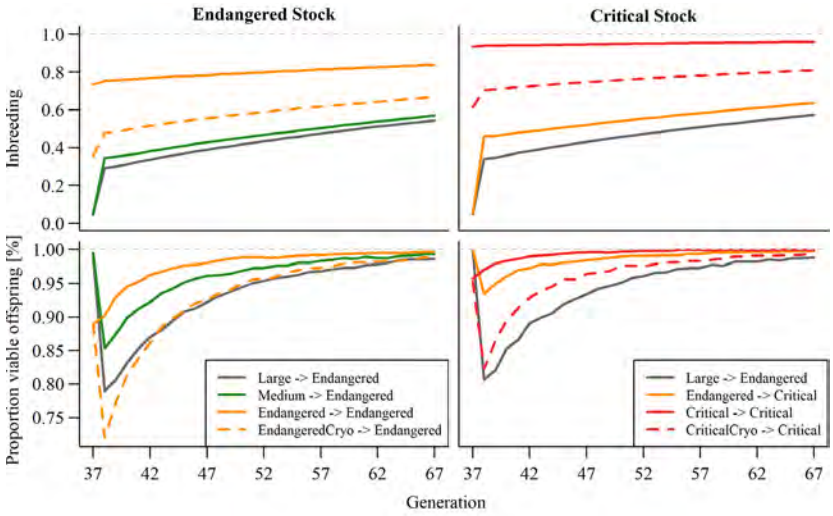


Figure 2. Degree of inbreeding derived from IBD of simulated genotypes and proportion of viable offspring for the genetic rescue scenarios of S3 for the Endangered (n=36) and Critical stock (n=15).

Discussion

Our results give an insight into how inbreeding and strongly deleterious variations interact in a framework of inbred livestock populations when genetic rescue is performed to restore genetic diversity. Strong inbreeding with resulting selection against deleterious variants reduced their number and frequency substantially, but genetic rescue led to decreased offspring viability due to homozygous lethal genotypes in the remaining loci in the short term. Extinction after rescue happened in E and C, but remained a rare event. Compared to the wild population scenario simulated by Kyriazis *et al.*(2021), genetic load in all donor stocks was limited due to fairly efficient purging. The main finding is the effect that introgression

from larger sub-populations was worse compared to usage of similarly endangered stocks, as the first had a substantially higher genetic load. The detrimental effect of using gene-bank semen could be attributed to the larger historical population size at sampling and shows that careful characterisation of deleterious variation is important when using older samples. Simulation seems to be an appropriate tool to assess such effects and has been used before (Wang *et al.*, 1999), but knowledge about the genetic structure of deleterious mutations is paramount to obtain reliable results. We studied strongly lethal variants, which might be purged as demonstrated, but variants with mild effects could become inherent (Charlesworth and Charlesworth, 1987; Hedrick and Kalinowski, 2000) and be more important in long term viability. To better account for the effect of deleterious variation in the context of conservation breeding, novel genomic approaches to assess deleterious variation, such as homozygous haplotype deficiency or sequence guided identification (Derks *et al.*, 2021) might be of high impact when routinely employed.

References

Bosse, M., Megens, H., Derks, M. F. L., Cara, Á. M. R., & Groenen, M. A. M. (2019). *Evolutionary Applications*, *12*(1), 6–17. https://doi.org/10.1111/eva.12691

Charlesworth, D., & Charlesworth, B. (1987). *Annual Review of Ecology and Systematics*, *18*(1), 237–268. https://doi.org/10.1146/annurev.es.18.110187.001321

Charlesworth, D., & Willis, J. H. (2009). *Nature Reviews Genetics*, *10*(11), 783–796. https://doi.org/10.1038/nrg2664

Derks, M. F. L., Gross, C., Lopes, M. S., Reinders, M. J. T., Bosse, M., *et al.* (2021). *Genomics*, *113*(4), 2229–2239. https://doi.org/10.1016/j.ygeno.2021.05.017

FAO. (2012). *Cryoconservation of Animal Genetic Resources*. https://www.fao.org/3/i3017e/i3017e00.htm

Frankham, R. (2015). *Molecular Ecology*, *24*(11), 2610–2618. https://doi.org/10.1111/mec.13139

Hedrick, P. W., & Kalinowski, S. T. (2000). *Annual Review of Ecology and Systematics*, *31*(1), 139–162. https://doi.org/10.1146/annurev.ecolsys.31.1.139

Howe, K. L., Achuthan, P., Allen, J., Allen, J., Alvarez-Jarreta, J., *et al.* (2021). *Nucleic Acids Research*, *49*(D1), D884–D891. https://doi.org/10.1093/NAR/GKAA942

Kyriazis, C. C., Wayne, R. K., & Lohmueller, K. E. (2021). *Evolution Letters*, *5*(1), 33–47. https://doi.org/10.1002/evl3.209

Pook, T., Schlather, M., & Simianer, H. (2020). *G3 Genes|Genomes|Genetics*, *10*(6), 1915–1918. https://doi.org/10.1534/g3.120.401193

Robinson, J. A., Brown, C., Kim, B. Y., Lohmueller, K. E., & Wayne, R. K. (2018). *Current Biology : CB*, *28*(21), 3487-3494. e4. https://doi.org/10.1016/J.CUB.2018.08.066

Soule, M. E. (Ed.). (1987). *Viable Populations for Conservation*. Cambridge University Press. https://doi.org/10.1017/CBO9780511623400

Wang, J., Hill, W. G., Charlesworth, D., & Charlesworth, B. (1999). *Genetical Research*, *74*(2), 165–178. https://doi.org/10.1017/S0016672399003900

250. Genebank collections in relation to risk status in livestock species

G. Leroy[1*], H. Blackburn[2], C. Danchin-Burge[3], F. Tejerina Ampudia[4], S.J. Hiemstra[5], C. Ligda[6], P. Boettcher[1] and R. Baumung[1]

[1]Food and Agriculture Organization, viale delle Terme de Caracalla, 00153 Roma, Italy; [2]United States Department of Agriculture, 1111 South Mason St., 80521 Fort Collins, USA; [3]IDELE, 149 rue de Bercy, 75012 Paris, France; [4]MAPA, C-Almagro 33, 28010 Madrid, Spain; [5]Centre for Genetic Resources, the Netherlands (CGN), Wageningen University & Research, P.O. Box 338, 6700 AH Wageningen, the Netherlands; [6]Hellenic Agricultural Organization, VRI, P.O. Box 60 272, 57001 Thessaloniki, Greece; gregoire.leroy@fao.org

Abstract

This study investigated national genebanking priorities based on information reported in the Domestic Animal Diversity Information System (DAD-IS) and feedback provided by gene bank managers. Only 3.2% of national breed populations are reported to have enough material collected to allow breed reconstitution in case of extinction. Depending on species, region and demographic risk status, there are substantial differences in quantities of germplasm collected when comparing breeds with a high risk status vs demographically important breeds. Most genebanks have small budgets and use pragmatic approaches for collecting germplasm by partnering with breeding organisations and/or artificial insemination companies. They also work with a variety of breed associations to implement collection strategies to ensure that the breadth of genetic diversity is captured. Future collection growth within countries will be driven by national programs that incorporate stakeholder participation, collection use, the scope of diversity captured; and will depend upon levels of public and private funding.

Introduction

Ex situ conservation by genebanks is an approach that provides insurance against losses of animal genetic resources in the field, caused by breed replacement, genetic erosion and/or emergencies (Smith, 1984; Paiva *et al.*, 2016). Currently, two indicators of the Sustainable Development Goals (namely 2.5.2 and 2.5.1b) consider *in situ* and *ex situ* conservation of animal genetic resources, respectively. Indicator 2.5.2 is defined as the proportion of local breeds (i.e. being reported to exist only in one country) classified as being at risk of extinction. Indicator 2.5.1b considers the number of local breeds stored within a genebank collection with an amount of genetic material (semen, embryos, oocytes, etc.) stored sufficient to reconstitute the breed in case of extinction (FAO, 2012; Leroy *et al.* 2019). Those indicators are based on official data provided by countries to the Domestic Animal Diversity Information System (DAD-IS).

The aim of this study was to assess to what extent genebank collection strategies in livestock consider risks related to genetic diversity, both at species and breed level. To investigate this, DAD-IS data from 197 countries and territories were analysed and augmented with feedback provided by some countries in Europe and North America.

Materials & methods

Data from 15,121 national breed populations belonging to 37 species and species crosses, reported by 197 countries and territories, were extracted from DAD-IS in December 2021. Depending on the information provided, the cryoconservation status of a national breed population can fall into four categories: (1) no information (no information provided on cryomaterial); (2) no material (no material collected, according to the country); (3) not sufficient (material collected, but with an amount insufficient to allow breed reconstitution in case of extinction); and (4) sufficient. Countries can indicate whether they consider the quantities of material stored to be 'sufficient', based on their own conservation objectives and their local

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_250, © G. Leroy *et al.* 2022

capacity in reproductive biotechnologies, as these factors can vary substantially across breeds, species and countries. Otherwise, sufficiency is determined by default, based on internationally agreed criteria considering species, numbers of samples and donors and type of material stored (frozen semen, embyros, oocytes; FAO, 2012).

The cryoconservation status of national breed populations were analysed as the dependent variable in a multinomial logistic regression model (R multinom function) with the following explanatory categorical variables: (1) species (Cattle, Chicken, Goat, Horse, Pig, Rabbit, Sheep, Other avians, Other mammals); (2) region (Africa, Asia, Europe and the Caucasus, Latin America and the Caribbean, Near and Middle East, North America); (3) the breed's geographical classification (Local, Regional transboundary, International transboundary); and (4) the breed's risk classification (Unknown, Not at risk, Vulnerable, Endangered, Critical, Extinct). The risk classification is defined based on demographic thresholds endorsed by the FAO Commission on Genetic Resources for Food and Agriculture (FAO, 2013).

Data from 11 European countries collected through the ERFP (European Regional Focal Point) Ad Hoc Action were used to assess the challenges on developing national strategies for cryoconservation. In addition, five countries from Europe and North America also provided open comments on the question of genetic risk in relation to topics such as the criteria used for selecting breeds sampled for cryoconservation, the amount of material requested, and the sampling strategy at breed level.

Results

On the 15,121 national breed populations reported in DAD-IS, 47.8% are reported without information regarding their cryoconservation status, 43.5% are reported to not have any cryo-material collected, 5.5% are reported to have material collected, but with insufficient amount, and only 3.2% have enough material collected to allow breed reconstitution in case of extinction. Among species, mammals and especially cattle (9.5% of national breed populations with not sufficient and 6.7% with sufficient material), showed larger genebank collections compared to avian species ($P<0.001$). When compared to other regions, Asia, Europe and the Caucasus and North America, showed larger proportions of national breed populations with material stored, either not sufficient (1.4 vs 5.1, 8.2 and 13.4%, respectively) or sufficient (2.2 vs 3.8, 3.2 and 9.2%, respectively) ($P<0.001$). By contrast, cryoconservation status did not vary substantially by geographic status of the breed (local, regional transboundary, or international transboundary breeds) with proportion being 5.9, 4.4 and 5.2% for not sufficient, respectively, and 2.8, 3.1 and 3.6% for sufficient respectively ($P<0.01$). When considering risk status (Figure 1), national breed populations with unknown risk status (i.e. no information provided on their population over the last past years), were the most likely to not have information reported on their cryoconservation status (55.4%) ($P<0.001$). Risk status (population size) seemed to be both a driver (increased need) and an obstacle (greater difficulty in accessing donors) for material collection in small populations. The proportion of national breed populations with material collected was greatest for 'Vulnerable' populations. Among larger (i.e. 'Not at risk') populations, a smaller

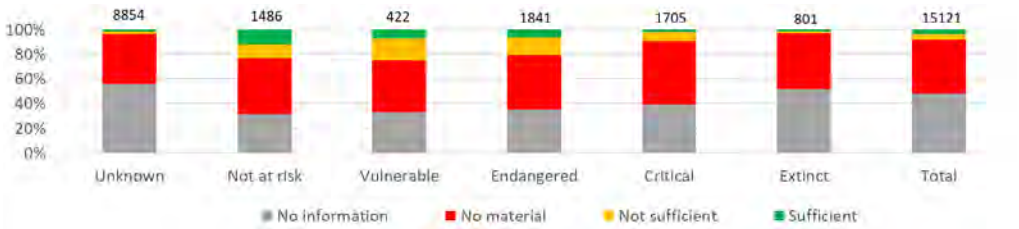


Figure 1. Collection status of national breed populations by risk status.

proportion of breeds had material collected, but the material was sufficient for a greater proportion of these populations, in part due to the economic importance of these populations and the ability to more effectively acquire germplasm. 'Endangered' populations were collected at a rate similar to those 'Not at risk', but the proportion with sufficient material was much less. Practical factors seemed to strongly limit collection of material from 'Critical' populations.

From the ERFP dataset, obstacles related with funding and policy issues were considered as the main drawbacks in development national genebank collections (Table 1). Among them, lack of funding, lack of infrastructures, low interest of breeders' societies and low priorities in governmental policies were ranked as the main ones. Technical issues link to specific species and genetic diversity aspects (low number of males in certain breeds) were ranked as medium obstacles (8.8 and 8.5 on 0-20 scale, respectively). When providing open input, national coordinators indicated that genetic and demographic risk status is only one among many factors for the selection of breeds to be collected. They reported that national genebanks use opportunities to collect material already available in AI centers (i.e. often from mainstream breeds) or through research projects, while trying at the same time to encourage and support breeding associations from local breeds to implement collection strategies. At the breed level, depending on information available (geographic distribution, pedigree and/or genomic information), donor animals are sampled to display a high genetic variability, using for instance optimal contribution selection (the Netherlands) or Ward's Minimum Variance method (USA), but often a more pragmatic approach.

Table 1. Obstacles reported by European countries for the development of ex-situ conservation strategies ranked from 1 (less important) to 20 (most important).[1]

	Obstacles	Rank (s.e.)	Obstacles	Rank (s.e.)	
Funding	**Lack of public funding for *ex situ* conservation**	16.9 (1.1)	Private companies/AI centers not involved	9.5 (1.8)	Organizational
	Unexploited funding opportunities	9.7 (1.5)	Low level of public awareness, in the farmer's community	9 (2.1)	
	No stable funding source	9.4 (2.5)	No integration of *ex situ* and *in situ* conservation activities	8.7 (1.6)	
			Weak structure of breeders associations / poorly developed breeding programs	8.2 (2.6)	
Policies	**Low interest of breeder societies in *ex situ* conservation activities**	12.5 (1.6)	**Lack of infrastructure and technical capacity**	13.3 (1.5)	Technical
	***Ex situ* conservation of animal genetic resources is low priority in governmental policies**	11.5 (2)	Gaps in methods for freezing semen and insufficient fertility in some species	8.8 (2.3)	
	Poor/lack of coordination among actors	10 (2.5)	Extremely small number of males in the population intended to be preserved	8.5 (1.7)	
	Absence of regulations on *ex situ* conservation	8.9 (2.6)	Lack of awareness on EUGENA activities	6.8 (1.9)	
	Bureaucratic burdens in financial planning and agreements	6.6 (1.8)	Lack of capacities at sub-country level	6.7 (2.2)	
	Low coordination among countries in case of transboundary breeds	6.2 (1.9)	Difficulties in the development of new biotechnologies	5.5 (1.3)	
	Legal restrictions in collections on the field or use of old material.	5.5 (1.8)	Problems with storage of duplicate collection	4.5 (1.6)	

[1] In bold the 5 most important obstacles.

Discussion & conclusions

Analysis of DAD-IS data shows that breeds not at risk are well represented in national genebank collections and that breeds at risk on average have limited amounts of genetic material stored in genebanks. Feedback provided by national coordinators indicate that national genebanks take into consideration more than genetic risk based upon a breed's threat level (e.g. socio-economic importance of a breed, or unique phenotypes). Although it is important for any breed and breeding program to develop *ex situ* collections, countries have different priorities and opportunities to support that. In practice, the level of development of genebank collections depends on national priorities, funding, stakeholder participation and the available infrastructure, institutions and capacities. When viewed globally, substantial germplasm collections have been developed in Europe and North America (Paiva *et al.*, 2016) and these resources present interesting research opportunities in the area of genomics and reestablishing lost genotypes (Dechow *et al.*, 2020). But national genetic resource programs must strive to increase the utilization of these collections to the research community and industry, as public and private goods. By increasing utilization of these resources, the genebank raises awareness of the collection's utility, which in turn can justify and facilitate the expansion of the collection.

References

Dechow C.D., Liu W.S., Specht L.W., and Blackburn H. (2020). J. Dairy Sci. 103(5): 4510-4516.

FAO (2012) Cryoconservation of animal genetic resources. FAO Animal Production and Health Guidelines No. 12. Rome.

FAO (2013) In vivo conservation of animal genetic resources. FAO Animal Production and Health Guidelines. No. 14. Rome

Leroy G., Boettcher P., Besbes B., Danchin-Burge C., Baumung R., & Hiemstra S.J. (2019). Diversity 11(12): 240. https://doi.org/10.3390/d11120240

Paiva, S.R., McManus, C.M., Blackburn, H.D. 2016. Conservation of animal genetic resources – A new tact. Livest. Sci. 193(2016): 32-38.

Smith C. (1984). Livest. Prod. Sci. 11(1): 37-48.

Part 16

Challenges – use of whole genome sequence information

251. Pangenomes of haplotype-resolved assemblies enable population-scale genotyping of cattle structural variation

A.S. Leonard* and H. Pausch

Animal Genomics, ETH Zurich, Universitaetstrasse 2, 8006 Zurich, Switzerland; alexander.leonard@usys.ethz.ch

Abstract

Haplotype-resolved, long read assemblies provide access to previously unresolved complex structural variation and enable pangenomic approaches to analyse diversity across multiple cattle breeds. We assess structural variation present in pangenomes integrating 12 HiFi-based haplotype-resolved and 4 older taurine cattle assemblies. Pangenomes produce approximately 15% more structural variants than linear reference-guided calling from sequencing reads and is especially beneficial for detection of variants larger than 50 Kb (238 compared to 75). We leverage population-scale short read data to investigate the prevalence of selected structural variants across multiple breeds. We find good consensus between the read-to-graph alignment approach of minigraph and the k-mer approach of Pangenie. We characterize structural variants within *PRDM9* and at QTL for mammary gland morphology and carcass traits nearby *TBX3* and *ENSBTAT00000082324* that might contribute to complex trait variation.

Introduction

Long read sequencing revolutionised genome assembly, greatly improving the contiguity of assemblies from kilobases to megabases. More recent improvements in the accuracy of long read sequencing, such as Pacific Biosciences HiFi or new Oxford Nanopore Technologies (ONT) chemistry, have further enabled assembly of challenging genomic regions, particularly those that are highly repetitive or share high sequence similarity. While updating reference genomes is beneficial (Aganezov *et al.* 2021), required validation efforts, manual curation, and lifting over existing results limit motivation to do so. As such, reference genomes, including the current Bos taurus reference (ARS-UCD1.2; Rosen *et al.*, 2020), commonly are lacking sequence in these challenging regions, hindering conventional methods like aligning sequencing reads to the reference. On the other hand, *de novo* assembly and pangenome integration allows for newly resolved genomic regions to be accessible to downstream analysis (Crysnanto *et al.* 2021, Leonard *et al.* 2021).

High sequencing and assembling costs are still limiting factors to producing sufficient high quality genomes for conducting population-scale studies at the assembly level. However, there are huge datasets available for many species, both public and private, for individuals either sequenced with short reads or array-genotyped. Recent tools like Pangenie (Ebler *et al.*, 2020) combine these different sources of DNA variation data, determining the spectrum of variation from haplotype-resolved assemblies, and then genotyping that variation across many samples sequenced with short reads. Such approaches are the first step towards understanding complex variation previously inaccessible to genomic analysis at population-scale.

Materials & methods

We used 16 cattle assemblies representing multiple breeds of taurine cattle of which 12 were assembled from HiFi reads (five Brown Swiss [BSW], six Original Braunvieh [OBV], one Piedmontese), three from Pacific Biosciences continuous long reads (one each of Hereford, Angus, and Highland), and one from ONT reads (Simmental [SIM]). A read-based structural variant (SV) panel was constructed with pbsv (Pacific Biosciences), calling on each sample and merging with bcftools (Danecek *et al.* 2021). An assembly-based variant panel was constructed using the assemblies-to-pangenome pipeline of Pangenie. For comparison, both panels were filtered with 'abs(ILEN)≥50' and intersected using bcftools. Graph pangenomes were constructed with minigraph (Li *et al.*, 2020) with the ARS-UCD1.2 reference genome as the backbone.

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_251, © A.S. Leonard and H. Pausch 2022

Short paired-end reads (2×150bp) from 49 samples (20 BSW, 20 OBV, 5 SIM, 2 Fleckvieh [FV], and 2 Holstein [HOL]) were aligned with the minigraph short-read preset to estimate the number of reads aligned to each node. Short read sequencing samples were genotyped against the panel with Pangenie using default parameters. The per-sample genotyped vcf files were merged using bcftools. The principal component analysis (PCA) plot was generated with plink2 (Chang *et al.* 2015).

Results

The minigraph pangenome, covering the 29 autosomes, contained 133,725 nodes with an average length of 19.1 Kb. There was approximately 65 Mb of non-reference sequence identified, in line with expectations of a taurus-only bovine pangenome given previous observations (Crysnanto *et al.* 2021). There were 47,324 bubbles across the autosomes, of which 1,554 overlapped with coding sequence annotated in the ARS-UCD1.2 reference genome. Several of these bubbles coincide with previously identified QTL for complex traits: a mammary gland morphology QTL nearby *TBX3* on BTA17 (Pausch *et al.* 2016, Fang and Pausch 2019), a carcass trait QTL nearby *ENSBTAT00000082324* on BTA5 (Fang and Pausch 2019) and insertions in *PRDM9* on BTA1. The structural variation present in the pangenome and population coverage are shown in Figure 1. The small number of assemblies prevent immediate association testing between the bubbles and traits of interest. However, signals of non-reference sequence are clearly visible for some samples, particularly when reads span node-edge-node pairs.

The approximate mapping approach of Minigraph can result in suboptimal pangenomes in complex regions, which interferes with read mapping. In the case of a QTL nearby *ENSBTAT00000082324*, node 5477 effectively contains 342 novel bases followed by 4,006 bases with 98% sequence similarity to node 1,067. This potentially suboptimal graph structure might lead to ambiguous or incorrect alignment of short reads, and erroneously suggest support for the reference or non-reference node.

Alternative VCF-based pangenome models, like that of Pangenie, first require a variant panel. These panels can be created through read-based calling (e.g. pbsv) or assembly-based calling (e.g. minimap2 with paftools.js). Assembly-based calling is more intensive, requiring higher coverage and computational resources to first produce the assemblies and then call variants, but the greater contiguity and correctness allows more accurate calling of large and/or complicated structural variation compared to using reads

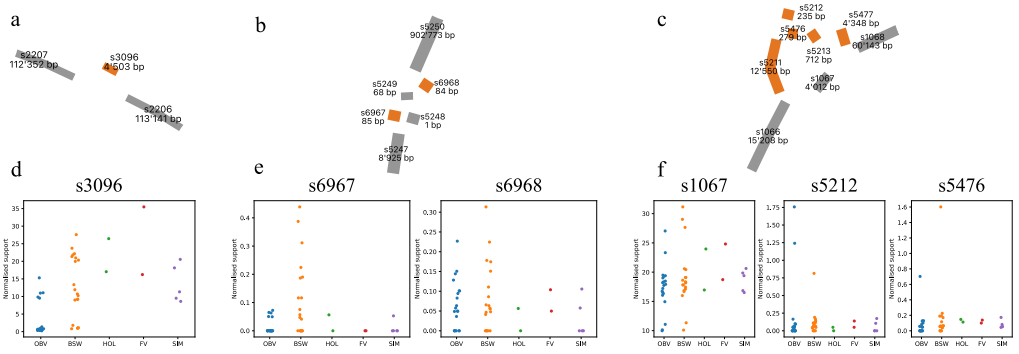


Figure 1. Population-scale short read support for structural variation alleles in minigraph pangenomes. a-c) Bandage plots of pangenome subgraphs for QTL nearby *TBX3*, *PRDM9*, and *ENSBTAT00000082324*. Reference and non-reference nodes are respectively coloured grey and orange. d-f) Short read support per node normalized by sample coverage.

directly. Given the same set of HiFi reads across 6 samples, we find assembly-based calling outperforms read-based calling (Table 1). Over half of the identified SVs are common to both read-based and assembly-based variant calling, but the former yielded a larger number of records, particularly ultra-long variants. Some read-based variants appear as fragmented versions of assembly-called variants, supporting that read-based variants may incorrectly call sufficiently large variation.

After genotyping the 49 samples against the assembly-based panel with Pangenie (containing both small and large variants), we assess the relationship between different input samples. We find good separation in the PCA of the merged sample VCF, shown in Figure 2, indicating the assembly-based variant panel does capture genomic differences across multiple breeds, and this can be leveraged into short read sequencing of much larger cohorts.

There was good consensus between Pangenie and minigraph for the genotyped SVs of interest (Table 2), despite the former using k-mers and variant panels while the latter uses alignment and assemblies directly. The minigraph approach is more suitable to manual inspection, and validates that the more scalable method of Pangenie is sufficiently accurate.

Table 1. Structural variation identified across HiFi cattle samples.

Method	Insertions	Deletions	Mean size (bp)	N>50 kb	N>100 kb
Assembly-based	19,810	12,435	2,252	238	25
Read-based	14,268	11,398	1,894	75	3
Mutual	20,905	13,766	1,079	6	0

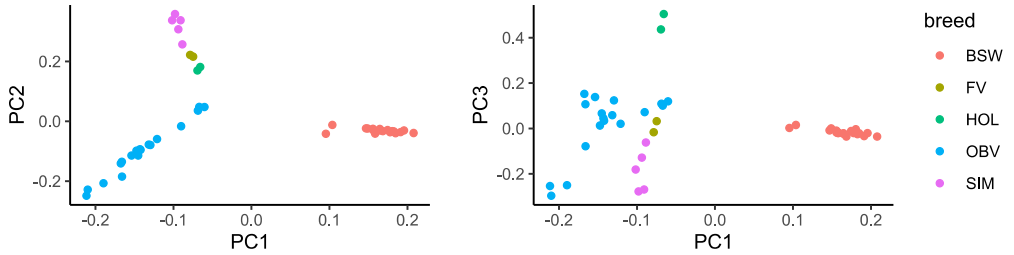


Figure 2. PCA plot for Pangenie genotyped samples. Breed information is not used in the PCA, with colours being assigned manually to breeds.

Table 2. Reference and alternate variant genotyping in minigraph and Pangenie.

Variant	Genotyper	BSW	OBV[1]	FV	HOL	SIM	Consensus
TBX3	minigraph	5/15	15/5	0/2	0/2	0/5	1.0
	Pangenie	5/15	13/5	0/2	0/2	0/5	
ENSBTAT00000082324	minigraph	3/17	9/11	0/2	0/2	0/5	0.96
	Pangenie	2/18	8/10	0/2	0/2	0/5	
PRDM9	minigraph	8/12	15/5	2/0	1/1	4/1	0.82
	Pangenie	10/10	13/5	0/2	0/2	2/3	

[1] Two OBV samples failed with Pangenie.

Simple variants, like an insertion near *TBX3*, are easier to genotype as they appear as large, novel insertions. More complex variants, like the two 84 bp insertions in the coding sequence of *PRDM9,* are more challenging to genotype correctly, as they are repetitive and have high sequence similarity with nearby regions.

Discussion

Population-scale analysis of DNA variation currently still requires cheaper variant detection methods like array-genotyping or short read sequencing over long read *de novo* assembly. Detecting variation through pangenomes and then genotyping additional samples enables analysis in challenging or complex regions that are inaccessible without high quality assemblies. Here we show that different approaches, like that of minigraph+alignment or Pangenie, can effectively genotype variation across large cohorts of short read sequenced samples. Still, repetitive or duplicated regions impair both read alignment and k-mer analysis and may require long read sequence or *de novo* assemblies to confidently assess.

The VCF output of Pangenie can be directly used in conventional downstream analysis, simplifying adoption, but the VCF panel model is substantially weaker than graph pangenome models for representing complex or nested variation. However, it remains unanswered if genotyping variation at the largest (i.e. one allele per variation combination) or at the smallest (i.e. separately assessing a deletion contained inside an insertion) bubble level is more biologically informative. One such example is the carcass trait SV, where a near identical 12.5 Kb insertion is present in four distinct alleles, potentially skewing the analysis when the alleles primarily differ by variation in later nodes. Such questions warrant further investigation as we look to genotype medium-scale short read cohorts and impute into population-scale cohorts.

References

Aganezov S., Yan, S. M., Soto, D. C., Kirsche M., Zarate S., *et al.* (2021). BioRxiv. https://doi.org/10.1101/2021.07.12.452063

Chang, C. C., Chow, C. C., Tellier, L. C., Vattikuti, S., Purcell, S. M, *et al.* (2015). GigaScience 4(1). https://doi.org/10.1186/s13742-015-0047-8

Crysnanto, D., Leonard, A. S., Fang, Z.-H., and Pausch, H. (2021). PNAS 118(20). https://doi.org/10.1073/pnas.2101056118

Danecek, P., Bonfield, J. K., Liddle, J., Marshall, J., Ohan, V., *et al.* (2021). GigaScience 10(2). https://doi.org/10.1093/gigascience/giab008

Ebler, J., Clarke, W. E., Rausch, T., Audano, P. A., Houwaart, T., *et al.* (2020). BioRxiv. https://doi.org/10.1101/2020.11.11.378133

Fang, Z. H., Pausch, H. (2019). BMC Genomics, 20(695). https://doi.org/10.1186/s12864-019-6066-6

Leonard, A. S, Crysnanto, D., Fang, Z.-H., Heaton, M. P., Vander Ley, B. L., *et al.* (2021). BioRxiv. https://doi.org/10.1101/2021.11.02.466900

Li, H., Feng, X., & Chu, C. (2020). Genome Biology, 21(1). https://doi.org/10.1186/s13059-020-02168-z

Pacific Biosciences, pbsv. Available at: https://github.com/PacificBiosciences/pbsv.

Pausch, H., Emmerling, R., Schwarzenbacher, H., Fries, R. (2016). Genet Sel Evol 48(14). https://doi.org/10.1186/s12711-016-0190-4

Rosen, B. D., Bickhart, D. M., Schnabel, R. D., Koren, S., Elsik, C. G., *et al.* (2020). GigaScience 9(3). https://doi.org/10.1093/gigascience/giaa021

252. Comparison of methods for building pangenome graphs

D. Crysnanto*, A. Leonard and H. Pausch

Animal Genomics ETH Zurich, Universitätstrasse 2, 8092, Zurich, Switzerland; danang.crysnanto@usys.ethz.ch

Abstract

Graph-based pangenomes could mitigate the bias from the use of a single reference genome. However, the impact of different pangenome integration approaches on the properties of the resulting graph remains unknown. In this study, we compared three methods for building pangenome graphs from multiple genome assemblies. Minigraph performs an approximate mapping of assemblies to a backbone genome to construct the pangenome. In contrast, Cactus and pggb apply reference-free base level alignment to build pangenome graphs. Our results show that pangenome graphs constructed via base-level alignment contain 40% more small variations than the minigraph pangenome. Cactus, pggb, and minigraph uncover an almost identical set of structural variations. However, the breakpoints and allelic paths inferred from minigraph's approximate mapping tend to be less precise than graphs constructed via base-level alignment. Taken together, our study informs on the optimal strategy for building informative pangenome graphs that are now being conducted for many species.

Introduction

It becomes increasingly evident that a single haploid reference genome cannot capture the full species diversity. Pangenomes integrate genomic resources generated from multiple individuals into a more informative reference (Yang *et al.*, 2019). Graph-based structures are commonly used to represent sequence diversity in pangenomes, where distinct alleles form nodes in the graph (Eizenga *et al.*, 2020). More recent approaches construct pangenomes from multiple genome assemblies (Crysnanto *et al.*, 2021; Leonard *et al.*, 2021). However, building multi-assembly graphs can be computationally demanding and the resulting graphs are often convoluted and difficult to interpret. The approximate mapping algorithm implemented in minigraph is one possibility to efficiently construct multi-assembly graphs. Minigraph incorporates large structural changes and requires a single graph backbone, which is typically the reference genome (Li *et al.*, 2020). Recently, Cactus (Armstrong *et al.*, 2020) and pggb (pangenome graph builder) (Garrison, 2021) have been proposed as alternative methods to construct pangenomes through reference-free base-level alignment. However, the properties and differences of graphs resulting from various pangenome approaches are vastly unknown.

Materials & methods

Twelve Bovinae assemblies were used for the graph construction: ARS-UCD1.2 (current reference genome) (Rosen *et al.*, 2020), Angus (Koren *et al.*, 2018), Highland (Rice *et al.*, 2020), Brahman (Koren *et al.*, 2018), Yak (Rice *et al.*, 2020), Simmental (Heaton *et al.*, 2021), Bison (Heaton *et al.*, 2021), Brown Swiss (Leonard *et al.*, 2021), Original Braunvieh (Leonard *et al.*, 2021), Piedmontese (Leonard *et al.*, 2021) and Nellore (Leonard *et al.*, 2021). All assemblies were soft-masked using RepeatMasker. The ARS-UCD1.2 assembly was used as the backbone of the graphs and the order of including the other assemblies was determined based on their mash distance to the backbone (Figure 1a). To limit computational complexity, all analyses are based on bovine chromosome 19.

Three different graph construction methods were applied: minigraph, pggb, and the Cactus Progressive pangenome pipeline. Minigraph extends the minimizer-mapping of minimap2 to a graph that is computationally efficient because it avoids base-level alignment. Cactus and pggb are computationally more demanding because they perform base-level alignment. Pggb consists of three phases: base-level alignment

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_252, © D. Crysnanto *et al.* 2022

with *wfmash*, graph induction with *seqwish*, and normalization with *smoothxg*. The pggb pipeline was run with the recommended parameters for a gigabase-sized genome with a repeat content similar to the human genome. The final *.smooth.gfa version of the graph was used for the subsequent analysis. The Cactus pipeline was run with progressive alignment using the guide tree from the mash-based distance. The final HAL alignment was converted into the vg graph format using *hal2vg*. Variations from the different graphs were determined relative to the ARS-UCD1.2 coordinates using snarls decomposition with *vg deconstruct*.

Results

Pggb and Cactus produced pangenomes that contain 5.5 and 5.7 Mb non-reference sequences, respectively, which were mostly found in structural variants longer than 50 bp. They also contained 2.2 (pggb) and 2.5 (Cactus) Mb non-reference bases due to approximately 1.7 million SNPs and Indels smaller than 50 bp. Most of these variants were missing in the minigraph pangenome. 75.58, 69.05 and 64.98% of the bases contained in non-reference nodes larger than 50 bp identified by Cactus, pggb, and minigraph, respectively, were repetitive. A large fraction (>25%) of the repeat elements belonged to L1 elements. Interestingly, 800 kb non-reference sequence of the pggb and Cactus pangenome graphs were satellites. The overall length of large non-reference nodes was smallest in the minigraph pangenome because it did not contain satellites in the non-reference sequences larger than 50 bp. Replacing the ARS-UCD1.2 backbone with more complete assemblies (e.g. Brown Swiss and Original Braunvieh) reduced the amount of non-reference sequences (~2.2 Mb) and slightly increased the proportion of repeat-masked sequences (~66.2%) but still did not recover satellites in the minigraph pangenome.

Pangenomes constructed using Cactus and pggb contained substantially more large variations (4729) than those constructed using minigraph (3027). However, most structural variations were identified by all pangenomes (Figure 1b). Using a 50 bp threshold to accommodate imprecise breakpoint resolution, pggb and Cactus had the lowest and Cactus and minigraph had the highest breakpoint discrepancy. We find that Cactus and pggb had the highest (86.03%), whereas pggb and minigraph had the lowest (82%) intersection of exact paths traversed by the assemblies.

Table 1. Properties of three pangenome graphs constructed for bovine chromosome 19.

Statistics	Unit	Cactus	pggb	minigraph
Total nodes	n	4,838,791	4,469,595	10,419
Node length	bp	69,008,903	68,607,291	65,888,880
Reference nodes	n	3,088,293	2,908,218	6,695
Reference node length	bp	63,274,710	63,061,455	63,449,741
Non- reference nodes	n	1,750,498	1,561,377	3,724
Non-reference node length	bp	5,734,193	5,545,836	2,439,139
Small non-reference node length [1]	bp	2,477,931 (43.21%)	2,204,141 (39.74%)	14,299
Large non-reference nodes length[1]	bp	3,256,262 (56.79%)	3,341,695 (60.26%)	2,424,840
Maximum memory usage	Gb	34.7	31.2	2.6
Average memory usage	Gb	4.8	4.1	1.9
CPU time	h	10.4	78.4	0.21
Wall-clock time (10 threads)	h	1.82	6.16	0.21

[1] small: <50 bp; large: >50bp.

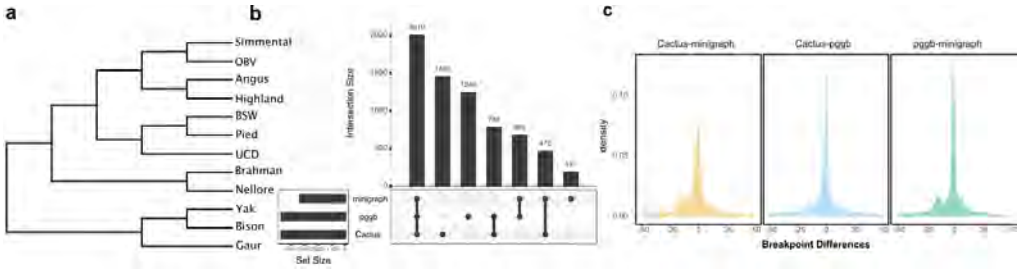


Figure 1 (a) Genetic relationship between 12 assemblies based on the mash distance (b) Overlap of the structural variations discovered among three tools within 50 bp of the breakpoints (c) Density of the differences in breakpoints of the overlapped variations.

Our findings indicate that pggb and Cactus pangenomes more often resolve bubbles with multiple paths that differ only in small variations, whereas these are collapsed into a single allele in minigraph pangenomes (e.g. Figure 2a). Additionally, minigraph pangenomes frequently contain bubbles with suboptimal topology that compromise the re-alignment of assemblies thereby preventing unambiguous assignment of paths (Figure 2b). These findings suggest that the approximate mapping in minigraph as well as the base-level alignment of pggb and Cactus uncover large mutational events quite accurately. However, only the pangenomes built using pggb and Cactus resolve complex bubbles that contain multiple paths that are largely similar.

Discussion

Here we built bovine pangenome graphs from twelve assemblies using either fast approximate mapping or computationally more demanding base-level alignment. We showed that minigraph pangenomes lack 40% of the total variations present in the pggb and Cactus graphs. There was generally good agreement across different pangenomes in terms of large variations. However, breakpoints are less precise and re-alignment of assemblies suffers from suboptimal topology in the minigraph pangenomes likely resulting from collapsed variations due to approximate mapping. We suspect that these limitations hinder the broad and universal applicability of approximate pangenome graphs beyond structural variant detection (Ebert

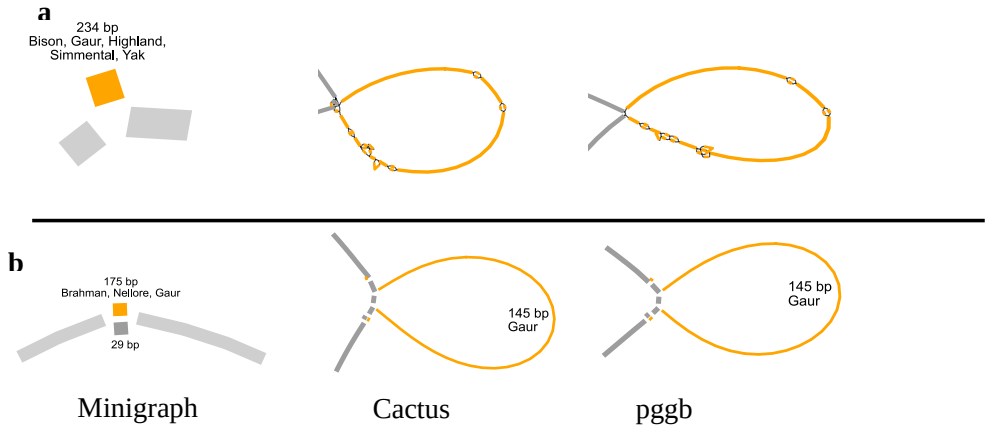


Figure 2. Visualization of allelic variation in three different pangenome graphs. Grey and orange color denote reference and non-reference nodes, respectively.

et al., 2021). A promising pangenome construction pipeline has recently been proposed that exploits the accuracy of full-base level alignment and the computational efficiency of approximate mapping (minigraph-Cactus pipeline) (Armstrong *et al.*, 2020). Our findings suggest that this strategy might increase the general applicability of pangenome graphs.

References

Armstrong J., Hickey G., Diekhans M., Fiddes I. T., *et al.* (2020) Nature 587(7833):246–251. https://doi.org/10.1038/s41586-020-2871-y

Crysnanto D., Leonard A. S., Fang Z. H., and Pausch H. (2021) Proceedings of the National Academy of Sciences of the United States of America 118(20):1–29. https://doi.org/10.1073/pnas.2101056118

Ebert P., Audano P. A., Zhu Q., Rodriguez-Martin B., *et al.* (2021) Science 7117eabf7117. https://doi.org/10.1126/science.abf7117

Eizenga J. M., Novak A. M., Sibbesen J. A., Heumos S., *et al.* (2020) Annual Review of Genomics and Human Genetics. https://doi.org/10.1146/annurev-genom-120219-080406

Garrison E. (2021). https://github.com/pangenome/pggb

Heaton M. P., Smith T. P. L., Bickhart D. M., Vander Ley B. L., *et al.* (2021) Journal of Heredity 1–8. https://doi.org/10.1093/jhered/esab002

Koren S., Rhie A., Walenz B. P., Dilthey A. T., *et al.* (2018) Nature Biotechnology 36(12):1174–1182. https://doi.org/10.1038/nbt.4277

Leonard A. S., Crysnanto D., Fang Z., Heaton M. P., *et al.* (2021) bioRxiv.

Li H., Feng X., and Chu C. (2020) Genome Biology 21(265):1–19. http://arxiv.org/abs/2003.06079

Rice E. S., Koren S., Rhie A., Heaton M. P., *et al.* (2020) GigaScience 9(4):. https://doi.org/10.1093/gigascience/giaa029

Rosen B. D., Bickhart D. M., Schnabel R. D., Koren S., *et al.* (2020) GigaScience 9(3):. https://doi.org/10.1093/gigascience/giaa021

Yang X., Lee W. P., Ye K., and Lee C. (2019) Genome Biology 20(1):. https://doi.org/10.1186/s13059-019-1717-0

253. Expanding the genomic toolkit: what does Oxford Nanopore sequencing have to offer?

A. Hess[1*], A. Caulton[2], R. Clarke[2], R. Brauning[2], K. McRae[2] and S. Clarke[2]

[1]University of Nevada, Reno, Agriculture, Veterinary & Rangeland Sciences, 1664 N. Virginia St. mail stop 202, Reno, NV 89512, USA; [2]AgResearch, Invermay Agricultural Centre, Private Bag 50034, Mosgiel 9053, New Zealand; ahess@unr.edu

Abstract

The challenge to provide energy-rich protein to the world while reducing the environmental impact is one of the largest challenges facing the animal industry today. To face this challenge, novel tools need to be adopted for methods to identify animals that perform optimally in their environment. This includes, but is not limited to, utilizing more complex variation, epigenetics, and microbes present within the host. A major hurdle in utilizing these data lies in the development of cost-effective and high-throughput methods for data capture. We argue that Oxford Nanopore Technologies is well-suited to fit this purpose. A computational approach called adaptive sampling couples long-read sequencing with targeted sequencing, affording the potential to capture multiple sources of information specifically and simultaneously in a sequencing run while reducing costs. This paper will discuss the potential benefits currently being explored to integrate this tool into the animal industry.

Sequence-based genotyping fills a niche within animal breeding

Sustainable progress in animal performance can be achieved by estimating and selecting on an individual's genetic merit. Utilizing pedigree allowed for the use of expected relationships based on the probability that two individuals shared alleles because they were identical by decent. This was further expanded to include the use of DNA markers dispersed across the genome, referred to as genomic selection. The use of genotype data has been widely adopted in the animal industry because it captures relationships between individuals at a higher resolution than using expectations from pedigree, resulting in more accurate predictions of an individual's genetic merit. Adoption of genomic information into animal industries has largely been accomplished through the development and use of SNP genotyping arrays; however, other 'minor' livestock species have found success utilizing other low-cost sequence-based platforms, e.g. genotyping-by-sequencing, for genomic analyses.

Expansion into using whole-genome sequence (WGS) has been explored to capture an individual's genetic merit more accurately. In particular, the use of WGS data is expected to increase genomic prediction accuracy in multi-breed genetic evaluations because the causative variant can be captured directly, whereas SNP arrays largely rely on linkage disequilibrium between the marker and the QTL (Moghaddar *et al.*, 2019). Traditional sequencing methods have been limited due to the costs of the technology when considering wide-scale adoption. Instead, a common approach is to build a reference set of animals that have reliable genotype calls from WGS data that are used to impute SNP array genotypes to WGS-level genotypes.

Low-pass sequencing has been developed as a balance between cost and reliability of genotype calls. Using this approach, the animal of interest is sequenced at <1× coverage across the genome, allowing for more accurate genotype calls when imputing the genotype (Snelling *et al.*, 2020). Low-pass sequencing is a cost-competitive alternative to SNP arrays and provides a straightforward approach to impute functional variant genotypes, without the upfront investment and inherent selection biases associated with developing SNP arrays (Matukumalli *et al.*, 2009). However, this approach is still limited by how representative the reference

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_253, © A. Hess *et al.* 2022

animals are of the target population, and a method to capture specific variants more accurately may still be necessary (Snelling *et al.*, 2020).

While genomic analysis using SNPs has been an invaluable tool for more accurately selecting animals with improved performance, moving forward, it is anticipated that our prediction equations will be expanded to include more information than SNPs by expanding to include DNA methylation, metagenomics, and structural variation, among others. The utility of this information will be dependent on the ability of this information to result in sustainable progress in the improvement of livestock species. This will be partially dependent on the ability to capture heritable genetic variation in these datatypes, which can be used for genetic selection purposes. Recently, the theoretical framework for estimating the heritable genetic component for these datasets when combined with traditional SNP information was developed (Christensen *et al.*, 2021). This is promising when considering how to integrate multiple sources of big data (i.e. large in dimensionality and complex). However, the challenge for the industry when considering widespread implementation is to develop tools that will be able to seamlessly capture multiple sources of information in a high-throughput and cost-effective manner.

Oxford Nanopore Technologies (ONT) have the potential to address this need, whereby long reads are produced to aid in accurate capture of more complex variation. Furthermore, they can identify base modifications (such as DNA methylation), making it possible to collect information on epigenetic markers. This technology can also base call sequences in real-time and decide whether to continue to read the sequence through a process called adaptive sampling (Payne *et al.*, 2021a), allowing for an efficient way to capture specific information. Therefore, with further development of this technology, there is potential to utilize ONT to simultaneously collect multiple sources of information on the same animal for genomic selection.

Long reads offer the potential to capture more complex variation

Long read sequencing generates reads that are thousands of base pairs in length, enabling the resolution of complex variation that troubles short read sequencing, such as copy number variation, or large insertions or deletions of sequence in the genome. The ability to accurately capture copy number variation will allow selection on more complex variation, which may improve prediction accuracy, particularly when the trait is under strong control of such a variant (El Hamidi *et al.*, 2018). Copy number variants (CNV) that differ by breed seem to be associated with adaptation. This has been shown in pigs and cattle where CNV regions under differential selection between breed and subspecies, respectively, are associated with disease resistance, heat tolerance, metabolism, muscle development and reproduction (Wang *et al.*, 2015; Hou *et al.*, 2012; Hu *et al.*, 2020; Stafuzza *et al.*, 2019), suggesting the potential of utilizing this information for genomic selection.

A major challenge for selecting on complex variation lies in the ability to accurately genotype the individuals. For example, the polled locus is a major locus that influences the development of horns, which is caused by an insertion in the *RXFP2* gene in sheep. However, to date, prediction accuracy has not improved by imputing the variant using SNP genotypes, presumably because the SNPs were not in high LD with the inserted sequence (Duijvesteijn *et al.*, 2018). The inability to accurately impute complex variation is common across the genome (Rafter *et al.*, 2020), suggesting direct capture of major CNVs would provide an improvement over the use of SNPs. Long-read sequencing can more accurately genotype copy number and complex structural variation by providing longer reads spanning a region of interest. Adaptive sampling (Payne *et al.*, 2021a) provides a means to specifically capture complex variants of interest to be considered when making selection decisions.

Oxford Nanopore machines capture epigenetic modifications

Epigenetics is a modular means for an individual to respond to changes in the environment, which can have an impact on the welfare of the individual. Variation in methylation status has a heritable genetic component, which has been estimated as 0.19 on average for DNA methylation measured at probes with no known SNPs in humans (McRae *et al.*, 2014) and as 0.20 for global DNA methylation rate in sheep (Hazard *et al.*, 2020). Epigenome-wide association studies are being explored for disease outcomes in humans (Rakyan *et al.*, 2011) and more recently in pigs for meat quality traits (Wang *et al.*, 2020). The methylation status at the loci identified in these studies, in addition to already known imprinted loci such as *callipyge* (Cockett *et al.*, 2005), may be utilized to improve the ability to predict individual performance.

Epigenetics can also be used as an indicator of the animal's ability to cope with environmental stressors. The concept of age prediction using epigenetic clocks has recently gained attention in livestock species (Caulton *et al.*, 2022; Hayes *et al.*, 2021). Deviation of chronological age and the age estimated by the epigenetic clock can be used as a predictor of biological age, which in turn could be incorporated into animal selection decisions. More specifically, early selection of animals with a 'reduced biological age' could result in animals with improved longevity through resilience to stressors such as a changing climate (Clarke *et al.*, 2021).

Epigenetic markers captured by long-read ONT sequencing could be used in animal breeding schemes either as a heritable molecular phenotype (e.g. biological age) or used to estimate the epigenomic similarity between individuals to better account for their historic environments (Clarke *et al.*, 2021). There is evidence that the methylation data provided by an ONT machine is more sensitive than the gold-standard bisulfite sequencing (Hayes *et al.*, 2021), but sequencing full genomes for age prediction is still cost prohibitive. Adaptive sequencing (Payne *et al.*, 2021a) may be employed to enrich target regions and reduce costs.

Adaptive sampling can target multiple organisms simultaneously

Metagenomic sequencing has gained attention lately as way to predict individual performance. Rumen microbial community profiles have been associated with environmentally important phenotypes and have been shown to be able to predict individual performance (Hess *et al.*, submitted-a). Microbial abundance has also been shown to have a heritable genetic component (Hess *et al.*, submitted-b), suggesting the ability to make sustained progress by selecting on the abundance of certain microbes. Similarly, capturing a microbial profile at the site of infection could be used to characterize complex polymicrobial diseases, such as pneumonia and mastitis, better which may improve our ability to predict the genetic merit of an individual through the improved definition of the disease afforded by increased resolution of the pathogens responsible for disease.

Adaptive sampling (Payne *et al.*, 2021a) can be used to deplete or enrich specific regions of a genome, or certain species. It has been used to specifically enrich for specific species of interest (Martin *et al.*, 2021), as well as deplete genomic sequence that is expected to be at high prevalence in the sample (e.g. Ong *et al.*, 2022). Therefore, samples can be interrogated to simultaneously screen for microbes of interest, either because they provide a benefit (e.g. reduce the carbon footprint) or a detriment (e.g. cause more severe disease) to the animal.

Towards an integrated approach to data capture

High-throughput, robust, and cost-effective assays need to be developed to successfully investigate, establish, and incorporate multiple datatypes into routine genetic merit predictions. A single platform, such as ONT, that can generate multiple types of desirable information, including complex variation, methylation, and pathogen load, will reduce the cost and turn-around time to the producer. Furthermore, ONT sequencing is also portable, potentially allowing for rapid analyses to be performed on-site. Through

targeted sequencing using adaptive sampling, high-confidence genomic data can be produced for specific targets, with low-depth sequence data being produced outside of the target region. Because background sequencing will still be captured due to the sequencing process, low-depth WGS can be obtained on an individual with higher read support in specific regions of interest, allowing for genomic prediction using the same data. This should provide the sufficient information to make accurate selection decisions (Lamb *et al.*, 2021) while capturing other relevant information that can be used to provide molecular phenotypes or can be used as predictors. The ability to couple adaptive sampling with barcoding (Payne *et al.*, 2021b) has the potential to increase throughput and reduce costs while maximizing the information collected on an individual. However, the throughput that can be achieved and strategies to identify the most suitable candidates for this approach need to be addressed before integration into animal breeding practices.

References

Caulton, A., Dodds, K. G., McRae, K. M., Couldrey, C., *et al.* (2022). Genes, 13(1), 96.

Christensen, O. F., Börner, V., Varona, L., & Legarra, A. (2021). Genetics, 219(2), iyab130.

Clarke, S., Caulton, A., McRae, K., Brauning, R., *et al.* (2021). Animal Frontiers, 11(6), 90-94.

Cockett, N. E., Smit, M. A., *et al.* (2005). Genetics Selection Evolution, 37(1), 1-18.

Duijvesteijn, N., Bolormaa, S., *et al.* (2018). Genetics Selection Evolution, 50(1), 1-11.

El Hamidi, A. H., Utsunomiya, Y. T., Xu, L., *et al.* (2018). BMC genomics, 19(1), 1-8.

Hayes, B. J., Nguyen, L. T., *et al.* (2021). Frontiers in genetics, 12, 760450-760450.

Hazard, D., Plisson-Petit, F., *et al.* (2020). Frontiers in Genetics, 11, 1717.

Hess, M., Hodgkinson, H, Hess, A., *et al.* (submitted to BMC Genomics)

Hess, M., Johnson, P., Brauning, R., *et al.* (submitted to Genetics Selection Evolution)

Horvath, S., & Raj, K. (2018). Nature Reviews Genetics, 19(6), 371-384.

Hou, Y., Liu, G. E., *et al.* (2012). Functional & integrative genomics, 12(1), 81-92.

Hu, Y., Xia, H., Li, M., Xu, C., Ye, X., *et al.* (2020). BMC genomics, 21(1), 1-11.

Lamb, H. J., Hayes, B. J., Randhawa, I. A., *et al.* (2021). Plos one, 16(12), e0261274.

Martin, S., Heavens, D., Lan, Y., Horsfield, S., Clark, M. D., & Leggett, R. M. (2021). bioRxiv.

Matukumalli, L. K., Lawley, C. T., Schnabel, R. D., *et al.* (2009). PloS one, 4(4), e5350.

McRae, A. F., Powell, J. E., Henders, A. K., *et al.* (2014). Genome biology, 15(5), 1-10.

Moghaddar, N., Khansefid, M., *et al.* (2019). Genetics Selection Evolution, 51(1), 1-14.

Ong, C. T., Ross, E. M., Boe-Hansen, *et al.* (2022). Journal of Animal Science, 100(1), skab344.

Payne, A., Holmes, N., Clarke, T., *et al.* (2021). Nature biotechnology, 39(4), 442-450.

Payne, A., Munro, R., Holmes, N., Moore, C., Carlile, M., & Loose, M. (2021). bioRxiv.

Rafter, P., Gormley, I. C., Parnell, A. C., *et al.* (2020). BMC genomics, 21(1), 1-10.

Rakyan, V. K., Down, T. A., *et al.* (2011). Nature Reviews Genetics, 12(8), 529-541.

Snelling, W. M., Hoff, J. L., Li, J. H., Kuehn, L. A., *et al.* (2020). Genes, 11(11), 1312.

Stafuzza, N. B., de Oliveira Silva, R. M., *et al.* (2019). BMC genomics, 20(1), 1-11.

Wang, H., Wang, C., Yang, K., Liu, J., Zhang, Y., *et al.* (2015). PLoS One, 10(7), e0131522.

Wang, K., Wu, P., Chen, D., Zhou, J., Yang, X., *et al.* (2020). https://doi.org/10.21203/rs.3.rs-26863/v1

254. Successful trio-based reverse genetic screen in an endangered local cattle breed

I.M. Häfliger[1*], F.R. Seefried[2] and C. Drögemüller[1]

[1]*Institute of Genetics, University of Bern, Bremgartenstrasse 109a, 3000 Bern, Switzerland;*
[2]*Qualitas AG, Chamerstrasse 56, 6300 Zug, Switzerland;* irene.haefliger@vetsuisse.unibe.ch

Abstract

We scanned the genomic data of almost one thousand Evolèner cattle, representing the by far smallest local cattle population of Switzerland, whereof only pedigree records are available. A reverse genetic screen using 94k SNP markers in 585 trios mapped recessive deleterious loci using homozygous haplotype deficiency. We found two haplotypes, EH1 and EH2, with moderate allele frequencies of >0.10. Mining for candidate variants was carried out by linkage analysis of the predicted haplotype status and whole-genome sequencing variant catalogue of seven Evolèner bulls. This led to the detection of two perfectly linked missense variants affecting conserved residues: *GBE1*:p.Arg437Gln for EH1, and *LRRC8A*:p.Ala73Val for EH2. No homozygous animals were observed in >5,100 cattle of various breeds including Evolèner. The presented study showed that very limited data can lead to the identification of candidate variants and thereby help to improve reproduction success in an endangered indigenous breed of cattle.

Introduction

The name of the Evolèner (EV) breed originates from the village called Evolèner in the municipality Val d'Hérens in the Valais; however, genetically the breed evolved from the local Swiss breed Eringer (in French Hérens). EV cattle are characterized by their small body size and by white markings on their foreheads, while their basic coat colour ranges from red to black. The breeds current population is spread across Switzerland and includes 1,175 animals (Dec 31, 2021), of which 823 animals are registered in the herdbook (A. Barenco, personal communication), whereof a proportion of 44% is genotyped. Together with genotypes from previous generations, almost 1000 genotype records were available based on a public SNP array encompassing 150k markers.

Recessive lethal variants had been identified by mapping haplotypes showing significant deviation from the Hardy-Weinberg equilibrium (HWE) in population-wide single nucleotide polymorphism (SNP) array data and subsequent whole-genome sequencing (WGS) data analysis of selected individuals (e.g. Schwarzenbacher *et al.*, 2016; VanRaden *et al.*, 2011). Usually, this approach is applied on large, economically important populations, where the SNP data is generated under the umbrella of genomic selection. Therefore, it is more challenging for smaller populations to identify harmful recessive alleles that impair fertility. Recently, population-wide SNP genotypes have been generated within the framework of a national genetic diversity project aiming at conserving the EV breed. Even though the power for haplotype detection was limited due to the relatively small dataset, the trio-based approach applied allowed the identification of two haplotypes showing significant depletion in homozygosity and the associated candidate causal variants affecting genes important for reproductive success.

Materials & methods

Genotype and pedigree data was provided by the breeding organisation (swissherdbook) and included all genotyped EV cattle born later than 2009 as well as their ancestors. Altogether 964 genotype records were available based on a publicly available SNP array encompassing 150K markers. The genomic positions relate to the latest cattle reference sequence ARS-UCD1.2.

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_254, © I.M. Häfliger *et al.* 2022

SNP data analysis. As the initial dataset included different batches, genotype data was imputed using Fimpute v3 software (Sargolzaei *et al.*, 2014) with default parameters. SNP data was pruned before imputation on animal and on marker level. Samples were used if the sample call rate was higher than 0.95. Marker were kept in the dataset if each of the following criteria were achieved: minor allele frequency >0.01, call rate per SNP>0.99, and <1% mendelian errors. The final dataset contained 94,345 markers and 964 animals. Imputed genotypes were screened for haplotypes showing a deviation from HWE by depletion of homozygosity in a dataset that includes complete trios (sire, dam and offspring; n=585). Thereby a 50 SNP sliding window approach with overlap within the software snp1101 was applied (Sargolzaei, 2014). Within each region the most significant haplotype that also passed the Bonferroni correction ($P<1.68E^{-4}$) was used for further analysis in WGS data.

WGS data. For each of the selected haplotypes, diplotypes were predicted for the genotyped population to represent an individuals' haplotype status. Aiming at three carrier animals selected per haplotype, five bulls were chosen for whole-genome sequencing. WGS data was prepared for these five samples and additional 196 available samples from other projects as previously described (Häfliger *et al.*, 2021; data published in the European Nucleotide Archive: PRJEB18113). Further genomic information was provided by the international consortium of the 1000 Bull Genomes project run 9, which included 5,116 animals from various breeds (Hayes and Daetwyler, 2019). Finally, the combined WGS dataset encompassed 5,317 genomes, including eight Evolèner cattle.

Variant interpretation. A linkage disequilibrium (LD) approach was performed to identify haplotype-associated variants in WGS data by using plink v1.9 (Purcell *et al.*, 2007). LD (r2) was calculated using the diplotype for each haplotype and WGS-derived protein-changing genotypes for seven EV animals that had both types of genomic information available. For improving the understanding of associated variants, the effects of the amino acid changes were estimated using PROVEAN (Choi *et al.*, 2012) and conservation scores PhyloP and PhastCons (Pollard *et al.*, 2010; Siepel *et al.*, 2005).

Results

With the above described approach two different haplotypes were identified that showed significant deviation from HWE: Evolèner Haplotype 1 (EH1), on chromosome 1 between the positions 29,769,321 and 30,160,645 and Evolèner Haplotype 2 (EH2) on chromosome 11 from 97,965,320 to 99,366,886. EH1 and EH2 were expected to be observed in homozygous state in the analysed dataset in 13 and 16 animals, respectively. The haplotype frequency for EH1 and EH2 in EV was estimated to be >0.1, while a positive or negative trend was not observed over the last 14 year of births. Table 1 provides a summary of the two identified missense variants that show perfect LD ($r^2=1$) with the haplotypes.

Firstly, the EH1-associated variant in *GBE1* shows high conservation scores at DNA level (PhyloP=7.4; PhastCons=1) and the predicted amino acid exchange is most likely deleterious (PROVEAN score=-3.7), as it is situated in the functionally important catalytic domain of the GBE1 enzyme (Figure 1A). Secondly, the EH2-associated genomic variant in *LRRC8A* is not highly conserved across species (PhyloP=0.19;

Table 1. Candidate causal variants for deficiency of homozygotes in Evolèner.

Haplotype	Gene	OMIM	Transcript[1]	Coding DNA change	Predicted protein change
EH1	*GBE1*	607839	NM_001122729.1	c.1310G>A	p.Arg437Gln
EH2	*LRRC8A*	608360	NM_001076807.1	c.218C>T	p.Ala73Val

[1] According to the NCBI Annotation Release 106.

PhastCons=0.002) and is predicted to be neutral (PROVEAN score=-0.49) on protein level, although the residue is located in the first extracellular loop of the LRRC8A protein (Figure 1B).

Both candidate variants were validated in 5,317 individuals with WGS data (Table 2). The EH1-associated *GEB1* variant could only be observed in EV animals, while the EH2-associated LRRC8A variant was observed in EV, as well as in Eringer and Gelbvieh cattle. These results lead to the conclusion that only the EH1-associated variant is breed specific, but both variants could not be observed in homozygous state in more than 5K animals.

Discussion

By applying a reverse genetic approach on a limited number of 595 completely genotyped trios, the SNP data of less than 1000 animals led to the identification of two homozygous deficient haplotypes, and the WGS data of seven animals pinpoint to two candidate causative variants. As far as we know, this is the first successful analysis based on such a limited dataset, concerning a small population for which not any phenotypic information is available. The EH1-related candidate variant in the essential *GBE1* gene, represents a most likely embryonic lethal mutation (Froese *et al.*, 2015). The gene is associated with the recessive, sometimes fatal, glycogen storage disorder type IV known in human, horse, and cat (OMIM: 607839; OMIA: 000420-9796, 000420-9685). Variants in the catalytic domain had been shown to destabilize the enzyme (Froese *et al.*, 2015). The exactly corresponding missense variant in the human *GBE1* gene is reported as rs770819412 and predicted to be likely disease causing; however, no homozygous variant carriers were observed and thereby no phenotype was assigned. These indications support our hypothesis, that we cannot observe

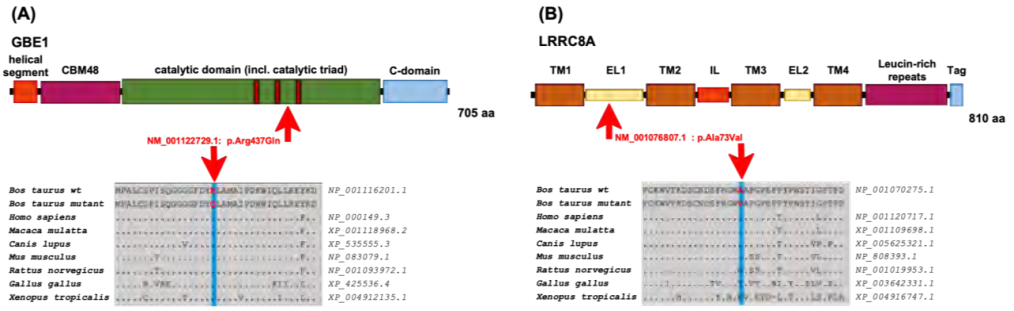


Figure 1. Protein domains and multiple species alignments of the amino acid sequences around the candidate variants. (A) The EH1-related variant in the *GBE1* gene (Froese *et al.*, 2015) and (B) the EH2-related variant in *LRRC8A* (Choi *et al.*, 2021). The abbreviated domains in LRRC8A mean transmembrane domain (TM), extracellular loop (EL), and intracellular loop (IL).

Table 2. Genotype distribution of the candidate variants in all whole-genome sequencing data, including Evolèner cattle.

	WGS data			
Variant	**ref/ref**	**ref/var[1]**	**var/var**	**./.**
EH1 (*GEB1*:c.1310G>A)	5,312	5	0	0
EH2 (*LRRC8A*: c.218C>T)	5,187	10	0	120

[1] Carrier animals of the EH1-associated variant are all of the breed EV; carrier animals of EH2-associated variant are of the breeds EV (4), Eringer (1), Gelbvieh (4), or unknown breed (1).

homozygous individuals or EH1-related cases, due to its embryonic lethal effect. The EH2-related candidate variant in *LRRC8A* gene, represents a likely candidate for a fatal postnatal phenotype. The protein is an essential compound of the volume-regulated anion channel, that maintains cellular homeostasis and is associated with the dominant disorder agammaglobulinemia 5 in human (OMIM: 608360; Choi *et al.*, 2021). In Lrrc8a knockout mice, increased pre- and postnatal mortality, growth retardation and tissue anomalies were observed (Kumar *et al.*, 2014). The extracellular loops were shown to be important oxidation sites during immune responses (Choi *et al.*, 2021). In conclusion, we identified two novel haplotypes and candidate variants potentially affecting fertility in the Evolèner population. Both protein-changing variants affecting candidate genes for developmental defects were never observed homozygous in more than 5,000 animals, and show perfect LD with the deficient homozygous haplotypes. Further steps include validation using farm questionnaires and *in vivo* data. Nevertheless, avoidance of risk matings can be immediately performed due to the ongoing comprehensive genotyping strategy in that population.

References

Choi H., Rohrbough J.C., Nguyen H.N., Dikalova A., and Lamb F.S. (2021) Oxidant-resistant LRRC8A/C anion channels support superoxide production by NADPH oxidase 1. J. Physiol. 599, 3013–3036. https://doi.org/10.1113/JP281577

Choi Y., Sims G.E., Murphy S., Miller J.R., and Chan A.P. (2012) Predicting the functional effect of amino acid substitutions and indels. PLoS One 7, e46688. https://doi.org/10.1371/journal.pone.0046688

Froese D.S., Michaeli A., McCorvie T.J., Krojer T., Sasi M., *et al.* (2015) Structural basis of glycogen branching enzyme deficiency and pharmacologic rescue by rational peptide design. Hum. Mol. Genet. 24, 5667–5676. https://doi.org/10.1093/hmg/ddv280

Häfliger I.M., Seefried F.R., Spengeler M., and Drögemüller C. (2021) Mining massive genomic data of two Swiss Braunvieh cattle populations reveals six novel candidate variants that impair reproductive success. Genet. Sel. Evol. 53, 1–16. https://doi.org/10.1186/s12711-021-00686-3

Hayes B.J., and Daetwyler H.D. (2019) 1000 Bull Genomes Project to map simple and complex genetic traits in cattle: Applications and outcomes. Annu. Rev. Anim. Biosci. 7, 89–102. https://doi.org/10.1146/annurev-animal-020518-115024

Kumar L., Chou J., Yee C.S.K., Borzutzky A., Vollmann E.H., *et al.* (2014) Leucine-rich repeat containing 8A (*LRRC8A*) is essential for T lymphocyte development and function. J. Exp. Med. 211, 929–942. https://doi.org/10.1084/jem.20131379

Pollard K.S., Hubisz M.J., Rosenbloom K.R., and Siepel A. (2010) Detection of nonneutral substitution rates on mammalian phylogenies. Genome Res. 20, 110–121. https://doi.org/10.1101/gr.097857.109

Purcell S., Neale B., Todd-Brown K., Thomas L., Ferreira M.A., *et al.* (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. Am. J. Hum. Genet. 81, 559–575. https://doi.org/10.1086/519795

Sargolzaei M. (2014) SNP1101 User's Guide. Version 1.0. Guelph HiggsGene Solut. Inc.

Sargolzaei M., Chesnais J.P., and Schenkel F.S. (2014). A new approach for efficient genotype imputation using information from relatives. BMC Genomics 15, 478. https://doi.org/10.1186/1471-2164-15-478

Schwarzenbacher H., Burgstaller J., Seefried F.R., Wurmser C., Hilbe M., *et al.* (2016) A missense mutation in TUBD1 is associated with high juvenile mortality in Braunvieh and Fleckvieh cattle. BMC Genomics 17, 400. https://doi.org/10.1186/s12864-016-2742-y

Siepel A., Bejerano G., Pedersen J.S., Hinrichs A.S., Hou M., *et al.* (2005) Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. Genome Res. 15, 1034–1050. https://doi.org/10.1101/gr.3715005

VanRaden P.M., Olson K.M., Null D.J., and Hutchison J.L. (2011) Harmful recessive effects on fertility detected by absence of homozygous haplotypes. J. Dairy Sci. 94, 6153–6161. https://doi.org/10.3168/jds.2011-4624

255. How does a harsh environment make wild and domestic animals evolve similarly?

G. Romashov[1], A. Yurchenko[1], N. Yudin[1], and D. Larkin[1,2*]

[1]Institute of Cytology and Genetics, Siberian Branch of Russian Academy of Sciences, Lavrentyeva ave. 10, 630090, Novosibirsk, Russia; [2]Royal Veterinary College, Royal College St, NW1 0TU, London, United Kingdom; dmlarkin@gmail.com

Abstract

This study has focused on detecting putative convergently evolved or selected sites in proteins of the northernmost Yakut cattle breed and a set of 12 hibernating, cold adapted, and deep diving mammals. A novel *de novo* assembled, and annotated Yakut cattle genome has been utilised as a reference. Over 8000 orthogroups have been identified and analysed with the *PCOC* software. After an additional filtering, 18 proteins were identified including a previously reported *nebulin related anchoring protein*. Twelve out of the 18 sites were present as SNPs in the 1000 Bull Genomes Project dataset of which five demonstrated high FST (>0.3) between the Yakut and other cattle breeds. Our results imply that convergent evolutionary events are rare between breeds and wild animals, but the same ancestral alleles could be selected in parallel in domestic animal breeds and wild species expressing similar phenotypes.

Introduction

Similar ecological niches lead to similar phenotypes and adaptions in different animals. In a few known cases there are convergent nucleotide changes involved in these adaptations in distinct species. Well-known examples include echolocation in whales and bats in mammals involving the same convergent amino acid mutation in the plestin protein (Liu *et al.*, 2014) and adaptation to different water light conditions in fish involving the same mutation in rhodopsin protein (Hill *et al.*, 2019). Our recent work on the northernmost cattle from Siberia (Yakut cattle) has revealed the phenomenon of convergent nucleotide evolution in breeds of domestic animals and wild species living in similar harsh environments and/or expressing similar phenotypes. We found the same amino acid change in a nebulin related anchoring protein (NRAP) in the northernmost cattle and 16 species of hibernating or entering torpor, cold-adapted, and deep-diving mammals which was absent in all other cattle breeds and other *Bovinae* species from the 1000 Bull Genomes project dataset (Buggiotti *et al.*, 2021). This change was occurred during the last 500-800 years and became nearly fixed in the Yakut cattle population. In the present work we extended the search for nucleotide convergent changes to the whole genome assembly and *de novo* annotated proteins of the Yakut cattle, proteins of 12 hibernating/cold-adapted/deep-diving mammals, and five species not expressing these phenotypes (outgroups). Our results have confirmed amino acid change in the *NRAP* gene and pointed to 17 additional genes with coding variants putatively selected in parallel in the northernmost cattle and wild animals.

Materials & methods

Sequencing the Yakut cattle genome. A Yakut cattle sample have been obtained from Juha Kantanen, Natural Resources Institute Finland (Luke), Finland. DNA was extracted using 100G Qiagen High Molecular DNA Kit. 10X Genomics sequencing has been done at Novogene, Cambridge. Assembly was generated using Supernova assembler on an HPC cluster. The 10x Genomics assembly was then merged with the Oxford Nanopore reads (~1X) using LRScaf (https://github.com/shingocat/lrscaf) to produce a hybrid Yakut cattle assembly. The scaffolds were further assembled into chromosomes using a reference-assisted approach (RagTag; https://github.com/malonge/RagTag) against the reference cattle genome (ARC-UCD1.2). Protein predictions have been done using GeMoMa software (v.1.7; http://www.jstacs.de/index.php/GeMoMa), using taurine cattle, indicine cattle, yak, bison, and buffalo gene annotations.

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_255, © G. Romashov *et al.* 2022

Species list and protein sequences processing. The following species and phenotypic groups have been used in the analysis: deep diving (*Odobenus rosmarus divergens, Physeter catodon, Eumetopias jubatus, Mirounga leonine),* hibernating (*Myotis brandtii, Marmota marmota marmota, Microcebus murinus, Chrysochloris asiatica),* cold-resistant *(Vulpes lagopus, Rangifer tarandus, Ovis nivicola, Ursus maritimus, Odobenus rosmarus divergens)* and outgroups (*Homo sapiens, Loxodonta africana, Canis lupus familiaris, Pteropus alecto, Bos taurus* (Hereford)). Coding sequences (CDSs) for all species have been downloaded from NCBI RefSeq or predicted with GeMoMa (v. 1.7). From the isoforms of the same protein the longest sequence has been used. The protein sequences have been obtained with PRANK (http://wasabiapp.org/software/prank/) using *-convert -translate* options from gff files. Orthogroups were identified with OrthoFinder (v. 2.5.2; https://github.com/davidemms/OrthoFinder) and additionally filtered to ensure each species had maximum one sequence in each orthogroup and each orthogroup contained ≥3 orthologous sequences, gaps were removed. Nucleotide sequences within each orthogroup were aligned with PRANK (v. 170703).

Detection of convergently evolving amino acid residues. PCOC method looking both at exact same amino acids changes and shifts in amino acid type preferences (Rey *et al.*, 2018) has been applied. For detection of the putative convergent changes between phenotypic groups the detection threshold has been set to 0.8. To identify convergent changes between the Yakut cattle and the ingroup species, all amino acids shared between the Yakut cattle and Hereford reference genome were filtered out.

Results

The hybrid (10X Genomics and Oxford Nanopore) Yakut cattle genome was assembled to N50=15.1 Mbp. A total of 3,805 scaffolds were ordered along 30 cattle chromosomes and the mitochondrial genome. The final Yakut cattle annotation contained 39,412 transcripts. Of these 22,323 were used in the analysis. A total of 8,939 orthogroups each containing one Yakut cattle transcript and the transcripts from at least 14 other mammalian species were generated.

The PCOC algorithm has been applied to the set of 8,939 orthogroups. After filtering for the number of species present in each ingroup and the outgroup, 18 putative convergent events were found in 18 genes. The largest number of putative convergently evolving amino acids with the Yakut cattle have been detected in the deep-diving animals (14), followed by hibernating (2) animals (Figure 1). This analysis did detect the previously found *NRAP* convergent change and identified putative events in 17 additional genes.

Comparison to the 1000 Bull Genomes (Run 9) variant database showed that *NRAP* event is indeed Yakut-cattle specific confirming its convergent formation. Four additional variants had FST>0.3 when Yakut cattle was compared to other taurine breeds, suggesting differences in allele frequencies between Yakut and taurine cattle and a higher frequency in Yakut cattle for the allele present in the deep diving, hibernating or cold adapted animals. This implies no convergent evolution but a parallel selection for the same ancestral amino acids in Yakut cattle and wild animals different from most other cattle breeds. The functional analysis of 18 genes containing convergent and selected in parallel changes revealed two genes previously found under selection in the Siberian human populations (*PLA2G4F* (FST=0.98) and *PAFAH1B2* (FST=0.01) (Hsieh *et al.*, 2017)), a gene involved in resistance to environmental toxicants (*FNBP4* (FST=0.63) (Zhu *et al.*, 2015)), genes involved in controlling the heart function (*NRAP* (FST=1.0) and *POPDC2* (FST=0.3) (Vasilescu *et al.*, 2018)), in regulation of blood pressure and insulin secretion (*SNX19* (FST=0.53) (Yang *et al.*, 2019)), and shortening of long bones (*WDR60* (FST=0.95) (McInerney-Leo *et al.*, 2013)).

Table 1. Genes with the convergently formed or selected in parallel amino acids in the Yakut cattle and deep diving, hibernating, and cold-adapted mammals (FST[3]0.3[1]).

Gene	Function	FST[1]	PCOC[2]
PLA2G4F	Lipid metabolism, cold adaptation	0.98	DD, HB
FNBP4	Resistance to environmental toxicants	0.63	DD
NRAP	Muscle contraction, heart function	1.00	HB
POPDC2	Heart rate	0.30	DD
SNX19	Blood pressure, insulin secretion	0.53	DD
WDR60	Long bone shortening, nervous system	0.95	HB

[1] Against the 1000 Bull Genomes project (Run9), taurine dataset.

[2] DD – deep diving animals, HB – hibernating animals.

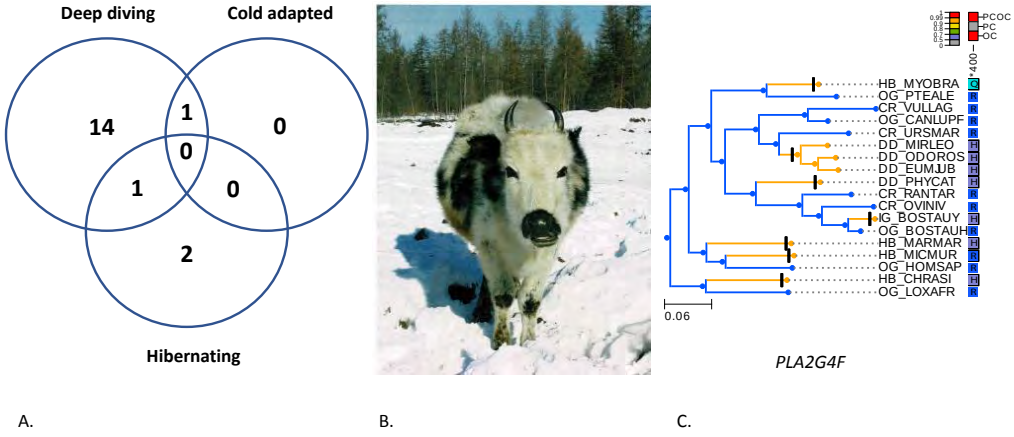


Figure 1. Analysis of convergent evolution and parallel selection in the Yakut cattle and mammalian species. (A) Number of convergently evolved and selected in parallel amino acids shared with three phenotypic groups and the Yakut cattle. (B) An image of the Yakut cattle. (C) Output of the PCOC algorithm for *PLA2G4F*. The histidine amino acid is present in Yakut cattle, deep diving (DD) and hibernating (HB) mammals, but not in Hereford (OG_BOSTAUH) and cold-adapted animals (CR).

Discussion

The aim of this study was to evaluate to what extent convergent nucleotide evolution has shaped the genome of the northernmost cattle from Siberia (Yakut cattle) utilising breed-specific *de novo* genome assembly and *de novo* protein annotations. In agreement with our previous study, we identified a missense mutation in *NRAP* which has occurred independently in Yakut cattle and the cold-adapted, hibernating, and deep-diving mammals (Buggiotti *et al.*, 2021). The current study has, however, pointed to a list of several additional genes which could have convergently evolving amino acid sites in the Yakut cattle and other animals. Five of these genes, however, had no SNPs in the 1000 Bull Genomes Project. Mutations in three of these genes were confirmed in 20 Yakut and other breed individuals aligned against our Yakut cattle assembly while two were transcript alignment artefacts. The fact that FST values for all three confirmed variants were <0.3 suggests that they could have a relatively minor effect on the Yakut cattle as a breed.

Our study has been successful in finding 12 additional mutations which were present in the 1000 Bull Genomes dataset. However, only five of these sites had a substantial difference in the allele frequency between the Yakut cattle and other taurine breeds ($FST^3$0.3). None of the mutations were Yakut-cattle specific, suggesting that these variants were likely selected in parallel in the Yakut cattle and mammals with similar phenotypes rather than occurred independently in the Yakut cattle and other mammals. Importantly, all Yakut cattle alleles were present in other *Bovinae* species (1000 Bull Genomes dataset) implying their likely ancestral origin. Yakut cattle alleles for two genes (*SNX19* and *POPDC2*) were more frequent in non-taurine *Bovinae* species than in cattle breeds.

Our study demonstrates that *de novo* convergent nucleotide evolution in breeds of domestic and wild species could be a rare phenomenon, in line with rare functional convergent nucleotide changes in different species, but the *parallel selection* for alternative ancestral alleles could be more widespread. The use of the population datasets is crucial to distinguish between a true signal and noise. Both the convergently evolving and selected in parallel alleles should be considered when studying the biology of wild animals and livestock breeds and designing the new generations of locally adapted livestock.

Acknowledgements

The work was funded by the Kurchatov Genomics Center of IC&G (075-15-2019-1662).

References

Buggiotti, L., Yurchenko, A. A., Yudin, N. S., vander Jagt, C. J., Vorobieva, N. v, *et al.* (2021). Demographic History, Adaptation, and NRAP Convergent Evolution at Amino Acid Residue 100 in the World Northernmost Cattle from Siberia. *Molecular Biology and Evolution*, *38*(8), 3093–3110. https://doi.org/10.1093/molbev/msab078

Hill, J., Enbody, E. D., Pettersson, M. E., Sprehn, C. G., Bekkevold, D., *et al.* (2019). Recurrent convergent evolution at amino acid residue 261 in fish rhodopsin. *Proceedings of the National Academy of Sciences of the United States of America*, *116*(37), 18473–18478. https://doi.org/10.1073/pnas.1908332116

Hsieh, P., Hallmark, B., Watkins, J., Karafet, T. M., Osipova, L. P., *et al.* (2017). Exome Sequencing Provides Evidence of Polygenic Adaptation to a Fat-Rich Animal Diet in Indigenous Siberian Populations. *Molecular Biology and Evolution*, *34*(11), 2913–2926. https://doi.org/10.1093/molbev/msx226

Liu, Z., Qi, F.-Y., Zhou, X., Ren, H.-Q., and Shi, P. (2014). Parallel sites implicate functional convergence of the hearing gene prestin among echolocating mammals. *Molecular Biology and Evolution*, *31*(9), 2415–2424. https://doi.org/10.1093/molbev/msu194

McInerney-Leo, A. M., Schmidts, M., Cortés, C. R., Leo, P. J., Gener, B., *et al.* (2013). Short-Rib Polydactyly and Jeune Syndromes Are Caused by Mutations in WDR60. *The American Journal of Human Genetics*, *93*(3), 515–523. https://doi.org/10.1016/j.ajhg.2013.06.022

Rey, C., Guéguen, L., Sémon, M., and Boussau, B. (2018). Accurate Detection of Convergent Amino-Acid Evolution with PCOC. *Molecular Biology and Evolution*, *35*(9), 2296–2306. https://doi.org/10.1093/molbev/msy114

Vasilescu, C., Ojala, T. H., Brilhante, V., Ojanen, S., Hinterding, H. M., *et al.* (2018). Genetic Basis of Severe Childhood-Onset Cardiomyopathies. *Journal of the American College of Cardiology*, *72*(19), 2324–2338. https://doi.org/10.1016/j.jacc.2018.08.217

Yang, J., Villar, V. A. M., Rozyyev, S., Jose, P. A., and Zeng, C. (2019). The emerging role of sorting nexins in cardiovascular diseases. *Clinical Science*, *133*(5), 723–737. https://doi.org/10.1042/CS20190034

Zhu, J., Dubois, A., Ge, Y., Olson, J. A., and Ren, X. (2015). Application of human haploid cell genetic screening model in identifying the genes required for resistance to environmental toxicants: Chlorpyrifos as a case study. *Journal of Pharmacological and Toxicological Methods*, *76*, 76–82. https://doi.org/10.1016/j.vascn.2015.08.154

256. Functional and association analyses of sequence data in a region of OAR 20 associated to paratuberculosis in sheep

S. Casu[1*], M.G. Usai[1], T. Sechi[1], C. Maestrale[2], S. Miari[1], S. Salaris[1], C. Ligios[2] and A. Carta[1]

[1]*AGRIS Sardegna, Research Unit: Genetics and Biotecnology, località Bonassai, 07040 Olmedo, Italy;* [2]*Istituto Zooprofilattico Sperimentale della Sardegna G. Pegreffi, Via Duca degli Abruzzi 8, 07100 Sassari, Italy;* *scasu@agrisricerca.it*

Abstract

The aim of this study was to search for mutations influencing immune response of sheep against *Mycobacterium avium* subsp *Paratuberculosis* within a region from 23 to 35 Mb on chromosome 20 previously detected in a nucleus flock of Sarda sheep. Functional annotation and association analysis of polymorphisms identified in the whole genome sequences of 56 animals and imputed into around 3,000 ewes were realized. Accurate imputation led to more precisely localize the quantitative trait locus between 23.77 to 25.83 Mb, in a known 'markers desert' of the OvineSNP50 chip. No significant variants showed functional effects on the genes mapping to this region, most of which belong to the major histocompatibility complex (MHC) II. However, we cannot exclude that such variants may reveal some functional meaning with the evolution of the assembly and the annotation of the sheep genome, especially in this region where accurate mapping of MHC genes is challenging.

Introduction

Paratuberculosis (PTB) is a contagious slowly developing disease, caused by *Mycobacterium avium* subsp. *Paratuberculosis* (MAP) that affects mainly ruminants, with adverse effects on animal welfare and productivity. No treatment exists and the benefits of vaccination are controversial. Genetic variability of susceptibility is documented (Sechi *et al.*, 2011; Küpper *et al.*, 2012) and several quantitative trait loci (QTL) or polymorphisms in candidate genes have been identified, especially in dairy cattle (Sanchez *et al.*, 2020). In sheep, genetic studies are fewer and often based on candidate gene approach or small size experiments (Moioli *et al.*, 2016). Thanks to a genotyped population of around 3,000 ewes with repeated records of serological Elisa tests against MAP antibodies, we were able to detect a genomic region on chromosome (OAR) 20 significantly associated to PTB resistance (Usai *et al.,* this congress). Full genomic sequences of 56 influential animals were available. The aim of this study was to further investigate this region by the functional annotation and association analysis of imputed polymorphisms to search for candidate genes and causative mutations.

Materials & methods

Whole genome sequences (WGS) were available for 56 animals from the experimental population genotyped with the OvineSNP50 chip. Twenty-four animals were resequenced with a coverage of 12X and 32 with a coverage of 30X. WGS data were processed with a pipeline implemented with Snakemake (Köster, 2018), and developed at CRS4 (Center For Advanced Studies, Research and Development in Sardinia https://www.crs4.it/)). The quality control was performed with FastQC (Andrews, 2010). Trimmed reads were aligned, using the BWA (v0.7.15) aligner (Li and Durbin, 2009), to the *Ovis aries* reference genome v4.0 (https://www.ncbi.nlm.nih.gov/assembly/GCF_000298735.2). Joint variant calling was performed using the GATK (v4.0.11.0) software (McKenna *et al.*, 2010). We focused on a large region on OAR20 (from to 23.12 to 35.13 Mb) surrounding the significant peak identified by linkage disequilibrium linkage analysis (Usai *et al.*, this congress). In this region, 217,248 single nucleotide polymorphisms (SNPs) were identified and imputed to the phenotyped ewes with 50k SNP data. The probabilities of carrying the reference allele (R) in the paternal ($P(h^p=R)$) and maternal ($P(h^m=R)$) gametes were predicted. First, gametes of sequenced animals were phased, on the base of the genotype information from sequencing and identity-by-descent

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_256, © S. Casu *et al.* 2022

probability (IBD) between gametes at the neighbouring SNP 50k. Then, at each WGS polymorphism of the parental gamete, we inferred the probabilities of carrying R based on the gametic phase of sequenced animals and the IBD between gametes of sequenced animals with the phenotyped ewes (Usai *et al.*, 2019). Accuracy of imputation was assessed by 56 leave-one-out cross validations as the correlation for each left out animal between the allele dosage from imputation (calculated as $P(h^p=R)+P(h^m=R)$) and the allele dosage from sequencing. The association analysis was then ran by regressing the phenotypes on the allele dosage. The analysed phenotype was the status of the animals, defined as 'susceptible' if the animal had at least one positive Elisa test along the recording period. An F-test was performed to calculate the *P*-values of each tested position. Functional annotation of the WGS SNPs surrounding the most significant location was then performed using the NCBI 4.0 sheep genome annotation release 102 and the SnpEff software v4.3.t (Cingolani *et al.*, 2012). Finally, the genes with the most significant SNPs or with SNPs predicted to have HIGH impact or missense effect, annotated by orthology to human genes from the OrthoDB v10 data base, were analysed using the KEGG and gene ontology (GO) biological process terms data bases with the web-based software WebGestalt (Liao *et al.*, 2019).

Results

The accuracy of imputation ranged from 0.843 to 0.984 across animals (0.928 on average) with an average information content across gametes of 0.86±0.02 (from 0 for only 42 SNPs to 0.89). The most significant location ($-\log_{10}$(*P*-value)=20.2) was at 25,253,298 bp, 393 kb from the LDLA peak. We focused on the interval (23.77 to 25.83 Mb), empirically defined by the extreme bounds of genes with SNPs showing $-\log_{10}$(*P*-value)>15, that included 35,506 SNPs. SNPeff provided 167,562 effects: 10 and 77% concerned intergenic and intronic regions. The most significant variants ($-\log_{10}$(*P*-value)>15) were classified as modifier, since they were either intergenic or intronic. In particular, among the 20 most significant positions, 18 were intergenic between ELOVL5 (elongation of very long chain fatty acids protein 5) and LOC101108696 (SLA class II histocompatibility antigen, DQ haplotype D alpha chain-like) genes, one affected the fifth intron of LOC101106976 (glutathione S-transferase A4-like) gene and one was intergenic, between DQA (HLA class II histocompatibility antigen, DQ alpha 2 chain) and LOC101120871 (H-2 class II histocompatibility antigen, E-S beta chain-like). The most highly significant variants mapping within coding regions had only LOW impact on transcripts, being synonymous polymorphisms that affected the genes DQA, DQB (boLa class II histocompatibility antigen, DQB*0101 beta chain) and LOC101109492 (SLA class II histocompatibility antigen, DQ haplotype D alpha chain). Twelve and 383 variants were predicted to have HIGH or MODERATE impact effect on the multiple transcripts of several genes. The twelve HIGH impact variants were identified in stop codons or splice sites of BTNL2 (butyrophilin like 2), PKHD1 (PKHD1 ciliary IPT domain containing fibrocystin/polyductin), LOC101110277 (butyrophilin-like protein 1) and LOC101106976 (glutathione S-transferase A4-like) genes. However, all of them showed low significance ($-\log_{10}$(*P*-value)<5). DQA, BTNL2 and PKHD1 genes also harbour the most significant missense variants ($-\log_{10}$ (*P*-value)>9). Overall, 28 out of 31 protein coding genes in the interval harbour variants that showed HIGH impact, missense effect or high statistical significance.

Functional classification of these genes showed that the top ranked biological functions were all related to immune response processes: GO:0050852: T-cell receptor signaling pathway; GO:0060333: interferon-gamma-mediated signaling pathway; and GO:0019886: antigen processing and presentation of exogenous peptide antigen via MHC class II. Moreover, the top ranked KEGG pathway was Inflammatory bowel disease, a term that includes chronic gastrointestinal inflammatory disorders of humans such as ulcerative colitis and Crohn disease. Thirteen genes enriched this pathway, two of which encode for interleukins (IL17F and IL17A) and 11 for MHC class II molecules (DQA, DQB, LOC101108696, LOC101109220, LOC101109492, LOC101109747, LOC101119856, LOC105603927, OVAR-DRB1, OVAR-DRB3). All these 11 genes also enriched the three most represented GO processes.

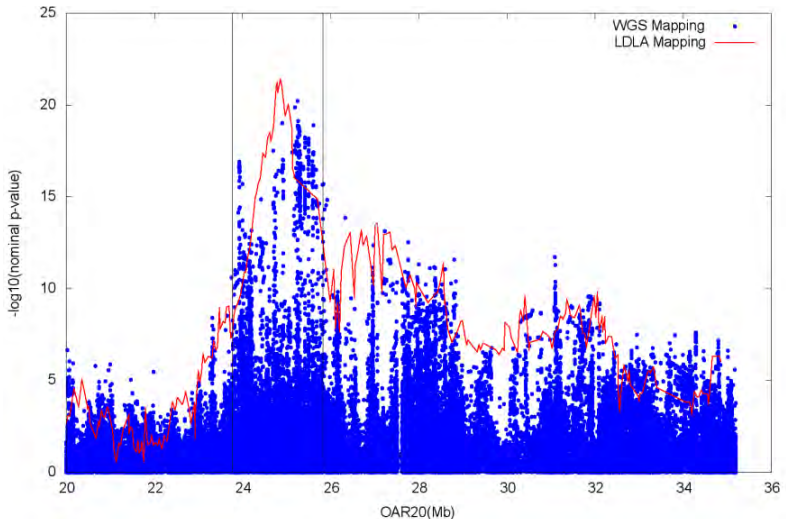


Figure 1. Test statistics profile of the LDLA analysis and Manhattan plot of the association analysis based on WGS data. Vertical grey lines define the interval investigated with the functional annotation.

Discussion

The initially investigated region (23.12-35.13 Mb) partially overlaps the region on bovine chromosome 20 where Sanchez *et al.* (2020) identified several significant SNPs in cattle. However, the candidate gene IER1 suggested by Sanchez *et al.* (2020), mapping between 27255853-27257183bp on OAR20 in Oar4.0 assembly, was far from the most significant position. The peak of the WGS association analysis falls in a known 'SNP desert' in the OvineSNP50 chip (Dicks *et al.*, 2021) where the MHC class IIa genes (DQA, DQB, OVAR-DRB1, OVAR-DRB3) are supposed to occur. Association between MHC alleles and susceptibility to MAP have already been suggested in sheep (Reddacliff *et al.*, 2005) and differential expression of MHC class II genes have been observed in experimentally infected (David *et al.*, 2014) or subclinically affected (Verschoor *et al.*, 2010) cows. MHC II genes are involved in both innate and adaptive immune system. Mutations that modify the expression of these genes or the ability to activate T cells have a systemic effect on the host, causing an inappropriate immune response that could change the course of the infection (Kravitz *et al.*, 2021). Although Gossner *et al.* (2017) did not evidence differential expression of MHCII genes in ileocaecal lymph node of ewes with paucibacillary or multibacillary PTB, Purdie *et al.* (2019) found a differential expression of MHC-II DQ alpha and beta genes in 'resilient' sheep. Combining association analysis with the functional annotation of SNPs within the target region was not able to identify causative mutations in this study. However, it cannot be excluded that highly significant variants located outside coding regions may reveal functional meaning with the evolution of the annotation of the sheep genome.

References

Andrews, S (2010) FastQC. https://www.bioinformatics.babraham.ac.uk/projects/fastqc

Cingolani P, Platts A, Wang LL, Coon M, Nguyen T *et al.* (2012). *Fly*, 6(2) 80-92. https://doi.org/10.4161/fly.19695

David J, Barkema HW, Guan LL, and De Buck J (2014). *Vet.Res.* 45(1) https://doi.org/10.1186/s13567-014-0096-5

Dicks KL, Pemberton JM, Ballingall KT, and Johnston SE (2021). *G3:-Genes Genom-Genet* 11(10). https://doi.org/10.1093/g3journal/jkab200

Gossner A, Watkins C, Chianini F, and Hopkins J (2017). *Sci Rep* 7(1), 46695 https://doi.org/10.1038/srep46695

Köster J and Rahmann S (2018). *Bioinformatics*, 34(20), 3600. https://doi.org/10.1093/bioinformatics/bty350

Kravitz A, Pelzer K and Sriranganathan N (2021). *Front Vet Sci*, https://doi.org/10.3389/fvets.2021.721706

Küpper J, Brandt H, Donat K, and Erhardt G (2012). *J Dairy Sci*, 95(5), 2734-2739. https://doi.org/10.3168/jds.2011-4994

Li H and Durbin R (2009). *Bioinformatics*, *25*(14), 1754-1760. https://doi.org/10.1093/bioinformatics/btp324

Liao Y, Wang J, Jaehnig EJ, Shi Z, and Zhang B (2019) *Nucleic Acids Res*, 47(W1), W199-W205. https://doi.org/10.1093/nar/gkz401

McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K *et al.* (2010). *Genome Res*, *20*(9), 1297-1303. https://doi.org/10.1101/gr.107524.110

Moioli B, D'Andrea S, De Grossi L, Sezzi E, De Sanctis B, *et al.* (2016). *Anim Prod Sci*, 56(7), 1046-1055. https://doi.org/10.1071/AN14826

Purdie AC, Plain KM, Begg DJ, de Silva K, and Whittington RJ (2019). *Sci Rep*, *9*(1). https://doi.org/10.1038/s41598-019-44670-w

Reddacliff LA, Beh K, McGregor H, and Whittington RJ (2005). *Aust Vet J*, 83(7), 435-441. https://doi.org/10.1111/j.1751-0813.2005.tb13087.x

Sanchez MP, Guatteo R, Davergne A, Saout J, Grohs C, et al (2020). *Genet Sel Evol 52*(1). https://doi.org/10.1186/s12711-020-00535-9

Sechi S, Salaris S, Ligios C, Ponti N, *et al.* (2011) In Proc. 19th ASPA Congr., Cremona, Italy

Usai MG, Casu S, Ziccheddu B, Sechi T, Miari S, *et al.* (2019). In Proc. 23rd ASPA Congr, Sorrento, Italy

Verschoor CP, Pant SD, You Q, Kelton DF, and Karrow NA (2010). *Vet. Immunol. Immunopathol.,* 137(1-2), 1-11. https://doi.org/10.1016/j.vetimm.2010.03.026

257. Discovery of a genetic variant in serum amyloid A that might protect chickens from amyloid arthropathy

C. Falker-Gieske[1*], N.-F. Paul[1], M. Spourita[1], K. Gustmann[1], J. D. Gilthorpe[2] and J. Tetens[1]

[1]Department of Animal Sciences, Georg-August-University, Burckhardtweg 2, 37077 Göttingen, Germany; [2]Department of Integrative Medical Biology, Umeå University, 901 87 Umeå, Sweden; clemens.falker-gieske@uni-goettingen.de

Abstract

Amyloid arthropathy in chickens is a type of AA amyloidosis with serum amyloid A (SAA) depositions in the knee joints of affected birds. The disease is triggered by bacterial infection, mainly with *Enterococcus faecalis*, and counteracted with the administration of antibiotics. Amyloid arthropathy has a serious impact on animal welfare and the laying hen industry. Brown layers are commonly affected whereas white layers are resistant to the disease. By comparing whole genomes of commercial lines of brown and white egg-laying chickens, we discovered a missense mutation in the SAA gene of white layers, which is predicted to cause an arginine to serine exchange at position 90 in the full-length protein. We cloned and overexpressed this SAA.R90S isoform along with the wild-type SAA in chicken hepatocellular carcinoma (LMH) cells, which will serve as a model system to investigate the influence of SAA-R90S on the molecular pathogenesis of amyloid arthropathy.

Introduction

Amyloid arthropathy is a disease in chickens caused by bacterial infection, in the majority of cases with *Enterococcus faecalis* (Landman *et al.*, 1998). A significant impact on animal welfare, contribution to the overuse of antibiotics, and as a result significant economic losses for the laying hen industry have been reported (Blanco *et al.*, 2016). The main characteristic of amyloid arthropathy is depositions of aggregated serum amyloid A (SAA) in large joints of affected animals (Landman *et al.*, 1998). This is most likely initiated by the release of SAA from the liver in large quantities during an acute phase response to bacterial infection (McAdam and, 1976, Morrow *et al.*, 1981). With white egg-laying chickens being resistant to the disease amyloid arthropathy mostly affects brown layers (Peperkamp *et al.*, 1997). It has been previously reported that no sequence variation in the *SAA* gene exists between the two chicken breeds (Ovelgönne *et al.*, 2001). Noteworthy, in that study, the *SAA* coding sequence of only one animal of each breed was sequenced. With this experimental setup, it is not possible to draw conclusions on the population scale. Consequently, past studies did not pay much attention to sequence variation in the *SAA* gene and rather focused on other aspects of the immune system. High CD8+ T cell counts in synovial tissue and peripheral blood were observed in challenged brown layers. On the contrary, an increase in T cell proliferation and macrophage activity was observed in white layers (Zekarias *et al.*, 2000). The severity of the disease was correlated with elevated levels of IL-1β, IL-6, and TNF-α (Sevimli *et al.*, 2008). Here we describe the discovery of a missense variant in the SAA gene that might confer disease resistance to white egg-laying chickens and present a cellular model system to study the impact of this mutation on the molecular pathogenesis of amyloid arthropathy.

Materials & methods

Variant calling from whole-genome sequencing data. Mapping and variant calling with GATK v. 4 (McKenna *et al.*, 2010) from whole-genome sequences of 25 white and 25 brown layers (BioProject PRJEB30270) were performed as described in our previous study (Falker-Gieske *et al.*, 2020) with genome version GRCg6a. Plink v. 1.9 (Chang *et al.*, 2015) was used to calculate FST values and variant effect prediction was performed with snpEff (Cingolani *et al.*, 2012).

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_257, © C. Falker-Gieske *et al.* 2022

Vector construction. A chicken SAA cDNA was isolated from chicken hepatocellular carcinoma (LMH) cells. Amplification of the cDNA was performed with Phusion High Fidelity DNA Polymerase (Thermo Scientific) with the primers GgSAA-*Nhe*I-F1 CTCATCgctagcATGAGGCTCTGTATCTGCTTCG and GgSAA-*Xho*I-R1: CTCATCctcgagGTATTTGTTAGGAAGTCCTGCGG (restriction sites in lower case letters). The cDNA was cloned into pJET1.2/blunt with the CloneJET PCR Cloning Kit (Thermo Scientific). The Q5 Site-Directed Mutagenesis Kit (NEB) was used to introduce the SAA mutation G270C (rs739601959). The restriction enzymes *Nhe*I and *Xho*I were used to clone the SAA and SAA-G270C cDNAs into pcDNA3.1(-) (Thermo Scientific).

Cell culture. To create stable cell lines, LMH cells were transfected with pCDNA3.1(-)-GgSAA-WT, pCDNA3.1(-)-GgSAA-G270C, and pCDNA3.1(-) (empty vector). Lipofectamine 2000 (Thermo Scientific) was utilized to introduce plasmid DNA into cells. For that cells were seeded at a density of 6×10^5 cells on a 6-well plate. After 24h the cells were treated with 2.5 µg of DNA with 15 µl of lipofectamine 2000. To select cells that stably expressed the plasmid vector cells were supplemented with 400 µg/ml G418 (Thermo Scientific) 48h post-transfection. The G418 concentration was lowered to 200 µg/ml after two weeks.

Real-time quantitative PCR. For RNA isolation 1×10^6 LMH cells were seeded on T25 flasks. After 48 h cells were processed with the RNeasy Mini Kit and QIAshredder columns (QIAGEN) according to the manual. The QuantiNova Reverse Transcription Kit (QIAGEN) was used to produce cDNA from 1 µg of RNA per sample. qPCR was performed with 10 ng of cDNA with the following primers: SAA forward primer CCATGCTCGTGGCAATTACG, SAA reverse primer CCTGAAGCGGTTGGGATCG, GAPDH forward primer ACTGTCAAGGCTGAGAACGG, and GAPDH reverse primer CATTTGATGTTGCTGGGGTC.

Results

To investigate the genetic basis underlying the difference in the susceptibility to amyloid arthropathy we performed a comparative analysis of 25 white and 25 brown layer genomes (Ni *et al.*, 2017). We calculated Wright's F_{ST} estimate via Weir and Cockerham's method for chromosome 5 since it harbors the *SAA* gene. The average F_{ST} for chromosome 5 was 0.355. We found several fixed variants in the SAA gene with F_{ST} values above 0.8 (Table 1).

Table 1. Fixation indices (F_{ST}) of variants located in the gene encoding chicken serum amyloid A in brown and white egg-laying chickens (values below 0.1 not shown).

Chromosome	Variant ID	Position (bp)	FST	Variant type
5	rs739601959	12686863	0.935	missense variant
5	rs312867488	12687022	0.857	intron variant
5	rs734537979	12687387	0.895	intron variant
5	rs16466917	12687415	0.918	intron variant
5	rs313607895	12687436	0.836	intron variant
5	rs314059120	12687544	0.857	intron variant
5	rs314828146	12687829	0.853	intron variant
5	rs15664672	12687908	0.849	intron variant
5	rs15664674	12687939	0.809	intron variant
5	rs15664676	12687944	0.809	intron variant
5	rs15664678	12687946	0.809	intron variant
5	rs16466918	12687989	0.836	splice acceptor variant
5	rs16466919	12688216	0.794	intron variant

The variant with the highest F_{ST} value (0.935) was the missense variant rs739601959 (GgSAA-G270C) in exon 4 of the *SAA* gene. Variant effect prediction revealed that this variant causes an arginine to serine amino acid exchange at position 90 (SAA.R90S) in the full-length SAA Protein. Although it was classified with a tolerable SIFT score of 0.4 we sought to further investigate this variant. The allele was not present in brown egg-laying chickens whilst 19 white layers were homozygous and three were heterozygous carriers. To create cell lines that stably overexpress SAA we cloned the SAA cDNA from a cDNA library that was prepared from LMH cells. We successfully introduced the GgSAA-G270C mutation by site-directed mutagenesis and sub-cloned both cDNAs into pCDNA-3.1(-) (Figure 1).

A total of seven stable LMH cell lines were generated by lipofection to overexpress the following constructs: pCDNA3.1(-) (LMH-empty-vector), pCDNA3.1(-)-SAA (LMH-SAA.WT pools 1-3) and pCDNA3.1(-)-GgSAA-G270C (LMH-SAA.R90S pools 1-3). Stable SAA expression was validated with qPCR and revealed comparable expression levels between SAA.WT (average $-\Delta Cq=3.64\pm1.15$, n=3) and SAA.R90S cell lines (average $-\Delta Cq=2.51\pm1.18$, n=3), whereas expression in SAA.WT cells was slightly higher.

Discussion

Currently, no other treatment than antibiotics is available to counteract bacteria-induced amyloid arthropathy outbreaks in chicken flocks. Since there is a clear difference between brown and white egg-laying chickens in disease susceptibility the underlying cause for brown layers to contract the disease is most likely genetic. In a comparative whole-genome approach, we discovered a fixed allele in white egg-laying chickens that might confer or contribute to their resistance to amyloid arthropathy. The genetic variant

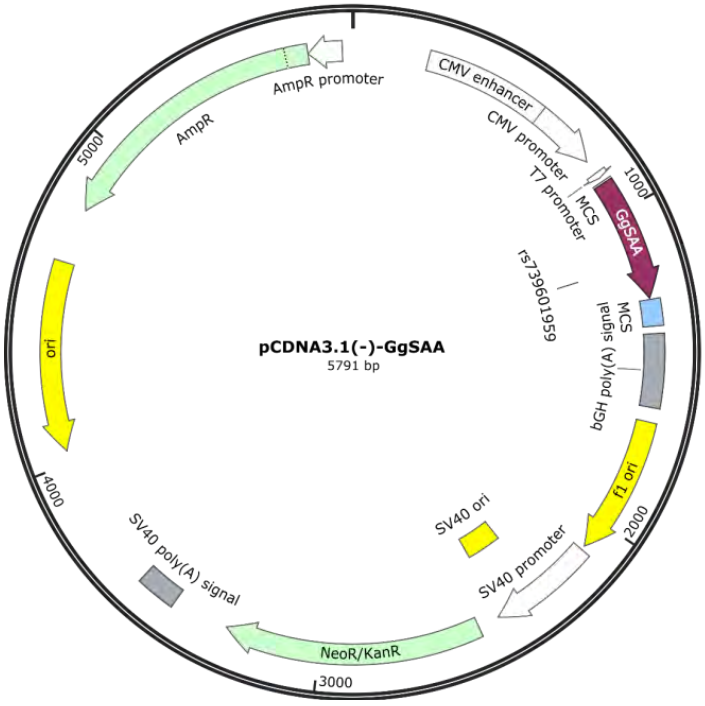


Figure 1. Vector map of the DNA plasmid used for stable overexpression of the wild type and R90S isoforms of serum amyloid A (SAA) in LMH cells. The SAA missense variant rs739601959 was introduced by site directed-mutagenesis.

rs739601959 is located in exon 4 of the SAA gene and leads to arginine to serine mutation at position 90 in the full-length protein. By creating stable chicken liver cell lines overexpressing the SAA.R90S and the wild type isoform we created a basis for the molecular characterization of this missense mutation regarding its potential function in amyloid arthropathy resistance. In humans, several variants have been correlated with lower SAA serum levels (Griffiths *et al.*, 2019). A similar effect might lead to lower SAA serum concentrations in white layers, which would explain the absence of SAA aggregates in their knee joints after challenging them with *E. faecalis* (Peperkamp *et al.*, 1997). Furthermore, variants of SAA1 in humans lead to different levels of SAA protein synthesis in the liver and other organs but also aberrant pathogen responses in the acute phase (reviewed in (De Buck *et al.*, 2016)). The functional characterization of the SAA.R90S isoform is currently ongoing work and involves immunofluorescent staining of cells, intra- and extracellular protein quantification of SAA, aggregation assays of recombinant SAA, and transcriptome analysis by RNA sequencing.

References

Blanco AE, Barz M, Icken W, Cavero D, Mazaheri A *et al.* (2016). Worlds Poult Sci J 72(03):495–508. https://doi.org/10.1017/S0043933916000453

Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM *et al.* (2015). Gigascience 4:7. https://doi.org/10.1186/s13742-015-0047-8

Cingolani P, Platts A, Le Wang L, Coon M, Nguyen T *et al.* (2012). Fly (Austin) 6(2):80–92. https://doi.org/10.4161/fly.19695

De Buck M, Gouwy M, Wang JM, van Snick J, Opdenakker G *et al.* (2016). Curr Med Chem 23(17):1725–1755. https://doi.org/10.2174/0929867323666160418114600

Falker-Gieske C, Iffland H, Preuβ S, Bessei W, Drögemüller C *et al.* (2020). BMC Genet 21(1):114. https://doi.org/10.1186/s12863-020-00920-9

Griffiths K, Maxwell AP, McCarter RV, Nicol P, Hogg RE *et al.* (2019). Ir J Med Sci 188(4):1175–1183. https://doi.org/10.1007/s11845-019-01996-8

Landman WJ, Gruys E, Gielkens AL (1998). Avian Pathol 27(5):437–449. https://doi.org/10.1080/03079459808419367

McAdam, Keith P. W. J., Sipe JD (1976). J Exp Med 144(4):1121–1127

McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K *et al.* (2010). Genome Res 20(9):1297–1303. https://doi.org/10.1101/gr.107524.110

Morrow JF, Stearman RS, Peltzman CG, Potter DA (1981). Proc Natl Acad Sci U S A 78(8):4718–4722. https://doi.org/10.1073/pnas.78.8.4718

Ni G, Cavero D, Fangmann A, Erbe M, Simianer H (2017). Genet Sel Evol 49(1):8. https://doi.org/10.1186/s12711-016-0277-y

Ovelgönne JH, Landman WJ, Gruys E, Gielkens AL, Peeters BP (2001). Amyloid 8(1):41–51

Peperkamp NH, Landman WJ, Tooten PC, Ultee A, Voorhout WF *et al.* (1997). Vet Pathol 34(4):271–278. https://doi.org/10.1177/030098589703400402

Sevimli A, Misirlioğlu D, Yağci A, Bülbül A, Yilmaztepe A *et al.* (2008). Vet Res Commun 32(7):499–508. https://doi.org/10.1007/s11259-007-9034-6

Zekarias B, Landman WJ, Tooten PC, Gruys E (2000). Veterinary Immunology and Immunopathology 77(1-2):55–69. https://doi.org/10.1016/S0165-2427(00)00233-6

258. Genetic background of the porcine immune system using SNP chip data and whole genome sequencing data

C.M. Dauben[1*], C. Große-Brinkhaus[1], H. Henne[2], A.K. Appel[2] and E. Tholen[1]

[1]Institute of Animal Science, University of Bonn, Endenicher Allee 15, 53115 Bonn, Germany; [2]BHZP GmbH, An der Wassermühle 8, 21368 Dahlenburg-Ellringen, Germany; cdau@itw.uni-bonn.de

Abstract

Over the last years, there is a stronger focus on the porcine immune system in the context of animal welfare and health. Therefore, this study aims to improve the understanding of the genetic architecture of several components of the immune system and immunocompetence of piglets. SNP chip data and whole genome sequencing (WGS) data are combined to deepen the knowledge in a Landrace and Large White population. Animals for re-sequencing experiments were selected using a combined pedigree- and haplotype-based approach to enable the accurate imputation of genotyped animals from both populations to WGS level. Both breeds share the majority of biallelic variants (13.6 million), identified by WGS data. Additionally, 4.5 (Landrace) and 5.0 (Large White) million breed-specific variants were investigated. All these variants will be further used to characterize both populations with a special focus on immune relevant genomic regions and performing Genome-wide association studies with imputed WGS data.

Introduction

The porcine immune system and the implementation of immunocompetence are still a major challenge in pig breeding and production. From the piglets' perspective, immunocompetence is characterised by health and immune traits (e.g. Flori *et al.*, 2011).

A number of studies have shown the genetic foundation of several immune-related traits by estimating moderate to high heritabilities (e.g. Clapperton *et al.*, 2009; Flori *et al.*, 2011; Ponsuksili *et al.*, 2016). Until now, information on 34,342 quantitative trait loci (QTL) is stored in the AnimalQTLdb (Release 46, December 2021, https://www.animalgenome.org/cgi-bin/QTLdb/SS/summary) of which 3,639 are assigned to the immune capacity, 2,900 to blood parameters and 619 to disease susceptibility. A lot of these studies have investigated single nucleotide polymorphism (SNP) chip array-based data. A recent study indicated genomic regions of interest for haematological parameters, cytokines and haptoglobin in a Landrace (LR) and Large White (LW) population under non-challenging conditions (Dauben *et al.*, 2021).

In recent years, technologies in applied genomics have been constantly improved and marker density has increased. Especially the sequencing of a subset of animals and imputation of the population with low marker density to a higher density is a cost-saving and promising strategy to reveal genetic effects in pig breeding populations (e.g. van den Berg *et al.*, 2019).

The aim of this study is to improve the understanding of the genetic architecture of the porcine immune system in LR and LW by investigating health- and immune-related traits in piglets. Therefore, SNP chip data is combined with whole genome sequencing (WGS) data to obtain detailed insights into genomic regions of interest and imputed WGS (iWGS) data is further used in Genome-wide association studies (GWAS).

Materials & methods

Phenotypes and SNP chip data. Phenotypic data was recorded for 24 traits with a link to health and immunity in a total of 535 LR piglets and 461 LW piglets. Four of the traits were excluded from further

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_258, © C.M. Dauben *et al.* 2022

analyses due to low expression, resulting in a variation close to zero. Details are given in Dauben *et al.* (2021). Trait complexes and examples for specific traits investigated are listed in Table 1. Piglets were reared on farms of the German breeding organisation BHZP GmbH under high hygienic standards.

Blood and tissue samples were analysed using the PorcineSNP60v2 BeadChip (Illumina, San Diego, CA, USA). Genetic distances led to breed-specific analyses. Breed-specific quality control on autosomes was performed considering the call rate for individuals and markers (95%), minor allele frequency (MAF) (1%) and linkage disequilibrium was estimated in windows of 3 kb.

Genome-wide association studies – SNP chip data. Methods for GWAS using SNP chip data are described in the study of Dauben *et al.* (2021) in detail. In brief, the R-package GenABEL (Aulchenko *et al.*, 2007) with a generalized linear model was used to identify associations between SNP genotypes and phenotypic information on all parameters except for the cytokines, which were analysed with ASReml® (Gilmour *et al.*, 2015). Population stratification was adjusted with the Genomic Control approach (Devlin and Roeder, 1999). Multiple testing was taken into account by correction for a modified false discovery rate (Storey, 2002).

Whole genome sequencing (WGS) data. A subset of 57 (LR: 28, LW: 29) animals was selected for next generation re-sequencing. Selected animals are expected to represent the LR and LW populations with the overall aim of enabling an accurate imputation of SNP chip data to higher density. The selection process followed a combined approach of pedigree and genotype data under the restriction of practical conditions as a follow-up study within the pigFit project (Dauben *et al.*, submitted).

Whole genome DNA sequencing (Illumina HiSeq 4000) was performed in a paired-end approach with a read length of 150 bp. Quality control of raw data was performed using FastQC (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/). Adapter clipping, trimming and quality thresholds were applied using Trimmomatic v.0.39 (Bolger *et al.*, 2014). Alignment to the reference genome *Sscrofa* 11.1 was performed using the Burrows-Wheeler Aligner (BWA, bwa version 0.7.17) (Li and Durbin, 2009), followed by identifying PCR duplications (Picard, https://broadinstitute.github.io/picard) and base quality score recalibration (GATK version v.3.8-1-0) (McKenna *et al.*, 2010). For variant calling, GATK v.3.8-1-0 (McKenna *et al.*, 2010) was used. Variants for further analyses were restricted to biallelic SNPs and variant filtering followed the filtering criteria QD<2.0, MQ<40.0, MQRankSum<-12.5, SOR>3.0 and ReadPosRankSum<-8.0.

Variant statistics and Annotation – WGS data. Handling of vcf files and the calculation of variant statistics, MAF and population parameters was performed with e.g. PLINK v.1.9 (Chang *et al.*, 2015, https://www.cog-genomics.org/plink/1.9/) and VCFtools (Danecek *et al.*, 2011). Annotation was performed using the Ensembl variant effect predictor (VEP version 105) (McLaren *et al.*, 2016).

Table 1. Phenotype measurements related to health and the immune system of piglets.

Complete blood count	Differential blood count [%]	Cytokines [ng/ml]	Acute phase proteins [mg/ml]
e.g.	e.g.	Interferon-γ	Haptoglobin
White blood cells [G/l]	Lymphocytes	Interleukins[1]	
Red blood cells [T/l]	Monocytes	Tumor necrosis factor-α	
Platelets [G/l]	Neutrophils		

[1] Interleukin (IL)-10, IL-12, IL-1β, IL-4, IL-6, IL-8

Imputation and Genome-wide association studies – WGS data. In ongoing steps, imputation of animals with 60k SNP chip data is performed to WGS level. Therefore, only variants on autosomes are used and imputation is performed using FImpute v3 (Sargolzaei *et al.*, 2014). Based on iWGS data, GWAS is going to be performed and results will be compared to that of GWAS using SNP chip data under consideration of the biological function.

Results

Genome-wide association studies – SNP chip data. GWAS on 17 immune relevant traits identified 477 breed-specific SNP markers in 535 LR and 461 LW piglets (Dauben *et al.*, 2021). Three traits did not show any significance. Results revealed highly informative genomic regions and candidate genes across the whole genome. In addition, putative pleiotropic effects were detected on *Sus scrofa* chromosome (SSC) 5, SSC12, SSC15, SSC16 and SSC17. Partly, regions are in accordance with QTL regions that were already published.

Preprocessing, variant calling and variant statistics – WGS data. In total, WGS data of 57 (LR: 28, LW: 29) pigs were available. Alignment was performed against the reference genome *Sscrofa* 11.1 and alignment rate was high (>99%) in both of the breeds. The overall coverage was 12.1 and varied across the chromosomes, but was consistently higher in LR. The highest coverage was observed on SSC16.

After filtering, 13.6 million variants were detected across the LR and LW populations. Additionally, 4.5 million (LR) and 5.0 million (LW) variants were exclusively detected in one of the breeds. In LR, 95.2% of all these variants were located on the porcine autosomes. In LW, this proportion was slightly lower (95.1%). Average distance between the detected variants on autosomes is 128 bp (LR) and 125 bp (LW) with clear chromosomal differences. The average Ts/Tv ratio across the whole genome was 2.33 and 2.32 for biallelic variants in LR and LW, respectively. Further investigations are ongoing.

Imputation and Genome-wide association studies – WGS data. First results indicate high variation of imputation accuracy across and within the chromosomes. In addition, higher-frequency variants (>5%) showed higher imputation accuracies, compared to low-frequency variants (<5%). In next steps, the detailed investigation of imputation results and GWAS on iWGS data will follow.

Discussion

Assuming haematological parameters and cytokines to have an effect as biomarkers of the immune reaction (Fossum, 1998; Ponsuksili *et al.*, 2016), our study has a focus on the underlying genetic background of these trait complexes.

The number of variants identified in this study is on a comparable level to previously reported results in commercial pig populations (e.g. Cai *et al.*, 2020). Investigation of variants is still in progress and detailed statistics will be worked out in further analyses steps. Especially variants within the splice region, stop gained or start and stop lost variants are of particular interest, potentially causing loss-of-function (Derks *et al.*, 2019). These variants will be further analysed in the context of breed-specificity and associations with immune-related parameters.

The LR and LW populations were already investigated regarding significant associations using SNP chip data (Dauben *et al.*, 2021). These analyses revealed highly informative and breed-specific regions, which are of interest with a putative link to immune-related parameters in piglets. In the next step, GWAS on iWGS data is performed to validate and further characterize regions of interest, associated with the porcine immune system. First results from imputation steps revealed regions with highly accurate imputed variants, depending on MAF. As expected from imputation steps using the SNP chip data to validate the composition

of the reference panel (Dauben *et al.*, submitted), the accuracy of some of the low-frequency variants was sufficiently high to be included in association analyses.

Acknowledgements

This study was performed within the 'pigFit' project which was supported by funds of the German Government's Special Purpose Fund held at Landwirtschaftliche Rentenbank (FKZ28-RZ-3-72.038).

References

Aulchenko Y.S., Ripke S., Isaacs A., and Van Duijn C.M. (2007) Bioinformatics 23(10):1294–1296. https://doi.org/10.1093/bioinformatics/btm108

Bolger A.M., Lohse M., and Usadel B. (2014) Bioinformatics 30(15):2114–2120. https://doi.org/10.1093/bioinformatics/btu170

Cai Z., Sarup P., Ostersen T., Nielsen B., Fredholm M., *et al.* (2020) J Anim Sci 98(7):skaa229. https://doi.org/10.1093/jas/skaa229

Chang C.C., Chow C.C., Tellier L.C.A.M., Vattikuti S., Purcell S.M., *et al.* (2015) Gigascience 4(1):s13742-015-0047-8. https://doi.org/10.1186/s13742-015-0047-8

Clapperton M., Diack A.B., Matika O., Glass E.J., Gladney C.D., *et al.* (2009) Genet Sel Evol 41:54. https://doi.org/10.1186/1297-9686-41-54

Danecek P., Auton A., Abecasis G., Albers C.A., Banks E., *et al.* (2011) Bioinformatics 27(15):2156–2158. https://doi.org/10.1093/bioinformatics/btr330

Dauben C.M., Pröll-Cornelissen M.J., Heuβ E.M., Appel A.K., Henne H., *et al.* (2021) BMC Genomics 22:717. https://doi.org/10.1186/s12864-021-07997-1

Derks M.F.L., Gjuvsland A.B., Bosse M., Lopes M.S., van Son M., *et al.* (2019) PLoS Genet 15(3):e1008055. https://doi.org/10.1371/journal.pgen.1008055

Devlin B., and Roeder K. (1999) Biometrics 55:997–1004. https://doi.org/10.1111/j.0006-341X.1999.00997.x

Flori L., Gao Y., Laloë D., Lemonnier G., Leplat J.-J., *et al.* (2011) PLoS One 6(7):e22717. https://doi.org/10.1371/journal.pone.0022717

Fossum C. (1998) Domest Anim Endocrinol 15(5):439–444. https://doi.org/10.1016/S0739-7240(98)80001-5

Gilmour A.R., Gogel B.J., Cullis B.R., Welham S.J., and Thompson R. (2015) ASReml User Guide Release 4.1 Structural Specification. Available at: https://asreml.kb.vsni.co.uk

Li H., and Durbin R. (2009) Bioinformatics 25(14):1754–1760. https://doi.org/10.1093/bioinformatics/btp324

McKenna A., Hanna M., Banks E., Sivachenko A., Cibulskis K., *et al.* (2010) Genome Res 20:1297–1303. https://doi.org/10.1101/gr.107524.110

McLaren W., Gil L., Hunt S.E., Riat H.S., Ritchie G.R.S., *et al.* (2016) Genome Biol 17:122. https://doi.org/10.1186/s13059-016-0974-4

Ponsuksili S., Reyer H., Trakooljul N., Murani E., and Wimmers K. (2016) PLoS One 11(7):e0159212. https://doi.org/10.1371/journal.pone.0159212

Sargolzaei M., Chesnais J.P., and Schenkel F.S. (2014) BMC Genomics 15:478. https://doi.org/10.1186/1471-2164-15-478

Storey J.D. (2002) J R Stat Soc Series B Stat Methodol 64(3):479–498. https://doi.org/10.1111/1467-9868.00346

van den Berg S., Vandenplas J., van Eeuwijk F.A., Bouwman A.C., Lopes M.S., *et al.* (2019) Genet Sel Evol 51:2. https://doi.org/10.1186/s12711-019-0445-y

259. Deciphering the origins of Neolithic sheep from northern Iberian Peninsula

M. Serrano[1*], A. Fernández[1], C. González[1], J. Martínez-Blanch[2], M. Moreno-García[3], M. Rojo-Guerra[4] and J.H. Calvo[5]

[1]INIA-CSIC, Ctra. De la Coruña km 7,5, 28040 Madrid, Spain; [2]ADM Biopolis, Parque Científico Universitat de Agustín Escardino, 9, Carrer del Catedràtic Agustín Escardino Benlloch, 46980 Paterna, Valencia, Spain; [3]IH-CSIC, C. de Albasanz, 26, 28037 Madrid, Spain; [4]Universidad de Valladolid, Plaza del Campus s/n, 47011 Valladolid, Spain; [5]CITA-ARAID, Av. Montañana, 930, 50059 Zaragoza, Spain; malena@inia.es

Abstract

Domestication events in livestock can be studied by comparing modern DNA with ancient DNA (aDNA) from archaeological faunal remains. Investigating the phylogeny history of sheep remains from the Neolithic Spanish site of Cova de Els Trocs, provides an opportunity to gain some insights about their origins. In this work, aDNA recovered from 15 ovine specimens dated between 6000 to 4000 years BP were sequenced by whole genome sequencing (WGS), and mapped against the *O. aries* modern mitogenome. Complete mitogenomes were restored for 12 ancient samples. Bayesian Inference (BI) and Maximum Likelihood (ML) phylogenetic trees fitting the HKY+G+I model were performed including the mitogenomes of 12 ancient and 74 modern sheep and 4 wild species. All Neolithic samples belong to the mitochondrial haplotype B. Four ancient sheep constituted a clade (89%) with the Viena sheep breed (Karelia), suggesting a complex demographic history of their origins.

Introduction

Archaeological evidence points to the Neolithic as the moment in time of sheep introduction in Europe along the Mediterranean and Danube routes (Zeder 2017). Although mtDNA studies of worldwide sheep breeds have provided some clues about the history of sheep domestication and dispersion, questions such as population colonization, demographic history and patterns of gene flow to Europe, Asia and Africa (Lv *et al.* 2015) are yet unsolved. Current advances in paleo-genetics and aDNA recovering techniques have allowed deepening into the genetic origins and dispersion of sheep breeds by comparing DNA sequences of ancient specimens with those of modern breeds. In this work, aDNA from 15 *Ovis* recovered in the site Cova dels Trocs (6000 to 4000 years BP) were sequenced and mapped against the *O. aries* modern mitogenome. A phylogenetic analysis including complete mitogenomes from 12 Neolithic and 74 modern sheep, and 4 wild species were performed to elucidate the origin of ancient specimens.

Materials & methods

Ovine fossil material, samples processing and DNA extraction and purification. Cova de Els Trocs, situated at above 1,500 m.a.s.l., on the Iberian central Pyrenees, is one of the few sites with a fully documented Neolithic stratigraphic sequence spanning from the 6th to the last third of the 4th millennium cal BC (calibrated years before Christ). In this site, a high number sheep remains have been recovered, revealing the great involvement in herding activities by the human groups from this territory (Tejedor-Rodríguez *et al.* 2021). Fifteen sheep jaws and maxillary from Cova dels Trocs (Huesca Spain), which came from three different occupation phases: 5 from T1 (5800-6100 years BP), 5 from T2 (5500-5800 years BP) and 5 from T3 (3900-4600 years BP years), was the available material. To minimize contamination with modern DNA, two DNA extraction replicates were performed by using new equipment and reagents in two independent laboratories. One of the replicates was used for WGS, while the other one for Sanger sequencing. All steps such as bone/tooth cutting and surface sanding, grinding and solutions preparation, and DNA extraction in sterile conditions, were carried out in separate places and times. Teeth roots were

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_259, © M. Serrano *et al.* 2022

crushed for 90 seconds at maximum frequency within a ball mill with liquid nitrogen in a mill MM400 (Retsch, Germany). DNA extraction were performed following a modified protocol of Huel (Huel *et al.* 2012). DNA purification was carried out with QIAquick PCR purification kit (Qiagen Inc. USA).

DNA libraries and WGS. Libraries for next-generation sequencing (NGS) were built with the aDNA extracts using the Accel-NGS Methyl-Seq DNA Library Kit and Uniq Dual Indexing kit (Swift Biosciences) following manufacturer's protocol. This kit is specially indicated for sequencing aDNA when retention of fragments containing uracil nucleotides, resulting from damage, is not desired. PrePCR and PostPCR labours were conducted in different spaces to avoid contamination. In order to validate methodology, 4 ancient samples were repeated by using Meyer libraries (Meyer 2010). Sequencing was performed on the Illumina NextSeq 550 platform to obtain between 11 and 62 million 75nt paired-end reads per library.

SANGER sequencing. Twelve overlapping amplicons were amplified, assembling a total length of 4000 bp of the mtDNA after Sanger sequencing (positions 1 to 1,079, and 13,696 to 16,616 of the NC_001941.1 sheep mitochondrial genome). Amplicon sizes varied between 395 and 679 bp. As for WGS, PrePCR and PostPCR were performed in different labs.

Ancient DNA authentication; reads processing, mapping and variant calling. Sample reads were trimmed of adapter sequences and low quality bases by AdapterRemoval v.2.3.1 (Schubert *et al.* 2016) and mapped to the sheep mitochondrial genome (NC_001941.1) using BWA v. 0.7.17 (Li and Durbin 2009). Duplicates were removed with Picard Tools, and the genome coverage was estimated by Samtools v.1.10 (Li 2011). MapDamage 2.0 (Jónsson *et al.* 2013) was used to evaluate fragment lengths and nucleotide misincorporation patterns. Two different algorithms were used to produce the VCF files for mitochondrial positions: bcftools mpileup and GATK UnifiedGenotyper (DePristo *et al.* 2011). All the mutations reported were visually inspected in the Biomatters IGV software v2.3.66 (Robinson *et al.* 2011) to rule out artefacts due to misincorporations in low coverage regions. Finally, Samtools v1.10 was used to create the consensus sequence for all the samples.

Phylogenetic Analysis. Complete mtDNA sequences from Neolithic, modern and wild sheep were aligned using the Clustal Ω program. Evolutionary models were tested with MEGA-X. BI phylogenetic tree was constructed with MrBayes v3.1.2 using the Hasegawa-Kishino-Yano HKY+G+I model, running 4 simultaneous Markov chains for 20 million generations starting from a random tree. Sampling frequency was set to 1/500 discarding as burning the first 25% trees. Bootstrap values for the ML tree were generated with 1000 replicates using MEGA-X.

Results

Twelve of the 15 samples available yielded enough aDNA for sequencing. Table 1 shows average results of aDNA sequencing. MapDamage shows classical patterns of post-mortem damage in Neolithic samples, such as 5' C*T deamination and T missincorporations. However, 3' G*A deamination could not be detected due to the Adaptase technology used in the Accel-NGS Library Kit, which adds a low complexity polynucleotide tail to the 3' end of each fragment. The final mitochondrial fragment length of the Neolithic samples after correction and trimming was 16,616 bp (positions 1 to 16,616 of the sheep mitogenome).

Alignments of nucleotide sequences obtained by Sanger and WGS showed 100% identity, indicating that no contamination occurred in the aDNA extraction and amplification processes. Phylogenetic analysis (Figure 1) assigned the 12 ovine ancient samples to the mitochondrial haplotype B. No clustering was found between the aDNA samples based on their chronology. The distance matrix of Hamming dissimilarity showed that Neolithic sheep from Trocs were very close to Finnsheep, Lacaune, Rasa Aragonesa and Finn-

Table 1. Results of ancient DNA sequencing and mapping.

Parameter	No. paired-end sequences	Average Length	Average Quality	MT properly mapped	MT mean depth	% aMT recovered
Mean	16,569,276	149.9	36.1	2,660	33.6	0.017
SD	3,766,339	0.4	0.1	1,847	24.1	0.012
Min	11,360,788	148.8	35.9	550	6.3	0.002
Max	23,089,914	150.3	36.2	6,966	89.2	0.043

MT = mitochondrial; aMT = ancien mitochondrial

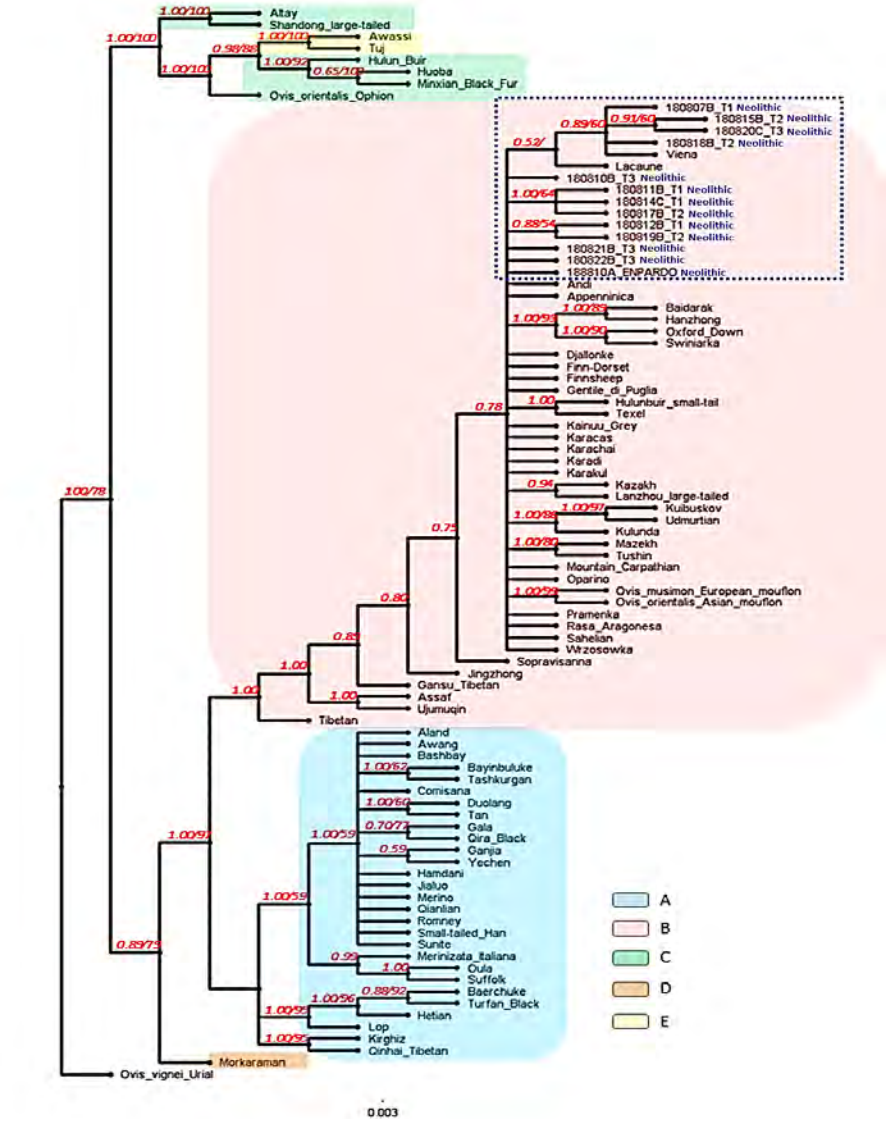


Figure 1. Phylogeny of Neolithic, modern and wild sheep inferred from 90 complete mitogenomes using BI and ML methods. On nodes: posterior probability/bootstrap values

Dorset breeds. Phylogenetic tree shows that 4 samples from Trocs formed a clade with the Viena sheep breed (Karelia, Russia).

Discussion

Accel-NGS Library Kit is suitable to work with aDNA, however its particular chemistry does not allow detecting in an optimal way the post-mortem damage patterns. However, by using Meyer libraries, both deamination patterns, 5' C$^{\circledast}$T and 3' G$^{\circledast}$A, could be detected. Ancient fragment lengths were longer than the average ~55 bp expected (Skoglund *et al.* 2012), probably due to the good preservation conditions of the Trocs cave. The small distance of all Neolithic samples from Trocs to Finnish breeds suggest a North European origin. Four Neolithic samples constituted a clade with an 89% of posterior probability with the Viena sheep, and ancient breed from Karelia, Russia (Meadows *et al.* 2005). In 2006, Tapio *et al.* found the haplotype A in the old native breeds of Nordic Countries and Russian Karelia, suggesting a previously unrecognized migration of sheep to northern Europe through Russia. This fact could indicate that some of the specimens found in Trocs, could have an origin in sheep populations that migrated to northern Europe, despite they exhibit the B haplotype.

References

DePristo M.A., Banks E., Poplin R., Garimella K.V., Maguire J.R. *et al.* (2011) Nat Genet. 43(5): 491-498. https://doi.org/10.1038/ng.806

Huel R., Amory S., Bilić A., Vidović S., Jasaragić E. *et al.* (2012) Methods Mol. Biol. 830: 185-198. https://doi.org/10.1007/978-1-61779-461-2_13

Jónsson H., Ginolhac A., Schubert M., Johnson P.L., Orlando L. (2013) Bioinformatics 29(13):1682-1684. https://doi-org/10.1093/bioinformatics/btt193

Li H. (2011) Bioinformatics 27(21) 2987-93. https://doi.org/10.1093/bioinformatics/btr509

Li H. and Durbin R. (2009) Bioinformatics, 25, 1754-1760. https://doi.org/10.1093/bioinformatics/btp324

Lv F.H., Peng W.F., Yang J., Zhao Y.X., Li W.R. *et al.* (2015) Mol. Biol. Evol. 32(10): 2515-2533. https://doi.org/10.1093/molbev/msv139

Meadows J.R.S., Li K., Kantanen J., Tapio M., Sipos W. *et al.* (2005). Journal of Heredity 96(5): 494-501. https://doi.org/10.1093/jhered/esi100

Meyer M, Kircher M. (2010) Cold Spring Harb Protoc. https://doi.org/10.1101/pdb.prot5448

Robinson, J.T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E.S. *et al.* (2011). Nature biotechnology, 29(1), 24–26. https://doi.org/10.1038/nbt.1754

Schubert M., Lindgreen S., Orlando L. (2016). BMC Research Notes, 9: 88. https://doi.org/10.1186/s13104-016-1900-2

Skoglund P., Malmström H., Raghavan M., Storå J., Hall P. *et al.* (2012) Science. 336 (6080): 466-9. https://doi.org/10.1126/science.1216304

Tapio M, Marzanov N, Ozerov M, Cinkulov M, Gonzarenko G *et al.* (2006) Mol. Biol. Evol. 23: 1776–1783. https://doi.org/10.1093/molbev/msl043

Tejedor-Rodríguez C., Moreno-García M., Tornero C., Hoffmann A., García-Martínez de Lagrán Í. *et al.* (2021) PlosOne 16, e0244139. https://doi.org/10.1371/journal.pone.0244139

Zeder M.A. (2017) Out of the Fertile Crescent: The dispersal of domestic livestock through Europe and Africa. In Boivin N., Crassard R. and Petraglia M. (Eds) Dispersal and Species Movement. Cambridge University Press. https://doi.org/10.1017/9781316686942.012

260. Prioritisation of variants from whole-genome sequence data for validation of their association with traits of interest

E. Molinero*, R.N. Pena, J. Estany and R. Ros-Freixedes

Departament de Ciència Animal, University of Lleida, Agrotecnio-CERCA Center, 25198 Lleida, Catalonia, Spain; eduard.molinero@udl.cat

Abstract

Whole-genome sequencing is a powerful tool for identifying genetic variants associated to complex traits. However, the large number of variants in linkage disequilibrium hinders the identification of relevant variants. The aim of this study is to assess if the prioritisation of variants from sequencing data of a reduced number of animals using sequential multi-layer criteria (association analysis, fixation index, predicted variant effect, and gene ontology) is a suitable strategy for identifying variants associated with traits of interest that can later be validated in a larger number of individuals. Using data from 205 pigs, we have designed a custom genotyping array with 240 variants from 193 candidate genes that will be validated in a larger number of pigs (~1000) to quantify the validation rate of this discovery-preselection-validation strategy and define which criteria are the most informative.

Introduction

Whole-genome sequencing is a powerful tool for dissecting the genetics that underlies complex traits because whole-genome sequence data include rare and population-specific variants, including causal variants. The identification of variants associated with complex traits not only contributes to a deeper knowledge of biological and physiological processes, but also has implications for the genetic evaluation of selection candidates. Such variants could increase the accuracy of gene-assisted or genomic predictions, which would in turn increase the rate of genetic gain per unit of investment in the breeding programs (vanRaden *et al.*, 2017), as well as increase the persistence of prediction accuracy across generations and across breeds (Iheshiulor *et al.*, 2016). Nevertheless, due to the large number of variants in linkage disequilibrium, the challenge resides on identifying relevant variants, ideally the causal ones, or representative tag variants. Across livestock species, many variants have been described as associated with a wide range of traits of interest, but in practice few of them have been properly validated and there is still much genetic variance left unexplained. The integration of genomic data with other omics data has been explored as a strategy to enhance the identification of causal variants (Flitman *et al.*, 2021; Mountjoy *et al.*, 2021).

In pigs, we have previously identified and validated two variants in the *LEPR* and *SCD* genes that affect fat content and composition (Ros-Freixedes *et al.*, 2016). With the purpose of identifying additional variants associated with those traits, we designed a strategy based on the sequencing of a small number (205) of pigs for variant discovery, the preselection of a large number (240) of variants according to sequential multi-layer criteria, and then the validation of the preselected variants in a much larger (~1000) number of samples from a referenced biobank. The aim of this study is to assess, using the fat content and composition traits as a model, if this discovery-preselection-validation strategy is suitable for identifying additional variants associated with complex traits and to test which criteria are the most effective.

Materials & methods

Animals and phenotypes. We sequenced 205 Duroc pigs of the same line, born from 123 sires and 188 dams. Pigs were grown in 14 batches from 2002 to 2019 complying with a common data recording and tissue sampling protocol (Ros-Freixedes *et al.*, 2016). The animals were raised from 75 days of age until slaughter at 210 days under similar commercial conditions. All pigs were slaughtered in the same abattoir, where

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_260, © E. Molinero *et al.* 2022

carcass weight, backfat thickness and loin thickness were recorded using and on-line ultrasound automatic scanner (AutoFOM, SFK-Technology, Denmark). We collected samples of *gluteus medius* muscle. The intramuscular fat content and fatty acid composition were determined by quantitative gas chromatography. Fatty acids were expressed as percentages relative to total fatty acid content. The proportion of individual and total saturated, monounsaturated, and polyunsaturated fatty acids were calculated.

Sequencing and variant discovery. We isolated genomic DNA form the *gluteus medius* muscle for sequencing on NovaSeq 6000 (Illumina, San Diego, CA, USA) in paired-end mode. The average realized sequencing coverage was 7.9x. We mapped the reads to *Sscrofa11.1* reference genome (GenBank accession: GCA_000003025.6) and we called SNPs and indels. We focused the analyses on the variants with call rates above 0.9 and with minor allele frequency equal or greater than 0.20, so that all genotypes were sufficiently represented in the 205 pigs, by using PLINK1.9 (Chang *et al.*, 2015). Although variants in non-coding regions can have regulatory functions (Koufariotis *et al.*, 2018), we focused on variants located in the transcribed region of the genes, defined here to include introns and the 500-bp upstream untranslated region, because it is difficult to predict the effect of variants in non-coding regions based on the DNA sequence alone.

Prioritisation of variants. Using the data of the 205 sequenced pigs, we assessed the association of the variants with carcass backfat thickness and intramuscular fat content and fatty acid composition in *gluteus medius* muscle using GEMMA 0.96 software (Zhou and Stephens, 2012) to fit a univariate linear mixed model accounting for a genomic relationship matrix. Phenotypes were adjusted for fattening batch (14 levels), the *SCD* genotype (3 genotypes), the *LEPR* genotype (3 genotypes), and the slaughter age. The genotypes for *LEPR* and *SCD* were included in the model because of their major impact on the studied traits (Ros-Freixedes *et al.*, 2016). Also, we calculated the fixation index (F_{ST}) for each variant and trait using two subsets of 40 pigs that included, respectively, the individuals with the highest and the lowest phenotypic values of the trait. We selected variants with F_{ST} greater than 0.15. Candidate variants were examined for their predicted variant effect using the Ensembl VEP tool (McLaren *et al.*, 2016). Finally, we identified all genes from the Ensembl database that had been annotated with at least one gene ontology term that is putatively related to the trait or to some underlying biological process. The set of tested variants was sequentially refined based on the association analyses results, F_{ST}, predicted variant effect, and gene ontology.

Results

We called a total of 19,046,451 variants across the 205 pigs, of which 7,379,521 variants passed the filters on minor allele frequency and call rate, and, of these, 3,159,752 were located in the transcribed region of 12,835 annotated genes. This was the set of variants that was refined sequentially based on the association analyses, F_{ST}, predicted variant effect, and gene ontology. Based on gene ontology terms, 875 of these genes (which contained 179,974 of the latter variants) were related to processes potentially involved in body fatness, fatty acid composition or feed intake regulation, among others (here referred to as 'fat-related genes').

As a result of the association analysis (significance threshold of 10^{-6}), we detected 108 associated variants in 16 genes. However, most of these variants were predicted to be intronic or in genes without any gene ontology terms related to our target traits (Table 1). Nonetheless, two of those genes (*EFR3A* and *IL18BP*) were located in quantitative trait loci associated with fat deposition and fatty acid composition that had been previously described. Due to the limited sample size used and the polygenic nature of the traits, we used a less stringent significance threshold as a first step in the prioritisation of the variants. With a significance threshold of 10^{-4} we detected 5,220 variants in 516 genes, of which a small subset were missense variants or variants located in fat-related genes.

Table 1. Number of variants and genes prioritised using the four criteria[1].

Predicted variant effect[2]	Missense variants			Exonic variants			Variants in transcribed region		
P-value[3]	10^{-6}	10^{-5}	10^{-4}	10^{-6}	10^{-5}	10^{-4}	10^{-6}	10^{-5}	10^{-4}
No. of variants in annotated genes	0 (0)	1 (1)	15 (5)	0 (0)	14 (3)	143 (28)	108 (33)	772 (239)	5,220 (1,248)
No. of variants in fat-related genes	0 (0)	0 (0)	1 (0)	0 (0)	0 (0)	15 (3)	0 (0)	9 (7)	129 (39)
No. of annotated genes	0 (0)	1 (1)	12 (5)	0 (0)	10 (2)	82 (20)	16 (14)	91 (52)	516 (191)
No. of fat-related genes	0 (0)	0 (0)	1 (0)	0 (0)	0 (0)	8 (3)	0 (0)	4 (3)	28 (12)

[1] Results from the association analysis are shown in columns for three significance thresholds (*P*-value). Within parentheses, number of significant variants (and genes that contain them) with F_{ST}>0.15. Results by predicted variant effect and gene ontology-based class are shown by columns and rows, respectively.

[2] Exonic variants include variants located in exons and 5'-UTR, 3'-UTR of the gene, while the 'transcribed region' also includes introns and 500 bp upstream the untranslated region.

[3] Significance thresholds used for the association analyses. Number of variants (and genes that contain them) with an association below the specified significance threshold for at least one of the tested traits.

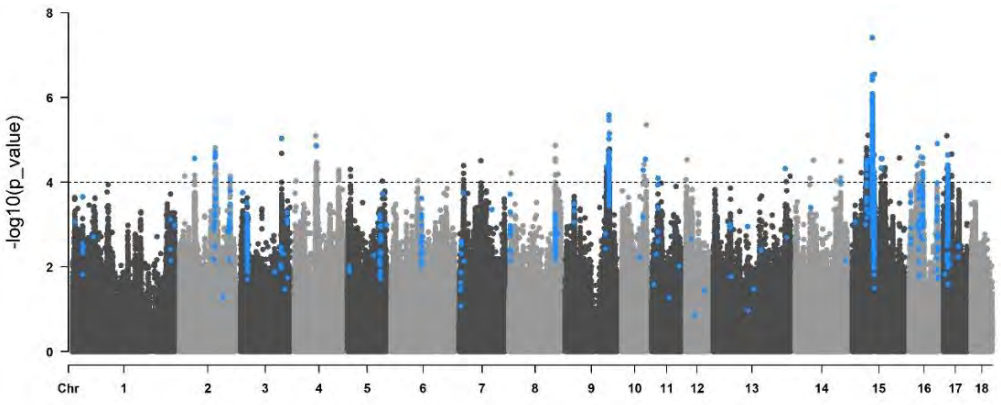


Figure 1. Manhattan plot for the trait with the strongest associations. In grey, the results of the association analysis. In blue, variants with F_{ST}>0.15.

As expected, the F_{ST} mean was close to zero because we calculated it between two subsets from the same population that diverged for a single trait. However, 9,598 variants in 706 genes had F_{ST} values greater than 0.15, of which, 219 variants in 41 genes had gene ontology terms related to our traits of interest. When comparing the results from F_{ST} and the association analysis, we observed that the overlap between both methods was far from complete but that the most relevant signals were captured by both (Figure 1).

Discussion

By sequentially refining a large set of variants from whole-genome sequence data using multiple layers of information (association analysis, F_{ST}, predicted variant effect, and gene ontology), we were able to preselect multiple candidate variants. It has been previously demonstrated that population statistics are efficient criteria for preselecting trait-relevant variants since the F_{ST}-based method, for example, is capable to differentiate weak associations from the spurious ones (Ling *et al.*, 2021). In our case, by using both association analyses and F_{ST} were we able to narrow down not only the candidate variants but also the candidate genes that contained them, particularly when using less stringent significant thresholds for the association analyses. A limitation of our strategy is that due to the limited number of sequenced animals,

we could detect only variants with high minor allele frequency and missed variants with lower minor allele frequency that could have larger effects. We did not consider variants located in non-coding regions, although they can have regulatory functions (Koufariotis *et al.*, 2018). Also, we relied on currently available functional data (like gene ontology), which can be incomplete or inaccurate.

Both *SCD* and *LEPR* appeared as the strongest candidate genes when we did not correct for their effects in the model. This can be interpreted as another evidence of the reliability of this strategy for variants with strong effects. However, the situation is more complex when targeting additional variants with weaker effects. A gene that ranked highly and was preselected using this strategy was *AGPAT5*. We have already validated a missense mutation from this gene in over 1000 animals with an effect on intramuscular fat content and fatty acid composition (Molinero *et al.*, 2021). Using the criteria described here we have designed a custom genotyping array with 240 variants from 193 candidate genes that will be used on ~1000 pigs from our biobank to quantify the validation rate of this whole discovery-preselection-validation strategy and to define which criteria are the most informative.

References

Chang C.C., Chow C.C., Tellier L.C., Vattikuti S., Purcell S.M., *et al.* (2015) Gigascience 4:7. https://doi.org/10.1186/s13742-015-0047-8

Flitman R.S., Khalili B., Kutalik Z., Rueedi R., Brümmer A., *et al.* (2021) J Proteome Res 20(11):5103-5114. https://doi.org/10.1021/acs.jproteome.1c00585

Iheshiulor O.O.M., Woolliams J.A., Yu X., Wellmann R., and Meuwissen T.H.E. (2016) Genet Sel Evol 48:15. https://doi.org/10.1186/s12711-016-0193-1

Koufariotis L.T., Chen YP.P., Stotjard P., and Hayes B.J. (2018) BMC Genom 19:237. https://doi.org/10.1186/s12864-018-4617-x

Ling A.S., Hay E.H., Aggrey S.E., and Rekaya R. (2021) BMC Genom 22:26. https://doi.org/10.1186/s12863-021-00979-y

McLaren W., Gil L., Hunt S.E., Riat H.S., Ritchie G.R.S., *et al.* (2016) Genome Biol. 17:122. https://doi.org/10.1186/s13059-016-0974-4.

Molinero E., Pena R.N., Estany J., and Ros-Freixedes R. (2021) Proc. of the 38th ISAG, virtual.

Mountjoy E., Schmidt E.M., Carmona M., Schwartzentruber J., Peat G., *et al.* (2021) Nat Genet 53(11):1527-1533. https://doi.org/10.1038/s41588-021-00945-5

Ros-Freixedes R., Gol S., Pena R.N., Tor M., Ibáñez-Escriche N., *et al.* (2016) PLoS One 11(3):e0152496. https://doi.org/10.1371/journal.pone.0152496

VanRaden P.M., Tooker M.E., O'Connell, J.R., Cole J.B., and Bickhart D.M. (2017) Genet Sel Evol 49:32. https://doi.org/10.1186/s12711-017-0307-4

Zhou X., and Stephens M. (2012) Nat Genet 44(7):821-824. https://doi.org/10.1038/ng.2310

261. Genome-wide association analyses and genomic prediction for pork meat quality traits using whole-genome sequence

M. Heidaritabar[1*], A. Huisman[2], M.C.A.M. Bink[2], P. Charagu[2] and G. Plastow[1]

[1]Department of Agricultural, Food and Nutritional Science, University of Alberta, Livestock Gentec, T6G 2H1, Edmonton, AB, Canada; [2]Hendrix Genetics, P.O. Box 114, 5830 AC Boxmeer, the Netherlands; heidarit@ualberta.ca

Abstract

Genome-wide association studies and genomic prediction for several pork carcass and meat quality traits were performed for 1,377 purebreds (PBs) and 954 crossbreds (CBs) using both whole-genome sequence (WGS) and medium-density panels. More associated regions were detected for traits in PBs than in CBs. WGS identified a small number of novel genomic regions in the PBs including three genomic regions for belly bend on chromosomes (SSC) 7, 8 and 13, two for muscle PH (SSC2 and 6), a different region on SSC2 for fresh loin Minolta b*(FMCOL), and one SSC10 for FMCOL Minolta L*. For both within- and between-population predictions, WGS did not improve prediction accuracy compared with a medium-density panel, except for slight improvement for some traits.

Introduction

Pork quality is an economically important trait in the swine industry, and it is one of the main breeding objectives in today's pig breeding programs. Moderate to high heritability of carcass and meat quality traits shows that in addition to the environmental factors, genetic factors control pork meat quality traits (Miar *et al.* 2014). Understanding the complex genetic mechanisms underlying carcass and meat quality traits will support the genetic improvement of these traits. Validated genomic regions can be used in genomic selection schemes to increase selection accuracy and potentially reduce generation interval, particularly in multi-population or across-population genomic evaluations which are applicable in swine breeding programs with crossbreeding schemes to accelerate genetic gain.

Whole-genome sequence (WGS) is expected to improve identification of associated regions (either extending candidate regions or identifying novel regions), because all of the causative single nucleotide polymorphisms (SNPs) are expected to be included. The variance associated with low frequency causative SNPs is expected to be captured using WGS. Our main objectives were therefore to investigate whether the use of WGS will detect novel associated regions in pigs compared with lower density panels. We also investigated how much genomic prediction accuracy is gained with WGS compared to medium-density panels for meat and carcass traits.

Materials & methods

Data. The total number of pigs with both imputed WGS genotypes and phenotypes was 1,377 for purebreds (PBs) and 954 for crossbreds (CBs, three-way cross Duroc-Landrace-Yorkshire). Imputation (using Beagle V5.0) from a medium-density marker panel (50K or 60K) to WGS was based on the sequence from 215 pigs from multiple breeds (Duroc, Large White, Maxter Pietrain). SNPs with the following features were excluded from both imputed WGS and medium-density panels: minor allele frequency <0.01, genotype call rate <0.95, and duplicated or unknown map positions on reference genome (Sus scrofa 11.1). Moreover, imputed genotypes were filtered based on the imputation reliability (cut-off threshold <0.8). After filtering out SNPs with low imputation accuracy, the average imputation accuracy (allelic DR^2 produced by Beagle) was 0.918 for ~10 million SNPs across the autosomes of CBs, and 0.955 for ~9.5 million SNPs of PBs.

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_261, © M. Heidaritabar *et al.* 2022

Meat and carcass quality traits included 15 meat colour traits (Minolta L*, a*, and b* for fat (FCOL)[1], and fresh ham gluteus medius (GMCOL), quadriceps femoris muscle (QFCOL), ham iliopsoas muscle (ICOL), and loin longissimus dorsi muscle (FMCOL)), peak shear force (PSF), fresh loin muscle PH (PH), drip loss (DL), marbling score (MS), and several belly traits. Our analyses were conducted for CBs, PBs, and combined CBs and PBs.

Association analysis. The model used for GWAS was a single marker mixed linear association model implemented in GCTA (Yang *et al.* 2014):

$$y = 1\mu + Zu + g + e \quad (1)$$

where y was the vector of phenotypic records (corrected for significant fixed effects such as sex, slaughter date, birthdate). u was the additive effect of the SNP to be tested for association, Z was a vector containing the SNP genotype coded as 0, 1, and 2. g was a vector of random polygenetic effects, and e was a vector of random residual effects. It was assumed that $g \sim N(0, G\sigma_g^2)$ and $e \sim N(0, I\sigma_e^2)$, where σ_g^2 and σ_e^2 were the genetic and residual variances, respectively. G was the genomic relationship matrix based on genotypes. For all traits, we defined the quantitative trait loci (QTL) regions according to the definition described by van den Berg *et al.* (2019): Firstly, SNPs on each chromosome were ranked based on their -log10 *P*-values. Secondly, starting with the SNPs with the largest -log10 *P*-value, all significant SNPs that exceeded the FDR of 10% and surrounding SNPs within a 0.5 Mb region to the left and right of the SNP were assigned to that particular QTL region.

Genomic prediction. For each population, the youngest pigs formed the validation population, which included 177 for PBs and 111 for CBs. The breeding values of the validation animals were predicted using the results from the oldest animals used as the training population. Accuracy was assessed as the correlation between corrected phenotypes and estimated breeding values divided by the square root of heritability for each trait. Predictions were based on genomic BLUP. Different prediction scenarios were assessed in order to find the optimal genomic selection scenarios to maximize accuracy. Prediction scenarios included: within CBs and within PBs; between-population predictions: (1) prediction of CBs using PBs or combined CBs and PBs; (2) prediction of PBs using CBs or combined CBs and PBs.

Results & discussion

Genome-wide associations. Of the 15 colour traits, we did not detect significant regions for most traits, regardless of panel density. Generally, more associated regions were detected for traits in PBs than in the CBs, and WGS identified only a small number of novel genomic regions in the PBs (Table 1). Moreover, across all datasets (for pork colour traits), we identified more QTL regions with WGS (n=58) compared with medium-density SNP panels (n=22). However, the added value of WGS compared with the medium-density panels was more for detecting extended QTL regions compared with SNP panel arrays, and not for detecting novel QTL regions. For non-colour traits, associated regions were found for PSF (CBs) and PH (both CBs and PBs), and marbling score (PBs). For PSF, PH, and MS, the most pronounced associated region, detected by both panel densities, were on SCC2, SCC15, and SCC1, respectively. When the combined dataset was used for GWAS for pork colour traits, many more QTL regions (11 for medium-density panels and 37 for WGS) were identified for FMCOL Minolta a*, FMCOL Minolta b*, GMCOL Minolta b*, and ICOL Minolta b*, compared with single population GWAS. This suggests that the added value of WGS was more in detecting novel regions compared with medium-density SNP panels when larger

[1] Each of the 5 colors (FCOL, FMCOL, GMCOL, QFCOL, ICOL) has 3 Minolta (L*, a*, b*) values (in total, 15 color traits.)

reference populations are available. This could be because with the larger sample size, the effect of causative mutations on polygenic quantitative traits might be estimated more accurately. Most of the identified QTL regions were also reported in other GWAS studies that used the same colour traits (Zhang *et al.* 2015; Yang *et al.* 2017). The most significant QTL region reported by Zhang *et al.* (2015) was located on SSC15 spanning 133 to 134 Mb which explained 3.51 to 17.06% of genetic variance for PH and some colour traits such as FMCOL Minolta a* (Figure 1). This region is very close to the previously reported gene *PRKAG3* controlling both meat PH and colour in pigs. This region on SCC15 was identified by both densities and the three datasets (CBs, PBs, and combined data). Our results provide additional insights for the genomic regions affecting meat and quality traits. However, overall the use of WGS for GWA studies was relatively disappointing in terms of the identifying new QTL, most likely due to power issues.

Genomic prediction. Simulations showed that if variation from causative variants can be captured with WGS, and exploited in genomic prediction, accuracy may be increased by 2-30% (Druet *et al.* 2014). Nevertheless, in this present study, there was generally no additional gain in prediction accuracy of pig carcass colour traits using WGS over medium-density panels for both within and across population scenarios. Moreover, pre-selection of the variants based on significant variants detected by GWAS did not improve accuracy of the predictions (results not shown).

Within population predictions. For within PB genomic prediction, WGS did not provide more accurate breeding values with the exception of QFCOL Minolta b* (from 0.26 to 0.29) and ICOL Minolta b* (from 0.31 to 0.37). Within CB genomic prediction, the use of WGS did not improve the prediction accuracy with the exceptions of FCOL Minolta b* (increase from 0.10 to 0.12) and QFCOL Minolta L* (from 0.22 to 0.23).

Between population predictions: prediction of CBs. When PBs were used as training for predicting breeding values of CBs, using WGS more accurate breeding values were found for FCOL Minolta a* (from 0.15 to 0.28), ICOL Minolta b* (from 0.49 to 0.54), ICOL Minolta L* (from 0.25 to 0.35). Using combined CBs and PBs as training for CB predictions, the improvement in prediction accuracy was only observed for FCOL Minolta a* (from 0.59 to 0.62), FMCOL Minolta b* (from 0.11 to 0.14), FMCOL Minolta L* (from 0.04 to 0.10), GMCOL Minolta b* (from 0.20 to 0.31), and ICOL Minolta L* (from 0.19 to 0.24).

Between population predictions: prediction of PBs. When CBs or combined CBs and PBs were used as training for PB predictions, WGS did not improve accuracy of predicting breeding values compared with a medium-density panel, except for slight improvement in prediction accuracy for some of the pork colour traits. Possible reasons for the lack of improvement when WGS is used over the medium-density panels include the following: (1) with small number of individuals in the training population, effects of QTL may be estimated with large errors (Druet *et al.* 2014); (2) in the case of within-population, long extension linkage disequilibrium may prevent the precise localization of quantitative trait nucleotides when all sequence variants are fitted simultaneously (van Binsbergen *et al.* 2015); and (3) many different linear combinations of variants may exist and result in equally accurate genomic predictions for the same set of

Table 1. Novel genomic regions detected by WGS in PB pigs.

Trait	Genomic region in Mega base pairs (Mb)
FMCOL Minolta b*	SSC2 (142.79-144.77)
FMCOL Minolta L*	SSC10 (38.12-38.28)
PH	SSC2 (146.65-147.87), SSC6 (168.26-168.33)
Belly bend	SSC7 (111.48-118.67), SSC8 (124.05-124.13), SSC13 (110.97-123.61)

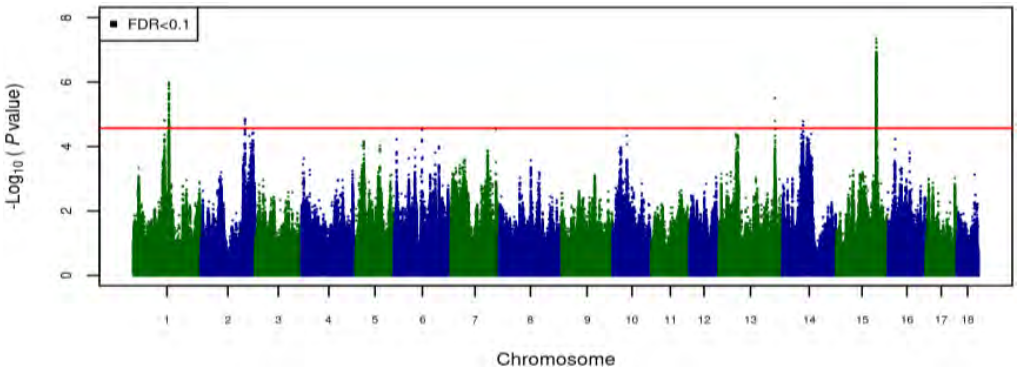


Figure 1. GWAS for FMCOL Minolta a* in combined CBs and PBs. −log10 *P*-values are plotted against genomic position for 18 autosomes. The horizontal red line represents a False Discovery Rate (FDR) of 10%.

phenotypes. Thus, to capture the full advantage of WGS, it is suggested to first preselect markers with a potential causal effect on the trait of interest and use only those preselected variants for prediction.

Conclusions

We conclude that any added value of WGS relative to lower density marker panels for GWAS will likely require larger training populations than those used here. For most pork meat quality traits, WGS hardly improved the accuracy of prediction compared with the medium-density SNP panel. Differences between traits in the effect of panel density and prediction scenario on genomic prediction accuracy can mainly be due to underlying genetic architecture of the traits. For between population predictions, we expect that the results to be specific for the current data set with low relationship between selection candidates and training data (PB and CB populations).

References

Druet T., Macleod I.M., and Hayes B. J. (2014) Heredity 112(1):39-47.

Miar, Y., Plastow G.S., Moore S.S., Manafiazar G., Charagu P., *et al.* (2014) J. Anim. Sci. 92: 2869-2884.

Yang J., Zaitlen N.A., Goddard M.E., Visscher P.M., and Price A.L. (2014) Nat. Genet. 46(2):100-106

van Binsbergen R., Calus M.P.L, Bink M.C.M., van Eeuwijk F.A., Schrooten C., *et al.* (2015) Genet. Sel. Evol. 47:71.

Van den Berg S., Vandenplas J., van Eeuwijk F.A., Bouwman A.C., Lopes M.S., *et al.* (2019) Genet. Sel. Evol. 51:2.

Yang T., Wang Z., Miar Y., Bruce H., Zhang C., *et al.* (2017) Can.J.Anim.Sci. 97(4):721-733.

Zhang C., Wang Z., Bruce H., Kemp R.A., Charagu, P., *et al.* (2015) BMC Genet. 16:33.

262. A conditional multi-trait sequence GWAS of heifer fertility in tropically adapted beef cattle

M. Forutan[1*], B. Engle[1], M.E. Goddard[2] and B.J. Hayes[1]

[1]Centre for Animal Science, Queensland Alliance for Agriculture and Food Innovation, The University of Queensland, Brisbane, QLD 4072, Australia; [2]Agriculture Victoria, AgriBio, Centre for AgriBioscience, Bundoora, Victoria, Australia; m.forutan@uq.edu.au

Abstract

Reproductive efficiency is major driver of profitability of beef production in northern Australia. Failure of heifers to achieve pregnancy results in a considerable loss to producers and accounts for most of the reproductive costs incurred. Availability of genomic information has provided a noteworthy opportunity to enhance the efficiency of selection, especially for traits such as fertility with low heritability. Here, we used imputed whole genome sequence data and up to 2,119 Brahman heifers in a stepwise conditional multi-trait GWAS (CM-GWAS) to identify pleiotropic putative causal variants and genes associated with fertility traits. We identified multiple genomic regions affecting fertility traits, some previously reported and some novel. Candidate gene findings included the genes *ARHGEF28, RNF150*, *EPHA6*. When we investigated the genes closest to the most significant SNP, they included *CNTNAP4* gene that encodes a member of the neurexin protein family and *PLAG1* gene.

Introduction

Over the past decade, genome-wide association studies (GWAS) have been widely used to identify associations of single-nucleotide polymorphisms (SNPs) with complex traits in livestock, especially in dairy and beef cattle (Visscher *et al.* 2017). Using whole genome sequence data can improve the power of GWAS since the causal variants should be included in the sequence data. However, even using whole genome sequence data, the identification of the causal variants for a complex trait is still difficult. This is due to the small effect size of most causal variants and the linkage disequilibrium (LD) between variants. Consequently, there are too many variants in high LD, any one of which could be the cause of the variation in phenotype. Causal variants are often pleiotropic, i.e. affecting more than one trait, so multi-trait analysis might result in higher power to detect quantitative trait loci (QTL) and greater precision in mapping them (Bolormaa *et al.* 2014).

In this study, the stepwise conditional multi-trait GWAS (CM-GWAS) analysis was performed to identify significant variants and genes associated with four fertility-related traits, including a binary trait (heifer pregnancy status) and three continuously distributed traits including fetal age in weeks, measured via manual palpation at pregnancy diagnosis; heifer age at first calving, defined as the number of days between the birthdate and calving date; and days to calving, defined as the number of days between the date of bull turn out at the beginning of the breeding season and calving date.

Materials & methods

Phenotypic data. For this study, lifetime productivity of a Central Queensland Brahman cow herd was assessed. Born between 1981 and 2015, these cows and heifers were part of a stud herd that has been developed with a heavy emphasis on fertility, where failure to produce a calf was the primary culling criterion. Four fertility-related traits were considered. In these herds, heifer pregnancy status was recorded as a binary trait (1 = successful, 0 = unsuccessful) indicating whether a heifer was able to conceive prior to three years of age. Pregnancy success was determined at time of yearly pregnancy check for all heifers born in 2011 and later. For heifers born prior to 2011, calving records were used to determine pregnancy success. Among those heifers that had both a pregnancy test and calving record available, 6% experienced

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_262, © M. Forutan *et al.* 2022

pregnancy loss after pregnancy check, making calving success a good approximation in cases where pregnancy check records were unavailable. Fetal age in weeks (hef_wks_preg) was recorded via manual palpation at pregnancy diagnosis for all heifers born in 2011 and later. Age at first calving (AFC) was only available for heifers with a recorded birth date and was calculated as the difference in days between first calving and birth. Days to calving (DTC) is a routinely recorded trait in Australia Brahmans and is defined as the number of days between the date of bull turn out at the beginning of the breeding season and calving date. Heifer days to calving was recorded as the number of days between first calving and bull exposure.

Genotyping and quality control. Heifers were genotyped with the BovineSNP50 BeadChip (Illumina, San Diego, CA). A detailed description of the genotype quality control was given by Hayes *et al.* (2019). Genotypes were imputed up to 728,785 SNPs (Bovine HD array) using the findhap4 software (VanRaden *et al.* 2013), and a panel of 4650 individuals from relevant breeds, including Brahman (300), Droughtmaster (300), Santa Gertrudis (250) and composites (1000) that were genotyped with the Bovine HD array. All genotypes were then imputed to 31,140,417 million whole-genome sequence variants using the 1000 Bull Genomes Run8, TaurIndicus reference (Hayes and Daetwyler 2019), with 600 Holsteins and 400 Simmental animals removed to avoid over-representation of these genomes in the imputation. Eagle (Loh *et al*., 2016) was used for phasing and Minimac3 software (Das *et al*., 2016) for imputation.

Genome-wide association analysis. A linear mixed model was performed using the GCTA software (Yang *et al.* 2011), fitting each sequence variant as a covariate, one at a time, and testing for association with each trait as follows:

$$\boldsymbol{y} = \mathbf{1}_{\boldsymbol{n}}\mu + \boldsymbol{X}\beta + \boldsymbol{Z}g + \boldsymbol{W}_{\boldsymbol{i}}\alpha_i + \boldsymbol{e},$$

where y is the vector of phenotypic values of the animals, 1_n is an n × 1 vector of 1s (n=number of animals with phenotypes), μ is the overall mean, X is an n × x matrix of fixed covariates, β is a length × vector of fixed effects, Z is a design matrix for the random additive genetic effects; g is a vector of random additive genetic effects assumed to be distributed as ~N(0, $G\sigma_g^2$), where G is the genomic relationship matrix (GRM) calculated from high-density genotypes using the GCTA software. W_i is a vector of genotypes for each animal at the i-th variant, α_i is the corresponding variant effect, and e is a random vector of length n as ~ N (0, σ_e^2I), where σ_e^2 represents non-genetic variance due to non-genetic effects assumed to be acting independently on animals. The choice of fixed covariate effect for continuous and binary traits was done using *lm* and *glm* function in R, respectively. For all four traits, year of birth and contemporary group were constantly considered as fixed covariates. Moreover, for trait AFC, calving success defined as 0 and 1, based upon whether they joined at two or three years of age was considered as fixed effect. Also, for trait DTC, the effect of heifer age of joining was fitted as continuous covariate fixed effect. The GRM was generated using variants with minor allele frequency (MAF) higher than 0.01 (609,878 SNPs) in the HD dataset.

We performed a conditional multi-trait meta-analysis (CM-GWAS) according to a previously described approach (Bolormaa *et al.* 2021) using whole genome sequence (WGS) variant effects estimated from four single-trait GWAS to identify pleiotropic variants that affected fertility trait. The multi-trait meta-analysis (M-GWAS) X^2 statistic with 4 degrees of freedom (equal to the number of traits analysed) was calculated as below:

$$X^2 = t_i V^{-1} t_i$$

where t_i is a vector of the signed t-values of the effects of the i-th sequence variants for the 4 traits and V^{-1} is the inverse of the 4 × 4 correlation matrix where the correlation was calculated over all estimated

sequence variants effects (signed t-values) between each pair of traits. The CM-GWAS approach cycles back and forward between the single-trait GWAS for all traits and M-GWAS to re-test variants conditional on jointly fitting the most significant putative causal variants from independent QTL (where we defined significant as $P<5\times10^{-6}$). To determine independent sequence variants, first the most significant M-GWAS variant from each chromosome was selected and added to the list of putative causal variants. If the pairwise LD between this variant and any other significant variant on the same chromosome was greater than 0.1, these other variants are considered as potentially tagging the same causal variant and were not considered as independent QTL for this cycle. Then, from the remaining significant variants in LD, r^2 less than 0.1, the next most significant variant was selected on each chromosome, LD was tested between these and the remaining significant variants and so on, until no more significant variants identified in this cycle.

Results & discussion

There was no indication of inflation of the test statistic due to population structure or other confounding effects for any of the traits (Figure 1B). To be able to identify independent putative causal variants we used the CM-GWAS. We selected only the most significant ('top') variants from each cycle of the CM-GWAS and identified 144 independent putative causal variants across the genome ($P<5\times10^{-8}$). Candidate genes included the genes *ARHGEF28, RNF150*, *EPHA6* located in intronic region of chromosomes 20, 17, and 1, respectively. When we investigated the genes closest to the most significant SNP, they included *CNTNAP4* gene that encodes a member of the neurexin protein family. Members of this family function in the vertebrate nervous system as cell adhesion molecules and receptors. Moreover, we discovered the same mutation as the one Bouwman *et al.* (2018) identified in a very large meta-analysis of height in cattle using imputed sequence data on more than 50k animals in *PLAG1* gene (14: 23300304). In this study, new potential causative mutations affecting fertility traits have been identified that should be incorporated into commercial SNP arrays. This should increase the accuracy of genomic breeding values (GEBV) for these traits, just as a similar approach has increased accuracy of GEBV in sheep and dairy cattle (Moghaddar *et al.* 2019, Xiang *et al.* 2019).

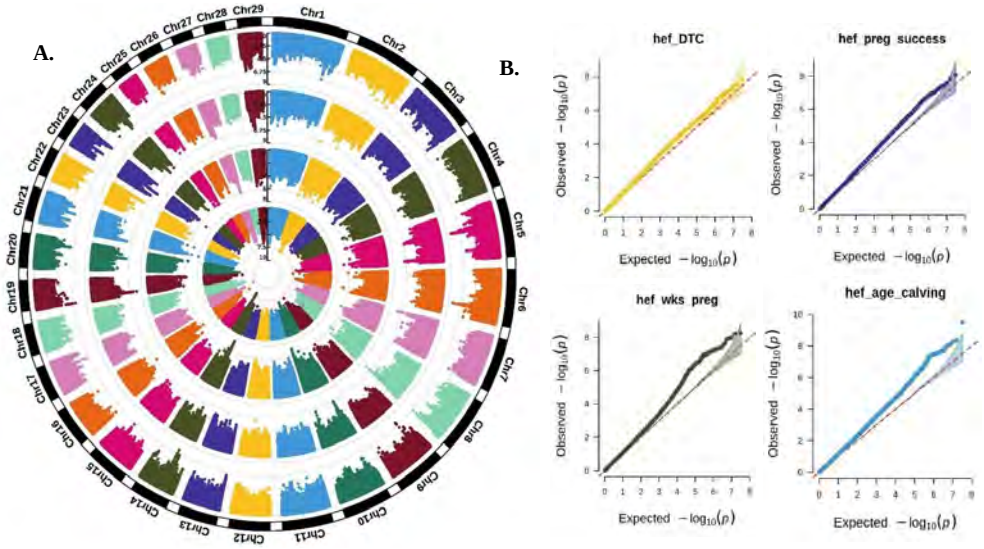


Figure 1. (A). Manhattan plot of single-trait GWAS analysis for four fertility traits of heifer day to calving (her_DTC), heifer pregnancy success (hef_preg_success), heifer weeks pregnant (hef_wks_preg) and heifer age of calving (hef_age_calving) in *Bos indicus* cows. (B). Quantile-quantile (QQ) plot of the single-trait GWAS shown in the Manhattan plot for 4 fertility traits.

References

Bolormaa S., Pryce J.E., Reverter A., Zhang Y., and Barendse W., *et al.* (2014) PLoS Genet 10(3): e1004198.
Bolormaa S., Swan A.A., Stothard P. *et al.* (2021) Genet Sel Evol 53, 58.
Bouwman A.C., and Daetwyler H.D. *et al.* (2018) Nat Genet. 50(3):362-367.
Das S., Forer L., Schönherr S., Sidore C., and Locke A.E., *et al.* (2016) Nat Genet. 48:1284–7.
Hayes B.J., and Daetwyler H.D. (2019) Annual review of animal biosciences 7: 89-102.
Hayes B.J., Corbet N.J., Allen J.M., Laing A.R., and Fordyce G. *et al.* (2019) J Anim Sci. 97:55–62.
Loh P.R., Danecek P., Palamara P.F., Fuchsberger C., and Reshef Y.A. *et al.* (2016) Nat Genet. 48:1443–8.
Moghaddar N., Khansefid M., van der Werf J.H.J, Bolormaa S., and Duijvesteijn N., *et al.* (2019) Genet Sel Evol. 5:51(1):72.
VanRaden P.M., Null D.J., Sargolzaei M., Wiggans G.R., and Tooker M.E., *et al.* (2013) J Dairy Sci. 96(1):668-78.
Visscher P.M., *et al.* (2017) Am. J. Hum. Genet. 101, 5–22.
Xiang R., Berg I.V.D., MacLeod I.M., Hayes B.J., Prowse-Wilkins C.P., *et al.* (2019) PNAS 116(39):19398-19408.
Yang J., Lee S.H., Goddard M.E., and Visscher P.M. (2011) Am. J. Hum. Genet. 88:76-82.

263. Genomic prediction of purebred Merinos from an admixed terminal population using selected sequence variants

N. Moghaddar[1*] A.A. Swan[2], P.M. Gurman[2], L. Li[2], and J.H.J. van Der Werf[1]

[1]University of New England, NSW, 2351, Australia; [2]AGBU, a joint venture of NSW Department of Primary Industries and University of New England, 2351, Armidale, Australia; n.moghaddar@une.edu.au

Abstract

The objective of this study was to evaluate genomic prediction accuracy of meat quality traits in purebred Merinos from an admixed terminal sire breed population, including terminal and Merino crosses, using sequence variants selected in the admixed terminal sire breed population. Results showed an increase in genomic prediction accuracy for carcase fat in Merinos compared to using a common 50k genotypes by using significant SNPs selected from the admixed population. This was supported by GWAS results indicating several common significant QTL regions for carcase fat between the two populations. Results for intramuscular fat showed a decrease in prediction accuracy in Merinos by using significant SNPs selected from the admixed population and in line with this GWAS showed fewer significant common QTL regions between the two populations. Results of this study suggest that improvement in the accuracy of across-population genomic prediction by using selected sequence variants varies between traits and depends on the degree of similarity between populations at significant chromosomal regions affecting the trait.

Introduction

Whole genome sequence data covers a large number of genome-wide genetic variants in which a part of them are at or in significant association with quantitative trait loci affecting the genetic variance of a production trait. Literature shows higher prediction accuracy by including selected sequence variants in prediction within population (Van den Berg *et al.* 2016; Moghaddar *et al.* 2019). For some breeds there are limited data recorded for production traits, making it difficult to detect predictive SNPs within breed. However, data from other breeds or admixed populations may be available, and the question is under which conditions these data can be used for SNP selection and genomic prediction. This study aimed to investigate the accuracy of genomic prediction for two meat quality traits in purebred Merinos from an admixed terminal sire breed sheep population (terminals) using selected sequence variants from a genome-wide association study for genomic prediction in terminals.

Materials & methods

Phenotypes for carcase fat (CCFAT) and intramuscular fat (IMF) in Merinos and terminals were extracted from routine genetic evaluation. These data were collected in a mixture of research (Information Nucleus Flocks (Van der Werf *et al.* (2010)) and MLA resource flock) and industry flocks (Sheep Genetics). The number of terminal sheep with both phenotypes and genotypes used in this study was 13,008 and 12,926 for CCFAT and IMF respectively. The number of validation animals (Merino population) were 1,438 and 1,432 for CCFAT and IMF respectively. The Merino population consisted of purebred animals and the terminal population was a multi-breed/admixture of terminal sire breeds including predominantly Poll Dorset, White Suffolks and composite terminals, including crossbreds with maternals and Merinos.

Imputed whole genome sequence data were available from previous studies and were used in this study. Dscription of the genotype imputation is provided in Bolormaa *et al.* (2019). Briefly, research and industry data with different sets of low and moderate density genotypes (12k, 15k, 50k) were imputed to a common 60k genotype set based on a large imputation reference set, then the 60k genotypes were imputed to high-density genotypes (500k: HD) using a 2,266 multi-breed reference set. Subsequently, all animals with HD

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_263, © N. Moghaddar *et al.* 2022

genotypes were imputed to sequence level using 726 multi-breed animals as a reference set. The final set of sequence data covered 36,934,305 variants on 26 autosomes after standard quality control and filtration for variants with low imputation accuracy (r<0.63).

Pre-corrected phenotypes were obtained from Sheep Genetics routine evaluation analyses, which are run separately for Merino, maternal and terminal sire breeds. The pre-corrected phenotypes ($y^* = \hat{u} + res$) used in this analysis were calculated from multi-trait estimated breeding values ($\widehat{u}$) and residual phenotypes adjusted for all effects in the model (*res*). These effects included the fixed effect of contemporary group (flock, year, sex, measurement date and management group), birth and rearing type, age of dam, age at measurement and weight at measurement. In addition to breeding values, the random effects adjusted for included maternal effects, permanent environmental effect for repeated observations of certain traits, and genetic groups.

Genome-wide association analysis was performed on terminals according to the single SNP mixed model regression method based on the following equation $y^* = Xb + Zu + e$ in Gemma V0.96 program (Zhou *et al.* 2014). In this equation y^* refers to the pre-corrected phenotypes, b refers to fixed effect of mean and allele substitution effect of the investigated SNP, u refers to the random additive genetic effect of the animal fitted by genomic relationship matrix (G), and e is the residual effect. G was calculated using 50k genotypes based on the Yang *et al.* 2010 method, and X and Z are incidence matrices relating fixed and random effects to phenotypes. Genomic prediction was performed with GBLUP based on the following equation $y^* = Xb + Zu + e$ implemented in the MTG2 program (Lee *et al.* 2016). In this equation y^* refers to the pre-corrected phenotypes, b refers to animal group (terminal or Merino), u refers to the random additive genetic effect of the animal which were fitted by genomic relationship matrix (G), and e is the residual effect. G was calculated based on Yang *et al.* 2010 using common 50k genotypes, whole genome sequence data, SNPs with non-zero effect and SNPs with non-zero effect pruned for LD, and X and Z are incidence matrices relating fixed and random effects to phenotypes.

Results

Results of genomic prediction are reported in Figure 1. Common 50k genotypes showed an accuracy of 0.35 and 0.32 for CCFAT and IMF, respectively. Whole genome sequence data showed a similar or lower prediction accuracy than the 50k genotypes. The prediction accuracy in cases when SNPs with non-zero effect derived from GWAS on terminals were used, was increasing according to stronger *P*-value threshold in CCFAT and showed a maximum of 0.04 absolute value of increase in accuracy, however, prediction accuracy in IMF was decreasing by using non-zero effect SNPs with stronger *P*-value threshold derived from terminals. LD pruned SNPs in both CCFAT and IMF showed a similar accuracy to unpruned SNPs and the bias of prediction was in line with accuracy and was increasing with more accurate genomic prediction.

Results of GWAS for CCFAT and IMF using imputed whole genome sequence data (Figure 2 and 3) showed more similarity between significant regions in CCFAT between terminals and Merino but showed more different significant regions for IMF between terminals admixed population and Merinos. More notably two highly significant regions for IMF in terminals on OAR18 and OAR2 were not associated with IMF in Merinos.

Discussion

This study investigated the predictability of genomic breeding values of two common meat quality traits, including carcase fat and intramuscular fat, in Merino sheep based on prediction from an admixed terminal sheep population and using selected variants in admixed terminal population. A common set

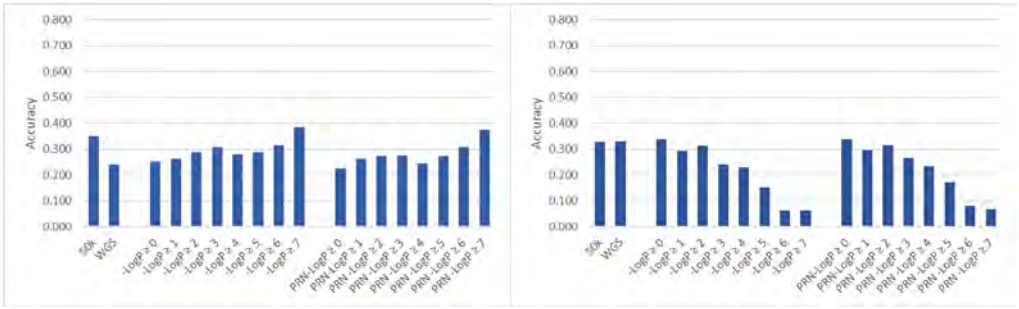


Figure 1. Accuracy of genomic prediction in different scenarios for CCFAT (left) and IMF (right). 50k:50k genotypes, WGS: whole genome sequence, -Log*P*≥0 to -Log*P*≥7 are SNPs filtered based on their *P*-value, PRN: SNPs filtered based on *P*-values and pruned for LD (window size 5,000, sliding window size=100).

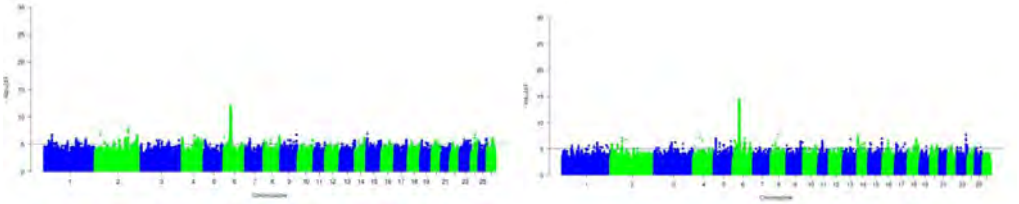


Figure 2. Results of GWAS on CCFAT in terminals (left) and in Merino population (right).

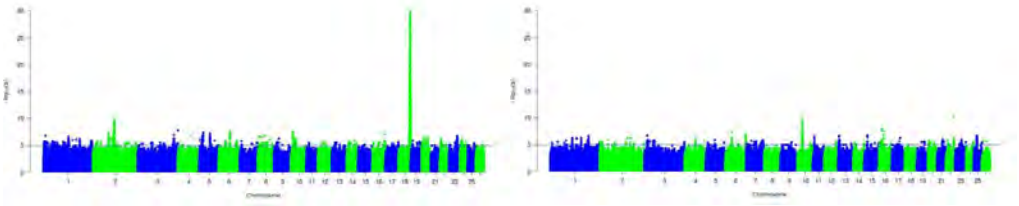


Figure 3. Results of GWAS on IMF in terminals (left) and in Merino population (right).

of 50k SNPs, whole sequence data and non-zero effect sequence variants selected in terminals were used in GBLUP prediction. Whole genome sequence data showed similar or even lower prediction accuracy to 50k which was because of very large number of SNPs provided by sequence data of which the majority are in very low linkage disequilibrium to QTLs. This result was in line with previous findings (Moghaddar *et al.* 2019). Different sets of selected SNPs with non-zero effect according to loose, moderate or stronger cut-off thresholds in terminal population GWAS showed an increasing trend of accuracy in CCFAT but a decreasing trend of accuracy in IMF. This was mainly because of more similar significant QTL regions controlling CCFAT between Merino and terminals which was revealed in GWAS results including highly significant region on OAR6 and different QTL regions affecting IMF between terminals and Merinos, more importantly on OAR 18 and 2 and in general small number of selected SNPs with stringent threshold like –Log *P*-val≥7 fo IMF. Generally low prediction accuracy in this study (less than 0.4) is also because of lower degree of genetic relationship between Merino validation set and terminals as reference set.

Conclusions

This study showed that the genomic prediction accuracy of purebred Merinos from an admixed population is trait specific and depends on the genetic structure of the traits between populations and the degree of similarity between significant chromosomal regions between populations.

Acknowledgements

We acknowledge Sheep Genetics for providing industry data sets. The work was funded by Meat and Livestock Australia (MLA), project L.GEN.1815.

References

Bolormaa S., Chamberlain A.J., Khansefid M., Stothard P., Swan A.A., Mason B., & MacLeod I.M. (2019) Genet Sel Evol. 51: 1-17. https://doi.org/10.1186/s12711-018-0443-5

Gilmour A.R., Gogel B.J., Cullis B.R., Welham S., & Thompson R. (2015) Hemel hempstead: VSN international ltd. 10.4236/ajps.2016.712158

Lee, S.H., & J.H.J. van der Werf. 2016. Bioinformatics, 32(9): 1420–1422. https://doi.org/10.1093/bioinformatics/btw012

Moghaddar, N., M. Khansefid., J.H.J. van der Werf *et al.* (2019) Genet Sel Evol 51, 72. https://doi.org/10.1186/s12711-019-0514-2

Van den Berg, I., D. Boichard & M.S. Lun., 2016. Genet Sel Evol. 48: 83 https://doi.org/10.1186/s12711-016-0259-0

Van der Werf, J.H.J., B.P. Kinghorn, & R.G. Banks. 2010. Anim. Prod. Sci., 50(12): 998. https://doi.org/10.1071/AN10151

Yang J., B. Benyamin., B.P., McEvoy., *et al.*, 2010. Nature Genetics. 42(7): 565 https://doi.org/10.1038/ng.608

Zhou X., & Stephens M., (2014). Nature Methods 11, 407–409. https://doi.org/10.1038/nmeth.2848

264. The complete mitochondrial genome of a Peruvian creole cattle (*Bos taurus*) and its phylogenetic analysis

C.I. Arbizu[1*], R.D. Ferro-Mauricio[1], J.C. Chávez-Galarza[1], H.V. Vásquez[1], J.L. Maicelo[1], C. Poemape[1], J. Gonzales[2], C. Quilcate[1] and F-A. Corredor[1]

[1]Dirección de Desarrollo Tecnológico Agrario, Instituto Nacional de Innovación Agraria (INIA), Av. La Molina 1981, Lima 15024, Peru; [2]Laboratorio de Biología Molecular, Universidad Nacional de Frontera, Av. San Hilarión 101, Sullana 20103, Peru; carbizu@inia.gob.pe

Abstract

The population of Peruvian creole cattle (PCC) is decreasing mainly due to the introduction of more productive breeds in recent years. We report the complete mitochondrial genome sequence of a PCC bull for the first time. This genome was 16,339 bp in length with the base composition 31.43% A, 28.64% T, 26.81% C, and 13.12% G. It contained 13 protein-coding genes, 2 ribosomal RNA genes, 22 transfer RNA genes and a control region. Among the 37 genes, 28 were positioned on the H-strand and 9 were positioned on the L-strand. The most frequently used codons were CUA (Leucine), AUA (Isoleucine), AUU (Isoleucine), AUC (Isoleucine) and ACA (Threonine). Maximum likelihood analysis clearly demonstrated that PCC are strongly related to a native African breed, giving insights into the maternal ancestry of PCC. The annotated mitochondrial genome of PCC would serve as an important genetic data set for further breeding work and conservation strategies.

Introduction

The global population of cattle (*Bos taurus* and *Bos indicus*) is around 1.5 billion (McLeod, 2011), making it one of the most common livestock species. The cattle population in Peru is about 5.5 million head (MIDAGRI, 2019) of which 5% are creole cattle (M. Rosenberg, UC del Sur, pers. Comm.). PCC, in difference to other exotic breeds, have developed the ability to survive in environments with food limitations and under climatic factors that heavily affect their performance. Therefore, PCC are the basis for milk and beef production in the Peruvian highlands. In some Peruvian regions, PCC are also used for agricultural purposes and cultural events. In this work, for the first time, we sequenced the complete mitochondrial genome of a Creole bull from Peru.

Materials & methods

Sample collection and DNA extraction. A tail hair sample was collected from a single male specimen from the Andagua district, Castilla province in Arequipa (3,574 masl; -15.499548°, -72.359927°). This individual named as 'Tremendo Cachorro' was considered an 'Arequipa fighting bull' as it possessed exceptional fighting skills and was part of the traditional bullfight activity of Arequipa. We extracted genomic DNA with the Wizard® Genomic DNA Purification Kit (Wisconsin, USA) following the manufacturer's instructions. The quality and quantity of genomic DNA were assessed by agarose gel electrophoresis and NanoDrop One (Thermo Fisher Scientific, USA), respectively. Library preparation and high-throughput sequencing with Illumina HiSeq Sequencing System (Illumina, San Diego, CA) were conducted by Genewiz (New Jersey, USA).

Assembly, annotation and sequence analysis. Pair-end clean reads were obtained by PE 150 library and the Illumina HiSeq 2500 platform. Adapters and low-quality reads were removed using Trim Galore (Martin 2011). We used clean data to assemble the mitochondrial genome using GetOrganelle v1.7.2 pipeline (Jin *et al.* 2020). The annotations of the protein-coding genes (PCGs), transfer RNAs (tRNAs) and rRNA genes from mitogenome were performed using the automatic annotators of mitochondrial genes

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_264, © C.I. Arbizu *et al.* 2022

online: Geseq (Tillich *et al.* 2017) and MITOS 2 (Donath *et al.* 2019), and curated manually. The tRNA secondary structure was analyzed with tRNAscan-SE 2.0 (Chan *et al.* 2021). The codon usage bias was analyzed using MEGA X (Kumar *et al.* 2018). The circularized drawing of the mitochondrial genome was performed with OGDRAW 1.3.1. (Greiner *et al.* 2019).

Codon usage and tRNA analysis. Codon usage analysis was performed in Molecular Evolutionary Genetics Analysis 11 (MEGA 11) software (https://www.megasoftware.net/). For this purpose, each nucleotide sequence was cured manually, removing the STOP codon and only keeping the codons that synthesize amino acids. Then, 13 nucleotide sequences were concatenated using the Concatenate Sequence Alignment option and we evaluated the Codon Usage using RCSU option in MEGA 11. To predict the secondary structure of each tRNA, we used the tRNAscan-SE website server (http://lowelab.ucsc.edu/tRNAscan-SE/). We analysed the sequences in FASTA format and kept the parameters by default except the Sequence Source that corresponds to Mammalian mitochondrial.

Phylogenetic analysis. To ascertain the genetic relationship of PCC with other *Bos* species, we also included 32 complete mitochondrial genomes from *Bos* species available in GenBank. As outgroup, we used a species of the genus Bison (*Bison bison*) from the same subfamily Bovinae. We used GTR+GAMMA model of evolution to obtain the best-scoring ML tree, and then 1000 nonparametric bootstrap inferences were performed with RAxML v8.2.11 (Stamatakis 2014).

Results

The complete mitochondrial genome of PCC was 16.339 bp in length, which included 13 protein-coding genes, two ribosomal RNA genes, 22 transfer RNA genes and a control region (Figure 1). Most of the genes (28) were encoded within the heavy (H) strand, and 9 genes were positioned on the L-strand. The base composition of this genome was 31.43% A, 28.64% T, 26.81% C, and 13.12% G. The entire mitochondrial genome sequence was submitted to the GenBank database with accession number: OK135155. The associated Bioproject, Biosample, and SRA numbers are PRJNA763011, SAMN21419641, and SRR15883111, respectively. The relative synonymous codon usage (RSCU) analysis showed the highest utilization of CUA, AUA, AUU, AUC and ACA codons among PCGs. Our phylogenetic analysis showed that PCC clustered as expected in a well-supported clade that included *B. taurus* individuals (Figure 2). Interestingly, this PCC was within subclade 1 that contains most African breeds.

Discussion

We sequenced for the first time the complete mitochondrial genome of a Peruvian creole bull using next generation sequencing. The genome size and order was in agreement to previous work on native cattle (Liu *et al.*, 2020). The nucleotide composition of the PCC mitochondrial genome was biased towards adenine and thymine and was in agreement with other studies in *Bos* (Kamalakkannan *et al.*, 2020). This feature is a common characteristic for the bovine mitochondrial genome (Zhou *et al.*, 2019). Similar to other studies in the *Bos* genus, amino acids leucine, threonine, and isoleucine were the most abundant among the protein coding genes (Kamalakkannan *et al.*, 2020). It was possible to recover the three clades that were reported for *Bos* (Kamalakkannan *et al.*, 2020). In addition, our study provided further evidence about the influence of African and European founders. However, more samples from a wider geographic area in Peru, especially form the Andean region, needs to be analysed using both mitochondrial and nuclear DNA to confirm PCC origin. This newly generated sequence would serve for phylogenetic, evolutionary and conservation work for this orphan breed.

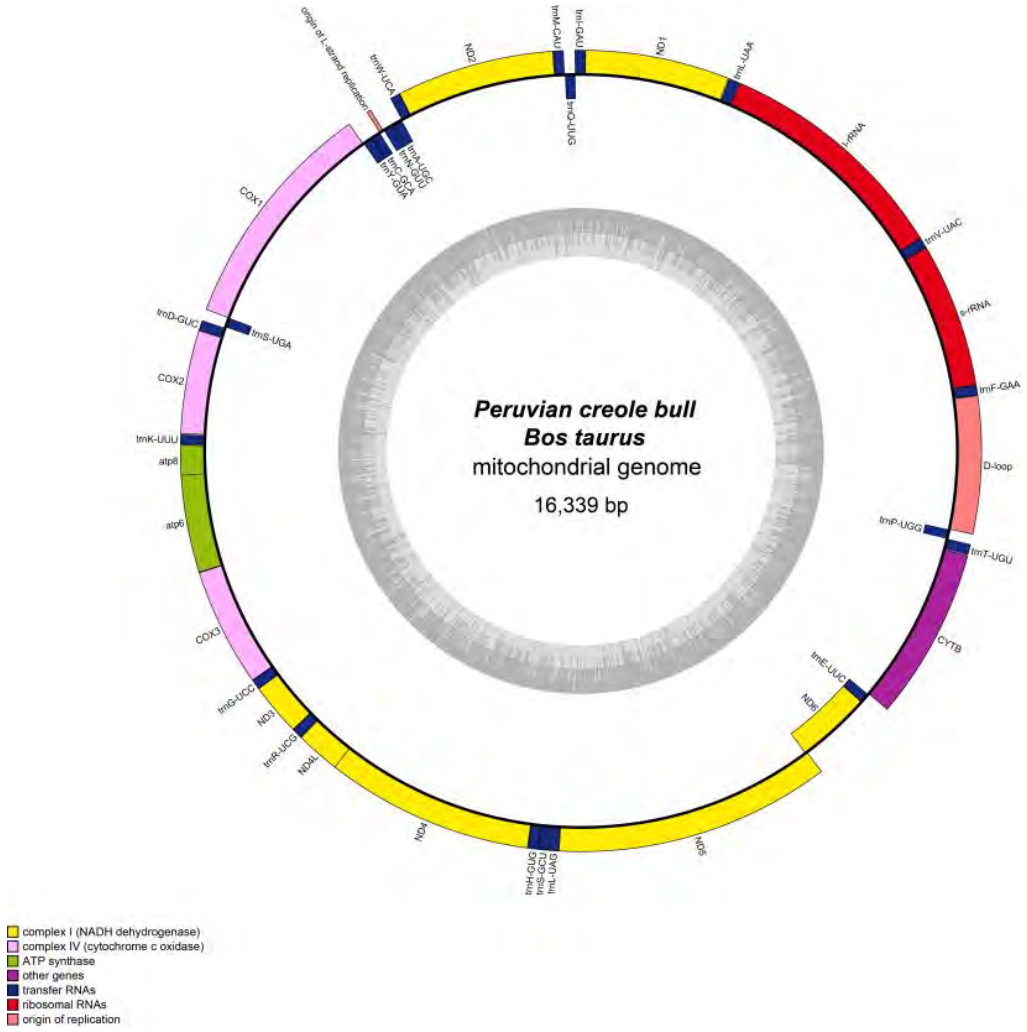


Figure 1. Representative map of the mitochondrial genome of the Peruvian creole cattle. Genes encoded by the heavy strand are shown outside the circle, while those encoded by the light strand are shown inside.

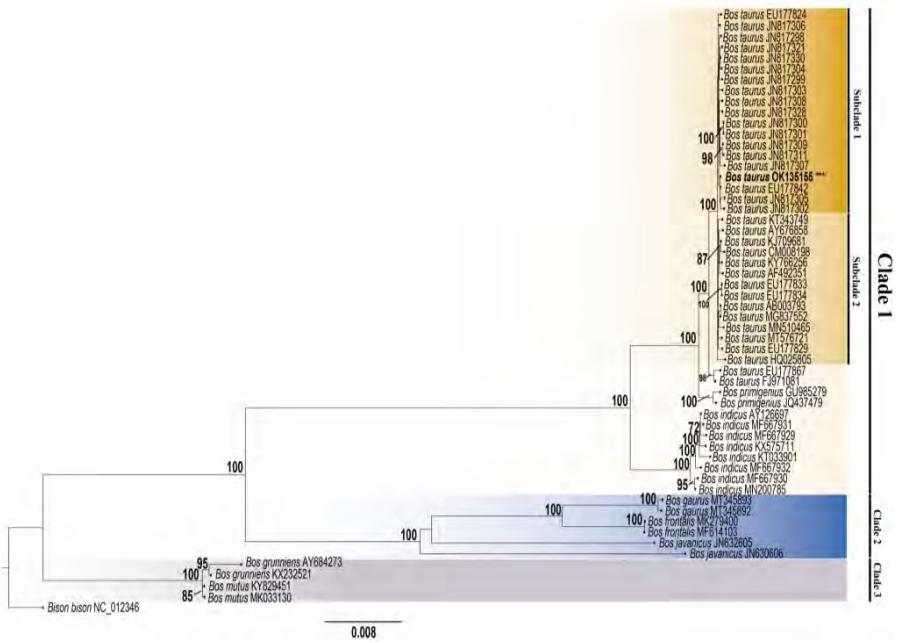


Figure 2. Phylogenetic relationship of *Bos* species inferred from mitochondrial genomic sequences. Maximum likelihood bootstrap support is shown only for branches with <72% support. The *Bison bison* mitochondrial genome sequence was used as an out-group.

References

Ministerio de Agricultura y Riego (MINAGRI), Anuario estadístico producción ganadera y avícola 2019. Available at: https://siea.midagri.gob.pe/portal/phocadownload/datos_y_estadisticas/anuarios/pecuaria/pecuario_2019.pdf

Jin J.J., Yu W.B., Yang J.B., Song Y., Depamphilis, C.W., *et al.* (2020). Genome Biol. 21(1):1-31. https://doi.org/10.1186/s13059-020-02154-5

Stamatakis A. (2014). Bioinformatics 30(9): 1312-1313. https://doi.org/10.1093/bioinformatics/btu033

Liu S-J., Lv J-Z., Tan, Z-Y., Ge X-Y. (2020). Mitochondrial DNA B Resour. 5(1):443-444. https://doi.org/10.1080/23802359.2019.1704639

Kamalakkannan R., Bhavana K., Prabhu V.R., Sureshgopi D., Singha H.S., *et al.* (2020). Sci. Rep. 10(1):1-11. https://doi.org/10.1038/s41598-020-68724-6

Zhou X., Wu X., Liang C., Guo X., Chu M., *et al.* (2019). Mitochondrial DNA B Resour. 4(1):1053-1054. https://doi.org/10.1080/23802359.2018.1546149

Chan P., Lin B., Mak A., Lowe T. (2021) Nucleic Acids Research 49(16):9077–9096. https://doi.org/10.1093/nar/gkab688

Donath A., Jühling F., Al-Arab M., Bernhart S., Reinhardt F., *et al.* (2019) Nucleic Acids Res. 47(20):10543–10552. https://doi.org/10.1093/nar/gkz833

Greiner S., Lehwark P., Bock R. (2019) Nucleic Acids Res. 47(W1):W59–W64. https://doi.org/10.1093/nar/gkz238

Kumar S., Stecher G., Li M., Knyaz C., Tamura K. (2018) Mol. Biol. Evol. 35(6):1547–1549. https://doi.org/10.1093/molbev/msy096

McLeod A. World livestock 2011 - livestock in food security. In: World livestock 2011 - livestock in food security. Food and Agriculture Organization of the United Nations (FAO); 2011.

Tillich M., Lehwark P., Pellizzer T., Ulbricht-Jones E., Fischer A., *et al.* (2017) Nucleic Acids Res. 45(W1):W6–W11. https://doi.org/10.1093/nar/gkx391

Martin, M. (2011). EMBnet. j. 17(1):10-12. https://doi.org/10.14806/ej.17.1.200

265. Unravelling the effect of structural variants from whole-genome sequence for depigmentation in Nellore cattle

G. Vargas[1*], H.H.R. Neves[2], N.A.M. Garzón[1], L.F.S. Fonseca[1], G.A. Fernandes Júnior[1], L.G. Albuquerque[1,3] and R. Carvalheiro[1,3]

[1]School of Agricultural and Veterinarian Sciences, Sao Paulo State University (UNESP), Via de Acesso Prof. Paulo Donato Castelane, 14884-900 Jaboticabal, SP, Brazil; [2]GenSys Associated Consultants, 90680-000 Porto Alegre, RS, Brazil; [3]National Council for Science and Technological Development, 71605-001 Brasilia, DF, Brazil; gi.vargas@hotmail.com

Abstract

The aim of this study was to detect structural variants (SVs) associated with skin depigmentation (DP) through genome-wide association study (GWAS), using whole-genome sequence of 81 Nellore sires. Phenotypic and pedigree data from 340,220 animals were used to estimate breeding values from single-trait analysis using a mixed linear animal model. The Parliament2 software was used to identify SVs. After filtering, a total of 46,850 deletions, 15,690 insertions, 7,937 inversions, and 30,446 duplications were identified within genic regions. GWAS for DP revealed 11 significant SVs located on chromosomes 4, 7, 8, 13, 15, 16, 17, 20 and 28; and harbouring 3 candidate genes (*MAML2*, *UMAD1* and *SAMD8*). The main gene identified was *MAML2*, which is involved in human skin pigmentation and is related to the Notch signalling pathway. Our results provide meaningful information about SVs, which may be useful in future assessment of variations associated to skin depigmentation in Nellore cattle.

Introduction

The genetic progress achieved by several breeding programs in Brazil can be credited in large part to the implementation of CEIP (Special Certificate of Identification and Production), a certificate of genetic superiority issued for the top 20-30% animals (young bulls and heifers). The animals eligible to receive the CEIP are evaluated for morphological and functional defects, including skin depigmentation (DP). The combination of dark pigmentation and white hair typical of Nellore cattle is a powerful feature to prevent skin lesions and subsequent complications that can impact the animal's performance. As a type of genetic variation, structural variation (SVs) often refers to structural differences in the genome of complex organisms, including gains and losses of DNA segments and rearrangements, which are inherited and may contribute to a considerable proportion of the differences between individuals for some traits. SVs have been associated with several productive traits and diseases in livestock species. In this regard, many SVs have also been identified and artificially selected in livestock such as a duplication of an allele of the *KIT* gene that results in a white phenotype in cattle, pigs, and sheep (Giuffra *et al.*, 2002; Gutiérrez-Gil *et al.*, 2007). Thus, the aim of this study was to detect structural variants associated with skin depigmentation in Nellore cattle by performing a genome-wide association study using whole-genome sequence data.

Materials & methods

Phenotypes and pedigree data. Phenotypic and pedigree information of 340,220 Nellore animals, collected between 2000 and 2018, were obtained from Alliance Nellore database (www.gensys.com.br). These animals belong to herds that systematically perform evaluations for morphological and functional defects or disorders.

Whole-genome sequencing dataset. A total of 81 influential Nellore sires were sequenced using the Illumina HiSeq X™ Ten and Illumina NovaSeq™ platforms (Illumina Inc., San Diego, USA). Sequencing of all samples was performed using the Illumina sequencing-by-synthesis technology at an overall average

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_265, © G. Vargas *et al.* 2022

sequence coverage of 14.5×, ranging from 7.8 to 26.3×. The SAMtools program (Li *et al.*, 2009) was used to index the sequences, and the alignment to the ARS-UCD1.2 reference genome was performed using HiSat (Kim *et al.*, 2015).

Variant Calling and Genotyping. Single sample SV discovery (including deletions, insertions, inversions, and duplications) was performed using Parliament2 (Zarate *et al.*, 2020) considering the following combination of callers CNVnator (Abyzov *et al.*, 2011), Lumpy (Layer *et al.*, 2014), Delly (Rausch *et al.*, 2012) and BreakDancer (Fan *et al.*, 2014). The BAM files and associated reference genome file (ARS-UCD1.2) were used to run Parliament2 on each sample via a single Docker image using the DNAnexus platform. A merged VCF file was prepared for each variant calling by merging individual VCF files of the 81 samples using the BCFtools merge function. The output file containing SVs was then transformed to 012 genotypes format using VCFtools (Danecek *et al.*, 2011), where 0 represented the homozygous genotype for the reference allele (no SV), 1 was the heterozygous genotype (heterozygous SV) and 2 was the homozygous genotype for the alternative allele (homozygous SV).

Statistical methods. In preliminary analyses, an additive threshold animal model including a fixed effect of contemporary group (CG) and a linear effect of age at recording as covariate was used to obtain the solutions of mixed model equations using THRGIBBS1F90 software (Misztal, 2014). CG were defined based on the concatenation of season of birth, sex, farm and management group at weaning and yearling, and date of measurement. The GWAS was performed for the 81 sequenced animals using the *fastGWA* (ultra-fast mixed linear model association) option implemented in the GCTA software (Yang *et al.*, 2011), fitting the following linear mixed model:

$$y = \mu + Xb + Zu + e \quad (1)$$

where y is the vector of pseudo-phenotypes (estimated breeding values); μ is the overall mean; X is the incidence matrix of the allele copies (coded as 0, 1, 2); b is the allele substitution effect; Z is the incidence matrix connecting the phenotypes to animals; u is the vector of additive genetic effect; and e is the vector of random residuals. The genomic relationship matrix (GRM) was derived as defined in Yang *et al.* (2010, 2011) based on SVs. According to the Bonferroni correction, the defined GWAS significant threshold was 0.05/N, where N refers to the number of SVs. Significant SVs were investigated for gene annotation using the Genomic Annotation in Livestock por positional candidate Loci (GALLO; http://github.com/pablobio/GALLO) package in R (Version 4.0.0; R Core Team, 2021).

Results & discussion

A total of 146,003 deletions, 52,208 insertions, 84,306 duplications, and 226,036 inversions were identified from the sequence data of the Nellore population, which jointly affect ~2.38 Mb of sequences (Table 1). In general, the number of deletions and inversions detected in each chromosome was greater than the number of duplications and insertions. In addition, the distribution of the SVs detected on each chromosome is quite uniform. Chromosome 1 had the highest density of structural variants, while chromosome 25 had the lowest. The annotation of detected variants was performed to further explore the distribution of SVs in genic regions. Within genic regions there were 46,850 (32.08%), 15,690 (30.05%), 7,937 (3.51%), and 30,446 (36.11%) identified deletions, insertions, inversions, and duplications, respectively. In addition, the number of total genes for deletions, insertions, inversions, and duplications contained in the Ensembl gene database of bovine were 6,483 (37.00%), 5,413 (30.89%), 2,104 (12.01%), and 1,966 (11.22%). The range of the number of deletions, relative to genes distributed on different chromosomes was from 24.23 to 52.87%; for insertions, from 18.30 to 45.97%; for inversions, from 5.08 to 19.4%; and for duplications, from 4.59 to 89.09%. Chromosomes 6, 24, and 28 had the highest number of genes with deletions, insertions, inversions,

and duplications. Single-trait GWAS revealed 11 significant SVs located on chromosomes 4, 7, 8, 13, 15, 16, 17, 20 and 28 (Table 2). A total of three candidate genes (*MAML2*, *UMAD1* and *SAMD8*) were annotated and harbouring three associated chromosomes. The gene protein *MAML2* belongs to the three-member MAML (mastermind-like) family that co-activates Notch receptor-induced transcription. Notch activation regulates skin homeostasis by balancing primarily growth arrest and progressive differentiation processes of keratinocytes, which are associated to skin pigmentation process by accumulating pigmented organelles including melanin (Benito-Martínez *et al.*, 2021). KEGG pathway analysis indicated that the genes with insertions were involved in 3 KEGG pathways (bta04330: Notch signalling pathway; bta04658: Th1 and Th2 cell differentiation; bta05165: Human papillomavirus infection).

Conclusions

In this study, we identified SVs from 81 Nellore sires, including deletions, insertions, inversions, and duplications. The annotation of SVs in different gene families suggests that SVs can be a relevant source of genetic variation for the breed's evolution. The *MAML2* was identified as an important functional candidate gene associated with skin pigmentation, which demonstrate the importance of this gene for the studied trait. The set of SVs reported here could be useful for further studies to investigate the incorporation of SV genotypes into genomic prediction in order to enhance selection and mating decisions aimed to reduce the incidence of skin depigmentation in Nellore.

Acknowledgements

To CNPq (#422863/2018-1), FAPESP (#18/20026-8; # 17/10630-2), and Brazilian Federal Agency for Support and Evaluation of Graduate Education (CAPES), in the scope of the Program CAPES-PrInt (#88887.571090/2020-00) for financial support.

Table 1. Summary of structural variants.

	Deletions	Insertions	Duplications	Inversions	Total
No.	146,003	52,208	84,306	226,036	508,553
Average size (bp)	686,225	1	1,013,137	683,547	2,382,910

Table 2. Significant SVs detected from genome-wide association study for depigmentation in Nellore cattle.

SV Type[1]	Chr[2]	Start (bp)	End (bp)	Var[3]	*P*-value	Candidate gene
DEL	13	80111745	80111794	0.5523	1.19E-08	-
DEL	16	16787955	16788618	0.3831	8.57E-08	-
INS	4	15831101	15831102	0.0999	1.18E-06	*UMAD1*
INS	4	82871449	82871450	0.1296	4.29E-06	-
INS	7	48306627	48306628	0.1488	9.71E-07	-
INS	8	57768936	57768937	0.1450	1.36E-06	-
INS	15	14161805	14161806	0.1166	2.35E-07	*MAML2*
INS	16	9833208	9833209	0.3869	8.31E-08	-
INS	17	25309994	25309995	0.1179	8.03E-07	-
INS	20	69296489	69296490	0.4099	2.43E-06	-
INS	28	30910133	30910134	0.1177	6.70E-06	*SAMD8*

[1] SV Type: deletion (DEL); insertion (INS).

[2] Chr: chromosome.

[3] Var: additive genetic variance.

References

Abyzov A., Urban A.E., Snyder M. and Gerstein M. (2011) Genome Res. 21(6):974-984. https://doi.org/10.1101/gr.114876.110

Benito-Martínez S., Salavessa L., Raposo G., Marks M.S., Delevoye C., *et al.* (2021) Integr Comp Biol 61(4):1546–1555. https://doi.org/10.1093/icb/icab094

Danecek P., Auton A., Abecasis G., Albers C.A., Banks E., *et al.* (2011) Bioinformatics. 27(15):2156-8. https://doi.org/10.1093/bioinformatics/btr330

Fan X., Abbott T.E., Larson D. and Chen K. (2014) Curr Protoc Bioinformatics. 45:15.6.1-15.6.11. https://doi.org/10.1002/0471250953.bi1506s45

Giuffra E., Törnsten A., Marklund S., Bongcam-Rudloff E., Chardon P., *et al.* (2002). Mamm. Genome 13:569–577. https://doi.org/10.1007/s00335-002-2184-5

Gutiérrez-Gil B., Wiener P. and Williams J.L. (2007) BMC Genet. 8:56. https://doi.org/10.1186/1471-2156-8-56

Kim D., Langmead, B. and Salzberg, S. (2015) Nat Methods. 12:357–360. https://doi.org/10.1038/nmeth.3317

Layer R.M., Chiang C., Quinlan A.R. and Hall I.M. (2014) Genome Biol 15:R84. https://doi.org/10.1186/gb-2014-15-6-r84

Li H., Handsaker B., Wysoker A., Fennell T., Ruan J., *et al.* (2009) Bioinformatics. 25(16):2078–2079. https://doi.org/10.1093/bioinformatics/btp352

Misztal, I. (2012). Available at: http://nce.ads.uga.edu/wiki/doku.php?id=application_programs.

Rausch T., Zichner T., Schlattl A., Stütz A.M., Benes V., *et al.* (2012) Bioinformatics. 28(18):i333-i339. https://doi.org/10.1093/bioinformatics/bts378

Yang J., Benyamin B., McEvoy B.P., Gordon S., Henders A.K., *et al.* (2010) Nat Genet. 42(7):565-569. https://doi.org/10.1038/ng.608.

Yang J., Lee S.H., Goddard M.E. and Visscher P.M. (2011) Am J Hum Genet. 88(1):76-82. https://doi.org/10.1016/j.ajhg.2010.11.011

266. On the advantage of identifying causal genetic variants for genomic prediction

T.H.E. Meuwissen[1*] and M.E. Goddard[2,3]

[1]Norwegian University of Life Sciences, Box 5003, 1432 Ås, Norway; [2]Agriculture Victoria, Bundoora, Australia; [3]Faculty of Veterinary and Agricultural Sciences, University of Melbourne, Parkville, Australia; theo.meuwissen@nmbu.no

Abstract

Genome Wide Association Studies have been successful in identifying Quantitative Trait Locus regions. However, the identification of causal variants (CVs) has proven difficult in many cases. One may hypothesize that knowing the CV would not improve the (genomic) prediction of phenotypes much since otherwise phenotypes would clearly point to the CV. The effect of (not) knowing the CV on local estimated breeding values (LEBV) in 200 kb regions surrounding the CV was estimated for 3 well-known CVs (in ABCG2, DGAT and GHR) in a large whole genome sequence dataset. Knowing the CV improved the accuracy of LEBV for all three CVs: for ABCG2 it was instrumental in pointing to the correct CV; the DGAT-CV was poorly imputed, and knowing the CV would result in improved genotyping; and GHR was poorly predicted by other SNPs in the region. In addition, knowing CVs is important for understanding trait-biology and for targeted gene-editing.

Introduction

Quantitative traits are generally complex in the sense that there are many genes, i.e. causal genetic variants, plus environmental effects that determine the trait phenotype. As an example there are 9,863 independent variants with significant ($P<5\times10^{-8}$) effects on human height (Visscher *et al.*, 2020). If many causal variants affect a trait then most of their effects must be relatively small, but some causal variants may have large effects. The distribution of effects sizes of causal variants can be modelled by a thick-tailed distribution such as a Gamma distribution with shape parameter <1 which includes more large effects than a Normal distribution. (Hayes and Goddard, 2001).

The identification of these causal variants with large effects is the aim of Genome Wide Association Studies (GWAS). GWAS have been successful in identifying many Quantitative Trait Loci (QTL), i.e. small chromosomal regions that explain genetic variance for a quantitative trait. However, identifying the causal variant (CV) within a QTL region is often difficult for a number of reasons:

1. The CV may not be included in the genotype data, even if Whole Genome Sequence (WGS) genotype data are used. For instance, the CV may not be included in the reference set of sequenced animals because it is too rare or hard to call as structural variants often are.
2. GWAS is based on Linkage Disequilibrium (LD), but many variants may be in high LD with the CV making it difficult to distinguish between these variants. If the CV and another non-causal variant (NCV) are in close LD, the NCV may show the higher GWAS signal due to an unfortunate sample of animals entering the GWAS. This problem may be reduced by increasing the sample size, but if the LD between the CV and NCV is 100%, a GWAS study cannot distinguish these two variants (since their genotypes are identical). If the LD is say 90%, a huge data set may be needed to ensure that the CV yields the highest GWAS signal.
3. Large WGS data sets are often obtained by genotype imputation, and genotype imputation errors of the CV genotypes may reduce the association between the CV genotypes and the trait.

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_266, © T.H.E. Meuwissen and M.E. Goddard 2022

Given that the identification of causal variants is problematic, the question arises whether the identification of the CV is important? From the perspective of understanding the trait biology, a sharp GWAS peak, without identification of the CV, may identify the gene that is affecting the trait, but it will not be able to answer the question what genetic change in (or around) the gene is causing the (positive or negative) trait effect. The latter also prevents studies investigating the mechanism by which the CV affects the trait. In addition, the CVs provide clear targets for gene editing (Jenko *et al.*, 2015).

However, here, we will take the perspective of genomic prediction, i.e. we are only interested in the accuracy of genomic prediction, and whether knowing the CV increases the accuracy of genomic prediction. Instead of performing computer simulations (where the CV is always known), we will address this question using real data and consider 3 known and well-studied CVs affecting milk traits in dairy cattle: the Y581S mutation in ABCG2 on BTA 6 (Cohen-Zinder *et al.*, 2005); the K323A mutation in DGAT1 on BTA 14 (Grisart *et al.*, 2004); and the F2Y mutation in GHR on BTA 20 (Blott *et al.*, 2003).

Materials & methods

Data. The dataset is described in detail in Meuwissen *et al.* (2021). Briefly, it consisted of WGS genotypes and daughter yield deviations (DYD; in case of bulls) or yield deviations (YD; in case of cows) for milk yield of 35,549 Holsteins, Jerseys, and Australian Red bulls and cows. Animals were either directly genotyped with the Illumina 800K BovineHD bead chip (HD), or first genotyped with the Illumina BovineSNP50K chip or a lower density SNP chip, and subsequently imputed to HD. All individuals were imputed to WGS using a reference population of Holstein, Jersey and Australian Red bulls and cows from Run 5 of the 1000 bulls genome project. After filtering out variants with a minor allele frequency (MAF) lower than 0.002 and LD pruning (r^2>0.9), 4,809,520 variants were retained for the analysis. All three mutations (in ABCG2, DGAT1 and GHR) were included in these data.

Comparison of alternatives. All analyses were performed by the BayesGC model (see Meuwissen *et al.*, 2021, for details), which is a Bayesian variable selection model that fits a genomic relationship matrix (GRM) as well as the variable selection term, which selects individual variants from the set of 4,809,520 variants using a Monte Carlo Markov Chain (MCMC) method. The GRM was based on HD SNP genotypes, which did not include the ABCG2, DGAT1, and GHR CVs. The following alternative analyses were compared:

- AV: all 4,809,520 variants were included in the analysis, including the 3 CV (treated the same as all other SNPs);
- AV-CV: the same as AV but excluding the CV (assessing the value of having the CV in the data or not);
- AV+CV: the same as AV except that the 3 CV were assumed known and fitted with a prior probability of 1, i.e. they were always included in the model;
- CV: fits only the three CVs with prior probability 1;
- TOP10: fits 10 SNPs with highest posterior probability in each of the three QTL regions, i.e. a total of 30 SNPs were fitted with prior probability 1.

The TOP10 analysis considers the situation where the QTL region is detected but it is uncertain which of the SNPs is the causal variant. In order to deal with this uncertainty, the model fits 10 top SNPs (the 10 SNPs with the highest posterior probability in a 200 kb region surrounding the CV). Since there are 3 QTL regions, this results in a total of 30 SNPs being fitted. It may be noted that the top-10 of the variates in the DGAT1 region did not contain the DGAT1-CV probably due to the poor imputation accuracy of this SNP (which was variable possibly because it is a 2-SNP mutation which may confuse some software), and possibly also due to other CV in the region affecting the association. In case of ABCG2, the ABCG2-CV ranked number 8 in its 200 kb region, i.e. it was not the most important SNP in the region. The GHR-CV ranked number 1 in its 200 kb region, and was clearly the most important SNP.

The bayesGC model requires a variance assumed for the fitted variants, which was a fraction of 0.0005 of the total genetic variance (Meuwissen *et al.*, 2021) for the AV, AV-CV and AV+CV analyses. For the CV analysis this factor of the genetic variance was assumed 0.01, which assumes that the CV explain a factor of 0.01 of the genetic variance each (and if the actual explained variance is bigger, this would hardly affect the accuracy of prediction). The TOP10 analysis assumes that each SNP explains 1/10 of the variance of the locus, and in order to agree with the CV analysis each SNP was assumed to explain a factor of 0.001 of the genetic variance.

Criteria for comparison. Because the effect of a single CV on the total EBV is often small, and to avoid investigating very small differences in accuracy, we concentrated on the local EBV (LEBV) of each of the 200 kb-regions surrounding the three CVs. As the criterion for comparison of two analyses we used the correlation between the LEBV coming from a part data set to those from the complete data set, following Legarra and Reverter (2018). This criterion estimates the ratio of the accuracies of the part data set to that of the complete data set. Here complete data means more SNPs included in the analysis or more prior information about these SNPs such as which is the CV.

Results

The accuracy of AV-CV versus AV shows it is only in the case of GHR important to include the CV in the BayesGC analysis (Table 1). The latter is presumably because the surrounding SNPs did not fully pick-up the GHR-CV effect and the accuracy of AV-CV LEBV is reduced by ~27%. The AV versus AV+CV comparison shows that in the case of GHR, the AV model fitted the GHR-CV with a high posterior probability anyway, so giving it a prior probability of 1 hardly affected the LEBV. For DGAT, probably most of the LEBV was due to the effect of other SNPs and fitting the DGAT-CV with a prior probability of 1 did not change this much. However, the ABCG2-CV was somewhat important to the LEBV of its region and giving it a prior probability of 1 did improve the accuracy of the LEBV. The TOP10 versus AV(+CV) comparisons show that fitting only the top-10 SNPs in a region is not sufficient in the case of ABCG2 and GHR. Other SNPs also seem to contribute significantly. In the case of DGAT, the top-10 SNPs could predict the LEBV of the entire region. The TOP10 versus CV comparison shows that the top10 SNPs could generally not predict the effect of the CV. The CV versus AV(+CV) comparisons show that the CV alone often does not explain all genetic variation that could be explained in the region. For the DGAT-CV this may be expected due to its imputation inaccuracy. The GHR-CV is an exception to this result where fitting its effect alone yielded virtually the same LEBV as fitting all SNPs in the region.

Table 1. Ratio of accuracies of LEBV for the alternative analyses.

Data		Ratio of accuracies:		
Partial	**Complete**	**ABCG2**	**DGAT**	**GHR**
AV-CV	AV	0.993	0.999	0.726
AV	AV+CV	0.805	0.9996	0.998
TOP10	AV	0.852	0.998	0.781
TOP10	AV+CV	0.758	0.998	0.754
TOP10	CV	0.154	0.531	0.730
CV	AV	0.094	0.551	0.996
CV	AV+CV	0.662	0.547	0.999

Discussion

The results showed that for some CVs a higher prediction accuracy was achieved when the genotype data included the CV whereas for others the surrounding SNPs were able to pick-up its effect. Knowing the CV (giving it a prior probability of 1 in the analysis) improved prediction accuracy for some CV, especially if the analysis had insufficient power to identify the CV. Generally, including all SNPs next to the CV in the analysis seemed to increase accuracy. Fitting the top-10 SNPs in a QTL region seemed generally to result in a reduced prediction accuracy compared to fitting all SNPs.

Albeit, for different reasons, the results indicate that for all three cases prediction accuracies would be improved by knowing the CVs: for the ABCG2-CV prediction accuracy increased because the variable selection analysis did not clearly identify this SNP as the CV; for the DGAT-CV the imputation accuracy was a limiting factor and knowing the importance of this CV would ensure that it would be amongst the set of genotyped SNPs; and without prior knowledge on the GHR-CV it may have been excluded from the analyses by any of the data quality control filters, but Table 1 shows that all analyses excluding the GHR-CV yield markedly reduced prediction accuracies. Thus, it seems important to identify the CVs, from a perspective of: (1) understanding the biology of phenotypes; (2) identifying targets for genome editing; and, as shown here, (3) to improve accuracies of genomic predictions.

Acknowledgements

Funding from Norwegian Research Council: 255297 & 309611.

References

Blott S., Kim J-J, Moisio S., Schmidt-Kuntsel A., Cornet A. *et al.* (2003) Genetics 163:253-66. https://doi.org/10.1093/genetics/163.1.253.

Cohen-Zinder M., Seroussi E., Larkin D.M., Loor J.J., Everts-van der Wind A. *et al.* (2005) Genome Res. 15:936-44. https://doi.org/10.1101/gr.3806705

Grisart B. Farnir F., Karim L., Cambisano N., Kim JJ. *et al.* (2004) Proc Natl Acad Sci 101:2398-403. https://doi.org/10.1073/pnas.0308518100

Hayes B.E. and Goddard M.E. (2001) Gen. Sel. Evol. 33 https://doi.org/10.1186/1297-9686-33-3-209.

Jenko J., Gorjanc G., Cleveland M.A., Varshney R.K., Whitelaw C.B. *et al.* (2015) Genet. Sel. Evol. 47-55. https://doi.org/10.1186/s12711-015-0135-3

Legarra A. and Reverter A. (2018) Gen. Sel. Evol. 50:53. https://doi.org/10.1186/s12711-018-0426-6

Meuwissen T:H.E., VandenBerg I., Goddard M.E. (2021) Gen. Sel. Evol. 53:19. https://doi.org/10.1186/s12711-021-00607-4

Visscher P.M. (2020) ICQG6, Brisbane, Australia, available at: icqg6.org/program/.

267. Fuzzy logic as a strategy for combining marker statistics to optimize preselection of high-density and sequence genotype data

A. Ling[1*], E.H. Hay[2], E. Hartono[3], S.E. Aggrey[3] and R. Rekaya[1]

[1]Department of Animal and Dairy Science, The University of Georgia, Athens, GA 30602, USA; [2]USDA Agricultural Research Service, Fort Keogh Livestock and Range Research Laboratory, Miles City, MT 59301, USA; [3]Department of Poultry Science, The University of Georgia, Athens, GA 30602, USA; asling@uga.edu

Abstract

Marker preselection based on criteria such as *P*-values, F_{ST} scores, and biological information has been proposed as a strategy to increase the functional relevance of preselected markers and reduce overparameterization and multicollinearity when using high-density and sequence genotype data in genomic evaluations. However, marker preselection based on a single criterion has largely failed to benefit predictions in both simulated and real data, and at worst has resulted in less accurate predictions when the trait of interest is lowly heritable with a complex genetic architecture (e.g. health and fertility traits). When multiple preselection criteria are available on markers, consolidating these criteria into a single composite score for preselection has the potential to increase the sensitivity to correctly detect relevant markers and benefit predictions. In this study, we investigate a fuzzy logic approach as a strategy for combining the criteria of *P*-values and F_{ST} scores for preselection of relevant markers in high-density (1.3 million biallelic markers) genomic data.

Introduction

The affordability and accessibility of high-density and sequence data has increased in recent years to allow practical use in genomic evaluations. It was initially expected that genomic prediction accuracy would benefit from such data by increasing capture of genetic variation (Meuwissen and Goddard, 2010), but practical results show little to no advantage relative to low- and moderate-density single nucleotide polymorphism (SNP) panels (Van Binsbergen *et al.*, 2015). This failure is likely due to a combination of excessive multicollinearity between SNP predictors and overparameterization of the association model.

SNP marker preselection may have the potential to improve the quality of genomic predictions by reducing multicollinearity and overparameterization and increasing the biological relevance of markers in the model. Several published studies evaluated single preselection criteria, including *P*-values (Veerkamp *et al.*, 2016), estimated marker effects (Ober *et al.*, 2012), F_{ST} scores (Chang *et al.*, 2018), and annotation missense status (Frischknecht *et al.*, 2018). In each of these studies, prediction accuracies outcompeted genome-wide SNP panels only when one or a few variants had a large contribution to the genetic variance of the trait of interest for both simulated and real data.

The limited success of single-criteria preselection to improve prediction accuracies is likely to be explained by the limited sensitivity of any single criteria to distinguish between relevant and non-relevant markers. All criteria will be subject to false positives and the impact of such markers on predictions is unlikely to be null. Ling *et al.* (2021) showed that while SNP preselection by estimated effect results in less false positives and more genetic variance explained than F_{ST} preselection, the false positives selected by estimated effect have a greater negative impact on predictions and can ultimately offset the benefit from capturing more genetic variance. Based on these observations, it is reasonable to hypothesize that a composite score combining these two preselection criteria may harness the benefits of both to ultimately yield higher accurate genomic predictions. Xiang *et al.* (2019) proposed a composite score for ranking variants from sequence data based on association and biological criteria and found that when preselected variants were combined with a

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_267, © A. Ling *et al.* 2022

genome-wide 630k SNP panel, differences in genomic prediction accuracies between high- and low-ranking variants were statistically significant for several economically important milk traits in Holstein.

Fuzzy logic models are a computationally efficient and flexible tool for decision making based on multiple criteria. They have been successfully implemented in a variety of engineering applications, such as self-driving vehicles and automatic rice cookers. In this simulation study, we evaluate the potential of a Sugeno fuzzy logic model to combine *P*-values and F_{ST} scores into a composite score for SNP preselection for use in genomic prediction.

Materials & methods

A simulation was conducted using QMSim (Sargolzaei and Schenkel, 2009). The simulated genome consisted of 29 chromosomes of 100 cM each, modeled after the autosomal cattle genome, and 1.3 million evenly-distributed SNP markers. 5,000 QTL were randomly distributed across 14 of 29 chromosomes. The simulated trait had a moderate heritability of 0.3. Training and validation populations consisted of 5k and 1k individuals from generations 1 to 5 and 6, respectively.

P-values were calculated as described in Aguilar *et al.* (2019) using the BLUPF90 suite of programs (Misztal *et al.*, 2018). SNP effects ($\hat{a}$) were calculated jointly from single-step GBLUP (ssGBLUP) solutions,

$$\hat{\boldsymbol{a}} = \frac{1}{2\sum_{i=1}^{m} p_i q_i} \boldsymbol{Z}'\boldsymbol{G}^{-1}\hat{\boldsymbol{u}}$$

where Z' is the transpose of the $n \times m$ matrix of genotypes, G^{-1} is the inverse of the genomic relationship matrix, $\hat{\boldsymbol{u}}$ is the vector of solutions of genotyped individuals from a ssGBLUP model

F_{ST} scores were calculated as in Chang *et al.* (2018). The two subpopulations were formed by taking 5% of individuals from each of the tails of the phenotype distribution, and then calculating as follows,

$$F_{ST} = \frac{H_T - H_S}{H_T}$$

where $H_T = 2 \cdot p_T \cdot q_T$, $H_S = \frac{H_{S1} \cdot n_{S1} + H_{S2} \cdot n_{S2}}{n_{S1} + n_{S2}}$ and $H_{Si} = 2 \cdot p_{Si} \cdot q_{Si}$,

where p_T and q_T are major and minor allele frequencies, respectively, in the total population; p_{Si} and q_{Si} are major and minor allele frequencies, respectively, in the i^{th} subpopulation; and n_{Si} is the number of individuals in the i^{th} subpopulation.

The Sugeno fuzzy logic model involves sorting the n inputs into c_i categories with fuzzy boundaries and incomplete membership in multiple classes as illustrated in Figure 1. Logic AND/OR rules are constructed around combinations of these categories across inputs. Given n inputs, there will be a maximum number of rules equal to $\prod_{i=1}^{n} c_i$. For each of m rules, a linear combination of the input criterion (x) is constructed, $y = b'x$, where b is a vector of parameters decided by the practitioner.

A given set of inputs is evaluated across all rules and then the contribution of each rule (y_j) to the final fuzzy output is weighted (w_j) by either the minimum (AND logic rule) or maximum (OR logic rule) of the class memberships across inputs, $y_{fin} = \frac{\sum_{j=1}^{m} w_j y_j}{\sum_{j=1}^{m} w_j}$.

Preselected marker sets were evaluated using a ssGBLUP model with BLUPF90 (Misztal *et al.* 2018) and genomic prediction quality measured by the Pearson correlation between true (u) and estimated ($\hat{u}$) breeding values.

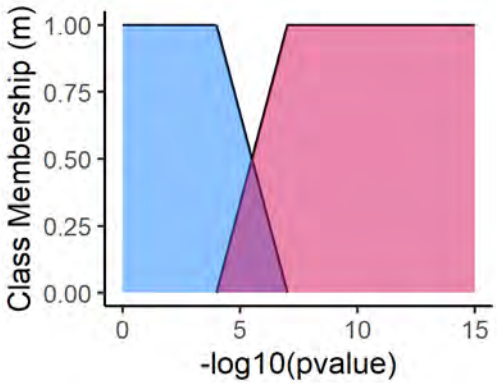


Figure 1. A fuzzy input criteria with 2 classes. The input value and class membership fall on the horizontal and vertical axes, respectively.

Results

For the present study, 2 criteria ($-\log_{10}$(*P*-value) and F_{ST} scores) with 3 classes each were implemented in a Sugeno fuzzy model. A heatmap of the fuzzy output is shown in Figure 2. The highest fuzzy score is yielded when both *P*-value and F_{ST} score take a high value. The lowest score is yielded when *P*-value is high and F_{ST} is low, illustrating the flexibility of the fuzzy approach.

For panels of 20k SNPs and smaller, fuzzy preselection yielded higher prediction accuracies than either $-\log_{10}$(*P*-value) or F_{ST} preselection individually, as shown in Figure 3. These gains were up to 6.3 and 1.9% higher than $-\log_{10}$(*P*-value) and F_{ST} preselection, respectively. For larger panels, both F_{ST} and fuzzy score preselection yielded similar accuracies while outperforming $-\log_{10}$(*P*-value) preselection by 1.5-2.2%.

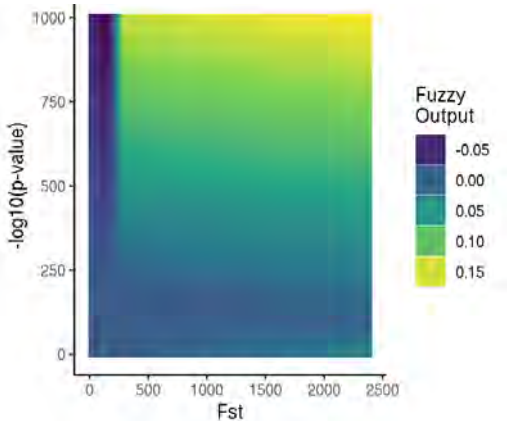


Figure 2. Heatmap of the Sugeno fuzzy model output for scaled values of $-\log_{10}$(*P*-value) and F_{ST} scores.

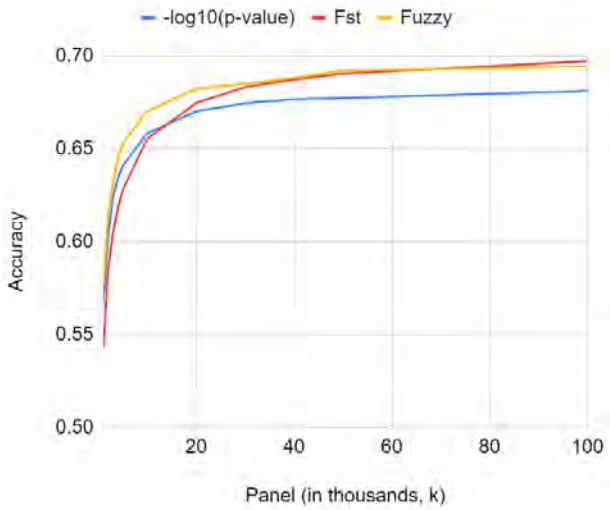


Figure 3. Trend of accuracy for -$\log_{10}$(*P*-value), F_{ST}, and fuzzy score preselection across panels of 1k to 100k SNPs.

Fuzzy preselection correctly distinguished more SNPs linked with QTL than -$\log_{10}$(*P*-value) or F_{ST} preselection for SNP panels of 40k and smaller as shown in Figure 4. Despite capturing more QTL-linked markers than fuzzy or F_{ST} score preselection for SNP panels of 50k and larger, -$\log_{10}$(*P*-value) preselection yielded lower accuracies than either of the other preselection criteria.

Discussion

The inferior accuracy of -$\log_{10}$(*P*-value) preselection relative to F_{ST} and fuzzy score preselection, despite capturing more QTL-linked markers, can be explained by the accumulation of SNP with large spurious associations. Estimated SNP effects and other statistics derived from it, including *P*-values, are highly

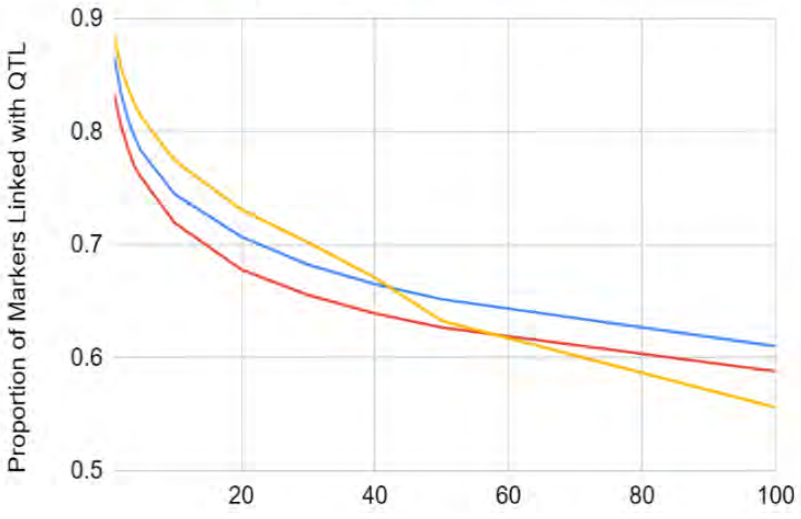


Figure 4. The proportion of preselected markers linked with QTL for -$\log_{10}$(*P*-value), F_{ST}, and fuzzy score preselection across panels of 1k to 100k SNPs.

sensitive to relevant markers, but also have a tendency to select markers with a large spurious association when a single dataset is used for both SNP discovery and model training, a phenomenon known as the Winner's Curse or Beavis effect (Xu, 2003). Ling *et al.* (2021) showed that accumulation of SNPs with large spurious associations could detract from the benefits of capturing more genetic variation and result in lower accuracies of genomic predictions.

Fuzzy score preselection yielded both higher prediction accuracies and sensitivity to detect truly relevant QTL for smaller preselected panels, and this may be due to the fuzzy model capturing the sensitivity of *P*-values to detect truly relevant QTL while identifying and excluding some of the most spuriously associated unlinked markers through contrast of *P*-values with F_{ST} scores. Ultimately, the strengths of both criteria are combined and their weaknesses balanced out.

Fuzzy preselection did not yield higher accuracies than F_{ST} scores for larger panels or outperform use of all markers. It should be noted we used only two criteria to calculate the fuzzy scores due to simulation constraints. In real data, additional prioritization criteria may be available, including prior information on variant-trait associations, gene expression data, and predicted mutation effect. Inclusion of additional criteria into the fuzzy approach has the potential to further increase capture of relevant markers and identification and exclusion of false positives, which may further improve genomic prediction accuracy.

References

Chang LY, Toghiani S, Ling A, Aggrey SE, Rekaya R. BMC Genet. 2018;19(1):4. https://doi.org/10.1186/s12863-017-0595-2.

Frischknecht M, Meuwissen THE, Bapst B, Seefried FR, Flury C, Garrick D, *et al.* J Dairy Sci. 2018;101(2): 1292–6. https://doi.org/10.3168/jds.2017-12890.

Ling A, Hay EH, Aggrey SE, Rekaya R. BMC Genomic Data (2021) 22:26. https://doi.org/10.1186/s12863-021-00979-y.

Meuwissen and Goddard. Genetics (2010) 185: 623–631. 10.1534/genetics.110.116590

Misztal, I., S. Tsuruta, D. A. L. Lourenco, Y. Masuda, I. Aguilar, *et al.* 2018. Manual for BLUPF90 family programs. University of Georgia. http://nce.ads.uga.edu/wiki/doku.php?id=documentation

Ober U, Ayroles JF, Stone EA, Richards S, Zhu D, *et al.* PLoS Genet. 2012;8(5):e1002685. https://doi.org/10.1371/journal.pgen.1002685.

van Binsbergen R, Calus MPL, Bink MCAM, van Eeuwijk FA, Schrooten C, *et al.* (2015). https://doi.org/10.1186/s12711-015-0149-x

Veerkamp RF, Bouwman AC, Schrooten C, Calus MP. Genet Sel Evol. 2016;48(1):95. https://doi.org/10.1186/s12711-016-0274-1.

Xiang R, van den Berg I., MacLeod IM, Hayes BJ, Prowse-Wilkins CP, *et al.* PNAS (2019) 116 (39) 19398-19408. https://doi.org/10.1073/pnas.1904159116

Xu S. 2003;165(4):2259–68. https://doi.org/10.1093/genetics/165.4.2259.

268. Genetic evaluations and genomic characteristics for local cattle using genome sequences, 50K and a specific SNP chip

K. May[1*], M.J. Wolf[1], G.B. Neumann[2], P. Korkuć[2], G.A. Brockmann[2] and S. König[1]

[1]Justus-Liebig University Gießen, Institute for Animal Breeding and Genetics, Ludwigstraße 21 b, 35390 Gießen, Germany; [2]Humboldt University zu Berlin, Albrecht Daniel Thaer Institute for Agricultural and Horticultural Sciences, Animal Breeding Biology and Molecular Genetics, Invalidenstraße 42, 10115 Berlin, Germany; katharina.may@agrar.uni-giessen.de

Abstract

This study compares the performance of the Illumina BovineSNP50 BeadChip (50K chip) with sequence data and a novel designed breed-augmented customized 200K SNP chip (200K chip) in the small local dual-purpose breed German Black Pied cattle (DSN). On such genomic databases, we applied genome-wide association analyses and estimated genetic parameters and breeding values for milk production, health indicator and calving traits. Accuracies of estimated breeding values were calculated using a cross-validation approach. We identified significant SNP marker associations and candidate genes with the specifically designed 200K chip, which were not detected with the commercial 50K chip. The annotated genes (e.g. *MGST1* for 305-day lactation milk yield) were confirmed based on whole-genome-sequence data, indicating breed-specific genomic mechanisms. Heritabilities and accuracies of genomic breeding values increased slightly by using the novel specifically designed 200K SNP chip compared with the 50K SNP chip.

Introduction

Marker density affects the power of genome-wide association studies (GWAS) and genomic estimated breeding value (GEBV) accuracy. The Illumina BovineSNP50 BeadChip with ~54.000 markers (hereafter referred to as 50K chip) is commonly used for GWAS and genetic evaluations in various cattle breeds worldwide. The 50K chip is designed on a reference panel of large commercial populations like Holstein Frisian [HF] and Simmental. The chip works particularly well in breeds closely related to these reference population (Matukumalli *et al.*, 2009). The German Black Pied cattle breed (DSN, German: Deutsches Schwarzbuntes Niederungsrind) is an endangered local dual-purpose breed with a small population size of ~2,500 cattle and considered as the founder breed of HF. However, only ~37,000 markers of the 50K chip are informative in DSN after filtering, which coincides with observations in other small cattle breeds (e.g. Hozé *et al.*, 2013). Moreover, the well described *DGAT1*gene with major effects on milk production traits in HF is one example for genes which are not associated with milk production traits in DSN in GWAS (Korkuć *et al.*, 2021), although DSN and HF are closely genetically related. Hence, we designed a customized 200K SNP chip for DSN including breed-specific variants (Neumann *et al.*, 2021). We hypothesize that a denser breed-specific SNP chip is helpful in GWAS to identify DSN-specific variants for a variety of traits. Wu *et al.* (2015) showed an increased power to detect significant associations with increasing marker density for mastitis in Danish Holsteins when comparing 50K, 777K and whole-genome sequence (WGS) data.

The objective of this study was to compare three marker densities (50K, 200K and WGS) in GWAS for milk production, health indicator and calving traits in DSN. Furthermore, we compared the accuracies of genomic breeding values obtained using different SNP chip arrays.

Materials & methods

Phenotypes. We considered first-parity DSN cows from six herds. In total, 2,020 records were available for 305-day milk yield (M305), fat yield (F305) and protein yield (P305), 1,638 records for test-day fat-to-

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_268, © K. May *et al.* 2022

protein ratio (FPR) and somatic cell score (SCS; log-transformed somatic cell count) as health indicators, and 2,606 records for calving ease (CE) and stillbirth incidence (SB). For FPR and SCS, we focused on the challenging early lactation period and included the first test-day record between 5 and 40 days in milk (DIM). The breed (DSN) percentage was larger than 90% for all cows according to an algorithm to clearly differentiate between DSN and HF (Jaeger *et al.*, 2018).

SNP chip design and imputation of genotypes. The customized DSN-augmented 200K SNP chip (Axiom® myDesign™ TG Array; hereafter referred as 200K chip) was designed as described by Neumann *et al.* (2021). Variants for the chip were selected from 20,587,181 sequence variants (SVs) detected from WGS of 304 DSN cattle. The chip contains DSN unique variants, variants associated with important traits of interest in DSN (e.g. disease resistance, milk yield, fertility), variants with low, high or moderate impact on the transcripts, DSN informative variants from the 50K chip as well as 175,537 SNPs and 8,618 indels corresponding to 103,801 haplotype blocks in DSN. In total, 300 DSN cattle were genotyped with the novel designed 200K chip and provided a reference panel for the imputation of 1,797 DSN 50K chip genotypes into 200K genotypes. The filtered WGS data included 16,175,216 SVs of 304 DSN and served as a reference panel for the imputation of 1,797 DSN 200K genotypes to WGS level. The imputation and filtering of SVs were performed in BEAGLE (Browning *et al.*, 2018). Quality control of the three marker densities (50K, 200K and WGS) was performed in PLINK (Purcell *et al.*, 2007). Depending on the trait category, the final number of markers after filtering were ~37,000 SNPs for 50K, ~125,000 for 200K and ~11,600,000 SVs for WGS.

Genome-wide associations and genetic parameter estimations. For the GWAS, we applied a single marker linear mixed model using the 'leaving one chromosome out' option in GCTA (Yang *et al.*, 2011) for all three marker densities (50K, 200K, WGS). The statistical model in matrix notation was:

$$y = X\beta + Zu + Ss + e \quad (1)$$

where y = vector of observations for M305, F305, P305, FPR, SCS, CE and SB; β = vector of fixed effects (herd-calving year-calving season and a linear regression on DIM and age at first calving for M305, F305 and P305; herd, test-day year-season and a linear regression on age at first calving for FPR; herd, test-day year-season, a linear regression on DIM and on fat percentage for SCS; herd-calving year-calving season and sex of calf for CE and SB); u = vector of polygenic effects with $u \sim N(0, G\sigma^2_u)$, with G denoting the genomic relationship matrix (VanRaden, 2008), and σ^2_u the polygenic variance; s = vector for marker effects; e = vector of random residuals; and X, Z, and S were incidence matrices for β, u, and s, respectively. We calculated an adjusted Bonferroni-corrected genome-wide significance threshold (*p*Bonf) with $P=0.05/n_{eff}$ with n_{eff} = effective number of independent SNPs/SVs. In addition, we considered a suggestive significance threshold *p*Sug with $p = 1/n_{eff}$. The -indep-pairwise option in PLINK (Purcell *et al.*, 2007) was used to calculate n_{eff} by excluding one SV of a SV pair in LD $r^2>0.5$ in a window size of 5000 SVs, which was shifted in an interval of 500 SVs. SNP effect correlations were calculated genome-wide and chromosome-wide for each trait for all overlapping SNPs/SVs between the three marker densities. Furthermore, model (1) was applied to estimate genetic parameters for all traits with all three marker densities using restricted maximum likelihood (REML) with the --reml function in GCTA.

Gene annotation. Potential candidate genes were queried and assigned to the associated SNPs/SVs using the current gene annotations from ENSEMBL (release 104). A gene was considered as a candidate gene if at least one SNP/SV with $P<p$Sug was located in the respective gene and/or within 150 kb up- and downstream. Physiological functions of candidate genes were studied in the ENSEMBL and KEGG databases.

Genomic breeding value estimation. Breeding values were estimated in two consecutive runs: (1) 50K chip vs 200K chip using 1,900 DSN; and (2) 200K chip vs 10 random selected 200K out of WGS data (random200K) using 1,980 DSN. Phenotypic data for M305 and F305 were pre-corrected for fixed effects according to model (1) in R. Afterwards, we applied an animal model for genomic breeding value estimations using pre-corrected phenotypes (residuals) in BLUPF90 (Misztal *et al.*, 2002). The statistical model was:

$$y = \mu + g + e \quad (2)$$

where y = vector for residuals for M305 and F305; μ = overall mean; g = vector of additive genetic effects with $\sim N(0, G\sigma^2_g)$, with G denoting the genomic relationship matrix (VanRaden, 2008), and σ^2_g the genomic variance; e = residual effects with $\sim N(0, \sigma^2_e)$, and σ^2_e residual variance. For cross-validation, we divided the dataset in a validation set (20% of cows) and training set (80% of cows). We selected the cows randomly in a procedure with 10 replicates using replacements. The accuracy was calculated for each of the 10 replicates as the correlation between the direct GEBV and residuals. Afterwards, the mean accuracy was calculated based on the 10 replicates for M305 and F305 for 50K, 200K and random200K.

Results

The numbers of associated SNPs/SVs and genes for the three marker densities per trait are given in Table 1. For M305, F305 and P305, we identified 23 SNPs with 50K, 84 SNPs with 200K and 950 SVs with WGS with $P<p$Bonf/pSug. For M305, 14.8% (4/27) of genes (*KLF3*, *LIMCH1, NSUN7, RALGAPA2*) overlapped for 50K and 200K. The genes with the largest number of SNPs/SVs located in a gene for M305 were *MGST1* (BTA5), *ADGRL3* (BTA6) and *GNAL* (BTA24), exclusively detected with 200K and WGS, but not with 50K (Figure 1).

For SCS, we identified 62 SVs with WGS but no SNP reached pSug for the 50K and 200K chips. For the calving traits CE and SB, we identified 11 genes with 50K and 23 genes with 200K. Among them, seven common genes were annotated with both SNP chips. Genome-wide and chromosome-wide SNP effect correlations between 50K and 200K were ≥0.99 (P≤0.001) for all traits, and ranged from 0.62 to 0.81 (P≤0.001) between WGS with 50K and 200K. The estimated heritabilities were quite similar for the three marker densities (Table 1).

Table 1. Number of associated markers (SNPs/SVs) and identified genes (G) in the GWAS and heritabilities (h^2) with standard error (SE) for the three marker densities.[1]

	50K		200K		WGS	
Trait	**No. SNPs/G**	**h^2**	**No. SNPs/G**	**h^2**	**No. SVs/G**	**h^2**
M305	11/11	0.40 (0.04)	64/20	0.39 (0.04)	724/48	0.41 (0.04)
F305	8/10	0.33 (0.04)	11/8	0.33 (0.04)	115/20	0.34 (0.04)
P305	4/6	0.34 (0.04)	9/7	0.33 (0.04)	111/11	0.35 (0.04)
FPR	4/4	0.13 (0.04)	10/6	0.14 (0.04)	52/12	0.13 (0.04)
SCS	0/0	0.11 (0.04)	0/0	0.12 (0.03)	62/2	0.13 (0.04)
CE	0/0	0.03 (0.02)	1/1	0.03 (0.02)	5/2	0.04 (0.02)
SB	1/1	0.04 (0.02)	2/1	0.05 (0.02)	10/8	0.05 (0.02)

[1] M305/F305/P305 = 305-day lactation yield for milk, fat and protein; FPR = fat-to-protein ratio; SCS = somatic cell score; CE = calving ease; SB = stillbirth.

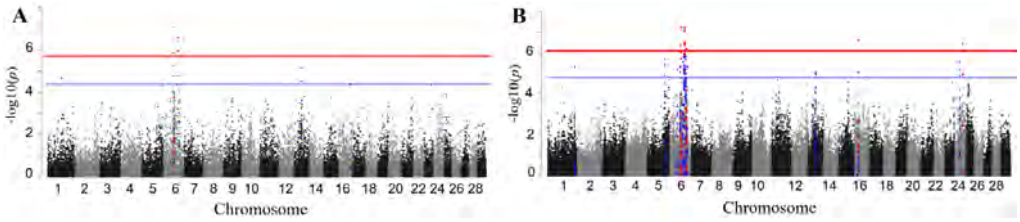


Figure 1. Manhattan-plots for 305-day lactation milk yield (M305); (A) 50K Illumina BeadChip and (B) novel designed DSN 200K SNP chip; markers above *p*Bonf (red line) are highlighted in red; markers above *p*Sug (blue line) are highlighted in blue; markers in a distance of 125 kb up- and downstream of associated markers are highlighted, too.

For GEBV run (1) the mean accuracy was 0.46 for M305 and 0.36 for F305 with the 50K chip, and 0.47 and 0.37 with the 200K chip, respectively. When comparing the 200K chip with random200K in run (2), mean accuracy was 0.46 for M305 and 0.34 for F305 with the 200K chip, and ranged from 0.446 to 0.450 for M305 and from 0.320 to 0.336 for the 10 random200K.

Discussion

Similar heritabilities and GEBV accuracies for different marker densities and traits showed a limited advantage when using the customized 200K chip compared to the 50K chip in genetic evaluations for the endangered DSN breed. Similarly, Erbe *et al.* (2012) indicated a small increase in genomic prediction accuracies of 0.01 in HF and 0.03 in Jerseys when using imputed high-density SNP panels compared to the Illumina 50K. However, since DSN and HF are closely genetically related, a breed-specific SNP chip might be more advantageous for small breeds that are distantly related with the large commercial cattle populations HF, Simmental and Brown Swiss. In agreement with the results by Wu *et al.* (2015), the power to detect candidate genes was improved when using the specifically designed 200K chip and WGS data compared to 50K. This might be due to stronger linkage disequilibrium between markers and quantitative trait nucleotides (QTN) for higher marker densities. SNP effect correlations were close to one between 50K and 200K, but moderate to high between WGS with both SNP chips. Hence, WGS is suitable to detect breed-specific variants affecting traits of interest in small-sized endangered cattle breeds, which can be integrated into genomic prediction models to improve prediction accuracies.

References

Browning, B.L. (2018) Am. J. Hum. Genet. 103:338-348. doi.org/10.1016/j.ajhg.2018.07.015
Erbe M. *et al.* (2012) J. Dairy Sci. 95:4114-4129. https://doi.org/10.3168/jds.2011-5019
Hozé C. *et al.* (2013) Sel. Evol. 45:33. https://doi.org/10.1186/1297-9686-45-33
Jaeger M., Scheper C., König S., Brügemann K. (2018) Züchtungskunde 90, pp. 262-279.
Korkuć P. *et al.* (2021) Front. Genet. 12:275. https://doi.org/10.3389/fgene.2021.640039
Matukumalli L.K. *et al.* (2009) PloS ONE 4:e5350. doi.org/10.1371/journal.pone.0005350
Misztal I., Tsuruta S., Strabel T. *et al.* (2002) Proc. of the 7th WCGALP, Montpellier, France.
Neumann G.B. *et al.* (2021) BMC Genomics 22:905. doi.org/10.1186/s12864-021-08237-2
Purcell, S. *et al.* (2007) Am. J. Hum. Genet. 81:559-575. https://doi.org/10.1086/519795
VanRaden, P.M. (2008) J. Dairy Sci. 91:4414-4423. https://doi.org/10.3168/jds.2007-0980
Wu X. (2015) Genet. Sel. Evol. 47:50. https://doi.org/10.1186/s12711-015-0129-1
Yang J. *et al.* (2011) Am. J. Hum. Genet. 88:76-82. https://doi.org/10.1016/j.ajhg.2010.11.011

269. Genomic prediction using whole-genome sequence data in intensely selected pig lines

R. Ros-Freixedes[1,2*], M. Johnsson[1,3], A. Whalen[1], C.Y. Chen[4], B.D. Valente[4], W.O. Herring[4], G. Gorjanc[1] and J.M. Hickey[1]

[1]The Roslin Institute and Royal (Dick) School of Veterinary Studies, The University of Edinburgh, Easter Bush, Midlothian, Scotland, United Kingdom; [2]Departament de Ciència Animal, Universitat de Lleida – Agrotecnio-CERCA Center, Lleida, Spain; [3]Department of Animal Breeding and Genetics, Swedish University of Agricultural Sciences, Uppsala, Sweden; [4]The Pig Improvement Company, Genus plc, Hendersonville, TN, USA; roger.ros@udl.cat

Abstract

We used a large pig dataset to assess the utility of whole-genome sequence data for genomic prediction. We imputed genotypes of 32.8 million variants for 396,100 individuals from seven commercial pig lines (17,224 to 104,661 per line). We used BayesR to perform genomic prediction for 8 real traits and 9 simulated traits with different genetic architectures. Sequence data improved prediction accuracy relative to the marker array, provided that the training set was sufficiently large and especially for traits with high heritability and low number of quantitative trait nucleotides. The most robust results were obtained when variants with statistically significant associations to the trait were preselected and added to the marker array, which yielded average improvements of prediction accuracy of 2.5 and 4.2 percentage points in within-line and multi-line scenarios, respectively, with training sets of around 80k individuals. More robust improvements could be achieved with larger training sets and optimised pipelines.

Introduction

Whole-genome sequence data (WGS) could improve genomic prediction accuracy and its persistence across generations and breeds because it is assumed to contain the causal variants. Early simulations indicated that using causal mutations from such data could increase prediction accuracy by up to 30% if causal variants with low minor allele frequency were captured (Druet *et al.*, 2014). Results with real data have been mixed so far, comprising either no relevant improvement of prediction accuracy compared to marker array data or small, and often non-robust, improvements (e.g. Zhang *et al.*, 2018; Song *et al.*, 2019). One of the strategies to exploit WGS that has been more successful to date consists in augmenting the available marker arrays with preselected variants from WGS based on their association with the trait of interest. In some cases, this strategy improved prediction accuracy by up to 9% (Al Kalaldeh *et al.*, 2019) or 11% (Lopez *et al.*, 2021). Thus, large data sets that capture most of the genome diversity in a population must be assembled so that the allele substitution effects are estimated with high accuracy for millions of variants. We generated WGS for 396,100 pigs from seven different intensely selected lines from diverse genetic backgrounds and with different numerical size. The objectives of this study were to use this dataset to assess the utility of WGS for genomic prediction compared to commercial marker array data and to identify scenarios in which WGS may become beneficial.

Materials & methods

Populations and sequencing strategy. We performed whole-genome re-sequencing of 6,931 individuals from seven commercial pig lines (Genus PIC, Hendersonville, TN) with a median individual coverage of 1.5×. Approximately 1.5% (0.9 to 2.1% in each line) of the pigs in each line were sequenced. Most pigs in each line were also genotyped with commercial marker arrays (GeneSeek, Lincoln, NE, USA).

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_269, © R. Ros-Freixedes *et al.* 2022

Genotype imputation. Genotypes were jointly called, phased, and imputed for a total of 483,353 pedigree-related individuals using the 'hybrid peeling' method implemented in AlphaPeel (Whalen *et al.*, 2018), which used all marker array and WGS that was available. Imputation was performed separately for each line using its complete multi-generational pedigree, which encompassed from 21,129 to 122,753 individuals. Individuals with low predicted imputation accuracy were removed before further analyses according to Ros-Freixedes *et al.* (2020). A total of 396,100 individuals remained, from 17,224 to 104,661 individuals for each line, with an average predicted imputation accuracy (individual-wise dosage correlation) of 0.97 (median: 0.98). We also excluded from the analyses variants with a minor allele frequency lower than 0.023, because their average predicted imputation accuracy (variant-wise dosage correlation) was lower than 0.90. After imputation, 32.8 million variants (14.5 to 19.9 million within each line) remained for downstream analyses, of which 9.9 million segregated across all seven lines.

Traits. We analysed data of 8 traits: average daily gain (ADG), backfat thickness (BFT), loin depth (LD), average daily feed intake (ADFI), feed conversion ratio (FCR), total number of piglets born, litter weight at weaning, and return to oestrus 7 days after weaning. Most pigs with records were born during the 2008-2020 period. Deregressed breeding values (dEBV) were used. To assist in the interpretation of results, we also created 9 simulated traits with different numbers of quantitative trait nucleotides (QTN; 100, 1000 or 10,000 QTN; sampled randomly across the called variants) and heritability levels (h^2; 0.10, 0.25 or 0.50).

Genomic prediction. For each line, we defined a testing set as the full-sib families from the last generation of the pedigree. The training set was defined as all those individuals that had a pedigree coefficient of relationship lower than 0.5 with any individual of the testing set. This design was chosen to mimic a realistic situation in which breeding companies evaluate the selection candidates available in the selection nucleus at the given time. We tested genomic prediction using variants from the marker array (referred to as 'Chip'; ~40k variants) or for sets of preselected variants from the WGS of a similar size: (1) LDTags: tag variants retained after pruning based on linkage disequilibrium; (2) Top40k: 40k variants preselected based on genome-wide single-marker regression analyses in the training sets (variants with the lowest *P*-value in each consecutive non-overlapping 55-kb windows along the genome); (3) ChipPlusSign: significant variants with ($P \leq 10^{-6}$; only the lowest *P*-value within each 55-kb window; average per trait: 309, range: 23 to 1,083) were preselected and added to Chip; (4) Functional: variants that were annotated as loss-of-function or missense; and (5) Rand40k: 40k variants chosen randomly. Genomic prediction was performed by fitting a univariate mixed model with BayesR (Erbe *et al.*, 2012). Prediction accuracy was calculated in the testing set as the correlation between the genomic estimated breeding value and the dEBV. Additionally, we considered multi-line (ML) training sets that were conformed by merging the training sets of each line and by performing all analyses adding the line as an effect.

Results

The performance of each set of predictor variants was not robust and differed for each trait and line, sometimes leading to no improvements or losses of prediction accuracy. The most robust results were obtained for ChipPlusSign (Figure 1). The size of the training set was one of the main factors that determined the capacity of variants from the WGS to improve prediction accuracy compared to Chip. The simulation results suggested that genetic architecture of the traits also conditioned the utility of WGS for improving prediction accuracy compared to Chip, as greater improvements were achieved for traits with high heritability and low number of QTN (Figure 2). However, for empirical polygenic traits, average improvements of prediction accuracy of up to 2.5 percentage points were observed with training sets of around 80k individuals. In the most successful scenarios, at least around 200 significant variants were added to Chip. With multi-line training sets, average improvements of prediction accuracy with ML-ChipPlusSign were of up to 4.2 percentage points (not shown). Preselecting an entirely new set of predictor variants from

WGS in Top40k proved more challenging than ChipPlusSign and did not perform much differently from just taking random variants from these windows, as in Rand40k. Selection based on functional variants and, more severely, LDTags reduced prediction accuracy in many cases.

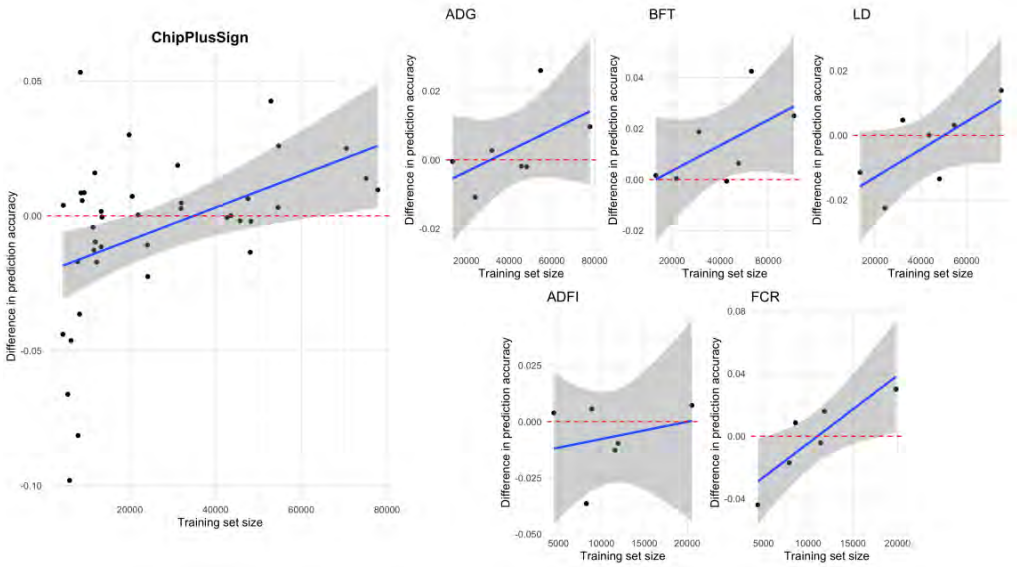


Figure 1. Genomic prediction accuracy with ChipPlusSign for the empirical traits. The difference of prediction accuracy between ChipPlusSign and Chip is shown, for all traits and lines (left) or by trait (right).

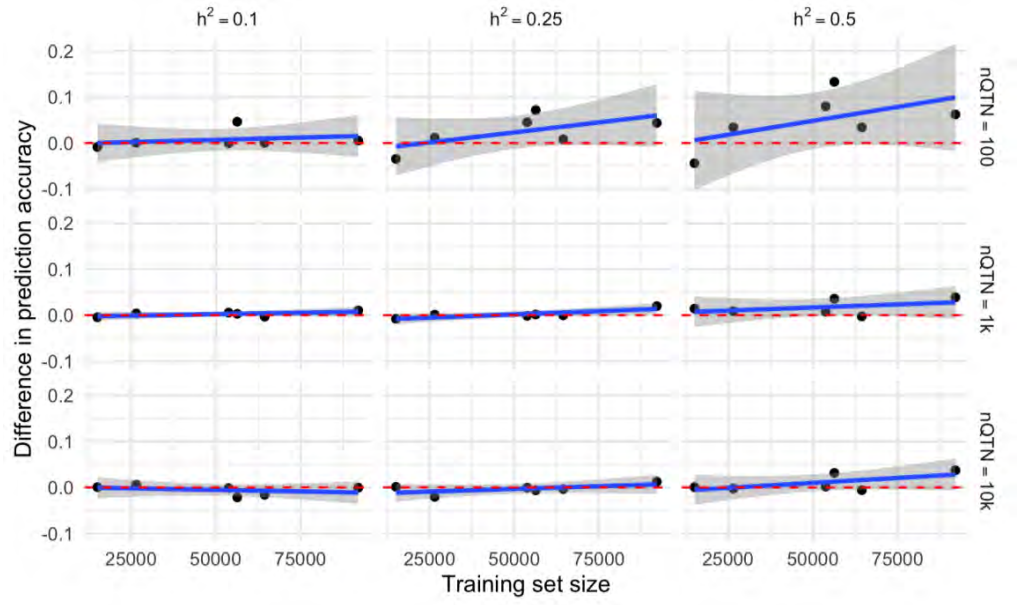


Figure 2. Genomic prediction accuracy with ChipPlusSign for the simulated traits. The difference of prediction accuracy between ChipPlusSign and Chip is shown.

Discussion

Our results evidenced the potential for WGS to improve genomic prediction accuracy in intensely selected pig lines, provided that the training sets are large enough. Improvements achieved so far were modest. On one hand, these limited improvements indicated that the strategies that we tested were likely suboptimal. On the other hand, the positive trend for the largest training sets indicated that we might have not reached the critical mass of data that is needed to leverage the potential of WGS, especially in scenarios where genomic prediction with marker arrays is already yielding high accuracy.

The modest performance of ChipPlusSign is also a consequence of the difficulty for fine-mapping causal variants through genome-wide association studies with WGS. Even though WGS allows the detection of a very large number of associations, problems such as false positives or *P*-value inflation also become more severe in a way that added noise might offset the added information. Therefore, WGS performed better with simple genetic architectures (i.e. traits with low number of QTN). This is consistent with expectations and simulation results (Clark *et al.*, 2011).

Multi-line training sets could be particularly beneficial for the use of WGS because they allow a larger training set with low relationship degree between the individuals. Previous simulations suggested that WGS might be the most beneficial with multi-breed reference panels (Iheshiulor *et al.*, 2016), especially for numerically small populations.

References

Al Kalaldeh M., Gibson J., Duijvesteijn N., Daetwyler H.D., MacLeod I., *et al.* (2019) Genet Sel Evol 51:32. 10.1186/s12711-019-0476-4

Clark S.A., Hickey J.M., and van der Werf J.H. (2011) Genet Sel Evol 43:18. 10.1186/1297-9686-43-18

Druet T., MacLeod I.M., and Hayes B.J. (2014) Heredity 112:39-47. 10.1038/hdy.2013.13

Erbe M., Hayes B.J., Matukumalli L.K., Goswami S., Bowman P.J., *et al.* (2012). J Dairy Sci 95:4114-29. 10.3168/jds.2011-5019

Iheshiulor O.O.M., Woolliams J.A., Yu X., Wellmann R., and Meuwissen T.H.E. (2016). Genet Sel Evol 48:15. 10.1186/s12711-016-0193-1

Lopez B.I.M., An N., Srikanth K., Lee S., Oh J.D., *et al.* (2021). Front Genet 11:603822. 10.3389/fgene.2020.603822

Ros-Freixedes R., Whalen A., Chen C.Y., Gorjanc G., Herring W.O., *et al.* (2020) Genet Sel Evol 52:17. 10.1186/s12711-020-00536-8

Song H., Ye S., Jiang Y, Zhang Z., Zhang Q., *et al.* (2019) Genet Sel Evol 51:58. 10.1186/s12711-019-0500-8

Whalen A., Ros-Freixedes R., Wilson D.L., Gorjanc G. and Hickey J.M. (2018) Genet Sel Evol 50:67. 10.1186/s12711-018-0438-2

Zhang C., Kemp R.A., Stothard P., Wang Z., Boddicker N., *et al.* (2018). Genet Sel Evol 50:14. 10.1186/s12711-018-0387-9

270. Single and multi-breed ssGBLUP using preselected variants from whole-genome sequence data in pigs

S. Jang[1*], R. Ros-Freixedes[2,3], J.M. Hickey[3], C.Y. Chen[4], W. Herring[4], I. Misztal[1] and D. Lourenco[1]

[1]University of Georgia, Edgar L. Rhodes Center for Animal and Dairy Science, 30602, Athens, USA; [2]Departament de Ciència Animal, Universitat de Lleida – Agrotecnio-CERCA Center, Lleida, Spain; [3]The Roslin Institute and Royal (Dick) School of Veterinary Studies, The University of Edinburgh, Easter Bush, Scotland, United Kingdom; [4]Genus PIC, 100 Bluegrass Commons Blvd, Suite 2200, Hendersonville, TN, USA; jsbng@uga.edu

Abstract

Whole-genome sequence data harbor causative variants that may not be present in regular SNP chips. We investigated the impact of using preselected variants from sequence data in single-breed (SB) and multi-breed (MB) ssGBLUP. For SB, three terminal pig breeds were evaluated for five traits. These three terminal breeds were combined for the MB analyses. The number of animals sequenced or imputed to sequence ranged from 42k to 105k, depending on the breed. We examined three genotype sets, namely, regular SNP chip, top markers in each 40k genomic window, and significant SNP based on GWAS combined with the chip. Maximum gains of 0.03 and 0.01 in accuracy were observed for SB and MB with selected sequence variants. The benefits of using sequence data for SB and MB predictions are limited even when information on many sequenced or imputed animals is available.

Introduction

Whole-genome sequence (WGS) data harbor millions of variants, possibly including causative variants that primarily affect the traits of interest. As WGS data are becoming available for many farm animal species, there is a question on whether this information can help increase the accuracy of genomic predictions (GP) beyond that already achieved with SNP chips. Certainly, sequencing the entire population is still unfeasible; therefore, the most efficient approach is to sequence important animals and impute the sequence to other genotyped animals (Ros-Freixedes *et al.*, 2020). Not all variants might be causative or in high linkage-disequilibrium (LD) with the causative ones; thus, using the entire WGS data would not benefit GP (Van Binsbergen *et al.*, 2015). Hence, genome-wide association studies (GWAS) can help identify the variants statistically associated with traits of interest or the variants with large effect sizes. The use of selected sequence variants showed marginal or no increase in prediction accuracy in several populations (VanRaden *et al.*, 2017; Fragomeni *et al.*, 2019); however, the number of individuals initially sequenced was usually small, and the imputation applied to less than 30k animals. Having larger reference populations with WGS could help increase GP accuracy, as well as combining breeds with smaller reference sizes (Song *et al.*, 2019). In this study, we investigated the impact of using preselected variants from imputed WGS (IWGS) on up to 105k animals in SB and almost 207k in MB genomic evaluations.

Materials & methods

Data. Datasets were provided by Genus PIC (Hendersonville, TN). We investigated average daily feed intake (ADFI), average daily gain (ADG), backfat (BF), ADG recorded in crossbred animals (ADGX), and BF recorded in crossbred animals (BFX) in three terminal pig breeds (B1, B2, and B3). A two-trait model was considered for ADFI and ADG (ADFI model), whereas a four-trait model was used for ADG, BF, ADGX, and BFX (GROWTH model). The total number of records and animals in the pedigree are in Table 1. The number of genotyped animals for B1, B2, B3, and MB was 60,474, 41,573, 104,661, and 206,684, respectively. Different SNP panels were used for SB and MB evaluations, which were termed Chip1 and Chip2, respectively. Chip1 was based on the GGP-Porcine HD BeadChip (GeneSeek, Lincoln, NE) and

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_270, © S. Jang *et al.* 2022

Table 1. Number of records and animals in the pedigree.

Breed	ADFI	ADG	BF	ADGX	BFX	Number of animals in pedigree
B1	35k	0.36M	0.34M	150k	149k	1.13M
B2	40k	0.30M	0.30M	158k	156k	0.84M
B3	64k	0.94M	0.86M	299k	247k	3.14M
MB	140k	1.60M	1.50M	578k	525k	5.28M

Chip2 was a MB imputed panel. After quality control, 35,786, 40,311, and 43,032 SNP remained for B1, B2, and B3 in SB, whereas 41,303 SNP remained in MB. For a fair comparison of SB and MB, we also used Chip2 for SB (Chip2SB), where 37,909, 42,897, 44,022 SNP remained for B1, B2, and B3.

Imputation. Genotypes were jointly called, phased, and imputed using the 'hybrid peeling' method implemented in AlphaPeel (Whalen *et al.*, 2018). This method used all marker array data available across the pedigrees and sequence data of 731, 760, and 1,856 pigs in B1, B2, and B3, respectively. Imputation was performed separately for each breed. Individuals with low predicted imputation accuracy were removed as described by Ros-Freixedes *et al.* (2020). A total of 60,474, 41,573, and 104,661 individuals with IWGS remained in each breed. These individuals were expected to have an average dosage correlation of 0.97 based on the imputation accuracy of 284 pigs that had both WGS (high coverage) and marker array data. Variants with a minor allele frequency lower than 0.023 were removed because their estimated dosage correlations were lower than 0.90 (Ros-Freixedes *et al.*, 2020).

Preselection of variants. Two preselected variant sets (Top40k and TopSign) were created from the IWGS data. Single-trait GWAS was carried out to identify the variants through Fast-LMM software (Lippert *et al.*, 2011). A total of 53,760, 37,147, and 92,012 animals were used for GWAS in B1, B2, and B3, respectively. For the MB, all animals were used in the GWAS, and a fixed breed effect was included in the model to account for breed differences. Top40k was constructed by selecting top markers in each 40k genomic window of IWGS. TopSign only included significant variants and was integrated into the chip data (ChipSign) for genomic predictions. We combined the selected variants for each trait in the ADFI and GROWTH models because the variants were prospected from single-trait GWAS. After quality control, from 78,714 to 80,899 SNP were used in Top40k and 36,305 to 44,734 in ChipSign for SB, depending on the breed and model. In the MB analyses, 51,297 and 49,738 SNP were retained in Top40k and 62,906 and 59,756 SNP in ChipSign, depending on the model.

Genomic predictions. Variance components were estimated for each breed separately in the SB scenario without using genomic information. Likewise, variance components from the MB scenario were estimated using MB data. All fixed effects in the model were made breed-specific in the MB analyses. For genomic predictions, we used ssGBLUP as implemented in the BLUPF90 family of programs (Misztal *et al.*, 2014). Predictive ability was calculated as the correlation between deregressed EBV (DEBV) and GEBV for the validation animals. The number of validation animals varied from 3,720 to 11,308, depending on the breed and model.

Results

Predictive abilities of both SB and MB evaluations are presented in Table 2. Using preselected variants for SB improved predictive ability up to 0.01 for ChipSign and 0.03 for Top40k compared to Chip1. However, when Top40k was used for MB, there was even a decrease of -0.05 (ADFI in B2) in predictive ability compared to Chip2. ChipSign showed a maximum of 0.01 gain and a -0.01 decrease. Comparisons between Top40k for SB and MB showed greater predictive ability when the top markers in each genomic window of

Table 2. Predictive ability for single-breed and multi-breed genomic predictions.

		SB			MB			
Breed	**Trait**[1]	**Chip1**	**Top40k**	**ChipSign**	**Chip2SB**	**Chip2**	**Top40k**	**ChipSign**
B1	ADFI	0.36	0.34	0.36	0.34	0.36	0.34	0.36
	ADG	0.49	0.45	0.49	0.46	0.46	0.43	0.47
	BF	0.60	0.59	0.61	0.60	0.59	0.57	0.60
	ADGX	0.33	0.33	0.34	0.31	0.30	0.29	0.30
	BFX	0.62	0.63	0.62	0.64	0.64	0.61	0.64
B2	ADFI	0.29	0.30	0.30	0.30	0.29	0.24	0.30
	ADG	0.48	0.50	0.49	0.47	0.45	0.44	0.46
	BF	0.60	0.60	0.61	0.58	0.56	0.56	0.57
	ADGX	0.18	0.21	0.18	0.22	0.21	0.19	0.20
	BFX	0.48	0.47	0.49	0.45	0.42	0.40	0.42
B3	ADFI	0.40	0.39	0.40	0.40	0.39	0.39	0.40
	ADG	0.50	0.51	0.51	0.50	0.49	0.50	0.50
	BF	0.59	0.60	0.60	0.59	0.59	0.59	0.59
	ADGX	0.26	0.28	0.27	0.27	0.26	0.27	0.27
	BFX	0.59	0.59	0.59	0.56	0.56	0.56	0.57

[1] ADFI = average daily feed intake; ADG = average daily gain; BF = backfat thickness; ADGX = average daily gain crossbreds; BFX: backfat thickness crossbreds.

40k SNP were selected based on SB GWAS. Differences in predictive ability became larger in breeds with less genotyped or sequenced/imputed animals. Moreover, Chip2SB outperformed Chip2 in most of the cases, indicating no benefits of MB genomic prediction.

Discussions

Our study is the first to report the predictive ability of ssGBLUP for large pig populations with IWGS data. Most of the studies have used less than 30k sequenced/imputed animals. We used up to 105k for SB and up to 207k animals for MB evaluations. Current results showed that the advantages of using preselected variants from IWGS highly depend on the trait, breed, and model (SB or MB) used for GWAS. For example, ADGX of B2 and B3 showed some improvement in predictive ability (16 and 8%) in SB when using Top40k, but for the same trait, B1 reported no gains. In contrast, predictive ability for MB was the same or decreased with Top40k, compared to Chip2, possibly meaning that those variants preselected based on MB GWAS are not as informative as the ones identified in the SB GWAS. This raises a question about the amount of information that has been used for selecting variants. Looking at the dimensionality of the genomic information can help to assess the number of genotyped animals needed to maximize the percentage of discoveries in GWAS (Jang *et al.*, 2021). According to Jang *et al.* (2021), populations with smaller effective size (Ne=20) require more data to capture causative variants, whereas for large populations (Ne=200), using the number of genotyped animals equal to that of the largest eigenvalues explaining 98% of the variance of the genomic relationship matrix suffices. Still, only a small proportion of the causative variants can be discovered if those genotyped animals do not have lots of progeny records.

Regarding MB evaluations, breeds with fewer genotyped animals (B1 and B2) observed a decrease in predictive ability when the models involved the top markers in each genomic window of 40k SNP. This might be because those Top40k SNP selected from MB GWAS could be related to causative variants in the largest breed (B3) but not in the small ones (B1 and B2). Larger breeds can easily overwhelm GWAS and GP

results when combined with smaller breeds. Using preselected variants from IWGS slightly improved the prediction accuracy when selected from SB GWAS and used for SB predictions. However, no benefits were observed in MB. Overall, the benefits of using sequence information for SB and MB genomic predictions may still be limited even when information on many sequenced or imputed animals is available.

References

Fragomeni B., Lourenco D., Legarra A., VanRaden P., and Misztal I. (2019) JDS 102(11):10012-10019. https://doi.org/10.3168/jds.2019-16262

Jang S., Tsuruta S., Leite N., Misztal I., and Lourenco D. (2021) Proc. of 72nd the EAAP, Davos, Switzerland

Lippert, C., Listgarten J., Liu Y., Kadie C.M., and Davidson R.I., *et al.* (2011). Nature methods 8(10):833-835. https://doi.org/10.1038/nmeth.1681

Misztal I., Tsuruta S., Lourenco D., Aguilar I., and Legarra A., *et al.* (2014) Manual for BLUPF90 family of programs. Athens: University of Georgia, USA

Ros-Freixedes R., Whalen A., Chen C.Y., Gorjanc G., and Herring W.O., *et al.* (2020). GSE 52:1-15. https://doi.org/10.1186/s12711-020-00536-8

Song H., Ye S., Jiang Y., Zhang Z., and Zhang Q., *et al.* (2019.) GSE 51(1):58. https://doi.org/10.1186/s12711-019-0500-8

Van Binsbergen R., Calus M.P., Bink M.C., Van Eeuwijk F.A., and Schrooten C., *et al.* (2015) GSE 47(1):71. https://doi.org/10.1186/s12711-015-0149-x

VanRaden P.M., Tooker M.E., O'connell J.R., Cole J.B., and Bickhart D.M. (2017) GSE 49(1):32. https://doi.org/10.1186/s12711-017-0307-4

Whalen A., Ros-Freixedes R., Wilson D.L., Gorjanc G., and Hickey J.M. (2018) GSE 50(1):1-15. https://doi.org/10.1186/s12711-018-0438-2

271. Application of low-pass sequencing to genomic prediction of egg quality in laying hens

A. Wolc[1,2*], J. Li[3], D. Lubritz[2], J. Arango[2], J. Fulton[2], P. Settar[2], K. Rowland[2] and H. Cheng[3]

[1]Department of Animal Science, Iowa State University, Ames, IA 50011, USA; [2]Hy-Line International, Dallas Center, IA 50063, USA; [3]Department of Animal Science, University of California, Davis, CA 95616, USA; awolc@iastate.edu

Abstract

Accuracy of genomic prediction was compared in two layer lines using medium density SNP chip and low-pass sequencing. Three thousand individuals representing two lines were sequenced for the project. Thirty-four genomic regions were identified as genome-wide significant across 11 egg quality traits and their genotypes were added to the chip. Multi-trait single-step animal model was used for genomic prediction. A slight increase in accuracy for some of the traits was obtained by adding the significant SNPs to those present on the chip. Fitting all sequence SNPs had no advantage for predictive ability over chip despite explaining 15% more of phenotypic variance in the Genome Wide Association Study. Further optimization of the utilization of sequence data and cost reduction are required to make low-pass sequence competitive to medium density SNP chips for genomic prediction.

Introduction

Genomic selection has been applied to poultry breeding for a decade providing benefits of increased accuracy of estimated breeding values and shortening of generation intervals. With the reduction in cost, implementation evolved from a combination of low- and high-density panels to population specific medium density panels (50-60,000 SNPs). With the development in sequencing technology the costs of obtaining whole genome sequences were greatly reduced. Low pass sequencing interrogates all genomic locations in contrast to chip genotyping which only considers those SNPs included on the chip. Low pass sequencing comes at a lower cost than standard sequencing but also at a lower accuracy of genotype calls than standard sequencing. Rubinacci *et al.* (2021) notes that the accuracy of genotype calls from low pass sequencing can be recovered by combining the genotype calling with imputation similar to the methodology used to combine information from low- and high-density chips. If relevant reference populations are used accurate genotype calls can be obtained with 1× sequences.

Some studies suggest that low-pass sequencing can improve genome wide association studies relative to chip data (Li *et al.*, 2021). Until now no studies using large scale sequencing were reported in the literature for laying chickens, therefore we investigated the impact of using whole genome sequence data and adding SNPs identified as significant in a Genome Wide Association Study (GWAS) to the 54k chip for genomic prediction.

Materials & methods

Data. One hundred males from 8 major lines were selected for sequencing at 4× coverage and combined with 124 samples from publicly available sequences to provide reference for genotype calling for low-pass (1×) sequenced individuals. Genotypes for 32 million SNPs were called. Vcf files were provided by Gencove with SNPs called using their proprietary algorithm. One generation of breeding males and two generations of breeding females were sequenced at 1× coverage for a total of 3,025 sequenced individuals, including 1,540 individuals from one White Leghorn (WL) line and 1,485 individuals from one Rhode Island Red line (RIR). Genotypes of males were used in single-step genomic prediction (they are sires of hens with phenotypes) but not in GWAS (they don't have own phenotypes). After quality control with minor allele

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_271, © A. Wolc *et al.* 2022

frequency >1%, 8.6 million markers were left in RIR and 7.6 million markers in WL line. Phenotypes are available for all of the sequenced females with more than 30,000 and 40,000 additional phenotyped relatives for RIR and WL line, respectively. Phenotypes used for the analyses include several egg quality traits such as shell strength, egg weight, yolk weight, albumen height and shell colour measured at different ages.

Validation of SNP calls. Accuracy of 1× genotype calls was validated by comparison to Axiom 54k SNP chip genotypes on the same individuals using absolute value of correlation for genotypes called as 0,1,2 and percentage of identical calls, and as percentage of identical calls for animals sequenced in duplicates.

GWAS. GWAS was performed using a mixed model in Genome Wide Complex Trait Analysis software (GCTA; Yang *et al.*, 2011) for WL and RIR lines separately. The following statistical model was used in the GWAS:

$$y_i = \mu + b\, x_i + g_i + e_i$$

where y_i is the phenotype adjusted by station and hatch effect for individual i; μ is the overall mean; b is the additive effect (covariate) of the candidate SNP to be tested for association; x_i is the SNP genotype indicator of the candidate SNP for individual i coded as 0, 1 or 2; g_i is the random polygenic effect for individual i based on the relationship matrix computed from genotypes and e_i is the residual. No obvious population stratification was observed based on the principle component analysis for both lines, thus it was not considered in the GWAS. The SNPs with *P*-value<0.00001 were added to the SNP panel for genomic prediction.

Genomic prediction. 212 females from WL line and 228 from RIR line had phenotypes removed when conducting the GWAS and genomic prediction for the purpose of validation. The data were analysed per line with AIREMLF90 software using chip data, sequence data or chip combined with significant SNPs from the sequence for the construction of H matrix.

A series of multi-trait animal models fitting fixed effects of hatch within generation, station processing the eggs, age of birds and random effects of animal and permanent environment were applied for genomic prediction. Variance components were estimated with pedigree and only BLUP was performed using H matrix in order to reduce computational requirements.

Accuracy was measured as correlation between the phenotype adjusted for fixed effects and predicted genomic breeding value divided by square root of heritability.

Results

Accuracy of genotyping with low-pass sequencing. Sequence was evaluated on 386 samples by comparing those genotypes to respective allele calls from 54k Axiom SNP Chip (currently a gold standard). Median correlation of the genotypes was 0.97 (SNPs with lower MAF tended to have lower correlation) with 98.4% calls being identical. For 96 test samples sequenced in duplicates, concordance of genotype calls around 98% was obtained compared to >99.5% for duplicates on chip.

Genomic prediction. Based on mixed model in GCTA software using only hens that had both phenotypic and sequence records sequence data explained on average 15% more variance than SNP chip. Across the egg quality traits 43 genomic regions were identified as significant at genome wide level.

In the first step, genomic prediction was performed using SNP chip genotypes. Based on the GWAS results, 7,063 and 7,216 SNPs were added to those present on the chip to conduct another set of genomic prediction. Slight improvement in accuracy was achieved for some of the traits (Table 1). The prediction accuracy based on 54k chip data was close to that based on low-pass sequencing, despite many more SNPs considered with sequencing data.

Discussion

Previous studies (Rubinachi *et al.* 2021, Kim *et al.* 2021, Li *et al.* 2021) used subsampling of high coverage sequence data to evaluate accuracy of low pass sequencing and showed that with exception of low frequency variants accurate calls can be obtained when genotype calling is combined with imputation. We used real samples that were genotyped and sequenced in duplicates under different IDs to validate low-pas sequencing and showed that the quality of information from low-pass sequence is slightly lower than from SNP chip but acceptable for GWAS and genomic prediction.

Mixed results were obtained by using sequence data for genomic prediction with several studies reporting very limited benefits (VanRaden *et al.*, 2016; Heidaritabar *et al.*, 2016). Yang *et al.* (2010) showed that the genetic variance does not depend on the number of SNPs when correcting for incomplete LD between SNPs, which explains why the genomic prediction based on SNP chip and sequencing data were similar.

Some increases in accuracy were observed when SNPs were weighted differentially based on their annotation (McLeod *et al.*, 2016) or smaller panel was designed by combining sequence based GWAS with annotation (Xiang *et al.*, 2021). Therefore, the next step will be to employ differential weighting of SNPs and other computational strategies.

Conclusions

Similar or slightly higher accuracy can be obtained in chicken using low pass (1×) sequence data compared to a medium density SNP chip however further optimization is needed to extract more information from the sequence data.

Table 1. Accuracy of GEBVs using SNP chip data (54k), SNP chip data enhanced with significant SNPs from sequence (54kPlusSig) or complete sequence (Seq) in two layer lines.[1]

	WL line			RIR line		
	54k	54kPlusSig	Seq	54k	54kPlusSig	Seq
AH1	0.77	0.77	0.69	0.56	0.55	0.53
AH2	0.53	0.51	0.48	0.69	0.68	0.70
BS1	0.27	0.29	0.35	0.48	0.41	0.48
BSP01	0.34	0.38	0.37	0.59	0.55	0.59
BSP02	0.46	0.53	0.47	0.59	0.56	0.58
C01	0.67	0.75	0.62	0.62	0.60	0.62
E3	0.65	0.65	0.65	0.50	0.47	0.43
EW1	0.45	0.48	0.45	0.31	0.31	0.30
EW2	0.47	0.48	0.50	0.25	0.24	0.24
YW1	0.51	0.52	0.45	0.3	0.31	0.34
YW2	0.53	0.54	0.52	0.35	0.34	0.32

[1] AH = albumen height; BS = breaking strength; C0 = shell color; E3,EW = egg weight; YW = yolk weight. Numbers 1 and 2 refer to data collected at early (around 26wks) and late (around 75wks) age of birds.

References

Heidaritabar M., Calus M.P., Megens H.J., Vereijken A., Groenen M.A., *et al.* (2016) J Anim Breed Genet. 133(3):167-79. https://doi.org/10.1111/jbg.12199

Kim, S., Shin, J. Y., Kwon, N. J., Kim, C. U., Kim, C., *et al.* (2021). Human genomics, 15(1), 58. https://doi.org/10.1186/s40246-021-00357-w

Li J. H., Mazur C. A., Berisa T., and Pickrell, J. K. (2021) Genome Res., 31(4), 529–537. https://doi.org/10.1101/gr.266486.120

MacLeod I.M., Bowman P.J., Vander Jagt C.J., Haile-Mariam M., Kemper K. E. *et al.* (2016) BMC Genomics. 17:144 https://doi.org/10.1186/s12864-016-2443-6

Rubinacci S., Ribeiro D.M., Hofmeister R. J., and Delaneau, O. (2021) Nature Genet., 53(1), 120–126. https://doi.org/10.1038/s41588-020-00756-0

VanRaden P.M., Tooker M.E., O'Connell J.R., Cole J.B., and Bickhart D.M. (2017). Genet Sel Evol., 49(1), 32. https://doi.org/10.1186/s12711-017-0307-4

Xiang R., MacLeod I.M., Daetwyler H.D. *et al.* (2021) Genome-wide fine-mapping identifies pleiotropic and functional variants that predict many traits across global cattle populations. Nat Commun 12:860. https://doi.org/10.1038/s41467-021-21001-0

Yang J., Benyamin B., McEvoy B. P., Gordon S., Henders A. K. *et al.* (2010). Nature Genet. 42(7), 565–569. https://doi.org/10.1038/ng.608

Part 17

Challenges – improving genomic prediction

272. Accuracy of prediction for a genomic evaluation in rotational crossbreeding scheme (Montbéliarde × Holstein × Red Danish)

P. Croiseau[1*], R. Saintilan[1,2], A. Baur[1,2], I. Croue[1,2], V. Ducrocq[1], J.R. Thomasen[3], E. Karaman[4], D. Boichard[1], H. Leclerc[1,2] and B.C.D. Cuyabano[1]

[1]Université Paris Saclay, INRAE, AgroParisTech, GABI, Domaine de Vilvert, 78350 Jouy-en-Josas, France; [2]ALLICE, Paris, France; [3]Viking Genetics, Ebeltoftvej 16, 8960 Randers SO, Denmark; [4]Center for Quantitative Genetics and Genomics, Faculty of Technical Sciences, Aarhus University, 8830 Tjele, Denmark; pascal.croiseau@inrae.fr

Abstract

In recent years, crossbreeding strategy has grown in dairy cattle farms at an international level. Breeders are interested in keeping crossbred cows in their herd both to combine the strengths of the pure breeds, compensate for their weaknesses and benefit from heterosis. However genetic tools are still lacking to manage these crossbred animals. In this study, we evaluate the performances of a genomic evaluation adapted for rotational crossbreeding schemes with real data. This genomic evaluation was applied to a population that includes pure-breed animals from Holstein, Montbéliarde, and Red Danish breeds, as well as crossbreds between these three breeds. The genomic evaluation approach was based on the estimation of breed specific SNP specific according to the Breed of Origin of the Alleles.

Introduction

Crossbreeding strategy in dairy cattle breeding schemes has significantly grown over the past years, and this trend may be reinforced in the future to face challenges linked to agroecology or to new constraints induced by climate change. Indeed, this strategy is particularly suitable to combine the strengths of the pure-breeds and to compensate for their weaknesses, being an efficient way for breeders to obtain more adaptable and robust animals, resulting in a more sustainable breeding system. Crossbreeding is also relevant to decrease inbreeding or, equivalently, generate heterosis.

Genomic evaluation in pure breed is common, it is developing in terminal crossbreeding, but only a few countries routinely evaluate crossbred animals in continuous crossbreeding programs. For most countries, breeders engaged in a crossbreeding process have information limited to purebred bulls and the raw performances of their crossbred females to manage their selection and mating. In this context, the literature dedicated to continuous crossbreeding relates, for the most part, to simulated data (Van Raden *et al.*, 2021; Eiriksson *et al.*, 2021; Karaman *et al.*, 2021) and rarely to applications on real data (Sevillano *et al.*, 2017).

In this study, we propose an application of genomic evaluation for continuous crossbreeding programs, through a SNP-BLUP based on Breed Origin of Alleles (BOA). This approach estimates specific SNP effect depending on its breed of origin. The rationale for this approach is that linkage disequilibrium between SNP and the causal mutation differs according to the breed; additionally, QTL effects may be breed dependent.

The data used in this analysis consisted of the first steps of a three-way crossbreeding scheme including Montbéliarde (Mo), Holstein (Ho) and Red Danish (RD) breeds. This data set was augmented by purebred data to ensure accurate estimation of within breed SNP effects. The procedure included several steps: imputation and phasing at the 50K level of all the animals; identification of BOA of each allele of crossbred animals; genomic evaluation on five production traits, including a cross-validation procedure. Accuracy of prediction and slope of regression were calculated.

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_272, © P. Croiseau *et al.* 2022

Materials & methods

Genotypes. The data used in this study originated from France and Nordic Countries (NAV). They consisted of 5,238 genotypes of crossbred animals, and 20,000, 22,265, and 6,866 genotypes of pure Mo, Ho, and RD animals, respectively. 53,498 autosomal SNP markers were retained from the Illumina 50K chips used routinely in France for genomic selection. Imputation was carried out into two steps. In the first step, genotypes of purebreds were imputed with FImpute (Sargolzaei *et al.*, 2014) with the pipeline used in the routine French national evaluation system and in the second step, these imputed genotypes were treated as known genotypes. In the second step, the crossbred genotypes were imputed with FImpute software using the purebred Ho, Mo, and RD as a reference, and without using any pedigree information.

BOA of alleles. The frequency (f) of each haplotype of n consecutive SNPs was estimated by counting within each pure breed. BOA i was selected if:

$$\frac{f_i}{\sum_{j=1}^{nbreed} f_j}$$

was higher than a given threshold (0.9). The whole genome was scanned by a sliding window moving one SNP at a time. The initial value of n was set to 16 (leading to 65,536 theoretical haplotypic combinations). When a haplotype origin remained undetermined, the process was iterated after n was divided by 2. Undetermined origins surrounded by identical breed origins were assigned to this breed. When n=1, the small proportion of finally remaining undetermined origins were allocated according to allelic frequencies. This procedure was implemented in the in-house BreedOrigin fortran software and applied to a target population of 5,238 crossbred animals. The results were compared to expected breed proportions based on pedigree.

Crossbreeding parameters. Fractions of heterosis (H) and recombination losses (R) were calculated for each crossbred animal from proportions of breed origins of their parents based on pedigree. Values of the H and R coefficients were calculated as in Dechow *et al.* (2007) using the following formulae:

$$H = 1 - \sum_{i=1}^{nb_{breed}} s_i d_i \text{ and } R = 1 - \sum_{i=1}^{nb_breed} \frac{(s_i^2 + d_i^2)}{2}$$

where s_i and d_i are the proportions of sire and dam genes from breed i, respectively. Heterosis and recombination losses were then estimated by regressing on H and R.

Phenotypes. The genomic evaluation was a two-step procedure. In a first step, a polygenic model was used to estimate all non-genetic effects (herd-year, age-year and year-month of calving), heterosis and recombination losses (as regression coefficients) and heterogeneous variances, as in Dezetter *et al.* (2015). Unknown parent groups accounted for breed of origin, and breed composition of crossbreds was accounted for by the pedigree. Yield deviations (YD) produced by this model correspond to performances adjusted for fixed and non-genetic random effects. Five production traits were analysed: lactation milk yield, protein yield, protein content, fat yield, and fat content. Two batches of evaluations were performed, the first one accounting for heterosis and recombination losses, and the second one ignoring them.

Crossbreed genomic evaluation accounting for BOA. To account for crossbred animals in a genomic evaluation model, we proposed an extension of the SNP-BLUP model where a SNP effect β_i is estimated for each breed as described in the following model:

$$y_i = \sum_{b=1}^{nb_breed} (p_{i,b}\mu_b) + \sum_{b=1}^{nb_breed} \left(\sum_{j=1}^{nb_SNP} (\beta_{i,j,b} X_{i,j,b}) \right) + e_i \quad (1)$$

where y_i is the YD of the animal i, μ is a vector of means defined within each breed, $p_{i,b}$ is the proportion of breed b in the genome of individual i, estimated with the BOA approach. $X_{i,j,b}$ is the allele content of SNP j that originates from the breed b for animal i, centered for the allelic frequency of the SNP in breed b:

$$X_{i,j,b} = \left(k_{i,j,b} - n_{i,j,b} f_{j,b}\right) \quad (2)$$

where $k_{i,j,b}$ and $n_{i,j,b}$ are the number of '2' alleles and the total number of alleles of the SNP j that originates from breed b for the animal i, respectively; $f_{j,b}$ is the frequency of allele '2' of the SNP j in breed b.

To assess the performance of the genomic evaluation, the data set was divided into a training dataset with both genotypes and phenotypic records used to estimate SNP effects according to their BOA, and a validation data set for which predicted breeding values were computed using the effects obtained with the training dataset and then compared to the observed phenotypes. The validation population consisted of 2,000 crossbred cows without progeny.

Results

Breed of origin of alleles. Results for breed composition based on pedigree information and BOA methodology are presented in Table 1. For genotyped crossbred animals, 94% of the alleles were assigned to a breed. 48% of markers were from Ho origin, 34% from RD origin and 13.6% from Mo origin. In comparison, based on pedigree information, the corresponding origins were 56.3, 29.8 and 11.8% respectively. Correlation between breed compositions for both methodologies ranged from 0.95 for the RD breed to 0.99 for the Mo breed.

Across breed genomic evaluation accounting for BOA. Correlations between YD and genomic breeding values and associated slope of regression are presented in Table 2. On the training population, all the correlations are around 0.80 for lactation milk yield, fat yield, and protein yield and around 0.90 for fat content and protein content. Adjusting for Heterosis weakly affected these correlations. The associated slopes of regression were slightly higher than 1 for all the traits. In the validation population, without

Table 1. Results of determination of Breed of Origin based on genomic or Pedigree information.

Breed of Origin	Breed composition with BOA		Breed composition based on pedigree		Correlation
	Mean	Sd	Mean	Sd	
Montbéliarde (Mo)	13.60%	20.00%	11.80%	19.20%	0.99
Holstein (Ho)	48.00%	19.40%	56.30%	20.20%	0.97
Red Danish (RD)	34.00%	17.80%	29.80%	17.80%	0.95

Table 2. Correlations and slopes of regression of the YD, adjusted or not for Heterosis for 5 production traits, on the estimated genomic breeding values, in the training and validation population.

Traits	Performance without heterosis effect				Performance adjusted for heterosis effect			
	Training population		Validation population		Training population		Validation population	
	Corr	Slope	Corr	Slope	Corr	Slope	Corr	Slope
Milk yield	0.85	1.11	0.38	1.62	0.84	1.12	0.39	1.66
Fat yield	0.81	1.16	0.37	1.58	0.80	1.19	0.38	1.64
Protein yield	0.80	1.18	0.34	1.63	0.78	1.20	0.35	1.69
Fat content	0.93	1.11	0.63	1.47	0.93	1.11	0.63	1.47
Protein content	0.94	1.08	0.59	1.50	1.00	1.08	0.59	1.50

integration of the Heterosis effect, correlations ranged from 0.34 for protein yield to 0.63 for fat content trait. The gain of correlation obtained after adjusting for Heterosis was marginal (+1 point for lactation milk yield, fat yield, and protein yield). Regarding the slopes of regression, they were overestimated both with and without adjustment of the performances for Heterosis (between 1.47 and 1.69).

Discussion

In this study, we propose an application on real data of a genomic evaluation in a rotational crossbreeding scheme based on a SNP-BLUP model accounting for BOA.

This approach required to impute and phase genotyping data of crossbred animals in order to predict BOAs for these animals. These steps were tested and an average error rate of about 1.5% for the predicted BOA was observed (data not shown). To complete this information, we compare in this study the breed composition of crossbred animals measured using the pedigree information of the animals and estimated from the genotyped animals. The very strong correlations obtained between these two approaches are completely consistent with previous tests (Table 1).

Finally, the genomic evaluations carried out have shown an honourable accuracy of prediction of around 0.30 for traits with a heritability of 0.3 and around 0.60 for traits with a heritability of 0.50 (Table 2). The adjustment of the performances for Heterosis did not impact the accuracy of the prediction of the breeding values.

Acknowledgements

This project has received funding from APIS GENE and from the European Union's Horizon 2020 research and innovation program 'GenTORE' under grant agreement no. 727213.

References

Christensen O.F., Madsen P., Nielsen B., & Su G. (2014) GSE 46(1):1-9. https://doi.org/10.1186/1297-9686-46-23.

Dechow C.D., Rogers G.W., Cooper J.B., Phelps M.I., & Mosholder A.L. (2007) JDS 90(7):3542-3549. https://doi.org/10.3168/jds.2006-889

Dezetter C., Leclerc H., Mattalia S., Barbat A. Boichard D. *et al.* (2015) JDS 98(7):4904-4913. https://doi.org/10.3168/jds.2014-8386

Karaman E., Su G., Croue I., and Lund M.S. (2021) GSE 53(1):1-15. https://doi.org/10.1186/s12711-021-00637-y

Sargolzaei M., Chesnais J.P., and Schenkel, F.S. (2014). BMC Genomics 15(1):1-12. https://doi.org/10.1186/1471-2164-15-478

Sevillano C.A., Vandenplas J., Bastiaansen J.W., Bergsma R., and Calus M.P. (2017) GSE 49(1):1-14. https://doi.org/10.1186/s12711-017-0350-1

VanRaden P.M., Tooker M.E., Chud T.C.S., Norman H.D., Megonigal Jr *et al.* (2020) JDS 103(2):1620-1631. https://doi.org/10.3168/jds.2019-16634

273. Genomic evaluation for two-way crossbred performance in cattle

Q. Mei[1,2*], H. Liu[3], T. Xiang[1] and O.F. Christensen[2]

[1]Huazhong Agricultural University, No. 1 Shizishan Avenue, Hongshan District, 430070 Wuhan, P.R. China; [2]Aarhus University, Blichers Alle' 20, 8830 Tjele, Denmark; [3]SEGES, Agro Food Park 15, 8200 Aarhus N, Denmark; quanshun@qgg.au.dk

Abstract

Dairy cattle production systems are mostly based on using purebreds, but recently crossbreeding has received increased interest. For genetic evaluation including crossbreds, ideally, models accounting for breed specific effects should perform better compared with simple models. A novel single-step best linear unbiased prediction method provides an opportunity to combine purebred and crossbred information when breed specific effects are accounted for. In this study, we applied this single-step method to evaluate two-way crossbred performance for average daily gain in crossbred beef on dairy. Results showed that the single-step method, which accounted for breed specific effects, had a higher predictive ability. We conclude that this method is useful in the genetic evaluation for crossbred performance in cattle production system.

Introduction

Crossbreeding is common practice in many livestock production systems, in particular for pigs and poultry. For dairy and beef cattle, the production systems are mostly based on using purebreds, but recently crossbreeding between dairy breed cows and beef breed bulls has received increased interest for a number of reasons (Berry 2021). In particular, the use of x-sorted dairy semen and herd specific strategies for the necessary number of replacement heifers has provided the option that some dairy cows in a herd can be inseminated with beef semen, and furthermore, meat production from crosses between dairy cows and beef bulls has lower environmental footprint than meat production from beef cattle.

Genomic selection (Meuwissen *et al.* 2001) is another potentially important factor to make the cattle meat production more efficient and with reduced environmental impact. Genomic evaluation in cattle has primarily been done within pure breeds, but ideally for crossbreeding with dairy cows, genomic breeding values on beef bulls should be estimated for such crossbred performance based on phenotypic records on crossbreds. Genomic models that split genomic effect of crossbreds into breed specific terms and account for the breed of origin of alleles (BOA) should in principle perform better than simpler models that do not take BOA into account, but this has not always been seen (Sevillano *et al.* 2017).

Commonly in genomic evaluation, for many reasons, not all relevant individuals are genotyped. For this purpose, so-called single-step genomic best linear unbiased prediction (ssGBLUP) was introduced (Legarra *et al.* 2014). Xiang *et al.* (2016) investigated an extension of ssGBLUP for crossbred performance in a terminal two-breed crossbreeding system for pigs, and concluded that models using marker genotypes of crossbreds improved accuracies of prediction for crossbred performances (Xiang *et al.* 2016). The approach investigated in that paper was based on BOA and so-called breed specific partial relationship matrices that combine genomic and pedigree-based relationships (Christensen *et al.* 2014).

The objective of this paper is to investigate predictive performance of a ssGBLUP model based on BOA compared to the usual ssGBLUP model that did not account for BOA for crossbred beef on dairy cattle.

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_273, © Q. Mei *et al.* 2022

Materials & methods

Data. All data sets were provided by SEGES cattle. In this study, 2,329 two-way crossbred calves with purebred Belgian Blue beef (BBL) sires and purebred Holstein dairy (HOL) dams were on test for 1 month. These calves had an average age of 210±36 days in the beginning of test, and 245±32 days in the end of test. During this period, feed intake was recorded for each animal, and body weight of each animal was recorded at both start and end of test period. Average daily gain (ADG) of each animal within this period was calculated as the increase in body weight divided by number of days between the two recordings. After data editing, 1,749 crossbred calves with ADG were retained.

Pedigree for crossbred animals was traced back three generations, and there were 13,612 animals in the pedigree. Among these animals, 36 BBL and 138 HOL were genotyped with either EuroG 10k Bead Chip or EuroG MD Bead Chip, and 1,137 crossbreds were genotyped with EuroG MD Bead Chip. In purebred animals, imputation from EuroG 10k Bead Chip to EuroG MD Bead Chip was done with software Beagle 5.2 (Browning and Browning 2007). Quality controls of genomic data with software Plink were as follows: it was first confirmed that no individuals had call-rate less than 90%; following that, SNPs with call-rate less than 90% were removed; SNPs with a minor allele frequency less than 0.05 were removed; SNPs deviating strongly from Hardy Weinberg equilibrium within breed (P<10-7) were also removed (Purcell *et al.* 2007). After quality control, 1,311 individuals and 45,288 SNPs were retained. The retained genotype data was phased with software Beagle 5.2 (Browning and Browning 2007).

Allele tracing. The detection of BOA in crossbred animals is necessary for constructing breed specific partial relationship matrix. In this study, for detecting BOA we used the recently developed AllOr algorithm (Eiríksson *et al.* 2021), which is based on the comparison of haplotypes in overlapping windows.

Statistical Model. Two ssGBLUP models for genomic prediction were evaluated. The first model is the usual ssGBLUP model (M_1):

$$y= Xb+Zu+e \quad (1)$$

where y is a vector of phenotypic records for ADG; Xb refers to the fixed effects including herd-year-season, sex, and covariate of body weight at start test; u is a vector of random additive genetic effects for each animal, and Z is the corresponding incidence matrix; e is a vector of random residual effects. The random effects follow normal distributions of $u\sim N(0,H\sigma_a^2)$, and $e\sim N(0,I\sigma_e^2)$, where H is the combined pedigree-based and marker-based relationship matrix, which can be constructed as previously reported (Legarra *et al.* 2014); I is the identity matrix; σ_a^2 is the additive genetic variance; σ_e^2 is the residual variance.

The second model is the partial ssGBLUP model (M_2):

$$y= Xb+Z_Au_A+Z_Bu_B+e \quad (2)$$

where y, Xb, and e are as in Equation 1; u_A is a vector of random additive genetic effects from BBL, u_B is a vector of random additive genetic effects from HOL; Z_A and Z_B are the corresponding incidence matrices. The random effects follow normal distributions of $u_A\sim N(0,H_A\sigma_A^2)$, and $u_B\sim N(0,H_B\sigma_B^2)$, where H_A and H_B are combined pedigree-based and marker-based breed specific partial relationship matrices, which can be constructed as previously reported (Christensen *et al.* 2014); σ_A^2 and σ_B^2 were additive genetic variances for crossbred performances of BBL and HOL, respectively. The above analyses were carried out using restricted maximum likelihood (REML) algorithm and best linear unbiased prediction (BLUP) in the software DMU (Madsen and Jensen 2013).

In order to reveal the differences in predictive abilities between models M_1 and M_2, data were split into a training dataset and a validation dataset by a cut-off date of October 20, 2020. The numbers of animals in the training and validation datasets were 1,399 and 350, respectively. Predictive abilities were calculated as the correlation between predicted total genetic values (g) and the corrected phenotypes (Y_c) in the validation dataset. Vector Y_c was calculated based on model M_1 with full dataset. In model M_1, g was equal to the additive genetic effects (g= u); in model M_2, g was equal to the sum of additive genetic effects from BBL and HOL (g= u_{A+} u_B). Furthermore, dispersion bias of genomic predictions for each model was assessed as the regression coefficient of Y_c on g, with an expected result of 1.

Results

Variance component estimation. Estimates of variance components and heritabilities are shown in Table 1. In model M_2, the additive genetic variances from BBL (σ_A^2) and HOL (σ_B^2) were equal to 0.011 and 0.013, and the additive genetic variance of crossbred animals was equal to 0.012 (half of the sum of σ_A^2 and σ_B^2), which was a little bit higher than in model M_1. The heritability in models M_1 and M_2 was 0.105 and 0.122, respectively.

Predictive abilities. Predictive abilities of models M_1 and M_2 are shown in Table 2. The Pearson correlation coefficients between the corrected phenotypes and total genetic values of crossbred animals in model M_2 was 0.059, being a little bit higher than in model M_1 where it was 0.052. In terms of regression coefficients of corrected phenotypes on the total genetic values, model M_1 deviated more from 1 compared with model M_2.

Discussion

In this study, the estimated heritability of ADG was low to moderate, which was a little bit lower than previously reported (Freetly *et al.* 2020). One potential reason could be that ADG in this study was calculated in a shorter time period compared to previous studies. Furthermore, Ahlberg *et al.* (2018) pointed out that during different periods, the correlations for each shortened test duration was different. In addition, the smaller dataset could also be another possible reason.

Results of predictive abilities indicated that partial ssGBLUP model accounting for breed-specific effects performs better than the usual ssGBLUP model in terms of accuracy and dispersion bias. This phenomenon

Table 1. Estimates of variance components and heritabilities in different models.[1]

Model	σ_a^2	σ_A^2	σ_B^2	σ_e^2	h_a^2	h_{AB}^{2*}
M_1	0.010			0.088	0.105	
M_2		0.011	0.013	0.086		0.122

[1] h_{AB}^2 = heritabilities of crossbred animals ($0.5(\sigma_A^2+\sigma_B^2)/[0.5(\sigma_A^2+\sigma_B^2)+\sigma_e^2]$).

Table 2. Predictive abilities for crossbred animals in the validation dataset in different models.

Model	$r(Y_c, g)$[1]	$b(Y_c, g)$[2]
M_1	0.052	0.733
M_2	0.059	0.774

[1] $r(Y_c, g)$ = Pearson correlation coefficients between Y_c and g.

[2] $b(Y_c, g)$ = Regression coefficients of Y_c on g.

has also been observed in other species (Xiang *et al.* 2016). The result presented here only focused on the prediction of total genetic values in crossbreds. In future study, we plan to investigate accuracy of estimated breeding value of BBL sires for crossbred performance when mated to a Holstein cow.

References

Ahlberg, C.M., K. Allwardt, A. Broocks, K. Bruno, L. McPhillips, *et al.* (2018) Journal of Animal Science 96(8):3043-3054. https://doi.org/10.1093/jas/sky209

Berry, D. (2021) Journal of dairy science 104(4):3789-3819. https://doi.org/10.3168/jds.2020-19519

Browning, S.R. and B.L. Browning (2007) The American Journal of Human Genetics 81(5):1084-1097. https://doi.org/10.1086/521987

Christensen, O.F., P. Madsen, B. Nielsen and G. Su (2014) Genetics Selection Evolution 46(1):1-9. https://doi.org/10.1186/1297-9686-46-23

Eiríksson, J.H., E. Karaman, G. Su and O.F. Christensen (2021) Genetics Selection Evolution 53(1):1-13. https://doi.org/10.1186/s12711-021-00678-3

Freetly, H.C., L.A. Kuehn, R.M. Thallman and W.M. Snelling (2020) Journal of Animal Science 98(1):skz394. https://doi.org/10.1093/jas/skz394

Legarra, A.,O.F. Christensen, I. Aguilar and I. Misztal (2014) Livestock Science 166:54-65. https://doi.org/10.1016/j.livsci.2014.04.029

Madsen, P. and J. Jensen (2013) Center for Quantitative Genetics and Genomics, Dept of Molecular Biology and Genetics, University of Aarhus, Research Centre Foulum, Box 50:8830.

Meuwissen, T.H.E., B.J. Hayes and M.E. Goddard (2001) Genetics 157(4):1819-1829. https://doi.org/10.1093/genetics/157.4.1819

Purcell, S., B. Neale, K. Todd-Brown, L. Thomas, M.A. Ferreira, *et al.* (2007) Am J Hum Genet 81(3):559-575. https://doi.org/10.1086/519795

Sevillano, C.A., J. Vandenplas, J.W.M. Bastiaansen, R. Bergsma and M.P.L. Calus (2017) Genet Sel Evol 49(1):75. https://doi.org/10.1186/s12711-017-0350-1

Xiang, T., B. Nielsen, G. Su, A. Legarra and O. Christensen (2016) Journal of Animal Science 94(3):936-948. https://doi.org/10.2527/jas.2015-9930

274. Improving accuracy of genomic prediction in distant populations by collecting sequence data over generations

M.Y. Nawaz[1,2*] and C. Gondro[1,2]

[1]Genetics and Genome Sciences Program, Michigan State University, East Lansing, MI 48823, USA; [2]Department of Animal Science, Michigan State University, East Lansing, MI 48823, USA; nawazmuh@msu.edu

Abstract

In this study we tested the hypothesis that accuracy of genomic prediction in distant populations can be significantly improved by having the actual QTLs in the SNP panels used for genomic evaluations. Three different simulation scenarios were created for within breed, crossbred and across-breed genomic prediction of Angus, Hereford and Brahman cattle with three sets of SNP panels consisting of only *SNP*, only *QTL*, or *SNP* and *QTL*. We found that when QTL were part of the SNP panel, the decline in prediction accuracy over generations was lower and the accumulation of data over generations led to higher accuracy gains compared to using only non-causal SNPs, especially for crossbred and across breed prediction. A wider adoption of sequence data for genomic prediction still seems a valid proposition to increase the accuracy of genomic prediction in cattle, especially for distantly related animals.

Introduction

Genomic prediction based on single nucleotide polymorphism (SNP) panels is highly dependent on the linkage (LD) between SNP markers and true QTLs regulating the trait. Due to recombination, the linkage between SNPs and QTLs decays across generations and, similarly, animals from different parts of the world or from different breeds also diverge in their genetic composition and LD patterns between markers (De Roos *et al.*, 2008). This results in reduced prediction accuracy across generations (Muir, 2007), across countries (Saatchi *et al.*, 2013) and across breeds as demonstrated by various studies in cattle (Hayes *et al.*, 2009; Kachman *et al.*, 2013; Mason *et al.*, 2012). It was predicted that whole genome sequence (WGS) information would increase the accuracy of prediction (Meuwissen and Goddard, 2010) as it would include most of the variants segregating in a population. Yet, cattle studies to date have reported minimal gain in prediction accuracy using WGS information when compared to medium and high density SNP genotyping panels (Calus *et al.*, 2016; Van Binsbergen *et al.*, 2015). These results were attributed to the high LD observed in cattle populations and an inability to accurately estimate SNP effects due to small sample sizes with millions of SNP markers. The objectives of this study were to: (1) quantify the decay in prediction accuracy as the genetic distances between the reference and test population increased over generations in beef cattle; (2) analyze the effect on prediction accuracy over generations if the true QTLs are part of the SNP panel; and (3) demonstrate the impact of accumulating cross-generational genomic information in the reference population.

Materials & methods

Medium density genotypes (50k) of 7,500 animals from Angus, Hereford and Brahman cattle were used as the base population for the simulation study. These genotypes were phased and imputed separately for each breed using the software *eagle*, converted to *vcf* format with *shapeit* and then into a numerical matrix with the R package *splitstackshape* (R version 4.0.2). The three datasets were subset down to 29,632 SNPs that were common across the three breeds. Matings were simulated using the forward simulator R package *epinetR*. First, a moderately heritable phenotype (h^2=0.4) regulated by 100 QTL was simulated under an additive model using the Angus cattle as a baseline. The SNP effects were extracted and used to create baselines for the Hereford and Brahman cattle. Crossbred F1 genotypes were then simulated using

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_274, © M.Y. Nawaz and C. Gondro 2022

the same software with equal contributions from the combinations of parental breeds. Once we had the baseline populations, we simulated matings under selection for 20 generations for each of the purebred and crossbred base populations. For this step we used the *runSim* function in *epinetR* with the following parameters: *truncSire*=0.3 (top 30% of males in the population with highest phenotypic value were selected during each generation), *truncDam*=0.8 (top 80% of females in the population with highest phenotypic value were selected during each generation), *breedSire*=50 (maximum number of offspring per sire). These parameters were chosen to represent reasonably well the current structure of the beef industry. Three different scenarios were created to evaluate genomic prediction: within a breed, in crossbreds and across breeds; and each was tested for three different SNP panels, namely SNP (all SNPs except the *causal* ones), QTL (only *causal* SNPs), and all SNP&QTL. For every scenario, we evaluated genomic prediction accuracy trends in two ways: (1) by increasing the number of generations between training and testing population; (2) by accumulating training populations backward from the testing population. Genomic prediction was performed by genomic best linear unbiased prediction methodology (GBLUP) using a genomic relationship matrix (G) to account for genomic relationships between animals (VanRaden, 2008).

Results

In all idealized scenarios where only the true QTLs were used for prediction, the accuracy was high, close to the theoretical limit of 0.63 and largely constant over generations (dotted lines in Figure 1A,1B,1C,1D,1E,1F). Results from *SNP* and *SNP and QTL* model were as follows.

Within breed and crossbred prediction. For genomic prediction of purebreds and crosses, accuracy of prediction was highest when the reference population was one generation apart from the test population. As the distance between reference and test population increased, the accuracy of prediction decreased for the first 10 generations until it hit a baseline level (Figure 1A,1C). When causal QTLs were included in the SNP set, the prediction accuracy was consistently higher, and the prediction accuracy decay was lower over generations in comparison to using only the SNPs. As expected, accuracies were higher in within breed predictions as compared to crossbred prediction.

Across breed prediction. Accuracy was low but constant over generations (Figure 1E). However, when the reference populations were accumulated over generations, we saw an increase in prediction accuracy for the first 2-3 generations (Figure 1F). From then on, the accuracy of prediction did not increase considerably for *SNP* model even though the size of reference population consistently increased due to addition of more generations. If causal QTL were included in the SNP set, the accuracy of prediction continued to increase with the addition of more generations in the reference dataset. Not surprisingly, the taurine breeds predicted each other better as compared to predicting Brahman or when using Brahman to predict the taurine breeds. This was true in scenarios which used only SNPs or both SNPs and QTLs were in the model. However, using only QTLs to predict the other breeds, Angus reference population performed the worst. Its prediction accuracy was even lower than Hereford predicting Brahman cattle and vice versa, which were genetically distant from each other. This result shows that even if we knew the true QTL, their effects may still not be accurately estimated in a reference population that is highly inbred.

Discussion

In this study our simulations demonstrated that if the actual QTLs are part of the SNP panel: (1) the prediction accuracy improved; (2) the decay in prediction accuracy over generations was smaller; and (3) the advantage of collecting genotype data over generations was higher. There was clear evidence that the prediction accuracy increased when the actual QTLs were part of the SNP data. The same is true for the value of accumulating genotypes over generations which is significantly higher when the QTLs are genotyped, although the gains are less pronounced in within breed predictions. This result has important implications

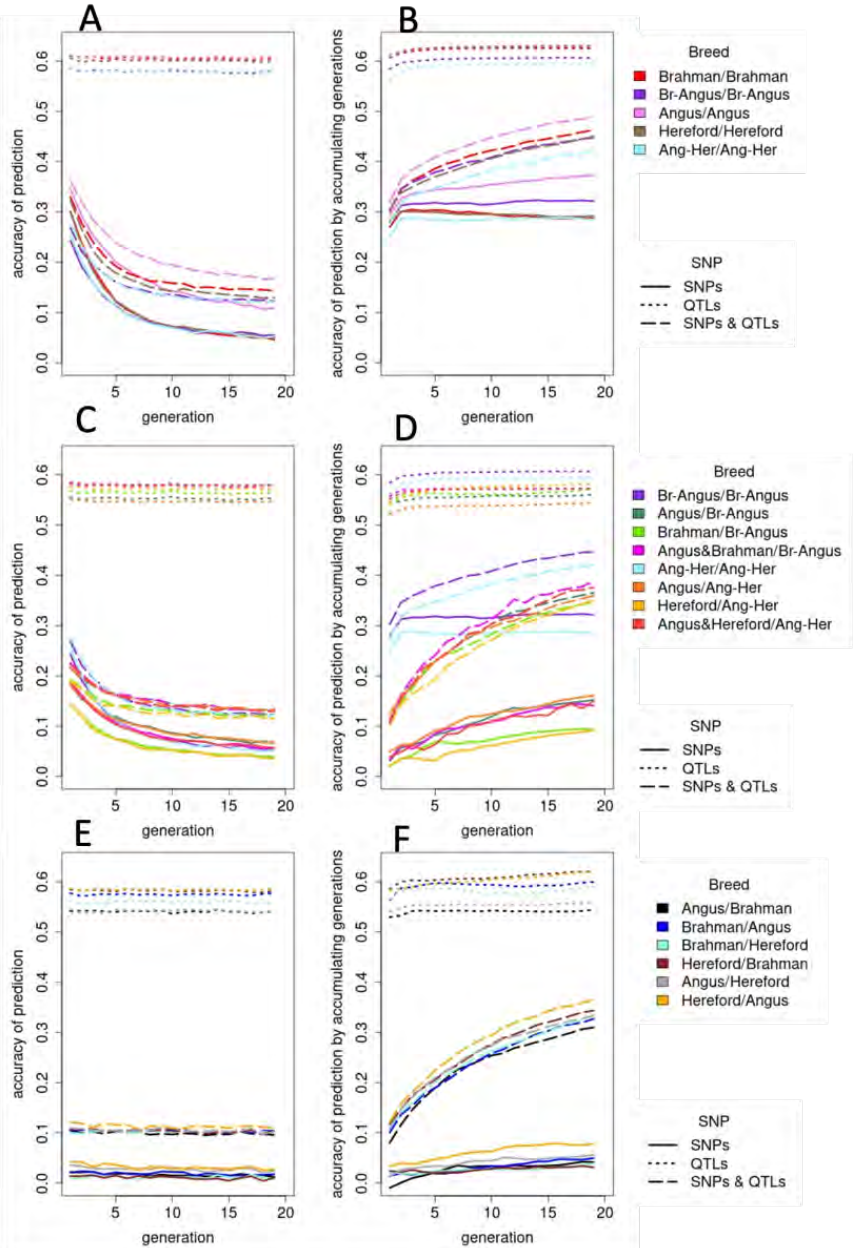


Figure 1. Accuracy of genomic prediction as a function of genetic distance in terms of number of generations between reference and test populations, and against cumulative generations backwards from the reference population in within breed (A,B), crossbred (C,D), and across breed (E,F) predictions.

for breeding companies and breed organizations that have been collecting genotype information for more than a decade now which could be imputed up to WGS to improve prediction accuracy – a strategy that could be especially relevant for across breed or multibreed evaluation programs since our results demonstrated that the benefit of including QTLs was more evident when there was a larger genetic distance between the

reference and target population. The three SNP sets used in this study serve as a proxy for three models. Firstly, SNPs without the QTLs represent widely used SNP chip data which does not include most of the causal QTLs for traits. The SNP and QTL panel represents whole genome sequence data which in principle should include most of the causal variants and other correlated SNPs. Here we should note that a limitation of this study is that it did not address the n << p problem which is frequently discussed when it comes down to the practical use of WGS data for cattle breeding – how well these results hold at the full sequence level with 20-40 million variants needs further studies. Finally, the QTL only panel represents an ideal situation where we can identify and utilize only those SNPs that are directly involved in a phenotype. Such exact SNP panels are still elusive and would require studies based on huge sample sizes at the WGS level. Nonetheless, some studies have already shown the benefits of selecting highly predictive features to increase prediction in distant populations (Raymond *et al.*, 2018; van den Berg *et al.*, 2016; Wientjes *et al.,* 2015). In summary, wider adoption of WGS for genomic prediction still seems a valid proposition to increase the accuracy of genomic prediction in cattle, especially for distantly related animals.

References

Calus M.P.L., Bouwman A.C., Schrooten C., and Veerkamp R.F. (2016) *Genet. Sel. Evol.* 48:1–19 https://doi.org/10.1186/s12711-016-0225-x

De Roos A.P.W., Hayes B.J., Spelman R.J., and Goddard M.E. (2008) *Genetics* 179, 1503–1512 https://doi.org/10.1534/genetics.107.084301

Hayes B.J., Bowman P.J., Chamberlain A.C., Verbyla K., and Goddard M.E. (2009) *Genet. Sel. Evol.* 41, 1–9 https://doi.org/10.1186/1297-9686-41-51

Kachman S.D., Spangler M.L., Bennett G.L., Hanford K.J., Kuehn L.A., *et al.* (2013) *Genet. Sel. Evol.* 45, 1 https://doi.org/10.1186/1297-9686-45-30

Mason B.A., Goswami S., Hayes B.J., Goddard M.E., Bowman *et al.* (2012) *J. Dairy Sci.* 95, 4114–4129 https://doi.org/10.3168/jds.2011-5019

Meuwissen T., and Goddard M. (2010) *Genetics* 185, 623–631 https://doi.org/10.1534/genetics.110.116590

Muir W.M. (2007) *J. Anim. Breed. Genet.* 124, 342–355 https://doi.org/10.1111/j.1439-0388.2007.00700.x

Raymond B., Bouwman A.C., Schrooten C., Houwing-Duistermaat J., and Veerkamp R.F. (2018) *Genet. Sel. Evol.* 50 https://doi.org/10.1186/s12711-018-0396-8

Saatchi M., Ward J., and Garrick D.J. (2013) *J. Anim. Sci.* 91, 1538–1551 https://doi.org/10.2527/jas.2012-5593

Van Binsbergen R., Calus M.P.L., Bink M.C.A.M., Van Eeuwijk F.A., Schrooten C., *et al* (2015) *Genet. Sel. Evol.* 47, 1–13 https://doi.org/10.1186/s12711-015-0149-x

van den Berg I., Boichard D., Guldbrandtsen B., and Lund M.S. (2016) *G3: Genes|Genomes|Genetics* 6, 2553–2561 https://doi.org/10.1534/g3.116.027730

VanRaden P.M. (2008) *J. Dairy Sci.* 91, 4414–23. https://doi.org/10.3168/jds.2007-0980

Wientjes Y.C.J., Veerkamp R.F., and Calus M.P.L. (2015) *BMC Genet.* 16, 1–15. https://doi.org/10.1186/s12863-015-0252-6-x

275. Comparison of genomic prediction methods for residual feed intake in broilers

Z.X. He[1,2#], S. Li[1#], W. Li[1], J.Q. Ding[1], M.Q. Zheng[1], Q.H. Li[1], A.G. Fahey[2], J. Wen[1], R.R. Liu[1*] and G.P. Zhao[1*]

[1]State Key Laboratory of Animal Nutrition; Key Laboratory of Animal (Poultry) Genetics Breeding and Reproduction, Ministry of Agriculture, Institute of Animal Sciences, Chinese Academy of Agricultural Sciences, Beijing 100193, China; [2]School of Agriculture and Food Science, University College Dublin, Dublin, Ireland; zhaoguiping@caas.cn; liuranran@caas.cn; [#]these authors contributed equally to this work

Abstract

Residual feed intake (RFI) is a measure of feed efficiency of animals. Previous studies have identified SNPs associated with RFI. The objective of this study was to compare the GBLUP model with GA-BLUP model included previously identified associated SNPs. The nine RFI-associated SNPs were obtained from the GWAS study on a discovery population, as preselection information. These models were analysed using ASREML software using a 5-fold cross-validation method on a validation population. With the genetic architecture (GA) matrix involved, which was conducted with the nine RFI-associated SNPs, the prediction accuracy of RFI was improved compared to the original GBLUP model. The prediction accuracy increased by 2% in GA-BLUP model with ω is 0.981 compare to GBLUP model. In conclusion, the GA-BLUP with the nine associated SNPs and an optimal ω can improve prediction accuracy for a specific trait compared to GBLUP.

Introduction

Genomic selection (GS) is a genetic improvement tool that utilises marker-assisted selection to predict the genetic merit of animal and is a genetic improvement tool used in animal breeding (Meuwissen *et al.* 2001). The best linear unbiased prediction (BLUP) is a general model in genetic and genomic selection used to estimate breeding values (Henderson 1975). A new method is named BLUP-given genetic architecture (GA, GA-BLUP). GA-BLUP is constructed using a trait-specific covariance matrix (T) by a weighted sum of the GA matrix and realized genomic relationship matrix (G) (VanRaden 2008; Zhang *et al.* 2015). The GA matrix was made by the nine pre-selection SNPs, which resulted from GWAS and fine mapping. A recent study showed that the GA-BLUP method that involved a machine learning method to combine prediction accuracy with computing efficiency(Yin *et al.* 2020).The GA-BLUP method could result in higher prediction accuracy of breeding values with the optimal parameter, weight (ω), applied. However, the optimal weight when combining GA matrix and G matrix needs to be determined.

As a critical factor representing feed efficiency, residual feed intake (RFI) is the difference between actual and expected feed intake, predicting by body weight and body weight gain (Henderson 1975). The advantage of RFI compared to other feed efficiency traits such as average daily gain is that you can select to decrease feed intake without connection to body weight and growth rate (Hoque *et al.* 2007; Crowley *et al.* 2010; Begli *et al.* 2016; Shirali *et al.* 2018). It is necessary to introduce the RFI-associated SNPs to genomic prediction and establish new approaches that result estimated breeding values with higher accuracy. In this study, two separate populations were used for GWAS discovery and GS validation, respectively. RFI-associated SNPs from the discovery population were used as the pre-selection information to build the GA matrix. The ω between the GA matrix and the G matrix was also tested and calculated based on the components explained variance.

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_275, © Z.X. He *et al.* 2022

Materials & methods

In total, 989 chickens were obtained from the fast-growing white-feathered pure line B. Line B is a synthetic line produced by Foshan Gaoming Xinguang Agricultural and Animal Industrials Co., Ltd. (Foshan, China), and it has been selected for high body weight and growth rate traits for seven generations. All individuals were housed in individual cages (length × width × height: 30×25×45 cm). The total feed intake for each broiler was the total feed consumption during the test period between 28 to 40 days of age, which was then used to calculate the average daily feed intake (ADFI). The individual body weight gain was calculated on the basis of body weight at 28 d (BW28) and (BW42), and was then used to derive the ADG. RFI was the difference between the observed and predicted ADFI. The predicted ADFI was calculated as formula (1).

$$ADFI = \mu + \beta_1 MWT + \beta_2 ADG + e \quad (1)$$

where μ represents the intercept, MWT represents the metabolic body weight at mid-test. ADG represents the average daily gain, β_1 and β_2 represent partial regression coefficients, and e represents the residual. The estimated e was taken as the measure of RFI.

Genotyping. The genotyping was described in detail in previous studies (Li *et al.* 2020; Li *et al.* 2021). The GWAS discovery population is the reference group at that generation, which contained 273 females and 244 males. The GS validation population was the candidate group including 472 males. A GWAS was performed for the RFI trait using a univariate linear mixed model (LMM) implemented in GEMMA version 0.98.1 software (https://github.com/genetics-statistics/GEMMA/releases) (Zhou and Stephens 2012). Gender was used as a covariate in LMM model. Nine RFI associated-SNPs were selected based on the reference group.

Genomic prediction. This study compared genomic prediction accuracy between GBLUP(2) and GA-GBLUP(3).

$$\mathbf{y} = \mathbf{Xb} + \mathbf{Zg} + \mathbf{e}; \quad \mathrm{Var}\begin{bmatrix}\mathbf{g}\\ \mathbf{e}\end{bmatrix} = \begin{bmatrix}\mathbf{G}\sigma_g^2 & 0\\ 0 & \mathbf{I}\sigma_e^2\end{bmatrix} \quad (2)$$

$$\mathbf{y} = \mathbf{Xb} + \mathbf{Zg} + \mathbf{Z}'g' + \mathbf{e}; \quad \mathrm{Var}\begin{bmatrix}\mathbf{g}\\ \mathbf{e}\end{bmatrix} = \begin{bmatrix}\mathbf{G}_{total}\sigma_g^2 & 0\\ 0 & \mathbf{I}\sigma_e^2\end{bmatrix} \quad (3)$$

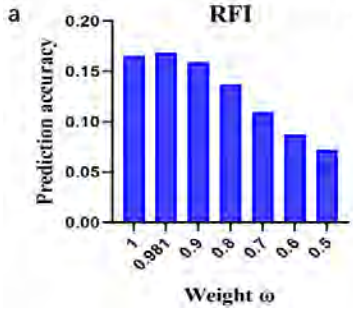
where b is the vector of fixed factor, g is the vector of random direct additive genetic effects, g' is the vector of discovered SNPs effects, e is the vector of random residual effects, and X Z and Z' are induced matrices. σ_g^2 and σ_e^2 are the additive genetic and residual error variances. Gtotal means the weighted sum of G and GA matrices, and G_total=ωG+(1-ω)GA, in which ω is the weight. The ω is calculated by the formula ω=(σ_g^2) / ($\sigma_g^2 + \sigma_{gSNP}^2$), where σ_g^2 and σ_{gSNP}^2 are the variance of 55K array and associated SNPs from GWAS, respectively.

The accuracies of the two different methods and the 5-fold cross-validation (CV) method were both conducted by the ASRmel v4.1 software package (Gilmour 2019). The 5-fold cross-validation method was replicated ten times to get the mean accuracies. For each fold, the same nine SNPs were for GA-BLUP.

Results

The comparison of the prediction accuracies is presented in Figure 1. Compared to the traditional GBLUP model, the prediction accuracy improved by 2% through the GA-BLUP model. The higher prediction accuracy could be explained by inclusion of the SNP information from the GWAS (Lourenco *et al.* 2015).

a



b

Weight ω	Prediction accuracy	Improvement [a]
1	0.165	
0.981	0.169	1.9%
0.9	0.159	-3.9%
0.8	0.137	-17.2%
0.7	0.109	-34.0%
0.6	0.087	-47.3%
0.5	0.072	-56.3%

Figure 1. The prediction accuracy changes with different weights of G matrix on RFI in GBLUP model. (a) Bar chart of prediction accuracy under different weights. (b) The details of the prediction accuracy and improvement.
[a] Improvement is compared the prediction accuracy with the weight equals to 1.

Discussion

This is the first time the poultry RFI was used to conduct GA-BLUP, and it showed improved prediction accuracy, which agreed with many previous studies (Gao *et al.* 2015; Zhang *et al.* 2015; Ni *et al.* 2017; Morgante *et al.* 2018). However, only 2% was found in this study. One possible reason could be the low-density SNP chip was used for GBLUP and few SNPs for conducting GA-BLUP. Another possible reason could be RFI is a low to moderate heritable trait which could cause the low prediction accuracy. A previous study reported that the different traits and the different number of SNPs used for GA could lead to 1-70% of prediction accuracy increase (Romé *et al.* 2021). In conclusion, RFI is a trait in which GA-BULP could increase the prediction accuracy with the SNPs from the GWAS for RFI

Another critical parameter is the ω in the processing of combining GA with the G matrix. In Holstein cow data set, GA in GBLUP showed an improved prediction accuracy compared to GBLUP with the ω range from 0.2 to 0.9% (Zhang *et al.* 2015). Even though the different population sizes could affect the optimal ω, the prediction accuracy increased then decreased with the increasing ω (Zhang *et al.* 2015; Liu *et al.* 2019). The prediction accuracy trend agreed with our result, which indicated that the optimal ω could cause an enormous prediction improvement. However, testing the most optimal ω could waste much time. Thus, a method to calculate the ω is needed. In this method, the genetic matrix's ω was calculated by the RFI variance component explained by G and GA matrix, and it showed a similar result with previous studies (Zhang *et al.* 2015; Ye *et al.* 2020). The calculated ω using the variance component was 0.981, which provided the highest accuracy in RFI. This result indicated that our ω calculation method was valid and efficient.

References

Begli H.E., Torshizi R.V., Masoudi A.A., Ehsani A. and Jensen J. (2016) Livestock Science 184(28-34. https://doi.org/10.1016/j.livsci.2015.11.018

Crowley J.J., Mcgee M., Kenny D.A., Crews D.H., Jr., Evans R.D. *et al.* (2010) Journal of Animal Science 88(3): 885-894. https://doi.org/10.2527/jas.2009-1852

Gao N., Li J., He J., Xiao G., Luo Y. *et al.* (2015) BMC Genet 16(120. https://doi.org/10.1186/s12863-015-0278-9

Gilmour A.R. (2019) 136(4): 262-272. https://doi.org/10.1111/jbg.12398

Henderson C.R. (1975) Biometrics 31(2): 423-447.

Hoque M.A., Suzuki K., Kadowaki H., Shibata T. and Oikawa T. (2007) Journal of animal breeding and genetics = Zeitschrift fur Tierzuchtung und Zuchtungsbiologie 124(3): 108-116. https://doi.org/10.1111/j.1439-0388.2007.00650.x

Li W., Liu R., Zheng M., Feng F., Liu D. *et al.* (2020) J Anim Sci Biotechnol 11(65. https://doi.org/10.1186/s40104-020-00469-8

Li W., Zheng M., Zhao G., Wang J., Liu J. *et al.* (2021) Genetics Selection Evolution 53(1): 13. https://doi.org/10.1186/s12711-021-00608-3

Liu X., Wang H., Hu X., Li K., Liu Z. *et al.* (2019) Front Plant Sci 10(1129. https://doi.org/10.3389/fpls.2019.01129

Lourenco D.A., Fragomeni B.O., Tsuruta S., Aguilar I., Zumbach B. *et al.* (2015) Genet Sel Evol 47(1): 56. https://doi.org/10.1186/s12711-015-0137-1

Meuwissen T.H., Hayes B.J. and Goddard M.E. (2001) Genetics 157(4): 1819-1829.

Morgante F., Huang W., Maltecca C. and Mackay T.F.C. (2018) Heredity (Edinb) 120(6): 500-514. https://doi.org/10.1038/s41437-017-0043-0

Ni G., Cavero D., Fangmann A., Erbe M. and Simianer H. (2017) Genet Sel Evol 49(1): 8. https://doi.org/10.1186/s12711-016-0277-y

Romé H., Chu T.T., Marois D., Huang C.H., Madsen P. *et al.* (2021) Journal of animal breeding and genetics = Zeitschrift fur Tierzuchtung und Zuchtungsbiologie 138(5): 528-540. https://doi.org/10.1111/jbg.12546

Shirali M., Varley P.F. and Jensen J. (2018) Genet Sel Evol 50(1): 33. https://doi.org/10.1186/s12711-018-0403-0

Vanraden P.M. (2008) Journal of Dairy Science 91(11): 4414-4423. https://doi.org/10.3168/jds.2007-0980

Ye S., Li J. and Zhang Z. (2020) J Anim Sci Biotechnol 11(1): 109. https://doi.org/10.1186/s40104-020-00515-5

Yin L., Zhang H., Zhou X., Yuan X., Zhao S. *et al.* (2020) Genome Biology 21(1): 146. https://doi.org/10.1186/s13059-020-02052-w

Zhang Z., Erbe M., He J., Ober U., Gao N. *et al.* (2015) G3 (Bethesda, Md.) 5(4): 615-627. https://doi.org/10.1534/g3.114.016261

Zhou X., and Stephens M. (2012) Nat Genet 44(7): 821-824. https://doi.org/10.1038/ng.2310

276. Impact of foreign phenotype data on single-step genomic evaluation of test-day protein yields for German Holsteins

H. Alkhoder*, Z. Liu, D. Segelke and R. Reents

IT solutions for animal production (vit), Heinrich-Schroeder-Weg 1, 27283 Verden, Germany;
hatem.alkhoder@vit.de

Abstract

Foreign phenotype data of bulls integrated in single-step evaluation may have an impact on genomic prediction. We set up three scenarios regarding the selection of integrated bulls from MACE evaluation: all bulls with daughters in any country (ALL), daughter-proven bulls with genotype data available (REF), and daughter-proven bulls linked to national cows or genotyped animals (LNK). For protein yield, GEBV correlation of validation bulls between the truncated and full evaluations differed little among the scenarios. However, regression coefficient of GEBV of the full on the truncated evaluation increased from 0.91 for scenario REF to 0.92 for LNK and to 0.96 for ALL. Scenarios REF and LNK resulted in higher GEBV mean and variance for young genomic bulls than scenario ALL. Results suggest that using MACE data of all bulls for national single-step evaluation leads to a lower inflation of genomic prediction.

Introduction

For the international dairy breed Holstein, national single-step genomic evaluation (Misztal *et al.* 2009) usually integrates foreign phenotype data, such as bull MACE EBV, to enhance the accuracy of genomic prediction. Daughter-proven bulls born before the implementation of genomic selection were genotyped incompletely or selectively in most countries. Because those genotyped bulls with daughters tend to be genetically superior to un-genotyped bulls of the same age, including only those old, genotyped bulls may bias the genomic prediction of young animals. The aim of this study was to investigate the impact of foreign phenotype data of bulls on national single-step genomic prediction in German Holsteins.

Materials & methods

For test-day protein yields in first three lactations, a multiple-lactation random regression test-day model has been applied in German Holsteins (Liu *et al.* 2004). Because a single-step SNP BLUP (ssSNPBLUP) model by Liu *et al.* (2014) was shown to be efficient (Vandenplas *et al.* 2021), the ssSNPBLUP model was chosen for routine single-step evaluation for German Holsteins (Alkhoder *et al.* 2022). The test-day data of national cows were jointly evaluated with deregressed MACE breeding values of foreign bulls together with effective daughter contribution of the bulls as weights. Genetic covariances between the national and MACE traits were derived using the genetic (co)variances of random regression coefficients (Liu *et al.* 2004). To avoid double counting the contribution of national daughters to those bulls, both deregressed EBV and effective daughter contribution of the bulls were adjusted for the national daughters. Genotype and the phenotype data were obtained from the official German evaluation in April 2021. There were 949,636 genotyped Holstein animals available. A total of 242,121,126 test-day records from 12,432,940 cows were jointly evaluated with 138,770 bulls with daughters outside Germany. Pedigree file for the genotyped or phenotyped animals contained e.g. 20 generations of ancestors for the youngest genotyped candidates born in 2021. The total number of animals in the pedigree file was 20,461,400. The number of unknown parent groups was 177, defined according to country origin, selection paths and birth year of the animals. The assumed proportion of genetic variance not explained by the SNP markers was 30% for the protein yield.

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_276, © H. Alkhoder *et al.* 2022

Scenarios of selecting bulls from MACE evaluation. Three scenarios regarding the selection of integrated bulls from MACE evaluation were considered: all bulls with daughters in any country (ALL), daughter-proven bulls with genotype data available (REF), and daughter-proven bulls linked to national cows or genotyped animals (LNK). Figure 1 shows the number of Holstein bulls with daughters evaluated in the April 2021 MACE evaluation of protein yield that were selected for the three scenarios. Prior to the implementation of genomic selection in 2009, less than 20% bulls were genotyped. The reference bulls REF had higher genetic level, i.e. MACE EBV, than the linked bulls LNK and much higher than the bulls ALL for the trait protein yield (data not shown here).

Genomic validation by data truncation. To simulate a forward prediction in four years, test-day data of national cows from the last four years were removed. The integrated MACE bulls born in last four years were deleted. In addition, national daughters of all the deleted bulls were removed to ensure no daughters of the deleted bulls were present in the truncated phenotype data set. Validation bulls were defined as genotyped Holstein bulls with daughters in at least 10 herds in Germany. There were 1,665 validation bulls born in 2013 through 2016. In the truncated phenotype data, there were 10,903,891 national cows with a total of 222,634,210 test-day records and 128,504 bulls for the scenario ALL.

A single-step SNP BLUP model for genomic evaluation. Identical fixed and random effects were used for the single-step evaluation as for the conventional test-day model evaluation (Liu *et al.* 2004). The software MiX99 (Strandén and Lidauer 1999) that was used for the single-step evaluation implemented the ssSNPBLUP model in a special way (Mäntysaari *et al.* 2018, unpublished data). The total number of equations or effects was 338,616,497 for the full data set under scenario ALL. Memory usage was about 347 Gb and solving the equations took about 1.63 minutes per round of iteration on a Linux server using 17 cores. Pre-defined convergence criteria were met within 3000 rounds for all the scenarios.

Three groups of genotyped animals were selected to compare the scenarios: genotyped AI bulls owned by German AI studs, genotyped male candidates without own daughters and genotyped females with or without own phenotype records. We focused on their genetic trends, GEBV variances and correlations between the truncated and full evaluations. Because the genotyped AI bulls had the highest selection intensity among the three groups, we are only showing the genomic evaluation results for a total of 12,249 genotyped German Holstein AI bulls born between 1998 and 2020.

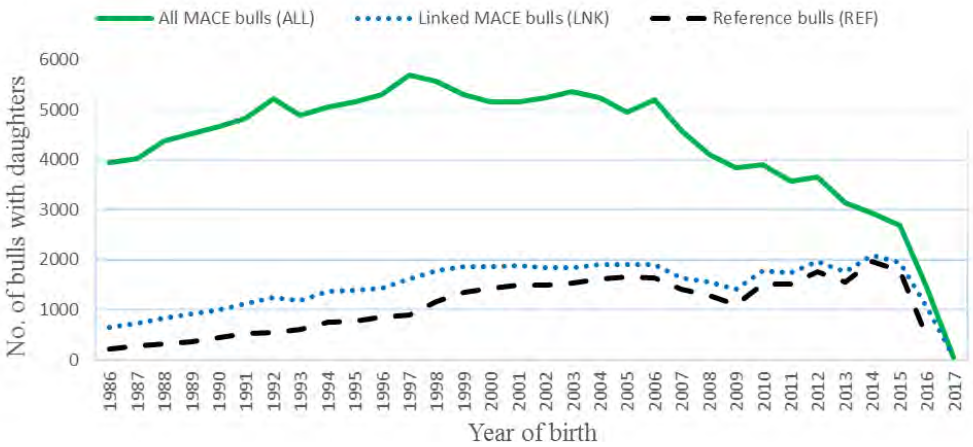


Figure 1. Number of Holstein bulls selected for the three scenarios (trait: protein yield).

Results & discussion

For either the full or truncated data set SNP effect estimates had a correlation higher than 0.99 between any two of the scenarios. Regression of the SNP effects of scenario ALL on either scenario REF or LNK was nearly 1 for all the cases. The three scenarios resulted in almost identical SNP effect estimates.

Genomic validation based on the validation bulls. The LR regression method (Legarra and Reverter 2018) was applied to GEBV of the validation bulls from the full and truncated data sets. Table 1 shows the genomic validation results according to the LR regression. The GEBV correlations of the validation bulls, shown in R^2 values, did not differ among the three scenarios. However, significant differences in the regression coefficient b_1 were observed among the scenarios. The regression coefficients indicate that selecting only genotyped bulls (scenario REF) led to the highest inflation in GEBV of the validation bulls among the three scenarios, whereas the smallest overestimation of candidate GEBV was found when all bulls evaluated in MACE (scenario ALL) were integrated into the national single-step evaluation. This finding may be explained by the fact that the genotyped bulls born before the introduction of genomic selection had higher genetic merits than those non-genotyped bulls in the same years. Additionally, the scenario ALL had the smallest b_0 value among all the scenarios.

GEBV correlations between the full and truncated evaluations. Figure 2 shows GEBV correlations for the genotyped German AI bulls between the full and truncated evaluation. It is clearly seen that using all bulls from MACE resulted in highest GEBV correlation between the full and truncated evaluations among the three scenarios.

Table 1. Regression of validation bulls' GEBV from the full on the truncated evaluation.

GEBV of protein yield (kg)	b_0	b_1	R^2
All MACE bulls (ALL)	0.47	0.96	0.71
Linked MACE bulls (LNK)	1.99	0.92	0.71
Reference bulls (REF)	2.13	0.91	0.70

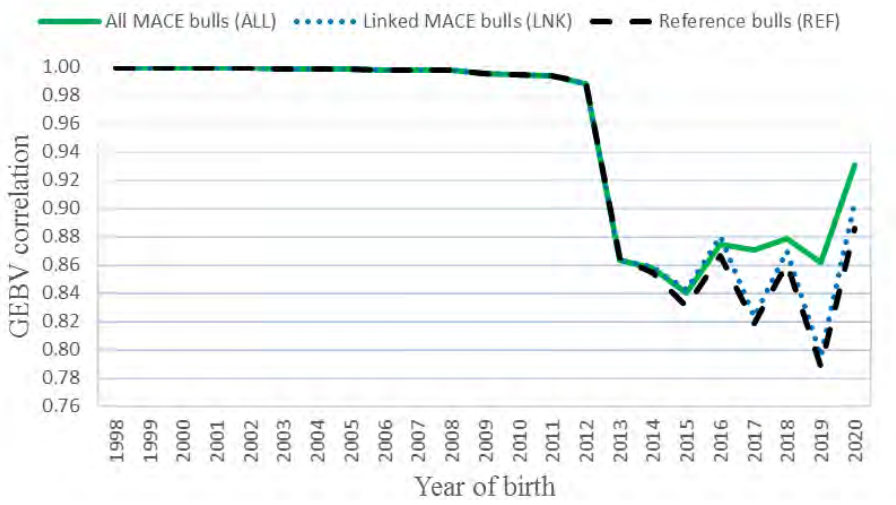


Figure 2. GEBV correlations between the full and truncated evaluations for the AI bulls.

Genetic trends in the genotyped German AI bulls. Genetic trend of scenario ALL, expressed in units of genetic standard deviation, is shown in Figure 3 as solid green line against the secondary Y axis. Genetic trend differences to scenario ALL are shown for scenario LNK in dotted blue line (LNK-ALL) and for scenario REF in dashed black line (REF-ALL). When only the reference bulls were used, an overestimation of GEBV for young AI bulls were evident. The trend difference of scenario LNK to ALL was small for the full data set, but a larger overestimation for the young AI bulls was shown in the truncated data set (unpublished data).

Besides the differences in genetic trends between the scenarios, we saw also slightly higher GEBV variances in young AI bulls, particularly for the truncated data set. Additionally, we calculated regression slopes of GEBV from the full on the truncated evaluation for the AI bulls by birth year. We found clearly lower regression slopes in young AI bulls for the scenario REF or LNK than scenario ALL, suggesting GEBV of young AI bulls were more inflated in the two scenarios REF and LNK.

In comparison to the genotyped AI bulls, genotyped female animals showed smaller GEBV differences between the scenarios. Like the genotyped AI bulls, the male candidates without own data had similar GEBV differences between the scenarios.

Our genomic validation showed a larger inflation, in terms of the regression slope, for scenario REF or LNK than scenario ALL, though the GEBV correlation between the full and truncated evaluation did not differ between the scenarios. For the young, genotyped animals, particularly for the AI bulls, scenarios REF or LNK resulted in a larger bias in genomic prediction.

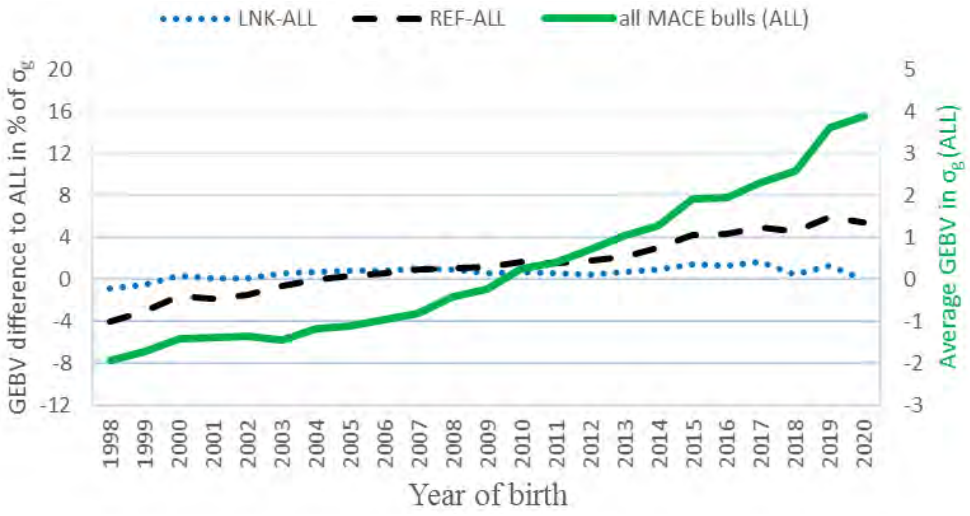


Figure 3. Genetic trends of protein yield in the genotyped AI bulls.

References

Alkhoder H., Liu Z., Segelke D., and Reents R. (2022). J. Dairy Sci. (in press)

Legarra A., and Reverter A. (2018). Genet. Sel. Evol. 50:53. https://doi.org/10.1186/s12711-018-0426-6

Liu Z., Goddard M.E., Reinhardt F., and Reents R. (2014). J. Dairy Sci. 97, 5833-5850. https://doi.org/10.3168/jds.2014-7924

Liu Z., Reinhardt F., Bünger A., and Reents R. (2004). J. Dairy Sci. 87:1896–1907. https://doi.org/10.3168/jds.S0022-0302(04)73348-2

Misztal I., Legarra A., and Aguilar I. (2009). J. Dairy Sci. 92, 4648-4655. https://doi.org/10.3168/jds.2009-2064

Strandén I., and Lidauer M. (1999) J. Dairy Sci. 82:2779-2787. https://doi.org/10.3168/jds.S0022-0302(99)75535-9

Vandenplas J., Eding H., and Calus M.P.L. (2021). J. Dairy Sci. 104:3298-3303. https://doi.org/10.3168/jds.2020-19460

277. Extend mixed models to multi-layer neural networks for genomic prediction including intermediate omics data

T. Zhao[1,2], J. Zeng[3] and H. Cheng[1*]

[1]Department of Animal Science, University of California Davis, 95616, Davis, USA; [2]Integrative Genetics and Genomics Graduate Group, University of California Davis, 95616, Davis, USA; [3]Institute for Molecular Bioscience, The University of Queensland, 4072, Brisbane, Australia; qtlcheng@ucdavis.edu

Abstract

With the growing amount and diversity of intermediate omics data complementary to genomics (e.g. gene expression), there is a need to develop methods to incorporate intermediate omics data into conventional genomic evaluation. We developed a new method named NN-LMM to model the multiple layers of regulation from genotypes to intermediate omics features, then to phenotypes, by extending conventional linear mixed models ('LMM') to multi-layer neural networks ('NN'). NN-LMM incorporates intermediate omics features by adding middle layers between genotypes and phenotypes. Linear mixed models (e.g. GBLUP, Bayesian Alphabet) can be used to sample marker effects or genetic values on intermediate omics features, and activation functions in neural networks can capture the nonlinear relationships between intermediate omics features and phenotypes. NN-LMM had significantly better prediction performance than the recently proposed single-step approach. Moreover, NN-LMM can handle various patterns of missing omics measures. NN-LMM has been implemented in an open-source package called 'JWAS'.

Introduction

The advances in high-throughput sequencing technology provide growing amount and diversity of multi-omics data complementary to genomics. The effects of genotypes on phenotypes can be mediated by multiple layers of omics features through mechanisms such as regulatory cascades from epigenome, to transcriptome, and to proteome (Ritchie *et al.*, 2015; Wu *et al.*, 2018). This multi-layer regulation works as a unified system to connect genome variations to the trait, and the relationships between different layers can be complex with interactions and nonlinear relationships (Kitano, 2002). A system of two linear models has been developed recently for genomic evaluation (Christensen *et al.*, 2021; Weishaar *et al.*, 2020), where one linear model describes how genotypes affect gene expression levels, and another describes how gene expression levels affect phenotypes. This system of two linear models was further extended for incomplete omics data based on single-step approach (Christensen *et al.*, 2021).

We developed a new method named NN-LMM (Zhao *et al.*, 2021a; Zhao *et al.*, 2021b) to model the multiple layers of regulation from genotypes to intermediate omics features, then to phenotypes, by extending conventional linear mixed models ('LMM') to multi-layer artificial neural networks ('NN'). NN-LMM incorporates intermediate omics features by adding middle layers between genotypes and phenotypes. Linear mixed models can be used to sample marker effects or genetic values on intermediate omics features, and nonlinear activation functions in neural networks are used to approximate the nonlinear relationships between intermediate omics features and phenotypes. Compared to other methods, NN-LMM allows various patterns of missing omics data, for example, individuals can have different missing omics features, and the assumption of nonlinearity between intermediate omics features and phenotypes may be more biologically realistic.

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_277, © T. Zhao *et al.* 2022

Materials & methods

The framework of NN-LMM incorporating intermediate omics data such as gene expression levels is shown in Figure 1(a). Genotypes affect the gene expression levels, then gene expression levels regulate the phenotypes. Linear mixed models can be applied to sample marker effects or genetic values on gene expression levels, and the non-linear activation function in neural networks will be used to capture the complex nonlinear relationships between gene expression levels and phenotypes. For an individual, the gene expression levels of the first two genes are 0.9 and 0.1, respectively, and the gene expression of the last gene is missing to be sampled. Individuals can have different missing gene expression levels. A detailed framework of NN-LMM is shown in Figure 1(b), and Markov chain Monte Carlo (MCMC) approaches are used to infer unknowns.

From middle layer (intermediate omics features) to output layer (phenotypes): non-linear activation function. Given all intermediate omics features, the phenotype of individual *i* is modeled as:

$$y_i = \mu^{(1)} + \sum_{j=1}^{l_1} w_j^{(1)} g\left(z_{i,j}\right) + e_i$$

where y_i is the phenotype, $\mu^{(1)}$ is the overall mean with flat prior, $z_{i,j}$ is the j-th intermediate omics feature, $g(.)$ is the activation function in neural networks, $w_j^{(1)}$ is the effect of $g(z_{i,j})$ on y_i with a normal prior, and e_i is the random residual with normal prior.

From input layer (genotypes) to middle layer (intermediate omics features): mixed models. Given all intermediate omics features, for individual *i*, the relationship between the j-th intermediate omics feature and genotypes can be written as a single-trait mixed model (e.g. Bayesian Alphabet) as:

$$z_{i,j} = \mu_j^{(0)} + \sum_{m=1}^{l_0} x_{i,m} w_{j,m}^{(0)} + \epsilon_{i,j}$$

where $z_{i,j}$ is the j-th intermediate omics feature, $\mu_j^{(0)}$ is its overall mean with flat prior, $x_{i,m}$ is the genotype covariate at locus m (coded as 0, 1, 2), $w_{j,m}^{(0)}$ is the marker effects of locus m on j-th intermediate omics feature, which can be sampled by various linear mixed models, ϵ_{ij} is the random residual with normal prior.

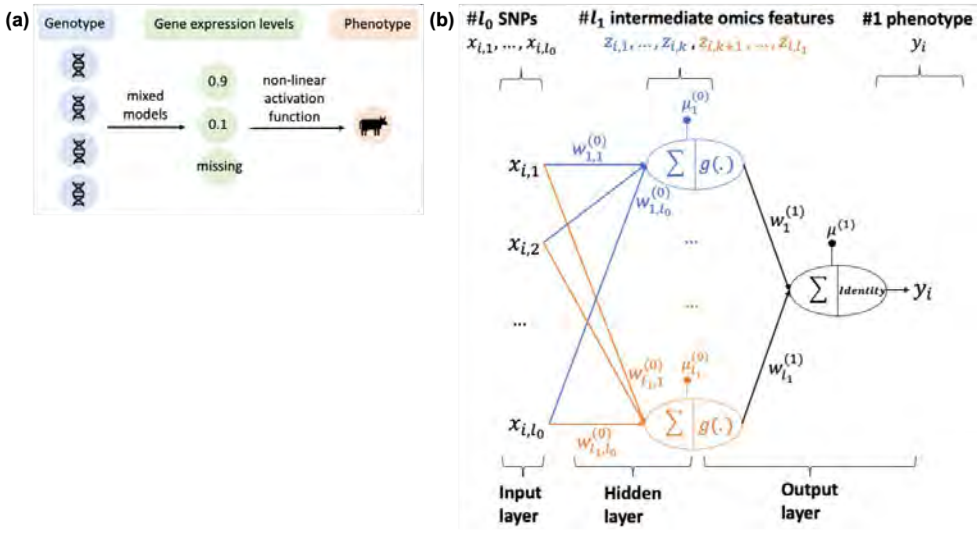


Figure 1. The framework of NN-LMM incorporating intermediate omics data.

Sample missing omics data by Hamiltonian Monte Carlo. Each missing omics feature of individual *i* will be treated as an unobserved intermediate trait to be sampled by Hamiltonian Monte Carlo (HMC) from its full conditional distributions. The introduction to the concepts underlying HMC can be found in Betancourt (2018).

Data analysis. To compare the prediction performance of NN-LMM to the single-step approach in Christensen *et al.* (2021), a linear activation function was used in NN-LMM. GBLUP was used to sample genetic values on intermediate omics features (i.e. NN-GBLUP). Simulated data from Christensen *et al.* (2021) were used. Note that in Christensen *et al.* (2021), a polygenic effect whose covariance matrix is defined by the pedigree or/and genotypes is also included in Equation 1. This part is ignored here for simplicity, and was subtracted from the simulated phenotypes. Two patterns of missing omics data were considered: missing omics pattern: (1): all omics data are completely missing for some individuals; missing omics pattern; (2) for each omics feature, some random individuals have no omics data. The single-step approach only works in the scenario (1), while NN-LMM allows both scenarios. Different proportions of missing omics data in the training dataset were considered, where 0% denotes the scenario where all omics features are measured on all individuals. 20 replicates were used for each scenario. We randomly sampled 5% individuals from the simulated data in Christensen *et al.* (2021) to have a subset of 1,055 individuals. The genotypic data consisted of 15,000 SNP markers observed for all individuals, and the intermediate omics data consisted of 1,200 omics features. The heritability of each omics feature was 0.61, and the heritability of the phenotypic trait was 0.337. More details about the simulation process are in Christensen *et al.* (2021).

We also simulated nonlinear relationships between intermediate omics features and phenotypes, where the logistic non-linear transformation was applied to the omics data as in Equation 1. In this case, the same heritability and variance components were applied.

Results

When all omics features were measured on all individuals, NN-GBLUP had similar prediction accuracies as the system of two mixed model equations in Christensen *et al.* (2021) (correlation r=0.999). When some omics data were missing, NN-LMM had equivalent or better prediction performance, as shown in Figure 2. Overall, the prediction accuracy decreased when the proportion of missing omics data increased. For missing omics pattern (1), when a small proportion of individuals had no omics data, NN-GBLUP (red solid line) had similar prediction performance as the single-step approach in Christensen *et al.* (2021) (blue solid line). However, when a large proportion of individuals had no omics data (e.g. >80%), NN-GBLUP had significantly higher prediction accuracies (pairwise t-test P-value<0.005). When >90% individuals had no omics data, the single-step approach performed even worse than the baseline (black dashed line), which was a conventional GBLUP model where no omics data were used. For missing omics pattern (2), when some random individuals had no omics measures for each omics feature, the prediction accuracy of NN-GBLUP (red dashed line) decreased with larger proportion of missing omics data, and eventually close to the baseline, whereas the single-step approach did not work for this scenario.

When the underlying relationships between omics features and phenotype was nonlinear, results verified that using the nonlinear sigmoid activation function in NN-LMM was significantly better than using the linear activation function.

Discussion

To sample marker effects on omics features, a naive multi-threaded parallelism has been implemented to employ multiple single-trait mixed models in parallel at each MCMC iteration. Parallel computing strategies such as those in Zhao *et al.* (2020) will be further studied.

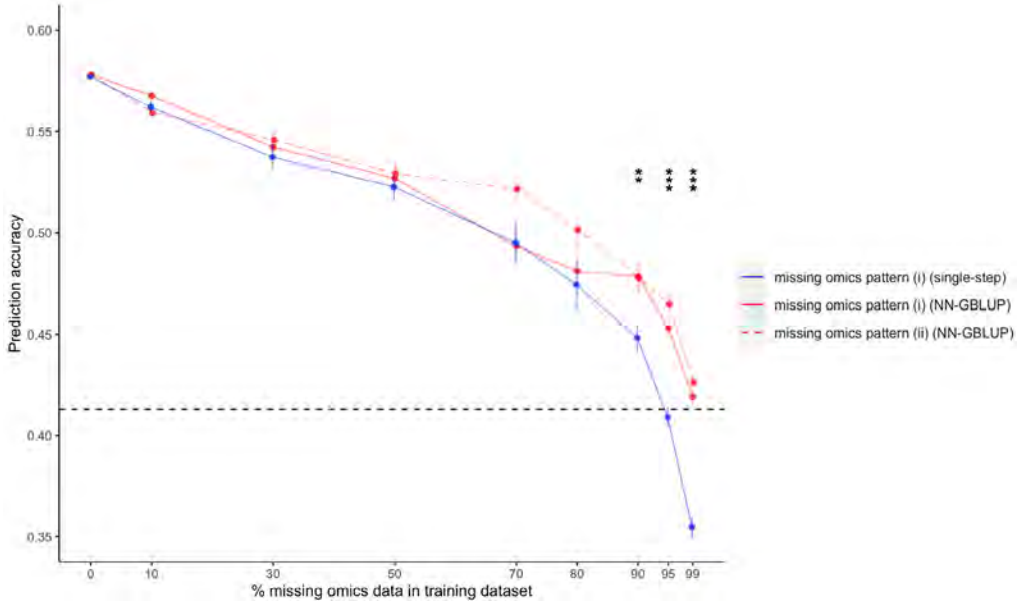


Figure 2. Prediction accuracies of NN-GBLUP with the linear activation function and the single-step approach in Christensen *et al.* (2021). The missing omics pattern (1): all omics features were not measured on some individuals, and missing omics pattern (2): for each omics feature, some random individuals had no omics measures. The horizontal black dashed linear is the conventional GBLUP model used as the baseline. Each dot represents the mean of prediction accuracies from 20 replications with its standard error. The asterisk symbol indicated a significantly better performance of NN-GBLUP.

References

Betancourt M. (2018) arXiv preprint. https://arxiv.org/abs/1701.02434

Christensen O.F., Börner V., Varona L., and Legarra A. (2021) Genetics 219(2):iyab130. https://doi.org/10.1093/genetics/iyab130

Kitano H. (2002) Nature 420(6912):206-210. https://doi.org/10.1038/nature01254

Ritchie M.D., Holzinger E.R., Li R., Pendergrass S.A., and Kim D. (2015) Nat. Rev. Genet. 16(2):85–97. https://doi.org/10.1038/nrg3868

Weishaar R., Wellmann R., Camarinha-Silva A., Rodehutscord M., and Bennewitz J. (2020) J. Anim. Breed. Genet. 137(1):14–22. https://doi.org/10.1111/jbg.12447

Wu Y., Zeng J., Zhang F., Zhu Z., *et al.* (2018) Nat. Commun. 9(1):1-14. https://doi.org/10.1038/s41467-018-03371-0

Zhao T., Fernando R., Garrick D., and Cheng H. (2020) Genet. Sel. Evol. 52(1):1-11. https://doi.org/10.1186/s12711-020-00533-x

Zhao T., Fernando R., and Cheng H. (2021a) G3. 11(10). https://doi.org/10.1093/g3journal/jkab228

Zhao T., Zeng J., and Cheng H. (2021b) BioRxiv preprint https://doi.org/10.1101/2021.12.10.472186

278. The holobiont 'predictome' of immuno competence in pigs

J. Calle-García[1*], Y. Ramayo-Caldas[2], L.M. Zingaretti[3], R. Quintanilla[2], M. Ballester[2] and M. Pérez-Enciso[1,4]

[1]Centre for Research in Agricultural Genomics (CRAG), 08193, Bellaterra, Spain; [2]Institute of Agrifood Research and Technology (IRTA), 08140, Caldes de Montbui, Barcelona, Spain; [3]Universidad Nacional de Villa María, Córdoba, Argentina. [4]ICREA 08010, Barcelona, Spain; joancallegarcia@gmail.com

Abstract

The objective of this work was to assess the ability of holobiont data (i.e. host's genotype and microbiota) to predict six immunity traits in 400 pigs. We propose the term 'predictome' to mean a systematic study of as many predictive methods as possible. With this spirit, we compare REML, Bayes C and Bayesian reproducing kernel Hilbert space (RKHS) regression with a wide range of priors. We assessed the performance of the models by partitioning the data into three disjoint sets. In total, we run 273 analyses per trait. We find that there does not exist a systematically best prediction method, although our results favor slightly Bayes C. By default, microbiota abundances should not be clustered. An holobiont model performs better than models using only genotype or microbiota data, i.e. use all data at your disposal.

Introduction

The role of gut microbiota composition in complex traits is well documented and it is currently a topic of utmost interest (Camarinha-Silva *et al.*, 2017; Difford *et al.*, 2018; Zhang *et al.*, 2020). However, there exist relatively little evidence on the actual advantage of combining genotype and microbial information for predictive purposes. Moreover, there is a plethora of statistical methods for prediction (Gianola, 2013), but we lack a systematic review on their actual behavior in real holobiont data. In this work, we perform a comprehensive holobiont prediction analysis for important traits related to the immune system in pigs. We call this analysis the 'predictome' to signify a comprehensive set of predictive tools and their performance.

Materials & methods

Data. We used genotype, phenotype and microbiota data fully described in Ballester *et al.* (2020), Ramayo-Caldas *et al.* (2021), except that here we included additional 16S sequence. Very briefly, data were available from 400 pigs for six immunocompetence traits: immunoglobulins IgM (IGM) and IgG (IGG), two acute-phase proteins, C-reactive protein (CRP) and haptoglobin (HP), gamma-delta T cells (GAMMADELTA) and lymphocyte phagocytic capacity (PHAGO_LYMPH). Data were corrected for fixed effects as described (Ballester *et al.*, 2020) and scaled. After quality control, 41,131 SNPs were retained from the original Porcine 70k GGP Porcine HD Array (Illumina).

Faecal 16S paired-end sequences were processed with QIIME2 (Bolyen *et al.*, 2019). Amplicon Sequence Variants (ASVs) were assigned using *DADA2* R package (Callahan *et al.*, 2016). ASVs present in less than three samples and representing less than 0.001% of the total counts were discarded. Several quality control measures were applied to ensure that 16S reads from both sequencing batches could be merged, e.g. both resulted in same Euclidean distances between samples and number of ASVs were congruent. The average number of reads per sample was over 136,000 and 2,945 ASVs were retained. We studied the effect in the prediction of clustering by phylogeny at the genus level and by abundance vs using individual ASV abundances. MEGA package and NJ algorithm was used for genus clustering, and *hclust* R package for abundance clustering. In either approach, 232 clusters of ASVs were obtained.

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_278, © J. Calle-García *et al.* 2022

Statistical analyses. We applied a wide range of models and algorithms to characterize the predictive ability for immunocompetence in pigs. Prediction was measured as the correlation between predicted and observed phenotype averaged over three partitions, each containing ~20% of different records. Three models were compared: a holobiont model that includes genotype (X) and abundance data (B), a genetic model fitting only genotypes and a microbiota model with abundance data only. For B, the ASVs were fitted individually or grouped by phylogeny or by abundance.

We considered three statistical methods: REML, Bayesian RKHS regression (equivalent to GBLUP), and Bayes C. REML was run using ASREML and Bayesian methods, using BGLR (Pérez and de los Campos 2014). In RKHS, a pseudo flat prior and the default mildly informative prior were compared. In Bayes C, the following prior probabilities of a variable entering the model were compared P=0.01, 0.001 and 0.0001 for X and P=0.1, 0.01 and 0.001 for B. Combinations of all models, algorithms, priors and clustering approaches were run for each partition and phenotype. In total, 273 analyses per phenotype were run.

Results

Frequentist or Bayes? Bayes C or GBLUP? Performance in prediction is an objective, pragmatic, and easy way of contrasting statistical approaches. Figure 1 shows that REML is not necessarily worse than Bayesian methods, except in lymphocyte phagocytic capacity. The method chosen was not too important except in some phenotypes like PHAGO_LYMPH, where a Bayes C modelling for both X and B outperformed other options, or haptoglobin where GBLUP methods prevailed.

Is it worth using holobiont data? In a recent simulation study (Pérez-Enciso *et al.* 2021), we argued that using an holobiont model could significantly increase predictive accuracy, provided both X and B contribute to phenotypic variance, and that a holobiont model was the best default option. Our results (Figure 2) confirm this conjecture, and we observed that B can improve prediction by a sizable margin, e.g. for HP. For CRP and PHAGO_LYMPH, the combined model X+B increased correlation by ~25% compared to models without B.

Figure 1. Predictive accuracy of the holobiont model by statistical method and trait across priors and ASV clustering. For each analysis combination, the first term (REML, RKHS, BayesC) refers to the algorithm modelling the genotype (X) and the second to the algorithm modeling the microbiota (B). For instance, RKHS.BayesC means genotype modeled as RKHS and abundances as Bayes C.

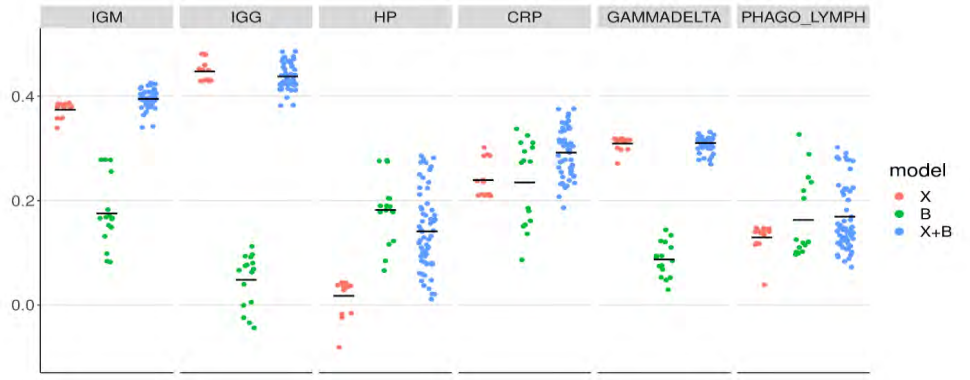


Figure 2. Correlation between predicted and observed phenotypes with genotype (X), microbiota (B) and holobiont (X+B) models. Each dot represents one analysis with different algorithm, prior and microbiota clustering combinations. The horizontal black line is the mean.

Should microbial abundances be grouped? Microbial abundances display highly leptokurtic distributions that require careful quality control, filtering and transformation steps. An option to minimize risks is to group them. Is that useful for predictive purposes? Again, there is not a single best option, although clustering worsens performance for most traits (Figure 3). Clustering by abundance yielded a 50% increase in HP prediction correlation, and clustering by phylogeny increased CRP results by 15%.

Discussion

The term 'hologenome' (Zilber-Rosenberg and Rosenberg 2008) was coined to describe the joint action of the genome and of the microbiota on a phenotype. We confirm that microbiota data can improve prediction, sometimes by a large margin, in important phenotypes related to immunocompetence. Here we systematically explored the optimal statistical approaches for prediction, what we have called 'predictome'. In all, we can conclude the following:

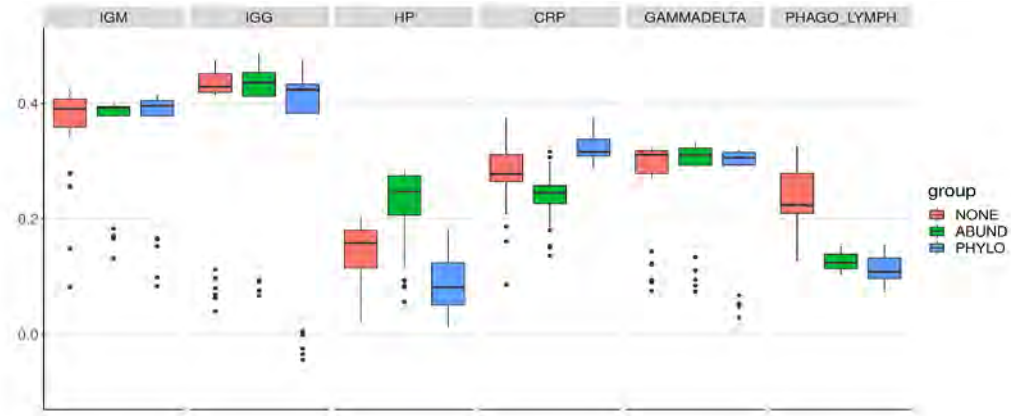


Figure 3. Microbiota clustering effect on prediction. 'NONE': no clustering; 'PHYLO': clustering at the genus level; 'ABUND': clustering by abundance.

1. There does not exist a systematically best prediction method, although our results favor Bayes C. Don't feel culprit though if you choose REML or Bayesian methods.
2. A holobiont model tends to perform better than either genotype or microbiota only models, i.e. use all data at your disposal.
3. By default, we recommend not clustering microbial abundances, despite their highly extreme distributions.

The explosion of statistical methods for holobiont based prediction has neither a clear 'winner' nor a clear 'loser'.

References

Ballester, M., Ramayo-Caldas, Y., González-Rodríguez, O., Pascual, M., *et al.* (2020). *Scientific Reports*, *10*(1), 1–15. https://doi.org/10.1038/s41598-020-75417-7

Bolyen, E., Rideout, J.R., Dillon, M.R., Bokulich, *et al.* (2019). *Nature Biotechnology*, *37*(8), 852–857. https://doi.org/10.1038/s41587-019-0209-9

Callahan, B.J., McMurdie, P.J., Rosen, M.J., Han, A.W., Johnson, *et al.* (2016). *Nature Methods*, *13*(7), 581–583. https://doi.org/10.1038/nmeth.3869

Camarinha-Silva, A., Maushammer, M., Wellmann, R., Vital, M., Preuss, S., *et al.* (2017). *Genetics*, *206*(3), 1637–1644. https://doi.org/10.1534/genetics.117.200782

Difford, G.F., Plichta, D.R., Løvendahl, P., Lassen, J., Noel, S.J., *et al.* (2018). *PLoS Genetics*, *14*(10), e1007580. https://doi.org/10.1371/journal.pgen.1007580

Gianola, D. (2013). *Genetics*, *194*(3), 573–596. https://doi.org/10.1534/genetics.113.151753

Pérez-Enciso, M., Zingaretti, L.M., Ramayo-Caldas, Y., & de los Campos, G. (2021). *Genetics, Selection, Evolution: GSE*, *53*(1), 1–20. https://doi.org/10.1186/s12711-021-00658-7

Pérez, P., & de los Campos, G. (2014). *Genetics*, *198*(2), 483–495. https://doi.org/10.1534/genetics.114.164442

Ramayo-Caldas, Y., Zingaretti, L.M., Pérez-Pascual, D., Alexandre, P.A., Reverter, A., *et al.* (2021). *Animal Microbiome*, *3*(1), 1–11. https://doi.org/10.1186/s42523-021-00138-9

Zhang, Q., Difford, G., Sahana, G., Løvendahl, P., Lassen, J., *et al.* (2020). *The ISME Journal*, *14*(8), 2019–2033. https://doi.org/10.1038/s41396-020-0663-x

Zilber-Rosenberg, I., & Rosenberg, E. (2008). *FEMS Microbiology Reviews*, *32*(5), 723–735. https://doi.org/10.1111/j.1574-6976.2008.00123.x

279. The long and short of Nanopore genomic prediction: the effect of read length on prediction accuracy

H.J. Lamb[1*], L.T. Nguyen[1], I.A.S. Randhawa[2], B.J. Hayes[1] and E.M. Ross[1]

[1]Queensland Alliance for Agriculture and Food Innovation, Centre for Animal Science, 306 Carmody Rd, 4067, St Lucia, Australia; [2]The University of Queensland, School of Veterinary Science, Building 8156 Outer Ring Rd, 4343, Gatton, Australia; harrison.lamb@uq.edu.au

Abstract

On-farm genomic selection using Oxford Nanopore Technologies' portable DNA sequencer is a potential method to accelerate genetic gain in extensive agricultural enterprises. Oxford Nanopore sequencing provides a number of potential benefits including no theoretical read length limit. Here we investigate the effect of read length on the accuracy of genomic prediction from Oxford Nanopore data by creating a range of different read length distributions with average read lengths ranging between 2,569 bp and 110 bp. Read length was found to have limited effect on the accuracy of genomic prediction for the tested range. However, average read lengths below 500 bp demonstrated a significant increase in unmapped reads and reduced sequencing coverage. Importantly, this means to maximise data recovery from Nanopore flowcells, on-farm DNA libraries should have average read lengths greater than 500 bp. This finding will inform the requirements of the on-farm molecular sample preparation, such as DNA extraction and library preparation.

Introduction

Oxford Nanopore Technologies' (ONT) MinION is a third-generation sequencing platform that uses a biological nanopore to interpret DNA sequences. The disruption in ionic current caused by nucleotides passing through the nanopore is interpreted using software to produce a sequence read (Jain *et al.* 2016). This method has two major benefits: firstly, data can be interpreted in real-time and secondly, there is no theoretical limit to the length of sequence reads (Jain *et al.* 2016). Instead, read length is determined by the type of biological sample used and the level of shearing during DNA extraction and library preparation. Methods which subject the DNA to increased levels of mechanical shearing result in more fragmented DNA libraries and shorter average read lengths. Generally, rapid DNA extraction protocols result in more fragmented DNA libraries as a result of the techniques used to speed up bottlenecks.

Long read lengths are able to map with greater specificity to reference genomes and therefore more accurately characterise complex genomic regions. We investigated the effect of read length on genomic prediction accuracy in cattle from ONT sequence data.

Materials & methods

Simulating read distributions. Sixteen Droughtmaster heifers were sequenced from tail hair on the MinION sequencing platform as described in (Lamb *et al.* 2021). The sequence data was sampled down to 0.5x sequencing coverage for each animal. All of the sequencing reads in each of the subsequent fastq files were then cut into halves, quarters and eighths, to create a range of read lengths for each sample. Each read length replicate was then assigned a read length group based on average read length: 0-499 bp, 500-999 bp, 1000-1,499 bp, 1,500-1,999 bp and >2,000 bp. A fifth distribution was also created to replicate the effect of Nanopore adaptive sampling. Nanopore adaptive sampling is a sequence enrichment method that uses software to enrich defined genomic regions. A by-product of this technique is the creation of a number of short ~450 bp read fragments, corresponding to non-target regions (Oxford Nanopore Technologies 2020). To simulate enrichment of 20% of the bovine genome, 80% of reads were fragmented into 450 bp read

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_279, © H.J. Lamb *et al.* 2022

fragments, while the remainder of reads were left whole. This produced a final read distribution referred to as the adaptive distribution and a total of 80 unique samples (each of the 16 samples having five different read length distributions).

Genotyping and imputation. The reads from each simulated distribution were aligned to the ARS-UCD v1.2 *Bos taurus* reference genome (Rosen *et al.* 2020) using Minimap2 v1.2 (Li 2018). Genotypes for the Bovine HD chip were imputed using QUILT v1.01 (Davies *et al.* 2021) and a reference panel of ~1,200 animals which represented a subset of animals from the 1000 bull genomes project (Hayes and Daetwyler 2019).

Estimating breeding values. Genomic estimated breeding values (GEBVs) for four traits: body weight (BW), hip height (HH), body condition score (BCS) and corpus luteum score (CL score) were then calculated using marker effects estimated in Hayes *et al.* (2019), as described in *Lamb et al. (2021).*

Read length linear models. Correlations between the Nanopore GEBVs and SNP array GEBVs were then calculated as well as prediction bias, the regression of SNP array GEBV ~ Nanopore GEBV using

$$\hat{g} \sim \hat{a} + e \quad (1)$$

Where $\hat{g}$ was a vector of SNP array GEBVs and $\hat{a}$ a vector of Nanopore GEBVs. Linear models were used to evaluate the effect of read length on GEBV accuracy. Read length was modelled as a continuous covariate using:

$$y \sim x + e \quad (2)$$

Where x was a vector of average read length values as a random effect and y was a vector of residuals from the regression of SNP array GEBV on Nanopore GEBV.

Results

The average read length for each sample ranged between 3,065 bp and 185 bp and the average read length for the adaptive distribution was 812 bp. The correlations between SNP array GEBV and Nanopore GEBV were between 0.92-0.99, 0.82-0.99, 0.95-0.99 and 0.90-0.99 for BW, HH, CL score and BCS respectively (Figure 1).

The prediction bias ranged between 0.77-1.04 for the adaptive read length distribution 1.03-1.13 for the smallest read length bin and 1.05-1.19 for the largest read length bin (Table 1). The largest prediction bias was seen in the adaptive (overprediction of BCS) and 1500-1999 read length distributions (underprediction of BW).

Read length had no effect (P>0.05) on the regression of SNP array GEBV ~ Nanopore GEBV when fitted as a continuous covariate nor as a categorical covariate using the six distributions. Read length did however have a significant effect (P<0.05) on the fraction of reads unmapped and the sequencing coverage (Figure 2). Average read lengths less than 500 bp had more unmapped reads and less sequencing coverage than larger average read lengths. Despite a similar fraction of unmapped reads, the adaptive distribution had far less mapped bases than the 0-499 bp distribution. This was caused by a larger average read length amongst the unmapped reads in the adaptive distribution than the 0-499 bp distribution. Despite the decrease in effective sequencing coverage, and no actual enrichment in the adaptive samples, no decrease in GEBV accuracy was observed.

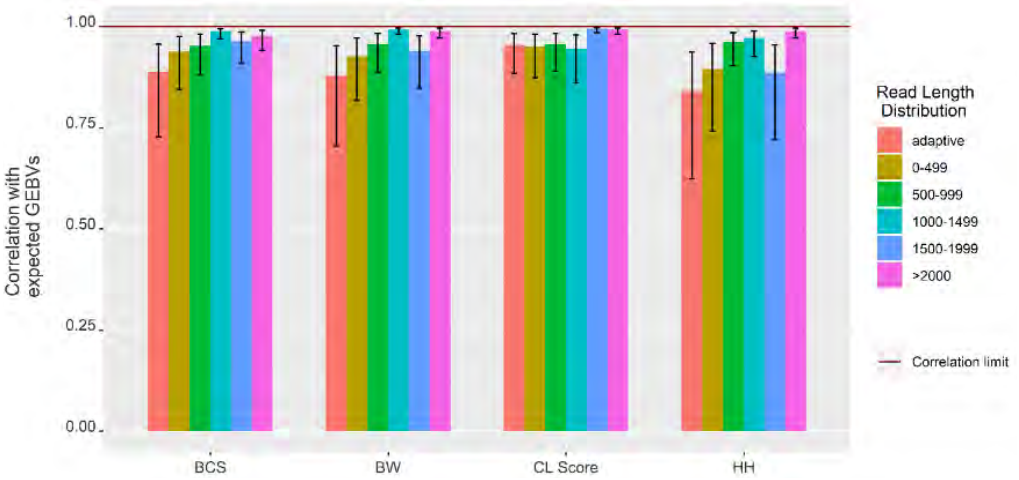


Figure 1. Corelations between SNP array GEBVs and Nanopore GEBVs for each of the 4 traits and the six different read length distributions.

Table 1. Prediction bias for each trait across the read length distributions.

Read length distribution	Body weight	Hip height	CL score	Body condition score
Adaptive	0.88	0.77	0.91	1.04
0-499	1.13	1.03	1.05	1.03
500-999	1.07	0.95	1.15	1.03
1000-1,499	1.09	1.05	0.91	1.04
1,500-1,999	1.21	0.97	0.87	1.13
>2,000	1.06	1.19	1.03	1.12

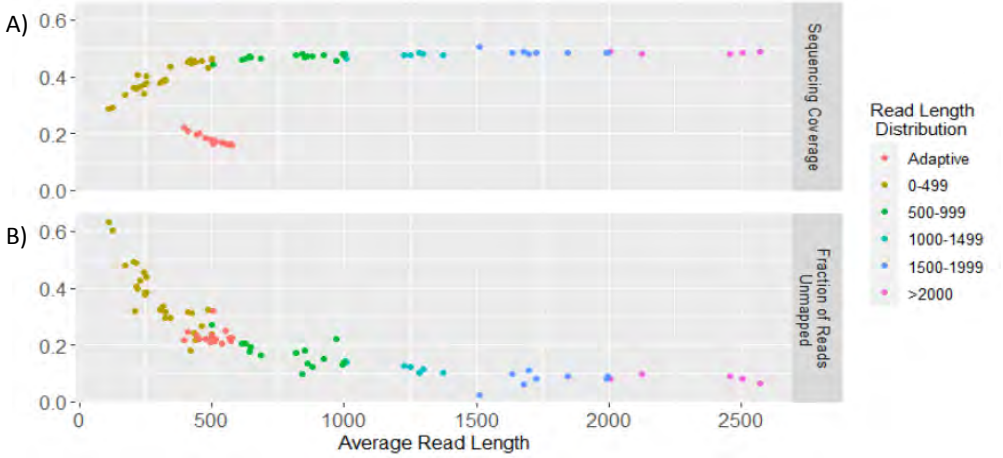


Figure 2. (A) Sequencing coverage plotted against the average read length for each analysis. (B) Fraction of reads unmapped plotted against average read length for each analysis.

Discussion

Despite average read lengths less than 500 bp having an effect on sequencing coverage, there was no effect on correlation or prediction bias. This suggests that despite losing data to unmapped reads, QUILT is still able to accurately impute genotypes from sequencing coverages as low as 0.2×.

Although 0.2× sequencing coverage was sufficient here for accurate genomic prediction, the decrease in sequencing coverage for average read lengths less than 500 bp will have implications when minimising per sample costs. For example, where samples are sequenced to 0.2× sequencing coverage, average read lengths less than 500 bp max cause the effective sequencing coverage to drop below the level required for accurate genomic prediction using QUILT. Therefore, the average read length for genomic prediction should be greater than 500 bp. This is shorter than average read lengths reported by a number of field sequencing studies with (Maestri *et al.* 2019) and without PCR amplification (Goordial *et al.* 2017).

A contributing factor to these results is that the positions of the bovine HD SNPs used in this study are primarily located in genomic regions with low complexity. This means even short read lengths (<500 bp) are able to map accurately for imputation from a large reference panel. If SNP effects in complex genomic regions are to be used in the future, longer read lengths may be required.

Overall, these results indicate that average read lengths greater than 500 bp from ONT sequencing will be sufficient for the purpose of calculating GEBVs. In the context of on-farm genomic predictions this means that little consideration towards read length will be required during DNA extraction and library preparation. Therefore, turnaround time can be prioritised without losing prediction accuracy from fragmented DNA.

References

Davies R.W., Kucka M., Su D., Shi S., Flanagan M. *et al.* (2021) Nat Genet (53):1104-1111. https://doi.org/10.1038/s41588-021-00877-0

Goordial J., Altshuler I., Hindson K., Chan-Yam K., Marcolefas E. *et al.* (2017) Front Microbiol (8):2594. https://doi.org/10.3389/fmicb.2017.02594

Hayes B.J., Daetwyler H.D. (2019) Annual Review of Animal Biosciences, Vol 7 (7):89-102. https://doi.org/10.1146/annurev-animal-020518-115024

Hayes B.J., Fordyce G., Landmark S. (2019) Genomic Predictions for fertility traits in tropical beef cattle from a multi-breed, crossbred and composite referecne population. In 'Australian Association for Animal Breeding and Genetics. Armidale'. Available at http://www.aaabg.org/aaabghome/AAABG23papers/69Hayes23282.pdf

Jain M., Olsen H.E., Paten B., Akeson M. (2016) Genome Biol (17):239. https://doi.org/10.1186/s13059-016-1103-0

Lamb H.J., Hayes B.J., Randhawa I.A.S., Nguyen L.T., Ross E.M. (2021) Plos One (16):e0261274. https://doi.org/10.1371/journal.pone.0261274

Li H. (2018) Bioinformatics (34):3094-3100. 10.1093/bioinformatics/bty191

Maestri S., Cosentino E., Paterno M., Freitag H., Garces J.M. *et al.* (2019) Genes (Basel) (10), https://doi.org/10.3390/genes10060468

Oxford Nanopore Technologies (2020) 'Nanopore Info sheet: Adaptive sampling.' Available at https://community.nanoporetech.com/info_sheets/adaptive-sampling/v/ads_s1016_v1_revd_12nov2020

Rosen B.D., Bickhart D.M., Schnabel R.D., Koren S., Elsik C.G. *et al.* (2020) Gigascience (9), https://doi.org/10.1093/gigascience/giaa021

280. Accuracy of genomic prediction of dry matter intake in Dutch Holsteins using sequence variants from meta-analyses

B. Gredler-Grandl[1*], B. Raymond[1], G.C.B. Schopen[1], P.K. Chitneedi[2], Z. Cai[3], C.I.V. Manzanilla-Pech[3], T. Iso-Touru[4], D. Fischer[4], S. Bolormaa[5], T.S. Chud[6,7], F.S. Schenkel[6], Y. Wang[8,9], C. Li[8,9], D. Hailemariam[8,9], B. Villanueva[10], A. Fernandez[10], C. Kuehn[2], G. Sahana[3], M.H. Lidauer[4], J.E. Pryce[5,11], O. González-Recio[10], G. Plastow[8], C.F. Baes[6,12], N. Charfeddine[13], Y. de Haas[1], R.F. Veerkamp[1] and A.C. Bouwman[1]

[1]Wageningen University & Research, P.O. Box 338, 6700 AH Wageningen, the Netherlands; [2]Institute for Genome Biology, Research Institute for Farm Animal Biology (FBN), Wilhelm-Stahl-Allee 2, 18196 Dummerstorf, Germany; [3]Center for Quantitative Genetics and Genomics, Aarhus University, Blichers Alle 20, 8830 Tjele, Denmark; [4]Natural Resources Institute Finland (Luke), Myllytie 1, 31600 Jokioinen, Finland; [5]Agriculture Victoria, AgriBio, Centre for AgriBioscience, 5 Ring Road, Bundoora, Victoria, 3083, Australia; [6]University of Guelph, Centre for Genetic Improvement of Livestock, Guelph, ON N1G 2W1, Canada; [7]Fast Genetics, Saskatoon, SK S7K 2K6, Canada; [8]Department of Agriculture, Food & Nutritional Science, University of Alberta, Edmonton, Alberta, Canada; [9]Lacombe Research and Development Centre, Agriculture and Agri-food Canada, Lacombe, Alberta, Canada; [10]Departamento de Mejora Genética Animal, Instituto Nacional de Investigación y Tecnología Agraria y Alimentaria, INIA-CSIC Crta. de la Coruña km 7.5, 28040 Madrid, Spain; [11]School of Applied Systems Biology, La Trobe University, Bundoora, Victoria, 3083, Australia; [12]University of Bern, Institute of Genetics, Vetsuisse Faculty, Bern 3001, Switzerland; [13]Spanish Holstein Association (CONAFE), Ctra. De Andalucía km 23600 Valdemoro, 28340 Madrid, Spain; birgit.gredler-grandl@wur.nl

Abstract

We evaluated the accuracy of biology informed genomic prediction for dry matter intake in 2,162 Dutch Holstein cows. Sequence variants were selected from meta-analyses including GWAS summary statistics for QTL and metabolomic QTL in several dairy and crossbred beef populations. Selected variants were prioritized in GBLUP models in a five-fold cross-validation. The accuracies were compared to genomic prediction based on routine 50k genotype data. The average accuracy for the 50k scenario was 0.683. Adding selected sequence variants in the GBLUP model did not improve the accuracies for dry matter intake. Next steps will include testing Bayesian variable selection methods to prioritize variants in genomic prediction for dry matter intake.

Introduction

Feed intake is an important trait, but large scale recording schemes on individual animals are costly and hamper implementations of genetic evaluations for dry matter intake (DMI). High prediction accuracies are difficult to achieve for traits like DMI, because the size of the reference population is often limited. Raymond *et al.* (2018) have shown a way to increase accuracy of genomic prediction (GP) in a multibreed evaluation when prioritising sequence variants from a meta-GWAS separate from the routine array genotypes. Additionally, functional information to pinpoint QTL regions may improve predictions. Metabolites in blood plasma represent intermediate phenotypes between the genome and transcriptome and final expressed phenotypes. Here, we evaluate the accuracy of biology informed GP for DMI in Dutch Holstein cows by prioritising variants from various meta-GWAS analyses.

Materials & methods

Meta-GWAS for QTL. To increase power of QTL detection for DMI, a meta-GWAS based on local GWAS results of eight populations was carried out using the METAL software package (Willer *et al.*, 2010). Local GWAS results for imputed sequence variants were available for five different Holstein populations in

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_280, © B. Gredler-Grandl *et al.* 2022

Australia (584 cows), Canada (588 cows), Denmark (495 cows), Germany (140 cows) and Spain (561 cows), one Finnish Red population from Finland (366 cows) and two crossbred beef populations in Germany (253 bulls of a Charolais × Holstein cross) and Canada (7,552 heifers and steers of Angus, Charolais, Kinsella Composite and crossbred animals thereof). Phenotypes used in the local GWAS were either de-regressed breeding values (DRBV) or raw phenotypes for DMI corrected for fixed effects across lactation. Variant effects and standard errors of the effect from the local GWAS were standardized based on the genetic standard deviation for DMI for each population. Variants with an imputation accuracy r^2<0.6, a minor allele frequency (MAF)<0.001 or an effect size >5 standard deviations apart from the mean were not considered in the meta-analysis. Three different meta-GWAS were carried out: a meta-analysis including all populations (ALL), Holstein populations only (HOL) and beef populations only (BEEF). The total number of sequence variants considered were 30,216,688, 19,647,876 and 27,839,929 for ALL, HOL and BEEF, respectively.

Meta-GWAS for mQTL. GWAS summary statistics of imputed sequence variants from three cattle populations (Holstein, Charolais × Holstein, mixed beef breed composite) were available for segregating blood plasma metabolomic QTL. Metabolites considered were amino acids, short and long-chain fatty acids and compounds from energy and protein metabolism functions. The METAL software package (Willer *et al.*, 2010) was used for the meta-analysis. Variant effects were standardized by the genetic standard deviation of plasma metabolite concentration. Variants with MAF<0.01 and/or imputation r^2<0.6 were discarded. The number of animals contributing to the meta-GWAS varied across metabolites between 241 and 1,103. The number of variants considered across metabolites were between 14,343,591 and 19,467,841.

SNP selection scenarios. To filter out non-causal and select independently associated variants from the QTL meta-analysis, a forward selection of variants based on the conditional and joint effect method (COJO) described in Yang *et al.* (2012) was conducted. Variants were selected using the following COJO model parameters: conditional and joint *P*-value threshold to declare a genome-wide significant variant of *P*=5e-3 (scencarios ALL3, HOL3, BEEF3) or *P*=5e-5 (scenarios ALL5, HOL5, BEEF5), collinearity between selected markers of 0.9 and window size of 10 Mb. *P*-value significance thresholds were chosen to achieve a reasonable number of selected variants for GP. The mQTL meta-GWAS resulted in 20,426 trait associated SNPs with $P<10^{-6}$ for 27 metabolites. Of those, all variants segregating in the target population for GP (see below) were used in the mQTL scenario. For the base scenario (50k) variants available on the Illumina Bovine snp50 v3 beadchip (Illumina Inc., San Diego, CA, USA) were selected from the imputed sequence genotypes of the target population for GP. The number of selected variants per scenario is shown in Table 1. Selected variants were used to derive the genomic relationship matrices (GRM) for estimation of heritability and genomic breeding values (GEBV) for DMI.

Genomic prediction. An independent data-set of 2,162 Dutch Holstein cows imputed to sequence, and with DRBV

for DMI with reliability ≥0.30 were used in GP models. For the base scenario (50k) and all scenarios where only one GRM was fitted the following model was run in the mtg2 software (Lee, 2016):

$$y = 1\mu + Wg + e \quad (1)$$

where y is a vector of DRBV for DMI for the reference set (missing for the validation set), 1μ is the overall mean, g is a vector of additive genetic effects for all cows, W is a design matrix linking g to DRBV in y and e is a vector containing residuals. g and e are assumed to be normally distributed with $g = N(0, GRM\sigma_g^2)$ and $e = N(0, I\sigma_e^2)$.

Table 1. Number of variants to build the GRM and estimated heritability of DRBV in all scenarios (average across 5 cross-validations).

Scenario	Number of variants	Heritability	SE
50k	37,179	0.849	0.024
ALL3	1,810	0.456	0.032
ALL5	166	0.142	0.026
HOL3	1,834	0.442	0.032
HOL5	63	0.086	0.026
BEEF3	1,746	0.473	0.032
BEEF5	321	0.240	0.030
mQTL	17,056	0.675	0.034
50k + ALL3	37,179 + 1,810	0.796[1] + 0.060[1]	0.037 + 0.030
50k + ALL5	37,179 + 166	0.853 + <0.001	0.025 + 0.007
50k + HOL3	37,179 + 1,834	0.825 + 0.026	0.034 + 0.026
50k + HOL5	37,179 + 63	0.849 + <0.001	0.025 + 0.005
50k + BEEF3	37,179 + 1,746	0.854 + 0.004	0.036 + 0.028
50k + BEEF5	37,179 + 321	0.858 + <0.001	0.026 + 0.012
50k + mQTL	37,179 + 17,056	0.844 + 0.008	0.025 + 0.008

[1] The first and second number represent the heritability estimates for 50k and the respective sequence variant set, respectively.

For scenarios with two different GRM (50k variants and selected sequence variants) the following model was used:

$$y = 1\mu + W_1 g_1 + W_2 g_2 + e \quad (2)$$

where subscripts 1 and 2 represent the first (base 50k) and second (selected sequence variants) GRM fitted, respectively, W_1 and W_2 are identical design matrices linking DRBV to the two genetic effects g_1 and g_2. As above, g and e are assumed to be normally distributed.

A five-fold cross validation was carried out. 400 of the 2,162 cows were randomly selected five times as validation and the remainder (1,762 cows) comprised the reference set. Accuracy of GP was measured as the correlation between the GEBV and DRBV in the validation set. The prediction bias was assessed by regressing DRBV of the validation cows on their GEBV.

Results

The heritability estimated in this study is based on DRBV and can therefore be interpreted as the explained genetic variance of DRBV. The estimates shown in Table 1 are similar between the 50k and all scenarios fitting two GRMs. When using selected sequence variants only, heritability ranged between 0.086 (HOL5) and 0.675 (mQTL).

The accuracy of GP for all scenarios is presented in Table 2. All scenarios fitting two GRMs resulted in similar accuracy as in the 50k scenario only. When using selected variants only, the highest accuracy was observed when using variants from the beef meta-GWAS (BEEF3) and using all populations (ALL3). GEBV were overestimated in all scenarios fitting two GRMs. GEBVs in ALL5 and HOL5 were least biased.

Table 2. Accuracy and slope of the regression of DRBV on GEBV for all genomic prediction scenarios (average across 5 cross-validations).

Scenario	Accuracy	SE	Slope
50k	0.683	0.027	0.909
ALL3 only	0.557	0.035	0.957
ALL5 only	0.363	0.043	1.014
HOL3 only	0.549	0.035	0.973
HOL5 only	0.204	0.048	0.992
BEEF3 only	0.568	0.034	0.983
BEEF5 only	0.469	0.039	1.068
mQTL only	0.475	0.039	0.914
50k + ALL3	0.686	0.027	0.936
50k + ALL5	0.682	0.027	0.932
50k + HOL3	0.683	0.027	0.933
50k + HOL5	0.683	0.028	0.932
50k + BEEF3	0.653	0.029	0.935
50k + BEEF5	0.682	0.027	0.932
50k + mQTL	0.682	0.027	0.931

Discussion

We present accuracies of GEBV for DMI for different sequence variant selection scenarios. No clear advantage in accuracy could be found when adding relevant sequence variants in GBLUP models fitting two GRMs. Reasons may be related to the genetic architecture of the trait DMI in Holstein. Furthermore, variants selected in the different scenarios have generally a very low MAF and imputation errors and linkage disequilibrium can influence the result of GP. Here, we used GBLUP for estimation of GEBV. Bayesian variable selection methods may potentially result in higher accuracy. Further studies will be undertaken in the near future testing Bayesian variable selection methods in a larger Dutch Holstein population and across dairy and beef breeds with DMI phenotypes.

Acknowledgements

This work is carried out within the BovReg project, which has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement no. 815668, and the global dry matter initiative gDMI. Canadian authors acknowledge additional funding support from Genome Alberta and Alberta Agriculture and Forestry, project GAB-A3GP37, as well as funders, partners and collaborators involved in the Resilient Dairy Genome Project (http://www.resilientdairy.ca).

References

Lee S.H, and Van der Werf J.H (2016) Bioinformatics 32(9):1420-1422. https://doi.org/10.1093/bioinformatics/btw012

Raymond B, Bouwman A.C., Wientjes Y.C.J., Schrooten C., Houwing-Duistermaat J. *et al.* (2018) Genet. Sel. Evol. 50:49. https://doi.org/10.1186/s12711-018-0419-5

Willer C.J., Li Y., and Abecasis G.R. (2010) Bioinformatics 26(17):2190-2191. https://doi.org/10.1093/bioinformatics/btq340

Yang J., Ferreira T., Morris A.P., Medland S.E., Madden P.A., *et al.* (2012) Nat. Genet. 44(4):369-375. https://doi.org/10.1038/ng.2213

281. Genomic prediction of bull fertility in Italian Brown Swiss cattle

H.A. Pacheco[1*], M. Battagin[2], A. Rossoni[2], A. Cecchinato[3] and F. Peñagaricano[1]

[1]Department of Animal and Dairy Sciences, University of Wisconsin-Madison, Madison, WI 53706, USA; [2]Italian Brown Breeders Association, Bussolengo, Verona 37012, Italy; [3]Department of Agronomy, Food, Natural Resources, Animals and Environment, University of Padova, Legnaro, Padua 35020, Italy; hpacheco@wisc.edu

Abstract

Bull fertility has been recognized as an important factor influencing reproductive success in dairy cattle. The objective of this study was to assess the feasibility of predicting bull fertility in Italian Brown Swiss cattle using genomic data. Dataset consisted of 1,102 Italian Brown Swiss bulls with sire conception rate records and 480k single nucleotide polymorphisms (SNP). The analyses included linear and Gaussian kernel-based models fitting either all SNP or also including markers with large effect. Predictive performance was evaluated using five-fold cross-validation. The entire SNP set exhibited predictive correlations around 0.19. Interestingly, the inclusion of two major non-additive markers as fixed effects achieved predictive correlations around 0.32. Therefore, the inclusion of markers with large effect markedly improved the prediction of bull fertility. Our findings have the potential to help the dairy industry to improve Brown Swiss sire fertility through accurate genome-guided decisions.

Introduction

Improving reproductive performance remains one of the most important goals for the dairy industry. Service sire has been recognized as an important factor affecting herd fertility. Recent studies have shown that the reduced ability of low-fertility bulls to establish pregnancy is multifactorial, including not only the fertilization process but also early embryo development (Ortega *et al.* 2018). Interestingly, there are substantial differences in sire conception rate (SCR) between high- and low-fertility bulls, with 10-20% conception rate differences between extreme bulls (Rezende *et al.* 2018; Pacheco *et al.* 2021). Therefore, it is clear that bull fertility should be considered in breeding schemes aimed to improve reproductive efficiency.

There is growing evidence that dairy bull fertility is influenced by genetic factors. Indeed, our team has identified individual genes and gene-sets strongly associated with SCR in both Holstein and Jersey bulls. We have also reported promising results for predicting SCR values using genomic data in these two breeds. However, to the best of our knowledge, no study to date has explored the feasibility of predicting service sire fertility in Brown Swiss cattle using genomic data. Therefore, our first objective was to assess the prediction of bull fertility in the Italian Brown Swiss cattle population using 480k SNP markers spanning the entire genome. Our second objective was to evaluate the potential benefits of incorporating markers with large effect into the genomic prediction models.

Materials & methods

Phenotypic and genotypic data. Phenotypic data consisted in sire conception rate (SCR) records from a total of 1,102 Italian Brown Swiss bulls. Sire conception rate records were calculated using cow field data, considering factors related to the bull under evaluation and factors (nuisance variables) associated with the cow that receives the unit of semen (Pacheco *et al.* 2021). Sire conception rate values ranged from -22.3 to 9.9%, and the number of breedings ranged from 50 to 8,110. Genotype data for 481,839 single nucleotide polymorphism (SNP) markers located in autosomal chromosomes were available for all the 1,102 Brown Swiss bulls with SCR records, kindly provided by the Italian Brown Breeders' Association (Anarb; Italy).

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_281, © H.A. Pacheco *et al.* 2022

Whole genome scan. The relevance of non-additive effects, namely dominance, recessive and overdominance effects, on service sire fertility was evaluated on a genome-wide scale using a two-step approach. In the first step, the following model was fitted, y = Xb + Zu + e, where y is the vector of SCR records, b is the vector of fixed effects, u is the vector of random animal effects, and e is the vector of random residual effects. The random effects were assumed to follow a multivariate normal distribution with $u \sim N(0, G\sigma_u^2)$ and $e \sim N(0, I\sigma_e^2)$, where σ_u^2 and σ_e^2 are the animal additive genetic and residual variances respectively; G is the genomic relationship matrix, and I an identity matrix. In the second step, the following model was fitted for every SNP, $y = X\beta + X_{SNP}\beta_{SNP} + \epsilon$, where X_{SNP} is the design matrix for the genetic marker under study and β_{SNP} is the regression coefficient or SNP effect. Every SNP genotype was coded using single numeric variables as (0, 1, 1), (0, 0, 1) and (0, 1, 0) for testing dominance, recessive and overdominance effects, respectively. All these analyses were performed using the R package MixABEL (Aulchenko *et al.* 2007).

Genomic prediction. The predictive power of either the entire SNP dataset or the alternative model, which includes markers with large effect as fixed effects, was evaluated using reproducing kernel Hilbert spaces (RKHS) regression models (Morota and Gianola 2014). The analyses were implemented under the kernel-based regression model, y = Xb + Kα + e, where y is the vector of phenotypic records (SCR values); b is the vector of fixed effects including a general intercept (μ); X is the design matrix relating fixed effects to SCR records; K is an $n \times n$ kernel matrix indexed by the SNP genotype matrix; α is the vector of RKHS regression coefficients and *e* is the is the error term. Kernel-based genomic prediction models were fitted using either linear or Gaussian reproducing kernels (K). The linear (K_L) kernel is equivalent to the well-known additive genomic relationship matrix. On the other hand, the Gaussian (K_G) kernel can capture nonadditive effects that might be important for phenotypic prediction of bull fertility. The models were implemented in a Bayesian framework using Gibbs sampling. All the analyses were performed using the R package Bayesian Generalized Linear Regression (Perez and de los Campos 2014). The predictive ability of the different kernel models was assessed by five-fold cross-validation with 10 repetitions. The ability to predict yet-to-be observed SCR values was assessed using the Pearson product-moment correlation coefficient (CORR) and the mean squared error of prediction (MSEP).

Results

Whole genome scan. Figure 1 shows the results of the whole-genome single-marker scan under the recessive model. Two genomic regions, located on BTA6 and BTA26, showed marked recessive effects on SCR (Figure 1B). Figure 1B shows the distribution of SCR values for the two SNP loci with marked recessive effects, rs133071278 and rs41601831. Notably, these box plots demonstrate that the BB genotypes have lower SCR values than genotypes AA and AB. No region showed pure dominance or overdominance effects.

Genomic prediction. Figure 2 shows the predictive performance of the linear kernel-based regression models fitting the entire set of SNP markers, 'Base', or the models that include the two significant recessive markers as fixed effects, 'Base + Recessive Markers'. Note that the 'Base' model (GBLUP model) exhibited an average correlation between observed and predicted SCR values of 0.197, and MSEP equal to 22.03. The 'Base + Recessive Markers' model delivered predictive correlation around 0.322 and MSEP around 20.3. Gaussian kernel models did not outperform linear kernel models in terms of predictive performance (data not shown).

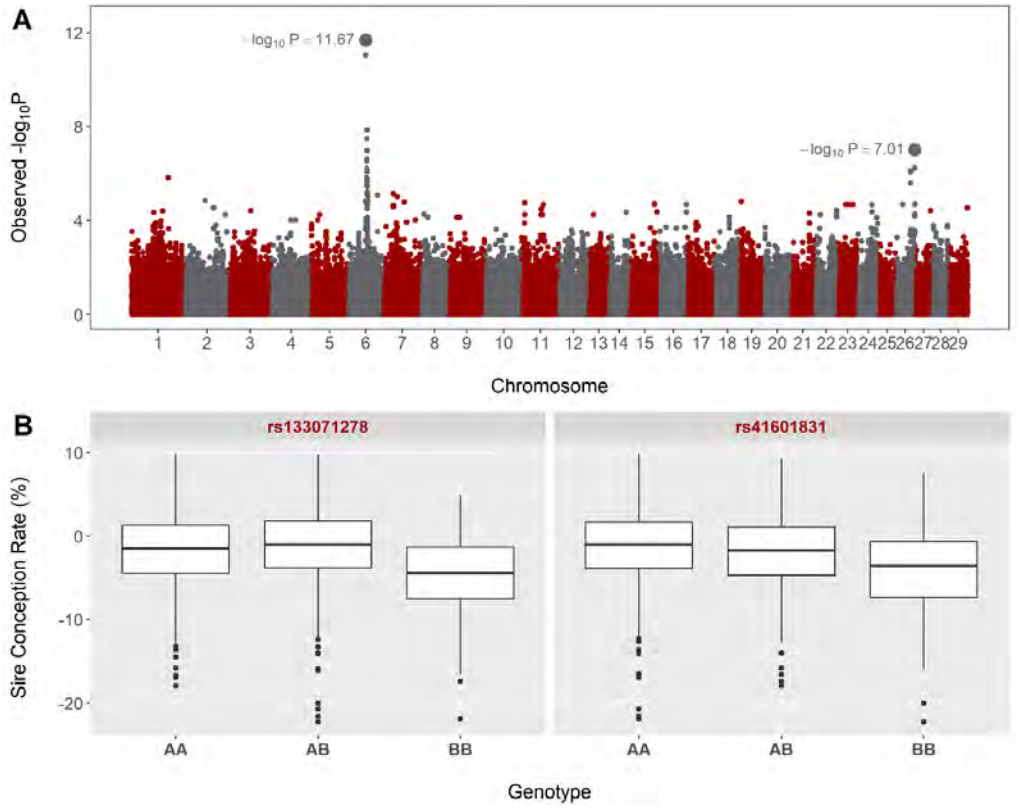


Figure 1. Whole-genome scan for recessive effects on Sire Conception Rate. (A) Manhattan plot showing the SNP significance across the bovine genome for recessive effects on Sire Conception Rate. (B) Boxplots showing the observed differences in Sire Conception Rate for each genotype of the two significant recessive markers.

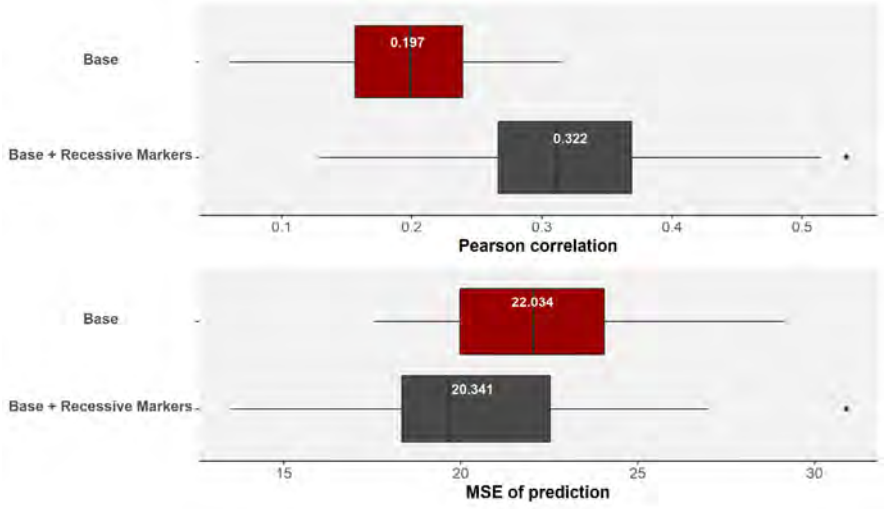


Figure 2. Predicting ability of alternative whole-genome predictive models. Predictive correlation (top) and mean squared error of prediction (bottom) evaluated for each model.

Discussion

The relevance of non-additive effects on Italian Brown Swiss bull fertility was evaluated, and two major genomic regions with large recessive effects were identified. Genomic prediction models including either linear or Gaussian kernels fitting either all SNP or including markers with large effect were evaluated. Notably, model predictive performance was largely improved by including the two markers with large recessive effect as fixed effects, representing an increase in predictive correlation of about 63%. The use of genomic prediction models that include markers with large effect is a promising alternative to the standard approach. Overall, our findings may help to improve reproductive performance of dairy herds through accurate genome-guided decisions on bull fertility.

References

Aulchenko Y.S., Ripke S., Isaacs A. and Van Duijn C.M. (2007) Bioinformatics 23(10)**:** 1294-1296. https://doi.org/10.1093/bioinformatics/btm108

Morota G., and Gianola D. (2014) Frontiers in Genetics 5(363. https://doi.org/10.3389/fgene.2014.00363

Ortega M.S., Moraes J.G.N., Patterson D.J., Smith M.F., Behura S.K. *et al.* (2018) Biology of Reproduction 99(6)**:** 1244-1254. https://doi.org/10.1093/biolre/ioy141

Pacheco H.A., Battagin M., Rossoni A., Cecchinato A. and Peñagaricano F. (2021) Journal of Dairy Science. https://doi.org/10.3168/jds.2021-20332

Perez P., and De Los Campos G. (2014) Genetics 198(2)**:** 483-495. https://doi.org/10.1534/genetics.114.164442

Rezende F.M., Dietsch G.O. and Peñagaricano F. (2018) Animal Genetics 49(5)**:** 393-402. https://doi.org/10.1111/age.12710

282. On the use of SNPs of large effect to improve prediction accuracy in pigs

M.S. Lopes[1,2*], M. Derks[1,3], M. van Son[4], A.B. Gjuvsland[4], C.A. Sevillano[1], E. Grindflek[4] and E.F. Knol[1]

[1]Topigs Norsvin Research Center, P.O. Box 43, 6640 AA Beuningen, the Netherlands; [2]Topigs Norsvin, Visconde do Rio Branco 1310, 80.420-210 Curitiba, Brazil; [3]Wageningen University & Research, P.O. Box 338, 6700 AH Wageningen, the Netherlands; [4]Norsvin, Storhamargata 44, 2317 Hamar, Norway; marcos.lopes@topigsnorsvin.com

Abstract

The aim of this study was to develop a strategy for selecting SNPs that could receive higher weight than other SNPs in a GBLUP approach due to their expected association with important pig phenotypes. In addition, we aimed to investigate if such a strategy yields improved prediction accuracy compared to a traditional GBLUP approach. Four prediction accuracy scenarios were evaluated using three production traits in two pig populations. Our results show that adding extra weight to SNPs that are expected to be (close to) causal variants in the G matrix increases the prediction accuracy of genomic prediction. The advantage of weighted G matrix compared to a traditional one is not very large, but the added value is consistent.

Introduction

The first practical use of genomic information in animal prediction models was the application of marker-assisted selection (MAS) where markers of large effect were identified in linkage-based studies using microsatellite markers and then applied in selection. The true benefits of genomics, however, became more pronounced only after the development of dense single nucleotide polymorphism (SNP) panels. With SNPs, the association studies moved from linkage-based to genome-wide association studies (GWAS), and MAS was replaced by genomic selection (Meuwissen *et al.*, 2001), which consists of the estimation of genomic breeding values using a large number of SNPs spread across the whole genome. With GWAS, the power to detect new QTL has increased and SNPs tightly-linked to the causal variants have been identified (van Son *et al.*, 2019). However, GWAS findings have not been extensively exploited for selection purposes as the focus has been mainly on genomic selection. Different methods for genomic selection have been developed, including the extensive Bayesian alphabet (Habier *et al.*, 2011) which allows to put more emphasis on SNPs with large effects. For practical application, however, the most used method in pig breeding is the so-called 'single-step' genomic evaluation (Misztal *et al.*, 2009), which can be considered as an extension of the genomic BLUP (GBLUP) assuming that all SNPs contribute equally to the trait (infinitesimal model). Nevertheless, it has been shown that a finite number of genes control quantitative traits (Hayes and Goddard, 2001) and an approach to give different weights to SNPs when building a genomic relationship (G) matrix has been proposed (Zhang *et al.*, 2014). Therefore, further improvement of traditional genomic prediction models could be achieved by weighting SNPs differently in the G matrix if they are, for example, significant in GWAS (Zhang *et al.*, 2014) or are expected to be a causal variant identified using a porcine Combined Annotation Dependent-Depletion (pCADD) tool (Derks *et al.*, 2021). The aim of this study was to develop a strategy for selecting SNPs that could receive higher weight than other SNPs in the G matrix due to their expected association with important pig phenotypes. In addition, we aimed to investigate if such a strategy yields improved prediction accuracy compared to a traditional GBLUP approach.

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_282, © M.S. Lopes *et al.* 2022

Materials & methods

GWAS. Our first step was to perform a GWAS using all traits available in the Topigs Norsvin breeding program. In total, we evaluated 113 traits using five pig populations: Large White, Landrace, Pietrain, Duroc and a Synthetic line. The single-SNP GWAS was ran within population using a single-trait linear animal model in GCTA (Yang *et al.*, 2011). The response variables for most traits were pre-corrected phenotypes for all non-genetic effects using all data available in the routine genetic evaluation of Topigs Norsvin using MiXBLUP (ten Napel *et al.*, 2017). For binary traits, maternal and indirect genetic effects and traits that are an index (e.g. selection index), the response variables were breeding values with a reliability greater than 0.50. Breeding values and reliability estimates were also extracted from the routine evaluation of Topigs Norsvin using MiXBLUP (ten Napel *et al.*, 2017). The number of animals per trait and per line ranged from ~1,300 to ~40,000. Significant association was declared when the SNP presented a *P*-value $<1.0\times10^{-10}$. All significant SNPs located within 5 Mb from another significant SNP were considered to belong to the same QTL region and only the most significant SNP of each QTL region was selected for the further steps.

pCADD. The second step was to select whole-genome sequence variants based on pCADD scores as described in Derks *et al.* (2021). The pCADD score estimates a probability of having a functional (deleterious) impact for each variant. This probability is based on a machine-learning procedure comparing sequence data across species and including annotation data of known functional elements. The score is trait-independent but depends on the quality of the annotation in the region of interest. We selected variants based on their high pCADD score across different annotations as well as variants that are underlying known QTL regions (Derks *et al.*, 2021).

Genotypes. Animals used in the GWAS were genotyped using Illumina Neogen medium-density SNP chips (50K or 80K) (Lincoln, NE, USA). Part of these animals (most influential boars) were also genotyped using the Affymetrix 660K high-density SNP chip (Santa Clara, CA, USA). All animals had their genotypes imputed towards the medium density chip using Fimpute v3 (Sargolzaei *et al.*, 2014) before performing the GWAS. In addition, ~2,000 SNPs were selected from the GWAS and pCADD analyses (2K set) and placed in an Illumina Neogen custom 25K SNP chip together with ~23,000 SNPs (23K set) from the 50K chip that were segregating across all Topigs Norsvin populations or were present on X or Y chromosome. After genotyping over 100,000 animals across all populations with the 25K SNP chip, another round of imputation was performed towards a panel of SNPs that combines all SNPs from the 50K and 25K SNP chips. This new imputed dataset was used for the validation step. Quality control was performed at each level of imputation as described in van Son *et al.* (2019).

Validation. We selected significant SNPs from the GWAS for 113 traits in five populations as well as SNPs based on high pCADD scores which are trait independent. The reason for this is that our final goal is to have a set of SNPs that can be used in multi-trait and multi-population scheme. However, for the validation in this study, we evaluated only the traits backfat (BF) at the end of the test period (~120 kg), average daily gain (DG) and feed intake (FI) during the test period (~25-120 kg) from two populations (Large White and Landrace). From each population we selected the youngest 6,000 animals that had pre-corrected phenotypes for all three traits. The 1000 (very) youngest animals from each population, all genotyped with the 25K SNP chip, were taken as our validation set. The remaining 5,000 animals, genotyped with both 25K and 50K, were taken as our reference (training) set. Four prediction accuracy scenarios were evaluated: 1. 23K, which used a standard G matrix using the same SNPs as the routine genetic evaluation (23K set); 2. 2K, which uses a standard G matrix using only the GWAS and pCADD SNPs (2K set); 3. 25K, which uses a standard G matrix built from a combined set with SNPs from the 23K and 2K sets; 4. w25K, which uses a weighted G matrix using the same SNPs as the 25K scenario. Accuracy of prediction of each scenario was assessed as the correlation between the pre-corrected phenotype and estimated breeding values of the

1000 validation animals from each population. The G matrices of all scenarios were built using calc_grm (Calus and Vandenplas, 2016). To define the weights used in the w25K scenario, we assumed that the 2K set and the 23K set of SNPs can explain the same amount of variance. Therefore, the sum of the weights of the 2K set was equal to the sum of the weights of the 23K set, which means that each 2K SNP has about 20 times heavier weight than the 23K SNPs in the G matrix. We applied the same weights for each SNP independently of the evaluated trait and population. In a standard G matrix, all SNPs have the same weight. Different levels of weights were evaluated (e.g. the sum of the weights of all 2K set was half of the sum of the weights of the 23K set), but the results were similar to those obtained with equal weights and will therefore not be presented.

Results

We selected 260 significant SNPs from the GWAS and 876 with high pCADD scores. The GWAS SNPs explained up to 16% of the phenotypic and 40% of the genetic variance of the evaluated traits. After including them in the custom 25K SNP chip, we obtained 220 GWAS and 784 pCADD SNPs working properly and segregating in both the Large White and the Landrace populations. Therefore, the final 2K set was composed of 1,004 SNPs, while the final 23K set was composed of 21,117 autosomal SNPs also segregating in both populations. The highest prediction accuracies were observed for scenario w25K, except for FI in the Landrace population (Table 1). However, the relative increase in prediction accuracy of the w25K scenario compared to the traditional 23K was not that high, ranging from 0.5 to 4%.

Discussion

Our results show that selecting SNPs that are expected to be (close to) causal variants, including these SNPs in a custom SNP chip and weighting them heavier than standard SNPs in the G matrix might be beneficial to increase the prediction accuracy in the evaluated populations. Although the advantage of w25K scenario compared to the traditional one (23K) is not very large, the added value is consistent. A similar situation has been shown by Khansefid *et al.* (2020) who showed that a custom panel enriched with, or close to, causal mutations yield higher prediction accuracy than their traditional scenario. It is also interesting to observe that using only 1,004 SNP selected from the GWAS or based on pCADD scores (2K scenario), we already obtain high accuracies that are not too far from those obtained with more than 21,000 SNPs (23K scenario). Further, this is a work in progress, and we expect to improve it by investigating other strategies for selecting causal variants and improving the strategy of weighting the SNPs in the G matrix. For practical reasons, we will be looking for strategies that allow improvements in a multi-trait and multi-population scheme. However, working within trait and within population, the added value of a weighted G matrix might be larger due to the differences in allele frequency and linkage disequilibrium across populations.

Table 1. Accuracies of prediction.

Population	Trait	Scenarios[1]			
		23K	2K	25K	w25K
Large White	Backfat	0.474	0.387	0.481	**0.492**
	Average daily gain	0.330	0.228	0.334	**0.340**
	Average feed intake	0.470	0.371	0.474	**0.480**
Landrace	Backfat	0.453	0.412	0.460	**0.473**
	Average daily gain	0.440	0.386	0.444	**0.452**
	Average feed intake	0.522	0.423	**0.525**	0.524

[1] Scenarios: 23K= standard G matrix used in the routine genetic evaluation; 2K= standard G matrix using only the GWAS and pCADD SNPs; 25K= standard G matrix using a combined set with SNPs from 23K and 2K; w25K= weighted G matrix using a combined set with SNPs from 23K and 2K. Numbers in **bold** indicate the highest accuracy of each line.

References

Calus M., and Vandenplas J. (2016) Animal Breeding and Genomics Centre, Wageningen UR Livestock Research.

Derks M.F., Gross C., Lopes M.S., Reinders M.J., Bosse M., *et al.* (2021) Genomics 113(4):2229-2239. https://doi.org/10.1016/j.ygeno.2021.05.017

Habier D., Fernando R.L., Kizilkaya K., and Garrick D.J. (2011). BMC bioinformatics 12(1):186. https://doi.org/10.1186/1471-2105-12-186

Hayes B., and Goddard M.E. (2001). Genetics Selection Evolution 33(3):1-21. https://doi.org/10.1186/1297-9686-33-3-209

Khansefid M., Goddard M.E., Haile-Mariam M., Konstantinov K.V., Schrooten C., *et al.* (2020). Frontiers in genetics 11. https://doi.org/10.3389/fgene.2020.598580

Meuwissen T., Hayes B., and Goddard M. (2001). Genetics 157(4):1819-1829. https://doi.org/10.1093/genetics/157.4.1819

Misztal I., Legarra A., and Aguilar I. (2009). Journal of dairy science 92(9):4648-4655. https://doi.org/10.3168/jds.2009-2064

Sargolzaei M., Chesnais J.P., and Schenkel F.S. (2014). BMC Genomics 15:478. https://doi.org/10.1186/1471-2164-15-478

ten Napel J., Vandenplas J., Lidauer M., Stranden I., Taskinen M., *et al.* (2017) MiXBLUP Manual 202. Available at: https://www.mixblup.eu/download.html

van Son M., Lopes M.S., Martell H.J., Derks M.F., Gangsei L.E., *et al.* (2019) Frontiers in Genetics 10, 272. https://doi.org/10.3389/fgene.2019.00272

Yang J., Lee S.H., Goddard M.E., and Visscher P.M. (2011). American Journal of Human Genetics 88(1):76-82. https://doi.org/10.1016/j.ajhg.2010.11.011

Zhang Z., Ober U., Erbe M., Zhang H., Gao N., *et al.* (2014). PloS one 9(3):e93017. https://doi.org/10.1371/journal.pone.0093017

283. Including environmental variables in genomic models for carcass traits in Hanwoo beef cattle

B.O. Makanjuola[1*], G. Rovere[1], B.C.D. Cuyabano[2], S.H. Lee[3] and C. Gondro[1]

[1]Department of Animal Science, Michigan State University, 474 S Shaw Ln, East Lansing, MI 48824, USA; [2]French National Institute for Agriculture, Food and Environment (INRAE), Génétique Animale et Biologie Intégrative (GABI), 78350 Jouy-en-Josas, France; [3]Division of Animal and Dairy Science, Chungnam National University, Daejeon 305764, Korea; makanjuo@msu.edu

Abstract

The objective of this study was to account for environmental variables to model non-genetic random effects for carcass traits in Hanwoo beef cattle. This model was compared to the traditional model that uses farms as an independent component to model non-genetic random effects. Heritability estimates for all carcass traits in our study ranged from 0.33 to 0.39 when using independent farm effects, whereas with the use of environmental variables, estimated heritabilities dropped and ranged from 0.11 to 0.31. In addition, prediction accuracy increased from a range of 0.39 to 0.45 to a range of 0.50 to 0.58 for all traits when modelling farm effects with environmental variables. This result suggests that environmental variables explain a considerable amount of the observed phenotypic variance, and that not accounting for such environmental variation could upwardly bias heritability estimates. Furthermore, an improvement in the model prediction accuracy was observed when accounting for environmental variables.

Introduction

The impact of climate change resulting from increasing fluctuations in environmental variables and rising temperatures is a major concern for livestock production. Adverse effects of rising temperatures are observed in the form of reduced productivity, performance, growth, and development of animals exposed to such challenging environments (Collier and Gebremedhin, 2015). Traditionally, environmental effects are accounted for by fitting the herd as an effect in genetic evaluation models. With this model, it is traditionally assumed that different herds are completely independent from each other with no similarities. However, herds in closer proximity could be more similar in terms of climatic and geographic factors when compared to more geographically distant herds. Therefore, accounting for relationships of environmental variables between production sites in genetic evaluation models could improve the estimates of genetic components in economically important traits. Improvement in the genetic evaluation of smallholder breeding programs using spatial relationships between herds has already been reported (Selle *et al.,* 2020). In addition, Cuyabano *et al.* (2021) recently showed that using GPS coordinates to create relationships between herds led to an increase in the reliability of the predicted genomic breeding values for carcass traits in beef cattle, when compared to fitting herd effects as independent. Nowadays, public weather stations regularly collect environmental variables and the data is easily accessible for general use. This source of information has been shown to be more accurate than on-farm weather information (Freitas *et al.,* 2006) and can be used to develop a relationship structure between herds based on their weather conditions. Furthermore, temperature-humidity index (THI) as environmental descriptor has been successful in modelling genotype-by-environment interactions with reaction norms (Chen *et al.,* 2021). This present study aimed to: (1) use environmental variables obtained from public weather stations to create a climate relationship matrix between herds; (2) evaluate the impact on estimated variance components by modelling herd effects through the climate relationship structure and (3) assess prediction accuracy for the predicted genomic breeding values.

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_283, © B.O. Makanjuola *et al.* 2022

Materials & methods

Animal and phenotypic data. Data from a total of 4,168 Korean Hanwoo beef cattle was used in this study. There were 3,810 steers and 358 bulls born in 2014 and 2015, all slaughtered in 2017. For this study, four carcass traits were analysed: carcass weight (CWT), backfat thickness (BFT), eye muscle area (EMA) and marbling score (MS). Measurements were taken 24h after refrigeration following slaughter. For BFT and EMA, records were measured at the 12-13th rib junction in millimeters and squared centimeters, respectively. MS was measured visually by a trained evaluator on an ordinal scale that ranged for 1 to 9. The 4,168 animals were raised on a total of 124 finishing farms, which were spread across South Korea.

Genotypes and weather data. All animals used in this study were genotyped with the 50K Illumina (San Diego, CA, USA) Bovine SNP50V2BeadChip array. Quality control measures were performed to retain only SNP present on the autosomal chromosomes with a call rate >90% and minor allele frequency >1%. After quality control, a total of 43,749 SNP genotypes were retained for further analyses. Weather information was available from 78 public weather stations that were close to the farms. To each farm, weather information from the nearest weather station was assigned. The distances between the farms and the weather stations ranged from 0 to 30 km with approximately 45% of the farms sharing the same location as the weather station. The environmental variables used for this study were monthly averages of the maximum temperature (T_{max}) and of the relative humidity (RH_{min}), spanning from 2013 to 2018. Temperature and relative humidity were then combined into THI using the following formula (NRC, 1971):

$$THI = (1.8T_{max} + 32) - [(0.55 - 0.0055RH_{min}) \times (1.8T_{max} - 26)] \quad (1)$$

THI was considered a suitable environmental variable to use because a previous study already reported that THI influences the modulation of the genetic components of these traits in Hanwoo (Chung *et al.,* 2020).

Genomic evaluation. Genetic parameters were estimated for all traits using both of the following linear mixed models:

$$y = Xb + Zg + Wfarm + e, \quad (2)$$

$$y = Xb + Zg + Wst + e, \quad (3)$$

where y is a vector of phenotypic records for CWT, BFT, EMA and MS; b is a vector of fixed effects that included the effects of sex, age at slaughter, herd size and slaughter date; g is a vector of random additive genetic effects; farm is a vector of non-genetic and independent random effects of farm; st is a vector of non-genetic and correlated random effects of farm; e is a vector of random residuals. X, Z and W are incidence matrices that link respectively the fixed effects, random additive genetic effects and non-genetic random farm effects to the phenotypes. The assumptions posed to the random effects were: $g \sim N(0, G\sigma_g^2)$, $farm \sim N(0, I\sigma_{farm}^2)$ and $st \sim N(0, E\sigma_{st}^2)$. σ_g^2 is the additive genetic variance; σ_{farm}^2 and σ_{st}^2 are the non-genetic variance of the herd effect assuming the farms as independent and correlated, respectively. G is the genomic relationship matrix (VanRaden, 2008), I is an identity matrix, and E is the covariance matrix of the farms using the closest weather station information. E was created using the standardized differences between THI values of the farms. First, a matrix D was calculated using the differences between THI values, and the matrix was then scaled by dividing the values by the maximum absolute THI difference, to ensure that the elements of the matrix ranged within the [-1,1] interval (i.e. $D^* = D / \max\{|D_{ij}|: i,j = 1, \ldots, st\}$. Finally, the covariance matrix E was rescaled to have a diagonal of ones and defined as $E_{ij} = 1 - D_{ij}$ with values ranging from $-1 \leq E_{ij} \leq 1$. With this parameterization, stations that are more similar to each other have values closer

to one and the more dissimilar stations will have values closer to minus one. All variance components estimation and genomic predictions were performed using WOMBAT software (Meyer, 2007).

Model predictive accuracy. The predictive accuracy of the different models in predicting the genomic breeding values was performed using a 5-fold cross validation. To evaluate accuracy, the whole data was randomly divided into five subsets of approximately equal size. For each fold of the cross validation, four out of the five subsets (i.e. 80% of data) were assigned as the training set and the remaining subset (i.e. 20% of the data) was used as the testing set. Replication of the 5-fold cross validation procedure was performed 20 times and resulted in a total of 100 estimates. Prediction accuracy in the testing set was calculated as the Pearson correlation between predicted genomic breeding values of the test group and observed phenotypes adjusted for fixed effects ($r = cor(\hat{g}_{test}, y_{test} - X_{test}\hat{b})$).

Results

Table 1 presents a summary of the phenotypic records, with descriptive statistics for each trait.

The proportions of variance explained by each random effect using the two different models for each trait are presented in Table 2, along with the prediction accuracy of the predicted genomic breeding values. Heritability estimates for BFT, EMA, CWT and MS were 0.34, 0.35, 0.39 and 0.38, respectively, when fitting the farm as independent effects. These estimates are in line with other studies that used similar models with heritability estimates that ranged from 0.33 to 0.42 (Chung *et al.,* 2020; Cuyabano *et al.,* 2021). Conversely, using the THI to model the farm effects as correlated resulted in a drop in heritability estimates to 0.31, 0.11, 0.29 and 0.18 for BFT, EMA, CWT and MS, respectively. To the best of our knowledge, no study has reported using a covariance matrix created from THI values from public weather stations. Nevertheless, our results are in accordance with those from a study that used GPS coordinates to create a covariance matrix between farms (Cuyabano *et al.,* 2021). These results suggest that environmental variables and geographic location explain a substantial proportion of the phenotypic variation observed. Prediction accuracies obtained were 0.39, 0.40, 0.45 and 0.40 for BFT, EMA, CWT and MS, respectively, when farm was fitted as an independent effect, and increased to 0.50, 0.54, 0.58 and 0.56 for BFT, EMA, CWT and MS, respectively, when fitting farms with a covariance matrix based on environmental variables, described in this study by the THI.

Discussion

Our study compared modelling herd effects considering farms as correlated or independent. A similar study has already been performed for Hanwoo beef cattle, using GPS coordinates to build correlation between farms (Cuyabano *et al.,* 2021). We innovate with our study by using environmental information from public weather stations, in the form of THI, to build this correlation between farms. Our results showed that a considerable proportion of the phenotypic variation can be explained by correlated environment, in agreement to what was reported by Cuyabano *et al.* (2021), and that this leads to a reduction in the

Table 1. Descriptive statistics of analysed traits including the number of records (N), mean, standard deviation, minimum and maximum of the observations.

Trait	N	Mean	Standard deviation	Minimum	Maximum
BFT	4,168	14.33	5.09	2	47
EMA	4,168	96.74	12.05	34	155
CWT	4,168	442.02	54.95	159	682
MS	4,168	6.44	1.76	1	9

Table 2. Estimates of variance proportions explained by additive genetics (h^2), non-genetic effects of farm or weather station ($(farm/st)^2$) and residual (e^2) and predictive accuracy (r) with their respective standard errors (se).

Trait	Model	$h^2 \pm se$	$(farm/st)^2 \pm se$	$e^2 \pm se$	r
BFT (mm)	g + farm	0.34±0.03	0.03±0.01	0.63±0.03	0.39
	g + st	0.31±0.04	0.13±0.08	0.57±0.06	0.50
EMA (cm^2)	g + farm	0.35±0.03	0.07±.02	0.58±0.03	0.40
	g + st	0.11±0.03	0.69±0.08	0.19±0.05	0.54
CWT (kg)	g + farm	0.39±0.03	0.04±0.01	0.58±.03	0.45
	g + st	0.29±0.05	0.29±0.12	0.43±0.07	0.58
MS	g + farm	0.38±0.03	0.05±.01	0.57±0.03	0.40
	g + st	0.18±0.02	0.54±0.01	0.27±0.01	0.56

additive genetic variance, and consequently to lower heritability estimates. This finding may suggest that not accounting for environmental variation could be upwardly biasing heritability estimates. Additionally, the highest prediction accuracy for all traits was obtained when using the model that accounted for the THI covariances between farms. Such improvement was not observed by Cuyabano *et al.* (2021), suggesting that actual weather information may be more accurate to estimate the random effects than GPS coordinates. Further research is warranted to demonstrate that environmental variables can be used to improve models in another independent population.

References

Chen S.Y., Freitas P.H.F., Oliveira H.R., Lázaro S.F., Huang Y. J., *et al.* (2021) Genetics Selection Evolution 53(1):1-18. https://doi.org/10.1186/s12711-021-00645-y

Chung Y., Lee S. H., Lee H. K., Lim D., van der Werf J., *et al.* (2020) Frontiers in Genetics, 11:1632. https://doi.org/10.3389/fgene.2020.576377

Collier R. J., and Gebremedhin K. G. (2015) Annu. Rev. Anim. Biosci. 3:513-532. https://doi.org/10.1146/annurev-animal-022114-110659

Cuyabano B. C. D., Rovere G., Lim D., Kim T. H., Lee H. K., *et al.* (2021) Animals 11(7):2050. https://doi.org/10.3390/ani11072050

Freitas M. S., Misztal I., Bohmanova J., and West J. (2006) Livestock Science 105(1): 223-228. https://doi.org/10.1016/j.livsci.2006.06.011

Meyer K. (2007) J Zhejiang Univ. Sci. B. 8:815-821. http://doi.org/10.1631/jzus.2007.B0815

NRC. (1971) *A guide to environmental research on animals*. National Research Council, Washington, DC: National Academies, USA.

Selle M. L., Steinsland I., Powell O., Hickey J. M., and Gorjanc G. (2020) Genetics Selection Evolution 52(1): 1-17. https://doi.org/10.1186/s12711-020-00588-w

VanRaden P. M. (2008) J Dairy Sci 91(11): 4414-4423. https://doi.org/10.3168/jds.2007-0980

284. Considering chromosomal trait correlations improves accuracy of genomic prediction

C. Schmidtmann[1*], D. Segelke[2], J. Bennewitz[3], J. Tetens[4] and G. Thaller[1]

[1]Christian-Albrechts-University Kiel, Institute of Animal Breeding and Husbandry, Hermann-Rodewald-Straße 6, 24118 Kiel, Germany; [2]Vereinigte Informationssysteme Tierhaltung w.V. (vit), Heinrich-Schröder-Weg 1, 27283 Verden, Germany; [3]University of Hohenheim, Institute of Animal Science, Garbenstraße 17, 70599 Stuttgart, Germany; [4]Georg-August-University Göttingen, Division of Functional Breeding, Department of Animal Sciences, Burckhardtweg 2, 37077 Göttingen, Germany; cschmidtmann@tierzucht.uni-kiel.de

Abstract

Genetic correlations at chromosomal level between milk and fat yield in dairy cattle were calculated from SNP effects. The results revealed positive genetic correlations for all chromosomes (r_g=0.327-0.792) except for *Bos taurus* autosome (BTA) 14, where genetic correlation between traits was negative (r_g=-0.794). Two bivariate GBLUP models were fitted – one (Model A) with a single genomic relationship matrix and the other (Model B) with two distinct genomic relationship matrices, one constructed from SNPs of all chromosomes except BTA14 and one constructed only from SNPs on BTA14, where major gene *DGAT1* is located. Accuracy of genomic prediction was evaluated by fivefold cross-validation. Model B showed on average higher relative accuracy of 3.56% for milk yield and 3.67% for fat yield compared to Model A. The presented approach accounts for chromosomal differences in genetic correlations between traits, thereby improving accuracy of genomic prediction.

Introduction

Genetic correlation indicates average similarity between two traits emerging from a shared genetic basis. Patterns of genetic sharing can differ across the genome. Local genetic correlations, i.e. correlations between two traits in particular genomic regions, may differ from a genome-wide estimate. Knowledge about such local genetic correlations can improve the accuracy of genomic prediction. The first objective of this study was to calculate local genetic correlations between traits at chromosomal level. The second aim was to predict genomic breeding values using a bivariate setting and taking into account information on differences in chromosomal genetic correlations between traits. The predictability of this approach was evaluated by cross-validation.

Materials & methods

Data. Genotypes (45,613 SNPs) of 103,496 cows with phenotypic information for traits 305d-milk yield (MKG) and 305d-fat yield (FKG) in 1st lactation recorded between 2015 and 2020 were available. SNPs were excluded from further analysis when they had a call rate lower than 90%, a minor allele frequency lower than 1% or they deviated from Hardy-Weinberg equilibrium at a threshold $P<10^{-5}$. After quality control, 43,665 SNPs remained. For MKG and FKG, SNP effects from routine genetic evaluation were provided by Vereinigte Informationssysteme Tierhaltung (vit), Verden, Germany.

Chromosomal additive genetic (co)variances. For each animal and the two traits MKG and FKG, chromosomal breeding values (CEBV) were calculated as the sum of individual SNP effects on each chromosome. Chromosomal additive genetic variance ($\sigma^2_{g_{CEBV}}$) was calculated as the variance of CEBV across all animals using $\sigma^2_{g_{CEBV}} = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$, where n is the number of animals, x_i is the CEBV of each animal for trait i and $\bar{x}$ is the average CEBV of each chromosome across all animals.

Chromosomal additive genetic covariance between traits was calculated using:

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_284, © C. Schmidtmann *et al.* 2022

$cov(CEBV_{MKG}, CEBV_{FKG}) = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})$,

where n is the number of animals, x_i and y_i are CEBV of each animal for MKG and FKG, respectively. $\bar{x}$ and $\bar{y}$ are the average CEBV of each chromosome across all animals for MKG and FKG.

Chromosomal genetic correlations. The genetic correlation (r_g) at chromosomal level was calculated as $r_{g_{MKG,FKG}} = cov(CEBV_{MKG}, CEBV_{FKG})/ (\sigma_{CEBV_{MKG}} * \sigma_{CEBV_{FKG}})$, where $cov(CEBV_{MKG}, CEBV_{FKG})$ is the covariance between CEBV of MKG and FKG and $\sigma_{CEBV_{MKG}}$ and $\sigma_{CEBV_{FKG}}$ are the standard deviations of CEBV.

Estimation of breeding values. Two bivariate GBLUP animal models were used to estimate genomic breeding values (EBV) for 5,000 animals. Model A was $y = 1_n\mu + Zg_A + e$, where y is a vector of phenotypes MKG and FKG that were adjusted for fixed effects herd-year-season and age of 1st calving. 1_n is a vector of ones, μ is the overall mean, Z is a design matrix for the random genetic effect, g_A is a vector of additive genetic effects, and e is a vector of random residual terms. g_A was assumed to follow $g_A \sim N(0,\mathbf{G}\sigma^2_{gA})$, where G is a genomic relationship matrix (GRM) calculated from all SNPs (n=43,665). Model B included two random genetic terms and was $y = 1_n\mu + Zg_{B1} + Zg_{B2} + e$. The first random genetic effect g_{B1} was distributed $g_{B1} \sim N(0,\mathbf{G}\sigma^2_{gB1})$, where G is a GRM built from SNPs on all chromosomes except of those on BTA14 (n=42,141). The second random genetic effect g_{B2} followed $g_{B2} \sim N(0,\mathbf{G}\sigma^2_{gB2})$, where G is a GRM built only from SNPs on BTA14 (n=1,524). All GRMs were constructed according to VanRaden (2008). Values of variance components for variance-covariance-matrices of genetic effects were determined from (co) variances of computed CEBVs:

$$\sigma^2_{gA} = \begin{pmatrix} 358288 & 3711 \\ 3711 & 334 \end{pmatrix}, \sigma^2_{gB1} = \begin{pmatrix} 280901 & 5666 \\ 5666 & 256 \end{pmatrix} \text{ and } \sigma^2_{gB2} = \begin{pmatrix} 77387 & -1955 \\ -1955 & 78 \end{pmatrix}$$

Mixed models were solved in R (R Core Team, 2021) using the package *sommer* (Covarrubias-Pazaran, 2016).

Cross-validation. Fivefold cross-validation with ten replicates was performed. Prediction accuracy ($r_{(pheno, EBV)}$) was defined as correlation between adjusted phenotypes and estimated breeding values in the validation set. Additionally, mean squared error (MSE) was computed as the average squared deviation of estimated breeding value from the adjusted phenotype over all individuals.

Results

Correlation of SNP effects. The global correlation between MKG and FKG over all chromosomes was positive (r_g=0.623); however, genetic correlations at chromosomal level differed (Figure 1). Weakest and strongest positive correlation was found for BTA20 and BTA9, respectively (r_{gBTA20}=0.327, r_{gBTA9}=0.792). BTA14 showed a strong negative correlation between MKG and FKG (r_{gBTA14}=-0.794). Single weak correlations between chromosomes are shown on the off-diagonals.

Estimated breeding values. For Model A, the average correlation between breeding values (EBV) of MKG and FKG was r_{EBV}=0.407. For Model B, two EBV per trait and animal were estimated (Table 1). A moderate positive correlation was found between $EBV1_{MKG}$ and $EBV1_{FKG}$ (r_{EBV}=0.654), the correlation between $EBV2_{MKG}$ and $EBV2_{FKG}$ was negative (r_{EBV}=-0.531). Weak positive correlations were observed between $EBV1_{MKG}$ and $EBV2_{MKG}$ (r_{EBV}=0.127) and between $EBV1_{FKG}$ and $EBV2_{FKG}$ (r_{EBV}=0.051).

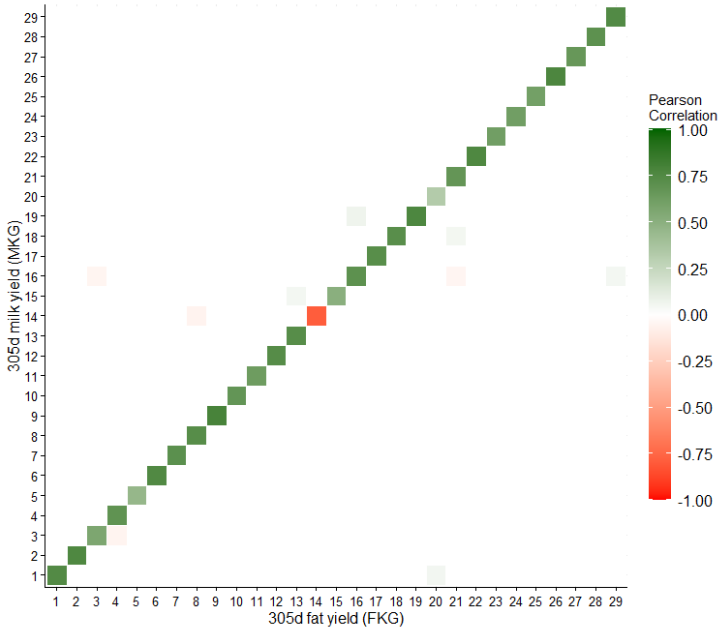


Figure 1. Correlations between chromosomal breeding values for 305d milk yield and 305d fat yield. Values between -0.05 and 0.05 are displayed in white.

Table 1. Correlations (r_{EBV}) between breeding values (EBV1: all chromosomes except BTA14; EBV2: BTA14) for milk yield (MKG) and fat yield (FKG) from Model B.

	$EBV2_{MKG}$	$EBV1_{FKG}$	$EBV2_{FKG}$
$EBV1_{MKG}$	0.127	0.654	0.015
$EBV2_{MKG}$		0.042	-0.531
$EBV1_{FKG}$			0.051

Cross-validation. For both traits, correlations between adjusted phenotypes and EBV in the validation set were higher for Model B (MKG: $r_{(pheno, EBV)}$=0.524, FKG: $r_{(pheno, EBV)}$=0.450) compared to Model A (MKG: $r_{(pheno, EBV)}$=0.506, FKG: $r_{(pheno, EBV)}$=0.434). The results revealed a relative increase in accuracy of 3.56% for MKG and 3.67% for FKG when using Model B. Moreover, the average MSE of Model B (MKG: MSE=1,584.9 kg, FKG: MSE=1.951 kg) was lower for animals in the validation set and for both traits compared to Model A (MKG: MSE=1,619.7 kg, FKG: MSE=1.979 kg).

Discussion

This study showed that genetic correlations between MKG and FKG differed at chromosomal level. A strong negative correlation was found on BTA14, which deviated from the positive correlations on remaining chromosomes. The major candidate gene diacylglycerol acyltransferase (*DGAT1*) located on BTA14 encodes a protein involved in fat metabolism and is a well-known quantitative trait locus for milk production traits (Grisart *et al.*, 2002, Winter *et al.*, 2002). The unfavorable association between milk and fat yield on BTA14 has been previously described in different studies (Grisart *et al.*, 2004, Thaller *et al.*, 2003). Given the two separate genetic effects in Model B, chromosome-specific (co)variances of traits were modelled more precisely compared to Model A, which included only one genetic effect.

Table 2. Mean of accuracy ($r_{(pheno, EBV)}$) and mean squared error (MSE) from 10 replicates of fivefold cross-validation for animals in training- and validation-set and traits 305d-milk yield (MKG) and 305d-fat yield (FKG), standard errors in parenthesis.

	MKG				FKG			
	$r_{(pheno, EBV)}$		MSE (in kg)		$r_{(pheno, EBV)}$		MSE (in kg)	
	training	validation	training	validation	training	validation	training	validation
Model A	0.846 (0.007)	0.506 (0.025)	194.9 (0.438)	1,619.7 (1.235)	0.834 (0.007)	0.434 (0.032)	0.248 (0.015)	1.979 (0.041)
Model B	0.839 (0.007)	0.524 (0.026)	199.2 (0.339)	1,584.9 (1.241)	0.825 (0.008)	0.450 (0.027)	0.255 (0.012)	1.951 (0.045)
Δ	-0.007	+0.018	+4.3	-34.8	-0.009	+0.016	+0.007	-0.028

Positive correlations were observed between the estimated breeding values EBV1 and EBV2 from Model B for both MKG and FKG, indicating that breeding values are not completely independent. Such correlations can imply existing interchromosomal linkage disequilibrium, possibly induced by directional selection.

The presented approach allows to easily account for differences in genetic architecture of traits at the chromosomal level. Thus, accuracy of genomic prediction improved. The approach will be extended to other pairs of traits including direct health traits of dairy cattle.

Acknowledgements

This study is part of the project 'QTCC: From Quantitative Trait Correlation to Causation in dairy cattle' and was supported by the German Research Foundation, DFG.

References

Covarrubias-Pazaran G. (2016) PLoS ONE 11(6):e0156744. https://doi.org/10.1371/journal.pone.0156744

Grisart B., Coppieters W., Farnier F., Karim L., Ford C., *et al.* (2002) Genome Res. 12:222-231. https://doi.org/10.1101/gr.224202

Grisart B., Farnir F., Karim L., Cambisano N., Kim J.-J., *et al.* (2004) Proc Natl Acad Sci USA 101(8):2398-2403. https://doi.org/10.1073/pnas.0308518100

R Core Team (2021), The R Project for Statistical Computing. Available at: https://www.r-project.org.

Thaller G., Krämer W., Winter A., Kaupe B., Erhardt G., *et al.* (2003) J Anim Sci. 81(8):1911–1918. https://doi.org/10.2527/2003.8181911x

VanRaden P.M. (2008) J. Dairy Sci. 91(11):4414-4423. https://doi.org/10.3168/jds.2007-0980

Winter A., Krämer W., Werner F.A.O., Kollers S., Kata S., *et al.* (2002) Proc Natl Acad Sci USA 99(14):9300-905. https://doi.org/10.1073/pnas.142293799

285. Effects of truncation and false positives in selection of markers for genomic prediction

Z. Loh*, J.H.J van der Werf and S. Clark

School of Environmental and Rural Science, University of New England, Armidale, 2351, NSW, Australia; zloh@myune.edu.au

Abstract

Association studies are often used to select genetic markers for genomic prediction, requiring truncation with a balance between power and false positives. Using simulation, the aim for this study is to test the effects of truncation on prediction accuracy and particularly the impact of false positives in SNP selection. Bonferroni and Benjamini-Hochberg False Discovery Rate methods were tested. Our study suggested that except for polygenic traits, truncation with a more lenient threshold such as the Benjamini-Hochberg False Discovery Rate increases the genomic prediction accuracy. In an inbred population, false positives could contribute positively to the accuracy especially for a oligogenic trait, although further study would be needed to generalize this result. Our study suggested that for a polygenic trait, all markers should be included in genomic prediction, and if SNP selection were to be applied a lenient threshold for truncation would be desirable.

Introduction

The development of high-density marker panels has allowed genomic prediction of breeding values estimation in livestock. The large number of markers also increased the computational intensity (Brøndum *et al.*, 2015) although most of the markers share no correlation with any causal loci (Moghaddar *et al.*, 2019). Thus, van der Berg *et al.* (2014) suggested that these markers can be truncated and only use selected markers ('top SNPs') using Genome-Wide Association Study (GWAS) with no significant negative impact on prediction accuracy, allowing the prediction to be conducted on smaller set of markers. Results were mixed however; Frischknecht *et al.* (2018) reported reduced accuracy when using top SNPs alone, compared to inclusion of all SNPs, while Chang *et al.* (2018) reported increased accuracy with top SNPs. A study for the possible causes of such mixed results, as well as effects of truncation threshold stringency would be desirable.

An important aspect to be considered in marker selection is the threshold used in GWAS. The most common threshold is from the Bonferroni multiple test correction, which has been criticized for its overly stringent threshold. The Benjamini-Hochberg False Discovery Rate (BH-FDR), was developed as a less stringent threshold, increasing power but also the false positive rate (Huang *et al.*, 2018). Studies on the effects of false positives when using them for genomic prediction are lacking however.

The aim of this study is to test the effects of truncation in marker selection and the associated effects of false positives on the genomic prediction accuracy. Truncation as defined by the Bonferroni and BH-FDR methods were tested.

Method

The experiment was conducted using simulation in R (version 3.6.1) and Python (version 3.9.7).

Simulation of base and parent genotype and phenotype. The base population was simulated by coalescent using the R package AlphaSimR (Gaynor *et al.*, 2021) with the command 'runMacs2' with six chromosomes of one Morgan each. The mutation rate was set at 10^{-5} per individual. To simulate a

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_285, © Z. Loh *et al.* 2022

population structure commonly found in livestock, the effective population size of the base population was set to 100. Via coalescence, a base population of 6,000 individuals was generated, each with 20,000 SNPs The tested parameters, along with their default and alternative values, are provided in Table 1. When a parameter is tested, default values were used for all other parameters.

The base population is then used to simulate a six-generation selection program. From the base population, an equal number of males and females were sampled randomly to be included in the selection program, with the total number of individuals sampled provided in Table 1. This served as the founder for the selection program. For each breeding pair, two offspring were generated for the next generation. The population size remained unchanged throughout the simulation. The Housworth-Stahl recombination model was used, with chiasma distance following a gamma distribution with shape parameter a=2 and rate parameter b=4. Selection was applied to sires only, with the top 50% of all sires been selected to propagate to the next generation, with each sire mated to two dams. An additive genetic model was assumed with heritability of the trait phenotype set at 0.3. A number of QTL was simulated, with their effect size following a gamma distribution with the parameters provided in Table 1.

Genotypes and phenotypes from the fifth generation after the base, denoted as parent generation, were used as training dataset, and those of sixth generation, denoted as offspring genotype and phenotype respectively, were used as test dataset.

GWAS and determination of top markers. To determine the top SNPs to be used in genomic prediction, genome-wide single SNP regression (GWSSR) of parent phenotypes on genotypes was used. Markers with minor allele frequency less than 0.05 were excluded from GWSSR. The threshold was calculated using either Bonferroni or BH-FDR multiple test correction, with alpha=0.05 for both methods. The number of true and false positive markers for both correction methods was recorded. Two types of false positives were identified: correlated false positives (FPC), defined as false positives in linkage disequilibrium (LD) (r^2>0.1) with at least one QTL, and uncorrelated false positives (FPU) that do not have correlation with any QTL.

Genomic prediction. The parent generation data was used to predict the breeding values of the offspring. Using the parent and offspring genotypes, a genomic relationship matrix (GRM) was built based on the methodology by VanRaden (2008), and based on parental generation phenotypes, the estimated breeding values (EBV) of the offspring were calculated using MTG2 (version 2.22) by Lee and van der Werf (2016). The prediction accuracy was defined as the correlation between the EBV and TBV of the offspring.

To test the effects of truncation method for top SNP, we compared different sets of markers used for genomic prediction. We used the selected top-SNPs and to remove the effect of varying number of markers included in the GRM and to ensure the prediction covered the entire genome, we added null markers uncorrelated with any QTL up to a total of 10k. To test the effects of false positives, two tests were conducted for each correction method, with one that utilized all the positives (AP), and the other with FPU excluded, hence only included true positives, and null, negative and uncorrelated markers were used in place of FPU.

Table 1. Parameters and value tested in this experiment.

Parameter tested	Default values	Alternative values
Parent sample size	2,500	500
QTL effect size distribution	Γ(0.3,1)	Γ(0.8,1)
Number of QTL	100	2,500

As control, additional tests were conducted on 10k randomly chosen markers (RND) and 10k markers including all true QTL (ALQ). This experiment as repeated 50 times for each scenario.

A pairwise t-test was conducted to compare the accuracies from different approaches, and for comparison between accuracies between two parameters, an independent Welch t-test was used. A comparison was deemed significantly different if $-\log_{10}(p)$ exceeded 3.0.

Results

The number of true and false positives from both correction methods are shown in Table 2, and the prediction accuracies are provided in Table 3.

Compared to all other methods, genomic prediction that include all QTL has the highest accuracy. Compared to 'RND', truncation for top SNPs generally was also associated with an increased accuracy. The only exception is for the polygenic trait (i.e. large number of QTL), where truncation was associated with a decreased accuracy.

Compared to the Bonferroni method, the use of BH-FDR for top SNP selection generally yielded significantly increased accuracy in genomic prediction in both all positives and true positives GRMs. This increase was associated with an increased number of positives associated with the trait. Interestingly, for both correction methods, the presence of uncorrelated false positives generally significantly increases the accuracy of prediction, although such increment due to false positives was not as significant when the trait is polygenic however (under Bonferroni correction, $-\log_{10}(p)=3.9$ for polygenic trait, compared to $-\log_{10}(p)=18.8$ for default condition).

Table 2. Number of true (TP), correlated false (FPC) and uncorrelated false (FPU) positives for each correction method and different simulated parameter values.

Parameter values	Bonferroni			BH-FDR		
	TP	FPC	FPU	TP	FPC	FPU
Default values[1]	8.4	39.0	173.5	13.7	65.8	523.3
Parent sample size =500	2.1	9.3	24.5	4.1	20.3	86.0
QTL effect size distribution= Γ(0.8,1)	10.2	39.1	152.1	18.4	79.4	575.4
Number of QTL =2,500	23.0	59.9	42.0	124.9	357.1	323.6

[1] The default values were provided in Table 1.

Table 3. Prediction accuracy from all markers, as well as all and true positives from each correction method under varying simulated parameter values.

Parameter values	All markers		Bonferroni		BH-FDR	
	RND[2]	ALQ	AP	TP	AP	TP
Default values[1]	0.730[a]	0.775[b]	0.771[c]	0.701[d]	0.778[e]	0.694[f]
Parent sample size =500	0.565[a]	0.599[a]	0.569[a]	0.531[b]	0.582[a]	0.531[b]
QTL effect size distribution= Γ(0.8,1)	0.728[a]	0.764[b]	0.755[c]	0.700[d]	0.766[b]	0.697[d]
Number of QTL =2,500	0.742[a]	0.755[a]	0.700[b]	0.697[c]	0.725[d]	0.721[e]

[1] The default values were provided in Table 1.

[2] Abbreviation: randomly chosen markers (RND); all QTL included (ALQ); all positives (AP); true positives (TP).

[3] All methods used 10k markers.

[4] Values with a different superscript with each rows were significantly different.

Discussion

Under default parameter values, truncation for top SNPs was associated with an increased prediction accuracy.

This could be caused by the genetic architecture of the traits. For an oligogenic trait (i.e. small number of QTL), only a small proportion of the markers have a correlation with a QTL, thus fewer markers are predictive. Compared to randomly chosen markers, truncation for top SNPs preferentially selects for markers that are likely associated with the trait variation, thus increasing the accuracy of prediction. This is further facilitated by the large variance explained by the detected QTL, further increasing the accuracy. For polygenic traits, a large portion of the markers have a correlation with QTL. Thus for randomly chosen markers, a large portion of the markers with predictive values would be included, increasing the accuracy, and limited gain is made by selecting them from GWAS. Given that many traits are polygenic, this study revealed the limited applicability of truncation in genomic prediction unless a very large dataset is made available. Between correction methods, the use of BH-FDR as truncation threshold yielded significantly higher accuracy than Bonferroni method. This could be attributed to the more lenient threshold of BH-FDR, which allowed more TP and FPC to be captured, thus increasing the prediction accuracy. This suggested that if truncation is to be applied in genomic prediction, a more lenient threshold might be favourable.

The inclusion of uncorrelated false positives in the selected top markers resulted in an increased accuracy in genomic prediction. One possible explanation for such observation was that these uncorrelated false positives, while not neighbouring to any true QTL, shared similar genotype structure with a QTL. This is evident from the LD structure of training data, where distant blocks of loci that shared relatively high LD with a locus have been observed. This structure can be linked to the high level of inbreeding within the population (Gondro, 2015). While not physically linked to a QTL, these loci might have added weight to the true QTL, increasing the accuracy. If this explanation is true, this supported the argument by Goddard (2009) that high accuracy can be achieved for an inbred population. Further study would be required on testing the generalizability of this observation under varying LD structure, and the role of FPU under these changes.

In conclusion, this study suggested that selection for predictive SNPs improves the prediction accuracy for oligogenic trait, but not for polygenic traits. If truncation was to applied in the genomic prediction, a more lenient threshold such as those proposed by BH-FDR would be desirable. While the inclusion of FPU increases the accuracy of prediction in this study, further study into the generalizability of such observations into other scenarios would be desirable. Further study should also focus on the optimal threshold for truncation given certain genetic architecture.

References

Brøndum R.F., Su G., Janss L., Sahana G., Guldbrandtsen B. *et al.* (2015). J. Dairy Sci. 98(6):4107-4116. https://doi.org/10.3168/jds.2014-9005

Chang L., Toghiani S., Ling A., Aggrey S.E., Rekaya R. (2018). BMC Genet 19(4) https://doi.org/10.1186/s12863-017-0595-2

Frischknecht M., Meuwissen T.H.E., Bapst B., Seefried F.R., Flury C., Garrick D. *et al.* (2018). J. Dairy Sci. 101(2):1292-1296. https://doi.org/10.3168/jds.2017-12890

Gaynor R.C., Gorjanc G., Hickey J.M. (2021). G3-Genes Genom Genet 11(2) https://doi.org/10.1093/g3journal/jkaa017

Goddard M. (2009). Genet. Sel. Evol. 136: 245-257 https://doi.org/10.1007/s10709-008-9308-0

Huang Q.Q., Ritchie S.C., Brozynska M., and Inouye M. (2018). Nucleic Acids Res. 46(22) https://doi.org/10.1093/nar/gky780

Lee S.H., van der Werf J.H.J. (2016). Bioinformatics 32(9):1420-1422 https://doi.org/10.1093/bioinformatics/btw012

Moghaddar N., Khansefid M., van der Werf J.H.J., Bolormaa S., Duijvesteijn N. *et al.* (2019). Genet. Sel. Evol. 51(72). https://doi.org/10.1186/s12711-019-0514-2

Norman A., Taylor J., Edwards J., Kuchel H. (2018). G3-Genes Genom Genet 8(9):2889-2899. https://doi.org/10.1534/g3.118.200311

VanRaden P.M. (2008). J. Dairy Sci. 91(11):4414-4423. https://doi.org/10.3168/jds.2007-0980

286. Genomic selection for breeding values under identity disequilibrium

R.J.C. Cantet[1,2*], B.K. Angarita Barajas[2], N.S. Forneris[1,2] and S. Munilla[1,2]

[1]Facultad de Agronomía, Universidad de Buenos Aires, Argentina; [2]Instituto de Investigaciones en Producción Animal (INPA) CONICET, 1417 CABA, Argentina; rcantet@agro.uba.ar

Abstract

We present formulae and properties of the additive recursive model for breeding values called the 'ancestral regression through parents', which combines pedigree and genomic data to account for identity disequilibrium due to differential contributions of the grandparents to the gametes forming the individual. At the end, we present an algorithm to predict breeding values of genotyped animals without own or progeny phenotypes measured.

Introduction

When fitting the 'ancestral regression' (AR) to the genetic evaluation of growth traits of Brangus beef cattle from South American countries, we have found computationally efficient to implement fittting the 'ancestral regression through parents' (PAR): a recursive model for breeding values (BV) that accounts for 'identity disequilibrium' (Weir and Cockerham, 1969) due to differential grandparental contributions to the animal's genome (Veller *et al.*, 2020). We have fitted PAR to a system of almost three million equations and found consistent predictions and accuracies between 2020 and 2021. In this presentation we give a summary of PAR, emphasizing on the properties of the covariance matrix of BV and show how to perform genomic selection.

Methods

Identity disequilibrium and the ancestral regression through parents (PAR). When predicting BV with the 'animal model' (Henderson, 1984), genetic similarity or relatedness between the BV of animals X and Y has been measured by twice the probability of identity by descent (*P*(IBD), Malecot, 1969). That is to say when animals X and Y have equal homologous alleles at site *i* in the genome because they are copies from an allele in a common ancestor to both animals. Weir *et al.* (1980) defined single 'site' (base, allele, gen, chromosome, individual) IBD with the random variables X_i and Y_i that indicate the location of the *i*-th homologous site at the genomes of X and Y, respectively. Thus, $P(\text{IBD}) = P(X_i \equiv Y_i)$ is equal to one when X_i and Y_i are IBD, and it is zero otherwise. However, genome inheritance is in 'segments' (Thompson, 2013) or multiple sites and $P(\text{IBD}) = P(X_1 \equiv Y_1, .., X_n \equiv Y_n)$. Under random mating and gametic equilibrium all sites are independent such that:

$$P\left(X_1 \equiv Y_1, ..., X_j \equiv Y_j, ..., X_N \equiv Y_N\right) = \prod_{j=1}^{N} P\left(X_j \equiv Y_j\right) \approx P\left(X \equiv Y\right) \quad (1.1)$$

The term on the far-right results from the fact that the *P*(IBD) under gametic equilibrium is 'equal at all sites'. In practice, genome inheritance is in 'segments' (Thompson, 2013) and not by single sites. Hence, we have to move to *P*(IBD) for non-independent sites, i.e. we have to consider linkage and linkage disequilibrium. For the *P*(IBD) at two sites, J.B.S. Haldane (Weir and Cockerham, 1969) observed the joint *P*(IBD) is always greater than single site *P*(IBD):

$$P\left(X_1 \equiv Y_1, X_2 \equiv Y_2\right) > P\left(X_1 \equiv Y_1\right) P\left(X_2 \equiv Y_2\right) \quad (1.2)$$

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_286, © R.J.C. Cantet *et al.* 2022

In turn, expression (1.2) allowed Weir and Cockerham (1969) to define 'identity disequilibrium' to the following difference, which is also the variance of IBD for two sites or loci (Weir *et al.*, 1980):

$$P\left(X_1 \equiv Y_1, X_2 \equiv Y_2\right) - P\left(X_1 \equiv Y_1\right) P\left(X_2 \equiv Y_2\right) = P\left(X_1 \equiv Y_1, X_2 \equiv Y_2\right) - P\left(X \equiv Y\right)^2 \tag{1.3}$$

Identity disequilibrium increases with inbreeding and linkage (Weir and Cockerham, 1969). Thus, the P(IBD) = $P(X_1 \equiv Y_1,.., X_n \equiv Y_n)$ at a segment is a 'multivariate binary' random variable that enables us to generalize identity disequilibrium to a large number of sites by means of theorem 5.1 in Esary *et al.* (1967), which can be translated here as that the P(IBD) from n sites is always greater or at least equal to the P(IBD) at n – 1 sites taken together, which in turn is greater than the P(IBD) at n – 2 sites taken together to eventually reach the single locus P(IBD). This is possible because multivariate binary random variables with 'non-negative covariance' among them are 'associated' and thus display 'positive quadrant dependence' when the 'covariance between BV is non-negative'. Under gametic equilibrium this is easy to prove because $\text{cov}(a_X, a_Y) = A_{XY}\sigma_A^2$ and, by definition, both terms are non-negative. The fact that BV are positive quadrant dependent Gaussian variables ensures that animals with larger (smaller) BV than the mean will tend to have progeny with larger (smaller) BV than the mean (Esary *et al.*, 1967; Bølviken, 1982).

We will call IDAR to the identity disequilibrium caused by unequal contributions of the grandparental genomes in the gametes forming the individual. Veller *et al.* (2020) proposed the parameter P_G as a measure of IDAR. This is the fraction of genome of a 'grandsire' in the gamete originated in parent G. The parameter space is $0 \leq P_G \leq 1$, and 'it is identifiable only with genomic information' from dense chips of SNP markers, originated in animals of at least three generations (Cantet *et al.*, 2017). When $P_G = 0$, the genome in the gamete is originated completely from the granddam, whereas $P_G = 1$ means that the genome in the gamete is originated completely from the grandsire. The expected value under gametic equilibrium is $E(P_G) = 0.50$ (Veller *et al.*, 2020) and grandparents contribute genome in equal parts in the gamete of parent G. It can be shown that the parameters of our 'ancestral regression' (Cantet *et al.*, 2017) carry IDAR, and the parametric equivalences are equal to $0.5\,(P_S - 0.5) = \beta_S$ and $0.5\,(P_D - 0.5) = \beta_D$. The equation for the AR (Cantet *et al.*, 2017) is:

$$a_X = 0.5\,a_{*S} + 0.5\,a_{*D} + \beta_S\left(a_{*SS} - a_{*DS}\right) + \beta_D\left(a_{*SD} - a_{*DD}\right) + \phi_{*X} \tag{1.4}$$

Notation is S for sire, D for dam, SS for paternal grandsire, DS for paternal granddam, SD for maternal grandsire and DD for maternal granddam. All BV in (1.4) are in gametic equilibrium (*), whereas the IDAR is contained in the term $\beta_S\,(a_{SS_} - a_{DS}) + \beta_D\,(a_{SD_} - a_{DD})$.

The recursive equation for BV under PAR parallels the 'animal model' but with unequal parental contributions to the 'information' for the BV of X due to IDAR:

$$a_X = \rho_S\,a_S + \rho_D\,a_D + \phi_X \tag{1.5}$$

The parameters ρ_S and ρ_D are the respective path coefficients of a_S and a_D in a_X, with parameter space $0.25 \leq \rho_S, \rho_D \leq 0.75$, and oftentimes ρ_S and ρ_D will have different values. Compared to (1.4), all BV in (1.5) are under IDAR. In other words, whereas in the AR the animal is regressed to the BV of parents and grandparents, in PAR – as in the animal model – the BV is 'regressed' to parental BV that are now under IDAR. The ρ parameters are shown to be the unique solution of the following system of equations:

$$\begin{bmatrix} 1 + F_S - \boldsymbol{\Sigma}_{S,G}\,\boldsymbol{\Sigma}_G^{-1}\,\boldsymbol{\Sigma}_{G,S} & \Sigma_{S,D} - \boldsymbol{\Sigma}_{S,G}\,\boldsymbol{\Sigma}_G^{-1}\,\boldsymbol{\Sigma}_{G,D} \\ \boldsymbol{\Sigma}_{D,S} - \boldsymbol{\Sigma}_{D,G}\,\boldsymbol{\Sigma}_G^{-1}\,\boldsymbol{\Sigma}_{G,S} & 1 + F_D - \boldsymbol{\Sigma}_{D,G}\,\boldsymbol{\Sigma}_G^{-1}\,\boldsymbol{\Sigma}_{G,D} \end{bmatrix} \begin{bmatrix} \rho_S \\ \rho_D \end{bmatrix} = \begin{bmatrix} \boldsymbol{\Sigma}_{S,X} - \boldsymbol{\Sigma}_{S,G}\,\boldsymbol{\Sigma}_G^{-1}\,\boldsymbol{\Sigma}_{G,X} \\ \boldsymbol{\Sigma}_{D,X} - \boldsymbol{\Sigma}_{D,G}\,\boldsymbol{\Sigma}_G^{-1}\,\boldsymbol{\Sigma}_{G,X} \end{bmatrix} \tag{1.6}$$

The Σ are relatedness coefficients under PAR for the 4 × 4 covariance matrix Σ_G among the four grandparents of X, $\Sigma_{G,S}$, $\Sigma_{G,D}$ and $\Sigma_{G,S}$ are vectors of the PAR coefficients between the grandparents with the sire, the dam and X, respectively, also with PAR coefficients $\Sigma_{S,D}$, $\Sigma_{S,X}$ and $\Sigma_{D,X}$ among them. Following the logic of Wright (1922), the quantity $2\,\rho_S\,\rho_D\,\Sigma_{SD} = F_X$, i.e. the inbreeding coefficient of X under IDAR. Under gametic equilibrium, $\rho_S = \rho_D = 0.5$ and $\Sigma_{SD} = A_{SD}$, $F_X = 0.5\,A_{SD}$. The variance of the residual breeding value (RBV), $\mathrm{Var}(\phi_X)$ is equal to:

$$\mathrm{Var}\left(\phi_X\right) = \left[1 - \rho_S^2\left(1 + F_S\right) - \rho_D^2\left(1 + F_D\right)\right]\sigma_A^2 \tag{1.7}$$

Again, when $\rho_S = \rho_D = 0.5$, $\mathrm{Var}\left(\phi_X\right) = 0.5\left[1 - 0.5\left(F_S + F_D\right)\right]\sigma_A^2$ as in the animal model.

Cox and Grimmett (1984) proved a general central limit theorem for 'associated' random variables under very mild requirements that gives the distribution of BV under PAR: if $\sigma_A^2 > 0$ and $\mathrm{cov}(a_X, a_Y) \rightarrow 0$ when the number of sites grows large, then the BV are asymptotically multivariate normal. Independently, Tallis (1989) used asymptotic theory of 'positive quadrant dependent' random variables to prove that when sites are 'linked' the distribution of BV remains to be multivariate normal.

Under PAR the inverse covariance matrix of BV can be written as A^{-1} (Henderson, 1984) such that $\boldsymbol{\Sigma}^{-1} = \left(\boldsymbol{I} - \boldsymbol{B}'\right)\boldsymbol{D}^{-1}\left(\boldsymbol{I} - \boldsymbol{B}\right)$. All elements in any row of B are equal to zero except for ρ_S and ρ_D in the columns pertaining to the parents of the animal, whereas D is a diagonal matrix whose elements are shown inside the square brackets in (1.7). As the parental path coefficients are all positive, Σ^{-1} has the property that all 'partial correlations' (i.e. the PAR relatedness coefficients) are non-negative (Bølviken, 1982). In addition, the 'number of non-zeroes' in the coefficient matrix of the mixed model equations has the same number in Σ^{-1} as in A^{-1}. It is worth mentioning that Σ^{-1} is calculated by simply modifying the algorithm of Henderson (1984) for A^{-1}, by replacing the 0.5 values in the contributions to A^{-1} by ρ_S or ρ_D, and 0.25 values by the product of ρ_S and ρ_D, or the squared values of these path coefficients. Mixed model equations for a single trait under IDAR following a PAR model for BV then are:

$$\begin{bmatrix} \boldsymbol{X}'\boldsymbol{X} & \boldsymbol{X}'\boldsymbol{Z} \\ \boldsymbol{Z}'\boldsymbol{X} & \boldsymbol{Z}'\boldsymbol{Z} + \boldsymbol{\Sigma}^{-1} \end{bmatrix}\begin{bmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\boldsymbol{a}} \end{bmatrix} = \begin{bmatrix} \boldsymbol{X}'\boldsymbol{y} \\ \boldsymbol{Z}'\boldsymbol{y} \end{bmatrix} \tag{1.8}$$

It is important to indicate that AR and PAR are not equivalent models in the sense of having the same covariance matrix of the phenotypes (Henderson, 1984).

Covariances between breeding values under PAR. Once the r parameters are estimated, the additive covariance under PAR between the BV of an ancestor (A) and a descendant (X) is calculated by a recursion to the BV of X in the following manner:

$$\Sigma_{A,X} = \rho_{S_X}\,\Sigma_{A,S} + \rho_{D_X}\,\Sigma_{A,D} \tag{2.1}$$

At the same time, when neither individual is an ancestor of each other the covariance under PAR requires expanding the BV of both individuals, and it is equal to:

$$\mathrm{cov}\left(a_X, a_Y\right) = \rho_{S_X}\,\rho_{S_Y}\,\mathrm{cov}\left(a_{S_X},\, a_{S_Y}\right) + \rho_{S_X}\,\rho_{D_Y}\,\mathrm{cov}\left(a_{S_X}, a_{D_Y}\right) + \rho_{D_X}\rho_{S_Y}\,\mathrm{cov}\left(a_{D_X},\, a_{S_Y}\right) + \rho_{D_X}\rho_{D_Y}\,\mathrm{cov}\left(a_{D_X},\, a_{D_Y}\right) \tag{2.2}$$

Specific expressions for the covariance between full-sibs (*FS*), paternal-half-sibs (*PHS*) and maternal half-sibs (*MHS*) are calculated by properly replacing in (2.1) with the appropriate expressions described some lines before. Respective covariances for *FS*, *PHS* and *MHS* are:

$$\text{cov}(FS) = \left[\rho_{S_X}\ \rho_{S_Y}\ (1+F_S) + \left[\rho_{S_X}\ \rho_{D_Y} + \rho_{D_X}\rho_{S_Y}\right]\boldsymbol{\Sigma}_{SD} + \rho_{D_X}\rho_{D_Y}\ (1+F_D)\right]\sigma_A^2 \tag{2.3}$$

$$\text{cov}(PHS) = \left[\rho_{S_X}\ \rho_{S_Y}\ (1+F_S) + \rho_{S_X}\ \rho_{D_Y}\ \boldsymbol{\Sigma}_{S_X D_Y} + \rho_{S_Y}\ \rho_{D_X}\ \boldsymbol{\Sigma}_{S_Y D_X}\right]\sigma_A^2 \tag{2.4}$$

$$\text{cov}(MHS) = \left[\rho_{S_X}\ \rho_{D_Y}\ \boldsymbol{\Sigma}_{S_X D_Y} + \rho_{S_Y}\ \rho_{D_X}\ \boldsymbol{\Sigma}_{S_Y D_X} + \rho_{D_X}\rho_{D_Y}\ (1+F_D)\right]\sigma_A^2 \tag{2.5}$$

Genomic selection for animals without own or progeny phenotypes. Suppose X is an animal with neither phenotype nor progeny, but with genotype. After calculating BLUP(a) under PAR, the breeder can calculate BLUP(a_X) by predicting all terms in the 'parental regression' $a_X = 0.5a_S + 0.5a_D + \phi_X$. For individual X, first calculate the RBV of the i^{th} collateral (c_i = sibs, cousins, uncles, great-uncles, etc) of animal X as $\hat{\phi}_{C_i} = \hat{a}_{C_i} - \rho_{S_{C_i}}\hat{a}_{S_{C_i}} - \rho_{D_{C_i}}\hat{a}_{D_{C_i}}$. Then, using mixed model theory (Henderson, 1984) we write:

$$\text{BLUP}\left(\phi_X\right) = \text{cov}\left(\phi_X, \hat{\boldsymbol{\phi}}_C'\right)\left[\text{Var}\left(\hat{\boldsymbol{\phi}}_C\right)\right]^{-1}\hat{\boldsymbol{\phi}}_C \tag{3.1}$$

Full and half-sibs will be the most informative collaterals of X and relevant covariances are displayed in (2.3)-(2.5). We can show that the covariance between the RBV of X in the 'animal model' with the RBV of collateral c_i under IDAR is $\text{cov}\left(\phi_X, \hat{\phi}_{C_i}'\right) = \boldsymbol{\Sigma}_{X,C_i} - \boldsymbol{A}_{X,C_i}$. Hence, on calling t to the number of collaterals of X, expression (3.1) can be written as:

$$\hat{\phi}_X = \text{BLUP}\left(\phi_X\right) = \sum_{i=1}^{t}\left[\frac{\left(\boldsymbol{\Sigma}_{X,C_i} - \boldsymbol{A}_{X,C_i}\right)}{\text{Var}\left(\phi_{C_i}\right)}\hat{\phi}_{C_i}\right] \tag{3.2}$$

The prediction in (3.2) is written and can be programmed using scalar algebra. Finally, the predictor of a_X for genomic selection purposes is equal to:

$$\text{BLUP}\left(a_X\right) = 0.5\hat{a}_{S_X} + 0.5\hat{a}_{D_X} + \hat{\phi}_X \tag{3.3}$$

References

Bølviken E. (1982). Scand. J. Statist. 9: 49-58. http://www.jstor.org/stable/4615855

Cantet R.J.C., García-Baccino C.A., Rogberg-Muñoz A, Forneris N.S., Munilla S. (2017). J Anim. Breed Genet. 134:224–231. https://doi.org/10.1111/jbg.12271

Cox J.T. and Grimmett G. (1984). Annals Prob. 12:514-528. https://doi.10.1214/aop/1176993303

Esary J.D., Proschan F., Walkup, D.W. (1967). Annals Math. Stats. 38:1466-1474. https://doi.10.1214/aoms/1177698701

Henderson C.R. (1984). U. Guelph, Canada.

Malecot, G. M. (1969). W. H. Freeman, San Francisco, USA.

Tallis, G. M. (1989). J. Math. Biol. 24:233-235. https://doi.org/10.1007/BF00275635

Veller C., Edelman N.B., Muralidhar P., Nowak M.A. (2020). Genetics. 216:985–994. https://doi.org/10.1534/genetics.120.303680

Weir B.S. and Cockerham C.C. (1969). Genetics. 61:923-940. https://doi.10.1093/genetics/61.4.923

Weir, B.S, Cockerham, C.C., Reynolds, J. (1980). Heredity 45:351–359. https://doi.org/10.1038/hdy.1980.77

Wright, S. (1922). Amer Nat. 56:330-338. https://doi.org/10.1086/279872

287. Effect of informative polymorphism on accuracy of genomic prediction and heritability

M.S. Tahir[1*], B.S. Olasege[1], K. Wockner[1], W.L.A. Tan[1], L.R. Porto-Neto[2] and M.R.S. Fortes[1]

[1]*School of Chemistry and Molecular Biosciences, University of Queensland, Brisbane, QLD 4072, Australia;* [2]*CSIRO Agriculture and Food, Brisbane, QLD 4067, Australia;* s.tahir@uq.edu.au

Abstract

Early fertility traits are linked with beef cattle profitability, and so they have become target traits for selective breeding. Informative single nucleotide polymorphism (SNP) can potentially increase the accuracy of genomic predictions for target traits. Multi-trait meta-analysis was applied on two reproductive traits measured early in life: pregnancy outcome after the first mating opportunity (PREG1) and rebreeding within the first two mating opportunities (REB) to identify informative SNP associated with these traits. The accuracy of genomic prediction using informative SNP was compared to using the bovine high density (HD) SNP chip in full. Accuracies were calculated in five-fold cross validations, through genomic best linear unbiased prediction (GBLUP) methods. The accuracy of genomic prediction for PREG1 was similar for both SNP sets: 0.18 for the informative SNP set, 0.19 for the HD SNP set. However, the accuracy of genomic prediction for REB was higher when the informative SNP set was used (0.36), in comparison to the HD SNP set (0.19).

Introduction

Genomic selection for puberty-based reproductive traits measured early in life can contribute to increasing cow fertility, and therefore to the profitability of beef farming. Easy to measure puberty-based reproductive traits are heritable, ranging from low to moderate heritability (Cavani *et al.*, 2015; Porto-Neto *et al.*, 2015; Tahir *et al.*, 2021). The heritability of a trait directly impacts the accuracy of its genomic prediction (Goddard, 2009). The use of informative variants in genomic analyses can increase the accuracy of genomic predictions (Sarup *et al.*, 2016; Van den Berg *et al.*, 2016). In this context, investigating the role of informative and non-informative SNP on estimation of the heritability and accuracy of genomic prediction can be insightful for puberty-based reproductive traits. A routine way of identifying informative SNP is to conduct genome-wide association studies (GWAS). However, the number of SNP passing the stringent criteria of significance in a GWAS is usually small for traits with low heritability. An alternative is to use multi-traits meta-analysis which has demonstrated identification of increased number of SNP associated with similar traits in different studies (Bolormaa *et al.*, 2014; Melo *et al.*, 2018).

Keeping in mind the advantage of multi-traits meta-analysis to identify informative SNP, we applied this methodology on two similar puberty-based reproductive traits: pregnancy outcome after first mating opportunity (PREG1) and rebreeding score of first two mating opportunities (REB). The aim was to identify maximum informative SNP and use those in genomic prediction and estimation of heritability of these traits.

Materials & methods

Animals and phenotypes. We selected 5,414 animals for the trait PREG1 and 3,118 animals for the trait REB from Female Fertility Phenobank. Female Fertility Phenobank is a platform that maintains genotype and phenotype database for female beef cattle fertility traits. The selected animals belonged to different breeds and properties from Northern Australia.

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_287, © M.S. Tahir *et al.* 2022

Genotypes. The animals belonging to both traits were originally genotypes for 35K and 50K SNP chips. All the animals were imputed to HD SNP level (700K) using the software Eagle and Minimac as procedure described by Tahir *et al.* (2021).

Splitting animals into training and validation groups. To avoid the bias in downstream genomic prediction analysis, the populations of animals from both traits were split into five training and five validation groups (Table 1). Only training groups of both traits were used in estimation of SNP effects and identification of the informative SNP through processes described below.

GWAS. Genome-wide association studies for all training groups of each trait were performed using linear mixed model implemented in software GCTA (Yang *et al.*, 2011).

$$y = X\beta + Z\mu + S\alpha + e$$

where, X is vectors of fixed effects, β is estimates of fixed effects. Z is an incidence matrix of random polygenic effects; μ is the estimate of random polygenic effects ~N(0, G$\sigma\mu^2$) where G is the genomic relationship matrix (GRM) based on all SNP. Residual is represented as 'e' ~N(0, σe^2). S is an incidence matrix of genotypes and α is the estimate of SNP effects. A fixed effect combined for information of property and year of birth was used in the model. Similarly, to correct for breed differences, the top two principal components were also used as covariates.

Multi-trait meta-analysis. The SNP effects and their standard error obtained from GWAS of corresponding training groups of both traits were used to perform multi-trait meta-analysis using SNP & Variation Suite v8.x Golden Helix (Golden Helix).

$$\text{Multi-trait } \chi^2 = t'_i \, V^{-1} \, t_i$$

where, χ^2 is multi-trait test statistics approximately distributed as chi-squared, t_i is a vector 2 × 1 of the $_i$th SNP effects divided by their respective standard errors, t'_i is the transpose vector of t_i, and V^{-1} is an inverse of the 2 × 2 correlation matrix of the correlations between the *t*-values.

From multi-trait meta-analysis, all the SNP associated with both traits, simultaneously (empirical *P*-value<9.0×10^{-5}) were selected and termed as informative SNP sets. The informative SNP were extracted from HD SNP set for all training groups of both traits. This process resulted in formation of five training groups with informative SNP sets.

Genomic prediction. The training groups of both traits were used to estimate GEBVs using informative and HD SNP sets through GBLUP method implemented in software GCTA.

$$y = X\beta + Z\mu + e$$

The GEBVs estimated through equation above were back solved to estimate the GBLUP-SNP effects for the training groups of both SNP sets using the software GCTA.

Accuracy of genomic prediction. The SNP effects from training groups of both traits were used to calculate the GEBVs of the validation sets of both traits. The accuracy of genomic prediction was calculated as:

Accuracy = r = correlation between GEBVs, and phenotypes adjusted for fixed effects.

Multiple GRM model. To account for random polygenic effects in the estimation of heritability of the traits with informative SNP set, the GRM of informative SNP was also modelled along with GRM of non-informative SNPs (HD SNPs minus informative SNPs), simultaneously.

Results

In single trait single training group GWAS, average number of SNP suggestively associated with PREG1 were 51 (empirical *P*-value<0.9×10^{-5}), while in case of REB, this number was 52. All training sets, through multi-trait-meta-analyses of both traits, identified an average of 516 SNPs (empirical *P*-value<0.9×10^{-5}) associated with both traits (Table 1).

The heritability estimated by HD SNP set was 0.17 (S.E 0.02) for PREG1 and 0.21 (S.E 0.03) for REB, while the heritability estimated by the informative SNP set was 0.21 (S.E 0.02) for PREG1 and 0.24 (S.E 0.02) for REB (Table 1). In single GRM models, the heritability explained by informative SNP set was more than the HD SNP set for both traits. In multiple GRM models, the heritability of PREG1 explained by informative SNP set was reduced to 0.19 while the heritability of REB explained by informative SNP set remained same as in single GRM model (Table 2).

For PREG1, informative SNP set produced similar accuracy of genomic prediction (0.18) as did HD SNP set (0.19). For REB, the informative SNP set, however, performed much better with accuracy of genomic prediction of 0.36 in comparison to HD SNP set which produced accuracy of genomic prediction of 0.19.

Discussion

This study aimed to investigate the effect of using informative SNP for the estimation of heritability and accuracy of genomic prediction of puberty-based reproductive traits in female beef cattle. In this study, the heritability estimates of PREG1 and REB by HD SNP set were comparable to the previous studies who reported heritabilities for similar traits such as pregnancy outcome after first fixed-time insemination (0.18)

Table 1. Estimates of heritability and accuracy of genomic prediction for PREG1 and REB using HD and informative SNPs.[1]

Trait	h^2	Accuracy	SNP set	SNPs (n)	Total population (n)	Training population (n)	Validation population (n)
PREG1	0.17	0.19	HD	669,396	5,414	4,331	1,083
PREG1	0.21	0.18	Informative	516	5,414	4,331	1,083
REB	0.21	0.19	HD	662,896	3,118	2,495	623
REB	0.24	0.36	Informative	516	3,118	2,495	623

[1] h^2 = heritability.

Table 2. Heritability estimates by informative SNP sets in multiple GRM model.

Trait	Multiple GRMs	Heritability explained
PREG1	Informative SNP	0.19
	Non-informative SNPs	0
REB	Informative SNP	0.24
	Non-informative SNPs	0

and rebreeding (0.22) (Cavani *et al.*, 2015; Porto-Neto *et al.*, 2015). The results indicate that the estimate of heritability of PREG1 by informative SNP set was slightly inflated which was confirmed by application of multiple GRM model (Table 1, 2). The estimate of better heritability by informative SNP set for REB seems unbiased as the informative SNP set resulted in same heritability in multiple GRM model as in single GRM model. Heritability of the traits impacts the accuracy of genomic prediction (Zhang *et al.*, 2014). In this study, the accuracy of genomic prediction was coherent with the estimates of heritability in case of PREG1 by both informative and HD SNP sets without any noticeable difference between performance of both SNP sets. However, in case of REB, the accuracy of genomic prediction was almost double using informative SNP set in comparison to HD SNP set. This increase in accuracy of genomic prediction of REB is very encouraging when compared to the other reports who reported small to moderate increase in accuracy of genomic prediction of fertility traits using biologically informed polymorphisms (Abdollahi-Arpanahi *et al.*, 2017; Nani *et al.*, 2019). Further exploration of these traits with larger dataset can confirm these findings with more confidence and can present these traits as suitable option to be adopted for genomic selection programmes in beef cattle industry.

References

Abdollahi-Arpanahi R., Morota G. and Penagaricano F. (2017) *J. Dairy Sci. 100*(12): 9656-9666. https://doi.org/10.3168/jds.2017-13288

Bolormaa S. Pryce J. E., Reverter A., Zhang Y., Barendse W. et al (2014) *PLoS Genetics 10*(3): e1004198. https://doi.org/10.1371/journal.pgen.1004198

Cavani L., Garcia D. A., Carreno L. O., Ono R. K., Pires M. P. el al (2015) *J. Anim. Sci. 93*(7): 3287-3291. https://doi.org/10.2527/jas.2015-8970

Goddard M. (2009) *Genetica 136*(2): 245-257. https://doi.org/10.1007/s10709-008-9308-0

Melo T. P., Fortes M. R. S., Bresolin T., Mota L. F. M., Albuquerque L. G. et al (2018) *J. Anim. Sci. 96*(10): 4087-4099. https://doi.org/10.1093/jas/sky289

Nani J. P., Rezende F. M., and Peñagaricano F. J. B. G. (2019) *BMC Genomics 20*(1): 1-10. https://doi.org/10.1186/s12864-019-5644-y

Porto-Neto L. R., Edwards S., Fortes M. R., Lehnert S. A., Reverter A. el al (2015) *J. Anim. Sci. 93*(11): 5119-5127. https://doi.org/10.2527/jas.2015-9401

Sarup P., Jensen J., Ostersen T., Henryon M., and Sorensen P. (2016) *BMC Genetics 17*(1):1-16. https://doi.org/10.1186/s12863-015-0322-9

SNP & Variation Suite ™ [Software]. Golden Helix, I. (Version 8.x) http://www.goldenhelix.com

Tahir M. S., Porto-Neto L. R., Gondro C., Shittu O. B., Wockner K. et al (2021) *Genes 12*(5): 768. https://doi.org/10.3390/genes12050768

Van den Berg I., Boichard D., Guldbrandtsen B., and Lund M. S. (2016) *G3: Genes 6*(8): 2553-2561. https://doi.org/10.1534/g3.116.027730

Yang J., Lee S. H., Goddard M. E., and Visscher P. M. (2011) *Am. J. Hum. Genet. 88*(1): 76-82. https://doi.org/10.1016/j.ajhg.2010.11.011

Zhang Y. D., Johnston D. J., Bolormaa S., Hawken R. J., and Tier B. (2014) *Anim. Prod. Sci. 54*(1): 16-24. https://doi.org/10.1071/an13016

288. Long-distance associations generate erosion of genomic breeding values of candidates for selection

D. Boichard[1], S. Fritz[1,2*], P. Croiseau[1], V. Ducrocq[1], B. Cuyabano[1] and T. Tribout[1]

[1]Université Paris-Saclay, INRAE, AgroParisTech, GABI, 78350 Jouy-en-Josas, France; [2]Allice, 75012 Paris, France; sebastien.fritz@inrae.fr

Abstract

Most validation studies of genomic evaluation observe inflation, i.e. regression coefficients of the later phenotypes on early predictions smaller than one. In this paper, we show that this is due to non-zero contributions of distant QTLs, especially those located in other chromosomes than the SNPs. These effects result from linkage disequilibrium (here measured with r^2) present in the reference data but eroded in the population of candidates. In six French dairy cattle breeds average r^2 is low across chromosomes but a substantial proportion of SNP pairs show r values higher than 0.05. A simulation study based on real genotypes from the Normande breed shows that 5% of the SNP effect's variance is explained by distant QTL with associations not maintained across generations. A real Holstein example illustrates extensive across chromosomes covariances between effects. It is thus recommended to erode SNP effects to compute unbiased genomic values of candidates to selection.

Introduction

In genomic evaluation, SNP effects are estimated in a reference sample and applied to selection candidates. This method is extensively used to select candidates at an early stage of their life or yet without phenotypic information. The standard interpretation is that a trait depends on QTLs, and those genetic markers in close LD to these QTL are good proxies for them. Implicitly, this assumes that estimated marker effects reflect those of the neighbouring QTL. Under this assumption, associations observed in the reference sample should be very similar in the next generation as short-distance LD erodes slowly due to recombination.

It is, however, well known that genomic evaluation efficiency is highly dependent on the relatively close relationship of the candidates to the reference sample (Habier *et al.*, 2007, 2013; Legarra *et al.*, 2008; Pszczola *et al.*, 2012). Many studies have shown the limited gain in accuracy in multi-breed evaluation (Erbe *et al.*, 2012; Hozé *et al.*, 2014), illustrating that distant reference data are not informative. Other studies have shown decrease in accuracy over generations when the reference sample is not updated (Sonneson *et al.*, 2009; Solberg *et al.*, 2009). Moreover, it has been observed that the absence of the parents in the reference sample directly influences the prediction accuracy of the selection candidates. All these results suggest that SNP effects erode as the distance between candidates and reference sample increases.

Despite these observations, validation studies of genomic evaluations are generally based on the regression of later performances on the early predictions. These studies frequently observe an inflation pattern, i.e. the regression coefficient is systematically below 1, meaning that later performances of the best candidates were below those initially predicted (and later performances, if any, of the worst candidates were above those initially predicted).

An interpretation is that some long-range LD exists, even between chromosomes, and, accordingly, many markers may capture some partial effects of supposedly unlinked QTL. Long-distance LD is much lower than short-distance LD, but the number of variants involved is much higher and their joint effects can explain a substantial proportion of the predicted genomic value. In this paper, we show that markers capture some effects of distant QTL due to long-distance LD. Because this long-range LD dissipates over

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_288, © D. Boichard *et al.* 2022

generations, the erosion of marker effects must be taken into account to predict the genomic value of the candidates.

Materials & methods

LD across chromosomes was measured in the female reference data of six French dairy cattle breeds (Holstein, Montbeliarde, Normande, Abondance, Tarentaise, Vosgienne). The size of these reference samples varied from 2,617 to 362,363. Vosgienne, Tarentaise, and Abondance are regional mountain breeds, Montbeliarde and Normande are large national populations (18 and 7% of the French dairy herd, respectively) and Holstein (70%) is international. One every 20 SNP was considered, resulting in a sample of ~3 million r^2 values for each breed (vs ~1.2 billion in total).

To study the impact of long-distance LD on SNP effects, we used the Normande population (n=69,220 genotyped cows with records). Only the first five chromosomes (ns=13,608 SNP) were considered. Two hundred (nq=200) additive QTLs were sampled at random among those SNPs with MAF>0.02 over the first four chromosomes whereas chromosome 5 remained empty of QTLs. Additive QTL effects were independently drawn from a normal distribution assuming a heritability of 0.3. Two scenarios were tested: (1) the SNP-BLUP model accounted for the 13,608 SNPs including the QTL, a situation believed to minimize the impact of long-distance LD; (2) the SNP-BLUP model did not include the QTL, a situation in which QTL effects are to be distributed on more markers. The contribution of each QTL to each SNP was estimated as follows. The SNP-BLUP equations can be written as $(M'M + \lambda I)\,\hat{s} = M'y$, with M the (N × ns) matrix of centered and scaled genotypes, s the vector of SNP effects, y the vector of phenotypes adjusted for fixed effects, and $\lambda = \sigma_e^2 / \sigma_s^2$. According to the simulation, the phenotype can be written as $y = Pq + e$, i.e. the sum of nq QTL effects and an error term, with P the (N × nq) matrix of genotypes at the QTL level and q the vector of true QTL effects. Therefore, the equations can be rewritten as $\hat{s} = (M'M + \lambda I)^{-1} M'(Pq + e)$. If c_i is line i (corresponding to SNP i) of $C = (M'M + \lambda I)^{-1}$, the contribution of QTL j to SNP effect i is $c_i M' P_j q_j$. These contributions were squared and summed within each of the 6 following categories of distance between QTL and SNP: (1) the QTL is the SNP; (2) the distance between the QTL and the SNP is less than 5 Mb; (3) this distance is between 5 and 20 Mb; (4) the QTL and the SNP are on the same chromosome but more than 20 Mb apart; (5) the QTL and the SNP are on different chromosomes; category (6) corresponded to the special case of SNPs on chromosome 5 with no QTL. Thirty replicates were simulated.

This way to estimate contributions by accumulating squared effects by individual SNP does not account for the covariances between SNP effects. This method is therefore only an approximation (De los Campos *et al.*, 2015). However, the long-distance covariances have a major impact and reinforce the contribution of distant markers. They are illustrated by the part of DGV variance explained by haplotypes of increasing size. These results were from the French Holstein evaluation for resistance to paratuberculosis (Sanchez *et al.*, 2022, WCGALP).

Results

Table 1 presents different LD statistics across chromosomes in the reference samples of six French dairy cattle breeds. Average r^2 values are small, which at first glance may suggest that across chromosomes effects are negligible. These values decreased when the size of the breed (reference sample, number of females in the breed, or effective size) increased. Note, however, the parameter of interest here is r instead of r^2, as the impact of a QTL on a SNP effect is proportional to r. The proportion of SNP pairs with |r| greater than 5% varied from 1.5 to 33%, depending on the breed. The larger the reference sample, the lower this proportion. Nevertheless, even with a few percentuals, one to several thousand SNP have some correlations to QTLs (assuming that these r distributions are the same between SNP and QTL).

Table 1. Statistics of |r| and r^2 values across the 29 chromosomes in reference samples of six French dairy cattle breeds (selection of 1 every 20 SNP within chromosome).[1]

Breeds	# cows in reference sample	Mean(|r|) *	% |r|>0.03	% |r|>0.05	Mean(r^2)
Vosgienne	2,617	0.0420	54	33	0.0029
Tarentaise	3,788	0.0225	41	18	0.0015
Abondance	7,115	0.0268	35	15	0.0012
Normande	69,220	0.0206	25	7	0.00073
Montbeliarde	185,053	0.0173	18	4	0.00053
Holstein	362,363	0.0148	12	1.5	0.00038

[1] Statistics based on 2,812,741 to 3,231,80 r-values per breed.

Table 2 presents the relative contribution of the 6 classes of QTL-SNP distance, expressed in proportion of the total sum of squared SNP effects. In the model including the causal variants, these variants succeeded to capture the largest part of the genetic variability, leaving little contributions to the nearby markers. Contributions of distant QTLs were nearly zero. In the model without causal variants, more similar to a real situation, the distribution of contributions was spread on more distant markers, even across chromosomes, although the largest part of the variability was still captured by the closest markers. This result was obtained with only five chromosomes and the contribution to distant markers is likely to be even larger with 29 chromosomes, due to the multiplicity of the across chromosomes contributions.

These contributions at long-distance are the consequences of long-distance LD (M'P is a measure of the covariance between SNP and QTL). This basal LD is a function of the effective size Ne and remains more or less constant across generations on average at the genome level. Two opposite forces explain this steady-state LD: (1) The existing long-range LD is reduced by a 0.5 factor at each generation due to segregations; and (2) it is regenerated by drift (and possibly selection) at random. However, this is not true for a given pair of loci and the LD between 2 unlinked loci existing in the reference sample is reduced by a 0.5 factor.

Finally, Table 3 illustrates that the part of the DGV variability explained was very small with individual variants and gradually increased with haplotype size, showing the huge amount of positive covariances between marker effects. Even when entire chromosomes were considered, the sum of the 29 variances represented only 58% of the total, showing that 42% of the total variance originated from inter-chromosomal covariances.

Table 2. Relative contribution of each of the 6 classes of QTL-SNP pairs defined according to their distance. The contribution of a class is the sum of squared effects. Results on 30 replicates.

Classes of QTL-SNP relationship	Relative contribution (%)		# of QTL contributions to SNP
	Model including causal variants	Model without causal variants	
1: QTL=SNP	95.2	0.0	200
2: distance(QTL,SNP) <5 Mb	4.6	73.8	42,660±530
3: 5 Mb < distance <20 Mb	0.1	19.6	112,500±1,500
4: distance(QTL,SNP) >20 Mb	0.02	0.6	410,000±3,700
5: QTL and SNP on different chromosomes	0.02	2.4	1,671,000±4,500
6: SNP on chromosome 5 (without QTL)	0.02	3.5	485,200

Table 3. Proportion of DGV explained by haplotypes of various sizes in French Holstein.[1]

Interval length (# markers)	1 (single markers)	10	100	500	Entire chromosomes	Whole- genome
% variance explained	1	8.5	26	47	58	100
# intervals	53,469	5,360	548	121	29	1

[1] Variances were computed from partial DGV per interval in the evaluated population.

Conclusions

The practical consequences are important. In the overall prediction of candidates, the short distance contributions are rather stable, slightly eroded by recombinations (Dekkers *et al.*, 2021), but the long-distance contributions are divided by 2 at each generation. According to the relative weight of short and long-distance LD, the overall erosion may strongly fluctuate, but it is always important. Assuming this erosion factor per generation (ρ) is known, eroded direct genomic values (DGV) of candidate i can be estimated from the raw DGV of i and of its parents s and d: $DGV_i = 0.5\ (DGV_s + DGV_d) + (1-\rho)^k\ MS_i$, with MS being the Mendelian sampling component and k the number of generations between i and the reference sample on both paternal and maternal pathways. This formula is recursive, as sire's and/or dam's DGV should also be eroded if they do not belong to the reference.

Further investigation is needed to estimate this ρ erosion factor. It of course depends on basic LD in the population (i.e. on Ne and the genome length L). But it also probably depends on the genetic architecture of the traits (number and magnitude of QTL effects) and the model used (SNP-BLUP/GBLUP vs Bayesian models). However, one can anticipate that: (1) inflation factors $(1 - \hat{b})$ frequently observed are between 0.1 and 0.2 and give an idea of the importance of the phenomenon; (2) models including causal variants are more persistent and less subject to erosion; (3) breeding schemes with accelerated generations without updating reference data accumulate more erosion; (4) models including a residual polygenic component present less inflated predictions because they combine two estimates of the MS term, the genomic one which is inflated and a polygenic one which is equal to zero. We believe that accounting for erosion is more rigorous and accurate than adding an arbitrary polygenic effect.

References

Dekkers J.C.M., Su H.L., and Cheng J. (2021) Genet Sel Evol 53, 81. https://doi.org/10.1186/s12711-021-00675-6

de los Campos G., Sorensen D., and Gianola D. (2015) PLoS Genet 11, e1005048. https://doi.org/10.1371/journal.pgen.1005048

Erbe M., Hayes B.J., Matukumalli L.K., Goswami S., Bowman P.J., *et al.* (2012) J Dairy Sci 95(7), 4114-4129. https://doi.org/10.3168/jds.2011-5019

Habier D., Fernando R.L., and Dekkers J.C.M. (2007). Genetics 177, 2389-2397. https://doi.org/10.1534/genetics.107.081190

Habier D., Fernando R.L., and Garrick D.J. (2013) Genetics, 194, 597–607, https://doi.org/10.1534/genetics.113.152207.

Hoze C., Fritz S., Phocas F., Boichard D., Ducrocq V., *et al.* (2014) J Dairy Sci 97(6), 3918-3929. https://doi.org/10.3168/jds.2013-7761

Legarra A., Robert-Granie C., Manfredi E., and Elsen J.M. (2008) Genetics 180(1), 611-618. https://doi.org/10.1534/genetics.108.088575

Pszczola M., Strabel T., Mulder H.A., and Calus M.P.L. (2012) J Dairy Sci 95(1), 389-400. https://doi.org/10.3168/jds.2011-4338

Solberg T R., Sonesson A.K., Woolliams J.A., Odegard J., and Meuwissen T.H.E. (2009) Genet Sel Evol 41, 53. https://doi.org/10.1186/1297-9686-41-53

Sonesson A.K., Meuwissen T.H.E. (2009) Genet Sel Evol 41, 37. https://doi.org/10.1186/1297-9686-41-37

289. Definition of reliabilities for models with metafounders

M. Bermann[1], I. Misztal[1], D. Lourenco[1], I. Aguilar[2] and A. Legarra[3*]

[1]*Animal and Dairy Science, University of Georgia, Athens, GA 30605, USA;* [2]*INIA, Montevideo, Uruguay;* [3]*INRAE, 31526 Castanet Tolosan, France;* andres.legarra@inrae.fr

Abstract

For models with several base populations (Unknown Parent Groups or Metafounders), the usual definition of reliability is ill-posed. Here we propose to define reliability based on contrasts with one or several metafounders, leading to a sounder genetic interpretation. In the case of a single metafounder, our definition equals the definition of reliability for the classical animal model. This definition also allows expressing the reliability of contrasts of metafounders, which may be of interest to decide if their setup is estimable with sufficient reliability. All desired quantities can be obtained from elements of the inverse of the MME.

Introduction

Reliabilities (*Rel*) are usually defined as the square of the correlation between estimated and true breeding values (EBV and TBV) on conceptual repeated sampling. When different base populations exist, they are modelled as Unknown Parent Groups (UPG; fixed or random) or Metafounders (MF; random). The total EBV (u) can be seen as a sum of genetic levels of the different base populations plus random deviations. An EBV has a reference point with a value of '0'. In the animal model with neither UPG nor MF, the base population is infinite with a mean breeding value of 0, so the EBVs already refer to this base population. In fixed UPG, EBVs are non-estimable functions and the '0' depends on the particular generalized inverse used for solving. There is thus no meaningful base population the EBVs refer to. In the case of MF, the EBVs refer to an ideal population of maximum heterozygosity, which has a value of 0.

Reliabilities need to refer to a base population, which is the '0' point when using some form of BLUP (Tier *et al.*, 2018). In the case of UPGs, the fact that they are fixed effects means that the reliability cannot be properly defined or computed. For MF, reliabilities obtained referring to the ideal population made little genetic sense as this population doesn't actually exist. We show that a sensible thing to do is to refer reliabilities to one (or several) MF using a contrast. In the case of a single base population, these reliabilities are identical to the classical ones.

On the other hand, UPGs or MFs are essential for correct genetic evaluations when animals lack pedigrees or when crosses of different breeds / populations / countries exist. Some genetic evaluations fit as many as 400 UPGs. In practice, UPG/MF may include few animals with phenotypes; therefore, they may not be accurately estimated, leading to suboptimal selection decisions (Kennedy, 1981). We therefore propose to routinely evaluate reliabilities of contrasts among MF, to ascertain if MF are correctly set up and estimated.

In (with neither UPG nor MF) BLUP and ssGBLUP, 'classical' reliabilities for individual i can be obtained using explicit inverses of the Mixed Model Equations, e.g. $Rel^c = 1 - PEV/(A_{ii}\sigma_u^2)$ using prediction error variance (PEV). However, including UPG or MF in the model leads to practical and conceptual problems. When UPG are fit as fixed effects, the usual equation for reliability does *not* hold, because $Var(\hat{u}) \neq Cov(\hat{u}, u)$ when u includes the UPG (Henderson, 1984). The expression is approximately correct *only* when UPG are estimated with very high precision. This is the case (roughly) when a large number of first-generation offspring from UPG have records, and UPG are not confounded. When UPG are fit as random effects with variance σ^2_{UPG}, the equation holds, but the UPG variance component should be included in the denominator e.g. $Rel_i^c = 1 - PEV/(\sigma^2_{UPG} + A_{ii}\sigma_u^2)$.

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_289, © M. Bermann *et al.* 2022

On the other hand, MF are a comprehensive solution to model genetic relationships (Γ) between and within different base populations, for instance the increase of overall relationship within a breed over time (Legarra *et al.*, 2015). These are reflected in modified relationships $A^{(\Gamma)}$ and genetic variance $\sigma^2_{u(\Gamma)}$. By construction, MF are deemed to be more adequate than UPG, in particular for ssGBLUP applications. Because MF are random effects, in principle, one could use $1 - PEV^{(\Gamma)}/(A^{(\Gamma)}_{ii}\,\sigma^2_{u(\Gamma)}))$ to obtain reliabilities. However, empirical assessment showed that this equation underestimated reliabilities for proven bulls.

Materials & methods

The reliability from PEV in the MF case computed as $1 - PEV^{(\Gamma)}/(A^{(\Gamma)}_{ii}\,\sigma^2_{u(\Gamma)}$ refers to this ideal population, but we are not interested in this reliability as it has no genetic interpretation or use. It is more meaningful to refer the reliability of the EBV to a single, genetically well-defined population (Tier *et al.*, 2018). This population can simply be one of the MF. In a purebred population with MF modelling missing pedigree, it can simply be the oldest MF. Thus, instead of considering the EBV $\hat{u}_i^{mf}$, we consider the contrast $\hat{u}_i^{mf} - \hat{u}_{mf}^{mf}$. To clarify, we use the super-index *mf* for an EBV modelled with metafounders, whereas subindices refer generally either to individual *i* or to the reference metafounder *mf*.

Reliability of EBV with metafounders. Given the contrast of EBVs, we can define the reliability of the contrast as $Rel_i^{mf} = Rel\left(\hat{u}_i^{mf} - \hat{u}_{mf}^{mf}\right)$. This is equal to:

$$Rel_i^{mf} = Rel\left(\hat{u}_i^{mf} - \hat{u}_{mf}^{mf}\right) = 1 - \frac{Var\left(\left[u_i^{mf} - u_{mf}^{mf}\right] - \left[\hat{u}_i^{mf} - \hat{u}_{mf}^{mf}\right]\right)}{Var\left(\left[u_i^{mf} - u_{mf}^{mf}\right]\right)}$$
$$= 1 - \frac{PEV\left(\hat{u}_i^{mf}\right) + PEV\left(\hat{u}_{mf}^{mf}\right) - 2PEC\left(\hat{u}_i^{mf}, \hat{u}_{mf}^{mf}\right)}{\left(A^{(\Gamma)}_{ii} + A^{(\Gamma)}_{mf,mf} - 2A^{(\Gamma)}_{i,mf}\right)\sigma^2_{u(\Gamma)}}$$

This is the general equation to obtain the reliability of the contrast 'individual – reference MF' $(\hat{u}_i^{mf} - \hat{u}_{mf}^{mf})$ from the inverse of the MME. In the equation, $PEC\left(\hat{u}_i^{mf}, \hat{u}_{mf}^{mf}\right)$ is the prediction error covariance, and $A^{(\Gamma)}_{ij}$ are corresponding relationship coefficients (which in ssGBLUP will be from matrix $H^{(\Gamma)}$. Interestingly, we can define different reliabilities with different metafounders; for instance, in a complex, unbalanced cross, for the same animal, one can potentially define different reliabilities for each origin.

In the particular case of a single metafounder, it can be shown that $Rel_i^{mf} = Rel\left(\hat{u}_i^{mf} - \hat{u}_{mf}^{mf}\right)$ is equal to the classical 'animal model, no UPG' reliability $Rel_i^c = 1 - PEV/(A_{ii}\sigma^2_u)$ where PEV is from the 'classical' animal model, i.e. $Rel_i^{mf} = Rel_i^c$. This is intuitively expected.

Reliability of metafounders' contrasts estimation. Following the same logic, we define a reliability of contrasts of MF estimation, namely

$$Rel^{mf}_{mf(i),mf(j)} = Rel\left(\hat{u}^{mf}_{mf(i)} - \hat{u}^{mf}_{mf(j)}\right)$$
$$= 1 - \frac{PEV\left(\hat{u}^{mf}_{mf(i)}\right) + PEV\left(\hat{u}^{mf}_{mf(j)}\right) - 2PEC\left(\hat{u}^{mf}_{mf(i)}, \hat{u}^{mf}_{mf(j)}\right)}{\left(\Gamma_{i,i} + \Gamma_{j,j} - 2\Gamma_{i,j}\right)\sigma^2_{u-related}}$$

this allows to appraise the quality of MF estimates in a scale from 0 to 1, something very difficult to do with UPG. Indeed, Kennedy (1981) suggested to check the standard error of contrasts of UPG to avoid wrong selection decisions. It is worth noting that the definition of Rel^{mf} above is strictly the same as for animals.

Computing aspects. To obtain reliabilities in 'classical' animal models (neither UPG nor MF), only the PEV from the diagonal of the inverse of the MME are needed, usually from the sparse inverse. In our

proposal, to obtain Rel_i^{mf}, the vector of $PEC\left(\hat{u}_i^{mf}, \hat{u}_{mf}^{mf}\right)$ is needed. This vector can be obtained by solving the equation $(MME)x = y$, where y contains 1 in the position of the reference MF and 0 elsewhere; on the output, x contains PEC of the reference MF versus all the other unknowns in the MME. Because MF are random effects, this vector of PECs (and all PEVs in the procedure) is invariant to the usual lack of full rank of the MME.

Additionally, the relationships of the metafounders with the other individuals can be obtained by solving $\boldsymbol{A}^{(\Gamma)^{-1}}\boldsymbol{x} = \boldsymbol{y}$ where $y = 1$ in the location of the metafounder and 0 otherwise. The solution vector x contains the values of the row of $A^{(\Gamma)}$ corresponding to the reference MF, and hence all $A_{i,mf}^{(\Gamma)}$. A similar procedure can be used for SSGBLUP $H^{(\Gamma)}$, combined with Legarra *et al.* (2020) to obtain $diag(H^{(\Gamma)})$.

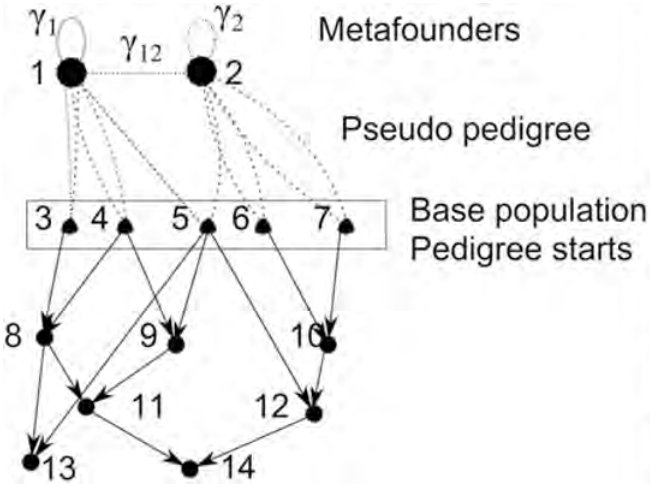
Example. The 12-individuals pedigree with 2 MF described in Legarra *et al.* (2015) (Figure 1) was considered, with:

$$h^2 = 0.5\ ,\ \boldsymbol{\Gamma} = \begin{pmatrix} 0.55 & 0.48 \\ 0.48 & 0.77 \end{pmatrix},$$

and one record per individual. The model includes an overall (fixed) mean. We computed Rel^{mf1}, Rel^{mf2} and Rel^{c} (in the last case, ignoring MF in the model).

Results

Numbers from the example are shown in Table 1. It can be seen that Rel_i^{mf} is roughly similar to Rel_i^{c} in this (rather artificial) example. It is interesting to note that the reliability is higher with respect to the MF who is *not* the ancestor of the individual, and this can be explained because the phenotype of an individual contributes both to its EBV and to the EBV of its ancestral MF. It is also of interest to see that the reliability of the contrast of metafounders is not very high.



Pedigree		
1	0	0
2	0	0
3	1	1
4	1	1
5	1	2
6	2	2
7	2	2
8	3	4
9	4	5
10	6	7
11	8	9
12	5	10
13	8	5
14	11	12

Figure 1. Example pedigree with metafounders.

Table 1. Reliability of EBV based on the contrast with each metafounder (Rel^{mf}) and from the 'classical' animal model with no MF or UPG (Rel^{c}).[1]

Individual	Rel^{mf1}	Rel^{mf2}	Rel^{c}
1		0.23*	
2	0.23*		
3	0.39	0.49	0.43
4	0.40	0.50	0.44
5	0.40	0.41	0.45
6	0.46	0.35	0.43
7	0.46	0.35	0.43
8	0.36	0.51	0.42
9	0.34	0.45	0.41
10	0.48	0.32	0.41
11	0.37	0.50	0.43
12	0.41	0.31	0.39
13	0.30	0.42	0.37
14	0.33	0.37	0.37

[1] An asterisk indicates reliability of the contrasts among the two metafounders.

Discussion

This work provides a rationale for the definition and computation of reliabilities in a model with several founder populations – something that had not been done before. Our proposal should give more meaningful reliabilities in populations with missing pedigrees or crossbreds. In addition, we emphasize that the computations involved are doable for medium-size data sets, and approximations can be used for larger ones. Moreover, the reliability of contrasts among MF may be used to verify whether differences across MF are estimated with enough precision, and therefore to help defining number and definition of MF.

Acknowledgements

This work received financing from European Unions' Horizon 2020 Research & Innovation Programme, Grant/Award Number: N°772787, Smarter.

References

Henderson C.R. (1984) Applications of Linear Models in Animal Breeding. University of Guelph, Guelph.

Kennedy B.W. 1981. J Dairy Sci 64:689–697. https://doi.org/10.3168/jds.S0022-0302(81)82631-8.

Legarra A., Christensen O.F., Vitezica Z.G., Aguilar I., and Misztal I.. (2015) Genetics 200:455–468. https://doi.org/10.1534/genetics.115.177014

Tier B., Meyer K., and Swan A.. (2018) Proc. of the 11th WCGALP, Auckland, New Zealand. http://www.wcgalp.org/proceedings/2018/implied-genetic-effects-relationships-and-alternate-allele-coding

290. A new metric to assess reference populations for genomic selection in Australian beef breeds

K.L. Moore*, M.H. Ferdosi, C.G. Girard, S.F. Walkom and D.J. Johnston

AGBU, a joint venture of NSW Department of Primary Industries and University of New England, 2351, Armidale, Australia; kmoore7@une.edu.au

Abstract

Research has shown that to maximise benefits (i.e. EBV accuracy and spread and decreased generation interval) of genomic selection, reference populations of genotyped and phenotyped animals need to be large, diverse in relationships and closely related to the selection generation. In practice, applying these design principles can be challenging as metrics to aid decision making are often not available. This paper applies a metric that objectively describes reference populations and their impact on accuracy for seven Australian beef populations. All populations for live weight and abattoir carcase had different levels of phenotyping, genotyping and overall size. Carcase traits had the lowest reference population size, and assessment of average relatedness in almost all the breeds showed there were herds that were not currently represented in the reference population. Generally, increasing reference size resulted in increased accuracy, but there were some exceptions. The metrics described in this paper are easy to apply and can assist in the construction of reference populations.

Introduction

Single-step GBLUP has been implemented into Australian BREEDPLAN genetic evaluations (Johnston *et al.* 2018). Reference population data to underpin genomics are generated by a number of projects, and are run by either breed societies or research organisations. Reference population design is critical and impacts the benefit obtained from genomics. The accuracy of genomic selection is directly impacted by reference population size, trait heritability, and effective population size (Goddard and Hayes, 2009), as well as relatedness amongst the reference animals, and relatedness to selection animals (Pszczola *et al.* 2011). As the relationship between the reference and selection population increases, smaller reference population sizes are required to achieve the same level of accuracy (Lee *et al.* 2017). Therefore, multiple design principals need to be balanced when constructing reference populations. Applying these design principals can be challenging for many practical reasons, including that there are limited quantitative metrics and tools available to make informed objective decisions on sire selections for building reference populations. The aim of this study was to develop a metric that describes and assesses reference populations for suitability for genomic selection on a breed, herd or individual level, and to explore the relationship between reference size and the realised accuracy observed in industry datasets from single-step genetic evaluations.

Materials & methods

Seven Australian beef populations represented a wide range of population structures and will be anonymised. Breeds A, B, D and G have established BREEDPLAN single-step genetic evaluations with breed F recently implementing single-step. Breed C and E are yet to implement single-step into their genetic evaluations. All metrics described in this paper were applied to animals born 2010 or later. Live weight and carcase reference populations were considered, with reference animals being those involved in structured projects and also qualifying industry animals. The reference population was defined as animals with both phenotype and genotype available, as well as genotyped but un-phenotyped sires with five or more progeny phenotypes not already included in the reference. For inclusion in the reference population, animals required at least one live weight record (200, 400 or 600 day) for the live weight reference population, and at least one of abattoir

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_290, © K.L. Moore *et al.* 2022

carcase records (carcase weight, P8 rump fat, rib fat, eye muscle area, retail beef yield or intramuscular fat) for the carcase reference population.

Numerator relationship matrices were constructed with unpublished AGBU nrmblock software written by Ferdosi based on algorithms by Aguilar *et al.* (2011) and Sargolzaei *et al.* (2005). For each animal, the average relatedness with all animals and then reference only animals for each trait was calculated based on the off-diagonal elements. To assess how trait specific reference populations represented the wider population, the average relatedness to reference animals (y-axis) was plotted against the average relatedness to all animals (x-axis) (Figure 1). This visual metric was quantified by the regression slope for all animals and for individual herds. If all animals were phenotyped and genotyped the figure (and slope) is a 1:1 line with no deviations. A regression slope close to 1 was considered optimum, <1 under and >1 over-represented in the reference population.

To explore the relationship between reference size and selection accuracy, single trait BLUP approximate accuracies from the BREEDPLAN genetic evaluations were used. The average live weight (200, 400 or 600 day) accuracy for genotyped but un-phenotyped animals was plotted against the reference size for each breed. To control against influencing factors (i.e. different effective population sizes) and considering only the larger reference populations, a series of phenotype knockouts were undertaken for breed B and D 400 day yearling weights. Phenotypes of whole contemporary groups were randomly removed from the analysis to achieve reference sizes between 1000 and 20,000 animals. The accuracy of genotyped but un-phenotyped animals was plotted against the reference size.

Results

Each beef population was different sized with varying levels of phenotyping and genotyping (Table 1). The percentage of genotyped animals were 4, 6, 8, 12, 16, 50 and 80% respectively for Breeds C, F, D, B, A, G and E. Reference sizes were greatest for live weight (3.5 to 68.3% of all animals) and smallest for abattoir carcase (0.0 to 3.7% of all animals) traits where recording tended to be limited to research projects.

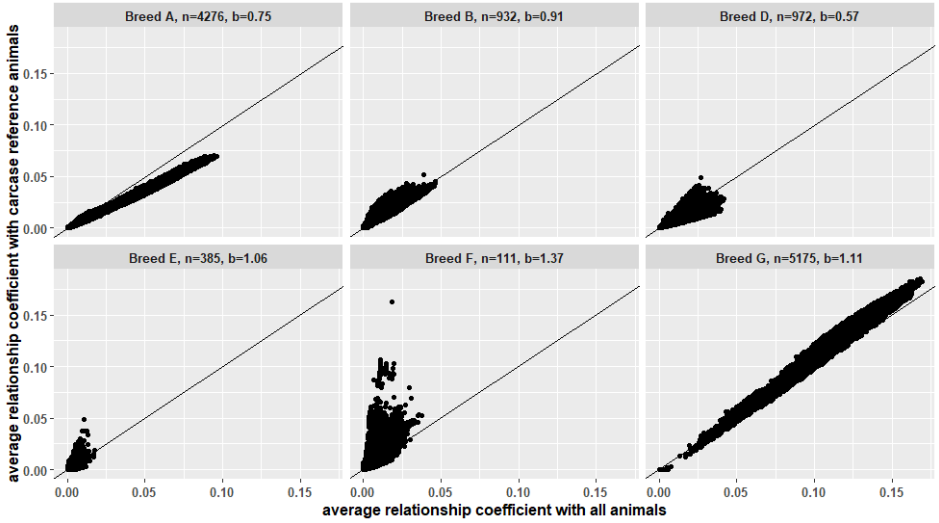


Figure 1. The average relatedness to all animals (x axis) compared with animals in the carcase reference population (y axis) for six Australian beef populations.

Table 1. Description of population and reference size and regression slope from average relationship plots for live weight (LWT) and abattoir carcase (CAR) traits.

	Population size				Regression slope			
	N animals	Reference size (herds)		N GT[1]	LWT		CAR	
Breed	2010+	LWT	CAR	herds	All[2]	Herds[3]	All[2]	Herds[3]
A	1,043,859	136,637 (662)	4,276 (17)	891	1.25	0.71	0.75	0.60
B	165,714	12,896 (73)	932 (5)	273	0.78	0.74	0.91	0.82
C	106,222	3,791 (125)	0 (0)	236	1.29	0.76	-	-
D	410,167	31,379 (390)	972 (12)	483	1.16	0.03	0.57	0.03
E	10,435	7,127 (7)	385 (7)	6	1.25	1.20	1.06	0.79
F	80,085	4,619 (24)	111 (2)	38	1.24	0.90	1.37	0.77
G	175,423	26,923 (133)	5,175 (40)	347	1.20	1.14	1.11	1.03

[1] GT=genotyped.

[2] All = slope for all animals.

[3] Herds= Minimum slope for an individual genotyped herd.

Slopes were calculated for all animals and individual herds, Table 1 reports the overall slope and the minimum herd slope. Larger reference groups were more representative of the general population. For breed G, the overall slope (1.11) and minimum herd slope (1.03) indicated a relevant reference population for genotyping herds. However, the minimum herd slope for un-genotyped herds (results not shown) were as low as 0.02. Breed D had the lowest slope both on a breed (0.57) and herd level (0.03) and along with Breed C showed the most variation in herd slopes (results not shown).

The size and structure of the reference population both impacted accuracy of genotyped un-phenotyped animals. Figure 2 shows that as reference size increased, EBV accuracy also increased. The trajectory was similar for all breeds, as well as when knock-out results were considered for Breeds B and D. Breed G accuracy was higher than expected given reference size, but Figure 1 showed the breed to have higher average relationship values. Breeds B, E and F all had similar accuracy yet reference sizes ranged from 4,619 to 12,896.

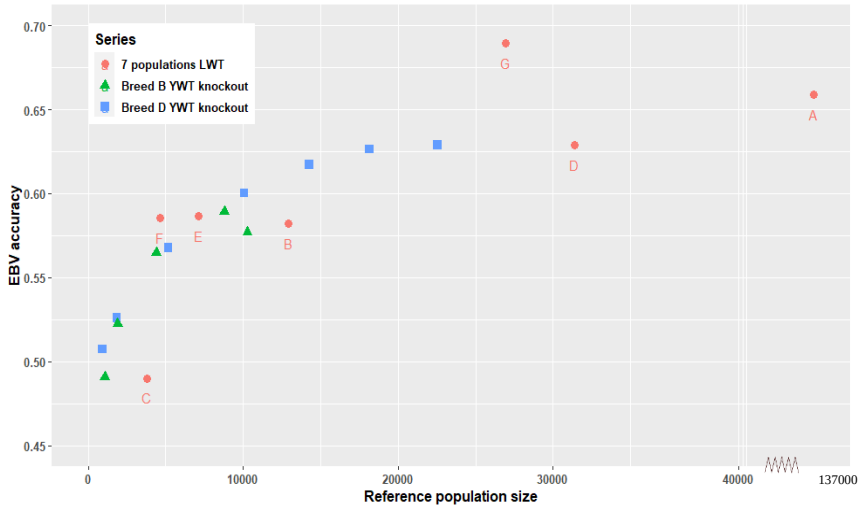


Figure 2. The relationship between reference size and average accuracy of genotyped only animals for live weight (LWT) from seven populations and yearling live weight (YWT) for Breed B and D knockout evaluations.

Discussion

The regression slope metric describes how well reference populations covered the general populat-ion. Larger refer-ence populations generally represent-ed the general population better. This may be more of a reflection on the reference size relative to the wider population size.

Breed B was much closer to 1. Breed D has a much larger population with more herds genotyping but overall fewer animals genotyped. This was also evident with Breed A, who had lower slopes despite having one of the numerically largest carcase reference populations, but when expressed as a proportion of the wider population, the carcase reference was one of the smallest. Results for Breed A show the population to be generally consistent with no herds identified with low slopes. The shape of the average relatedness figure was also informative with the ideal shape being a narrow cloud following the 1:1 line. Breed F had a large number of dots above the line indicating animals already heavily represented in the reference. This plot shape is more likely with a very small reference like we see with Breed F. Herds that genotype benefit from genomic selection due to their linkage to the reference population. The regression slope metric can identify herds with low slopes so that their genetics can either be included into reference data projects or they can use reference animals to build genetic links with their herd and ultimately increase the overall accuracy of genomic selection.

Breed G accuracy was higher compared to other breeds with similar sized reference populations, but Figure 1 shows the breed to have higher average relatedness values. Breeds B, E and F all had similar accuracy despite different reference numbers and proportions. A generally consistent relationship between size and EBV accuracy was observed. Compared to theoretical values (for similar heritability) reported by Goddard and Hayes (2009), the accuracy values we observed from industry datasets were higher when the reference was small (≤2,000) but lower with larger reference populations (≥10,000). For yearling weight, after 10,000 reference animals the observed accuracy tended to plateau rather than continuing to increase as theoretical equations have predicted. Further research is required to understand these differences.

These metrics are an effective and easy way of describing reference populations in the context of the general population and at individual herd level. They provide important information to assist in enhancing the design of the reference population. Breeds can utilise these metrics to identify animals/herds that are not connected to the reference and respond appropriately. These metrics do not consider data quality, so that would need to be considered separately. This simple quantitative metric can help maximise the benefits of the reference population and may be the first objective measurement available to breed societies for managing their reference populations. The next stage of the research will be to investigate the relationship between the relatedness to reference metric with EBV accuracy.

Acknowledgements

This work was supported by Meat and Livestock Australia (L.GEN.1904) and thanks to Angus Australia, Australian Brahman Breeders Association, Australian Wagyu Association, Charolais Society of Australia, Herefords Australia Ltd, North Australian Pastoral Company and Santa Gertrudis Breeders Association and their members for access to their data.

References

Aguilar I., Misztal I., Legarra A. and Tsuruta S. (2011) J. Anim. Breed. Genet. 128(6):422–428. https://doi.org/10.1111/j.1439-0388.2010.00912.x

Goddard M.E. and Hayes B.J. (2009) Nat. Rev. Genet. 10(6):381–391. https://doi.org/10.1038/nrg2575

Johnston D.J., Ferdosi M.H., Connors N.K., Boerner V., Cook J., *et al.* (2018) Proc. of the 11th WCGALP, Auckland, New Zealand

Lee S.H., Clark S. and Van Der Werf J.H.J. (2017) PLoS ONE 12(12):1–22. https://doi.org/10.1371/journal.pone.0189775

Pszczola M., Strabel T., Mulder, H.A. and Calus, M.P.L. (2012) J. Dairy Sci. 95(1):389–400. https://doi.org/10.3168/jds.2011-4338

Sargolzaei M., Iwaisaki H. and Colleau J. J. (2005) J. Anim. Breed. Genet. 122(5):325–331. https://doi.org/10.1111/j.1439-0388.2005.00538.x

291. Decreasing computing cost of categorical data analysis

J. Hidalgo[1*], I. Misztal[1], S. Tsuruta[1], M. Bermann[1], A. Garcia[2], K. Retallick[2] and D. Lourenco[1]

[1]*Department of Animal and Dairy Science, University of Georgia, Athens, GA 30602, USA;*
[2]*Angus Genetics Inc., St. Joseph, MO 64506, USA; jh37900@uga.edu*

Abstract

Threshold models used in the analysis of binary traits such as calving ease provide solutions in liability scale that are easily transformed into probabilities. However, these models are prone to convergence and computing time issues because of the nonlinear equations involved. Although linear models provide a high correlation to liability solutions and are alternative to overcome the issues above, the lack of a transformation from linear to probability scale may prevent using linear models. We derived formulas for transforming linear solutions to probabilities using conditional expectations and variances. This transformation was evaluated by correlating solutions in probability scale for calving ease obtained using threshold or linear models. The transformation from linear solutions to probabilities was equivalent to its counterpart. The correlation for all animals, sires, or animals with records was 0.99. Combining linear models with the new transformation provided an optimal ranking of animals and reduced computing time by five-fold.

Introduction

The nonlinear equations involved in the threshold models increase complexity and computing time in contrast to linear models (Meijering and Gianola, 1985; Weller, 1988). In the genomics era, convergence issues and increased computing time have become significant challenges. As an illustration, the evaluation for calving ease at the American Angus Association which uses a bivariate threshold-linear model with calving ease (1.8M phenotypes) and birth weight (9.4M phenotypes) takes 5 h for a classic BLUP run and 52 h with genomic information (1.12M genotyped animals). This represents a 10-fold increase in computing time, even when using BLUP solutions as initial values in the genomic evaluation.

Although threshold models are more theoretically appropriate for analyzing binary and categorical traits, several studies reported limited advantages over linear models (Varona *et al.*, 1999; Ramírez-Valverde *et al.*, 2001). For example, using a sire model to analyze dystocia and calf mortality in dairy cattle, Weller *et al.* (1988) reported correlations between solutions from the threshold and linear models ≥0.96, and ranking changes for sires were small contrasting both models. Furthermore, the robustness of linear models to departure from normality increases with progeny group size (number of records), number of categories, and incidence (Meijering and Gianola, 1985). A neat feature of threshold models is that solutions in the liability scale are easily transformed into probabilities so genomic estimated breeding values (GEBV) of animals represent the probability of expressing the category of primary interest. However, such a transformation is not well defined for solutions from linear models. Therefore, formulas were derived for transforming GEBV from linear into probability scale using conditional expectations and variances for direct and maternal effects. The computing time and the correlation between GEBV in probability scale for calving ease obtained using threshold or linear models were evaluated.

Materials & methods

Data. The American Angus Association (St. Joseph, MO 64506, USA) provided the data including 11M animals in the pedigree, 9.2M birth weight phenotypes (BW), and 1.8M calving ease phenotypes (CE) recorded as a binary trait, where 1 = easy calving and 2 = assisted calving. A total of 964,444 animals were genotyped for 54k SNP, of which 39,739 remained after quality control.

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_291, © J. Hidalgo *et al.* 2022

Model and computations. Analyses were performed using single-step genomic BLUP (Aguilar *et al.*, 2010) with the algorithm for proven and young (Misztal *et al.*, 2014a), including 22k animals in the core. Two bivariate animal models for CE and BW were implemented: a linear-linear animal model and a threshold-linear animal model:

$$\begin{bmatrix}\mathbf{y}_{ce}\\ \mathbf{y}_{bw}\end{bmatrix}=\begin{bmatrix}\mathbf{X}_{ce} & 0\\ 0 & \mathbf{X}_{bw}\end{bmatrix}\begin{bmatrix}\mathbf{b}_{ce}\\ \mathbf{b}_{bw}\end{bmatrix}+\begin{bmatrix}\mathbf{Z}_{ce} & \mathbf{0}\\ \mathbf{0} & \mathbf{Z}_{bw}\end{bmatrix}\begin{bmatrix}\mathbf{u}_{ce}\\ \mathbf{u}_{bw}\end{bmatrix}+\begin{bmatrix}\mathbf{T}_{ce} & \mathbf{0}\\ \mathbf{0} & \mathbf{T}_{bw}\end{bmatrix}\begin{bmatrix}\mathbf{m}_{ce}\\ \mathbf{m}_{bw}\end{bmatrix}+\begin{bmatrix}\mathbf{W}_{ce} & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{bmatrix}\begin{bmatrix}\mathbf{cg}_{ce}\\ \mathbf{0}\end{bmatrix}+\begin{bmatrix}\mathbf{e}_{ce}\\ \mathbf{e}_{bw}\end{bmatrix} \quad (1)$$

in which y, b, u, m, cg, and e are vectors of phenotypes, fixed effects [contemporary group (only for BW), sex, age of dam, and sex × age of dam interaction; the latter two effects only for CE], additive direct genetic effects, additive maternal genetic effects, random contemporary group effects (only for CE), and random residuals; X, Z, T, and W are incidence matrices for b, u, m, and cg, respectively.

In the threshold-linear animal model, it was assumed that the CE phenotype for individual i (y_{ce_i}; i = 1, 2, …, n) is the expression of an underlying continuous random variable, the liability (L_{ce_i}) of individual i. If L_{ce_i} exceeds an unknown fixed threshold (t), the $y_{ce_i} = 2$ (assisted calving), and $y_{ce_i} = 1$ (easy calving) otherwise. We assumed that the liability was normally distributed with mean vector θ and unit variance:

$$L_{ce_i}|\boldsymbol{\theta} \sim N\left(\mathbf{s}_i'\boldsymbol{\theta},1\right) \quad (2)$$

where θ' = (b', u', m', cg') is a vector of fixed and random effects, and $\mathbf{s}_i'$ is a row incidence vector linking θ to the records.

The response of CE was modelled with the following distribution:

$$p(\mathbf{y}_{ce}|\mathbf{L}_{ce}) = \prod_{i=1}^{n}\left[I\left(L_{ce_i} > t\right)I\left(y_{ce_i} = 2\right) + I\left(L_{ce_i} \le t\right)I\left(y_{ce_i} = 1\right)\right] \quad (3)$$

where I is an indicator function that takes the value of 1 if the condition specified is true, otherwise 0. Variance components were estimated using the average information restricted maximum likelihood algorithm (linear-linear model) or the Gibbs sampling methodology (threshold-linear model). All computations were done using the BLUPF90 suite of programs (Misztal *et al.*, 2014b). For both models, direct breeding values were transformed to probabilities assuming independence between direct and maternal breeding values (independent transformation; Equation 4) or conditioning on maternal breeding values (conditional transformation; Equation 5):

$$P_i = 1 - \Phi\left(\frac{u_i - \mu_u}{\sigma_u}\right) \quad (4)$$

$$P_i = 1 - \Phi\left(\frac{u_i - \mu_{u_i|m_i}}{\sigma_{u|m}}\right) \quad (5)$$

where P_i is the probability of easy calving for animal i, Φ is the standard cumulative distribution function, u_i is the breeding value of animal i; u_u is the mean of breeding values; σ_i is the standard deviation of breeding values; $\mu_{u_i|m_i} = \mu_u + \left[\sigma_{u,m}(\sigma_m^2)^{-1}(m_i - \mu_m)\right]$ is the conditional expectation of the direct breeding value given the maternal breeding value, for animal i; $\sigma_{u|m} = [\sigma_u^2 - (\sigma_m^2)^{-1}(\sigma_{u,m})^2]^{0.5}$ is the conditional standard deviation of the direct breeding values given the maternal breeding values. Spearman rank correlations and the genetic trends were used as indicators of the agreement between probabilities from both models.

Results

Computing time for the threshold-linear and the linear-linear models was 145 and 32 hours, respectively. Spearman rank correlation between direct GEBV obtained with threshold or linear models was 0.988 before

and after the transformation to the probability scale assuming independence between direct and maternal breeding values. The correlation was 0.988 before the transformation, and it was marginally improved (0.991) using the conditional transformation. The correlations were the same for all animals in the pedigree, sires, and animals with CE records. Hence, the linear model provided an optimal ranking of all animals, i.e. the ranking was virtually the same as when using threshold models. Furthermore, the transformation of solutions to probabilities in the linear-linear model was equivalent to that in the threshold-linear model. Figure 1a shows the distribution of direct GEBV in liability and linear scales. Figure 1b shows the distributions in the probability scale. The transformation from the linear scale to the probability scale not only yielded an equivalent ranking of animals, but the breeding values were also in the same range and with nearly identical distributions; therefore, genetic trends obtained by both models were equivalent (Figure 1c), and the agreement was improved using the conditional transformation. These results suggest that using linear models does not lead to significant losses in selection response in this population.

Discussion

Misztal *et al.* (1988) estimated that the cost of a threshold model analysis would be 3 to 5× larger than that of a linear model; their results agreed with the current results–computing time in the threshold-linear model was ~5× larger than computing time in the linear-linear model. However, it is essential to highlight that we used the solutions from a non-genomic, threshold-linear analysis as initial values in the genomic threshold-linear analyses to accelerate convergence. Without the latter strategy, differences would be more prominent.

The transformation of GEBV from liability and linear scales to probability scale relies on the standard cumulative distribution function. If the distribution of solutions has an extreme departure from normality, the probabilistic statements based on linear solutions assuming normality can be distorted. In this study, GEBV obtained in the linear scale was normally distributed, leading to a similar distribution of GEBV in the probability scale from both linear and threshold models. Additionally, the conditional variance reduces the dispersion of breeding values improving the agreement between distributions and genetic trends.

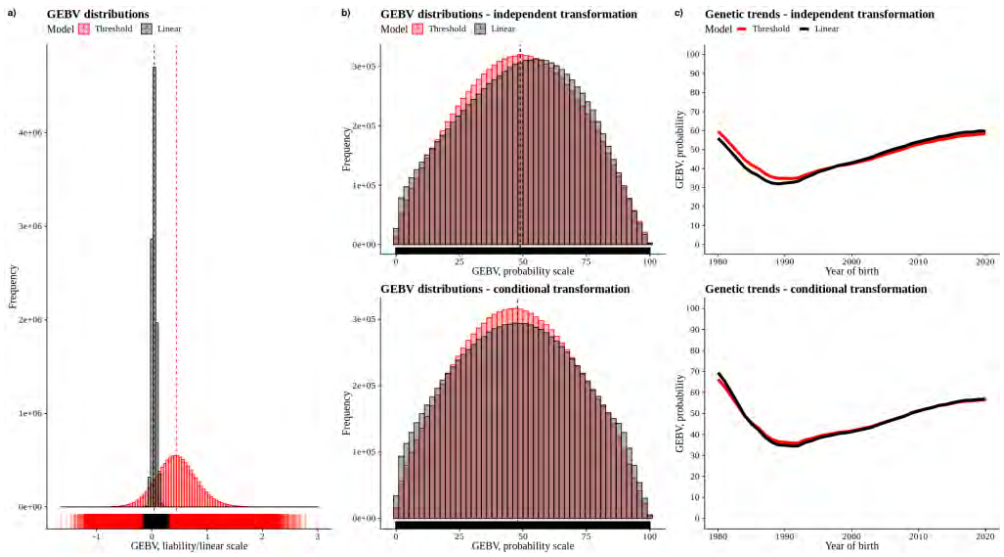


Figure 1. Distributions of direct genomic estimated breeding values (GEBV) and genetic trends in American Angus contrasting linear and threshold models for calving ease.

Robustness of linear models to departure from normality increase with the amount of data (Meijering and Gianola, 1985). In this population, the average progeny number with CE records per sire was 31, and 84% of the heifers had progeny with CE records. In addition, a large amount of BW records (9.2M), which is highly (0.82) correlated with CE, increase the accuracy of GEBV for CE, as also reported by Varona *et al.* (1999) and Ramírez-Valverde *et al.* (2001). These factors lead to a high correlation (0.99) between GEBV for CE obtained from threshold-linear and linear-linear models.

In this study, the correlation between GEBV in probability scale from linear and threshold models was the same despite the number of progeny per sire with CE phenotypes. Such a concordance suggests that large amounts of data in this population provide a similar ranking for all the animals, not only for sires with large progeny group size. The latter contrasts with Meijering and Gianola (1985) and Ramírez-Valverde *et al.* (2001), who, using non-genomic analyses, found higher correlations for sires with larger progeny group size with records for CE. A possible explanation is that genomic information increases the GEBV accuracy for all genotyped animals as every genotyped animal with CE phenotypes influences the direct genomic value of all genotyped animals. In summary, implementing a linear-linear model for the genomic evaluation of calving ease and birth weight, together with the new transformation to the probability scale, represents a feasible alternative to overcome convergence issues and long computing time without harming the genetic progress. The solution presented here may enable large-scale evaluations with multiple categorical or binary traits, which is impractical under the threshold methodology.

References

Aguilar I., Misztal I., Johnson D. L., Legarra A., Tsuruta S., *et al.* (2010) J. Dairy Sci. 93(2):743–752. https://doi.org/10.3168/jds.2009-2730

Lourenco, D. A. L., Tsuruta S., Fragomeni B. O., Masuda Y., Aguilar I., *et al.* (2018) Proc. Of the 11th WCGALP, Auckland, New Zealand. 11.495.

Meijering, A., and Gianola D. (1988) Genet. Sel. Evol. 17(1):115–131. https://dx.doi.org/10.1186%2F1297-9686-17-1-115

Misztal, I., Legarra A., and Aguilar I. (2014a) J. Dairy Sci. 97(6):3943–3952. https://doi.org/10.3168/jds.2013-7752

Misztal, I., Gianola D., and Foulley J. L. (1988) J. Dairy. Sci. 72(6):1557–1568. https://doi.org/10.3168/jds.S0022-0302(89)79267-5

Misztal, I., Tsuruta S., Lourenco D.A.L., Aguilar I., Legarra A., *et al.* (2014b) Manual for BLUPF90 family of programs. Available at http://nce.ads.uga.edu/wiki/lib/exe/fetch.php?media=blupf90_all2.pdf

Ramírez-Valverde, R., Misztal I., and Bertrand J. K. (2001) J. Anim. Sci. 79(2):333–338. https://doi.org/10.2527/2001.792333x

Varona, L., Misztal I., and Bertrand J. K. (1999) J. Anim. Sci. 77(8):2003–2007. https://doi.org/10.2527/1999.7782003x

Weller, J., Misztal I., and Gianola D. (1988) J. Dairy. Sci. 71(9):2491–2501. https://doi.org/10.3168/jds.S0022-0302(88)79836-7

292. Within and across breed single step genomic prediction for somatic cell count including foreign information

R. Mrode[1,2*], M.W. Winters[3] and M. Coffey[1]

[1]Scotland Rural College, Edinburgh EH9 3JG, United Kingdom; [2]International Livestock Institute, Box 30709, Nairobi, Kenya; [3]Agriculture and Horticulture Development Board, Stoneleigh Park, Kenilworth, Warwickshire, CV8 2TL, United Kingdom; raphael.mrode@sruc.ac.uk

Abstract

Current UK genetic evaluations for somatic cell count (SCC) are based on an across breed repeatability animal model. Genomic prediction for SCC was investigated using ssGBLUP within the Holstein (HOL) breed only and on an across breed basis. Also, impact of incorporating de-regressed MACE proofs for foreign bulls on accuracy of predictions was investigated. In the within HOL breed analysis, accuracy of bull predictions increased by about 50% to 0.66 with ssGBLUP and by 70% to 0.76 with including foreign information. Results from the across breed evaluations for HOL were similar to results from the within breed analysis. Increases in cow accuracies were low at about 13% with including foreign information. However, there were no increases in accuracies for the other smaller breeds in the across breed evaluations, due may be to the limited number of their bulls in the validation set with genotypes. Evaluations from the various models were highly correlated.

Introduction

Currently in the UK, conventional genetic evaluations for somatic cell count (SCC) are based on an across breed repeatability animal model using up to 5 lactations with unknown parent groups (UPGs) handled by the QP transformations. The de-regressed proofs from this model are subsequently fed into the multi-step SNPBLUP model that incorporates de-regressed MACE proofs from INTERBULL for genotyped foreign bulls. Genotyped cows have not yet been incorporated into the training population for the multi-step genomic SNPBLUP. However, the simultaneous use of pedigree, phenotypic and genomic information for genomic prediction utilizing a single-step approach (ssGBLUP) would allow genotyped cows to be part of the training population and should, in theory, overcome some of the problems of preselection bias associated with the multi-step SNPBLUP. Therefore, this study examines the use of ssGBLUP for genomic prediction of SCC, initially within the Holstein (HOL) breed and then in an across breed setting involving the Ayrshire (AYR), Jersey (JER) and Guernsey (GUE). In addition, it examines the impact of incorporating de-regressed MACE for foreign bulls on the accuracy of genomic prediction.

Materials & methods

The SCC data for the official UK conventional genetic evaluations and MACE from INTERBULL for the August 2021 run and corresponding genotypes for the genomic prediction, were used for these analyses. The summary of the data showing number of cows, records, pedigree and genotypes are in Table 1. SCC was analysed as $\log_e$SCC and the conventional model (AM1) fitted herd-year-season, lactation, age nested within parity, month of calving as fixed effects and random effects of herd-sire interaction, permanent environment and animal. A pedigree of five generations was used in the analysis with a total of 209 UPGs defined by year of birth, breed, country of origin and the four paths of dam of cows and bulls, and sire of bulls and cows. In the ssGBLUP, 79,051 SNPs used were derived by USDA (Al-Khudhair, *et al.*, 2021). Initially, only HOL was analysed (ssGBLUP-HOL1) with the G matrix computed from genotypes of 9,214 and 45,547 UK bulls and cows, with UPG fitted for both A and the pedigree relationship matrix for genotyped animals (A_{22}) (Cesarani, *et al.*, 2021). Thus, the inverse of the H matrix was computed as:

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_292, © R. Mrode *et al.* 2022

$$\mathbf{H}^{-1} = \mathbf{A} + \begin{bmatrix} 0 & 0 & 0 \\ 0 & \mathbf{G}^{-1} - \mathbf{A}_{22}^{-1} & -\mathbf{A}_{22}^{-1}Q_2 \\ 0 & -\mathbf{Q}_2'\mathbf{A}_{22}^{-1} & \mathbf{Q}_2'\mathbf{A}_{22}^{-1}\mathbf{Q}_2 \end{bmatrix}$$

Foreign information was included for a total of 31,633 HOL bulls with genotypes and MACE proofs, using two slightly different approaches. For bulls with UK daughters included in their MACE, an adjustment was made to avoid double counting of this domestic information. The procedure described by Guarini *et al.* (2019) was implemented as: $PTA_{adj} = \frac{[(DE_M+k)\times PTA_M]-[(DE_D+k)\times PTA_D]}{(DE_M-DE_D)+k}$ and $Rel_{adj} = \frac{DE_M - DE_D}{(DE_M - DE_D)+k}$; where PTA_M and PTA_D are PTAs from MACE and domestic evaluations respectively;. DE_M and DE_D are corresponding daughter equivalents calculated from their respective reliabilities for PTA_M and PTA_D; and k is a variance ratio computed as $(4\text{-}h^2)/h^2$. The de-regressed PTA_{adj} multiplied by two were used as the response variable in the analysis for these bulls with a weighting of corresponding daughter equivalents (DE) computed from Rel_{adj}. However, for foreign bulls with no domestic information included in MACE, their de-regressed PTAs using a full pedigree and multiplied by two were used in the analysis, weighted by their corresponding DEs computed from MACE reliability. To handle the increased number of genotypes by including foreign information, the inverse of G inverse was computed using the APY approach, with 35,000 random bulls with genotypes used as core animals. The model with foreign information included was termed as ssGBLUP-HOL2.

The ability to predict the future performance of young candidates was used for validation of accuracy of genomic prediction. The validation cows comprised of 282,728 cows with 357,591 SCC records born in 2017 and onwards and their records were set to missing in the validation run. Yield deviations (YD) were computed for validation cows by adjusting their records for all fixed effects in the model. Validation bulls were sires with all their UK daughters born within the validation years for validation cows. The YD of validation cows were further adjusted for half the breeding value of their dams and averaged by sires, to compute estimated daughter yield deviations (DYD) for the validation bulls. The accuracy of predictions was computed as the correlation between (G)EBV and YD or DYD for validation cows or bulls. Regression coefficients of YD or YD on (G)EBV were used to measure the inflation of predictions in validation cows and bulls.

Across breed predictions. The conventional repeatability animal model used for HOL was implemented including other breeds; AYR, JER and GUE and various crossbreeds (AM2). This involved 69,737,471 cows with 17,572,710 SCC records with a pedigree consisting of 9,063,330 animals (Table 1). The same UPG structure was utilized since animals from some of the other breeds were dams of cows in the Holstein evaluation. Similarly, ssGBLUP was implemented for the across breed model (ssGBLUP-ALL1) resulting in the inclusion of an additional 1,583 bulls and 6,402 cows with genotypes from the AYR, JER and GUE in the construction of G_{22}. Moreover, foreign information was included for an additional 10,924 foreign bulls for these breeds with genotypes and MACE proofs as described for the HOL only evaluations. The ssGBLUP with foreign information included was termed ssGBLUP-ALL2. The validation bulls and cows were defined as for the within HOL analysis and comprised of 311,633 cows with 393,492 SCC records born in 2017 and onwards and their records were set to missing in the validation run. Similarly, accuracies and measures of inflation of predictions were computed as described for the within HOL analysis. The heritability of 0.11 was used in all analyses within the HOL breed and in the across breed evaluations. All analyses were carried out using the MiX99 software (Strandén *et al.*, 2017).

Table 1. Summary statistics of number of animals, records and genotypes.

	Evaluations	
	Holstein breed only	Across breed
Cows	6,453,896	6,973,747
Records	16,245,062	17,572,710
Pedigree size	8,420,844	9,063,330
UK bull genotypes	9,214	10,797
Cow genotypes	45,547	51,949
Foreign bull genotypes	31,633	38,130

Results

Accuracies and measures of inflation of predictions. The estimates of correlations between (G) EBVs and DYDs for bulls or YDs for cows in the validation set are shown in Table 2 for HOL from the various models. In the within HOL run, the accuracy of genomic prediction for bulls increased from 0.44 with the AM1 by about 50% to 0.66 with ssGBLUP. However, the inclusion of foreign information increased accuracies by 72% relative to the AM1 to 0.76. The pattern of results for HOL from the across breed evaluations was similar to results from the within breed analysis. In terms of measures of inflation of predictions, the regression coefficients (Table 2) were close to unity apart from the ssGBLUP in the within HOL analysis where there was some under-prediction. The accuracies of genomic prediction were much lower for cows at 0.23 from the AM1 and increased slightly to 0.26 with ssGBLUP and with foreign information included. In general, the regression coefficients were close to unity for cows from all models.

The number of bulls with 50 UK daughters in the validation set for the other breeds were too few to obtain any reliable results. Using only bulls with at least 10 UK daughters resulted in a total of 149 bulls across the three breeds, AYR, JER and GUE in the validation set. The accuracy of predictions using these 149 bulls was 0.43 from AM2 but decreased to 0.42 from ssGBLUP-ALL1 with no foreign information included and further to 0.30 when foreign information was incorporated. The regression coefficients were 1.02 for both AM2, ssGBLUP-ALL1 and lower at 0.84 for ssGBLUP-ALL2. In terms of the accuracies of predictions for cows, the accuracies were in generally similar to those obtained to HOL from the AM2 but ssGBLUP with or without inclusion of foreign information did not result in any increases in accuracies.

Table 2. Estimates of correlations between (G)EBVs and DYDs of bulls and YDs of cows in validations data set and corresponding estimates of regression coefficients of DYDs or YDs on (G)EBVs.[1]

Model	Bulls		Cows	
	Correlation	Regression	Correlation	Regression
AM1	0.44	1.00	0.23	0.95
ssGBLUP-HOL1	0.66	1.20	0.24	1.01
ssGLUP-HOL2	0.76	0.98	0.26	0.99
AM2	0.43	0.98	0.23	0.96
ssGBLUP-ALL1	0.66	1.23	0.24	1.06
ssGBLUP-ALL2	0.77	1.00	0.26	1.00

[1] AM1= Within HOL repeatability animal model; AM2= Across breed repeatability animal model; ssGBLUP-HOL1 and ssGBLUP-HOL2= within HOL ssGBLUP without (1) and with (2) foreign information; ssGBLUP-ALL1 and ssGBLUP-ALL2= Across breed ssGBLUP without (1) and with (2) foreign information.

Genetic trends and correlations among various evaluations. The correlation between (G)EBVs for bulls from the AM1 and ssGBLUP-HOL1 was 0.99, 0.98 between AM1 and ssGBLUP-HOL2 and 0.99 between ssGBLUP-HOL1 and ssGBLUP-HOL2. Corresponding estimates for cow evaluations were all 0.99. The estimates of correlation for bulls and cows for the other breeds in the across breed evaluations followed the same pattern with correlations generally between 0.97 to 0.99. Estimates of genetic trends for HOL from all the models within and across breed evaluations were very similar.

Discussion

Increased accuracy of prediction has been one of the major advantages of genomic prediction and an increase of about 50% in accuracy of predictions for the HOL have been observed with ssGBLUP from the within and across breed evaluations. Cesarani *et al.* (2021) reported similar ranges of increases in reliabilities of bulls implementing ssGBLUP for milk, fat and protein yields in the USA. However, Guarini *et al.* (2019) reported lower increases of about 10 to 19% for milking temperament and speed. However, the largest increase in accuracy for HOL of about 70% resulted in the inclusion of foreign information. Guarini *et al.* (2019) reported similar increases in accuracy of about 90% with inclusion of foreign information for milking temperament and speed. However, the levels and increases in accuracies for cow predictions were low and similar to those reported by Cesarani, *et al.* (2021) for productive traits in USA cows. Similar to the results of Guarini, *et al.* (2019), the inclusion of foreign information generally improved the bias of genomic predictions with regression coefficients being closer to unity. However, there were no improvements in accuracy for both bulls and cows for other breeds in the across breed evaluations with ssGBLUP or inclusion of foreign information. The limited number of bulls with genotypes in the validation set may account for this and this will further be investigated.

However, correlations and estimates of genetic trends showed that evaluations from the various models were consistent.

References

Al-Khudhair A., VanRaden P.M., Null D.J., and Li B (2021) J. Dairy Sci. 104(4):4478–4485. https://doi.org/10.3168/jds.2020-19260

Cesarani A., Masuda Y., Tsuruta S., Nicolazzi E.L., VanRaden P.M., *et al.* (2021) J. Dairy Sci. 104(9):5843–5853 https://doi.org/10.3168/jds.2020-19789

Guarini A.R., Lourenco D.A.L., Brito L.F., Sargolzaei M., Baes C.F., *et al.* (2019) J. Dairy Sci. 102(9):8175–8183. https://doi.org/10.3168/jds.2018-15819

Strandén, I., Matilainen, K., Aamand, G.P. and Mäntysaari, E.A. (2017). J Anim Breed Genet. 134:264-274. https://doi.org/10.1111/jbg.12257

293. Using convolutional neural networks for image-based genomic prediction in mice

B.C. Perez[1*], A. Savchuk[1], P. Duenk[2], M.P.L. Calus[2] and M.C.A.M. Bink[1]

[1]Hendrix Genetics, P.O. Box 114, 5830 AC Boxmeer, the Netherlands; [2]Wageningen University & Research, Animal Breeding and Genomics, P.O. Box 338, 6700 AH Wageningen, the Netherlands; bruno.perez@hendrix-genetics.com

Abstract

Convolutional neural networks (CNN) are well suited image recognition tools for their ability to recognize latent patterns from images. Here, we investigated whether CNN can be used for genomic prediction. We created genomic images from genotype data and used them to predict phenotypes in mice. This approach was compared with traditional GBLUP and gradient boosting machine (GBM) models. For the two traits analysed, CNN was competitive in terms of predictive performance. The resolution of genomic images impacted model performance where, for this dataset, optimum results were obtained with 100×100 pixels. These first results demonstrate the potential of genomic images for genomic prediction using CNNs and merit investigation on adding layers of information to further increase accuracy of prediction.

Introduction

In recent years, machine learning algorithms have appeared as promising tools for genomic prediction. These algorithms do not require prior assumptions about the underlying genetic architecture of traits, being capable of capturing these patterns from training data (Pérez-Enciso and Zingaretti 2019). For these reasons, machine learning algorithms could yield better predictive performance than traditional linear models like GBLUP. Convolutional Neural Networks (CNN; fully connected layers and convolutional/pooling filters) are algorithms that are widely used in image recognition tasks for their ability to identify latent patterns and features from images (Trevisan *et al.* 2020). Visual representations of genotype data as predictors could help to improve performance of CNN models for genomic prediction of complex phenotypes (Galli *et al.* 2021). The objective of this research was to investigate the predictive performance of CNN models using genomic information transformed into images, benchmarked with a parametric linear method and an alternative machine learning method.

This project is part of EuroFAANG (https://eurofaang.eu), a synergy of five Horizon 2020 projects that share the common goal to discover links between genotype to phenotype in farmed animals and meet global Functional Annotation of ANimal Genomes (FAANG) objectives.

Materials & methods

Diversity outbred (do) mouse dataset. The DO mice data comprising 835 animals were obtained from The Jackson Laboratory (Bar Harbor, ME). The animals originated from 6 non-overlapping generations (4, 5, 7, 8, 9 and 11) in which males and females were represented equally. Based on anticipated differences in genetic architecture of traits, we selected circulating cholesterol at 12 weeks (CHOL) and fat percentage at 19 weeks (FATP, Table 1). Prior to the analyses, phenotypes were pre-corrected for fixed effects of generation, sex, cage, and diet. All animals had genotype data available for 50,122 SNP markers.

Genotype data formats. We transformed genotypes from tabular format into a genomic image using the DeepInsight algorithm (Python implementation, available at http://www.alok-ai-lab.com) proposed by Sharma *et al.* (2019). This algorithm applies a similarity measuring technique, here we used t-SNE, to obtain a 'topology graph' based on the similarity between genomic markers. In this resulting graph,

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_293, © B.C. Perez *et al.* 2022

Table 1. Trait abbreviation, number of records, heritability estimate, and anticipated genetic architecture of circulating cholesterol and fat percentage.

Trait	No. records	Heritability	Genetic architecture
CHOL	832	0.29	Evidence of epistasis
FATP	834	0.37	Highly polygenic

each point represents a marker. If two or more markers are strongly related due to linkage disequilibrium patterns, these will be placed at similar coordinates. After that, this graph is transformed into an image, one uniquely per individual, and the genomic marker information, e.g. 0, 1, or 2, is mapped to its corresponding pixel position (Figure 1). At this stage, each pixel in the genomic image may contain information from one or more SNP markers (in case these were collapsed due to high similarity) and colours for each pixel indicate A1A1 (light grey, 0), A1A2 (medium grey, 1) and A2A2 (dark grey, 2) genotypes, or no information (black). We compared three different image resolutions: 50×50 pixels (CNN50), 100×100 pixels (CNN100 and 150×150 pixels (CNN150) (Figure 1).

Genomic best linear unbiased prediction (GBLUP). For its well stablished and wide use in routine genomic prediction, GBLUP was used as a benchmark for comparisons. The implementation of this model was performed using the BGLR package (Perez and de los Campos, 2014).

Gradient boosting machine (GBM). The GBM model is an ensemble learning technique that combines gradient-based optimization and boosting techniques (Friedman, 2001). It applies an iterative process of assembling 'weak learners' and therefore, can be expressed as a linear combination of multiple models. The algorithm is also able to prioritize informative predictors while ignoring uninformative ones. Optimal hyperparameters were defined using a grid search (described in Perez *et al.* 2021) on an inner validation comprised of a random sample of 20% observations from the original training set. Results reported come from the best performing model for each trait analysed. The GBM model was implemented using the h2o.ai R package (Click *et al.* 2016).

Convolutional neural networks (CNN). The input layer for the CNNs consisted of the pixel matrix of the genomic image for an individual. A basic network architecture was proposed (described below) and kept the same for both traits. At trait level, a simple hyperparameter search was performed on an inner validation set (20% of the training set) on one replicate, and the best combination of hyperparameters

Figure 1. Illustration of genomic images for a single individual using resolutions of 50×50 (*left*), 100×100 (*middle*) and 150×150 (*right*) pixels, reshaped to a common format for visual comparison.

(lowest mean squared error) from this one search was used for all other tested resolutions in 10 replicates. For the network architecture chosen, the input data was first passed through three convolutional layers, followed by a maximum pooling layer, a dropout layer, a flattening layer, a dense (i.e. fully connected) layer and finally to the output layer containing one node with the predicted trait value. We used the rectified linear activation unit (ReLU) as the activation function in the hidden layers. To reduce the time and memory requirements, models were trained using a fixed batch size (32) and run for a maximum of 400 epochs. The CNN were implemented in Python 3.6 using Tensorflow 2.0 and Keras libraries (Gulli and Pal, 2017).

Predictive performance. Predictive performance was determined with a forward-prediction validation scheme, in which models were trained using data from older individuals (generations 4,5,7,8 and 9; n=643) and prediction was carried out for younger animals (generation 11; n=192). The predictive performance was evaluated by the correlation between model solutions and pre-corrected phenotypes on validation animals (hereafter termed accuracy). For GBM and CNN, for each trait results were reported for the models that obtained best results in the model-specific grid search analysis previously mentioned.

Results

Predictive accuracies varied from 0.221 (CNN50) to 0.301 (GBM) for CHOL and from 0.285 (CNN50) to 0.352 (GBLUP) for FATP (Figure 2). For the root mean-squared error (RMSE), results varied from 0.78 (GBM) to 0.84 (CNN50) for CHOL and from 0.71 (GBLUP) to 0.76 (CNN50) for FATP. The resolution of the genomic image had an impact on both predictive accuracy and error, with CNN100 showing the best results for both traits. The CNN50 yielded poorest results, while CNN150 performed slightly worse than CNN100.

Discussion

In the present study our current CNN model was slightly outperformed by either GBM or GBLUP for CHOL and FATP, respectively (Figure 2). Nevertheless, our results show the potential of using genomic images for phenotype prediction using CNN models as prediction performance for CNN was better than GBLUP for CHOL and better than GBM for FATP. It must be noted here that for time/computational constraints, the network architecture was kept simple and the hyperparameter tuning for the CNN model was limited to a simple search. A more detailed procedure could help to improve performance of CNN

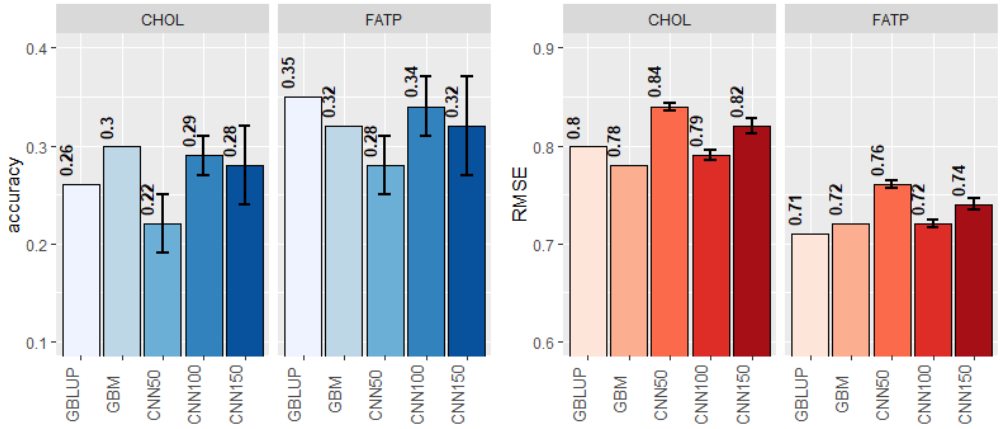


Figure 2. Predictive accuracy & RMSE for GBLUP, GBM and CNN models on genomic images of 50×50 (CNN50), 100×100 (CNN100) and 150×150 (CNN150) pixels.

models. In a case study in maize hybrids, Galli *et al.* (2021) observed superior predictive performance for CNN (using genomic images) over GBLUP (relationship matrix). A common limitation in both studies was the size of datasets (n=904 in Galli *et al.* (2021); n=835 in the present study). Future studies using larger datasets are required to assess performance of genomic prediction using images on routine of breeding programs or other commercial applications.

The resolution of genomic images had major impact on predictive performance. In the present study, the resolution of 100×100 pixels (CNN100) yielded best results and weakening results were found for CNN50 and CNN150. This indicates that there is an optimum image resolution, most likely dependent on the number of SNP genotypes available for prediction and the size and structure of the population studied. For a fixed number of variants, lower resolutions will merge too much information making it harder for the model to capture some patterns in data. Higher resolutions on the other hand can accommodate detailed and complex patterns but will require much bigger training sets to achieve a good predictive performance. These encouraging results and the flexibility of CNN models deserve further investigation to include additional layers of information, such as genome annotations, that can increase accuracy of prediction.

Acknowledgements

The GENE-SWitCH project received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement no. 817998.

References

Click C., Malohlava M., Candel A., Roark H., and Parmar V. (2016) *Gradient Boosted Models with H2O*. Available at: 'http://h2o-release.s3.amazonaws.com/h2o/master/3568/ docs-website/h2o-docs/booklets/GBMBooklet.pdf'

Friedman, J. H., (2001) Ann. Stat. 29:1189-1232.

Galli G., Sabadin F., Yassue R.M., Souza C.G., Carvalho H.F., *et al.* (2021) Research Square. https://doi.org/10.21203/rs.3.rs-840380/v1

Gulli A. and Pal S. (2017). *Deep learning with Keras*. Packt Publishing Ltd.

Perez B.C., Bink M.C.A.M., Churchill G.A., Svenson K.L., and Calus M.P.L. (2021) bioRxiv 2021.08.02.454826. https://doi.org/10.1101/2021.08.02.454826

Pérez P. and de los Campos G. (2014) Genetics. 198:483-495. https://doi.org/10.1534/genetics.114.164442

Pérez-Enciso M. and Zingaretti L.M. (2019) Genes 10:1–19. https://doi.org/10.3390/genes10070553

Sharma A., Vans E., Shigemizu D., Boroevich K.A., and Tsunoda, T. (2019) Sci Rep 9:1–7. https://doi.org/10.1038/s41598-019-47765-6

Trevisan R.G., Pérez O., Schmitz N., Diers B., and Martin N. (2020) Remote Sensing 12:3617. https://doi.org/10.3390/genes10070553

294. Selection for robustness: exploring the value of genomic prediction, reaction norm models and phenotyping strategies

M. Ghaderi-Zefreh[1*], A. Doeschl-Wilson[1], V. Riggio[1,2], O. Matika[1,2] and R. Pong-Wong[1]

[1]The Roslin Institute and R(D)SVS, University of Edinburgh, Easter Bush Campus, EH25 9RG, United Kingdom; [2]Centre for Tropical Livestock Genetics and Health (CTLGH), Roslin Institute, University of Edinburgh, Easter Bush Campus, EH25 9RG, United Kingdom; mghaderi@ed.ac.uk

Abstract

Animals with high production potential and simultaneously high resilience to environmental challenges are vital for sustainable livestock production. A simulation study was carried out to: (1) evaluate the influence of genomic data, statistical models, and phenotyping strategies on prediction accuracies for resilience and production potential; and (2) assess the impact of different selection strategies on genetic improvement in these traits. Prediction accuracies were found to be compromised when families were clustered in similar environments or when the environmental challenge level was unknown. Selection on individuals' performance alone could simultaneously improve resilience if performance records of individuals were obtained across a wide range of environments, whereas a narrow range may decrease resilience. Simultaneous genetic improvement in resilience and production potential strongly benefits from the use of genomic evaluations, reaction norm models, and phenotyping in a wide range of environments.

Introduction

Robust animals are highly resilient (i.e. remain productive in challenging conditions) and have a high production potential (Knap and Doeschl-Wilson, 2020). To genetically improve robustness in animals, accurate estimates of breeding values for production potential and resilience are needed. These can be estimated using reaction norm models based on performance records from different environmental conditions. For linear reaction-norms, the estimated breeding value (EBV) for the intercept represents the production potential, and the inverse of the slope represents resilience (Knap and Su, 2008). One of the impediments in improving robustness using such models is that the environmental challenge level is often unknown and that proxies need to be used instead (Knap and Doeschl-Wilson, 2020).

The objectives of this study were: firstly, to evaluate the influence of genomic data, statistical models, and phenotyping strategies on the accuracy of genetic prediction for resilience and production potential. Secondly, to assess the impact of different selection strategies on genetic improvement in these traits. The data are simulated assuming a sheep population, but the results can be generalised to other species.

Materials & methods

Simulation of data. The phenotypes were simulated assuming a linear reaction norm model where the performance is affected by two components: one related to the performance under ideal rearing condition (i.e. no challenge) and another influenced by the ability of the animal to withstand a challenge. Here, these components will be denoted as the production potential and resilience, and both are assumed to be under genetic and environmental control.

Hence, the performance (y) of individual j in environment k with a challenge level X_k is:

$$y_{j,k} = (m_0 + a_{0j} + e_{0j,k}) + X_k\,(m_R + a_{Rj} + e_{Rj,k}) \quad (1)$$

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_294, © M. Ghaderi-Zefreh *et al.* 2022

where m_0 is the mean performance of the population in the absence of any challenge and, m_R is the average reduction of performance due to the challenge level X_k; a_{0j} and a_{Rj} are the breeding values of individual j for production potential and resilience; $e_{0j,k}$ and $e_{Rj,k}$ are the respective random environmental deviations for each trait. Both traits were assumed with a heritability of 0.1, phenotypic variance of 1.0, zero environmental correlation and an antagonistic genetic correlation of -0.5. Values for m_0 and m_R were arbitrary chosen to be 10 and -3, respectively and X values ranged between 0 and 2 (i.e. average performance in best and worst environment differ by six phenotypic deviation).

The true breeding values were sampled by simulating the genome of each individual in the population (see info about population structure below). To assume a sheep species, the genome was distributed on 26 chromosomes, each with 1,500 segregating biallelic loci in linkage disequilibrium (500 QTL affecting both traits, and 1000 in the SNP panel used to calculate the genomic relationship matrix). Linkage disequilibrium was simulated by sampling the base animals using a mutation-drift algorithm (Meuwissen *et al.*, 2001). The offspring' genome was simulated by gene dropping haplotypes following Mendelian inheritance rules, given their parents' genotypes. The environmental deviations (e_0 and e_R) for each individual and environment k were sampled from a normal distribution.

The level of the environmental challenge that an individual was exposed to (X) was simulated by assuming ten farms, each farm being a subset (10%) of the entire range of challenge levels. Once individuals were allocated to a farm, their specific challenge level were sampled within the range of that farm (i.e. all individuals in the same farm have different challenge level), and their phenotype were calculated from Equation 1.

Genetic evaluation. Best linear unbiased predictor (BLUP) or genomic BLUP (GBLUP) (Strandén and Garrick, 2009) assuming the random regression model as in Equation 1 was used to obtain genomic estimated breeding values (GEBV) for both production potential and resilience. If the level of the environmental challenge was assumed to be unknown (i.e. not measured), a reaction norm approach (Knap and Su, 2008) was implemented where the unknown challenge levels are first estimated using the average farm effects. The farm effect is estimated with the same data using a simpler model ignoring the resilience component. The farm effect estimates are rescaled, and all individuals with performance within a farm are assumed to have the same challenge level, equal to the scaled value of the effect of that farm.

Scenarios compared. To assess the factors affecting the GEBV accuracy, a population of half sib families was assumed, where 100 grandsires were mated to 1,100 granddams to produce 100 sires and 1000 dams, which in turn, were mated to produce 1000 offspring. Only the offspring generation were assumed to have performance records. We considered three different scenarios how families were distributed across environmental conditions: random, clustered, and assortative. In the random scenario, offspring of a sire are randomly allocated to different farms. In the clustered scenario, all offspring of a sire are allocated to the same farm. In the assortative scenario, a correlation between genetic merit and environment challenge was assumed, so that offspring of sires with the highest EBV were allocated to farms with the lowest challenge levels. The effect of knowing or not the challenge level was also assessed.

To assess the response to selection for resilience, we assumed a population of 8,100 individuals (2,700 males and 5,400 females) per generation, where the best 270 males and 2,700 females are selected to produce the next generation. Individuals were randomly distributed across the entire range of environments. The genetic evaluation was done via GBLUP implementing the reaction norm model (where the estimated farm effects were used as challenge levels, see genetic evaluation section above). All candidates were genotyped and had one phenotypic record. Thereafter, a selection index was defined as $I = (1 - \alpha) \times \hat{a}_0 + \alpha \times \hat{a}_R$, where

α ($0 \leq \alpha \leq 1$) is the weight given to resilience. The selection scheme was done over 10 discrete generations. We also considered scenarios where performance records were evaluated in a narrow range of environment (i.e. X values between 0 and 1). The results were compared with the case where candidates were evaluated using a conventional genetic model that does not account for the term influenced by the level of environmental challenge.

Results

Impact of genomic information, unknown environmental challenge level and distribution of families across environments on prediction accuracies. As expected, genetic predictions for production potential and resilience were significantly more accurate with GBLUP than with BLUP (70-163 and 79-114%, respectively). Figure 1 shows the effect of unknown challenge level and the distribution of families on the prediction accuracies for production potential and resilience. Both traits have reduced accuracy when the environmental challenge level is unknown. In general, the accuracy of GEBVs for production potential are higher than that of GEBVs for resilience but also more sensitive to the distribution of individuals across environmental conditions. When the individuals were allocated at random across the whole range of available environmental conditions, the accuracy of GEBVs was highest (between 6-75% higher for production potential and 4-17% for resilience).

Response to selection. Figure 2 shows the genetic gain for production potential and resilience after ten generations of selection using different strategies and when the data are obtained in a wide or a narrow range of environmental challenge levels. When the selection index has all the weight on one component only (i.e. α=0 or α=1), a genetic loss for the other component was observed due to the assumed negative genetic correlation. When the information comes from a wide range of environmental conditions, selection based on the conventional model (which ignores the resilience component) yielded genetic gain for both traits, despite their negative genetic correlation (solid circles in Figure 2). However, when the range of environmental conditions were narrowed down, selection ignoring resilience yielded genetic gain for production potential, but resilience deteriorated. Using a selection index based on the GEBVs of the reaction norm analysis, both traits can be improved regardless of the range of environmental conditions where the animals were measured.

The trend of the selection response when varying α depended on the range of environments where the animals were phenotyped. Phenotypes obtained from a narrower range of environmental conditions generally led to higher gain for production potential (except for large index weight on resilience, α>0.84), but reduced genetic gain for resilience. Genetic gain in resilience could only be obtained when the index weight α was sufficiently large (α>0.12 and α>0.42 corresponding to the wide and narrow environmental range, respectively).

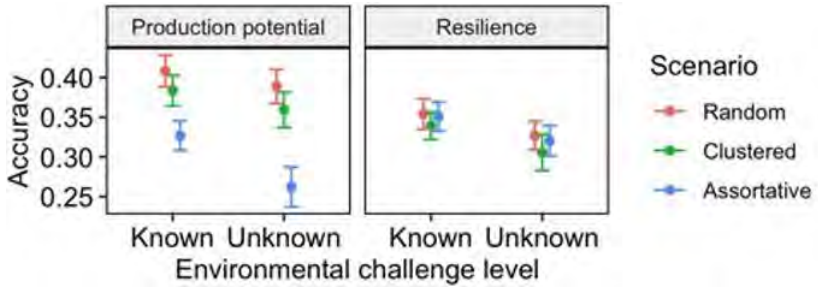


Figure 1. Effect of unknown environmental challenge level and distribution of families across environments on the prediction accuracy of production potential and resilience.

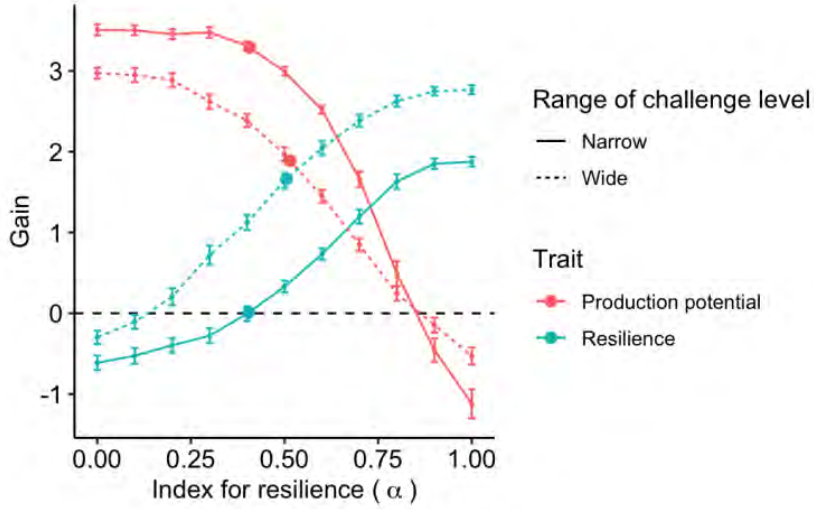


Figure 2. Genetic gain for production potential and resilience after ten generations of selection with different index weight α for resilience (error-bars indicate 95% C.I.). The results for the conventional model, which ignores the resilience component, are shown with solid circles.

Discussion and conclusion

The study showed that genomic prediction is beneficial to improve genetic prediction for production potential and resilience. The GEBV accuracies for both traits are strongly affected by the distribution of phenotyped individuals across environments and knowledge of the challenge level. Whilst the prediction accuracies for resilience was generally lower than for production potential, they were less sensitive to the distribution of phenotypes. Our results on the response to selection show the potential benefits for improving resilience using a random regression/reaction norm model in the evaluation. Conventional genetic evaluations ignoring resilience can improve both resilience and production potential if candidates are evaluated under a wide range of environmental conditions, even when their genetic correlation is negative. However, this is not the case when the performance data come from a narrow range of environmental conditions. This problem can be avoided with the reaction norm model. Hence, reaction norm model and selection index with appropriate weights have great potential for simultaneous improvement of both production potential and resilience.

Acknowledgements

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 772787 (SMARTER) and from the Roslin Institute Strategic Programme grant (ISP3.1, BBS/E/D/30002275).

References

Knap P.W., and Doeschl-Wilson A. (2020) Genet. Sel. Evol. 52:60. https://doi.org/10.1186/s12711-020-00580-4

Knap P.W., and Su G. (2008) Animal 2(12):1742-1747. https://doi.org/10.1017/S1751731108003145

Meuwissen T.H.E., Hayes B.J., and Goddard M.E. (2001) Genetics. 157(4):1819-1829. https://doi.org/10.1534/genetics.116.189795

Strandén I., and Garrick D.J. (2009) J. Dairy Sci. 92(6):2971-2975. https://doi.org/10.3168/jds.2008-1929

Part 18

Methods and tools: genomic prediction

295. Benchmark SNPs selection strategies to achieve optimum prediction in small population

E. Mancin*, C. Sartori, B. Tuliozi, G.G. Proto, R. Mantovani

University of Padova, Department of Agronomy, Food, Natural resources, Animals and Environment – DAFNAE, Viale dell'Universita' 16, 35020 Legnaro (PD), Italy; enrico.mancin@studenti.unipd.it

Abstract

In recent decades, genomic selection has become a routine practice for genetic evaluations of most livestock species and breeds. However, its use is still limited in small local populations. The main reasons are the lower interest to invest in these breeds, and the poor accuracy of their genomic predictions due to the low number of animals. Bayesian variable selection regression models (i.e. Bayes B/C) appear to improve the accuracy of genomic prediction from small populations, but their ability to account for selection remains unclear, especially when using time-series cross validations. In the present work, we evaluated in terms of accuracy and bias the impact of constructing the genomic matrix using only a subset of informative SNPs. Analysis was performed on Rendena cattle, a small local cattle population of North Italy.

Introduction

One of the major challenges of genomic best linear unbiased predictor (GBLUP) is the breeding values estimation in small populations, such as those of local breeds. Indeed, different methods have been exploited to feature single nucleotide polymorphisms (SNPs) selection or assume heterogeneous SNP variance (Gianola *et al.*, 2009; Habier *et al.,* 2013). However, properties linked to the possible bias of SNPs selection were never considered.

For this reason, many studies have 'relaxed' the assumptions of the G matrix by trying to assign different weights to each SNP. However, there are two main drawbacks in this approach: (1) although these methods have greater accuracy, they result in greater dispersion and bias; (2) in a single step GBLUP (ssGBLUP) framework, an excessive SNPs shrinkage can lead to an incompatibility between the genomic and pedigree matrices.

In the present study, we tried an alternative strategy which consists of constructing the G matrix by removing the non-informative SNPs. Indeed, Fragomeni *et al.* (2017) demonstrated that accuracy of GBLUP can substantially improve by removing SNPs close to quantitative trait loci (QTLs), especially in small populations. According to Fragomeni *et al.,* this is due to the shrink effect of those SNPs on QTLs. Since position of QTLs are unknown on real populations, different variable selection approaches to select SNPs for the construction of the G matrix. Then prediction of ssGBLUP using these sets of informative SNPs was benchmarked. Analysis was conducted on Rendena breed. Rendena is small local cattle (~4,000 cows) reared in the North-Eastern Italian Alps and in the adjacent plains. The breed shows appreciable levels of fertility and longevity, and the ability to adapt to marginal areas and graze during summer in high Alpine pastures. Rendena is mainly bred for milk improvement and meat (35%).

Materials & methods

Study subject. We used the performance test database of Rendena cattle, and the target trait was *in vivo* dressing percentage. The study considered 1,691 young bulls with only phenotypic information, 1,739 animals with only genotypic information, and 687 animals with both phenotypic and genotypic information. After quality control (MAF<0.05, LD prune, SNP missing rate <0.1) 85,331 SNPs were used for the analysis.

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_295, © E. Mancin *et al.* 2022

Statistical analysis. All genetic predictions were computed using BLUPF90 family program (Aguilar *et al.*, 2018). In the present work we compared 'classical' models (pedigree BLUP, Single step Genomic BLUP and Weighted ssGBLUP) with ssGBLUP using only a set selected SNP. Three categories of SNPs selection were performed: randomly even spaced SNPs, based on allele substitution effect, and using variable selection algorithm

Pedigree BLUP (PBLUP). The equation for PBLUP is written as follows:

$$\begin{bmatrix} X'X & X'Z \\ X'X & Z'Z+A^{-1}\frac{\sigma_e^2}{\sigma_a^2} \end{bmatrix} \begin{bmatrix} \hat{\mu} \\ \hat{a} \end{bmatrix} = \begin{bmatrix} X'y \\ Z'y \end{bmatrix} \quad (1)$$

Where y is the vector of phenotypes, X represents the incidence matrix for systematic effects, b is the vector of fixed effects of contemporary group (142 levels) and the parity order of the mother (four classes). Z is the incidence matrix of random genetic additive effects, *a* represents the vector of these effects (EBVs) and e is the random residual. Normal assumption was assumed for random effect. In the Pedigree BLUP model, covariances of random genetic effect were sampled from a distribution $N(0, A\sigma_a^2)$, where A is the Identical by Descendent (IBD) matrix. Variance components were estimated under PBLUP tracing back all generation in the pedigree.

Single step genomic BLUP (ssGBLUP). The equation for ssGBLUP is equal to (1) but A is substituted by H (Auguilar 2010).

$$H^{-1} = A^{-1} + \begin{bmatrix} 0 & 0 \\ 0 & G^{-1} - A_{22}^{-1} \end{bmatrix} \quad (2)$$

where A^{-1} and A_{22}^{-1} represent the inverses of the IBD matrix for all individuals and only genotyped animals, respectively. G is the genomic relationship matrix (VanRaden, 2008)

$$G = \frac{MM'}{2\sum p_i(1-p_i)} \quad (3)$$

where M is a matrix of SNP content centered by twice the current allele frequencies, and p_i is the allele frequency for the i^{th} SNP.

ssGBLUP with selected SNPs. Then ssGBLUP was performed by constructing G using only a set of SNPs. Different selection methods for selecting SNPs were used, as following:

Selection of top SNPs. In this strategy the best 500, 1000, 30,000 SNPs were based on their allele substitution, calculated as (Gualdrón Duarte *et al.,* 2014):

$$\hat{g} = \frac{1}{2\sum pq} M'G^{-1}\hat{a}_{22} \quad (4)$$

Selection of even spaced SNPs. In this strategy 500, 1000 and 30,000 even spaced SNPs were randomly selected.

Selected by Feature selection algorithm.

1. Lasso Regression (LASSO), further information see Tibshirani, 1996, compute with R package (https://cran.r-project.org/web/packages/glmnet/index.html).
2. Recursive Feature elimination using Support vector machine (RFESVM), hand wrote R code following what was reported in Sanz *et al.,* 2018.

3. Extreme Gradient Boost (XGBOOST), further information see Cheng *et al.* 2016 and compute with h20 R environment (https://cran.r-project.org/web/packages/h2o/index.html).
4. Spike-and-Slab LASSO (SSLASSO); for information see Ročková and George, 2018.

The hyperparameter of those models were estimated via cross validation by using young animals as test population (genotyped animals born after 1-1-2015), and mean squared error was used as the parameter to select the best set of hyper parameters

Weighted ssGBLUP (WssGBLUP). In addition, WssGBLUP was also considered, and nonlinear A SNP weight was also performed. For simplicity, only two WssGBLUP predictions were reported: (1) using the best iteration CT=1.105; and (2) best iteration of CT=1.250; all SNPs were also considered in this model. For further details of this strategy see *Fragomeni et al.* (2019).

Cross validations. LR cross validation has been used (Legarra and Reverter, 2018). Two validation cohorts have been considered (young bulls born after 1-1-2013 and young bulls born after 1-1-2016).

Results

Bar plot of Figure 1 reported accuracy measured and dispersion of prediction models. All genomic models had better accuracy than PBLUP. Slight increase of accuracy was observed for nonlinear models with CT=1.105, with highest increase being CT=1.250. Interestingly, no substantial differences were observed for models that used random selection of SNPs. On the other hand, a substantial increase of accuracy of prediction was observed for the SNPs selected based on alleles substitution effects, especially when number of SNPs was low. However, the main goal was to observe the impact that the feature selection algorithm had on ssGBLUP. On these topics the impact of XGBOOST appeared to be marginal for both accuracy and dispersion; lasso regression presented similar performance, although a slight increment in accuracy was observed. RFESVM and SSLASSO presented better accuracy values.

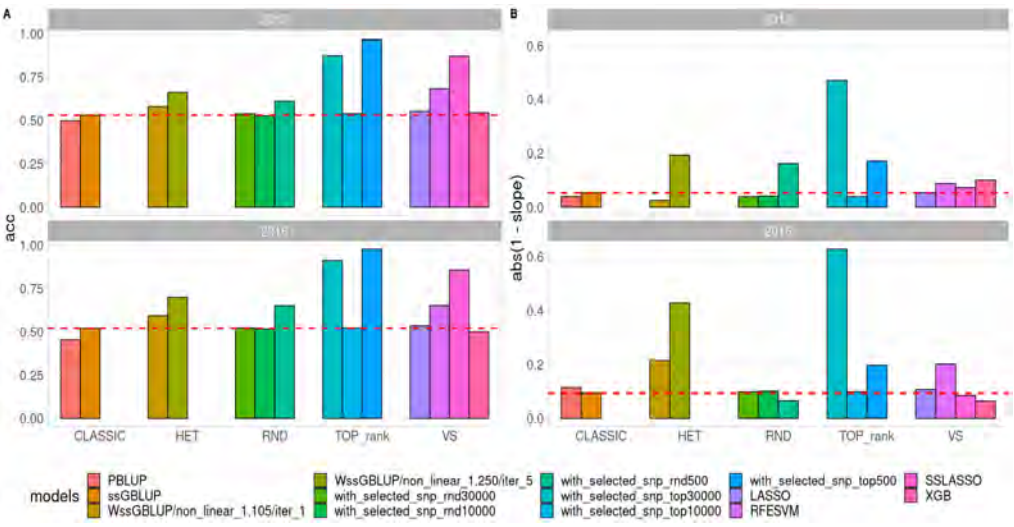


Figure 1. Accuracy (A) and dispersion (B) of prediction using the two-validation datasets from 2013 and 2015.

Discussion

We demonstrated that selection of SNPs based on allele substitution effect (Selection of top SNPs) or WssGBLUP models were not a promising strategy since the increase of accuracy of these models relates to an increase of dispersion. On the other hand, selection of random even spaced SNPs did not affect accuracy and dispersion. However, the variable selection models, especially SSLASSO or LASSO can be considered the best for this trade-off since they increase accuracy of prediction without affecting the dispersion. For that reason, we demonstrated that using ad-hoc algorithms for SNPs selection can be a suitable strategy for small populations. From these studies SSLASSO appears as a better strategy for selecting the most informative SNPs. However, there are still some concerns about: (1) its impact on SNPs selection over the time; (2) its efficacy in multi-traits models.

References

Aguilar, I., *et al.*, (2018). Proc. 10th WCGALP, Vancouver, Canada

Chen, T., and Guestrin, C. (2016). XGBoost: Proc of the 22nd SIGKDD. New York, NY, USA.

Fragomeni, B. O., Lourenco, D. A. L., Legarra, A., VanRaden, P. M., and Misztal, I. (2019). J. Dairy Sci. 102, 10012–10019. https://doi.org/10.3168/jds.2019-16262.

Fragomeni, B. O., Lourenco, D. A. L., Masuda, Y., Legarra, A., and Misztal, I. (2017). Genet. Sel. Evol. 49, 1–11. https://doi.org/10.1186/s12711-017-0335-0.

Gianola, D., de los Campos, G., Hill, W. G., Manfredi, E., and Fernando, R. (2009). Genetics 183, 347–363. https://doi.org/10.1534/genetics.109.103952.

Gualdrón Duarte, J. L., *et al.*, (2014). BMC Bioinform 15, 1–11. https://doi.org/10.1186/1471-2105-15-246.

Habier, D., Fernando, R. L., and Garrick, D. J. (2013). Genetics 194, 597–607. https://doi.org/10.1534/genetics.113.152207.

Legarra, A., and Reverter, A. (2018). Genet. Sel. Evol. 50, 1–18. https://doi.org/10.1186/s12711-0180426-6.

Ročková, V., and George, E. I. (2018). J. Am. Stat. Assoc. 113, 431–444. https://doi.org/10.1080/01621459.2016.1260469.

Sanz, H., Valim, C., Vegas, E., Oller, J. M., and Reverter, F. (2018). BMC Bioinform. 19, 432. https://doi.org/10.1186/s12859-018-2451-4.

Tibshirani, R. (1996). J. R. Stat. Soc. Ser. B 58, 267–288. Available at: http://www.jstor.org/stable/2346178.

VanRaden, P. M. (2008). J. Dairy Sci. 91, 4414–4423. https://doi.org/10.3168/jds.2007-0980.

296. Genetic correlations among longevity, fertility, udder health and type traits including or not genomic information

R. Rostellato*, L.-H. Maugan and V. Ducrocq

Université Paris Saclay, INRAE, AgroParisTech, GABI, 78350 Jouy-en-Josas, France; roberta.rostellato@inrae.fr

Abstract

The aim of the study was to estimate genetic correlations among functional longevity, conception rate of heifers and of lactating cows, somatic cell score, clinical mastitis, udder depth and udder balance in Montbéliarde breed. Bivariate random regression sire models including a pedigree or a genomic relationship matrix were used to compute (co)variance components. Genetic correlations obtained including genomic information were consistent with those estimated using the pedigree relationship matrix. Conception rate, somatic cell score, clinical mastitis and udder depth showed moderate to high genetic correlations with functional longevity, suggesting that they can be good indirect predictors of the length of productive life when included in a combined genomic evaluation. In addition, the large genetic correlations among the two udder health traits and udder depth indicate that they can improve the prediction of the other traits in a multi-trait single-step model.

Introduction

Functional longevity (FL) is defined as the ability of a cow to avoid being culled for other reasons (e.g. illness or infertility) than low milk production (Ducrocq *et al.*, 1988). Dairy cow's longevity can be expressed as the length of productive life, that is the time between the date of the first calving and the culling or death date. Similarly to other functional traits, FL is increasingly considered in breeding programs because of growing concerns about animal welfare. As reported in Rostellato *et al.* (2021), it is affected by factors that are known to reflect resilience, such as fertility problems and diseases, suggesting that FL can be used as an indicator of long-term resilience.

In French dairy cattle, FL is analysed using a Weibull proportional hazard model. The heritability estimated for the trait using such model is small and accuracy of estimated breeding values (EBVs) of young bulls is low given that the majority of their daughters are censored (Ducrocq and Sölkner, 1998). To overcome those drawbacks and to include EBVs for FL in total merit index, a combined polygenic evaluation was developed (Ducrocq *et al.*, 2001). This methodology consists in an approximate multi-trait animal model, where several genetically correlated traits are included and used as predictors of the other traits. In addition to FL, the traits considered are female fertility, udder health and type traits (GenEval, 2021).

Variance and covariance components used in the current combined polygenic evaluation have been estimated ignoring genomic information. Hence, the objective of the present study was to estimate genetic correlations (r_g) among FL, reproductive, udder health and udder type traits using a pedigree (A) relationship matrix or a genomic (G) relationship matrix.

Materials & methods

Data. The data were extracted from the French bovine national database. The dataset included 1,308,066 Montbéliarde cows, born between 2000 and 2018, from 2,345 sires with a minimum of 30 daughters. The investigated traits were FL, conception rate of heifers (CR_h) and of lactating dairy cows (CR_c), somatic cell score (SCS), clinical mastitis (CM), udder depth (UDE), and udder balance (UDB).

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_296, © R. Rostellato *et al.* 2022

Sires of cows were genotyped using the BovineSNP50 BeadChip (Illumina Inc., San Diego, CA). Markers exhibiting call rate <0.9 or minor allele frequency <5% and monomorphic single nucleotide polymorphisms (SNPs) were discarded. After quality control, 37,133 SNPs remained.

The sire's pedigree was traced back for 3 generations and it included 5,946 animals.

Estimation of genetic correlations. The investigated traits are described by very different statistical models (e.g. FL is analysed using a survival model). It makes the use of a multiple trait animal model not feasible to compute r_g among traits. Hence, pre-adjusted performances (i.e. phenotypes corrected for all non-genetic effects) adjusted for the dam's EBV and their associated weight were computed for each cow and each of the investigated traits by means of an accurate model as described in GenEval (2021). Then, these corrected records were used as pseudo-phenotypes in a simple and unique model for all traits, whatever the complexity of the original univariate model of analysis, as proposed by Ducrocq *et al.* (2001).

Variance and covariance components were estimated using bivariate random regression sire models. To take into account heterogeneous residual variance for each pre-adjusted performance due to the different amounts of information (weight), the statistical model was as follows:

$$y^{\#}_{i,m} = \upsilon_{i,m} y^{*}_{i,m} = \upsilon_{i,m} \times by_m + \upsilon_{i,m} \times a_{i,m} + \varepsilon_{i,m} \quad (1)$$

where $\upsilon_{i,m}$ is the square root of weight for the trait i and animal m, $y^{*}_{i,m}$ is the pre-adjusted performance corrected for the dam's EBV of trait i and animal m, by_m is the birth year of cow m, $a_{i,m}$ is the additive genetic value of animal m for trait i, and $\varepsilon_{i,m}$ is a residual. The fixed effect of birth year is to take into account changes in genetic due to selection on correlated traits.

Two distinct analyses were carried out in order to estimate (co)variance components using A matrix or G matrix, respectively.

Variance and covariance components were estimated using the average information REML algorithm implemented in AIREMLF90 program (Misztal *et al.*, 2014).

Results

Genetic correlations among the investigated traits including or not genomic information are presented in Table 1.

Comparison of estimates including or not genomic information. Genetic correlations estimated using the A matrix were consistent to those estimated with G matrix. The inclusion of genomic information gave higher r_g among FL, SCS and CM, as well as between the two fertility traits. Only r_g estimated using G matrix are presented and discussed hereinafter.

Genetic correlations between FL and the other investigated traits. Genetic correlations between FL and the two fertility traits were favourable. Conception rate of heifers exhibited a lower r_g compared to CR_c. High and favourable (i.e. lower culling risk when lower SCS or CM) r_g were found between FL and the two udder health traits. Udder depth and UDB exhibited a moderate and favourable r_g with FL (i.e. lower culling risk when shallow and/or balanced udder).

Genetic correlations among fertility, udder health, and udder type traits. Genetic correlation between CR_h and CR_c was moderately positive. Conception rate of heifers did not show significant genetic relationships with udder health and udder type traits. In contrast, conception rate of lactating cows exhibited moderate and favourable r_g with SCS, CM and UDE (higher conception rate when low SCS or CM or UDE). A very high and positive r_g was found between the two udder health traits. Moreover, SCS and CM were positively correlated with UDE. A negative and moderate r_g was estimated between the two udder type traits.

Discussion

Comparison of estimates including or not genomic information. The increase of the magnitude of r_g for FL, fertility and udder health traits when genomic information was taken into account was related to higher genetic covariances with respect to the estimates obtained using A matrix (data not shown). Similar results were also found by Gordo *et al.* (2016).

Genetic correlations between FL and the other investigated traits. In our study, we found that FL is genetically correlated with fertility, udder health, and udder type traits, suggesting that they are good indirect predictors of the length of productive life.

The favourable genetic relationships between FL and conception rate indicates that selection for a better fertility can prolong the length of productive life of dairy cows. In addition, it confirms the major importance of reproductive problems in culling-decision making (Olechnowicz *et al.*, 2016). Functional longevity exhibited a quite large relationship with SCS and CM. It suggests that the risk of being culled increased when the cow is more prone to develop sub-clinical or clinical mastitis. Hence, selective breeding for a better udder health improves the length of productive life and animal welfare. The r_g between FL and UDE indicates that a deep udder is associated with poor longevity, as reported in Larroque and Ducrocq (2001). The favourable genetic relationship between FL and UDB suggests that cows with good udder balance have a lower risk to be culled with respect to animals with poor UDB. It is in agreement with the study of Rostellato *et al.* (2021).

Table 1. Genetic correlations among the investigated traits using a pedigree relationship matrix (above diagonal) or a genomic relationship matrix (below diagonal).[1,2]

	FL	CR_h	CR_c	SCS	CM	UDE	UDB
FL	-	-0.21 (0.03)	-0.42 (0.03)	0.45 (0.02)	0.47 (0.03)	0.44 (0.02)	-0.29 (0.03)
CR_h	-0.20 (0.04)	-	0.34 (0.04)	-0.08 (0.03)	-0.02 (0.05)	-0.02 (0.03)	-0.1 (0.03)
CR_c	-0.43 (0.03)	0.44 (0.04)	-	-0.23 (0.03)	-0.23 (0.04)	-0.24 (0.03)	0.04 (0.03)
SCS	0.50 (0.02)	-0.08 (0.04)	-0.28 (0.03)	-	0.71 (0.02)	0.33 (0.02)	-0.11 (0.02)
CM	0.53 (0.04)	-0.04 (0.05)	-0.26 (0.05)	0.82 (0.02)	-	0.41 (0.03)	-0.08 (0.02)
UDE	0.42 (0.03)	0.08 (0.04)	-0.26 (0.03)	0.33 (0.02)	0.46 (0.04)	-	-0.39 (0.02)
UDB	-0.29 (0.03)	0.06 (0.04)	-0.05 (0.04)	-0.11 (0.03)	-0.09 (0.04)	-0.36 (0.02)	-

[1] FL : functional longevity (low risk to be culled → high risk to be culled), CR_h : conception rate of heifers (poor conception rate → good conception rate), CR_c : conception rate of lactating dairy cows (poor conception rate → good conception rate), SCS : somatic cell score (low somatic cell score → high somatic cell score), CM : clinical mastitis (few clinical mastitis events → several clinical mastitis events), UDE : udder depth (shallow udder → deep udder), UDB : udder balance (poor udder balance → good udder balance).

[2] Sample standard deviation within parentheses.

Genetic correlations among fertility, udder health, and udder type traits. Conception rate of heifers did not show a significant r_g with the investigated traits, except with CR_c. However, its inclusion in a multi-trait genomic evaluation can improve the prediction of conception rate in multiparous cows. A strong positive genetic relationship between the two udder health traits is reported in many other studies (e.g. Rupp and Boichard, 1999; Govignon-Gion *et al.*, 2016), confirming that SCS can be used as an indicator of CM. In addition, the positive r_g with UDE indicates that deep udders are associated with an increase of the number mastitis events and that this type trait can be used as a predictor of mastitis resistance.

Conclusions

Genetic correlations among functional longevity, conception rate, udder health and udder type traits estimated including genomic information were consistent with those obtained using a pedigree relationship matrix. The analysis presented in this study will be extended to other breeds and the results will be used in the development of a combined single-step evaluation in French dairy cattle.

Acknowledgements

The research leading to these results has received funding from European Union's Horizon 2020 research and innovation programme – GenTORE – under grant agreement no. 727213.

References

Ducrocq V., Quaas L.R., Pollack E.J., and Casella G. (1988) J. Dairy Sci. 71:3061-3070. https://doi.org/10.3168/jds.S0022-0302(88)79906-3

Ducrocq V., and Sölkner J. (1998) Proc of the 6th WCGALP, Armidale, Australia.

Ducrocq V., Boichard D., Barbat A., and Larroque H. (2001) Proc. of the 52th EAAP, Budapest, Hungary.

GenEval, Evaluation génétique des taureaux montbéliards (2021). Available at: https://www.geneval.fr/indexations-races-bovines

Gordo D.G.M., Espigolan R., Tonussi R.L., Júnior G.A.F., Bresolin T., *et al.* (2016) J. Anim. Sci. 94:1821-1826. https://doi.org/10.2527/jas.2015-0134

Govignon-Gion A., Dassonneville R., Baloche G., and Ducrocq V. (2016) Animal. 10(4):558-565. https://doi.org/10.1017/S1751731115002529

Larroque H., and Ducrocq V. (2001). Genet. Sel. Evol. 33:39-59. https://doi.org/10.1186/1297-9686-33-1-39

Misztal I., Tsuruta S., Lourenco D.A.L., Aguilar I., Legarra A., *et al.*, Manual for BLUPF90 family of programs (2014). Available at: http://nce.ads.uga.edu/wiki/lib/exe/fetch.php?media=blupf90_all2.pdf

Olechnowicz J., Kneblewski P., Jaśkowki J.M., and Wlodarek J. (2016) J. Anim. Plant Sci. 26:1533-1541.

Rostellato R., Promp J., Leclerc H., Mattalia S., Friggens N.C., *et al.* (2021) J. Dairy Sci. 104(12):12664-12678. https://doi.org/10.3168/jds.2020-19974

Rupp R., and Boichard D. (1999) J. Dairy Sci. 82:2198-2204. https://doi.org/10.3168/jds.S0022-0302(99)75465-2

297. Alternative SNP weighting for genomic prediction methods in the presence of causative variants

B.F. Santana[1*], M. Riser[1] and B. Fragomeni[1,2]

[1]University of Connecticut, 17 Manter Rd, 06269, Storrs, USA; [2]Institute for System Genomics, 67 North Eagleville Road, 06269-3003 Storrs, CT 06269-3003; USA; bruna.santana@uconn.edu

Abstract

Commercial implementations of genomic selection usually assume that SNP markers distributed across the genome explain the same proportion of the additive genetic variance, which may be erroneous. Furthermore, the accuracy of selection may be increased by including selected variants from sequence data. The presence of selected variants and alternative SNP weights may enhance accuracy, depending on the method applied to account for this information. This work aimed to study the impacts of weights and inclusions of causative variants in different methods using simulated data and alternative SNP weighting for different genomic selection methods. SNP weights were computed with the termed nonlinear A and quadratic weighting approaches. Accuracies increased with the inclusion of causal variants in the tested scenarios. Gains due to SNP weighting were more prominent in less polygenic scenarios and the nonlinear A approach was superior to the quadratics, however results from nonlinear A were dependent of the equation parameters.

Introduction

The introduction of genomic selection in dairy cattle breeding has led to expressive improvement in progeny transmitting ability's accuracy, especially in young bulls (Mc Hugh *et al.*, 2011). On the other hand, the accuracy of selection may still be increased with the adoption of technologies such as the inclusion of sequence variants (VanRaden *et al.*, 2017). Preceding work in simulated data has shown that accounting for causal variants might improve accuracy to its maximum (>0.99) (Fragomeni *et al.*, 2017). However, researchers were not able to reproduce those findings with real data, and the overall gains in accuracy from including selected variants are modest. Furthermore, depending on the genomic selection method used, the presence of causative variants might bring different results.

Most commercial applications of genomic selection rely on the single-step GBLUP (ssGBLUP) method (Aguilar *et al.*, 2010, Christensen and Lund, 2010). This approach uses phenotypes, genotypes, and pedigrees and allows the inclusion of non-genotyped individuals. Alternative to ssGBLUP, there are methods that rely on the calculation of pseudo-phenotypes and allow only the inclusion of genotyped animals, following the linear predictions with a genomic relationship matrix, proposed by VanRaden (2008), which will be hereby termed GBLUP. The default implementation of ssGBLUP assumes that all SNPs follow the same distribution, which may not be biologically accurate. Modifications on the genomic relationship matrix can be made to assume unequal SNP variances in both GBLUP and ssGBLUP. However, the nuances of weighting strategies are not fully understood and vary with trait, population, and method. For instance, Fragomeni *et al.* (2019) found that weights and causative variants increased accuracy in a GBLUP approach with homogeneous residual, but not with ssGBLUP.

The objective of the current study is to evaluate the impact of alternative SNP weighting in the accuracy of a multi-step method (GBLUP) and ssGBLUP using regular SNP, causal variants, and the combination of both in a single panel. Finally, the effects of weights in different genetic architectures were tested by simulating traits with different numbers of causative quantitative trait nucleotides (QTN).

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_297, © B.F. Santana *et al.* 2022

Materials & methods

The data used were simulated by the package AlphasimR (Gaynor *et al.*, 2020). Three traits with the heritability of 0.3 were assumed, and 100% of the genetic variance was accounted for 100, 1000, or 2,000 quantitative trait nucleotides. The effects of the causative QTNs were sampled from a Gamma distribution with a shape parameter of 0.4 and scaled internally for the simulated genetic variance.

Populations mimicking dairy population characteristics were simulated. Initially, founder genomes were generated by sampling 100k haplotype sequences across 29 diploid chromosomes mimicking the length of the bovine genome. The haplotypes were coded as either zero or one. A total of 10,000 loci sequences were simulated for each chromosome, in which 50k were selected as markers. The founder sample was composed of 5,000 individuals. The same recombination rate for females and males was assumed.

An initial historical population was created over 1000 generations of random mating from the founder population. This process allowed a mutation and drift balance while linkage disequilibrium was established. The breeding population started from a sample of the last generation of the historical population and was carried out for 20 generations. The number of males and females were selected to result in an effective population size of 120. At the end of the simulation process, 440,000 animals were available in the pedigree, wherein 16,800 of those individuals were genotyped.

Genotypes were available for all sires and randomly selected dams sampled from generations 16-19, plus 2,000 animals from generation 20, totaling 16,800 individuals. Phenotypes for animals in generation 20 were not included in the analysis, for they were used for validation. This process mimicked selection for young individuals.

Genomic estimated breeding values (GEBV) were computed using GBLUP and ssGBLUP. For GBLUP three pseudo-phenotypes were included in the analyses, varying from the raw phenotypes, the simulated true breeding value (TBV) with a noise sampled from a N~(0,0.2), and the TBV, termed GBLUP1, GBLUP2, and GBLUP3, respectively. SNP weights were calculated based on SNP effects backsolved from direct genomic value (DGV). The square of the SNP effect was called the quadratic weight aproach, and is similar to what was suggested by Zhang *et al.* (2010) and Wang *et al.* (2013). Additionally, the weights were also computed with the nonlinear A formula as described in VanRaden (2008). Different equation parameters were used on the nonlinear A approach. Accuracy was the correlation between TBV and GEBV for animals in generation 20, and bias was the slope of the regression of TBV on GEBV.

Results

Figure 1 includes the accuracies of prediction for the quadratic weights. This approach returned higher accuracy gain for the less polygenic traits. Additionally, models without causative variants experience losses in accuracy, except for the trait with 100 QTN.

Figure 2 includes the accuracies for the nonlinear A method. The accuracies varied according to the method, trait, and equation parameters. Accuracies either increased or remained stable with nonlinear A weights. However, when large numbers for the constant and limit equation parameters were tested, a drastic decline in the accuracies was observed (results not shown).

Discussion

The inclusion of selected sequence variants in SNP panels increasing prediction accuracy or ability have been reported in previous studies (Moghaddar *et al.*, 2019; VanRaden *et al.*, 2017; Xiang *et al.*, 2021). The correspondent accuracy gain depends on how validation was performed, and how variants were selected.

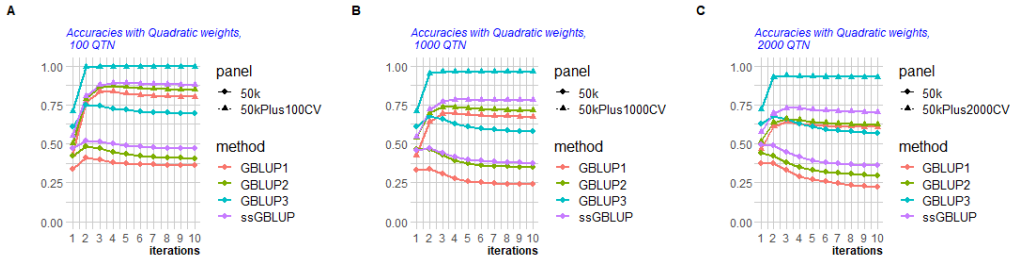


Figure 1. Accuracy of selection with quadratic weights for traits with 100, 1000, and 2,000 simulated QTN (A, B, and C, respectively) The panel included only SNPs (50k) or the inclusion of causative variants (50kplus).

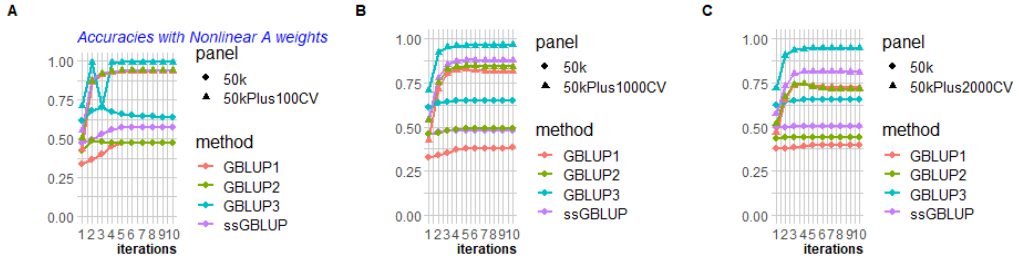


Figure 2. Accuracy of selection with nonlinear A weights for traits with 100, 1000, and 2,000 simulated QTN (A, B, and C, respectively) The panel included only SNPs (50k) or the inclusion of causative variants (50kplus).

Results from real data are still far from achieving what simulations have found, indicating that identifying all variants is a challenge.

Assumptions of non-uniform SNP weights led to expressive gains, especially when causative QTN were included in the panel. For the more polygenic traits, the nonlinear A weighting approach was superior to the quadratic method, in agreement with other authors (Fragomeni *et al.*, 2019; Lopez *et al.*, 2019). When the QTNs were not included in the panel, the quadratic weights reduced accuracy for the more polygenic traits, as expected based on the literature (Fragomeni *et al.*, 2019). However, when extreme values for nonlinear A equation parameters were used, it led to noticeable losses in accuracies, similar from what happens with quadratic weights. For the trait with 100 QTN, the difference between quadratic weights and nonlinear A weights was smaller than the other traits.

Less polygenic traits showing better responses due to the weight implementation is expected, since genes with major effects are more likely to occur in such case. Quadratic weights tend to shrink most of the SNP effects to 0, while increasing others dramatically. For that reason, only the trait affected by a small number of QTNs experienced gains in accuracy with this method. For polygenic traits, extreme shrinkage may remove information from important regions, which results in loss of accuracy.

The ssGBLUP method resulted in lower bias than the GBLUP approach (results now shown), which was expected and has been reported previously (VanRaden *et al.*, 2012). This may have occurred because ssGBLUP takes into consideration all available information simultaneously and also due to high pedigree completeness in this study. Additionally, the use of weights overall increased bias (results not shown), which was also expected (Lee *et al.* 2017, Lourenco *et al.*, 2017).

Finally, the proportional gain from the inclusion of variants and use of weights was higher in GBLUP than in ssGBLUP. Additionally, within GBLUP the method with the raw phenotypes experienced higher proportional gains in accuracy than the model that used the TBV. A possible explanation for those results is that models with less information depend more on the priors and agree with the findings of Fragomeni *et al.* (2019).

The inclusion of causative variants increased the accuracy of selection in all scenarios. The use of weights further increased such gains. The more polygenic traits performed better under more conservative weighting strategies, while the trait with 100 QTN achieved higher accuracy with extreme weight. The use of non-uniform weights may help elucidate features of trait architecture based on the pattern of accuracy variation across scenarios.

References

Aguilar, I., Misztal, I., Johnson, D. L., Legarra, A., Tsuruta, S., *et al.* (2010) J. Dairy Sci., 93(2): 743–752. https://doi.org/10.3168/jds.2009-2730

Christensen, O., & Lund, M. (2010) Genet Sel Evol, 42(2): 1–8. https://doi.org/10.1186/1297-9686-42-2

Fragomeni, B. O., Lourenco, D. A. L., Legarra, A., VanRaden, P. M., & Misztal, I. (2019) J. Dairy Sci., 102(11): 10012–10019. https://doi.org/10.3168/jds.2019-16262

Fragomeni, B. O., Lourenco, D. A. L., Masuda, Y., Legarra, A., & Misztal, I. (2017). Genet Sel Evol, 49(1): 1–11. https://doi.org/10.1186/s12711-017-0335-0

Chris Gaynor, R., Gorjanc, G., & Hickey, J. M. (2020) G3: Genes, Genomes, Genetics, 11(2): 1-5. https://doi.org/10.1093/G3JOURNAL/JKAA017

Lee, J.; Cheng, H.; Garrick, D.; Golden, B.; Dekkers, J.; *et al.* (2017) Genet. Sel. Evol. 49(2): https://doi.org/10.1186/s12711-016-0279-9

Lopez, B. I., Lee, S. H., Park, J. E., Shin, D. H., Oh, J. D., el al. (2019) Genes, 10(1019): 1–13. https://doi.org/10.3390/genes11091013

Lourenco, D.A.L.; Fragomeni, B.O.; Bradford, H.L.; Menezes, I.R.; Ferraz, J.B.S.; *et al.* (2017) J. Anim. Breed. Genet. 134(6): 463–471. https://doi.org/10.1111/jbg.12288

Mc Hugh, N., Meuwissen, T. H. E., Cromie, A. R., & Sonesson, A. K. (2011) J. Dairy Sci., 94(8): 4109–4118. https://doi.org/10.3168/jds.2010-4016

Moghaddar, N., Khansefid, M., Van Der Werf, J. H. J., Bolormaa, S., Duijvesteijn, N., *et al.* (2019) Genet Sel Evol, 51(72): 1–14. https://doi.org/10.1186/s12711-019-0514-2

VanRaden, P. M. (2012) Interbull Bulletin, (45).

VanRaden, P M. (2008) J. Dairy Sci., 91(11): 4414–4423. http://dx.doi.org/10.3168/jds.2007-0980

VanRaden, Paul M., Tooker, M. E., O'Connell, J. R., Cole, J. B., & Bickhart, D. M. (2017) Genet Sel Evol, 49(32): 1–12. https://doi.org/10.1186/s12711-017-0307-4

Xiang, R., MacLeod, I. M., Daetwyler, H. D., de Jong, G., O'Connor, *et al.* (2021) Nature Communications, 12(1): 1–13. https://doi.org/10.1038/s41467-021-21001-0

298. Including selected sequence variants in genomic predictions for age at first calving in Nellore cattle

G.S. Campos[1*], D.A. Silva[1], H.H.R. Neves[2], D. Lourenco[3], G.A.F. Júnior[1], L.F.S. Fonseca[1], L.G. Albuquerque[1] and R. Carvalheiro[1]

[1]Departamento de Zootecnia, Universidade Estadual Paulista 'Júlio de Mesquita Filho', Via de Acesso Professor Paulo Donato Castellane s/n, Vila Industrial, 14884-900 Jaboticabal, SP, Brazil; [2]GenSys Consultores Associados S/S, R. Guilherme Alves, 170, Conj. 304, Jardim Botânico, 90680-000 Porto Alegre, RS, Brazil; [3]University of Georgia, Department of Animal and Dairy Science, 425 River Rd, Athens, GA 30602, USA; gabrielsoarescampos@hotmail.com

Abstract

The objective of this study was to evaluate the gain in accuracy of genomic predictions by adding selected sequence variants to a high-density SNP panel and using weighted or regular single-step GBLUP (ssGBLUP) for age at first calving evaluations in Nellore cattle. Phenotypic records from 224,110 heifers from the Alliance Nellore database and 42,788 genotyped animals were used. All genotyped animals were imputed to whole-genome sequence based on a reference of 151 Nellore sires. A whole-genome association was performed, and 403 sequence variants were selected for inclusion in the high-density SNP panel. Accounting for sequence variants in ssGBLUP did not increase prediction accuracies and only marginally increased accuracy (1%) in weighted ssGBLUP. Our study indicates that genomic selection with ssGBLUP helps to improve prediction accuracy for age at first calving; however, only a marginal gain is observed by including selected sequence variants in the ssGBLUP.

Introduction

Age at first calving (AFC) plays a critical role in identifying heifers that will reach puberty earlier, especially under tropical conditions, and directly affects the herd productivity and profitability of beef cattle. Although genetic progress is slow for reproductive traits such as AFC, genomic information can help increase the prediction accuracy and, consequently, genetic gains. Research interest in whole-genome sequence (WGS) data has increased in the last few years. This data harbors causal mutations, which can be identified and included directly in genomic models without relying on linkage disequilibrium between markers and quantitative trait loci. Pérez-Enciso *et al.* (2015) showed, in a simulation study, that if all causative variants were known, the accuracy of genomic prediction would be perfect (i.e. equal to 1). However, including millions of genome sequence variants into genetic evaluation routines is not advantageous. A practical strategy could be variant selection to identify ones associated with traits of interest (VanRaden *et al.*, 2017).

One suitable method to consider genotyped and ungenotyped animals in genomic evaluation is ssGBLUP (Aguilar *et al.*, 2010). As ssGBLUP is an extension of GBLUP, which in turn is equivalent to SNP-BLUP, SNP effects can be backsolved from ssGBLUP genomic enhanced breeding values (GEBVs), and variance explained by SNPs can be calculated based on those effects. The variance explained by SNPs can then be used as weights in constructing the genomic relationship matrix (G) to account for the fact that SNPs may explain different proportions of variance. This approach may be helpful when causative variants from WGS data are included in the SNP chips used for genomic evaluations (Fragomeni *et al.*, 2017). The main objective of this study was to evaluate the gain in accuracy of genomic predictions using weighted or regular ssGBLUP when adding selected sequence variants to a high-density SNP panel for AFC evaluations.

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_298, © G.S. Campos *et al.* 2022

Materials & methods

Phenotypic, pedigree and genotypic data. AFC records from 224,110 Nellore females belonging to the Aliança Nelore database were used in this study. AFC average was 1,010±133.40 days. Contemporary groups (CG) were formed by heifers born in the same year and season, from the same farm, and raised in the same management group at birth, weaning, and yearling. From 1,348,164 animals in the pedigree data, 42,788 were either genotyped with the Illumina 777k BovineHD bead chip (HD) or a medium-density SNP chip (30k-80k SNPs). In the latter, genotypes were imputed to the HD panel. In addition, 151 influential Nellore sires were sequenced with the Illumina sequencing-by-synthesis technology at an overall average sequence coverage of 14.5× (Fernandes Júnior *et al.*, 2021). A quality control filtering step was implemented according to Fernandes Júnior *et al.* (2021), using the following exclusion criteria: quality by depth <2.0; Fisher Strand test >60.0; root mean square of the mapping quality score <40.0; ranked sum test for the distance of alleles from the end of the reads <-8.0; mapping qualities of reads <-12.5; and SOR >3.0. After this step, we excluded non-biallelic markers and also those with a minor allele frequency lower than 0.01. Marker genotypes with a phred-scaled confidence (a genotype quality score) less than 15 were treated as missing and those SNPs with missing values for more than 40 individuals (26.5% of the total population) were removed from the analyses. A total of 30,394,484 sequence SNPs in autosomal regions remained after quality control. All genotyped individuals were imputed to WGS using those 151 sires as a reference population. After imputation with the software FImpute3 (Sargolzaei *et al.*, 2014), we filtered out variants with minor allele frequencies lower than 0.001 and linkage disequilibrium values (r^2) greater than 0.9 using PLINK 2.0 (Chang *et al.*, 2015). After this step, 4,809,603 SNPs were retained for subsequent analysis.

Whole genome sequence association analyses. A model considering CG as a fixed effect, random additive genetic and residual effects were used to estimate variance components and estimated breeding values (EBVs) with regular BLUP. The EBVs were used to calculate deregressed breeding values as proposed by Garrick *et al.* (2009) and they were used as response variables for the whole-genome association analysis. The software GCTA (Yang *et al.* 2014) was used to estimate SNP effects via linear mixed models with a random polygenic effect based on G. We also performed a principal component analysis and included the first eight principal components as covariates in the model to account for possible population stratification. To prevent SNP double-counting, we utilized the option LOCO (leaving-one-chromosome-out) that excludes the chromosome on which the candidate SNP is located (Yang *et al.* 2014). The significant variants were selected by using a false discovery rate lower than 5% for multiple test correction. The selected variants (SV) not present in the HD panel were included for genomic prediction.

Genomic prediction. The AFC GEBV were computed using ssGBLUP with the same animal model described in the previous subsection. Two scenarios were tested to compute GEBV as follows; (a) using only HD SNPs, and (b) using HD SNPs and selected variants (HD+SV). Furthermore, the ssGBLUP was performed without and with weights iteratively, with G constructed with quadratic or nonlinear A weights. The weighted ssGBLUP was performed only for the scenario considering HD+SV. We chose five iterations for the weighted ssGBLUP as suggested by Fragomeni *et al.* (2019), and results are shown only for the last iteration. The pedigree BLUP was also applied to assess gains in prediction accuracy when using genomic information.

Validation. Validation was performed by the linear regression method proposed by Legarra and Reverter (2018), where young animals had no phenotypes in the reduced dataset (r) and complete records in the whole dataset (w). The prediction accuracy ($\widehat{acc}$), which considers the covariance between the (G)EBV in the whole and reduced datasets, was calculated as follows: $\widehat{acc} = \sqrt{cov[(G)EBV_w, (G)EBV_r]/(1-\bar{F})\sigma_u^2}$, where cov is the sample covariance, $\bar{F}$ is the average inbreeding coefficient of validation animals (i.e. young animals) and σ_u^2 is the additive genetic variance computed with the whole data set. The regression coefficient

(b_1) of (G)EBV$_w$ on (G)EBV$_r$ was used to assess the degree of inflation/deflation of pedigree and genomic predictions. Statistics were calculated for genotyped and non-genotyped young heifers and for genotyped young bulls that were at most two years old.

Results

The heritability for AFC was 0.06±0.004, which is lower than in other studies with Nellore cattle (Mota *et al.*, 2017). Regarding the whole genome association analysis, 403 significant variants were included in the HD SNP panel for genomic prediction. Among those significant variants, 143 are on chromosome BTA23.

The prediction accuracies of GEBV using regular or weighted ssGBLUP are presented in Table 1. In general, the GEBV accuracies ranged from 0.39 to 0.43 between the methods, and b_1 varied from 0.87 to 1.16 across the validation groups (young heifers and young genotyped bulls). The GEBV accuracies from ssGBLUP (weighted or not) outperformed those from BLUP, with an average increase in accuracy of 18% for young heifers and 48% for young genotyped bulls. Including SV in the regular ssGBLUP has not increased prediction accuracies and only marginally increased accuracy (1%) in weighted ssGBLUP with nonlinear A method (HD+SV).

Discussion

Even though AFC has low heritability, our study indicates that genomic selection with ssGBLUP can help to improve prediction accuracy and, therefore, accelerate genetic gains for AFC in this Nellore population. Using SV combined with SNP panels considerably increased prediction accuracy in simulation studies (Fragomeni *et al.*, 2017) when applying weighted ssGBLUP, compared to the regular ssGBLUP. However, we found only a marginal increase (1%) when including SV in weighted ssGBLUP with nonlinearA weights compared to the HD SNP panel. This result is similar to those from Fragomeni *et al.* (2019) for stature in US Holsteins. The authors compared ssGBLUP with and without weights and did not find any gain by including nearly 17,000 significant variants to a medium-density panel. According to Lourenco *et al.* (2020), the amount of information used in ssGBLUP overwhelms any *a priori* assumption on SNP effects, being less sensitive to weighting approaches for large data sets. These authors suggested that an ideal algorithm should assign large variance to causative SNPs while reducing the variance of nearby SNPs. Exploring methods to increase the power of detection of causal variants, as multi-trait meta-analysis or gene expression and genome annotation could help avoid spurious associations. The use of additional biological sources of information (i.e. multi-omics data) could also help to increase the likelihood of identifying mutations that contribute to the variance of complex traits and possibly, increase the prediction accuracies.

Table 1. Accuracy of (G)EBV and regressions coefficients (b_1) of (G)EBV in the whole data set on (G)EBV in the reduced data set when including or excluding selected variants (SV) in single-step GBLUP (ssGBLUP) for young heifers and young genotyped bulls for age at first calving.

Method[1]	Scenario[2]	Young heifers		Young genotyped bulls	
		Accuracy	b_1	Accuracy	b_1
BLUP	Pedigree	0.35	0.87	0.37	0.89
ssGBLUP	HD	0.42	0.88	0.56	0.93
ssGBLUP	HD+SV	0.42	0.88	0.56	0.93
WssGBLUP_quadratic	HD+SV	0.39	0.89	0.51	1.16
WssGBLUP_nonlinearA	HD+SV	0.43	0.87	0.57	0.95

[1] BLUP: regular pedigree BLUP; ssGBLUP: regular single-step genomic BLUP; WssGBLUP_quadratic: weighted single-step GBLUP using quadratic weights in the fifth iteration; WssGBLUP_nonlinearA: weighted single-step GBLUP with nonlinear A method in the fifth iteration.

[2] HD: high density SNP panel; SV: selected sequence variants (403 SNPs).

Acknowledgements

To 'Coordenação de Aperfeiçoamento de Pessoal de Nível Superior' (CAPES, finance code 001), and São Paulo Research Foundation (FAPESP # 17/10630-2, # 2018/20026-8, # 2019/06361-1 and # 2020/03975-6) for financial support, and to the Genetic Breeding Programs participating in the Aliança Nelore (www.gensys.com.br) for providing the data.

References

Aguilar I., Misztal I., Johnson D.L., Legarra A., Tsuruta S., *et al.* (2010) Journal of Dairy Science. 93:743–752. https://doi.org/10.3168/jds.2009-2730

Chang C.C., Chow C.C., Tellier L.C., Vattikuti S., Purcell S.M., *et al.* (2015) Gigascience, 4(1):s13742–015–0047–8. https://doi.org/10.1186/s13742-015-0047-8

Fernandes Júnior G.A., Carvalheiro R., de Oliveira H.N., Sargolzaei M., Costilla R., *et al.* (2021). Genetics Selection Evolution, 53:27. https://doi.org/10.1186/s12711-021-00622-5

Fragomeni B.O., Lourenco D.A., Masuda Y., Legarra A. and Misztal I. (2017) Genetics Selection Evolution, 49:59. https://doi.org/10.1186/s12711-017-0335-0

Fragomeni B.O., Lourenco D.A.L., Legarra A., VanRaden P. M. and Misztal I. (2019) Journal of Dairy Science, 102(11):10012-10019. https://doi.org/10.3168/jds.2019-16262

Garrick D.J., Taylor J.F., and Fernando R. L. (2009) Genetics Selection Evolution, 41:55. https://doi.org/10.1186/1297-9686-41-55

Legarra A., and Reverter A. (2018) Genetics Selection Evolution, 50:53. https://doi.org/10.1186/s12711-018-0426-6

Lourenco D., Legarra A., Tsuruta S., Masuda Y., Aguilar I., *et al.* (2020) Genes, 11(7):790. https://doi.org/10.3390/genes11070790

Mota R.R., Guimarães S.E.F., Fortes M.R.S., Hayes B., Silva F.F., *et al.* (2017) Journal of Animal Breeding and Genetics, 134(6):484-492. https://doi.org/10.1111/jbg.12299

Pérez-Enciso M., Rincón J. C., and Legarra A. (2015) Genetics Selection Evolution, 47:43. https://doi.org/10.1186/s12711-015-0117-5

Sargolzaei M., Chesnais J.P., and Schenkel F. S. (2014) BMC Genomics, 15:478. https://doi.org/10.1186/1471-2164-15-478

299. Increasing accuracy of single-step evaluations of functional longevity including information from correlated traits

L.-H. Maugan[1*], T. Tribout[1], R. Rostellato[1], S. Mattalia[1,2] and V. Ducrocq[1]

[1]Université Paris Saclay, INRAE, AgroParisTech, GABI, 78350 Jouy-en-Josas, France; [2]Institut de l'Elevage, UMT eBIS, 78350 Jouy-en-Josas, France; laure-helene.maugan@inrae.fr

Abstract

Breeding values for low heritable traits are difficult to estimate. Often, if traits are correlated to each other they should be analysed together in a multivariate genetic evaluation to improve accuracy. However, the correlated traits may not be described by the same model. To overcome the resulting complexity, a 'combined' genetic evaluation consisting of two successive evaluations has been implemented in France, resulting in more accurate breeding values. The aim of this study was to verify that a combined Single-Step evaluation is more accurate than a univariate Single-Step evaluation. To that end, genomic breeding values obtained with univariate and combined Single-Step evaluation were compared for functional longevity and five other correlated traits in 2016 and 2021. The average and the correlation between breeding values in 2021 and 2016 suggested that the combined Single-Step evaluation is more accurate than univariate Single-Step evaluations.

Introduction

Longevity is a complex trait with a low heritability. Genetic evaluation for this trait requires a complex survival model to account for highly nonlinear data and for the existence of censored records, corresponding to cows that are still alive (i.e. only a lower bound of their actual longevity is known). Another level of complexity is added by the fact that we are mostly interested in the functional longevity (FL), i.e. the ability to avoid involuntary culling regardless to the milk production level of a cow. Because of these intricacies, the genetic evaluation of French dairy cows on FL is based on a Weibull sire model (Ducrocq, 2005). Moreover, FL is correlated to other traits such as udder health, fertility and type traits that can be used as predictors. Unfortunately, all these traits are described by very different models, making difficult the development of a multiple trait evaluation of FL and of its predictors using raw performances. This is why a 'combined' genetic evaluation was developed (Ducrocq *et al.*, 2001) including FL and correlated traits. It relies on a two-step approach: first, univariate evaluations (or multivariate evaluations by group of traits described by a same model) are performed. For all traits, records are corrected for all fixed effects (and possibly for the heterogeneity of residual variances). As a result, each recorded animal gets one 'pseudo-record' for each observed trait and an associated weight is calculated. Then a multivariate genetic evaluation is carried out on these weighted pseudo-records, based on a very simple model with only one fixed effect, a year of birth effect, in order to account for selection on correlated traits not accounted for in the first step, and the individual genetic value of each animal (Govignon-Gion *et al.*, 2016). Such combined evaluations have major advantages: as a result, all animals have an estimated breeding value (EBV) on all traits and the construction of a total merit index is straightforward.

In the context of single step evaluations, the combination of Genomic Estimated Breeding Values (GEBVs) from separate genomic evaluations is not optimal. In order to illustrate the benefits of a combined Single-Step (CSS) approach and to evaluate low heritability traits such as FL and its predictors in comparison with several univariate SS (USS) evaluations, performances of six correlated traits were used in the Montbéliarde breed: FL, somatic cell score (SCS), conception rate in heifers (CRH) and lactating cows (CRC), clinical mastitis (CM) and udder depth (UD).

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_299, © L.-H. Maugan *et al.* 2022

Materials & methods

Data. The number of records available in Montbéliarde since 2002 and included in the analyses is listed in Table 1. Records were obtained from the French bovine national database.

Method. The FL pseudo-records were derived from the Weibull sire model mentioned above (Ducrocq, 2005). For the other traits, pseudo-records were derived from the SS evaluation results, applying the same univariate models as in the corresponding current official evaluations. Two SS evaluations were run: one considering all available performances (recorded from 2002 to mid-2021) and one considering truncated data after deletion of the performances recorded from 2016 to 2021 to mimic a SS evaluation performed on data available in 2016.

The objective was to verify whether the GEBVs obtained for young animals (2016 SS evaluations) predicted more accurately the CSS GEBVs of 2021 (supposedly the most accurate GEBVs), using a CSS instead of a USS. For this purpose, the GEBVs of the 10% best cows estimated using a CSS evaluation and all data available in 2021 were compared to the GEBVs calculated with a USS or a CSS, and using only data available five years before. This group of best cows included 1,972 cows born and genotyped in 2014, having their first performances in 2016 or later.

Model and software. The general Single-Step model equation for the first step of the CSS evaluation can be summarised as:

$$\mathbf{y}_{kj} = \sum_{i=0}^{n} \mathbf{b}_k + \mathbf{a}_j + \mathbf{e}_{kj} \quad (1)$$

with y_{kj} the vector of performances of animal j, b_k the corresponding vector of fixed effects k. For each trait, performances were modelled using the effects used in the current routine evaluation run by GenEval (https://www.geneval.fr/indexations-mutualisees), these models being adapted to each group of traits. a_j is the vector of additive genetic effects of animal j with $a = \{a_j\} \sim N(0, H \otimes G)$, G is the genetic (co)variance matrix between traits (in the case of multivariate evaluation), H is the genetic/genomic relationship matrix between animals, and e_{kj} is the vector of residuals for animal j with $e = \{e_j\} \sim N(0, I \otimes R)$, R being the residual (co)variance matrix between traits (in the case of multivariate evaluation). For UD and CM, phenotypic heterogeneous variances were accounted for using Meuwissen's approach (Meuwissen *et al.*, 1996).

GEBVs for USS_2016 and pseudo-records for each trait are estimated using model (1). These pseudo-records are used in the second step of the CSS evaluation of which the equation is:

$$y_{kj} = \mu_{year} + a_j + e_{kj}, \quad (2)$$

Table 1. Number of records for the investigated traits.[1]

	FL	SCS	CRH	CRC	CM	UD
From 2002 to 2021	1,781,415	4,095,686	2,123,962	5,566,297	2,028,701	1,142,654
From 2016 to 2021	553,930	1,342,868	505,578	1,775,346	689,419	351,915
% of erased performances[2]	31.1%	32.8%	23.8%	31.9%	34%	30.8%

[1] FL: functional longevity, SCS: milk somatic cell score, CRH: conception rate of heifers, CRC: conception rate of lactating cows, CM: clinical mastitis, UD: udder depth.

[2] i.e. ignored in the 2016 analyses.

where y_{kj} are the pseudo-records of animal j for trait k, μ_{year} is the birth year of the animal. The objective of μ_{year} is to adjust the additive genetic effects for potential selection on correlated traits not accounted for in the (univariate) first step. a_j and e_{kj} are the same as in (1) but in a multivariate context.

Unknown parent groups were used in the SS evaluations. They were defined according to the birth year of the animals with unknown parent.

Both Single Step evaluations were performed using the HSSGBLUP software (Tribout *et al.*, 2020).

Results & discussion

Average of GEBVs. In Figure 1, the average GEBVs of the top 10% cows increased between 2016 (CSS or USS) and 2021. This was expected, since the increase in accuracy due to additional information leads to a greater variability of GEBVs in 2021 and thus to higher GEBV means of the best females. Moreover, for CM, CRH and CRC in 2016 the average GEBV for USS is lower than for CSS, which could be explained again by higher accuracies obtained with the CSS. For the other traits, the difference is small. It could be explained for UD and SCS by the fact that the accuracy is not much improved using a combined evaluation, since the heritability of these traits is high enough (SCS: 15%; UD: 38%). For FL, the difference between average GEBVs in USS_2016 and CSS_2016 are about 0.02 while is it about 0 for SCS and UD. The theoretical heritability of FL is about 14.5% but is lower in practice because of censored animals (i.e. still alive) at the time of evaluation (GenEval, 2021).

Correlation between GEBVs. Correlations between GEBVs (Figure 2) are higher with the CSS than with the USS evaluation for all traits. It indicates that in 2016, CSS is more efficient than USS to predict CSS in 2021. These differences are small for the most heritable traits (SCS and UD) but are quite clear for the low heritable traits (CM, FL, CRC and CRH).

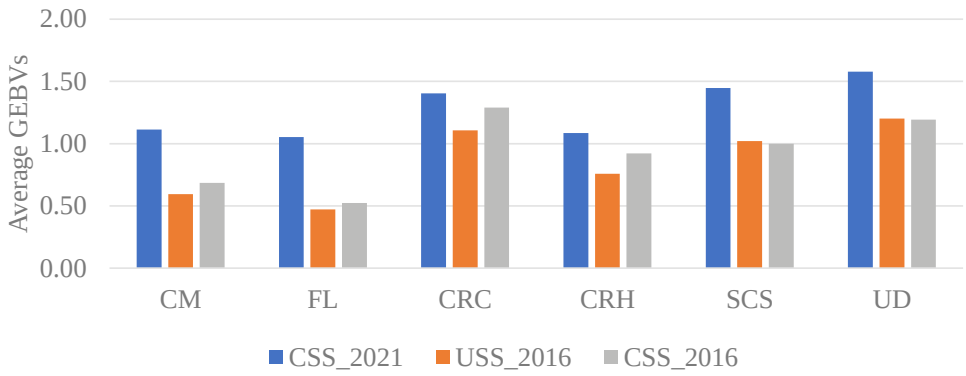


Figure 1. Average GEBVs obtained with the different evaluations for the top 10% cows and for the investigated traits. FL: functional longevity, SCS: milk somatic cell score, CRH: conception rate of heifers, CRC: conception rate of lactating cows, CM: clinical mastitis, UD: udder depth, CSS 2021: combined Single-Step evaluation of 2021, CSS 2016: combined Single-Step evaluation of 2016, USS 2016: univariate Single-Step evaluation of 2016. All results expressed in genetic standard deviation units.

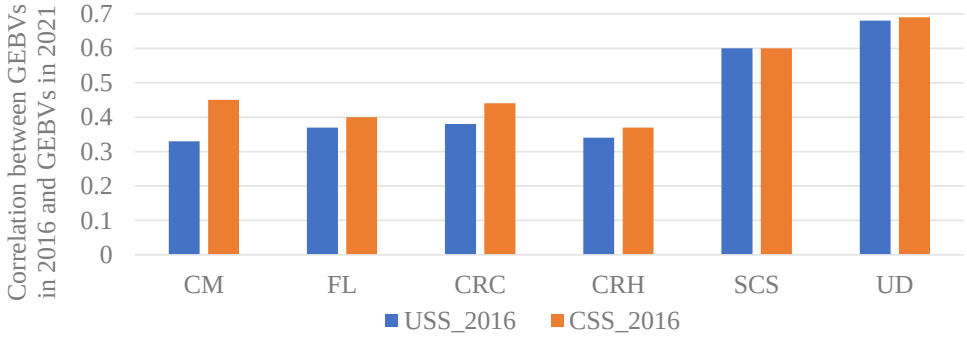


Figure 2. Correlations between GEBVs obtained in 2021 with those obtained in 2016 (USS or CSS) for the top 10% cows and for the investigated traits. FL: functional longevity, SCS: milk somatic cell score, CRH: conception rate of heifers, CRC: conception rate of lactating cows, CM: clinical mastitis, UD: udder depth, CSS 2021: combined Single-Step evaluation of 2021, CSS 2016: combined Single-Step evaluation of 2016, USS 2016: univariate Single-Step evaluation of 2016.

Conclusions

A combined (multivariate) SS evaluation leads to a gain in accuracy for GEBVs for low heritable traits. Even though differences between GEBVs in USS 2016 and CSS 2016 were expected to be higher, the results of this preliminary study are encouraging.

The genetic parameters (particularly genetic correlations) used in our study were estimated a few years ago and did not consider the genomic information on bulls. This could explain why relatively minor differences were found. Therefore, we plan to use updated genetic parameters and to complete the analyses with a study of the GEBV reliability of bulls and cows.

Acknowledgements

This study received funding from the Ministry of Agriculture, through the CASDAR (Compte d'Attribution Spéciale du Développement Agricole et Rural).

References

Ducrocq V. (2005) Anim. Sci. 80(3):249-256. https://doi.org/10.1079/ASC41720249

Ducrocq V., Boichard D., Barbat A., Larroque H. (2001). Implementation of an approximate multitrait BLUP evaluation to combine production traits and functional traits into a total merit index. Proc. of the 52nd EAAP, Budapest, Hungary.

GenEval, Evaluation génétique des taureaux Montbéliards (2021). Available at: https://www.geneval.fr/indexations-races-bovines

Govignon-Gion A., Dassonneville R., Baloche G., Ducrocq V. (2016) Anim. 10(4):558–565. https://doi.org/10.1017/S1751731115002529

Meuwissen T.H.E., De Jong G., Engel B. (1996) J. Dairy Sci. 79(2):310–316. https://doi.org/10.3168/jds.S0022-0302(96)76365-8

Tribout T., Ducrocq V., Boichard D. (2020) HSSGBLUP: a Single-Step SNP BLUP genomic evaluation software adapted to large livestock populations. Proc. of the 6th ICQG.

300. Multi-trait genomic estimation of genetic parameters for growth and carcass traits of Duroc pigs

E.C. Akanno[1,2*], D.M. Thekkoot[1], C. Zhang[1], C. Bierman[1], G. Plastow[2] and R.A. Kemp[1]

[1]Genesus Inc., 101 2nd St., Oakville, MB, R0H 0Y0, Canada; [2]Livestock Gentec, University of Alberta, 116 St & 85 Ave, Edmonton, AB, T6G 2R3, Canada; eakanno@genesus.com

Abstract

A total of 10,581 Duroc pigs with genotype, phenotype and pedigree were used to estimate genetic parameters for nine growth and carcass traits utilising pedigree- and genomic-based multi-trait animal models implemented in MTG2 software. Moderate to high heritability that ranged from 0.12±0.01 for average daily feed intake (FI) to 0.50±0.02 for carcass lean (CL) were observed. Genomic-based estimates were slightly larger than the pedigree-based estimates, on average, and had smaller standard errors. Genetic correlations from a genomic-based model were high for the following trait pairs; Ultrasound fat depth (UF) and Intramuscular fat percent (IF) (0.65), FI and Days to 120 kg (-0.63), UF and Carcass fat (CF) (0.97), IF and CF (0.63), CF and FI (0.59), Ultrasound lean depth and CL (0.91), and IF and Marbling score (0.65). Therefore, facilitating improvement in accuracy of genetic parameters by incorporating genomic information is encouraged.

Introduction

Genetic parameters as well as components of total phenotypic variance are essential for the design of pig genetic improvement programs (Akanno *et al.* 2014). Traditionally, genetic parameters are estimated using large, across-generation datasets with reliable pedigrees, however, traits that are difficult to measure such as feed intake and those measured after harvest such as carcass traits may lack sufficient data for accurate estimation. With the growing availability of genomic data, genetic parameter estimation can be improved by incorporating genomic information. In addition, the development of enhanced statistical models, improvement in computing capacity and the likelihood that population parameters may have evolved due to improved rearing conditions and selection may necessitate the re-estimation of genetic parameters from time to time (Wray and Visscher, 2008). The objective of this study was to compare a multi-trait pedigree- and genomic-based estimates of genetic parameters for combined growth and carcass traits in Duroc pigs.

Materials & methods

Data. A total of 10,872 Duroc pigs were genotyped with a custom Affymetrix Axiom® Porcine Genotyping Array (59,400 SNPs). Genotypes were filtered for 95% animal and SNP call rate and <0.02 minor allele frequency resulting in a total of 45,240 SNPs and 10,581 genotyped animals. Animals with missing genotype calls were imputed using FImpute software (Sargolzaei *et al.* 2014). Pedigree records were available on 13,241 individuals, 581 sires and 2,497 dams. Phenotypes were available for days to 120 kg (DA), ultrasound fat depth (UF), ultrasound lean depth (UL), average daily feed intake (FI), intramuscular fat percent (IF), hot carcass weight (HW), carcass fat depth (CF), carcass lean depth (CL) and marbling score (MS). MS was a subjective score from 1 to 6 and evaluated by a single, experienced technician. UF and UL were adjusted to 120 kg while HW was adjusted to 130 kg. The data were edited to remove records with more or less than 4 standard deviations from the mean.

Statistical analyses. Genetic parameters of nine growth and carcass traits were estimated using a multi-trait animal mixed model. The model utilised either a numerator relationship matrix (A) constructed from pedigree, or a genomic relationship matrix (G) constructed from 45,240 SNPs using $\mathbf{WW}'/2\sum p_i(1-p_i)$, where W contains centred genotypes codes and p_i is the minor allele frequency for locus i (VanRaden, 2008). The following mixed effect model in matrix notation (1) implemented in MTG2 version 2.22 (Lee *et al.* 2016) was applied.

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_300, © E.C. Akanno *et al.* 2022

$$y = Xb + Z_1l + Z_2a + e \quad (1)$$

where y is the vector of observations for multiple growth and carcass traits, b is the vector of varying fixed effects depending on the trait which includes sex applied to all traits, Improvest® (https://zoetis.cvpservice.com/product/view/1198451) applied to growth traits, ultrasound machine type applied to ultrasound fat and lean depth, and management groups. Management group for growth traits was defined as probe year, probe week, location, and technician while management group for carcass traits was defined as slaughter date and technician, l is the vector related to random litter effects ~N (0, L⊗I), a is the vector related to random direct additive genetic effects of animals ~N (0, G_a⊗A) for pedigree-based model or ~N (0, G_a⊗G) for genomic-based model and e is the vector related to random residuals ~N (0, R⊗I). X is the design matrix that relates the fixed effects to the observation, Z_1 is a design matrix that relates random litter effects to observations, Z_2 is a design matrix that relates random animal genetic effects to observations, L, G_a, and R, are litter, additive genetic and residual (co)variance matrices between traits, respectively, and I is identity matrix.

Different four-trait analyses were initially performed separately for growth and carcass traits and their results were found to be comparable to a nine-trait model that includes all traits. Thereafter, only a nine-trait multi-trait model was applied and reported. Total phenotypic variance (σ_p^2) was calculated $\sigma_p^2 = \sigma_l^2 + \sigma_a^2 + \sigma_e^2$, where the σ_l^2, σ_a^2 and σ_e^2 are the litter variance, additive genetic variance, and residual variance, respectively. Heritability (h^2) was computed as $h^2 = \sigma_a^2 / \sigma_p^2$. Genetic correlations ($r_g$) were estimated as $r_g = \sigma_{a_1 a_2} / \sqrt{\sigma_{a1}^2 \times \sigma_{a2}^2}$, where the σ_{a1}^2 and σ_{a2}^2 are the genetic variance estimates for traits 1 and 2, respectively, and $\sigma_{a_1 a_2}$ is the genetic covariance between two traits.

Results

The descriptive summary and estimates of h^2 from pedigree- and genomic-based models are presented in Table 1. Moderate to high h^2 ranged from 0.12±0.01 (FI) to 0.50±0.02 (CL) across the traits and models. Genomic-based model h^2 were slightly larger for UF, UL, CF, CL, and MS than estimates from a pedigree-based model. Also, the standard error (SE) of h^2 estimates for the genomic-based model were generally smaller than those of pedigree-based model. Traits measured on live animals (UF and UL) had similar h^2 to those measured on carcass (CF and CL) in both the pedigree- and genomic-based models.

Table 1. Number of records (N), mean, standard deviation (SD), heritability (h^2) ± standard error (SE) for pedigree- and genomic-based estimation in Duroc pigs.

Traits	N	Mean	SD	Pedigree h^2±SE	Genomic h^2±SE
Days to 120 kg (d)	10,859	162.27	11	0.16±0.02	0.16±0.01
Ultrasound fat (mm)	10,807	13.98	2.78	0.42±0.03	0.48±0.01
Ultrasound lean (cm^2)	10,840	70.24	5.24	0.43±0.03	0.48±0.01
Intramuscular fat percent	7,778	2.71	0.87	0.28±0.03	0.27±0.02
Average daily feed intake (kg/d)	8,387	2.93	0.62	0.16±0.02	0.12±0.01
Hot carcass weight (kg)	3,377	103.19	3.48	0.22±0.05	0.17±0.03
Carcass fat (mm)	3,395	15.88	3.74	0.44±0.04	0.49±0.02
Carcass lean (cm^2)	3,402	68.63	5.15	0.45±0.05	0.50±0.02
Marbling score	3,402	2.97	1.15	0.25±0.05	0.31±0.03

Table 2 shows the variance components and genetic correlations among the nine traits studied. Genetic variances from the genomic-based model were slightly greater than those from the pedigree-based model for majority of the traits (UF, UL, CF, CL, and MS). Similarly, the genetic covariances from the two models showed similar signs for most of the trait combinations except for a switch in sign between DA and HW. The estimates of genetic correlation from the pedigree- and genomic-based models were similar in magnitude and direction except for a switch in sign for DA and HW. Genomic-based model also had smaller ranges of SE for the (co)variances and genetic correlation than the pedigree-based model (Table 2). The litter and residual (co)variances from the pedigree- and genomic-based models were generally similar. Also, the signs of the covariance estimates from both models were generally the same except for some trait pairs in which the covariance estimates were close to zero.

Discussion

The moderate to high h^2 observed for most of the growth and carcass traits were within the range of estimates (0.10-0.47) reported by MacNeil and Kemp (2015) and Esfandyari *et al.* (2020) utilising a related purebred Duroc population. Genomic-based models gave slightly larger h^2 estimates for UF, UL, CF, CL and MS which corroborates the findings of Esfandyari *et al.* (2020) that compared the difference in estimates for pedigree-based and a single step genomic-based model (ssGBLUP). This is probably because genomic-based models included contribution of causal variants in linkage disequilibrium with SNP markers in the estimation process. Also, utilising the same size of data but including genomic information in the estimation

Table 2. Additive genetic variances on the diagonal, genetic covariance above diagonal and genetic correlation below diagonal for growth and carcass traits[1] in Duroc pigs[2].

	DA	UF	UL	IF	FI	HW	CF	CL	MS
DA	**13.68**	-1.66	3.07	-0.21	-0.38	0.06	-3.53	4.61	-0.61
	(13.98)	(-0.63)	(2.01)	(-0.06)	(-0.38)	(-0.09)	(-2.49)	(1.13)	(-0.60)
UF	-0.29	**2.36**	-1.30	0.39	0.09	0.46	3.19	-1.44	0.31
	(-0.10)	**(3.08)**	(-1.73)	(0.42)	(0.13)	(0.31)	(4.16)	(-1.76)	(0.25)
UL	0.27	-0.28	**9.15**	-0.21	-0.23	0.93	-1.93	8.24	-0.72
	(0.16)	(-0.29)	**(11.71)**	(-0.31)	(-0.20)	(1.41)	(-2.73)	(11.47)	(-0.93)
IF	-0.16	0.68	-0.19	**0.14**	0.02	0.06	0.52	-0.23	0.14
	(-0.05)	(0.65)	(-0.24)	**(0.14)**	(0.02)	(0.02)	(0.58)	(-0.36)	(0.15)
FI	-0.57	0.34	-0.42	0.28	**0.03**	-0.01	0.18	-0.24	0.04
	(-0.63)	(0.46)	(-0.36)	(0.34)	**(0.03)**	(-0.02)	(0.23)	(-0.16)	(0.03)
HW	0.01	0.25	0.25	0.13	-0.04	**1.50**	0.32	1.32	-0.08
	(-0.02)	(0.16)	(0.38)	(0.04)	(-0.11)	**(1.20)**	(0.15)	(1.27)	(-0.03)
CF	-0.44	0.95	-0.29	0.64	0.45	0.12	**4.73**	-2.36	0.39
	(-0.27)	(0.97)	(-0.32)	(0.63)	(0.59)	(0.05)	**(6.04)**	(-2.95)	(0.41)
CL	0.39	-0.29	0.84	-0.19	-0.41	0.33	-0.33	**10.48**	-0.91
	(0.08)	(-0.27)	(0.91)	(-0.26)	(-0.27)	(0.31)	(-0.33)	**(13.58)**	(-1.02)
MS	-0.31	0.38	-0.44	0.69	0.41	-0.12	0.33	-0.52	**0.29**
	(-0.26)	(0.23)	(-0.44)	(0.65)	(0.27)	(-0.04)	(0.27)	(-0.44)	**(0.39)**

[1] DA = Days to 120 kg (d), UF = ultrasound fat (mm), UL = ultrasound lean (cm^2), IF = intramuscular fat percent, FI = average daily feed intake (kg/d), HW = hot carcass weight (kg), CF = carcass fat (mm), CL = carcass lean (cm^2), MS = marbling score.

[2] Genomic-based estimates in brackets.

[3] Standard error for variances in the pedigree-based model ranged from 0.005-2.000 while genomic-based model ranged 0.003-1.360; standard error for covariances in the pedigree-based model ranged from 0.005- 1.132 while genomic-based model ranged from 0.004-0.827; standard error for genetic correlation in the pedigree-based model ranged from 0.005-0.151 while genomic-based model ranged from 0.009-0.103.

process decreased the SE of estimates for all the studied traits, thus, accuracy of estimates is improved for genomic-based model than for pedigree-based model.

The genetic, litter and residual variances and covariances among trait combinations are required for constructing complete variance-(co)variance matrices for joint genetic evaluation of growth and carcass traits. Estimates of the variance components observed in this study are consistent with estimates reported elsewhere (Akanno *et al.* 2014; Esfandyari *et al.* 2020). The high positive genetic correlation observed among traits measured in live animals (UF, UL, and IF) and those measured in carcass (CF, CL, and MS) are expected because of the linear relationship among these traits and has been previously observed (Esfandyari *et al.* 2020). In addition, estimates from the genomic-based model had smaller SE and thus more accurate than those from the pedigree-based model.

Conclusions

Estimation of genetic parameters can be very challenging due to requirements for sophisticated computing techniques and large datasets. It is even reassuring to obtain results that appear similar when applying different models and methods. Therefore, facilitating improvement in accuracy of genetic parameters by incorporating genomic information is encouraged.

References

Akanno E.C., Schenkel F.S., Quinton, V.M., Friendship R.M., and Robinson, J.A.B. (2013) Livest. Sci. 152:101–113. http://dx.doi.org/10.1016/j.livsci.2012.07.021

Esfandyari H., Thekkoot D., Kemp R., Plastow G., and Dekkers J. (2020) J. Anim. Sci. 98(12): skaa379. https://doi.org/10.1093/jas/skaa379

Lee S.H., and van der Werf J.H.J. (2016) Bioinformatics 32:1420-1422 https://doi.org/10.1093/bioinformatics/btw012

MacNeil M.D., and Kemp R.A (2015) Can. J. Anim. Sci. 95:155-159. https://doi.org/10.4141/cjas-2014-089

Sargolzaei M., Chesnais J.P., and Schenkel F.S. (2014) BMC Genomics 15:478. https://doi.org/10.1186/1471-2164-15-478

VanRaden P.M. (2008) J. Dairy Sci. 91(11):4414–4423. https://doi.org/10.3168/jds.2007-0980

Wray N., and Visscher P. (2008) Nature Education 1(1):29

301. Accounting for nuclear- and mito-genome in genetic evaluation and breeding of dairy cattle

G.M. Fortuna[1,2,3*], B.J. Zumbach[2], M. Johnsson[3], I. Pocrnic[1] and G. Gorjanc[1]

[1]The University of Edinburgh, The Roslin Institute, EH25 9RG, Edinburgh, United Kingdom; [2]Georg-August-Universität Göttingen, Albrecht-Thaer-Weg 3, 37075 Göttingen, Germany; [3]Swedish University of Agricultural Sciences, Gerda Nilssons väg 2, Uppsala, Sweden; g.mafra-fortuna@sms.ed.ac.uk

Abstract

Studies indicate that mito-genome variation impacts phenotypes in a range of species. In dairy cattle up to 5% of phenotypic variation for milk production has been associated to mito-genome variation. Bearing in mind that milk production is a very energy demanding process that inflicts systemic physiological changes, it is logical to expect that it depends on well-functioning mitochondria. Here we evaluated the impact of accounting for nuclear- and mito-genome variation in genetic evaluation and breeding of dairy cattle with the means of stochastic simulations. Results show that accounting for mito-genome variation can increase the accuracy of estimated breeding values by up to 0.04, particularly in females, and consequently increase genetic gain.

Introduction

Mitochondria are involved in many critical cellular processes via proteins encoded in nuclear-genome and mito-genome. While variation in nuclear-genome is driven by mutations, recombinations and combination of parental genomes, variation in mito-genome is driven by mutations that are inherited from dams to their offspring in haploid form without recombination. Dorji *et al.* (2021) recently reported a high level of variation in mito-genomes across globally important cattle breeds, with most of the variation attributed to within-breed component. Considering the critical role of mitochondria and the observed variation, it is expected that mito-genome variation impacts livestock production. Recent and past studies suggest that mito-genome accounts for up to five percent of phenotypic variation for yield traits in cattle (Boettcher *et al.*, 1996; Gibson *et al.*, 1997; Brajković, 2019). Previous studies have evaluated the impact of accounting for mito-genome variation in genetic evaluation in livestock using pedigree information, while this impact has not yet been evaluated with genomic information. Furthermore, given that mito-genome has a different mode of inheritance than nuclear-genome, we need to revise the definition of a breeding value that accounts for the inheritance of each genome in males and females.

The goal of this study was to evaluate the impact of accounting for nuclear-genome and mito-genome variation in estimating breeding values and using different definitions of breeding values and their estimates in a dairy cattle breeding programme with a stochastic simulation.

Materials & methods

Simulation setting. Dairy cattle breeding scheme was simulated using AlphaSimR R-package (Gaynor *et al.*, 2020). We simulated nuclear-genome and mito-genomo independently. To generate nuclear-genome chromosomal haplotypes, we used the option 'CATTLE' in 'AlphaSimR' with 10 diploid chromosomes (to reduce computation time) with 10^8 base pairs each, mutation rate of 2.5×10^{-8}, recombination rate 1×10^{-8}, and effective population size as in MacLeod *et al.* (2013). We chose 1000 loci per chromosome as SNP markers and another 1000 as QTL. For the mito-genome haplotypes we considered 1 haploid chromosome with 16,202 base pairs, mutation rate of 2.5×10^{-7}, no recombination, and effective population size of 1000 to obtain a level of diversity that matched values found in literature (Brajković, 2019; Xia *et al.*, 2019; Dorji *et al.*, 2021). We chose all polymorphic loci (on average 1084 across 10 replicates) in mito-genome as

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_301, © G.M. Fortuna *et al.* 2022

SNP markers and as QTL at the same time. We also evaluated a setting where only one locus was a QTL. Both simulated haplotypes were randomly sampled giving rise to nuclear-genomes and mito-genomes of founding individuals. The nuclear-genomes were passed between generation with recombination and combining two parental nuclear-genomes, while mito-genomes were passed only from mothers to their progeny without recombination. We defined one polygenic trait with heritability of 0.3, partitioned between nuclear-genome ($\sigma^2_{a_n} = 0.25$) and mito-genome ($\sigma^2_{a_m} = 0.05$) components in the base population with allele substitution effects of the QTL sampled from a Gaussian distribution that gave rise to targeted genetic variances. The trait was expressed only in cows for each lactation and generated as:

$$y_{ij} = \mu_i + a_{n,j} + a_{m,j} + p_j + e_{ij} \quad (1)$$

where y_{ij} is the phenotype of cow j in lactation i, μ_i is the population mean for lactation i, $a_{n,j}$, the value of nuclear-genome for animal j, $a_{m,j}$ the value of mito-genome for animal j, which was the same as for its mother, maternal grand-mother, etc., p_j the permanent environment effect of animal j ~N(0,0.10), and e_{ij} is the environmental effect ~ N(0,0.60).

Genetic evaluation. We analysed the phenotype data using the generative model (1) with pedigree- and genome-based information accounting for the mito-genome or not. In the pedigree-based model we assumed a_n ~ N(0, A0.25) with A being pedigree relationship matrix for nuclear-genome and a_m ~ N(0, I0.05) with I being an identity matrix of dimension equal to the number of different maternal founder lineages. In the genomic-based model we assumed a_n ~ N(0, H_n0.25) with H_n being the 'single-step' joint pedigree and genomic (SNP marker) relationship matrix for nuclear-genome (Aguilar *et al.*, 2010) and a_m ~ N(0, G_m0.05) with G_m being genomic (SNP marker) relationship matrix for mito-genome with dimension equal to the number of distinct mito-genomes, n_m. We calculated it as $G_m = MM^T / k$, where M is an $n_m \times n_s$ matrix of 0 (ancestral allele) and 1 (mutation) for n_m mito-genomes and its n_s polymorphic loci (we assume we know all these loci by sequencing mtDNA within pedigrees), $k = \sum_{l=1}^{n_s} p_l(1 - p_l)$, and p_l is the frequency of mutations. For both models we firstly assumed that p ~ N(0, I0.10) and e ~ N(0, I0.60) during a burn-in phase. Following we estimated the components and used the new estimates as fixed parameters for the remainder of simulation. We fitted the models with the BLUPF90 suite (Misztal *et al.*, 2018).

Breeding value definition. We defined breeding value of individual j in two ways: (1) as nuclear breeding value $a_{n,j}$; or (2) as the sum of nuclear and mitochondrial breeding value, $a_{n,j} + a_{m,j}$. The definition (2) is the correct definition for females because they transmit both nuclear-genome and mito-genome to the next generation. However, males potentially express the effect of mito-genome they inherited from their mother, but they only transmit their nuclear genome to the next generation. Hence, for males, the definition (1) is the correct definition.

Breeding scenarios. We evaluated sixteen scenarios driven by three factors across 10 replicates. The first factor was a breeding scheme: (1) progeny testing-based selection (PT); and (2) genomic selection (GS). The second factor was a modelling and selection scenario: (1) using a model without mito-genome effect and selecting both females and males on their nuclear breeding values (STANDARD); (2) using a model with mito-genome effect and selecting both females and males on their nuclear breeding values (BASELINE); (3) using a model with mito-genome effect and selecting both females and males on their nuclear plus mito breeding values (EXTREME); and (4) using a model with mito-genome effect and selecting females on their nuclear plus mito breeding values and selecting males on their nuclear breeding values (OPTIMAL). The third factor was assuming that all or just one polymorphic locus in mito-genome is a QTL. We evaluated all the scenarios with: (1) accuracy of genetic evaluation (as correlation between true breeding value and

estimated breeding value) in scenarios STANDARD vs BASELINE and (2) genetic gain in nuclear breeding value over 20 years in scenarios STANDARD, BASELINE, EXTREME, and OPTIMAL.

Results

We present the results separately for the following five categories: (1) heifers; (2) 1st lactation cows (cows1); (3) cows with 2 to 5 lactation records (cows2-5); (4) young bulls, males without progeny; and (5) proven bulls, males that have undertaken a progeny test or have been used as sires. For the GS scenarios, the female categories were expanded to show the difference between genotyped and non-genotyped animals. We present only the scenario where all polymorphic loci in mito-genome were QTL (the other scenario was qualitatively the same).

The accuracy of mito breeding values in both PT and GS was close to one for all animal categories (results not shown) which is expected given the lack of recombination. When evaluating the accuracy of the nuclear breeding values (Table 1), accounting for mito-genome variation increased accuracy from 0.01 to 0.04, depending on the animal category. In the PT scenario, the highest increase in accuracy was observed for 'Cows2-5' (+0.04). In the GS scenario, the 'Cows2-5 (not geno.)' was the category showing the best improvement (+0.04), followed by 'Young-bulls (geno.)' (+0.03).

The results presented in Figure 1 show that considering mitochondrial effect when estimating breeding values can increase genetic gain when both males and females are selected on their nuclear breeding values (BASELINE scenarios). When using the OPTIMUM scenario, results are comparable to that observed for the STANDARD scenario. Lastly, the EXTREME scenario seems to induce less gain than all other scenarios in both genomic and progeny-testing settings.

Discussion

Dairy breeders are basing their genetic evaluation and selection solely on nuclear breeding values. Gibson *et al.* (1997), however, defined genetic merit as the sum of nuclear and cytoplasmatic components. Southwood *et al.* (1989) performed a simulation to evaluate estimation of additive maternal and cytoplasmic variance. They found inflated estimates of additive genetic variance when ignoring cytoplasmic variance or both

Table 1. Accuracy of nuclear breeding values for different animal categories in the model with nuclear-genome (Model N) or nuclear-genome and mito-genome (Model N&M) with the pedigree-based model in progeny testing-based selection (PT) or with genomic-based model in genomic selection (GS).

Scenario	Category	Model N	Model N&M	Difference
PT	Heifers	0.47±0.02	0.48±0.01	0.02±0.02
	Cows1	0.61±0.02	0.63±0.01	0.03±0.02
	Cows2-5	0.54±0.04	0.58±0.01	0.04±0.04
	Young bulls	0.37±0.03	0.38±0.03	0.01±0.02
	Proven bulls	0.74±0.01	0.76±0.01	0.02±0.01
GS	Heifers (non-geno.)	0.43±0.01	0.44±0.01	0.01±0.01
	Heifers (geno.)	0.72±0.03	0.75±0.01	0.02±0.02
	Cows1 (non-geno.)	0.58±0.02	0.60±0.00	0.02±0.02
	Cows1 (geno.)	0.76±0.02	0.78±0.01	0.02±0.02
	Cows2-5 (non geno.)	0.52±0.03	0.56±0.00	0.04±0.04
	Cows2-5 (geno.)	0.71±0.03	0.73±0.01	0.02±0.03
	Young bulls (geno.)	0.69±0.02	0.72±0.02	0.03±0.03
	Proven bulls (geno.)	0.78±0.01	0.79±0.01	0.01±0.01

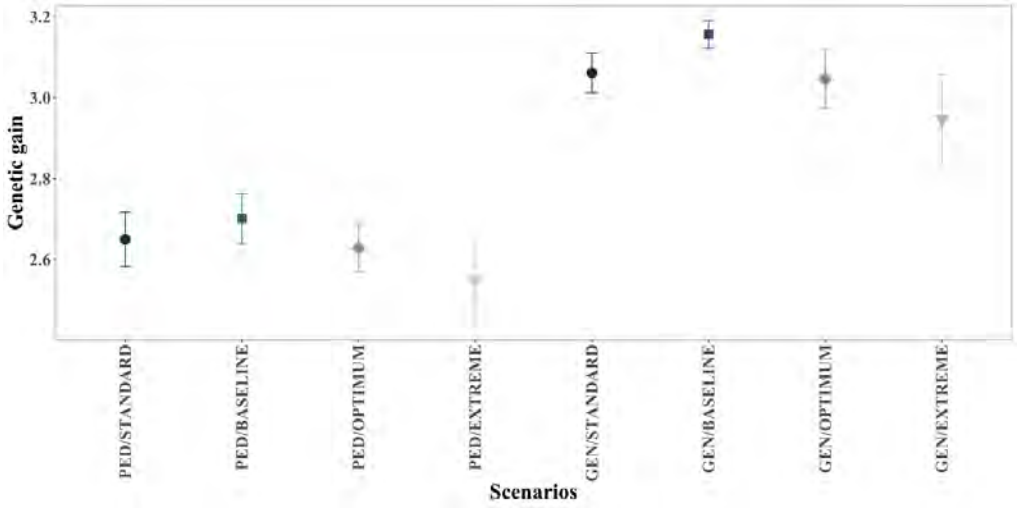


Figure 1. Average genetic gain and 95% confidence interval over 20 years of selection with progeny testing-based selection (PT) and genomic selection (GS) and four different scenarios of modelling and breeding value definition (STANDARD, BASELINE, EXTREME, and OPTIMAL).

the cytoplasmic and additive maternal effects in the analysis. Our study complemented this past work by extending it with genomic information in estimating breeding values and by clarifying the definition of breeding values in females and males when both nuclear- and mito-genome affect a trait.

References

Aguilar I., Misztal I., Johnson D.L., Legarra A., Tsuruta S. *et al.* (2010) J. Dairy Sci. 93(2):743-752. https://doi.org/10.3168/jds.2009-2730.

Boettcher P.J., Kuhn M.T and Freeman A.E. (1996) J. Dairy Sci. 79(4):663-675. https://doi.org/10.3168/jds.S0022-0302(96)76412-3.

Brajković V. Impact of mitogenome on milk traits in cattle. Available at: https://core.ac.uk/download/pdf/222819316.pdf (Accessed: 15 April 2021).

Dorji J., Jagt C.V., Chamerlain A., Cocks B., MacLeod I. and Daetwyler H. (2021) BioRxiv. https://doi.org/10.21203/rs.3.rs-957964/v1.

Gaynor R.C., Gorjanc G. and Hickey J.M. (2020) G3 11(2). https://doi.org/10.1093/g3journal/jkaa017.

Gibson J.P., Freeman A.E. and Boettcher P.J. (1997) Livest. Prod. Sci. 47(2):115–124. https://doi.org/10.1016/S0301-6226(96)00023-1.

MacLeod I.M., Larkin D.M., Lewin H.A., Hayes B.J., Goddard M.E. (2013) Mol. Biol. 30(9):2209-2223. https://doi.org/10.1093/molbev/mst125.

Misztal I., Tsuruta S., Lourenco D. A. L., Masuda Y., Aguilar I. *et al.* Manual for BLUPF90 family programs. Available at: http://nce.ads.uga.edu/wiki/doku.php?id=documentation

Southwood O.I., Kennedy B.W., Meyer K. and Gibson J.P. (1989) J. Dairy Sci. 72(11):3006-3012. https://doi.org/10.3168/jds.S0022-0302(89)79453-4.

Xia, X., Qu, K., Jia, P., Chen, Q. *et al.* (2019) Animals, 9(9):641. https://doi.org/10.3390/ani9090641.

302. Using residual regressions to quantify and map signal leakage in genomic prediction

B.D. Valente[1*], G. de los Campos[2,3,4], A. Grueneberg[2], C.Y. Chen[1], R. Ros-Freixedes[5,6] and W.O. Herring[1]

[1]The Pig Improvement Company, Genus plc, Hendersonville, TN 37075, USA; [2]Michigan State University, Department of Epidemiology & Biostatistics, East Lansing, MI 48824, USA; [3]Department of Statistics & Probability, East Lansing, MI 48824-1027, USA; [4]Michigan State University, Institute for Quantitative Health Science and Engineering, East Lansing MI 48824, USA; [5]The Roslin Institute and Royal (Dick) School of Veterinary Studies, The University of Edinburgh, Easter Bush, Midlothian, Scotland, United Kingdom; [6]Departament de Ciència Animal, Universitat de Lleida, Agrotecnio-CERCA Center, Lleida, Spain; bruno.valente@genusplc.com

Abstract

Prediction accuracy is commonly used to evaluate genomic prediction models, but it can be insufficient to describe how different models perform in terms of capturing (or missing) genetic signals. Here we propose using residual single-marker regression analyses to identify loci harbouring genetic signals that were not captured by a model – we refer to this phenomenon as signal leakage. The proposed approach was applied to average daily gain in pigs. Different prediction models were compared: pedigree-BLUP, ssGBLUP, Bayesian Ridge Regression (BRR) and BayesC. Genomic scans were based on either medium or ultra-high-density SNP panels. Signal leakage was strong and widespread in pedigree-BLUP. Including genomic data in ssGBLUP reduces the extent of signal leakage, but regional peaks of missed signal remain. BayesC performed better than BRR, indicating that over-shrinkage contributes to signal leakage. Residual regressions with ultra-high-density panel suggested that imperfect LD between SNPs and causal variants also contributes to signal leakage.

Introduction

Multiple factors (including the statistical model used, sample size, marker density, etc) can affect the ability of a genomic model to extract genetic signals from phenotypic data. Models' performances are often evaluated based on their ability to predict phenotypic values in testing data. However, because multiple factors can affect prediction accuracy, predictive ability alone can be insufficient metric for characterizing how models behave in terms of capturing genetic signals. For example, shrinkage methods may be better at capturing genetic signals in regions of very high LD and in those where there are multiple variants with small effects, while variable selection methods may be better at capturing signals in regions with large-effect variants. Likewise, increasing marker density with the use of sequence data may be more relevant in some regions than others.

Our research was originally motivated by the goal of identifying regions of the genome holding genetic signals that are missed by a model (hereinafter referred to as signal leakage). To this end, we propose using single-marker regression analysis applied to the residuals of this genetic model. The intuition behind the approach is simple: strong associations between residuals and SNP genotypes are suggestive of possible genetic signals missed by the predictor used to derive the residuals. These residual GWA can be applied to the same SNP panel used to derive predictions and residuals or to a different one. We use different panels and show that signals captured by one panel (e.g. imputed sequence-level SNP data) can be missed by a lower density panel. The residual regression approach can also provide insights into the effect of shrinkage on signal leakage.

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_302, © B.D. Valente *et al.* 2022

Materials & methods

Data sets. A total of 28,156 pigs from a single line with genotypes and phenotype records for average daily gain (ADG) were considered in this study. Phenotypes were pre-corrected for contemporary group and random litter effects. Pedigree information for 42,946 pigs was available for models that use this type of data. This study used a medium density (MD) SNP panel and an ultra-high density (UHD) SNP panel. The MD panel included 41,205 SNPs from a commercial marker array (GeneSeek, Lincoln, NE, USA) after quality control and imputation. The UHD panel was based on sequence-level SNP calling. A subset of the full population was sequenced for different depth levels. Variants called were then imputed for the remaining population that was genotyped for MD panels. Details of this approach are described by Ros Freixedes *et al.* (2020). The UHD panel contained ~18M SNPs after filters.

Analysis. We fitted five types of models to the aforementioned data: Pedigree BLUP (PBLUP), single-step Genomic BLUP (ssGBLUP), Bayesian Ridge Regression (BRR), and BayesC (a variable selection model with a point of mass at zero and a Gaussian slab). In all cases, the genotypes used to fit these models were those in the MD SNP panel. We fit ssGBLUP considering genotypes for either 50 or 100% of the animals (ssGBLUP_50 and ssGBLUP_100, respectively). The set of pigs with genotypes ignored in ssGBLUP_50 was randomly selected. The scenario ssGBLUP_50 provides a transition between PBLUP and ssGBLUP_100. The ssGBLUP (Legarra *et al.*, 2009) and PBLUP analyses were performed with the BLUPF90 software family (Misztal *et al.*, 2014). BRR and BayesC were fitted using the BGLR package in R (Habier *et al.*, 2011; Perez and de los Campos, 2014). Residuals were derived from the fitted models (i.e. phenotypes minus prediction). Subsequently we conducted marginal association tests between these residuals and genotypes of the MD and UHD panel, one SNP at a time. We determined significance using FDR<0.01. The association tests were conducted using the BGData package (Grueneberg and de los Campos, 2019).

Results

Results for all models under different panels are depicted in Figure 1. There is strong widespread leakage for PBLUP. Variants showing FDR<0.01 are spread across the genome, with very few regions devoid of them, especially in the UHD panel. These results reflect the amount of genetic signal missed by not including genomic information in the model.

The inclusion of genotype data under the ssGBLUP model strongly changes the leakage landscape, even though the pedigree still influences the model fit and predictors. Two major aspects of that change are the improved fitting of genome-wide (i.e. non-local) associations that were missed by PBLUP and the reduction of the strength of locally concentrated leakage peaks. That change is gradual with the inclusion of more genotypes in the data set. Some strong leakage regions in PBLUP (e.g. in chromosome 1) were solved by ssGBLUP under ssGBLUP_100, regardless of the panel used to evaluate it. This indicates that local modelling with differential prior weights might not be required to solve strong signals of local leakage. However, a couple of clear leakage peaks remain even when 100% of genotypes are included. One of them (in chromosome 2) is only detected using UHD. The strongest remaining leakage signal for ADG is in chromosome 17. Although its detected magnitude is stronger using UHD in the scan, it is still clear under MD, even if this same panel was used to fit the ssGBLUP. The association pattern resulting from the residuals of BRR is very similar to the ssGBLUP_100 scenario, with the latter showing the same peaks for both traits, albeit with slightly extra shrinkage.

The use of variable selection, as in BayesC, fully solves the strongest remaining leakage point in chromosome 17. This suggests that this region involves associations that are too strong to be allowed under Gaussian prior distributions but are efficiently accommodated by variable selection (and possibly other thick-tailed) priors. Under BayesC no leakage was detected with the MD panels used to build the predictors. However, scans

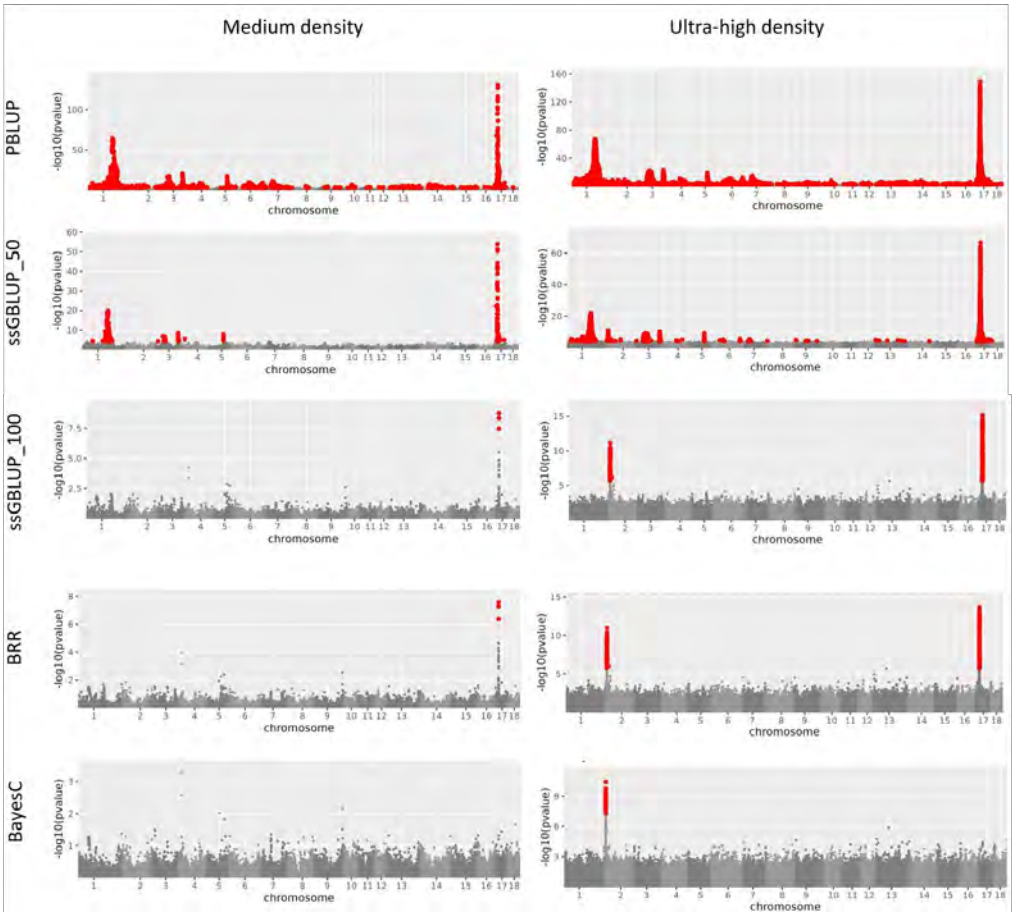


Figure 1. Manhattan plots depicting genomic signal leakage by model and SNP panel, with significant SNPs highlighted in red (FDR<0.01).

with the UHD panels indicated a leakage spot in chromosome 2 missed by the MD panel, suggesting that imperfect LD between the variants of the MD panel and causal variants may be generating signal leakage. Adding genomics and allowing thicker tails is not only insufficient to solve the leakage, but its magnitude is also hardly affected, indicating that this peak had a different nature from the one in chromosome 17.

Discussion and implications

Metrics typically used to compare genomic models are often based on an overall score, such as prediction correlation, AIC, BIC, etc. Such genome-wide summaries are affected by multiple competing factors and thus may not be enough to characterize how a model performs at capturing the genetic signals present in genomes for a trait. Here we present a complementary approach for evaluating and comparing models that allows characterizing genomic landscapes of signal leakage from a genetic predictor. Leakage can be widespread or concentrated in specific genomic regions. Peaks of leakage can result from different sources (e.g. over-shrinkage of marker effects, low LD between SNP panel used to build the predictor and a high association genomic region).

Relative to other methods to evaluate and compare models, the present approach potentially provides better insights about the existence or not of room for improvements, as well as directions on how problems detected using residual regressions can be addressed. For example, it is common to attempt to improve genomic predictors performance based on commercial SNP panels by adding sequence variants. Two common choices are either increasing the overall panel density or including SNPs from selected regions. The region selection involved in the latter is typically based on statistical information (e.g. GWAS), biological information (e.g. genes that are differentially expressed or assigned to relevant biological pathways), or a combination of both. However, these approaches are blind to the pattern of missed signals of the predictor that one wants to improve. For example, there is limited potential in adding variants pertaining to a few relevant regions if leakage is spread genome-wide with little regional concentration. Alternatively, even if leakage is more local, using GWAS or biological information may suggest the inclusion of variants from regions from which signals are already well captured by current predictors. The approach proposed here could describe the leakage pattern of a specific combination of panel and model and be used as a criterion to evaluate if/which genomic regions would present the best potential to harbour sequence variants that would improve this particular predictor.

References

Grueneberg A., de los Campos G. (2019) G3-GENES GENOM GENET, 9(5):1377–1383. https://doi.org/10.1534/g3.119.400018

Habier D., Fernando R. L., Kizilkaya K., and Garrick, D. J. (2011). BMC Bioinform, 12:186. https://doi.org/10.1186/1471-2105-12-186

Legarra A., Aguilar I. and Misztal I. (2009) J. Dairy Sci. 92:4656–4663. https://doi.org/10.3168/jds.2009-2061

Misztal I.S., Tsuruta, Lourenco D.A.L., Masuda Y., Aguilar I., *et al.* (2014) Manual for BLUPF90 family of programs. Available at http://nce.ads.uga.edu/wiki/lib/exe/fetch.php?media=blupf90_all2.pdf.

Perez P, de los Campos G (2014). Genetics, 198(2):483–495. https://doi.org/10.1534/genetics.114.164442

Ros-Freixedes R., Whalen A., Chen C. Y., Gorjanc, G. Herring, W. O., et al (2020). Genet. Sel. Evol. 52(1):17. https://doi.org/10.1186/s12711-020-00536-8

303. Capitalizing on complex annotations in Bayesian genomic prediction for a backcross population of growing pigs

F. Mollandin[1*], H. Gilbert[2], P. Croiseau[1] and A. Rau[1,3]

[1]INRAE, AgroParisTech, GABI, Université Paris-Saclay, 78350 Jouy-en-Josas, France; [2]GenPhySE, Université de Toulouse, INRAE, ENVT, 31320 Castanet Tolosan, France; [3]BioEcoAgro Joint Research Unit, INRAE, Université de Liège, Université de Lille, Université de Picardie Jules Verne, 80203 Estrées-Mons, France; fanny.mollandin@inrae.fr

Abstract

Prior biological information has the potential to guide and inform genomic prediction models, but the BayesRC approach is currently limited to the use of disjoint categorizations of genetic markers. We propose two novel Bayesian approaches to model cumulative (BayesRC+) or preferential (BayesRCπ) contributions of multiple biological categories for multi-annotated SNPs. We illustrate the performance of these approaches on data from a backcross population of growing pigs in conjunction with several different sets of annotations related to multiple production traits constructed using the PigQTLdb. On the two traits predicted, ADG and BFT, we observed improved prediction quality on ADG (up to 1.7-gain point) with both BayesRCpi and BayesRC+, and suitable annotation set.

Introduction

In plant and animal breeding, genomic prediction models have been widely developed and deployed in recent years to predict polygenic traits using genetic variants, typically single nucleotide polymorphisms (SNP). An interesting and potentially useful approach to improve upon existing genomic prediction models is to combine the use of genotype and phenotype data with prior biological information to better guide models. Most routinely used genomic prediction models are based on linear models, including notably genomic best linear unbiased prediction. Another family of models, known as the Bayesian alphabet (Habier *et al.*, 2011), uses a flexible set of assumptions about how individual SNPs contribute to the overall genomic variance. Among these, BayesR (Erbe *et al.*, 2012) assumes SNP effects arise from one of four groups (null, small, medium, or large variance) and has been shown to perform well for both prediction and quantitative trait loci (QTL) mapping (Moser *et al.*, 2015; Mollandin *et al.*, 2021). BayesRC extends BayesR to further incorporate prior biological information in the form of disjoint annotation categories (MacLeod *et al.*, 2016), but SNPs can only be assigned to a single annotation category. There thus remains a need for genomic prediction models able to capitalize on annotations of greater complexity, in particular those for which SNPs may potentially be assigned to multiple categories.

In this work, we present two novel extensions to BayesRC to deal with such complex, overlapping annotations, and we illustrate their utility on data from an experimental backcross population in growing pigs. This project is part of EuroFAANG (https://eurofaang.eu), a synergy of five Horizon 2020 projects that share the common goal to discover links between genotype to phenotype in farmed animals and meet global Functional Annotation of ANimal Genomes (FAANG) objectives.

Materials & methods

Bayesian genomic prediction with complex, overlapping annotations. The general statistical model for genomic prediction can be defined as:

$$\boldsymbol{y} = \boldsymbol{\mu} + \boldsymbol{X\beta} + \boldsymbol{e}, \quad e_i \sim N(0, \sigma_e^2) \tag{1}$$

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_303, © F. Mollandin *et al.* 2022

where $\boldsymbol{y}$ is a vector of phenotypes, μ an intercept, β the vector of SNP effects, X the centered and scaled marker matrix, and σ_e^2 the variance of the residuals e. We further assume that C = ($C_{i,j}$) denotes annotation categories, such that $C_{i,j}$ = 1 if SNP i is included in category j and 0 otherwise.

Using the four-component mixture of BayesR as a base, we propose two alternative models to account for overlapping annotations (where $\Sigma_j\, C_{i,j} > 1$ for some i). The first, BayesRCπ, defines a mixture-of-mixtures prior for SNP effects to assign multi-annotated markers to the single annotation category that maximizes its conditional likelihood:

$$\beta_i \sim \sum_{j \in C_{i,j}=1} p_{i,j}\left(\pi_{1,j}\delta(0) + \pi_{2,j}N(0, 10^{-4}\sigma_g^2) + \pi_{3,j}N(0, 10^{-3}\sigma_g^2) + \pi_{4,j}N(0, 10^{-2}\sigma_g^2)\right) \quad (2)$$

such that $\delta(0)$ represents the dirac function at 0, $\Sigma_k\, \pi_{k,j}$ for all annotations j, σ_g^2 the total additive genetic variance, and $p_{i,j}$ the annotation mixing parameter with $\Sigma_j\, p_{i,j} = 1$ for all i. The second, BayesRC+, instead defines a cumulative mixture prior across categories for SNP effects:

$$\beta_i \sim \sum_{j \in C_{i,j}=1}\left(\pi_{1,j}\delta(0) + \pi_{2,j}N(0, 10^{-4}\sigma_g^2) + \pi_{3,j}N(0, 10^{-3}\sigma_g^2) + \pi_{4,j}N(0, 10^{-2}\sigma_g^2)\right) \quad (3)$$

In both models, all mixing proportions are assumed to follow flat Dirichlet priors and σ_g^2 an inverse X^2 prior. A Gibbs sampler is used for inference as posterior distributions are not tractable. Both BayesRCπ and BayesRC+ have been implemented in the BayesRCO software in Fortran; additional details can be found in the User's Guide (https://github.com/fmollandin/BayesRCO).

Genotype and phenotype data from a backcross pig population. A backcross (BC) population between Large White (LW; 3/4) and Creole (CR;1/4) pigs was established as previously described (Gourdine *et al.*, 2019). BC (n=1,297 from 130 LW sows) growing pigs raised in two environments (tropical and temperate) were related via genetically related sows sired with the same 10 F1 × CR LW boars. A common trait recording protocol was used in the two environments for phenotypic data. Phenotypes were pre-corrected for age, sex, and farm; we focus here on measures at 23 weeks for backfat thickness (BFT) and average daily weight gain (ADG). Animals were genotyped using the Illumina Porcine 60k BeadChip array; markers with minor allele frequencies greater than 0.01 were retained for the analysis (corresponding to 46,908 and 46,881 markers for ADG and BFT, respectively). To establish the potential impact of our models on prediction accuracy, we used a sibling-structured 10-fold cross validation procedure. For the descendants from each sire in turn, we calculated the correlation between their observed corrected phenotypes and those predicted from models constructed on the descendants of the remaining 9 sires; validation correlations were averaged across the ten folds.

PigQTLdb annotations. Animal QTLdb (https://www.animalgenome.org/QTLdb) groups together curated results from genotype-phenotype association studies in several livestock species (Hu *et al.*, 2021). Cross-experiment QTL data from PigQTLdb (Release 45; SS11.1) for traits relevant to pig production were downloaded for eleven trait sub-hierarchy categories (anatomy, behavioral, blood parameters, conformation, fatness, fatty acid content, feed conversion, fowth, immune capacity, litter traits, reproductive organs). An additional 'other' category was created for markers not included in PigQTLdb. Genotyped markers in our data were subsequently assigned to one or more annotation categories using three different strategies: (1) using the position of known PigQTLdb markers; (2) using the extended position of known PigQTLdb markers, including the nearest up- and downstream neighbors ('extended PigQTLdb'); and (3) using the extended position of known PigQTLdb markers as before, where neighboring markers were allowed ambiguous assignment to both trait-specific and 'other' categories ('fuzzy extended PigQTLdb').

In the three annotation construction strategies, 1.3, 4.9 and 17.7% of markers were respectively assigned to two or more categories.

Results

We compared the prediction accuracy of BayesRCπ and BayesRC+ to that of BayesR (ignoring annotation categories) and BayesRC (where a single category is allowed per marker). For the latter, multi-annotated SNPs were randomly assigned to a single category. We notably observed different trends for the two traits (Table 1). For ADG, we remark a loss in prediction accuracy compared to BayesR for all annotation-based models with straightforward pigQTLdb annotations; however, extending these annotations to include neighboring markers (extended and fuzzy extended pigQTLdb) led to improvements in prediction quality, with a 1.7-point gain in correlation for BayesRCπ with extended annotations. On the other hand, for BFT the use of annotations, regardless of how they are constructed, did not appear to lead to a marked improvement in prediction. This suggests that categorizations constructed from PigQTLdb contribute little pertinent information for the genomic prediction of BFT in our data, or that environment-dependent categories should be added to the model to account for the significant GxE affecting this trait (Gourdine *et al.*, 2019).

Discussion

In this work we have proposed two new approaches, BayesRCπ and BayesRC+, to fully capitalize on complex, overlapping annotations in genomic prediction. Both methods showed promise for incorporating partially overlapping categories from pigQTLdb in genomic prediction for a growing pig population, although a gain in predictive accuracy was observed for only one (ADG) of the two traits considered here. We also compared three strategies for constructing prior biological categories by extending pigQTLdb annotations in various ways to include neighboring markers, which has the potential to better exploit linkage disequilibrium around relevant markers. Taken together, these results suggest that the incorporation of complex annotations can lead to modest gains in prediction performance in some cases, even for moderate marker density SNP chips, but such gains depend strongly on the choice and construction of annotations and are unlikely to be universal across traits.

This work is part of the GENE-SWitCH project that has received funding from the European Union's Horizon 2020 Research and Innovation Programme under grant agreement no. 817998. The financial support of the French National Agency of Research (ANR PigHeaT, ANR-12-ADAP-0015) is also gratefully acknowledged.

Table 1. Validation correlation for two traits in pig data for BayesRCπ and BayesRC+ with different annotation strategies, as compared to BayesR and BayesRC.

Method	Annotations	ADG; mean (SD)	BFT; mean (SD)
BayesR	-	0.213 (±0.081)	0.265 (±0.161)
BayesRC	PigQTLdb (random)	0.200 (±0.105)	0.265 (±0.159)
	Extended pigQTLdb (random)	0.225 (±0.098)	0.258 (±0.157)
BayesRCπ	PigQTLdb	0.200 (±0.100)	0.266 (±0.157)
	Extended pigQTLdb	0.229 (±0.095)	0.254 (±0.162)
	Fuzzy extended pigQTLdb	0.226 (±0.096)	0.262 (±0.159)
BayesRC+	PigQTLdb	0.207 (±0.097)	0.273 (±0.163)
	Extended pigQTLdb	0.227 (±0.095)	0.271 (±0.158)

References

Erbe, M., Hayes, B.J., Matukumalli, L.K., Goswami, S., Bowman, P.J., *et al.* (2012) Journal of Dairy Science 95(7): 4114–29. https://doi.org/10.3168/jds.2011-5019

Gourdine, J.-L., Riquet, J., Rosé, R. Poullet, N., Giorgi, M. *et al.* (2019) Journal of Animal Science 97(9): 3699–3713. https://doi.org/10.1093/jas/skz245

Habier, D., Fernando, R.L., Kizilkaya, K., and Garrick, D.J. (2011). BMC Bioinformatics 12(1): 186. https://doi.org/10.1186/1471-2105-12-186

Hu, Z.-L., Park, C.A., and Reecy, J.M. (2021) Nucleic Acids Research, gkab1116. https://doi.org/10.1093/nar/gkab1116

MacLeod, I.M., Bowman, P.J., Vander Jagt, C.J., Haile-Mariam, M., Kemper, K.E., *et al.* (2016) BMC Genomics. 17(1): 144. https://doi.org/10.1186/s12864-016-2443-6

Mollandin, F, Rau, A., and Croiseau, P. (2021) G3 Genes|Genomes|Genetics 11(11): jkab225. https://doi.org/10.1093/g3journal/jkab225

Moser, G., Lee, S.H., Hayes, B.J., Goddard, M.E., Wray, N.R. *et al.* (2015) PLOS Genetics 11(4): e1004969. https://doi.org/10.1371/journal.pgen.1004969

304. Predictive ability of genomic prediction in layers when including CADD scores as genome function information

M.C.A.M. Bink[1*], M.P.L. Calus[2], M.F.L. Derks[2], J. Visscher[1] and B.C. Perez[1]

[1]Hendrix Genetics, P.O. Box 114, 5830 AC Boxmeer, the Netherlands; [2]Wageningen University & Research, Animal Breeding and Genomics, P.O. Box 338, 6700 AH Wageningen, the Netherlands;
marco.bink@hendrix-genetics.com

Abstract

Accuracy of genomic prediction is key for genetic progress in breeding programs and including genome functional annotation may help. Here we report on the added value of Combined Annotation Dependent Depletion (CADD) scores to weigh SNPs in GBLUP analyses. Multiple transformations of CADD scores were considered and empirically validated on a layer dataset including 5 traits and 18K animals with 27K SNP genotypes. This initial analysis revealed that the use of (squared) CADD scores yielded marginally higher accuracy for 3 traits. We anticipate that the added value of CADD scores will increase by using a higher number of SNPs and ultimately whole genome sequence SNPs.

Introduction

Since the introduction of genomic prediction (Meuwissen *et al.*, 2001), animal breeding has been transitioning towards approaches that increasingly exploit genomic information. An emerging strategy involves weighting of SNPs that have a higher probability of being causal or being tightly linked to causal variants (Xiang *et al.*, 2021).

The combined annotation dependent depletion (CADD) (Rentzsch *et al.*, 2019) model that was developed to investigate SNPs in human populations, can score variants at coding and non-coding locations in the genome. Similar models have been deployed for livestock species. For chicken and pig, so-called cCADD and pCADD scores were created, as a tool to prioritize variants and evaluate sequences to highlight new sites of interest to explain biological functions that are relevant to animal breeding (Groβ *et al.*, 2020).

The objective of this study is to validate the added value of CADD scores to weight SNPs in genomic prediction on five traits in commercial layers. We explore multiple weighting strategies, i.e. transformations of CADD scores, in GBLUP models. This project is part of EuroFAANG (https://eurofaang.eu), a synergy of five Horizon 2020 projects that share the common goal to discover links between genotype to phenotype in farmed animals and meet global Functional Annotation of ANimal Genomes (FAANG) objectives.

Materials & methods

Chicken CADD scores. The chicken CADD scores were retrieved from Groβ *et al.* (2020). CADD is a machine learning tool originally developed to identify deleterious variants in the human genome. CADD uses a wide range of annotations to predict impact from sequence variation. Important annotations include conservation scores, gene model annotations, epigenomic data, and functional predictions. CADD combines these annotations to provide an impact score for any possible SNP variant in the genome. The higher the CADD score, the more likely the variant has impact. The scores can be used to prioritize genetic variants in genetic evaluation.

Weighting SNPs in genomic prediction. A typical genomic relationship matrix G is computed as: $G=XX^T/c$, equivalently written as: $G=XIX^T/c$, where X contains centred (and possibly scaled) SNP genotypes, c is a constant, and I is an identity matrix (VanRaden, 2007). This implies that all SNPs have

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_304, © M.C.A.M. Bink *et al.* 2022

an equal weight of 1. The computation can be generalized as: $G_w=XDX^T/c$, where D is a diagonal matrix, with weights for each SNP on the diagonal that are proportional to the variance that this SNP explains. The calc_grm program (Vandenplas and Calus 2020) was used to compute G and G_w.

Scenarios of weights. In this approach, the weight represents the *a priori* variance associated with a particular SNP. So, if one SNP has a weight of 1, and another one has a weight of 2, this means that the second SNP explains twice as much variance as the first one. The CADD scores do not represent variances and therefore three options are considered to translate CADD scores into weights, i.e. original (CADD), squared (CADDsq), and square root (CADDrt) (Figure 1).

To develop a benchmark for these scenarios, two options were considered: [A] all SNPs having equal weights (Equal); [B] random values (Random). For option [B] 10 replicates were generated to provide standard error estimates. The random values were generated by squaring random draws from $N(0,1)$. Within each scenario, all weights are scaled by dividing by the average weight, such that after scaling the mean weight is equal to 1.

Layers dataset. The dataset was provided by the layer division (ISA BV) of Hendrix Genetics and pertains to the pure line WA of the commercial layer breeding program. A total of 18,419 animals was genotyped with a 60K SNP array. These genotypes were curated on minor allele frequency (>0.05) and missing values (<10%). SNPs with unknown position or located at sex chromosomes were also discarded. After quality control, 26,841 autosomal SNPs were available for further analyses. Including ancestors of the genotyped animals yielded a pedigree of 21,060 animals. The dataset included five traits, i.e. shell breaking strength (BS), egg number (EN), egg weight (EW), hatchability (HT), and survival (SV). The number of records varied between 2,361 and 5,802 (Table 1). The dataset has been pre-corrected for all non-genetic effects to ease analyses. The pre-correction was executed using the models as used in the routine genetic evaluations. That is, the feature 'yielddev' in the MiXBLUP software (www.mixblup.eu) was used to produce yield deviations, i.e. the sum of estimated animal & residual effects. Variance components were estimated using the average information restricted maximum likelihood algorithm implemented in the AIREMLF90 software (Misztal *et al.*, 2014), and yielded heritability estimates ranging from 0.06 to 0.75 (Table 1). When omitting SNP genotypes, the pedigree-based accuracy of EBV was between 0.06 and 0.37 for the validation animals.

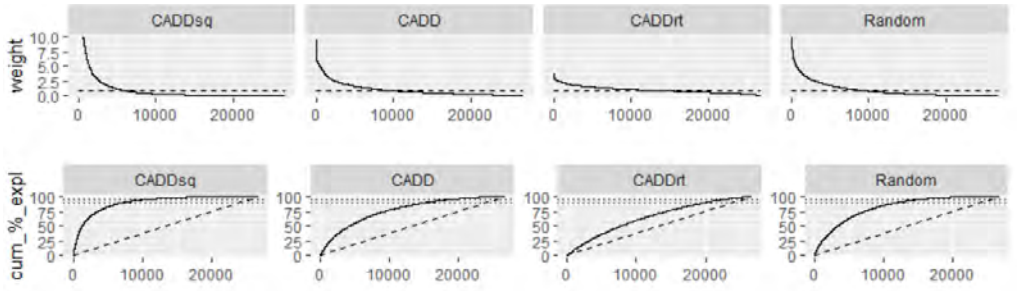


Figure 1. Distribution of weights (upper), and cumulative % of *a priori* explained genetic variances (lower) for CADD and random scores. Dashed lines pertain to scenario with equal weights, and dotted lines pertain to 90 and 95% explained variance.

Table 1. Numbers of records, heritability estimates, split dates for validation, and pedigree-based accuracy for validation animals, on five layers traits.

Trait	No. records	Heritability	CV hatch date	Accuracy$_{(ped)}$
Breaking strength (BS)	2,522	0.26	2014-11-04	0.25
Egg number (EN)	3,400	0.12	2014-11-04	0.11
Egg weight (EW)	5,802	0.75	2016-06-21	0.37
Hatchability (HT)	2,361	0.30	2016-06-21	0.20
Survival (SV)	3,536	0.06	2015-03-24	0.06

Validation set & accuracy. In animal breeding programs the main interest is in accurate breeding values on selection candidates. The forward validation scheme aligns best with this interest. Here, the dataset was split into a training set of 75% oldest animals and a validation set of 25% youngest animals. The patterns and numbers of records on the five traits differ substantially and therefore different dates were used to assign animals to subsets for each trait (Table 1). The validation set allowed a dissection in birthyears to assess impact of connectedness between training and validation animals with youngest animals least connected. Measure for accuracy was the correlation between estimated breeding values and pre-corrected phenotypes (on validation set). All data manipulation was done with dedicated R scripts (R Core Team, 2020).

Results

The weighting scenarios on CADD scores yielded accuracy estimates that were similar to those from Equal scenario (Figure 2) with slightly higher values for 3 traits (BS, HT, and SV). For these traits, the squared CADD scores yielded highest accuracies. Note that for the hatchability trait accuracies were not different from the Random scenario. For other traits the Random scenario always yielded the lowest accuracies. As expected, accuracies dropped for younger animals in the validation set (Figure 3) and these patterns were very similar for all scenarios (some not shown). The rate of drop in accuracy with age was trait dependent.

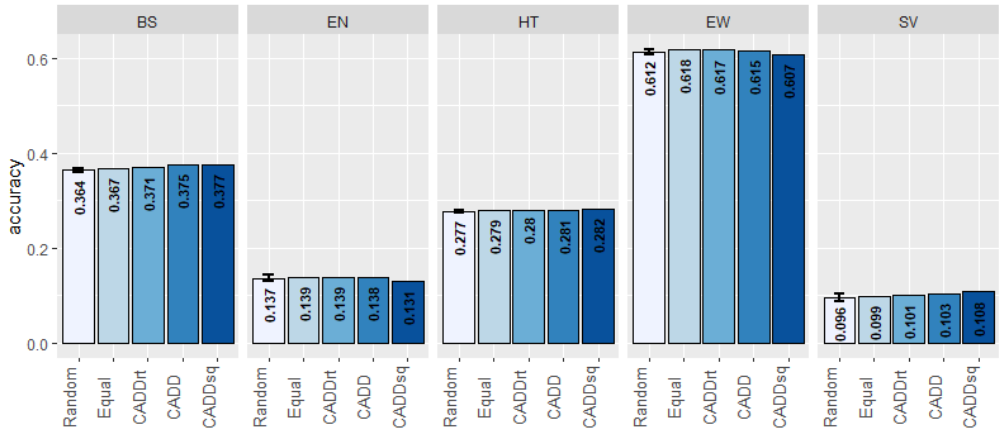


Figure 2. Accuracy of genomic prediction on validation animals for five traits and five weight scenarios (standard error bars are included for Random).

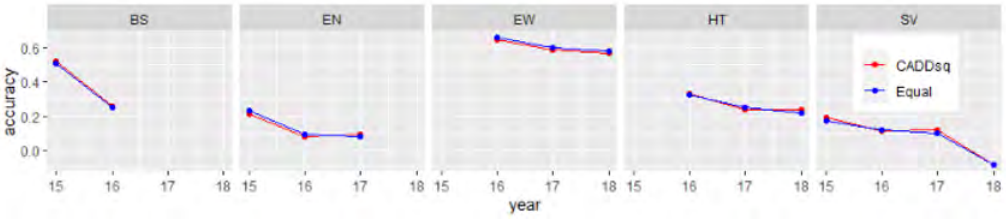


Figure 3. Accuracy estimates versus birthyear of validation animals.

Discussion

We reported on the use of CADD scores to weight SNPs in genomic prediction when using data from a 60K SNP array. These results did not show a consistent added value in predictive ability when using CADD scores to weight SNPs in GBLUP models for genomic prediction. Among the weighting scenarios some trends were observed. The square root CADD scores yielded accuracy closest to the Equal scenario while the squared CADD scores were always most different. This is consistent with the patterns of cumulative percentages explained genetic variance (Figure 1), where the square root CADD scores are most close to the Equal scenario. The trait-dependency of results points to potential differences in trait architecture with some traits highly polygenic and others pre-dominantly driven by fewer genes. This deserves further investigation. With whole genome sequence SNPs, it is known to be important to use a subset for genomic prediction. CADD enables selecting variants with impact, independent from the training data.

Acknowledgements

The GENE-SWitCH project received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement no. 817998.

References

Groβ C., Bortoluzzi C., de Ridder D., Megens H.J., Groenen M.A.M. *et al.* (2020) PLoS Genet 1:22. https://doi.org/10.1371/journal.pgen.1009027

Meuwissen T.H.E., Hayes B.J., and Goddard M.E. (2001) Genetics. 157(4):1819–29. https://doi.org/10.1093/genetics/157.4.1819

Misztal I., Legarra A., and Aguilar I. (2014) J Dairy Sci. 97(6):3943–3952. https://doi.org/10.3168/jds.2013-7752

Rentzsch P., Witten D., Cooper G.M., Shendure J., and Kircher M. (2019) Nucleic Acids Res. 47(D1):D886–D894. https://doi.org/10.1093/nar/gky1016

R Core Team (2020). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. (www.R-project.org)

VanRaden P.M., (2008) J Dairy Sci 91(11):4414-4423. https://doi.org/10.3168/jds.2007-0980

Vandenplas, J., and Calus M.P.L. (2020). Calc_grm – a program to compute pedigree, genomic, and combined relationship matrices. WUR-ABG, Wageningen Livestock Research.

Xiang R., MacLeod I.M., Daetwyler H.D., de Jong G., O'Conner E. *et al.* (2021) Nature Commun 12(1):860. https://doi.org/10.1038/s41467-021-21001-0

305. Across-sex genomic prediction using pre-selected SNP sets in cattle

B.S. Olasege[1,2], M.S. Tahir[1,2], L.R. Porto-neto[2], B.J. Hayes[3] and M.R.S Fortes[1,3*]

[1]The University of Queensland, School of Chemistry and Molecular Biosciences, St Lucia Campus, Brisbane, QLD, 4072, Australia; [2]CSIRO Agriculture and Food, St Lucia, QLD, 4067, Australia; [3]The University of Queensland, Queensland Alliance for Agriculture and Food Innovation (QAAFI), St Lucia Campus Brisbane, QLD, 4072, Australia; m.fortes@uq.edu.au

Abstract

Female fertility traits are often difficult and expensive to measure. Male fertility traits, however, are sometimes cheaper and easier to measure and are genetically correlated with female fertility. This study investigated the accuracy of genomic prediction using the SNP effects from male traits to estimate genomic breeding values for female traits. Three sets of genetically correlated male vs female traits in Brahman cattle were used, with varying trait correlation from low to high. A recently developed methodology termed correlation scan was used to interrogate the genome and identify SNPs that underpin the genomic correlations between male and female fertility. Some of the identified SNPs can be classified as 'correlation-drivers' for the genomic correlations. When these significant correlation-drivers were used in genomic predictions, we observed an increased accuracy, by as much as ~30%, compared to using the whole high-density SNP chip.

Introduction

Female fertility is a key driver of profitability of beef enterprise in tropical regions of the world (Taylor and Rudder 1986). However, most of the traits that define female fertility are often difficult and expensive to measure on commercial herds and are mostly available late in life (Hayes *et al.* 2018). These challenges make female fertility traits difficult to select for and make genetic progress (Hayes *et al.* 2018). Male fertility traits, for example, scrotal circumference, are relatively cheaper and easier to measure and are genetically correlated with expensive to measure female traits, such as age at puberty (Johnston *et al.* 2014; Olasege *et al.* 2021a). This implies that there are possibilities that some quantitative trait loci may affect both male and female traits in the same direction (Bolormaa *et al.* 2014).

Currently, there is a paucity of information on how accurately SNP effects from male fertility phenotypes can predict correlated female traits. Recently, Olasege *et al.* (2021b) developed a new methodology that can identify genomic regions driving and antagonizing correlated male and female fertility traits. This methodology uses best linear unbiased prediction (BLUP) solution of SNP effects to estimate the local correlations between studied traits using 500-SNPs in a 100-SNPs sliding window across each chromosome. The resulting estimates were used to infer drivers and antagonizers regions contributing to the overall genomic correlation between traits. However, the application of this methodology is yet to be tested on genomic predictions. Hence, the aim of this study was to examine the accuracy of across-sex genomic predictions using pre-selected SNPs with the new methodology compared to the whole high-density (HD) SNP chip.

Materials & methods

Traits, genotypes and estimated genomic correlations. The traits used in this study were a subset of male and female traits from our previous study (Olasege *et al.* 2021a). The traits were age at first corpus luteum (AGECL), scrotal circumference measured at 12-months (SC12) and serum levels of insulin growth hormone (IGF1 measured in bulls, IGF1b, or cows, IGF1c), all measured in Brahman (BB) population. Details about the pairwise genomic correlations for these traits are provided in Table 1. The selected

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_305, © B.S. Olasege *et al.* 2022

Table 1. Genomic correlation estimates for each pairwise trait used in this study.[1]

Pairwise traits	No of animals	Genomic correlation (S.E)	Degree of correlation
AGECL vs SC12	AGECL – 980 SC12 – 1,020	-0.32 (0.15)	Low
AGECL vs IGF1b	AGECL – 980 IGF1b – 964	-0.65 (0.13)	Moderate
IGF1c vs IGF1b	IGF1c – 995 IGF1b – 964	0.86 (0.11)	High

[1] AGECL, age at first *corpus luteum;* IGF1c, serum levels of insulin growth hormone (IGF1 measured in cows at 24-month, IGF1c, or in bulls at 6-month, IGF1b). SC12, scrotal circumference at 12-month. S.E; standard error.

pairwise traits represented three degrees of genomic correlation estimates i.e. low, moderate, and high. Briefly, across-sex genomic correlations were estimated through a bivariate analysis using linear mixed model fitted in GIBBS2F90 (Misztal *et al.* 2002). After quality control (i.e. excluding SNPs with minor allele frequency <5%), about 645K SNPs remained for BB datasets used for the analysis.

Identification and selection of correlation-driver set of SNPs. The correlation-driver set of SNPs were identified following the correlation scan methodology described by Olasege *et al.* (2021b). To avoid potential bias in downstream prediction analyses, we split the male and female data into training (80%) and validation set (20%) in a five-fold cross-validation. The male and female training sets were used to estimate genomic breeding values (GEBVs) for the respective male and female traits using GBLUP model:

$$y = Xb + Za + e \quad (1)$$

where *y* is the vector of phenotypic values, X is the incidence matrix of fixed effects, *b* is the vector of fixed effects, Z is the design matrix assigned to genomic breeding values, g is the vector of GEBVs for all animals and e is the vector of residuals. Vectors g and e are assumed to follow a normal distribution, thus $g \sim N(0, G\sigma^2 g)$ and $e \sim N(0, I\sigma^2 e)$. Matrix G is the genomic relationship matrix.

After this, we then back-solved these GEBVs for each male and female traits (training sets) to obtain SNP effects using the method of Strandén and Garrick (2009):

$$\hat{\mathbf{u}} = \frac{1}{\sum_{j=1}^{m} 2p_j(1-p)} \mathbf{X}'\mathbf{G}^{-1}\hat{\mathbf{a}} \quad (2)$$

Where û is the vector of estimated SNP effects; *m* is the number of SNPs; p_j is the allele frequency of the second allele of the j^{th} marker; X is a matrix with gene contents for all markers; G is a genomic relationship matrix and â is a vector of GEBVs.

The SNP effects from the training sets for each pre-selected pair of male and female traits were used to perform a correlation scan using 500 SNPs in a 100 SNPs sliding window across each chromosome, as described by Olasege *et al.* (2021b). Consequently, the set of SNPs that drove the genomic correlation between pairs of male and female traits were identified and were termed as 'correlation-drivers'. Table 2 shows the average number of the correlation-driver SNPs across the five-folds, as compared to the HD SNPs set.

Genomic prediction analysis. The SNPs identified as correlation-drivers in pairwise correlation scan analysis were selected from training datasets of male traits and their effects were estimated using GBLUP model. These SNP effects were used to calculate GEBVs for the corresponding validation sets for the female traits using the equation: $GEBV_f=Z_f g_{m,}$ where Z_f is the incidence matrix of the additive allele score of each SNP for each female and g_m is corresponding SNP effect for each SNP in the males. The same procedure was followed for the estimation of GEBVs for validation sets of female traits using whole HD SNPs. The prediction accuracy was calculated as correlation between $GEBV_f$ and their phenotypes adjusted for fixed effects, averaged across the five validation groups, divided by the square root of heritability of the female traits.

Results

Accuracy of genomic prediction. Figure 1 shows the accuracy of genomic predictions (correlation-driver vs HD SNP set) using the SNP effects from male phenotypes to predict female traits, having varying degree of genomic correlations with male traits (low, ~-0.3; moderate, ~-0.6; and high, ~0.9). For the correlation-driver set of SNPs, accuracies were 0.15, 0.27, 0.38 for the low, moderate, and highly correlated traits, respectively. For HD SNP sets, prediction accuracies were 0.11, 0.20, and 0.29 for these pairs of correlated traits. Furthermore, the accuracy of genomic prediction increased as the genomic correlation between traits increased. In summary, the correlation-driver set of SNPs, on average, improved accuracy of genomic prediction across-sex by 30%, irrespective of the degree of genomic correlation between the male and female traits.

Table 2. The average number of correlation-driver and HD SNPs for the pairwise traits.[1]

Pairwise traits	No of correlation-driver SNPs, averaged across the five-folds	No of HD SNPs
AGECL vs SC12	216K	645K
AGECL vs IGF1b	284K	645K
IGF1c vs IGF1b	341K	645K

[1] AGECL, age at first *corpus luteum*; IGF1c, serum levels of insulin growth hormone (IGF1 measured in cows at 24-month, IGF1c, or in bulls at 6-month, IGF1b). SC12, scrotal circumference measured at 12-month.

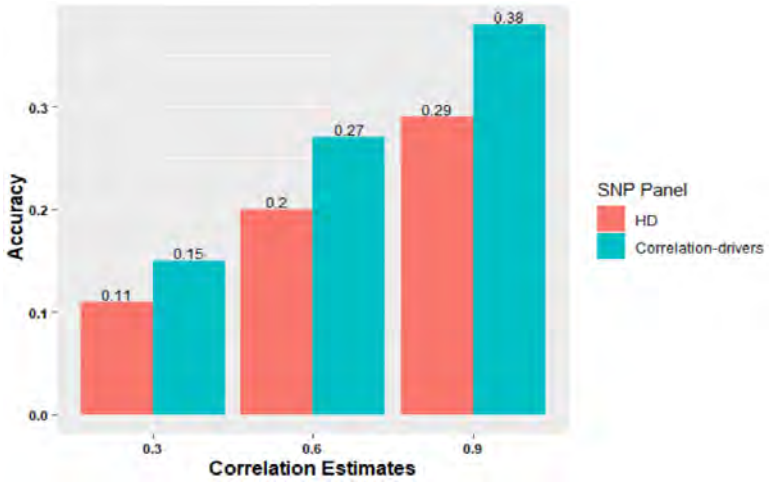


Figure 1. Comparison of the accuracy of genomic prediction (correlation-driver vs HD) using SNP effects from male traits to predict varying degrees of correlated female traits.

Discussion

The objective of this study was to investigate the possibilities of using cheap and easy-to-measure male fertility phenotypes to predict correlated female fertility traits that are difficult and expensive to measure. The results of our analyses using both the pre-selected correlation-driver and HD SNP sets show that female fertility traits can be predicted from correlated male fertility phenotypes. However, the pre-selected correlation-driver SNP sets demonstrated superiority in prediction accuracy (~30% increase) compared to the HD set of SNPs, irrespective of the degree of genomic correlation between the male and female traits. Although we used a relatively smaller sample size, these accuracies are likely to be further improved by using a larger sample size. Studies have shown that using pre-selected markers can improve genomic prediction accuracy (Raymond *et al.* 2018; Al Kalaldeh *et al.* 2019). The increase in accuracies for the correlation-driver SNPs set is expected since the direction of effects of these SNPs are the same for both male and female traits. It will be interesting to investigate the possibilities of using these set of SNPs in a different set of populations. This will be the next phase of research.

References

Al Kalaldeh M., Gibson J., Duijvesteijn N., Daetwyler H.D., Macleod I. *et al.* (2019) Genetics Selection Evolution 51(1)**:** 32. https://doi.org/10.1186/s12711-019-0476-4

Bolormaa S., Pryce J.E., Reverter A., Zhang Y., Barendse W. *et al.* (2014) PLoS Genet 10(3)**:** e1004198. https://doi.org/10.1371/journal.pgen.1004198

Hayes B.J., Corbet N.J., Allen J.M., Laing A.R., Fordyce G. *et al.* (2018) Journal of animal science 97(1)**:** 55-62. https://doi.org/10.1093/jas/sky417

Johnston D., Corbet N., Barwick S., Wolcott M.L. and Holroyd R. (2014) Animal Production Science 54(1)**:** 74-84. https://doi.org/10.1071/AN13044

Misztal I., Tsuruta S., Strabel T., Auvray B., Druet T. *et al.* (2002) BLUPF90 and related programs (BGF90), pp. 743-744 in *Proceedings of the 7th world congress on genetics applied to livestock production*.

Olasege B.S., Porto-Neto L.R., Tahir M.S., Gouveia G.C., Cánovas A. *et al.* (2021b) bioRxiv. https://doi.org/10.1101/2021.11.05.467409

Olasege B.S., Tahir M.S., Gouveia G.C., Kour J., Porto-Neto L.R. *et al.* (2021a) Animal Production Science. https://doi.org/10.1071/AN21097

Raymond B., Bouwman A.C., Wientjes Y.C., Schrooten C., Houwing-Duistermaat J. *et al.* (2018) Genetics Selection Evolution 50(1)**:** 1-14. https://doi.org/10.1186/s12711-018-0419-5

Strandén I., and Garrick D. (2009) Journal of dairy science 92(6)**:** 2971-2975. https://doi.org/10.3168/jds.2008-1929

Taylor W., and Rudder T. (1986) Reduction of sale age for export carcases: will it affect beef producers in tropical regions, pp. 379-382 in *Proceedings of the Australian Society of Animal Production*.

306. Genomic prediction with incomplete omics data

E. Karaman*, V. Milkeviych, Z. Cai, L. Janss, G. Sahana and M.S. Lund

Center for Quantitative Genetics and Genomics, Faculty of Technical Sciences, Aarhus University, 8830 Tjele, Denmark; emre@qgg.au.dk

Abstract

In animal breeding, there has been an increasing interest in investigating the added value of intermediate omics traits such as transcriptomes, metabolites and methylation patterns in genomic predictions. Such data are available only for small number of animals. The 'single-step genomic prediction' machinery, which was first proposed to combine pedigree information of a large number of individuals, and genomic information of a fraction of the population, can be useful to handle incomplete omics data. Such an approach, when applied to incomplete omics data scenarios, imply a simple linear relationship from genotypes to different omics traits, which in reality may be very complex. Little is known about the accuracy of genetic evaluations when the omics traits are generated for the whole population. Here, we present two different approaches to handle incomplete omics data, and investigate their impact on genomic predictions, using simulations.

Introduction

Genomic prediction relies on the estimation of the effects for tens of thousands of genetic variants (generally single nucleotide polymorphisms-SNPs) over the whole genome (Meuwissen *et al.*, 2001), and has received great attention in animal and plant breeding as well as in human genetics studies. Recently, research interests moved towards exploring the added value of intermediate omics traits, such as transcriptomes, metabolites, methylation patterns, etc. in genetic evaluations.

Christensen *et al.* (2021) provided formulas for breeding values required for genomic evaluation when intermediate traits are included, based on a set of two models: (1) a model relating intermediate omics traits to phenotypes; and (2) a model relating genotypes and intermediate omics traits. Such an approach has been used in transcriptome-wide association studies (TWAS) to discover gene-trait associations (Wainberg *et al.*, 2019). Analogous to conventional genomic predictions that estimates SNP effects in a reference population, TWAS first uses a reference population of individuals with SNP genotypes and gene expression data. This population is used to estimate the effect sizes of expression quantitative trait loci (eQTL). Second, expression levels are predicted for the individuals in an independent population, for which genotypes and phenotypes are available. Third, statistical associations are tested for their significance between predicted gene expression data and the trait (Wainberg *et al.*, 2019).

A limitation of incorporating intermediate omics traits in genomic predictions is that, they are available for only a fraction of the populations compared to the number of genotyped individuals. When omics data are available for only a subset of the population, a genomic relationship matrix between individuals with and without omics data can be used to predict (impute) the missing omics data (Christensen *et al.*, 2021). This, however, imply a simple linear relationship from genotypes to different omics traits, which in reality may be very complex. We hypothesize here that if the structure of the gene regulatory network (GRN) responsible for the complex trait can be inferred from the available omics data, the unobserved omics traits can be predicted more accurately, and in turn accuracy in genomic evaluations can be improved. In addition, assuming the consistency of underlying true GRN across breeds (Huang *et al.*, 2012), we further hypothesized that GRN estimated in one breed can be used to predict expression data in other breeds, and in turn for genomic predictions.

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_306, © E. Karaman *et al.* 2022

Materials & methods

Models. We used a two-level model as in Christensen *et al.* (2021) for genomic predictions.

$$\boldsymbol{y} = \mathbf{1}\mu + \mathbf{M}\boldsymbol{\alpha} + \boldsymbol{\epsilon} \tag{1}$$

$$\mathbf{m}_{\mathrm{i}} = \mathbf{1}\mu_{o,i} + \mathbf{g}_{\mathrm{i}} + \boldsymbol{e}_{\mathrm{i}},\ i = 1, \ldots, K \tag{2}$$

where $\mu_{o,i}$ is the mean of expression level i, m_{i} $(N \times 1)$ vector of expression levels, g_{i} is the vector of genetic effects with $\mathrm{g}_{\mathrm{i}} \sim N(0, \mathrm{G}\sigma_{\mathrm{g}}^{2})$ where G is a relationship matrix computed based on SNPs, and e_{i} is the vector of residuals with $e_{\mathrm{i}} \sim N(0, \mathrm{I}\sigma_{\mathrm{e}}^{2})$, for gene i. The μ is the phenotypic mean, M is an $(N \times K)$ matrix of expression levels, α is a $(K \times 1)$ vector of omics' effect sizes, and ε is environmental effect. Equation 2 is a linear mixed model (LMM) with an underlying assumption of a linear path between SNPs to expression levels through the G matrix, whereas from SNPs to expression levels there can be very complex relationships.

For GRN-based description of the relationship between omics expression levels and genetic effects, we replaced Equation 2 with a structural equation model (Equation 3) to describe the expression levels on K genes for individual j (Zhou and Cai, 2022):

$$\mathbf{m}_{\mathrm{j}} = \mathbf{b} + \boldsymbol{\Lambda}_1\mathbf{m}_{\mathrm{j}} + \boldsymbol{\Lambda}_2\mathbf{t}_{\mathrm{j}}^{*} + \boldsymbol{\varepsilon_j} \tag{3}$$

where b is a $(K \times 1)$ vector of general means, Λ_1 $(K \times K)$ and Λ_2 $\left(K \times (\sum_{k=1}^{K} n_k)\right)$ are matrices of the regulatory effects among genes behind expression levels, and among the candidate *cis*-eQTL, respectively,

$$\mathbf{t}_j^{*} = \begin{bmatrix} \mathbf{t}_{j1} \\ \vdots \\ \mathbf{t}_{jK} \end{bmatrix}$$

is a $\left((\sum_{k=1}^{K} n_k) \times 1\right)$ vector of genotypes of individual j where t_{jk} consists of centered genotype content at n_k candidate eQTL for gene k, b is a $(K \times 1)$ vector of means, and ε_j is a vector of residuals with $e_{\mathrm{i}} \sim N(0, \mathrm{I}\sigma_{\varepsilon}^{2})$. Rearranging Equation 3 we get:

$$\mathbf{m}_{\mathrm{j}} = \mathbf{b}^{*} + \boldsymbol{\Lambda}\mathbf{t}_{\mathrm{j}}^{*} + \boldsymbol{\varepsilon}_j^{*}$$

where $\mathbf{b}^{*} = (\mathbf{I} - \boldsymbol{\Lambda}_1)^{-1}\mathbf{b}, \boldsymbol{\Lambda} = (\mathbf{I} - \boldsymbol{\Lambda}_1)^{-1}\boldsymbol{\Lambda}_2$ is a $\left(K \times (\sum_{k=1}^{K} n_k)\right)$, and $\boldsymbol{\varepsilon}^{*} = (\mathbf{I} - \boldsymbol{\Lambda}_1)^{-1}\boldsymbol{\varepsilon}$.

Data simulation. Genotypic data were simulated using XSim (Cheng *et al.* 2015), based on an imputed (50k->HD->WGS) whole genome sequenced SNPs (WGS) dataset for 5,783 Holstein (HOL), and 1,305 Jersey (JER) cows. We restricted ourselves to a randomly selected 0.25 Morgan (1 Mb = 1 Morgan) from chromosome 25, for the proof of concept. In total ten generations were simulated, and population sizes were the same as initial population sizes. The numbers of males were 500 (HOL) and 100 (JER), and sex was assigned at random.

In order to simulate eQTL and gene expression data, we randomly selected 10 of the genes within the selected chromosome region, using the information in the Ensembl database (Howe *et al.*, 2021). Then, for each of the 10 genes, we randomly selected 20 SNPs as a list of candidate eQTL, from 1 kb up- and downstream of the gene's TSS site. Two out of the 20 SNPs per gene were selected at random as the cis-eQTL of the gene. Expression data of individual j were simulated with $\mathbf{m}_{\mathrm{j}} = (\mathbf{I} - \boldsymbol{\Lambda}_1)^{-1}\mathbf{b} + \boldsymbol{\Lambda}\mathbf{t}_{\mathrm{j}}^{*} + (\mathbf{I} - \boldsymbol{\Lambda}_1)^{-1}\boldsymbol{\varepsilon}_j^{*}$ (Zhou and Cai, 2022), where $\mathbf{g}_{\mathrm{j}} = \boldsymbol{\Lambda}\mathbf{t}_{\mathrm{j}}^{*}$ includes the breeding values for omics traits, and each were assumed

to have the same heritability, $h_m^2 = 0.6$. The variances in $\boldsymbol{\varepsilon}_j^*$ of $(\mathbf{I} - \boldsymbol{\Lambda}_1)^{-1}\boldsymbol{\varepsilon}_j^*$, were set according to this heritability. The weight of each gene's expression data on the final trait phenotype, α_i, was simulated from the unit normal distribution. Genomic breeding value of each individual was obtained as $\mathbf{u}_j = \mathbf{g}_j' \boldsymbol{\alpha}$, and the trait heritability was set at 0.1 or 0.5. All these procedures described above were replicated 10 times. The simulated network structure (Figure 1) was the same for both breeds, as well as the heritabilities, across all the replicates. The complexity of the simulated network, the number of genes, heritabilities of the gene expressions and the traits were varied, but only the results from the parameters described above were presented here.

Analyses. We used generations 9 and 10 as reference and validation populations, respectively. All individuals had genotypes, whereas only 300 and 100 individuals from generation 9 had gene expression data (data were masked), for HOL and JER respectively. Individuals in generation 9 had also phenotypes, and phenotypes of individuals in generation 10 were masked. First, we generated (imputed) the omics data to the whole population by fitting a LMM or GRN model. This was done by using HOL (or JER) individuals' available gene expression and genotype data, together with the genotypes of HOL or JER, for which the omics data will be imputed. Second, breeding values were predicted using those imputed omics data by fitting the model in Equation 1. We used JWAS package in Julia language (Cheng *et al.*, 2018) for the analyses using Equations 1 and 2, and ssemQr package (Zhou and Cai, 2022) in R language for the analyses using Equation 3. It is worth to note that the analyses rely on 20 candidate eQTL SNPs of each gene, and two of those were used to generate the gene expression data, and thereby the phenotypes. Accuracy in genomic predictions were measured as the correlation between the true and predicted genomic breeding values, for individuals in generation 10.

Results

The accuracies from different scenarios are presented in Table 1. When the heritability of the trait was 0.5, accuracies from GRN and LMM were ranged between 0.89-0.99 and 0.82-0.95, respectively. In general, when the imputations were within-breed, genomic predictions were the highest. Using HOL as the reference to impute gene expressions of JER led to accuracies of 0.91 and 0.86, using GRN and LMM, respectively, for heritability of 0.1. Those were 0.97 and 0.93, when the heritability was 0.5.

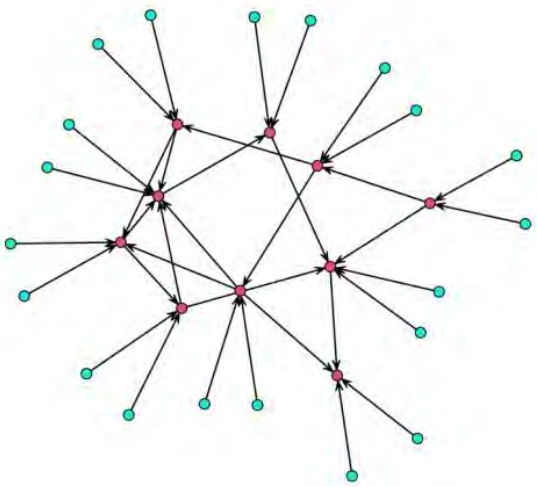


Figure 1. The structure of the simulated GRN. The red and blue circles stand for genes and eQTL.

Discussion

In this work, we have used a hierarchical modeling approach to describe the relationship between SNPs and phenotypes, which was earlier used for TWAS (Wainberg *et al.*, 2019) and recently for integrating intermediate omics traits in genomic prediction (Christensen *et al.*, 2021). So far, the approach has not been explored extensively for genomic prediction, and little is known about the accuracy of breeding value predictions when the omics traits are extended (imputed) to the whole population. We compared two approaches, LMM and GRN, to handle incomplete intermediate omics traits for genomic prediction, and showed that the LMM-based imputation approach was inferior to GRN-based approach, even when the simulated true GRN was not highly complex (Figure 1). Our results are only preliminary and shown here primarily for the purpose of illustrating the effectiveness of accounting for complex traits' genetic architecture, in genomic predictions integrating omics data. It is worth to note that the accuracies (Table 1) are generally much higher (above 0.8) than one might observe in real data. By design, all genes leading to ultimate trait phenotype was known, and the phenotypes were linearly related to steady-state expression levels of those genes. These conditions can hardly be met for real-life complex traits. Analogous to across breed genomic predictions, we used the estimated GRN for one breed to predict gene expression levels in another breed, though GRN is influenced by many factors. When gene expression and other relevant data for, for example genotypes of the two breeds, are available, it is more appropriate to use the estimated GRN in one breed to inform the estimation of GRN in the other, but not to replace it (Zhou and Cai., 2020).

Table 1. Prediction accuracies (standard errors).

Omics data	Prediction	Model for gene expression data			
		GRN		LMM	
		h^2=0.1	h^2=0.5	h^2=0.1	h^2=0.5
HOL	HOL	0.98(0.002)	0.99(0.000)	0.94(0.003)	0.95(0.003)
JER	JER	0.95(0.002)	0.98(0.002)	0.88(0.004)	0.94(0.001)
HOL	JER	0.91(0.007)	0.97(0.002)	0.86(0.008)	0.93(0.004)
JER	HOL	0.78(0.013)	0.89(0.009)	0.74(0.010)	0.82(0.011)

References

Cheng, H., Garrick, D., Fernando, R. (2015) G3 (Bethesda, Md.) 5(7):1415-1417. https://doi.org/10.1534/g3.115.016683

Cheng H., Fernando R., Garrick D. (2018a) JWAS: Julia implementation of whole-genome analysis software. World Congr Genet Appl Livest Prod, Auckland, NZ 11:859

Christensen, O.F., Börner, V., Varona, L., Legarra, A. (2021) 219(2):iyab130, https://doi.org/10.1093/genetics/iyab130

Howe, K.L., Achuthan, P., Allen, J., Allen, J., Alvarez-Jarreta, J *et al.* (2021) Nucleic Acids Res. 49(1):884-891 https://doi.org/10.1093/nar/gkaa942

Huang, W., Richards, S., Carbone, M. A., Zhu, D., Anholt, R. R., *et al.* (2012) Proc. of the National Academy of Sciences of the United States of America, 109(39):15553-15559. https://doi.org/10.1073/pnas.1213423109

Meuwissen T.H.E., Hayes B.J., and Goddard M.E. (2001) Genetics 157(4):1819-1829. https://doi.org/10.1093/genetics/157.4.1819

Wainberg, M., Sinnott-Armstrong, N., Mancuso, N., Barbeira, A.N., Knowles, D.A., *et al.* (2019) Nat Genet 51:592-599. https://doi.org/10.1038/s41588-019-0385-z2019

Zhou, X., and Cai, X. (2020) Bioinformatics 36(1):197–204. https://doi.org/10.1093/bioinformatics/btz5292020

Zhou, X., and Cai, X. (2022) Bioinformatics 38(1):149-156. https://doi.org/10.1093/bioinformatics/btab609

307. A Bayesian hierarchical model to integrate a mechanistic growth model in genomic prediction

H. Yu[1], J. van Milgen[2], E. Knol[3], R. Fernando[1] and J. Dekkers[1*]

[1]Iowa State University, Ames, IA 50011, USA; [2]PEGASE, INRAE, 35590 Saint-Gilles, France; [3]Topigs Norsvin, 6640 AA Beuningen, the Netherlands; jdekkers@iastate.edu

Abstract

Genomic prediction can improve the accuracy of estimated breeding values for traits driven by additive genetic effects within common settings but prediction of traits affected by non-additive genetic effects and G×E remains a challenge. Mechanistic growth models express growth performances in terms of nonlinear functional interactions between underlying latent traits and nutritional environmental effects. Assuming the latent traits are less affected by non-additive genetic effects and G×E, these models can capture certain non-additive genetic effects and G×E at the phenotype level and allow prediction at unobserved ages for longitudinal data, e.g. mature weight and mature feed intake. In this study, we developed a Bayesian hierarchical model to integrate a Gompertz model for body weight and feed intake into genomic prediction models for pigs. By predicting breeding values for biologically relevant underlying latent traits, these models have the potential to advance genetic improvement across populations and environments.

Introduction

Genomic prediction (GP) has facilitated genetic improvement by enabling more accurate estimated breeding values at an early age (Meuwissen *et al.*, 2001). While GP is effective in predicting traits dominated by additive genetic effects within common settings, challenges of predicting non-additive genetic effects and extrapolating to other conditions, such as to later ages or to other nutritional environments (G×E) remain. Previous attempts to address these challenges have mainly focused on statistical models, while augmentation of statistical models with biological information has received less attention. Developed from a nutrition perspective, mechanistic growth models (MGM) (e.g. Van Milgen *et al.*, 2008) describe growth performance as nonlinear functional interactions between underlying latent traits and nutritional environmental factors and can capture GxE and non-additive genetic effects that are created by the non-linear relationships embedded in the MGM. Following the pioneering ideas of Varona *et al.* (1997) and Bourdon (1998), several studies have sought to accommodate such MGM in genetic prediction (Doeschl-Wilson *et al.*, 2007; Cai *et al.*, 2012). Their successes were limited either by lack of appropriate statistical and computational methods or by the absence of genomic information. In plant breeding, crop growth models have recently been integrated into GP (Technow *et al.*, 2015; Messina *et al.*, 2018; Campbell *et al.*, 2020), leading to increases in prediction accuracies for performance in other environments compared to standard GP methods. Against this background, the objective of this study was to develop a Bayesian hierarchical model to integrate an MGM into GP for pigs.

Materials & methods

Bayesian hierarchical model for body weight and feed intake. The proposed model to integrate a MGM for body weight (BW) and daily feed intake (DFI) into GP is presented in Figure 1 and was implemented using the programming language Julia (Bezanson, Jeff, *et al.*, 2017). The MGM of the InraPorc model (van Milgen *et al.*, 2008), with a reparameterized Gompertz model for BW and a gamma function for DFI as a function of BW, was used to connect five latent traits to observed BW and DFI data. The five latent traits (i.e. DFI50, DFI100, Age115, Shape, and BW65, as defined in Figure 1, were chosen based on biological relevance, and they were modeled as a linear combination of an intercept (μ), contemporary group effects (CG), and additive SNP (α) and residual effects (e). Observed traits of BW and DFI were

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_307, © H. Yu *et al.* 2022

then modeled as a function of the latent variables through the MGM. While latent traits were modeled as additive effects without G×E, the MGM allows for certain non-additive genetic effects and G×E for BW and DFI by internally creating non-linearity and interactions among genetic and environmental effects. We implemented MCMC to sample the unknown parameters of the MGM using a multi-trait Bayesian regression model. Specifically, the μ, GC and α were sampled conditional on the latent traits using Gibbs sampling, while the five latent traits were jointly sampled using the Metropolis-Hastings algorithm, given the observed BW and DFI data. Variance components were obtained by Gibbs sampling.

Application. The model was applied to daily BW and DFI data from ~83 to ~186 days of age on 3,934 purebred boars in 97 contemporary groups, that were genotyped for 24K SNPs. Phenotypes of the 2,000 earlier born pigs were used as training and the remaining pigs as validation. Prediction bias and accuracy were estimated by forward cross-validation (CV) based on regression and correlation (divided by the square root of heritability) of observed BW and DFI adjusted for CG on predictions for the validation data. Because such CV does not allow evaluation of predictions for latent variables and for BW and DFI outside the observed age range, predictions were also evaluated using the linear regression (LR) method of Legarra and Reverter (2018), using predictions based on partial and whole data, which we applied to the non-linear models.

Results & discussion

Figure 2 shows GP for BW and DFI from 0 to 500 days for the validation data. Note that predictions extend outside the age range with phenotypes, which is facilitated by the MGM. The GP for BW and DFI plateaued at ~400 and ~300 days, respectively. Previous studies used a 2-step method to incorporate growth models into genetic evaluation (Kachman *et al.*, 1998; Vautier *et al.*, 2013), with latent variables estimated on an individual basis in a first step, using only that individual's data, followed by genetic analysis of the estimated latent variables as phenotypes. Our model integrates these two steps and leverages relationships among individuals to estimate the latent traits for all individuals, regardless of the amount of data available for each.

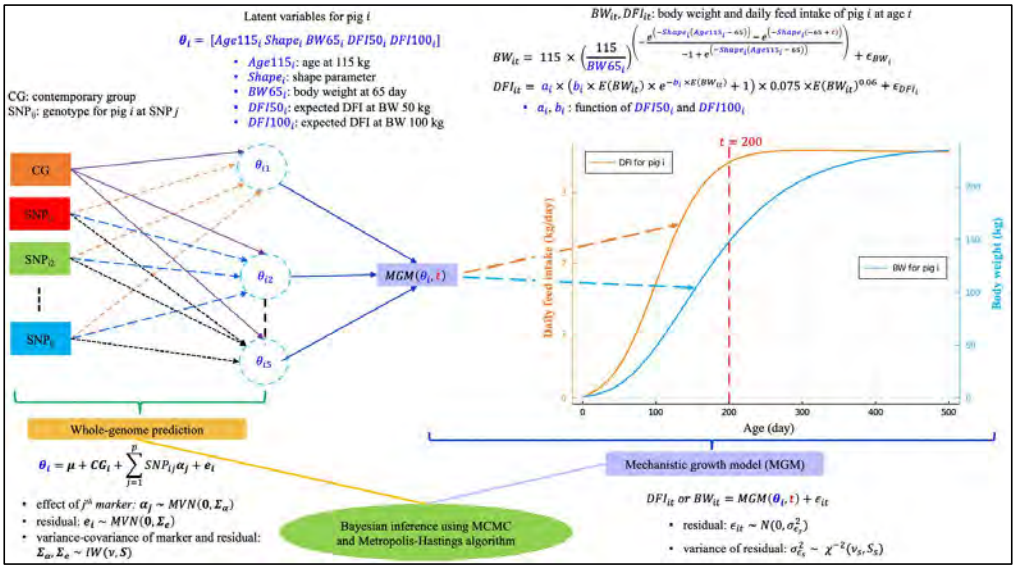


Figure 1. Structure of the Bayesian hierarchical model to integrate a mechanistic growth model (MGM) with five latent variables (blue font) for BW and DFI into genomic prediction.

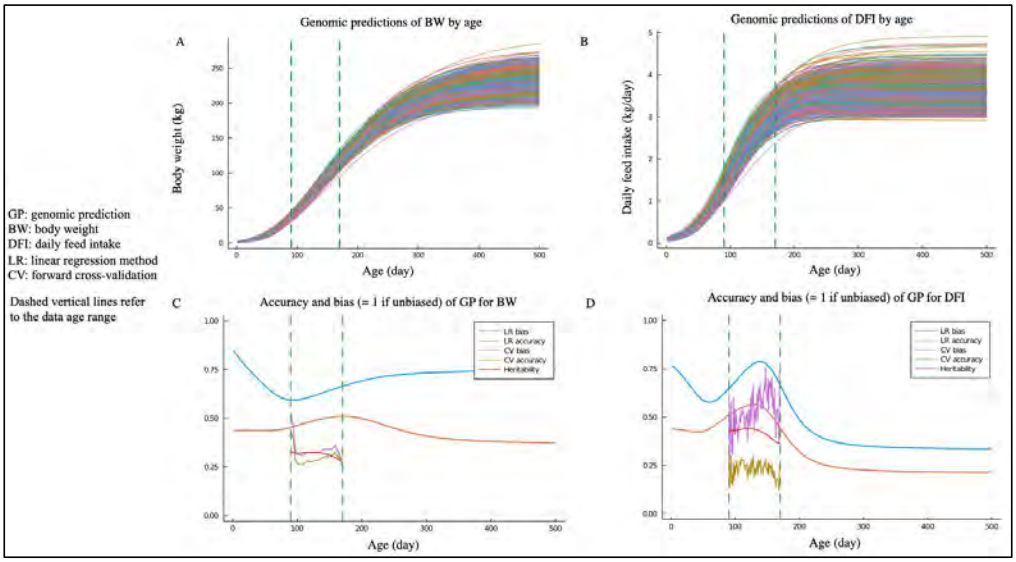


Figure 2. GP for BW (A) and DFI (B) for the validation data and estimates of bias and accuracy of GP of BW (C) and DFI (D) using the LR and CV methods.

Estimates of accuracy and bias for BW and DFI based on the LR and CV methods are also in Figure 2. Leveraging the LR method, we computed the accuracy and bias of GP of the latent traits (Table 1) and for BW and DFI outside the observed data (Figure 2). The LR method uses the relationship between predictions based on partial and whole data and can, therefore, estimate accuracy and bias of predictions of traits that do not have corresponding observed phenotypes (Legarra and Reverter, 2018). The latent traits have specific biological meaning related to growth processes based on the MGM of van Milgen *et al.* (2008). The GP for latent trait DFI100 had the highest accuracy (0.51), while the GP for Shape had the lowest accuracy (0.43). As Shape is related to the inflection point and mature weight, its low accuracy is likely because the BW and DFI data are from a very linear part of the growth curve. Prediction bias was found for all latent traits, ranging from 0.61 to 0.91, compared to the expectation of 1 for unbiased predictions.

Compared to the CV method, which only allows evaluation of predictions within the age range with phenotypes, the LR method resulted in much higher estimates of accuracy and lower estimates of bias (i.e. closer to 1) of predictions for BW and DFI for that age rage, although the trends with age were similar (Figure 2). Macedo *et al.* (2020) pointed out that the LR method can result in false estimates of bias when a seriously flawed model is fitted but that estimates of accuracy are robust. However, in a separate simulation study, we found that estimates of both bias and accuracy from the LR method can be biased when using a flawed model, in contrast to the CV method.

Table 1. Bias and accuracy of genomic predictions of latent traits based on the LR method.

	Age115	Shape	BW65	DFI50	DFI100
Bias (unbiased if = 1)	0.66	0.73	0.61	0.83	0.91
Accuracy	0.51	0.43	0.45	0.51	0.55

In conclusion, the proposed Bayesian hierarchical model allows GP for biologically relevant latent traits based on a MGM and can, thereby, accommodate certain non-additive genetic effects and G×E and allow prediction of traits outside the age range with data. Selection on GP for latent traits is expected to lead to greater rates of genetic improvement and the ability to better deal with G×E, if the latent traits are less affected by non-additive genetic effects and G×E. Whether they are, however, requires further investigation and potentially reparameterization of the latent variables. However, caution must be exercised when using the LR method to evaluate predictions as it may provide biased estimates of accuracy and bias when the model used is not valid for the data at hand.

Acknowledgements

Funded by USDA-NIFA grant # 2020-67015-31031.

References

Bezanson J., Edelman A., Karpinski S. and Shah V.B. (2017) SIAM review. 59(1):65-98. https://doi.org/10.1137/141000671

Bourdon R.M. (1998) J. Animal Sci 76(9):2308-2323. https://doi.org/10.2527/1998.7692308x

Cai W., Kaiser M.S., and Dekkers J.C.M. (2012) J. Animal Sci. 90(1):127-141. https://doi.org/10.2527/jas.2011-4293

Campbell M.T., Grondin A., Walia H., and Morota G. (2020) J. Experimental Botany 71(18):5669-5679. https://doi.org/10.1093/jxb/eraa280

Doeschl-Wilson A.B., Knap P.W., Kinghorn B.P., and Van der Steen H.A.M. (2007) Animal 1(4):489-499. https://doi.org/10.1017/S1751731107691848

Kachman S.D., Baker R.L., and Gianola D. (1988) Theor. Appl. Genetics 76(1):148-156. https://doi.org/10.1007/BF00288846

Legarra A., and Reverter A. (2018) Genet. Sel. Evol. 50(1):1-18. https://doi.org/10.1186/s12711-018-0426-6

Macedo F.L., Reverter A., and Legarra A. (2020). J. Dairy Sci., 103(1):529-544. https://doi.org/10.3168/jds.2019-16603

Messina C.D., Technow F., Tang T., Totir R., Gho C., *et al.* (2018) European J. Agronomy, 100:151-162. https://doi.org/10.1016/j.eja.2018.01.007

Meuwissen T.H.E., Hayes B.J., and Goddard M.E. (2001) Genetics 157(4):1819-1829. https://doi.org/10.1093/genetics/157.4.1819

Technow F., Messina C.D., Totir L.R., and Cooper M. (2015) PloS one 10(6):e0130855. https://doi.org/10.1371/journal.pone.0130855

Varona L., Moreno C., García Cortés L.A., and Altarriba J. (1997) Livest. Prod. Sci. 47(3):201-209. https://doi.org/10.1016/S0301-6226(96)01415-7

van Milgen J., Valancogne A., Dubois S., Dourmad J.Y., Sève B., *et al.* (2008) Animal Feed Sci. and Technology, 143(1-4):387-405. https://doi.org/10.1016/j.anifeedsci.2007.05.020

Vautier B., Quiniou N., van Milgen J., and Brossard L. (2013) Animal 7(8):1265-1273. https://doi.org/10.1017/S1751731113000554

308. Mega-scale mixed models for genome-wide prediction with thousands of high-throughput phenotyping traits

J. Qu[1], D.E. Runcie[2*] and H. Cheng[1*]

[1]*Department of Animal Science, University of California, Davis, One Shield Avenue, Davis, CA 95616, USA;* [2]*Department of Plant Sciences, University of California, Davis, One Shield Avenue, Davis, CA 95616, USA;* *deruncie@ucdavis.edu; qtlcheng@ucdavis.edu*

Abstract

Large-scale phenotypic data is widely available today due to advances in high-throughput phenotyping platforms. Genetic analyses of high-dimensional, highly correlated traits using conventional multivariate linear mixed models presents analytical and computational challenges. Here, we use a Bayesian sparse factor model to adapt multivariate linear mixed models to the application of genome-wide prediction with thousands of traits. We have implemented both kernel-based models (e.g. genomic best linear unbiased prediction (GBLUP)) and marker effects models (e.g. BayesC) in the Bayesian sparse factor model statistical framework. Our mega-scale mixed models show significant improvement for genetic value prediction in real data by leveraging information from thousands of traits jointly, particularly when using mixture priors on marker effects.

Introduction

With the recent advances in high-throughput phenotyping technologies, including multi-omics methodologies that are used for generating molecular phenotypes, statistical models that enable the joint analysis of large numbers of highly correlated traits are needed to enable efficient genomic selection. However, genomic analyses of high-dimensional, highly correlated trait data presents computational challenges. In the past several decades, multivariate linear mixed model (MvLMM) have been widely used for genetic analyses of two or more correlated traits (Henderson *et al.*, 1976). Since the number of covariance parameters in MvLMMs increases quadratically with the number of traits, and most algorithms used to fit MvLMMs require repeated inversions of the genetic and residual covariance matrices, MvLMMs can be susceptible to over-fitting and have a computational burden that grows cubically to quintically as the number of traits increases. To alleviate much of this computational burden, we re-parameterize MvLMMs as Bayesian sparse factor models to address these statistical and computational challenges.

We have implemented both kernel-based models (e.g. GBLUP) and marker effects models (e.g. BayesC) in Bayesian sparse factor models to accommodate genome-enabled analysis with thousands of traits. Based on the assumption that the covariances among large sets of traits can be explained by a small set of latent factors (e.g. through gene regulatory networks), MegaLMM (Runcie *et al.*, 2013, Runcie *et al.*, 2021) was developed to incorporate genomic data through one (or more) genomic relationship matrices (e.g. MegaGBLUP). These kernel-based models impose specific assumptions about the distribution and effect sizes of the underlying causal genetic variants, and does not allow direct inference on the effects of the causal loci. On the other hand, marker effects models such as the Bayesian Alphabet methods (Cheng *et al.*, 2015; Habier *et al.*, 2011; Meuwissen *et al.*, 2001) encode a wide range of different and more flexible prior distributions on the effect sizes of causal genomic variants and allow for inference of the causal variants themselves. In this paper, marker effects models are incorporated into the Bayesian sparse factor model framework (hereinafter referred to as MegaBayesianAlphabet) to analyze thousands of traits for genome-wide prediction and genome-wide association. We focus on the mixture prior in BayesC due to its incorporation of biologically meaningful assumptions. Extensions of MegaBayesianAlphabet using other mixture priors on marker effects should be straightforward. The objectives of this paper are to show the

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_308, © J. Qu *et al.* 2022

significant improvement in genomic prediction by leveraging information from thousands of traits and mixture priors on marker effects in MegaBayesianAlphabet.

Materials & methods

MvLMMs are extended as Bayesian sparse factor models by introducing K independent (unobserved) latent factors to account for the covariances among t observed traits ($t >> K$).

Model description and prior specification in MegaBayesianAlphabet. Variation among t observed traits is decomposed into two parts: the variation caused by dependencies on K independent latent factors, which induces correlations among the t observed traits, and the variation that is idiosyncratic to each trait. In MegaBayesianAlphabet, genetic values of latent factors are defined as a linear combination of all marker effects, and priors from the Bayesian Alphabet methods are assigned to the marker effects. The model specification of MegaBayesianAlphabet is described below.

$$Y = F\Lambda + X_1B_1 + X_{2R}B_{2R} + E_R \quad (1)$$

With

$$F = X_{2F}B_{2F} + E_F \quad (2)$$

where Y is a matrix of observations for n individuals on t traits, F is a matrix of latent factors for the n individuals across K latent factors, and Λ is a factor loading matrix whose elements describe the corresponding factor-trait relationships. The K latent factors in F are further decomposed into genetic effects and residual effects as shown in Equation 2. X_1 is an incidence matrix allocating the observations to fixed effects in B_1. Residuals conditional on latent factors are similarly decomposed into trait-specific genetic effects (i.e. $X_{2R}B_{2R}$) and trait-specific residual effects (i.e. E_R). The model structure of MegaLMM (e.g. MegaGBLUP) is similar to MegaBayesianAlphabet, except that the genetic effects in MegaGBLUP are expressed as a single random effect based on a genomic relationship matrix. If all sources of correlation among observed traits are explained by the latent factors, residuals conditional on these factors become uncorrelated. With the independence between latent factors, samples at each iteration of MCMC chains can be obtained simultaneously in parallel, which leads to significant decline in the computational cost of model fitting.

In this paper, the BayesC prior is used for the marker effects and we term this specific version of MegaBayesAlphabet: MegaBayesC. In MegaBayesC, the prior distribution of the marker effect at locus i for the k^{th} latent factor is shown as follows.

$$\mathbf{b}_{2F_{k(i)}} = \begin{cases} N\left(0, \sigma^2_{B2F_k}\right) & \text{probability}(1 - \pi_{F_k}) \\ 0 & \text{probability } (\pi_{F_k}) \end{cases} \quad (3)$$

where $0, \sigma^2_{B2F_k}$ is the variance of marker effects corresponding to factor k. Note that if marker inclusion probabilities for all factors were set to 1.0, the model is equivalent to RR-BLUP, and we term this specific version of MegaBayesianAlphabet: MegaRRBLUP. A BayesC mixture prior is also used for the elements of the factor loading matrix (Λ). To improve the identifiability of the model, we introduce an additional parameter (τ_k) to the included-variable variance that is stochastically increasing across factors. For the factor loading that describes the relationship between factor k and trait j, its prior distribution is shown as follows.

$$\mathbf{\Lambda}_{k,j} = \begin{cases} N\left(0, \tau_k^{-1}\sigma^2_{R_j}\right) & \text{probability}(1 - \pi_{\Lambda_k}) \\ 0 & \text{probability } (\pi_{\Lambda_k}) \end{cases} \quad (4)$$

Through this prior specification, an appropriate level of truncation on the rows of Λ is able to ensure that the contribution from additional factors beyond the truncation point is negligible (Bhattacharya *et al.*, 2011).

Data and analysis. Best linear unbiased estimators (BLUEs) of grain yield and reflectances from 62 wavelength bands collected with an aerial hyperspectral camera at each of 10 time-points during the growing season for 1,033 bread wheat lines were downloaded from CIMMYT Research Data (Krause *et al.*, 2019). In our analysis, the 620 hyperspectral BLUEs measured at 10 time points were used as secondary traits to improve the prediction of the genetic value of grain yield, which is served as a focal trait in our prediction scenario. Four different models were used to predict the grain yield: traditional single-trait GBLUP, MegaGBLUP, MegaRRBLUP, and MegaBayesC. In MegaRRBLUP and MegaBayesC, the probability of an element from Λ being zero was considered as known and set to be 0.9 to introduce sparsity and to shorten the time required for its convergence, while the probability of a marker having a null effect on a latent factor was considered as unknown and was estimated. Trait-specific genetic effects are not considered in these analyses. Cross-validation was used to evaluate the predictive performance of different models by masking the grain yield of half of the population during model fitting and comparing the masked values to model predictions. 20 replicates were done with different masked lines.

Results & discussion

Here, we tested if MegaBayesianAlphabet models could match or exceed the performance of MegaLMM in trait-assisted genomic prediction using data from a breeding trial of bread wheat. As a baseline, we performed conventional univariate GBLUP-based genomic value prediction as well. As shown in Figure 1, univariate GBLUP achieved a prediction accuracy of 0.43 in this dataset. MegaGBLUP was fitted in MegaLMM (Runcie *et al.*, 2021) using a single random effect to account for the genetic values among observed traits based on a genomic relationship matrix. The achieved prediction accuracy of MegaGBLUP was 0.69. MegaRRBLUP was fitted in MegaBayesianAlphabet and achieved a prediction accuracy of 0.68. No significant difference was observed between MegaGBLUP and MegaRRBLUP, which is expected due

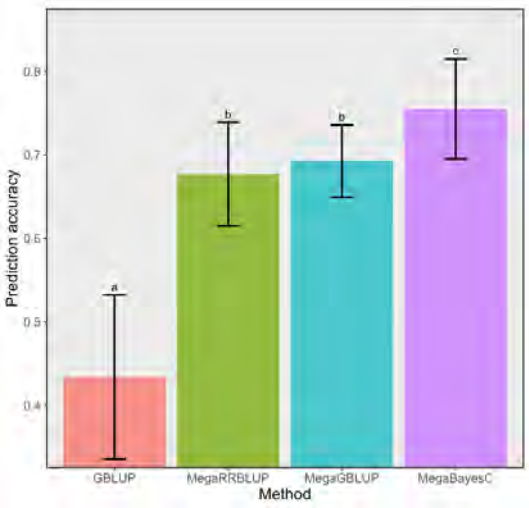


Figure 1. Genomic value prediction performance of four models for wheat yield. Bars show the mean prediction accuracy (± standard error) for each model, and letters show the statistical significance of mean difference between methods based on a paired t-test.

to the equivalence between RR-BLUP and GBLUP in terms of prediction of genetic values as reported in single-trait analysis (Meuwissen *et al.*, 2001; Habier *et al.*, 2007). MegaBayesC, instead of fitting all markers in the model as in MegaRRBLUP, used a mixture prior for the marker effects and achieved a significantly higher prediction accuracy than the other three methods, with an average accuracy of 0.75. These results show that the use of biologically meaningful prior on marker effects can further improve the genomic selection in breeding programs when large amounts of traits were used to assist the prediction of focal trait.

In MegaBayesianAlphabet, due to the independence among latent factors, and the independence among traits conditional on factors, marker effects can be efficiently sampled from a set of univariate BayesC models in parallel across traits and factors at each iteration of MCMC. The Gibbs sampling method used for estimating marker effects, however, can still be time-consuming for large amounts of SNP. To overcome the intractability problem faced by genome-enabled analysis incorporating both large-scale phenotypic data and genomic data, further investigation such as a pre-selection process to narrow down the number of SNPs is needed.

References

Bhattacharya A. and Dunson D.B (2011) Biometrika 98(2):291–306. https://doi.org/10.1093/biomet/asr013

Cheng H., Qu L., Garrick D.J., and Fernando R.L. (2015) Genet Sel Evol 47(1):1-7. https://doi.org/10.1186/s12711-015-0157-x

Habier D., Fernando R.L. and Dekkers J.C. (2007) Genetics 177(4):2389-2397. https://doi.org/10.1534/genetics.107.081190

Habier D., Fernando R.L., Kizilkaya K., and Garrick D.J.(2011) BMC bioinformatics 12(1):1-12. https://doi.org/10.1186/1471-2105-12-186

Henderson C. and Quaas R. (1976) Journal of Animal Science 43(6):1188-1197. https://doi.org/10.2527/jas1976.4361188x

Krause M.R., González-Pérez L., Crossa J., Pérez-Rodríguez P., Montesinos-López O., *et al.* (2019) G3 9(4):1231–1247. https://doi.org/10.1534/g3.118.200856

Meuwissen T.H.E., Hayes B.J., and Goddard M.E. (2001) Genetics 157(4):1819-1829. https://doi.org/10.1093/genetics/157.4.1819

Runcie D.E. and Mukherjee S. (2013) Genetics 194(3):753–767. https://doi.org/10.1534/genetics.113.151217

Runcie D.E., Qu J., Cheng H., and Crawford L. (2021) Genome Biology 22(1):1–25. https://doi.org/10.1186/s13059-021-02416-w

309. Nordic Holstein single-step test day model using metafounders and blending of foreign information

M. Koivula[1*], I. Strandén[1], G.P. Aamand[2] and E.A. Mäntysaari[1]

[1]Natural Resources Institute Finland (Luke), 31600 Jokioinen, Finland; [2]NAV Nordic Cattle Genetic Evaluation, 8200 Aarhus N, Denmark; minna.koivula@luke.fi

Abstract

The aim of the study was to compare the metafounders approach (MF) to conventional genetic unknown parent groups (UPG) and to estimate the effect of blending the external information. All results were compared also to results from the official EBV model (AM). The results showed that both the traditional UPG and the MF approach can be implemented in ssGTBLUP together with the blending of foreign information. However, the MF approach was slightly more effective in reducing bias in young bulls.

Introduction

The use of sires from international populations is common in dairy cattle breeding. The exchange of sire genotypes allows national breeding programs to properly integrate multi-trait cross-country assessments (MACE; Schaeffer, 1994) by Interbull (Uppsala, Sweden) into domestic evaluations. The integration of bull information by blending MACE yield indices into the Nordic test-day (TD) model was introduced by Pitkänen *et al.* (2020).

In single-step models, an unsolved model problem is how to build genomic (G) and pedigree (A_{22}) relationship matrices which meet the theoretical assumptions of the same scale and equal base populations. For example, base population allele frequencies (AF) can be used (VanRaden, 2008), and elements of G can be scaled and centered to have on average the same diagonal and off-diagonal elements as in A_{22} (Christensen, 2012). The metafounder (MF) approach proposed by Legarra *et al.* (2015) attempts to make the A and A_{22} matrices compatible with the G matrix. In the MF approach, allele frequencies (AF) equal to 0.5 for all markers are assumed in the G matrix (Garcia-Baccino *et al.* 2017) and the A matrix has MF or pseudo-individuals with self-relationships. The MF are like unknown parent groups (UPG) but allow a related base population with nonzero inbreeding coefficients (e.g. Legarra *et al.* 2015; Garcia-Baccino *et al.* 2017). The relationships within and between the MF are modelled by a Γ matrix, which thereafter is used in forming the pedigree-based relationship matrix (A_Γ).

The objectives of this study were to compare the performance of the MF approach to the corresponding model with UPG and to address the effect of integration of external information into Nordic Holstein test day (TD) model evaluations. The MF was applied, and MACE information was included in the single-step (model called GT_B_MF), and in the animal model (AM_B_MF). To compare method differences, we run separately also ssGTBLUP with MACE information and the QP transformation where UPGs were included with the QP transformation (GT_B), and ssGTBLUP without MACE information using either the QP transformation (GT) or the metafounders (GT_MF). All the results were compared also to results from the official EBV model (AM).

Materials & methods

The official Nordic HOL milk production evaluation data obtained from the Nordic Cattle Genetic Evaluation (NAV) were used in all analyses of this study. The official evaluation includes TD records from milk, fat and protein production from the first three lactations as multiple traits. The TD data included 8.5 million cows with a total of 173.7 million records and 10.9 million animals in the pedigree. There were

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_309, © M. Koivula *et al.* 2022

274,145 genotyped animals. After the same edits used in current NAV HOL genomic evaluations, 46,342 SNP markers on the 29 bovine autosomes were available for further analysis.

Models. The Nordic multiple traits reduced rank random regression TD model with heterogeneous variance (HV) correction (Lidauer *et al.*, 2015) was used. The model specifies 27 traits: 3 countries (DFS: Denmark, Finland and Sweden), 3 yield traits, 3 lactations. Each animal receives 15 solutions for the random regression effects. With the data used, the mixed model equations (MME) had 155.8 million animal equations and a total of 444.2 million equations. For the validation of bull (G)EBVs, the last four years of observations were removed from the full data.

All models were solved with MiX99 software (Strandén and Lidauer, 1999). The single-step model used ssGTBLUP with efficient computations (Mäntysaari *et al.*, 2017). In all single-step models, the genomic relationship matrix was based on VanRaden method 1 with a 30% residual polygenic proportion, AF of 0.5 for all the markers, and in UPG models diagonals in G were scaled by trace(A_{22})/trace(G).

Four single-step models were tested: GT, GT_B, GT_MF and GT_B_MF. In the GT and GT_B models, 176 genetic groups were included using QP transformation (detailed description in Koivula *et al.*, 2021). The GT_MF and GT_B_MF models used the MF approach. The MF approach needs a covariance matrix, denoted Γ, for the metafounders. The 176 UPGs were used as metafounders. The UPGs were based on 4 breed groups (HOL, RDC, JER, other) and 5 country of origin groups within Holstein (HOLdnk, HOLswe, HOLfin, HOLother, HOLred). Within each of these 9 sources, UPG were further grouped by birth year decade and by selection path when appropriate. For the Γ matrix of the 9 base metafounders, we defined a covariance function (CF) model (Kirkpatrick *et al.*, 1994) described in Kudinov *et al.* (2021). After solving the Γ matrix for the 176 MF, we computed the Γ matrix compliant inbreeding coefficients needed for the inverses A_Γ and A_{22}^Γ in the MME.

Foreign information was included in the GT_B and GT_B_MF models. The external information from MACE was included as described by Pitkänen *et al.* (2020). A bull was considered to have additional information in MACE when its reliability for milk, protein and fat indices in MACE were at least 0.01 units higher compared to the Nordic evaluation reliability. The integration process had three steps: (1) effective record contributions (ERC) in MACE and in the national evaluations were calculated for the selected bulls; (2) yield indices and ERC were used to calculate multi-trait deregressed proofs (DRPs) separately using the Nordic and the MACE data. Finally, (3) based on the two DRPs and the two ERCs, pseudo-observations were calculated for the selected 31,779 bulls and were included in the TD data. In addition, the official animal model (AM) and animal model AM_B_MF using blending with MF were performed.

Results & discussion

Table 1 illustrates the linear regression validation results (LR, Legarra and Reverter, 2018) from the different models for 524 DFS Holstein validation bulls. The level differences in the (G)EBV predictions were corrected by standardizing the mean of cows born 2007 to be the same in all models. The b_0 column has the mean difference (kg) between the full and reduced run (G)EBVs. This illustrates the realized bias between the full and reduced evaluation. The results indicate that the use of the MF approach led to a somewhat lower bias and higher coefficient of correlation (R^2) in the single-step evaluations. The regression coefficient (b_1) was on average 2% units higher for the MF models than for the QP models suggesting somewhat less overdispersion of GEBVs. These results support the earlier findings that the MF approach may improve the single-step evaluations. The inclusion of external information also improved the evaluations and together with the MF approach gave the best validation results. The blending of foreign information had, in general, a positive effect on the R^2 values but, on the other hand, does not improve the bias problem.

Table 1. Bull LR validation (Bulls=524) results. Regression coefficients (b_1) and coefficient of correlation (R^2) from different models. The b_0= mean(Full_(G)EBV − reduced_(G)EBV). The single-step models: unknown parent groups (UPG) using the QP transformation (GT) or the metafounder (MF) approach (GT_MF), UPG using the QP transformation and the blending of foreign information (GT_B), and the MF approach with the blending of foreign information (GT_B_MF). The animal models: the official model (AM), and the MF approach with the blending of foreign information (AM_B_MF).

	Model	b_0	b_1	R^2
Milk	AM	-101.7	0.84	0.32
	AM_B_MF	-109.86	0.89	0.35
	GT	-319.84	0.87	0.67
	GT_MF	-272.31	0.89	0.68
	GT_B	-306.32	0.86	0.69
	GT_B_MF	-267.07	0.88	0.71
Protein	AM	0.80	0.74	0.24
	AM_B_MF	-0.31	0.83	0.28
	GT	-11.10	0.81	0.63
	GT_MF	-9.71	0.83	0.64
	GT_B	-11.04	0.79	0.64
	GT_B_MF	-9.73	0.82	0.67
Fat	AM	-2.18	0.73	0.23
	AM_B_MF	-2.55	0.80	0.26
	GT	-16.16	0.82	0.64
	GT_MF	-14.67	0.85	0.65
	GT_B	-14.76	0.81	0.66
	GT_B_MF	-13.38	0.84	0.68

The genetic trend of protein for the genotyped Nordic Holstein bulls is presented in Figure 1. The figure shows that with the full data the genetic trends were similar with different single-step models, and the correlations between GEBVs varied from 0.985 to 1.00. The differences between the models can be observed in the reduced data runs. The most obvious difference is seen between the models using the QP transformation and the MF approach. The MF approach model does not increase the genetic trend of young bulls as much as the QP model. Also, the blending of foreign information increases the genetic trend of young bulls, but the increase is not as high when the MF approach is used. The animal model EBVs showed a much lower genetic trend compared to the single-step models, but with the MF approach and the inclusion of MACE information, the b_1 and R^2 also improved in the animal model.

Conclusions

It seems that both the traditional genetic groups and the MF approach can be implemented in ssGTBLUP together with the blending of foreign information. However, it seems that the MF approach can be slightly more efficient in reducing bias in young bulls.

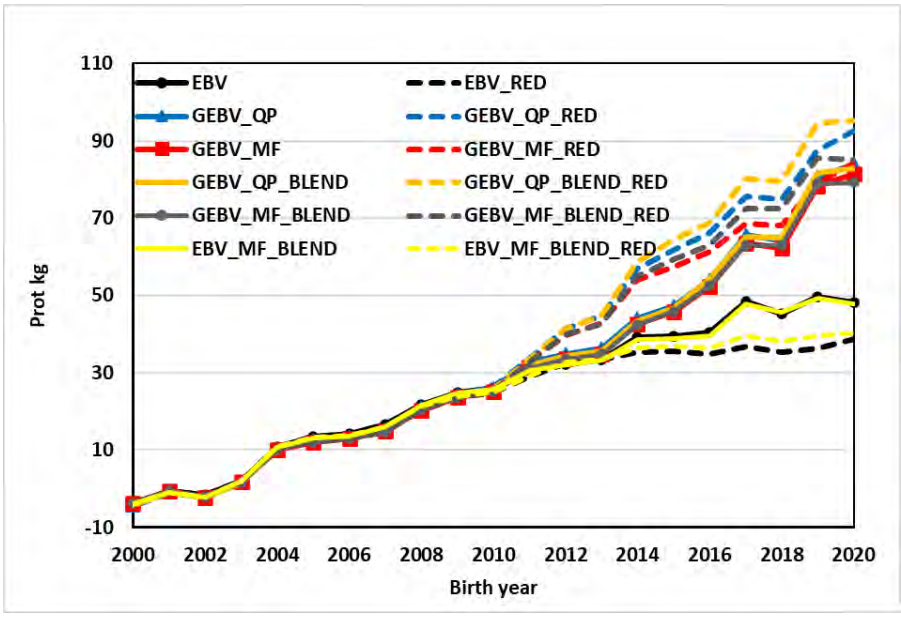


Figure 1. Genetic trends by birth year for protein from different models. The single-step models: unknown parent groups (UPG) using the QP transformation (GT) or the metafounder (MF) approach (GT_MF), UPG using the QP transformation and the blending of foreign information (GT_B) and the MF approach with the blending of foreign information (GT_B_MF). The animal models: the official model (AM), and the MF approach with the blending of foreign information (AM_B_MF). The *RED indicates the reduced data run.

References

Christensen O.F. (2012) Genet. Sel. Evol. 44: 37. https://doi.org/10.1186/1297-9686-44-37.

Garcia-Baccino C. A., Legarra A., Christensen O.F., Misztal I., Pocrnic I., *et al.* (2017) Genet. Sel. Evol. 49: 34. https://doi.org/10.1186/s12711-017-0309-2.

Kirkpatrick, M., Hill W.G. and Thompson R. (1994) Genet. Res. 64: 57–69. https://doi.org/10.1017/S0016672300032559.

Koivula, M., Strandén, I., Aamand, G.P, and Mäntysaari, E.A. (2021) J. Dairy Sci., 104: 10049-10058. https://doi.org/10.3168/jds.2020-19821.

Kudinov A.A., Koivula M., Strandén I., Aamand G.P. and Mäntysaari E.A. (2021) Interbull Bulletin 56:174-179.

Legarra, A., Christensen, O.F., Vitezica, Z.G. Aguilar, I. and Misztal, I. (2015) Genetics 200: 455-468. https://doi.org/10.1534/genetics.115.177014.

Legarra A. and Reverter, A. (2018) Genet. Sel. Evol, 50:53. https://doi.org/10.1186/s12711-018-0426-6.

Lidauer, M., J. Pösö, J. Pederson, J. Lassen, P. Madsen, *et al.* (2015) J. Dairy Sci. 98:1296–1309. https://doi.org/10.3168/jds.2014-8307.

Mäntysaari, E.A., Evans, R.D. and Strandén, I. (2017) J. Anim. Sci. 95:4728-4737. https://doi.org/10.2527/jas2017.1912.

Pitkänen, T.J., Koivula, M.,Strandén, I., Aamand, G.P. and Mäntysaari, E.A. (2020) EAAP Book of Abstracts, Vol. 26 Virtual Meeting.

Schaeffer L. R. (1994) J. Dairy Sci., 77: 2671-2678. https://doi.org/10.3168/jds.S0022-0302(94)77209-X.

Strandén I. and Lidauer M. (1999) J. Dairy Sci. 82:2779-2787. https://doi.org/10.3168/jds.S0022-0302(99)75535-9.

VanRaden P.M. (2008) J. Dairy Sci. 91:4414-4423. https://doi.org/10.3168/jds.2007-0980.

310. Effects of missing or wrong pedigree records on Single-Step predictions

E.C.G. Pimentel*, C. Edel, R. Emmerling and K.-U. Götz

Bavarian State Research Center for Agriculture (LfL), Institute of Animal Breeding, Professor-Dürrwaechter-Platz 1, 85586 Poing-Grub, Germany; eduardo.pimentel@lfl.bayern.de

Abstract

The Single-Step model combines information from pedigree, genotypes and phenotypes at the same time. Inconsistencies between genotypes and pedigree records lead to bias in predictions. This problem has been addressed so far at the level of overall statistics, such as reliabilities, regression coefficients and genetic trends. The purpose of this study was to investigate the effects of pedigree errors at the level of individual predictions. We used a small example pedigree to show how predictions become biased when a pedigree record is wrong or missing. We further used real data to illustrate the common case of wrong or missing maternal grandsire when the dam is not genotyped. When the pedigree record was wrong, estimated breeding values were not much affected but estimated mendelian sampling terms were biased upwards. When it was missing, both estimates were notably biased upwards. Further studies are needed to investigate the cumulative effects of this bias.

Introduction

Genetic evaluations with the Single-Step model combine simultaneously all information from pedigree, genotypes and phenotypes (Legarra *et al.*, 2009; Christensen and Lund, 2010). One issue of major importance in a Single-Step model is the compatibility between the pedigree and the genotypic information, which makes pedigree correctness as crucial as ever before.

Wrong pedigree records are a common problem in livestock populations. Visscher *et al.* (2002) reported pedigree error rates in dairy herds ranging from 5 to 15%. The availability of large-scale genotypic information created the opportunity to correct or fill in wrong or missing pedigree records. Therefore, problems at the level of parent-offspring relationships can be often avoided through checks for genotype conflicts and parent search. So the rate of missing or wrong sires of genotyped individuals should be very low when most or even all potential sires are genotyped, which is often a realistic scenario. In such situations, problems will most commonly occur in the case of missing maternal grandsires (MGS) if the dam is not genotyped, where attempts to discover the MGS often end up with multiple putative MGS.

Misztal *et al.* (2017) suggested the use of scaling factors to set different weights on the genomic and the numerator relationship matrix. According to Tsuruta *et al.* (2019) this could account indirectly for missing pedigree information. However, such scaling parameters operate in a very unspecific way. Tsuruta *et al.* (2019) investigated the effects of missing pedigrees on predictions from the Single-Step model in terms of overall bias and genetic trends. The objective of this study is to investigate and illustrate the effects of missing or wrong pedigree information on Single-Step predictions at the individual animal level.

Materials & methods

Example pedigree. The pedigree depicted in Table 1 was used to construct the covariance matrix H assumed in a Single-Step model:

$$\mathbf{H} = \begin{matrix} \mathbf{A_{11} - A_{12}A_{22}^{-1}A_{21} + A_{12}A_{22}^{-1}GA_{22}^{-1}A_{21}} & \mathbf{A_{12}A_{22}^{-1}G} \\ \mathbf{GA_{22}^{-1}A_{21}} & \mathbf{G} \end{matrix}$$

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_310, © E.C.G. Pimentel *et al.* 2022

Where A is the numerator relationship matrix, G the genomic relationship matrix and subscripts 1 and 2 stand for non-genotyped and genotyped animals, respectively. For the sake of clarity matrix G was set equal to A_{22}. Matrix H was subsequently decomposed in H=TDT'. Breeding values of all animals can be expressed as the product of T by the vector containing breeding values of base animals and mendelian sampling terms of animals with known parents (Kennedy *et al.*, 1988). Matrices H and T were computed for two genotyping scenarios and three scenarios with respect to the MGS of animal 10 when setting up A:

- using the correct pedigree (Table 1), in which animal 3 is the sire of dam 9;
- assigning a wrong sire (animal 2) to dam 9;
- assigning a missing sire to dam 9.

Matrix G was the same in all cases, i.e. equal to A_{22} with the correct pedigree. This way, G always contains the true relationship information, eventually not available in the pedigree.

Real data. Data from the December 2021 run of the German-Austrian routine Single-Step evaluation for conformation traits in Fleckvieh were used. Total number of animals in the pedigree was 4,734,659, from which 1,824,625 had an observation for udder score, for example. Missing pedigree records were assigned to unknown parent groups (UPG), defined according to origin, sex and birth year. Modifications in H to account for UPG are described in Macedo *et al.* (2020). The number of genotyped animals was 401,802. Matrix G^{-1} was set up using the algorithm for proven and young (APY) with 32,866 animals in the core set. Mixed model equations were solved using the software package MiX99 Release XI/2019 version 19.1129 (MiX99 Development Team, 2019). Estimated breeding values (EBV), parent averages (EPA) and mendelian sampling terms (EMS) from the routine run were used as reference and are hereafter referred to as the results obtained with the original pedigree. Ten genotyped animals that had a non-genotyped dam and a genotyped MGS were chosen, such that their genomic relationship coefficient to their MGS were at least 0.24. The MGS of five of them were replaced by UPGs and the other five by some non-genotyped bull born on close to the same year as the original MGS. Analyses were performed again, and results will be referred to as the ones with the modified pedigree. All EBVs were scaled to have a mean of 100 and a standard deviation of 12.

Results

Example pedigree. The row of matrix T corresponding to animal 10 is shown in Table 2 for all simulated scenarios using the example pedigree of Table 1. When the correct pedigree is used to set up A, the 10th row of T contains 0.50 for each parent and 0.25 for each grandparent of animal 10. In the scenario I, an

Table 1. Example pedigree with two genotyping scenarios I and II.

Animal	Sire	Dam	Is genotyped (I)	Is genotyped (II)
1	0	0	Yes	Yes
2	0	0	No	No
3	0	0	Yes	Yes
4	0	0	No	No
5	0	0	No	No
6	0	0	No	No
7	1	4	Yes	Yes
8	2	5	No	Yes
9	3	6	No	No
10	7	9	Yes	Yes

Table 2. Coefficients of the matrix T corresponding to animal 10 for different scenarios.

	Animal									
	1	2	3	4	5	6	7	8	9	10
Scenario I:										
Right MGS	0.25	0.00	0.25	0.25	0.00	0.25	0.50	0.00	0.50	1.00
Wrong MGS	0.25	0.25	0.23	0.25	0.00	0.23	0.50	0.00	0.47	1.00
Missing MGS	0.25	0.00	0.25	0.25	0.00	0.23	0.50	0.00	0.47	1.00
Scenario II:										
Right MGS	0.25	0.00	0.25	0.25	0.00	0.25	0.50	0.00	0.50	1.00
Wrong MGS	0.25	0.20	0.24	0.25	-0.06	0.24	0.50	-0.10	0.48	1.00
Missing MGS	0.25	0.00	0.25	0.25	0.00	0.23	0.50	0.00	0.47	1.00

additional coefficient of 0.25 to animal 2 appears when it is wrongly assigned as the MGS of 10, whilst the coefficient to the true MGS is still 0.23 due to the information available in G. This means that there could be almost a full contribution from the right MGS plus a full contribution from the false MGS to the breeding value of animal 10. Final effective contributions will of course depend on the structure of phenotypes and the availability of mendelian sampling information.

When the MGS is set to missing the coefficient to the true MGS (animal 3) is still 0.25, again because this information is available in G. In scenario II animal 8 is also genotyped. The coefficient to the wrongly assigned MGS drops to 0.20 and the coefficients to animal 8 and its dam are no longer 0. This means that the extent to which the information from the wrong MGS to the breeding value of 10 can be transmitted depends on which other animals in the pedigree are genotyped, i.e. on how much and which information is available in G.

Real data. Results from the tests with real data are presented in Table 3. In the cases where the MGS was assigned to a wrong bull, EBVs from the runs with the original and the modified pedigree did not differ much.

Table 3. Estimated breeding value (EBV), parent average (EPA) and mendelian sampling term (EMS) from the real data with the original and the modified pedigree.

Type of modification	Original pedigree			Modified pedigree		
	EBV	EPA	EMS	EBV	EPA	EMS
Wrong MGS	117.00	112.28	4.72	117.24	109.67	7.57
Wrong MGS	115.04	114.40	0.64	115.37	111.65	3.72
Wrong MGS	110.47	115.30	-4.83	111.05	111.28	-0.22
Wrong MGS	131.09	119.62	11.47	130.95	115.00	15.95
Wrong MGS	130.06	127.23	2.83	129.65	119.81	9.84
Missing MGS	117.50	115.83	1.67	118.80	108.93	9.87
Missing MGS	113.22	114.16	-0.94	118.43	113.33	5.10
Missing MGS	114.52	114.28	0.24	119.52	111.46	8.06
Missing MGS	132.83	125.82	7.01	135.78	121.56	14.21
Missing MGS	130.81	117.93	12.88	135.73	118.50	17.23

Due to the lower genetic level of the non-genotyped bulls assigned as wrong MGS compared with the original genotyped MGS, EPA were considerably lower with the modified pedigree. This consequently resulted in considerably larger EMS. When the MGS was set to missing, EBVs from the run with the modified pedigree were overestimated by a quarter or even close to half a standard deviation. The EMS was also overestimated, and this time not only due to lower EPA with the modified pedigree, as illustrated in the last row of Table 3.

Discussion

The exercise with the small example showed that the EBV of an animal may receive two contributions from the MGS when the information about it in A and G do not match. In scenario I, the amount contributed by the wrong MGS was 0.25, making it a full additional MGS contribution. Scenario II showed that depending on the structure of the genotyped population the amount contributed by the wrong MGS may decrease. We did the same exercise with other example pedigrees and other genotyping scenarios and observed that the extra contribution from the wrong MGS may become as low as 0.02 (data not shown). This could explain the small differences in EBV observed with the real data for wrong MGS. Nevertheless, differences in EMS were large and may well depict a common phenomenon in practical applications. A wrong MGS record would only remain in pedigree when a conflict test cannot be done, and the genetic level of such ungenotyped bulls should be usually low.

When the true MGS is genotyped but could not be unambiguously identified in a discovery trial, this information is missing in A but is available in G. Therefore, the coefficient to the true MGS in T was still 0.25 in the small example even though it was set to missing. In a model with UPG, final EBVs are the summation of the estimated animal effects and group solutions weighted by corresponding group contributions. If the information from the true MGS is accounted for in G, the extra contribution from an UPG supposed to replace him will cause information from the MGS being transmitted twice, incurring upward bias in the EBV. Results from the real data were in agreement with this expectation. There was not only a biased partition into EPA and EMS, but a notable bias in the EBV themselves.

The observed patterns of overestimation of EMS of animals with wrong or missing MGS could be at least partly an explanation for the inflation in prediction of young animals often reported in the literature (e.g. Tsuruta *et al.*, 2019). Further studies will be conducted to investigate the cumulative effect of the patterns observed here and its potential implications.

References

Christensen O.F., and Lund M.S. (2010) Genet. Sel. Evol. 42:2. https://doi.org/10.1186/1297-9686-42-2

Kennedy B.W., Schaeffer L.R., and Sorensen D.A. (1988) J. Dairy Sci. 71(Suppl. 2):17-26. https://doi.org/10.1016/S0022-0302(88)79975-0

Legarra A., Aguilar I., and Misztal I. (2009) J. Dairy Sci. 92(9):4656-4663. https://doi.org/10.3168/jds.2009-2061

Macedo F.L., Christensen O.F., Astruc J.-M., Aguilar I., Masuda Y. *et al.* (2020) Genet. Sel. Evol. 52:47. https://doi.org/10.1186/s12711-020-00567-1

Misztal I., Bradford H.L., Lourenco D.A.L., Tsuruta S., Masuda Y. *et al.* (2017) Interbull Bulletin 51:38-42.

MiX99 Development Team (2019) MiX99: A software package for solving large mixed model equations. Release XI/2019. Natural Resources Institute Finland (Luke), Jokioinen, Finland.

Tsuruta S., Lourenco D.A.L., Masuda Y., Misztal I., and Lawlor T.J. (2019) J. Dairy Sci. 102(11):9956-9970. https://doi.org/10.3168/jds.2019-16789

Visscher P.M., Woolliams J.A., Smith D., and Williams J.L. (2002) J. Dairy Sci. 85(9):2368-2375. https://doi.org/10.3168/jds.S0022-0302(02)74317-8

311. An alternative method to consider a reference population in Single-Step SNP BLUP model without separating genomic terms

T. Osawa[1*], T. Baba[2], Y. Masuda[3], T. Kawahara[2] and Y. Goto[2]

[1]National Livestock Breeding Center, Nishigo-mura, Fukushima, 961-8511, Japan; [2]Holstein Cattle Association of Japan, Hokkaido Branch, Sapporo, Hokkaido, 001-0015, Japan; [3]Rakuno Gakuen University, Ebetsu, Hokkaido, 069-8501, Japan; t0ohsawa@nlbc.go.jp

Abstract

This study proposed an alternative method to consider a reference population in the mixed model equations (MME) of single-step SNP BLUP (ssSNP-BLUP) model without separating genomic terms to estimate SNP effects. We adopted the preconditioned biconjugate gradient stabilized method to solve MME because this model was a nonsymmetric linear system. Overall conformation score in the first lactation for Japanese Holsteins was used to compare the original and modified ssSNP-BLUP models. Genotyped animals included in the dataset were 37,197 cows, 5,352 sires, 3,973 young bulls and 88,058 heifers. Three reference populations were defined as sires, cows and both. Our method can consider the reference population in the ssSNP-BLUP without separating genomic terms. The contribution of young animals to genomic predictions was small, and predictions varied when only sires or cows were considered in a reference population. On the other hand, considering the reference population may reduce the overestimation of the predictions.

Introduction

Currently, genetic evaluations of dairy cattle in many countries include large-scale SNP marker information, known as genomic evaluation. Most countries have applied a multi-step model for the genomic evaluation of dairy cattle (VanRaden., 2008), which utilizes deregressed estimated breeding values (EBV) obtained via prior conventional genetic evaluation based on pedigree information. The multi-step model is feasible to implement in conventional genetic evaluation systems. A major advantage of the multi-step method is its low computational cost. It uses only selected genotyped animals such as proven sires and cows with their own records as reference population to estimate SNP effects. However, the separate steps of estimating SNP effects and EBV during conventional evaluation cannot fully account for genomic preselection; therefore, genomic EBV (GEBV) can be biased (Patry and Ducrocq, 2011). In contrast to the multi-step model, the single-step genomic BLUP (ssGBLUP) model can evaluate genotyped and non-genotyped animals jointly, thus providing unbiased predictive value (Aguilar *et al.*, 2010). The ssGBLUP integrates the genomic relationship matrix (G) and the pedigree relationship matrix (A) into the hybrid matrix and replaces the pedigree relationship matrix in the mixed model equations (MME) of BLUP. On the other hand, Liu *et al.* (2014, 2016) developed a single-step SNP BLUP (ssSNP-BLUP) model equivalent to ssGBLUP in theory. The model does not require the constructions of G and its inversion and can estimate GEBV and SNP effects simultaneously. Liu *et al.* (2014) proposed a computing strategy by separating genomic terms from the MME of ssSNP-BLUP and solving two separate sets of equations. They also showed how to consider reference animals, as defined in the multi-step model by expanding the equations associated with genomic term. However, since the calculation of the two separated equations is repeated alternately, the program may not be easy to implement for complex models.

The objectives of this study were to propose a method that considers a reference population for ssSNP-BLUP model without separating the genomic terms to estimate SNP effects and compare the predictions using different reference populations.

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_311, © T. Osawa *et al.* 2022

Materials & methods

Equation for original ssSNP-BLUP model (ssORG) by Liu *et al.* (2014) can be written as follows,

$$\begin{bmatrix} \mathbf{X'R^{-1}X} & \mathbf{X'R^{-1}W_1} & \mathbf{X'R^{-1}W_2} & \mathbf{0} \\ \mathbf{W_1'R^{-1}X} & \mathbf{W_1'R^{-1}W_1} + \sigma_u^{-2}\mathbf{\Sigma^{11}} & \mathbf{W_1'R^{-1}W_2} + \sigma_u^{-2}\mathbf{\Sigma^{12}} & \sigma_u^{-2}\mathbf{\Sigma^{1g}} \\ \mathbf{W_2'R^{-1}X} & \mathbf{W_2'R^{-1}W_1} + \sigma_u^{-2}\mathbf{\Sigma^{21}} & \mathbf{W_2'R^{-1}W_2} + \sigma_u^{-2}\mathbf{\Sigma^{22}} & \sigma_u^{-2}\mathbf{\Sigma^{2g}} \\ \mathbf{0} & \sigma_u^{-2}\mathbf{\Sigma^{g1}} & \sigma_u^{-2}\mathbf{\Sigma^{g2}} & \sigma_u^{-2}\mathbf{\Sigma^{3g}} \end{bmatrix} \begin{bmatrix} \hat{\mathbf{b}} \\ \hat{\mathbf{u}}_1 \\ \hat{\mathbf{u}}_2 \\ \hat{\mathbf{g}} \end{bmatrix} = \begin{bmatrix} \mathbf{X'R^{-1}y} \\ \mathbf{W_1'R^{-1}y} \\ \mathbf{W_2'R^{-1}y} \\ \mathbf{0} \end{bmatrix} \quad (1)$$

where y is a vector of observations, $\hat{\mathbf{b}}$ is a vector of fixed effects, subscripts 1, 2 and g refer to groups of non-genotyped, genotyped animals and SNP effects, respectively, $\hat{u}_1$ and $\hat{u}_2$ are vectors of additive genetic effects of the non-genotyped and genotyped animals, respectively, $\hat{\mathbf{g}}$ is a vector of SNP effects, X is a design matrix for fixed effects, W_1 and W_2 are design matrices for additive genetic effects of non-genotyped and genotyped animals, respectively, R is $\sigma_e^2\mathbf{I}$, σ_u^2 and σ_e^2 are an additive genetic and residual variance.

$$\Sigma = \begin{bmatrix} \mathbf{\Sigma^{11}} & \mathbf{\Sigma^{12}} & \mathbf{\Sigma^{1g}} \\ \mathbf{\Sigma^{21}} & \mathbf{\Sigma^{22}} & \mathbf{\Sigma^{2g}} \\ \mathbf{\Sigma^{g1}} & \mathbf{\Sigma^{g2}} & \mathbf{\Sigma^{gg}} \end{bmatrix} = \begin{bmatrix} \mathbf{A^{11}} & \mathbf{A^{12}} & \mathbf{0} \\ \mathbf{A^{21}} & \mathbf{A^{22}} + \left(\frac{1}{k} - 1\right)\mathbf{A_{22}^{-1}} & -\frac{1}{k}\mathbf{A_{22}^{-1}Z} \\ \mathbf{0} & -\frac{1}{k}\mathbf{Z'A_{22}^{-1}} & \frac{m}{1-k}\mathbf{I} + \frac{1}{k}\mathbf{Z'A_{22}^{-1}Z} \end{bmatrix} \quad (2)$$

where A^{ij} are submatrices of the inverse of the pedigree relationship matrix, consisting of non-genotyped (1) and genotyped (2), k is the proportion of the additive genetic variance due to the residual polygenic effects (k was set to 0.1 in this analysis), and m is $2\Sigma p_j\,(1 - p_j)$, with p_j being the observed allele frequency of the j^{th} SNP marker and Z is a design matrix of regression coefficients on genotyped animals at all SNP markers (2-$2p_j$, 1-$2p_j$ or -$2p_j$ for genotype AA, AB, or BB of the j^{th} SNP marker).

Single-Step SNP BLUP model with the reference population. Liu *et al.* (2014) described the genomic term to estimate SNP effects with the reference population as follows.

$$\hat{\mathbf{g}} = \frac{1}{k}\frac{1-k}{m}\mathbf{IZ'FA_{22}^{-1}}\hat{\mathbf{a}}_2 \quad (3)$$

$$\hat{\mathbf{a}}_2 = \hat{\mathbf{u}}_2 - \mathbf{Z}\hat{\mathbf{g}} \quad (4)$$

where F is a filter matrix which is *diag*{1,0,0,1, … 1,1,1,0} to define genotyped animals as reference (=1) or not (=0). Substituting Equation 4 into Equation 3:

$$\hat{\mathbf{g}} = \frac{1}{k}\frac{1-k}{m}\mathbf{IZ'FA_{22}^{-1}}(\hat{\mathbf{u}}_2 - \mathbf{Z}\hat{\mathbf{g}}) \quad (5)$$

Equation 5 can be rearranged as follows:

$$\left[-\frac{1}{k}\mathbf{Z'FA_{22}^{-1}} \quad \frac{m}{1-k}\mathbf{I} + \frac{1}{k}\mathbf{Z'FA_{22}^{-1}Z}\right]\begin{bmatrix}\hat{\mathbf{u}}_2 \\ \hat{\mathbf{g}}\end{bmatrix} = \mathbf{0} \quad (6)$$

Replace $\mathbf{\Sigma^{g2}}$ and $\mathbf{\Sigma^{gg}}$ in Equation 2 by Equation 6 as follows:

$$\mathbf{\Sigma^{g2}} = -\tfrac{1}{k}\mathbf{Z'FA_{22}^{-1}} \text{ and } \mathbf{\Sigma^{gg}} = \tfrac{m}{1-k}\mathbf{I} + \tfrac{1}{k}\mathbf{Z'FA_{22}^{-1}Z} \quad (6)$$

Therefore, the equation of ssSNP-BLUP model with reference population (ssRP) is a nonsymmetric linear system.

Solving algorithm. ssORG is applicable with the preconditioned conjugate gradient method. However, ssRP results in a nonsymmetric linear system, and we adopted the preconditioned biconjugate gradient

stabilized method (Van der Vorst, 1992). The calculation programs introduced an 'iteration on data' algorithm (Strandén and Lidauer., 1999 and Tsuruta *et al.*, 2001) and used 10 OpenMP threads. The preconditioner matrix combined the Jacobi preconditioner and a second-level diagonal preconditioner (Vandenplas *et al.*, 2019). Assuming that Equation 1 is written as Cx = b, the convergence criterion is defined as $\|b - Cx\|^2 / \|b\|^2 < 10^{-14}$, which $\|\cdot\|$ is the 2-norm. In addition, the average of the last 30 iteration rounds must be less than 10^{-14} in the biconjugate gradient stabilized method.

Data. We used overall conformation score in the first lactation for Japanese Holsteins. The additive genetic variance, residual variance and heritability of the trait were 0.685, 1.852 and 0.27, respectively. The dataset comprised 762,650 cows with record and 8,745 sires with daughters; of those, 37,197 cows and 5,352 sires were genotyped. The dataset also included 88,058 genotyped heifers and 3,973 genotyped young bulls. After the quality control, 39,756 SNP markers were considered in this study. We used three reference populations: sires with daughters (ssRP_S), cows with the record (ssRP_C) and ssRP_SC with both of them.

Results

In ssORG, 821 iterations were needed to stratify convergency criteria, and the total computing time was 94 minutes. For ssRP_S, ssRP_C and ssRP_SC, although the number of iterations was less than ssORG (424 to 600), the total times was slightly longer (106 to 140 minutes).

Correlations and regression coefficients. Pearson correlations in the solutions between ssORG and ssRP and regression coefficients of ssORG on ssRP for SNP effects and GEBV for genotyped sires, cows, young bulls and heifers are showed in Table 1. Correlations in ssRP_SC were closed to 1.0, but ssRP_S and ssRP_C were lower than ssRP_SC. Therefore, although the contribution of young animals to the predictions was small, the composition of reference population affected the predicted values. Since the regression coefficients were smaller than 1.0 for almost any combinations, considering reference population may reduce the overestimation of the predictions.

Genetic trends of genotyped animal. Genetic trends of genotyped sires and young bulls for ssORG and ssRP are shown in Figure 1. Similar trends were obtained for both sires and young bulls in ssORG, ssRP_S and ssRP_SC. However, the genetic trends of ssRP_C were flattened in older generations of sires and young bulls, compared to other reference populations.

Table 1. Correlations between ssORG[1] and ssRP[2] and regression coefficients of ssORG on ssRP for SNP effects and GEBV for genotyped sire, cows, young bulls and heifers.

		Correlations			Regression coefficients		
Items	**n**	**ssRP_S**	**ssRP_C**	**ssRP_SC**	**ssRP_S**	**ssRP_C**	**ssRP_SC**
SNP effects	39,756	0.531	0.790	0.967	0.384	0.740	0.950
Sires	5,332	0.979	0.975	0.999	0.979	0.881	1.001
Cows	37,197	0.894	0.962	0.995	0.838	0.931	0.987
Young bulls	3,973	0.865	0.915	0.992	0.852	0.818	0.985
Heifers	88,058	0.880	0.936	0.993	0.838	0.883	0.972

[1] Original single-step SNP BLUP model.

[2] Single-step SNP BLUP model with reference populations (S=sires, C=cows and SC=sires and cows).

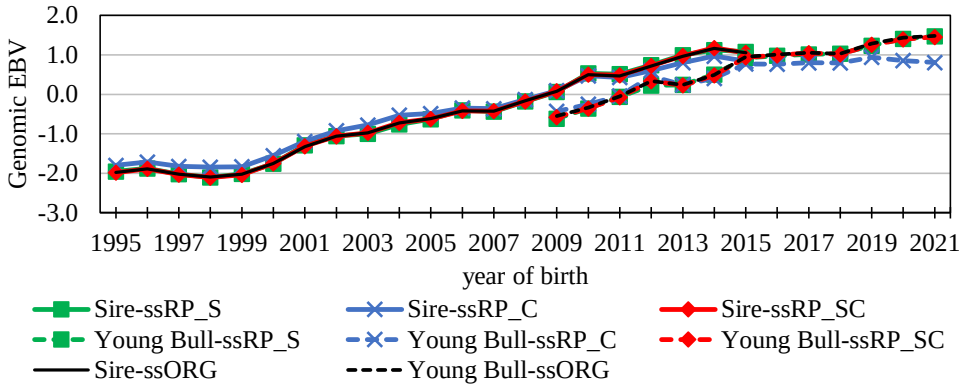


Figure 1. Genetic trends of genotyped sires and young bulls by models. ssORG = Original single-step SNP BLUP model, ssRP = Single-step SNP BLUP model with reference populations (S=sires, C=cows and SC=sires and cows).

Discussion

The proposed method can consider the reference animals without separating the genomic term from the MME in ssSNP-BLUP. This model is easily extensible to complex models e.g. multi-trait and random regression models (results is not shown). In addition, this method can be helpful to understand the influence of the reference population on GEBV and SNP effects in the single-step method. In the results of this analysis, although the contribution of young animals to both GEBV and SNP effects was small, the predictions varied greatly when only sires or cows were considered as reference population. Thus, the composition of the reference population may also need to be carefully considered in the single-step methods. Considering the reference population may reduce the overestimation of the predictions. However, the genetic trends were flattened when only cows were considered in reference population. This may be because the fact that genotyped cows were only relatively new generations. Further studies are needed to investigate the prediction accuracy using the method.

References

Aguilar I., Misztal I., Johnson D.L., Legarra A., Tsuruta S. and Lawlor T.J. (2010) J. Dairy Sci. 93:743–752. https://doi.org/10.3168/jds.2009-2730

Liu, Z., Goddard M.E., Hayes B.J., Reinhardt F., and F., and Reents R. (2016) J. Dairy Sci. 99:2016-2025. https://doi.org/10.3168/jds.2015-10394

Liu, Z., Goddard M.E., Reinhardt F., and Reents R. (2014) J. Dairy Sci. 97:5833-5850. https://doi.org/10.3168/jds.2014-7924

Patry C., and Ducrocq V. (2011) J. Dairy Sci. 94:1011–1020. https://doi.org/10.3168/jds.2010-3804

Strandén, I., and Lidauer M. (1999) J. Dairy Sci. 82:2779-2787. https://doi.org/10.3168/jds.S0022-0302(99)75535-9

Tsuruta, S., Misztal I., and Strandén I. (2001) J. Anim. Sci. 79:1166-1172. https://doi.org/10.2527/2001.7951166x

Van der Vorst H. (1992) SIAM J. Sci. and Stat. Comput. 13:631–644. https://doi.org/10.1137/0913035

Vandenplas J., Calus M.P.L., Eding H., and Vulik C. (2019) Genet. Sel. Evol. 51:30 https://doi.org/10.1186/s12711-019-0472-8

VanRaden, P. M, 2008. Efficient Methods to Compute Genomic Predictions. J. Dairy Sci. 91:4414-4423. https://doi.org/10.3168/jds.2007-0980

312. Accuracy of pedigree and genomic predictions of meat quality in a multi-breed cattle using single and two step model

S. Naderi[1], R. Evans[1*] and D.P. Berry[2]

[1]The Irish Cattle Breeding Federation, Highfield House, Bandon, Co. Cork, Ireland; [2]Department of Animal Bioscience, Animal and Grassland Research and Innovation Centre, Teagasc, Moorepark, Fermoy, Co. Cork, Ireland; revans@icbf.com

Abstract

The objective of this study was to develop single step genomic breeding values (ssGEBV) for meat tenderness, juiciness and flavour and compare with two-step genomic breeding values as well as non-genomic (i.e. traditional) breeding values (EBV); the comparison also involved testing the impact on the accuracy of the (G)EBVs by altering the number of half sibs in the training population. A residual polygenic single step GBLUP and a two-step genomic approach was applied to 5,262 genotyped and phenotyped animals. Including genomic information improved the prediction accuracy and this accuracy also improved as the number of half sib-progeny in the training population increased. Moreover, applying single step genomic approaches were associated with further improvements in prediction performance across all traits.

Introduction

Meat eating quality (MEQ) impacts repeat purchase of meat by consumers; this is particularly important for beef meat which is often more variable in quality than many other meat types. MEQ is often depicted by meat tenderness, juiciness, and flavour. Whereas heterogeneity is often viewed as an inconvenience, animal breeders require variability to achieve genetic gain. Nonetheless, genetic gain eventually translates into phenotypic changes which is a function of how well estimates of genetic merit translate into expected phenotypic differences but also the acceptance of stakeholders that (products from) individuals divergent in genetic merit will truly be reflective in phenotypic differences. Validation of genetic evaluations is therefore key to stakeholder acceptance. The accuracy of estimates of genetic merit can be improved using widely available genomic information. Integrating phenotypic, pedigree and genomic data to achieve high prediction performance is a modelling challenge. Therefore, the objective of the present study was to evaluate alternative strategies of combining these data to predict genetic merit and, by extension, phenotypic performance.

Materials & methods

Data. Tenderness, juiciness and flavour sensory data were available on 5,912 crossbred cattle consisting of 2,137 steers, 2,093 heifers and 1,682 young bulls. The main breed proportion of the animals was Angus (27%), Belgian Blue (5%), Charolais (12%), Holstein-Friesian (4%), Hereford (14%), Limousin (21%) and Simmental (8%). The phenotyped population was constructed to be as diverse as possible with the aim of a maximum of 20 progeny per sire. The generation of the meat sensory data including the standard operating procedures are described in detail by Judge *et al.* (2021) for a subset of the data set used. Each of the three traits were scored on a 1 to 10 scale: 1 = very tough and 10 = very tender; juiciness: 1 = not juicy and 10 = very juicy; beef flavour: 1 = no beef flavour and 10 = very strong beef flavour. Of the 5,912 animals with meat quality phenotypes, 5,262 were genotyped with the International Dairy and Beef genotyping platform version3 (IDBV3; 50,855 SNPs). The number of animals in the pedigree was 95,674.

Estimation of breeding values. Traditional breeding values (EBVs) as well as genomic breeding values (GEBVs) were estimated for three traits using the following linear mixed model:

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_312, © S. Naderi *et al.* 2022

$$y = Xb + Wu + e$$

where y is a vector of one of the sensory phenotypes for all animals in training set, X is the incidence matrix for fixed effects, b is the vector of fixed effects consisting of herd-by-date of slaughter, date-by-location of sensory analysis, gender of the animal, the age in months since slaughter of the meat sample, heterosis proportion class, month of analysis and breed proportion, W is the incidence matrix of random (genomic) effects, u is a vector of EBVs or GEBVs for all individuals, and e is a vector of random residuals. For the traditional genetic evaluation, it was assumed that $u \sim N\ (0,\ \boldsymbol{A}\sigma_u^2)$ where σ_u^2 is the additive genetic variance and A is pedigree relationship matrix. For the estimation of GEBV using the GBLUP model, the A matrix was replaced with a genomic relationship matrix. In the single-step GBLUP model, the A matrix was replaced with a H matrix where H was the combined relationship matrix (Christensen *et al.* 2012). The H matrix is constructed by blending the pedigree relationship matrix with the genomic relationship matrix) considering information from non-genotyped and genotyped animals simultaneously. The inverse of H-matrix was calculated according to Christensen and Lund (2010) and Aguilar *et al.* (2010):

$$\mathbf{H^{-1}} = \mathbf{A^{-1}} + \begin{bmatrix} 0 & 0 \\ 0 & \boldsymbol{G_w^{-1}} - \mathbf{A_{22}^{-1}} \end{bmatrix}$$

where A^{-1} was the inverse of the pedigree-based relationship matrix, $\mathbf{A_{22}^{-1}}$ was the inverse of the sub-matrix of the pedigree-based relationship matrix for genotyped animals and G_w was calculated as follow:

$$G_w = (\alpha \times G + \beta \times A_{22})$$

where G_w was the genomic relationship matrix, $\beta = 1 - \alpha$ where α=0.70. For all the computations, the MiX99 software (MiX99 development team, 2020) was used.

The GEBVs from the two-step genomic evaluation are the result of blending the Direct Genomic Value (DGV) which was calculated by multiplying the SNP effects with the animals' genotypes, traditional Parent Average (PA) and the EBV_{ref} from a traditional evaluation including only relationships among genotyped animals. The method for blending three sources of information (i.e. DGV, PA and EBVref.) to calculate the GEBV are described in detail by VanRaden *et al.* (2009).

Training and testing population design. Three different scenarios were explored on a group of validation animals to assess prediction accuracy. First a group of 49 'well' proven sires with at least 25 progeny was identified. To build a validation set, a random 12 progeny from each of the 49 sires was sampled (i.e. 588 animals) with the remaining progeny of these sires, who did not appear in the validation set, being hereafter referred to as the *remaining set*. In the first scenario (S_I), none of the progeny from the well-proven sires were included in the training population. In the second scenario (S_II), a random seven progeny of each well proven sire from the *remaining set* was added to the training set; in the third scenario (S_III), all the *remaining set* progeny were added to the training set; only the phenotypic records of validation set were masked in this scenario.

Prediction accuracy. The correlation between (G)EBV and the phenotypes (following adjustment for all non-genetic effects) of the validation animals was considered to represent the prediction accuracy at the animal level for all three scenarios. In addition, to assess prediction performance at sire level, the correlation between validation sires' (G)EBVs from S_I in which none of their progeny were in training set with their daughter yield deviation (DYD) was calculated. DYD were calculated based on the average yield deviation from the progeny of each sire including all phenotypic records (full data set).

Results & discussion

Validation results from (single)two-step genomic evaluation as well as traditional EBVs for three scenarios at the animal level are presented in Table 1. Including genomic information and applying a single or two step genomic approach improved the prediction accuracy for all three traits compared with the EBVs from the pedigree-based relationship matrix. For example, the prediction accuracy for flavour increased from 0.09 to 0.17 following the incorporation of genomic information into the model and using two-step approach (i.e. S_III) with a similar pattern detected for the two other traits. Further increases in prediction accuracy were observed by implementing the single step genomic approach compared with the two-step method for all three traits and scenarios. Aguilar *et al.* (2010) demonstrated that ssGBLUP is a quick alternative to estimate GEBV when phenotypes, genotypes and pedigree are jointly available arguing that the automatic and proper definition of weighting factors to blend different sources of information were the main advantage of implementing this approach. In addition, Legarra *et al.* (2014) indicated that the appropriate weight to integrate various source of information avoids double counting of contributions emerging from genetic relationships and phenotypic records.

By increasing number of half sibs in the training population (i.e. S_I to S_III), the accuracy increased for all three prediction approaches (i.e. traditional, (single)two-step GEBV). This improvement could be explained by increasing the genetic relationships between training and validation sets but also an enlargement of the training population size. De Los Campos *et al.* (2013) indicated that the genetic relationships between training and validation set as well as training population size were the most important factors affecting prediction performance for GBLUP.

Validation results at the sire level for the three different prediction approaches for S_I in which none of the identified sires' progeny were included in the training population are shown in Table 2. The results demonstrate that incorporating genomic information and applying the (single)two-step genomic approach improved the prediction accuracy for all three traits. For instance, the prediction accuracy for tenderness increased from 0.17 to 0.38 following the inclusion of genomic information in the model and applying a two-step genomic model.

Using single step genomic approaches generated further improvements in prediction accuracies for all three traits. Across all traits, the slope of DYD on (G)EBV of the validation sires where the expectation of the slope was 0.5, improved with the inclusion of genomic information in the model and improvement further when a single step genomic approach was applied compared to a two-step approach indicating ssGBLUP could reduce potential biases.

Table 1. Correlations between the (genomic) estimated breeding value and adjusted phenotype for the three different scenarios at the animal level from three different approaches.

	EBV			Two-step GEBV			Single step GEBV		
Traits	**S_I[1]**	**S_II[2]**	**S_III[3]**	**S_I**	**S_II**	**S_III**	**S_I**	**S_II**	**S_III**
Tenderness	0.03	0.09	0.13	0.09	0.11	0.15	0.11	0.14	0.17
Flavor	0.09	0.14	0.20	0.17	0.18	0.22	0.21	0.21	0.26
Juiciness	0.008	0.06	0.12	0.08	0.10	0.14	0.09	0.10	0.15

[1] S_I: none of the progeny from the well-proven sires were in the training population.

[2] S_II: a random seven progeny of each well proven sire from the *remaining set* was included in the training set.

[3] S_III: all the *remaining set* progeny were included in the training set.

Table 2. Correlations between the (genomic) estimated breeding value and daughter yield deviations (DYD) and the slope of the DYD on the (G)EBV from three different approaches.

Traits	Statistic	EBV	Two-step GEBV	Single step GEBV
Tenderness	Correlation	0.17	0.38	0.42
	Slope	0.12	0.31	0.39
Flavor	Correlation	0.13	0.36	0.40
	Slope	0.10	0.34	0.36
Juiciness	Correlation	0.16	0.34	0.41
	Slope	0.09	0.27	0.32

References

Aguilar I., Misztal I., Johnson D. L., Legarra A., Tsuruta S. and *et al.* (2010) J. Dairy Sci. 93 (2): 743–752. https://doi.org/10.3168/jds.2009-2730.

Christensen O. F. and Lund M. S. (2010). Genet. Sel. Evol. 42 (2). https://doi.org/10.1186/1297-9686-42-2.

Christensen O. F., Madsen P., Nielsen B., Ostersen T. and Su G. (2012). Animal 6 (10): 1565–1571. https://DOI.10.1017/S1751731112000742.

de Los Campos G., Hickey J. M., Pong-Wong R., Daetwyler H. D. and Calus M. P. L. (2013) Genetics. 193 (2): 327–345. https://doi.org/10.1534/genetics.112.143313.

Judge M. M., Conroy S., Hegarty P. J., Cromie A. R., Fanning R and *et al.* (2021). Meat Science, 172: 108371. https://doi.org/10.1016/j.meatsci.2020.108371.

Legarra A., Chistensen O.F., Aguilar I. and Misztal I. (2014). Livest. Prod. Sci. 166: 54–65. https://doi.org/10.1016/j.livsci.2014.04.029

MiX99 Development Team, (2020) MiX99: Natural Resources Institute Finland (Luke). Jokioinen, Finland. URL: http://www.luke.fi/mix99.

VanRaden P., Van Tassell C., Wiggans G., Sonstegard T., Schnabel R. and *et al.* (2009) J. Dairy Sci. 92 (1): 16-24. https://doi.org/10.3168/jds.2008-1514

313. Single-step genomic prediction in small breeds: Finncattle case

A.A. Kudinov[1*], M. Koivula[1], I. Strandén[1], G.P. Aamand[2] and E.A. Mäntysaari[1]

[1]Natural Resources Institute Finland (Luke), 31600 Jokioinen, Finland; [2]Nordic Cattle Genetic Evaluation (NAV), 8200 Aarhus, Denmark; andrei.kudinov@luke.fi

Abstract

Genomic prediction is of great interest in small breeds. When breed-wise populations cannot exchange data, the establishment of a multi-breed evaluation is a tempting option. Single-step genomic prediction is a useful tool when limited data is available. In the current study, we developed a single-step genomic prediction for the small indigenous Finncattle breed with or without data from national major breed Nordic Red Dairy cattle (RDC). The reliability of genomic prediction was assessed using modified Interbull and Legarra and Reverter validation approaches. Inclusion of the RDC data to Finncattle evaluations helped to improve the quality of prediction in traditional and genomic models. The highest prediction reliability for milk and protein was obtained using the multi-breed single-step model.

Introduction

Genomic prediction has been widely used as a two-step procedure in large dairy cattle populations since 2009. The use of DNA markers allows predicting the breeding value of a candidate animal with high reliability at an early age (VanRaden, 2020). However, in small breeds, the use of the two-step genomic prediction approach might be challenging due to the limited data available. Thus, the single-step method (Christensen and Lund, 2010) would be the best way to perform genomic prediction with a limited number of genotyped animals and phenotypic records (Song *et al.*, 2019). A widely reported procedure to increase the power of genomic prediction is the exchange of genomic, phenotypic, or MACE (Multi-trait Across-Country Evaluations, Interbull, Sweden) information between countries (Kudinov *et al.*, 2021a). Naturally, the exchange can only be done with the common breeds and cannot be used when the breed is unique. In such a situation multi-breed evaluation might be the solution. Usually when multi-breed evaluations are used, the solutions for SNP effects are typically estimated by breed as the two-step procedure. The applicability of the multi-breed single-step genomic prediction model has not been studied extensively.

Finncattle (FIC) is the local Finnish breed. Breeding values for FIC have been evaluated by Nordic Cattle Genetic Evaluation (NAV) in the same evaluation with Nordic Red Dairy Cattle (RDC) and Finnish Holstein (Fin HOL, Kudinov *et al.*, 2021b). Multi-breed data is used because many of the herds raise cows from the three breeds in the same facilities and feeding. The current genomic evaluation for RDC animals uses the two-step approach with pure RDC animals. Because genotyping of FIC breed has started only recently, the development of FIC genomic prediction models has become topical. Assuming the benefits and willingness of Nordic countries to run genomic prediction using a single-step model, the development of a multi-breed single-step model was started.

The focus of the current study was to assess the accuracy of genomic prediction for FIC cows with and without RDC phenotypic and genomic information. We performed validation of the model fit by comparing results from approaches suggested by Legarra and Reverter (2018) and modified Mäntysaari *et al.* (2010).

Materials & methods

Data. Pedigree and phenotypic records were obtained from the August 2020 NAV RDC milk and protein evaluations. Pedigree included 4,741,559 RDC, 1,047,697 Fin HOL, 36,133 FIC, and 101,144 other breed animals. There were 137 unknown parent groups (UPG) formed using breed by country by time classes.

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_313, © A.A. Kudinov *et al.* 2022

Phenotypic data was presented by 305-day milk and protein first lactation records from 3,468,530 RDC, 811,363 FIN HOL, and 27,810 FIC cows. Genomic data include genotypes from 187 FIC and 35,729 RDC bulls, and 756 FIC and 133,107 RDC cows. Raw genotypes were imputed and edited by NAV and included 46,914 SNP markers that are used in the current national multi-step genomic evaluations.

Statistical models. Single-step GBLUP and pedigree-based BLUP (PBLUP) were used to estimate breeding values. All models used the same pedigree and UPG, but the phenotypic and genomic data differed. Genomic prediction was performed using a single-step GTBLUP (Mäntysaari *et al.*, 2017; Koivula *et al.*, 2021) model with two different datasets: $ssGTBLUP_{FIC+RDC}$ used FIC and RDC data; $ssGTBLUP_{FIC}$ used FIC data only. Corresponding PBLUP predictions used either FIC and RDC data – $PBLUP_{FIC+RDC}$ or FIC data only – $PBLUP_{FIC}$. In the single-step models, UPG were substituted by the same number of metafounders (MF, Legarra *et al.*, 2015). The relationships between MF in the pedigree were modeled using a gamma matrix (Γ) of size 137 as presented in Kudinov *et al.* (2021b). The inverses of pedigree relationship matrix (A) and submatrix for genotyped animals (A_{22}) were built with MF inbreeding coefficients. Genomic relationship matrix (G) was built assuming a base population allele frequency = 0.5 and residual polygenic effect = 30%. Model runs were done with MiX99 software (Strandén and Lidauer, 1999).

The prediction power of the genetic evaluation models was examined using forward prediction validation. In the reduced data, records from 2017-2020 were omitted. The validation used 73 genotyped cows that had no record in the reduced milk and protein data but had records in the full data. Validation reliability of predictions was assessed by two methods: (1) production prediction validation (PPV) – regression of yield deviations (YD) of cows computed with full genomic and phenotypic data on (genomic) estimated breeding values ([G]EBV) computed using truncated data ($[G]EBV_{red}$; $YD - \mu_{[G]EBVred} = b_0 + b_1 ([G]EBV_{red} - \mu_{[G]EBVred})$, where $\mu_{[G]EBVred}$ is mean of $[G]EBV_{red}$). Similar to the Interbull GEBV validation test (Mäntysaari *et al.*, 2010), the YD of cow *j* was weighted by w_j computed as $w_j = ERC_j / (ERC_j + \lambda)$, where $\lambda = (1-h^2)/h^2$;h^2 is heritability of the trait and ERC_j is the effective record contribution (Přibyl *et al.*, 2013), the coefficient of determination (R^2) from the regression equation was adjusted by mean of the weights (R_w^2); (2) Linear-regression validation (LRV, Legarra and Reverter, 2018) of [G]EBV computed using the full data on $[G]EBV_{red}$ using formulae $[G]EBV - \mu_{[G]EBVred} = b_0 + b_1([G]EBV_{red} - \mu_{[G]EBVred})$.

Results & discussion

Table 1 presents validation results for the first lactation milk yield in 73 FIC cows obtained using the PPV and LRV approaches. In the PPV approach, the highest validation reliability (R_w^2) was obtained for the $ssGTBLUP_{FIC+RDC}$ model (0.52), as expected. Elimination of the RDC data from the single-step model ($ssGTBLUP_{FIC}$) led to a 0.12 decrease in R_w^2 (0.40). In both data scenarios, the single-step models showed higher validation reliabilities than their corresponding animal models. The increase was 0.12 and 0.07 for RDC+FIC and FIC data, respectively. The regression coefficient (b_1), indicating the bias, was over 1.0 for the $ssGTBLUP_{FIC+RDC}$ model due to underprediction of GEBV differences by the reduced data. In the LRV approach, the b_1 coefficient behaved like in PPV and the value was over one (1.05) in the $ssGTBLUP_{FIC+RDC}$ model. The highest correlation between the full and the reduced data predictions was observed in $ssGTBLUP_{FIC+RDC}$ suggesting a higher predictive ability of the model compared to the other methods. For protein (Table 2), in general, the R_w^2 values were higher than in milk. The b_1 coefficients were over one in PPV for both genomic models (1.07 and 1.03). In the LRV approach, b_1 was close to one in all models and the highest coefficient of correlations (*cor*) were in models with genomic data (0.83 and 0.78). The *cor* values for milk and protein were alike.

The addition of RDC phenotypic data to FIC evaluations improved the prediction reliability of the FIC traditional model. Further augmentation of the model by FIC and RDC genotypes enhanced prediction even further.

Table 1. Results from validation test in 73 FIC cows for 305d first lactation milk.

	PPV[2]			LRV[3]		
Model	b_0 **(kg)**	b_1	R_w^2	b_0 **(kg)**	b_1	*corr*
ssGTBLUP$_{FIC+RDC}$	-60	1.07	0.52	1.2	1.05	0.84
ssGTBLUP$_{FIC}$	-34	1.01	0.40	-21	0.99	0.78
PBLUP$_{FIC+RDC}$	-52	0.99	0.40	2.2	0.99	0.73
PBLUP$_{FIC}$	-20	0.92	0.33	-13	0.96	0.70

[1] b_0 = the bias; b_1 = the regression coefficient; R_w^2 = the coefficient of determination from PPV adjusted by the ERC; *corr* = coefficient of correlation.
[2] Prediction of full data YD by reduced data [G]EBV using linear model $YD - \mu_{[G]EBVred} = b_0 + b_1([G]EBVred - \mu_{[G]EBVred})$.
[3] Prediction of full data [G]EBV by reduced data [G]EBV using linear model $[G]EBV - \mu_{[G]EBVred} = b_0 + b_1([G]EBVred - \mu_{[G]EBVred})$.

Table 2. Results from validation test in 73 FIC cows for 305d first lactation protein.[1]

	PPV[2]			LRV[3]		
Model	b_0 **(kg)**	b_1	R_w^2	b_0 **(kg)**	b_1	*corr*
ssGTBLUP$_{FIC+RDC}$	-1.91	1.07	0.61	-0.03	0.99	0.83
ssGTBLUP$_{FIC}$	-0.11	1.03	0.49	0.14	0.98	0.78
PBLUP$_{FIC+RDC}$	-2.09	1.00	0.44	-0.23	0.98	0.72
PBLUP$_{FIC}$	0.19	0.98	0.42	0.32	0.97	0.71

[1] b_0 = the bias; b_1 = the regression coefficient; R_w^2 = the coefficient of determination from PPV adjusted by the ERC; *corr* = coefficient of correlation.
[2] Prediction of full data YD by reduced data [G]EBV using linear model $YD - \mu_{[G]EBVred} = b_0 + b_1([G]EBVred - \mu_{[G]EBVred})$.
[3] Prediction of full data [G]EBV by reduced data [G]EBV using linear model $[G]EBV - \mu_{[G]EBVred} = b_0 + b_1([G]EBVred - \mu_{[G]EBVred})$.

Both PPV and LRV approaches gave similar results for the analyzed data. The unfit of the PPV for complex data sets (low heritability, indirect genetic values) discussed in Legarra and Reverter (2018) was not applicable for our research. However, the performance of the LRV approach is computationally less challenging compared to PPV and could be prioritized. Implementation of the ssGTBLUP model for FIC and RDC Test-Day data is the next development step. In the analysis of the test day data, the LRV approach allows a straightforward way to perform prediction validation.

Conclusions

A multi-breed single-step model for FIC and RDC was successfully used in genomic prediction using 305-day first lactation milk and protein data. The genomic model with FIC and RDC data showed higher prediction reliability than using FIC data only.

Acknowledgements

We acknowledge Project 'Genomic evaluations for Western Finncattle' financed by Luke (Finland) and Finncattle Foundation (Finland). Data providers: Finnish Breeder Association (FABA, Finland), Nordic Cattle Genetic Evaluation (NAV, Denmark), and VikingGenetics (Denmark).

References

Christensen O.F., and Lund M. S. (2010) Genetics Selection Evolution, 42, 2. https://doi.org/10.1186/1297-9686-42-2

Koivula M., Strandén I., Aamand G.P., and Mäntysaari E.A. (2021) J. Dairy Sci. 104(9):10049-10058. https://doi.org/10.3168/jds.2020-1982

Kudinov A.A., Koivula M., Strandén I., Aamand G.P., and Mäntysaari E.A. (2021) Interbull Bulletin, 56, 174-179.

Kudinov A.A., Mäntysaari E.A., Pitkänen T.J., Saksa E.I., Aamand G.P. *et al.* (2021) J. Animal Breed. Genet. 00, 1– 12. https://doi.org/10.1111/jbg.12660

Legarra A., Christensen O.F., Vitezica Z.G., Aguilar I., and Misztal I. (2015). Genetics, 200(2), 455–468. https://doi.org/10.1534/genetics.115.177014

Legarra A., and Reverter A. (2018). Genetics Selection Evolution 50:53. https://doi.org/10.1186/s12711-018-0426-6

Mäntysaari E.A., Liu Z., and VanRaden P.M. (2010) Interbull Bulletin, 41,17–22.

Mäntysaari, E.A., Evans R.D., and Strandén, I. (2017) J. Animal Sci. 95:4728–4737, https://doi.org/10.2527/jas2017.1912

Přibyl J., Madsen P., Bauer J., Přibylová J., Šimečková M. *et al.* (2013) J. Dairy Sci. 96:1865-1873. https://doi.org/10.3168/jds.2012-6157.

Song H., Zhang J., Zhang Q., and Ding X. (2019) Frontiers in Genetics, 9, 730. https://doi.org/10.3389/fgene.2018.00730

Strandén I., and Lidauer M. (1999) J. Dairy Sci. 82:2779–2787. https://doi.org/10.3168/jds.S0022-0302(99)75535-9

VanRaden P.M. (2020) J. Dairy Sci. 103:5291–5301. https://doi.org/10.3168/jds.2019-17684

314. Challenges to implement an interim genomic prediction for young animals in Wallonia, Belgium

K. Wijnrocx[1*], R.R. Mota[1,2], M. Dufrasne[3] and N. Gengler[1]

[1]University of Liège – Gembloux Agro-Bio Tech, Passage des Déportés, 2, 5030 Gembloux, Belgium; [2]Council on Dairy Cattle Breeding, Northview Dr 4201, Bowie, MD 20716, USA; [3]Walloon Breeders Association, rue des Champs Elysées 4, 5590 Ciney, Belgium; katrien.wijnrocx@uliege.be

Abstract

Official Walloon genomic evaluation for the Holsteins is performed three times a year for several traits, by using a single-step GBLUP model, with a 'Bayesian' integration comprising both local and foreign breeding values. The need of a system to provide in between evaluations for young animals was stressed out by breeders. This study gives an overview of the method developed, it describes the challenges of its practical implementation, and finally it proposes some improvements. We identified that a crucial step for the correct interim values estimation is the polygenic term, which requires a deep pedigree knowledge.

Introduction

Official genomic evaluations for the Holstein population in the Southern part of Belgium (Walloon region) are performed for production, conformation, udder health and functional traits since 2015. Nearly all traits are submitted three times a year to the Interbull Centre (https://interbull.org/index, Uppsala, Sweden), which generates a Genomic Multiple Across Country Evaluation (GMACE) with local genomically enhanced breeding values (GEBVs). As outlined by Vandenplas *et al.* (2014) and Colinet *et al.* (2018), the Walloon local genomic evaluations are based on pseudo single-step GBLUP model (ssGBLUP) that used a 'Bayesian' integration of local Walloon breeding values. This method adds MACE values and are able to avoid double-counting by the subtraction of the Walloon contribution from MACE. By having evaluations only three times a year this generates a delay up to 4 months between the genotyping of an animal and its official GEBV evaluations. For example, an animal genotyped after the Interbull deadline has to wait for the next official evaluation to get its own GEBV. Thus, an interim system to perform genomic evaluations for young animals is of interest for breeders that will be able to make earlier selection decisions. Several authors have proposed different solutions this issue. A study of Lourenco *et al.* (2015) described an indirect method for the prediction of the GEBV of young animals by using ssGBLUP analyses. This method calculates the GEBV through the weighted combination of direct genomic value (DGV) and parent average (PA), which they adapted by including an algorithm and SNP effects blending and tuning (Lourenco *et al.*, 2018a). This approach however still generated some issues such as animals with records, and the weighing factor calculation.

An alternative approach was proposed by Pimentel *et al.* (2019), in which the GEBVs are directly split into a SNP based part and a polygenic part. Both parts were estimated separately, and then combined afterwards. The SNP part was estimated by using a 'SNP model' on the residual GEBVs from reference animals from ssGBLUP. These SNP effects were used subsequently to compute DGV, and the 'residual' term was the polygenic part. This approach however had some problems to transmit the polygenic term from reference animals to young animals especially when the distance between genotyped animals and selection candidates is not very close, i.e. when direct ancestors are not genotyped. Therefore, Mota *et al.* (2019) proposed an alternative where they combined the best of both methods to obtain an interim prediction for young Holstein animals in Wallonia. The new method was validated, and a test implementation was done by the Walloon Breeders Association (AWE). Some inconsistencies have arisen and resulted in low correlations between the predicted and calculated interim values. This manuscript will give a short outline

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_314, © K. Wijnrocx *et al.* 2022

of the method of Mota *et al.* (2019), its implementation by AWE, and challenges and necessary adjustments needed for practical use.

Materials & methods

Partitioning of the GEBVs and generation of polygenic terms. The modified approach that was proposed by Mota *et al.* (2019) followed the derivation of ssGBLUP by Christensen and Lund (2010) to split GEBV into genomic (DGV) and polygenic (PG) parts as described previously by Pimentel *et al.* (2019). In contrast to Pimentel *et al.* (2019), PG were extracted at first, and only then DGV were calculated. This step can be derived as follows: let's call û the vector of available genetic values (EBV respectively GEBV). Each of these values will represent the full genetic variance, and even for the non-genotyped animals, their PG can be approximated through $\hat{\mathbf{a}} = \mathbf{Cov(a,u)}\left(\mathbf{Var(u)}\right)^{-1}\left(\hat{\mathbf{u}} - \mathbf{1}\hat{\mathrm{m}}\right)$ where û is as aforementioned, and m̂ is a constant. The matrix $\left(\mathbf{Var(u)}\right)^{-1}$ should be equal to $\mathbf{H}^{-1}$ as defined by Legarra *et al.* (2009) and Christensen and Lund (2010). The covariance matrix Cov(a,u) needs to express the variance ratio between the variance explained by PG, hereafter called σ^2_a and the total genetic variance σ^2_u.

An alternative approach would be to write this hypothetical mixed model as $\hat{\mathbf{u}} = \mathbf{1}\hat{\mathrm{m}} + \mathbf{Z}\hat{\mathbf{a}} + \mathbf{e}$. The associated theoretical correct decomposition of the total (co)variances would be $\mathbf{Var(u)} = \mathbf{ZVar(a)Z'} + \mathbf{Var(e)} = \mathbf{ZAZ'}\sigma^2_a + \mathbf{G}\sigma^2_e$. Therefore, the model would be:

$$\begin{bmatrix} \mathbf{1'G^{-1}1} & \mathbf{1'G^{-1}Z} \\ \mathbf{Z'G^{-1}1} & \mathbf{Z'G^{-1}Z} + \mathbf{A^{-1}}k \end{bmatrix} \begin{bmatrix} \hat{\mathrm{m}} \\ \hat{\mathbf{a}} \end{bmatrix} = \begin{bmatrix} \mathbf{1'G^{-1}\hat{u}} \\ \mathbf{Z'G^{-1}\hat{u}} \end{bmatrix} \text{with } k = \sigma^2_e / \sigma^2_a \quad (1)$$

and perfectly equivalent to:

$$\hat{\mathbf{a}} = \sigma^2_a \mathbf{AZ'}\left(\mathbf{ZAZ'}\sigma^2_a + \mathbf{G}\sigma^2_e\right)^{-1}\left(\hat{\mathbf{u}} - \mathbf{1}\hat{\mathrm{m}}\right) = \mathbf{AZ'}\left(\mathbf{ZAZ'} + \mathbf{G}k\right)^{-1}\left(\hat{\mathbf{u}} - \mathbf{1}\hat{\mathrm{m}}\right) \text{ with } k = \sigma^2_e / \sigma^2_a \quad (2)$$

where

$$\hat{\mathrm{m}} = \left(\mathbf{1'}\left(\mathbf{ZAZ'} + \mathbf{G}k\right)^{-1}\mathbf{1}\right)^{-1}\left(\mathbf{1'}\left(\mathbf{ZAZ'} + \mathbf{G}k\right)^{-1}\hat{\mathbf{u}}\right) \quad (3)$$

In these equations, 1 is a vector of ones, A is the pedigree-based relationship matrix among known animals, G represents the hypothetical genomic relationships between all animals; m̂ is BLUE estimate of the overall mean; â is the vector of the PG values for all animals, which are to be estimated; $\boldsymbol{\sigma^2_e}$ is the residual variance (which is the genetic variance associated with the SNP effects (DGV)); σ^2_a is the genetic variance (explained by pedigree and therefore due to PG); Z is the incidence matrix linking â to û. The overall GEBV mean was added to the model to avoid a mean shift in the indirect predictions (Lourenco *et al.*, 2018b). It can be shown that these mixed model equations are a close approximation of the theoretically correct equations deriving polygenic contributions to GEBV. The proportion of additive genetic variance explained by SNPs was defined by our routine genomic evaluations, i.e. currently 60% for production and 65% for other traits. Therefore, $k = \sigma^2_e / \sigma^2_a = 0.60\sigma^2_u / 0.40\sigma^2_u$ or $0.65\sigma^2_u / 0.35\sigma^2_u$.

Solving (1) is not trivial therefore we solve these alternative equations:

$$\begin{bmatrix} \mathbf{1'1} & \mathbf{1'Z} \\ \mathbf{Z'1} & \mathbf{Z'Z} + \mathbf{A^{-1}}k \end{bmatrix} \begin{bmatrix} \hat{\mathrm{m}} \\ \hat{\mathbf{a}} \end{bmatrix} = \begin{bmatrix} \mathbf{1'\hat{u}} \\ \mathbf{Z'\hat{u}} \end{bmatrix} \text{with } k = \sigma^2_e / \sigma^2_a \quad (4)$$

where G was considered equal to an identity matrix and therefore all links between û considered pedigree based and â estimated in an equivalent approach as if this penalized regression $\hat{\mathbf{a}} = \mathbf{AZ'}\left(\mathbf{ZAZ'} + \mathbf{I}k\right)^{-1}\left(\hat{\mathbf{u}} - \mathbf{1}\hat{\mathrm{m}}\right)$ was used. The use of this approach that allowed the combination of both EBV and GEBV in û. After solving this model using all available solutions (i.e. for genotyped and non-genotyped animals) from the ssGBLUP

system, the vector $\hat{\mathbf{e}}$ for genotyped animals (i.e. reference animals) is considered equal to the vector of DGV ($\hat{\mathbf{d}}$) as follows:

$$\hat{\mathbf{d}} = \hat{\mathbf{e}} = \hat{\mathbf{u}} - (\mathbf{1}\hat{m} + \mathbf{Z}\hat{\mathbf{a}}) \quad (5)$$

In this adapted procedure, the generation of PG could be done for newly genotyped animals mixing their EBV with EBV/GEBV from relevant ancestors. As of such, GEBV a posteriori combining PG with SNP effects can be generated, which will potentially include non-reference animals own records.

Generation of SNP effects. The second step of the adapted method proposed by Mota *et al.* (2019) was the estimation of the SNP effects. This step applied the proposal of Lourenco *et al.* (2015), based on selection index theory, and reduced GEBV into $\hat{\mathbf{d}}$:

$$\hat{g} = DM'G^{-1}(\hat{d}) \quad (6)$$

where $\hat{g}$ is the vector of SNP effects, D is the diagonal matrix of standard variance weights for SNPs (an identity matrix here), M is a centred genotypes matrix (VanRaden, 2008), G^{-1} is the inverse genomic relationship matrix (MDM') among reference animals, and $\hat{\mathbf{d}}$ is a vector of reference population DGV values.

Implementation by the Walloon breeders association. Practical implementation of the interim genomic predictions for young animals was performed by AWE using a simplified, approximate procedure. This was done because new genotyped animals are constantly added to the population and a full single-step cannot yet be perform on a weekly or monthly basis. The 10k genotypes of young animals were imputed into common SNP list that combines 50k V1 and 50k V2 SNP chips. In the first step, the reduced polygenic term was obtained by using Equation 4. The interim GEBV for a young animal *i* can thus be calculated for each trait of interest as follows:

$$\text{Interim GEBV}_i = \text{mean} + \text{DGV}_i + \text{PG}_i = \hat{m} + \hat{a}_i + m_i\hat{g} \quad (7)$$

Note that $\hat{g}$ is calculated in the Walloon genomic evaluation centre and transferred to AWE, whereas $\hat{m}$ and $\hat{a}$ are calculated by AWE for the set of animals of interest. The vector of m_i representing the centred genotypes for animal *i* are computed by AWE using the allele frequencies used to generate $\hat{g}$.

Validation and mitigation. Pearson correlations were calculated for a set of 150 Holstein animals that had no own records in the previous official Walloon evaluation in order to validate the method. Correlations were calculated between the published GEBV from the current evaluation, and the calculated interim GEBV for these animals by the proposed method. The extraction of a three generations pedigree was found to cause biased results, and thus a full extraction was applied to overcome this issue and to be further used in the SNP effects estimation.

Results

Pearson correlations between the published and interim GEBV for a three generations and full pedigree for conformation traits (loin strength, udder balance, overall rump) and milk yield are shown in Table 1. Correlations for conformation traits were around 0.72 for the three generations pedigree depth, whereas values increased up to 0.87 by full extraction. Likewise, correlations for milk yield increased from 0.93 to 0.97.

Table 1. Pearson correlations between interim and published GEBVs for a subset of 150 young animals (AWE) for a three generation and full pedigree extraction.

Trait	3 generations pedigree	Full pedigree
Loin strength	0.72	0.87
Udder balance	0.72	0.84
Overall rump score	0.72	0.87
Milk yield	0.93	0.97

Discussion

This study demonstrated the challenge that arose when practically implementing interim genomic breeding values using a simplified system. For breeders, this system is of great interest as it allows earlier selection decisions. Our method, after adaptations, correctly implements genetic groups, a problem that was documented in literature (Lourenco *et al.*, 2015; Pimentel *et al.* 2019), since all animals in the pedigree are considered for the interim estimation. However, the implementation demonstrated suboptimal results when using few generations in the pedigree. The full pedigree extraction allows the estimation of less biased GEBVs. This also demonstrated the importance of the correct calculation of the PG term. Indeed, crucial for a correct implementation is a compatible $\hat{m}$ and $\hat{a}$, as calculated by the Walloon genomic evaluation centre (Equation 4) on the full set of animals, and $\hat{m}$ and $\hat{a}$ calculated and used by AWE (Equation 7). Correct calculation of m_i that represents the centred genotypes for animal i was ensured by providing the calculated allelic frequencies to AWE.

This research demonstrated that interim GEBVs can be applied successfully after few adjustments, to ensure the correct approximation of the PG part needed to calculate interim GEBVs. However, more research is necessary to keep improving the system.

References

Christensen O.F., and Lund M.S. (2010). GSE 42:1. https://doi.org/10.1186/1297-9686-42-2

Colinet F.G., Vandenplas J., Vanderick S., Hammami H., Mota R.R. *et al.* (2018). Animal 12(5):898–905. https://doi.org/10.1017/S1751731117002324

Legarra A., Aguilar I., and Misztal I. (2009). J. Dairy Sci. 92(9):4656-4663. https://doi.org/10.3168/jds.2009-2061.

Lourenco D.A.L., Legarra A., Tsuruta S., Moser D., Miller S., *et al.* (2018a). Interbull Bul. 53:48–53. https://journal.interbull.org/index.php/ib/article/view/1448

Lourenco D.A.L., Tsuruta S., Fragomeni B.O., Masuda Y., Aguilar I. *et al.* (2015). J. Anim. Sci. 93(6):2653–2662. https://doi.org/10.2527/jas.2014-8836

Lourenco D.A.L., Tsuruta S., Fragomeni B.O., Masuda Y., Aguilar I. *et al.* (2018b). Proc. of the 10th WCGALP, Vancouver, Canada.

Mota R.R., Nader S., Vanderick S., Colinet F.G., Gillon A. *et al.* (2019). Interbull Bul. 55:64–69. https://journal.interbull.org/index.php/ib/article/view/1472/1586.

Pimentel E.C.G., Edel C., Emmerling R. and Götz K.U. (2019). J. Dairy Sci. 102(4):3266-3273. https://doi.org/10.3168/jds.2018-15592

Vandenplas J., Colinet F.G., and Gengler N. (2014). GSE 46:1–15. https://doi.org/10.1186/s12711-014-0059-3

VanRaden P.M. (2008). J. Dairy Sci. 91(11):4414–4423. https://doi.org/10.3168/jds.2007-0980

315. New insights into the computation of the inverse of the numerator relationship matrix for genotyped animals

A. Maiwashe[1,2]

[1]ARC-Animal Production Institute, Private Bag X2, Irene, 0062, South Africa; [2]Department of Animal, Wildlife and Grassland Sciences, UFS, P.O. Box 339, Bloemfontein, 9300, South Africa; norman@arc.agric.za

Abstract

The revolution in genotyping, sequencing and statistical computing allowed use of big data in animal breeding programmes. Genomic selection (GS) through Single-Step GBLUP (sSGBLUP) is now the gold standard for national genetic evaluation although computational challenges remain. The sSGBLUP requires the inverse of genomic (G^{-1}) and pedigree matrices (A^{-1}_{22}). These matrices are structurally similar and computationally intensive. The objective of this study was to investigate the properties of the lower triangular matrix, T^{-1}, to identify opportunities for enhancing computational efficiency. We observed that the heuristic algorithm of Faux and Gengler to identify sparsity pattern of T^{-1} is consistent with the *d-separation* graph criterion used in causal inference. We further observed that some of non-zero elements of T^{-1} are correlated due to the violation of the causal Markov faithfulness assumption. The linear dependency provides an opportunity for efficient computation.

Introduction

Genomic selection is now commonplace in most national genetic evaluation systems in developed countries with some developing countries following suite (Burrow *et al.*, 2017). This technology has gradually attained some level of maturity but more work is still required on the statistical computing front (Misztal *et al.*, 2014).

Research to ensure that genomic selection is placed on solid theoretical foundation could be categorised into two broad areas: Firstly, the need for reconciliation of the existing quantitative genetic theory and the new theoretical developments especially in relation to the key theoretical assumptions e.g. base population of the genotyped and ungenotyped animals. Secondly, there is a need for development of the computational statistical machinery for efficient and cost-effective prediction of genomic breeding values. This paper focuses on the second line of research i.e. computational efficient statistical methods.

The computation of the inverse of the genomic and pedigree-based matrices has received considerable attention recently due to their integral role in sSGBLUP (e.g. VanRaden, 2008; Misztal *et al.*, 2009; Faux and Gengler, 2013). Other approaches for inclusion of genomic data which do not require the inverse of the genomic matrix have been developed. Faux and Gengler (2013) presented a practical method to compute A^{-1}_{22} which builds on the theory developed by Henderson (1976). This approach is appealing as it is aligned with the existing classical theory of Wright (1921) on causation and correlation.

As a result of recent theoretical development, we now have a good understanding of the properties of the non-zero elements of the inverse of the lower triangular matrix T^{-1} derived from full knowledge of the pedigree. However, our understanding of the properties of T^{-1} is limited when some animals are ungenotyped. The theoretical basis behind the high density of the T^{-1} for the genotyped animals is also less understood. The objective of this study was to investigate the properties of the A^{-1}_{22} and specifically the sparsity pattern of the lower triangular matrix (T^{-1}) in order to identify opportunities for enhancing computational efficiency.

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_315, © A. Maiwashe *et al.* 2022

Materials & methods

Current theory. The classical paper of Wright (1921) on correlation and causation has laid a solid foundation for the quantitative genetic theory. This paper presented a key concept of *path coefficient*. Without loss of generality and following the notation of Li (1956), the fundamental equation for computing correlations / covariance from *path coefficients* is $r_{XY} = \sum_i p_{Xi}p_{Yi} + \sum_{i \neq j} p_{Xi}r_{ij}p_{Yi}$. This equation gave rise to Wright's numerator relationship matrix (A) which can be presented as follows: A = TDT′ (Henderson 1976). For a scenario where animals have different sources of information e.g. pedigree data (denoted by 1) and genomic data (denoted by 2), Faux and Gengler (2013) partitioned the numerator relationship matrix as follows:

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}, \text{ where } \quad A_{22} = \begin{bmatrix} Z & y \\ y' & m \end{bmatrix},$$

Following partitioned matrix theory, the inverse of A_{22} is given by:

$$A_{22}^{-1} = \begin{bmatrix} Z^{-1} & 0 \\ 0' & 0 \end{bmatrix} + \frac{1}{s}\begin{bmatrix} -Z^{-1}y \\ 1 \end{bmatrix}[-y'Z^{-1} \quad 1],$$

where $b = z^{-1}y$ and -b is the last row of T^{-1}. The non-trivial activity in the computation of $\mathbf{A}_{22}^{-1}$ pertains to solving the linear system $b = z^{-1}y$. Faux and Gengler (2013) implemented efficient algorithms that avoids explicit computations involving zero entries through identification of the sparsity pattern of T^{-1}.

Identification of the sparsity pattern of T^{-1}. The heuristic algorithm of Faux and Gengler (2013) to identify the sparsity pattern of the T^{-1} uses pedigree data. The intuition behind this algorithm was briefly described in their paper but the underlying theory was not adequately addressed. In this paper, we appeal to graph theory and the structural causal model framework to provide an alternative basis for identifying sparsity pattern.

We use the toy pedigree example provided in Faux and Gengler (2013) for continuity (Figure 1). In graph theory, the pedigree is an example of a Directed Acyclic Graph (DAG). A DAG represents the underlying causal structure of a closed system. In a DAG, animals are represented by nodes and the causal relationship between them represented by arrows / vertices whose metrical value is given by a path coefficient.

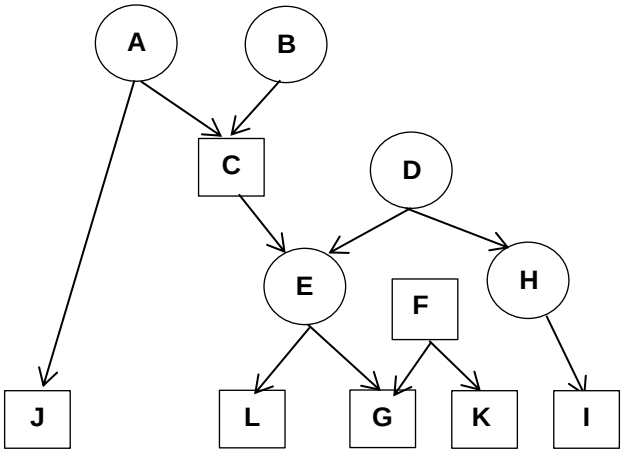


Figure 1. A toy example pedigree (adapted from Faux and Gengler 2013). Circles and squares represent ungenotyped and genotyped animals, respectively.

In a trivial case where all the nodes in a causal network (e.g. pedigree) are known, the covariance between nodes can be computed from the path coefficients and observed covariance. For example, an animal is independent of all other animals in the pedigree conditional on the knowledge of its parents. In a non-trivial case where some of the parents are unknown, as is the case in A_{22}, the notion of conditional independence is generalised through a *Markov blanket* which is a set of all immediate known ancestors of the animal, their offspring and mates. The Markov blanket is identified through a *d-separation* graphical criterion which is a process of traversing the graph to identify a set of animals that blocks all the connecting paths between animals (see Pearl 2010). If animal *X* is *d-separated* from animal *Y* then animals *X* and *Y* are conditionally independent. The set *S* separating *X* from *Y* is a *Markov blanket*. For the example pedigree in Figure 1, a set S = {*C*, *F*, *G*} is a Markov blanket of node *I*. In other words, when animals *D*, *E* and *H* are ungenotyped, animal *I* is conditionally correlated with animals *C*, *F* and *G*. On the other hand, when *H* is genotyped, animal *I* is conditionally uncorrelated with all animals in the pedigree. This example emphasises the fact that the *flow of information* in the pedigree depends on our knowledge of the information on each node.

Computation of path coefficients. The identification of sparsity pattern using *d-separation* indicates that the lack of complete knowledge about the pedigree, as is the case when some animals are ungenotyped, induces statistical covariance among animals. In the example above, animals *C* and *F* are conditionally correlated with animal *I* when its direct ancestors are ungenotyped even though these animals are conditionally uncorrelated based on the pedigree. The induced correlation results in linear dependency among path coefficients. We considered the numerator relationship of the example pedigree in Figure 1 to demonstrate linear dependence between path coefficients.

The linear system to compute the elements of T^{-1} of animal L in Figure 1 is as follows.

$$\begin{bmatrix} a_{cc} & a_{cf} & a_{cg} & a_{ci} \\ a_{fc} & a_{ff} & a_{fg} & a_{fi} \\ a_{gc} & a_{gf} & a_{gg} & a_{gi} \\ a_{ic} & a_{if} & a_{ig} & a_{ii} \end{bmatrix} \begin{bmatrix} b_{cl} \\ b_{fl} \\ b_{gl} \\ b_{il} \end{bmatrix} = \begin{bmatrix} a_{cl} \\ a_{fl} \\ a_{gl} \\ a_{il} \end{bmatrix}$$

Using the observed covariance, the above system is given as follows,

$$\begin{bmatrix} 1 & 0 & 0.25 & 0 \\ 0 & 1 & 0.50 & 0 \\ 0.25 & 0.50 & 1 & 0.06 \\ 0 & 0 & 0.06 & 1 \end{bmatrix} \begin{bmatrix} b_{cl} \\ b_{fl} \\ b_{gl} \\ b_{il} \end{bmatrix} = \begin{bmatrix} 0.25 \\ 0 \\ 0.25 \\ 0.06 \end{bmatrix}$$

Based on the above, $b_{fl} = -\frac{1}{2} b_{gl}$ and the following reduced linear system can be solved to obtain the path coefficients:

Table 1. A_{22} for animals in Figure 1.

	C	F	G	I	J	K	L
C	1	0	0.25	0	0.25	0	0.25
F	0	1	0.5	0	0	0.5	0
G	0.25	0.5	1	0.06	0.06	0.25	0.25
I	0	0	0.06	1	0	0	0.06
J	0.25	0	0.06	0	1	0	0.06
K	0	0.5	0.25	0	0	1	0
L	0.25	0	0.25	0.06	0.06	0	1

$$\begin{bmatrix} 1 & 0.25 & 0 \\ 0.25 & 1 & 0.06 \\ 0 & 0.06 & 1 \end{bmatrix} \begin{bmatrix} b_{cl} \\ b_{gl} \\ b_{il} \end{bmatrix} = \begin{bmatrix} 0.25 \\ 0.25 \\ 0.06 \end{bmatrix}; \text{ with } \begin{bmatrix} b_{cl} \\ b_{gl} \\ b_{il} \end{bmatrix} = \begin{bmatrix} 0.183 \\ 0.269 \\ 0.044 \end{bmatrix} \text{ and, therefore}$$

$$b_{fl} = -\frac{1}{2} b_{gl} = -0.134$$

Results

The results from this study shows that the linear system to obtain the path coefficients required for A^{-1}_{22} exhibit linear dependence as a result of ungenotyped ancestors that explains the observed relationships in the pedigree, A_{22}. For the pedigree considered here, the number of equations in the linear system was reduced from 4 to 3 which seem trivial. The linear dependence was induced by genotyped offspring of the hidden or ungenotyped ancestors of the animal in question. Given that these ungenotyped ancestors may have hundreds or thousands of genotyped offspring, in view of the typical breeding structure in livestock, it could be speculated that the reduction in the linear system could be significant and result in tangible computational efficiency of T^{-1} efficiency.

Discussion

Consideration of the linear system used in the computation of T^{-1} revealed important properties of the path coefficients. The linear dependence among path coefficients induced by functional relations among animals indicate that these are induced statistical correlations to ensure integrity of the linear system. Accordingly, the linear system ensures that the key underlying assumptions, such as closed nature of the causal system, is guaranteed. The results from this study provide important insights into the properties of the linear system which could be useful for computational efficiency in deriving A^{-1}_{22}. The attempt to introduce causal graphical models may stimulate adoption of causal analysis for inference in animal breeding.

References

Burrow H.M., Wolcott M.L., Maiwashe A., Makgahlela M.L., Hayes B.J., Rees J.G., and Bradfield M.J. (2017). Proc. Assoc. Advmt. Anim. Breed. Genet. 22:353-360.

Faux P., and Gengler N. (2013). Genet. Sel. Evol. 45:45. http://wwwgsejournal.org/content /45/1/45

Henderson, C.R. (1976). Biometrics 32: 69-83. http://www.jstor.org/stable/2529339

Legarra A., Aguilar I., and Misztal I. (2009). J. Dairy Sci. 92:4656-4663. https://doi.org/10.3168/jds.2009-2061.

Li C.C. (1956). Biometrics, 12(2): 190-210. http://www.jstor.org/stable/3001760

Misztal I., Legarra A., and Aguilar I. (2009). J. Dairy Sci. 92:4648-4655. https://doi.org/10.3168/jds.2009-2061.

Misztal I., Legarra A., and Aguilar I. (2014). J. Dairy Sci. 97:1-10. http://dx.doi.org/10.3168/jds.2013-7752

Pearl J. (2010). Int J Biostat. 6(2): 1-59. https://doi.org/10.2202/1557-4679.1203

VanRaden P.M. (2008). J. Dairy Sci. 91(11):4414-4423. https://doi.org/10.3168/jds.2007-0980

Wright S. (1921). J. Agric Res 20:557-585.

316. GAPLS – a novel method to genomically rank bulls for daughter reproductive performance

Y. Li[1*], S. Lehnert[1], L.R. Porto-Neto[1], R. McCulloch[1], S. McWilliam[1], P. Alexandre[1], J. McDonald[2], C. Smith[2] and A. Reverter[1]

[1]CSIRO Agriculture & Food, 306 Carmody Rd., St. Lucia, Brisbane, QLD 4067, Australia;
[2]MDH Pty Ltd, Cloncurry, QLD 4824, Australia; yutao.li@csiro.au

Abstract

Female reproductive traits directly impact the profitability of commercial beef herds. The ability to select herd bulls, based on the predicted reproductive performance of their female progeny, has the potential to significantly improve herd productivity. Genomic selection in beef cattle run in the extensive production systems of tropical and subtropical northern Australia is hindered by technical and economic barriers, including the difficulty of obtaining accurate records on reproductive performance of individual females. In this study we present a novel method, Genomic Attribution to Pregnancy and Lactation Status (GAPLS), to predict an individual bull's contribution to six categories of their progeny's 2nd joining pregnancy and lactation status. Using a simulated example and four commercial Brahman datasets, we demonstrate the merit of the method.

Introduction

In extensive production systems of tropical and subtropical northern Australia, uptake of genomic selection has been limited due to practical, financial and management complications associated with collecting individual performance records. In northern commercial cattle herds, following natural syndicate joining, heifers are mustered and grouped based on the result of their 2nd joining pregnancy and lactation status (PLS). Heifers are commonly grouped into six categories in these situations: 1. DNP = Dry and Not Pregnant; 2.WNP = Wet and Not Pregnant; 3. DEP = Dry and Early Pregnant; 4. DMP = Dry and Mid Pregnant; 5. DLP = Dry and Late Pregnant; 6. WEP = Wet and Early Pregnant. When heifers are genotyped, either individually or pooled as a group based on these categories, their DNA profiles become a valuable resource to explore genomic selection strategies for the improvement of fertility traits in beef cattle (Reverter *et al.*, 2016). However, the categorical nature of PLS coupled with departures from normal symmetry (e.g. PLS 3 and 4 not necessarily more abundant than 1 or 6) and the uncertainty about importance ranking of categories (e.g. PLS 2 not necessarily much worse than PLS 5), makes the analytical methodology to explore PLS rather cumbersome.

Here, we present Genomic Attribution to PLS (GAPLS), a novel and intuitive non-parametric proportion-based metric of genomic relatedness of a given individual animal (presumably a sire candidate for selection) across the six PLS categories.

Materials & methods

The basic concept of genomic attribution to PLS (GAPLS). Using the genotypes from individual bulls and a reference population of cows with genotypes and phenotypes for PLS, a genomic relationship matrix (GRM) can be computed as described by VanRaden *et al.* (2008). The GAPLS of a tested bull is defined for each PLS category as its average genomic relationship with the cows having that PLS category divided by its average genomic relationship across all cows. Numerically, for the i^{th} bull, its GAPLS to the j^{th} PLS category (j=1, 2, …6, corresponding to DNP, WNP, … WE, respectively) is computed as follows:

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_316, © Y. Li *et al.* 2022

$$\text{GAPLS}_{i,j} = \frac{\frac{1}{N_j}\sum_{k=1}^{k=N_C}[g_{i,k} \times \text{I}(k=j)]}{\sum_{j=1}^{j=6}\left(\frac{1}{N_j}\sum_{k=1}^{k=N_C}[g_{i,k} \times \text{I}(k=j)]\right)},$$

where: N_j = number of cows in the j^{th} PLS category, N_C = total number of cows in the reference population, $g_{i,k}$ = genomic relationship between the i^{th} bull and the k^{th} cow; I(k=j) = an indicator that takes a value of 1 if the k^{th} cow belongs to the jth PLS category, and 0 otherwise. By construction, the GAPLS of the i^{th} bull summed across all 6 j^{th} PLS categories is one.

Simulated example. For simple illustration purposes, consider a situation with 15 cows in the reference population and 2 candidate sires in the testing population (Table 1). The number of cows belonging to the PLS categories 1 to 6 is 2, 1, 3, 4, 1, and 4, respectively (second column of Table 1). Similarly, the genomic relationships between each cow and the hypothetical sires (Sire 1 and Sire 2) are given in the last two columns of Table 1.

In this scenario, the GAPLS of Sire 1 to PLS category 1 is:

$$\text{GAPLS}_{1,1} = \frac{\frac{1}{2}(0.1+0.0)}{\frac{1}{2}(0.1+0.0)+\frac{1}{1}(0.5)+\frac{1}{3}(0.2+0.1+0.1)+\frac{1}{4}(0.2+0.0+0.2+0.1)+\frac{1}{1}(0.0)+\frac{1}{4}(0.5+0.1+0.0+0.2)} = 0.0496$$

Similarly, the GAPLS of Sire 2 to PLS category 6 is:

$$\text{GAPLS}_{2,6} = \frac{\frac{1}{4}(0.3+0.4+0.3+0.4)}{\frac{1}{2}(0.1+0.0)+\frac{1}{1}(0.2)+\frac{1}{3}(0.0+0.1+0.0)+\frac{1}{4}(0.0+0.1+0.2+0.2)+\frac{1}{1}(0.2)+\frac{1}{4}(0.3+0.4+0.3+0.4)} = 0.375.$$

Table 2 presents the average GAPLS values for both sires across the 6 PLS categories based on the original data and after 1000 permutation tests of PLS. Since the first (1= DNP) and last two PLS categories (5=DLP and 6=WEP) indicate poor and good fertility, respectively, therefore Sire 2 would be preferred over Sire 1

Table 1. Hypothetical scenario with 15 cows with pregnancy and lactation status (PLS) and their genomic relationships ($g_{i,k}$) with two testing sires.

Cows		Sire's $g_{i,k}$	
K	PLS	Sire 1 (i = 1)	Sire 2 (i = 2)
1	6	0.5	0.3
2	3	0.2	0.0
3	4	0.2	0.0
4	1	0.1	0.1
5	6	0.1	0.4
6	3	0.1	0.1
7	6	0.0	0.3
8	1	0.0	0.0
9	4	0.0	0.1
10	5	0.0	0.2
11	2	0.5	0.2
12	4	0.2	0.1
13	6	0.2	0.4
14	3	0.1	0.0
15	4	0.1	0.2

because of its highest attribution value for category 6 (0.375). If using completely random data, i.e. a sire has no close relationship with any group of cows of six categories of PLS, the GAPLS of the sire to each PLS category is expected to equal 1/6 = 0.1667. Large deviations of GAPLS will be desirable when selecting sires because the PLS categories are unlikely to be equally represented in real scenarios. As shown in Table 2 (standard deviations in brackets), a robust measure of sampling variation can be quickly computed by permuting the PLS labels while maintaining the genomic relationships fixed.

Real data application. Datasets from four tropical Brahman cattle populations in north Queensland were used for the study. The first three populations contained 829 bulls, of which 114 were from the 2020 season (Bulls_114), 229 from 2021 (Bulls_229), and 486 are to be used in the 2022 mating season (Bulls_486). The fourth population consisted of 795 cows with PLS records. All samples were individually genotyped with 54,791 SNPs (Neogen Australasia GGP TropBeef 50K chip) and later imputed to a higher density of 529,260 autosomal SNPs. A GRM across all animals was constructed, and the GAPLS values of individual bulls were then estimated based on their genomic relationships with the 795 cows.

Results

Genomic relationships between bull and cow populations. The intensity difference in Figure 1A reveals that: (1) The animals in the Cows_795 set were more closely related among themselves than with the 829 bulls from the three bull populations; (2) The two bull populations (Bulls_114 and Bulls_229) had a closer relationship with each other than either had with the third bull population (Bulls_486); (3) The relationship between bulls and the cow population (Cows_795) was stronger for the Bulls_114 and Bulls_229 than for the Bulls_486.

GAPLS estimates for three bull populations. The closer genomic relationships between the first two bull populations (Bulls_114 or Bulls_229) and the 795 cows, allowed for large variations in average GAPLS values (Table 3, columns 2 and 3) in these two bull populations (ranging from 0.139 to 0.182 and 0.139 to 0.194, respectively). In comparison, the bull population Bulls_486 had a much narrower range (0.159-0.172). Accordingly, the standard deviations (STDs) of the GAPLS in the first two bull populations were also consistently larger than those in the third bull population. When clustering bulls based on their GAPLS values (Figure 1B), it is easy to identify the bulls with distinguished value differences (color gradient) across the six categories. For example, selecting the desirable bulls with the highest GAPLS values for 6_WEP (red color) and the lowest values for 1_DNP (green color).

Table 2. Average genomic attribution to pregnancy test outcome (GAPLS) values across the six PLS categories for the two sires after 1000 permutations of six PLS values.

PLS Category	Sire 1 (STD[1])	Sire 2 (STD)
1_DNP	0.0496 (0.115)	0.0536 (0.096)
2_WNP	0.4959 (0.137)	0.2143 (0.125)
3_DEP	0.1322 (0.100)	0.0357 (0.082)
4_DMP	0.1240 (0.098)	0.1071 (0.081)
5_DLP	0.0000 (0.144)	0.2143 (0.126)
6_WEP	0.1983 (0.091)	0.3750 (0.079)

[1] STD = standard deviation.

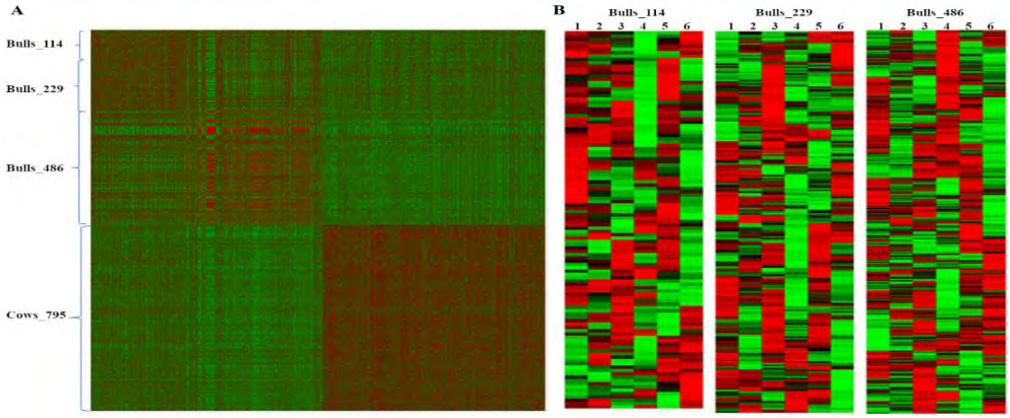


Figure 1. (A). Pairwise genomic relationships between animals of different populations. Each off-diagonal dot represents a genomic relationship matrix element between two animals. The red intensity represents closeness between animals. (B). Heatmap of individual bull's GAPLS values across six categories within each bull population.

Table 3. Average values of bull's genomic attribution to pregnancy and lactation status (GAPLS) across the six PLS categories for three bull populations.

Population	Bulls_114	Bulls_229	Bulls_486
PLS category	Mean (std)	Mean (std)	Mean (std)
1_DNP	0.182 (0.0572)	0.170 (0.0521)	0.172 (0.0463)
2_WNP	0.169 (0.0451)	0.161 (0.0436)	0.159 (0.0299)
3_DEP	0.176 (0.0567)	0.194 (0.0676)	0.178 (0.0499)
4_DMP	0.139 (0.0636)	0.139 (0.0608)	0.159 (0.0514)
5_DLP	0.178 (0.0594)	0.171 (0.0626)	0.170 (0.0461)
6_WEP	0.156 (0.0709)	0.165 (0.0675)	0.163 (0.0513)

Discussion

Genomic prediction for a phenotype like PLS, with six arbitrary categories, is challenging because traditional GBLUP models are inadequate. Firstly, it is hard to interpret the biological significance of a single GEBV that is obtained after treating the six PLS categories as a continuous trait. Secondly, how to define the scoring system for the six categories is open for discussion. For instance, if categories are coded from worst to best for fertility, WNP could be coded as 2 or as 5 and this has a significant impact on GBLUP results (Li *et al.* 2021). The newly proposed GAPLS provides an effective way of assessing individual sire's genomic contributions to six categories of PLS by using the standard GRM values between sires and the cows with PLS records. The merits of the method include: (1) results are easy to interpret, promoting adoption by commercial producers wanting to improve the reproductive performance of their herds; (2) the GAPLS method will not be influenced by the phenotype scoring system allowing producers to tailor selection strategies to their production system; (3) The population variation of GAPLS values will truly reflect the degree of relatedness between testing and reference populations; and, (4) The method could easily be extended to any phenotype of a categorical or ordinal nature.

Acknowledgements

The authors wish to acknowledge Meat and Livestock Australia (MLA) for their financial support to the project P.PSH.1211 'Validation of pooled DNA gEBV for Brahman commercial cow fertility', and Dr Brad Hine for his valuable suggestions to improve the paper.

References

Li Y., Porto-Neto L., McCulloch R., McWilliam S., Alexandre P., *et al.* (2021). Proc. Assoc. Advmt. Anim. Breed. Genet. 24:204-207. 52Li24204.pdf (aaabg.org)

Reverter A., Porto-Neto L.R., Fortes M.R., McCulloch R., Lyons R.E., *et al.* (2016). J Anim Sci. 94(10):4096-4108. http://dx.doi.org/10.2527/jas.2016-0675

VanRaden, P.M, 2008. J. Dairy Sci. 91(11):4414-4423. https://doi.org/10.3168/jds.2007-0980

317. Performance of identity-by-state and identity-by-descent genomic models for meat traits in a cross-pig F2 population

B.K. Angarita[1*], J.P. Steibel[2], R.J.C. Cantet[1] and N.S. Forneris[1]

[1]Universidad de Buenos Aires, Facultad de Agronomía, Departamento de Producción Animal, INPA-CONICET, Avenida San Martin 4453, C1417DSQ Buenos Aires, Argentina; [2]Michigan State University, Department of Animal Science and Department of Fisheries and Wildlife, 48824, East Lansing, MI, USA; angaritabarajas@uba.ar

Abstract

Dense genome-wide SNP chips allow computation of genomic relationships through different methods, either based on identity by state (IBS) or identity by descent (IBD). The latter have shown to increase the accuracy of both estimated relationships and predicted breeding values. Yet, it is unclear which approach would achieve higher performance in real data with a pedigree constrained to a few generations. Here we compare the performance of genomic models based on IBS (G_{IBS}) and IBD (G_{IBD}) for meat quality traits in an F2 crossbred pig population by estimating their predictive ability through cross-validation, compared with the pedigree model. The genomic based covariances had higher predictive ability when compared to A_{22}. A scarce increment in the predictive ability was found with G_{IBS} when compared to G_{IBD}, most likely due to the former method recovering relationships among founder animals.

Introduction

Prediction of breeding values (BV) in livestock genetic evaluation requires phenotypes and the pedigree through the inverse of the additive relationship matrix A. These relationships constitute twice the expected proportion of alleles identical by descent (IBD) shared by pairs of individuals. The availability of molecular markers allows estimating the so-called genomic relationship matrix G (VanRaden, 2008), which is used to estimate the genomic BV (GEBV). Different methods have been proposed to estimate genomic relationships (VanRaden, 2008; Yang *et al.*, 2010), being VanRaden's (2008) the most widely used. In this method, elements reflect the estimated proportion of alleles that are identical by state (IBS) between two individuals, as a deviation from the expected proportion of alleles shared in the population (Vela-Avitúa *et al.*, 2015). Alternative methods have been developed based in the proportion of the genome that is IBD among pairs of individuals. Han and Abney (2011, 2013) proposed a Hidden Markov Model for estimating the probability of IBD between pairs of individuals within a known pedigree. The method takes account of inbreeding and LD among markers. IBD based methods using genomic information have increased the accuracy of estimated relationships and predictions of breeding values (García-Baccino *et al.*, 2016; Jurcic *et al.*, 2021). The use of IBD assumes that the pedigree is complete. Unfortunately, it is not straightforward to think there will be an advantage of IBD over IBS in terms of predictive ability when the pedigree has low depth when using real population data. Therefore, this study compares estimates of predictive abilities through cross-validation, using IBD and IBS from records of meat quality and carcass composition of crossbred pigs with a pedigree constrained to a few generations.

Materials & methods

The experimental pig population was bred at the Michigan State University Swine Teaching and Research Center, East Lansing, MI. The pedigree and genotypes were available for three generations (F_0, F_1 and F_2) of a cross of Duroc × Pietrain. Briefly, the F_0 was formed by mating 16 Pietrain sows by artificial insemination to 4 Duroc boars. 6 males and 51 females were selected from F_1 as parents of the F_2 generation. F2 Pigs (n=948) from both sexes had phenotypes for 20 meat quality and carcass composition traits. The phenotypes were classified within five groups: primal cut weight, carcass measures, meat quality, laboratory analysis

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_317, © B.K. Angarita *et al.* 2022

and sensory panels analyses. Genotyping and editing procedures were previously reported by Casiró *et al.* (2017) and consisted of SNP data from two chips with different densities. All F_0 and F_1 animals, as well as 336 F_2 pigs were genotyped with the Illumina PorcineSNP60 BeadChip, which contained approximately 62,000 SNP. Additionally, 612 F_2 animals were genotyped at a lower density (8,836 SNP) using the GeneSeek Genomic Profiler – Porcine LD (GGP-Porcine LD, GeneSeek a Neogen Company, Lincoln, NE). The procedure for editing genotypes involved removing SNP with missing genotypes in all animals and those with Mendelian inconsistencies. Then, all missing genotypes in the 60K chip as well as those not scored in the low-density chip were imputed using the software FImpute with default parameters (Sargolzaei *et al.*, 2014). After the imputation, we additionally removed SNP with MAF<0.05, resulting in a dataset with records from 1,015 pigs having a genotype of 42,087 SNP each.

Relationship matrices. Three different methods were employed to estimate the relationship matrices. The first one was the additive relationship matrix A calculated with 1,037 animals contained in the pedigree and including the 1,015 genotyped animals. Only the submatrix A_{22}, with the additive relationship between genotyped individuals was used in this study. The second was the IBS method of VanRaden (2008) (G_{IBS}). Briefly, genotypes were expressed as allelic dosage and stored in standardized marker matrix M, with dimensions *n* (number of individuals) by *m* (number of loci). From here, G_{IBS} was computed by multiplying Z (the standardized marker matrix) by its transpose. The resulting matrix contains the estimates of the realized genomic relationships between any pair of pigs. The third was an IBD method proposed by Han and Abney (2011, 2013) and implemented in the 'IBDLD' v3.38.1 software. The algorithm employs a Hidden Markov Model for estimating the probability that a pair of individuals at a locus are IBD, given genomic and pedigree information. The output of the program is an empirical kinship coefficient per chromosome between pairs of individuals. Then, the actual genomic relationship between pairs of individuals for a genome of length *L* is calculated as twice the kinship coefficient weighted by the length of the chromosome. Because the resulting genomic relationship matrix may be indefinite, it was necessary to compute the nearest positive definite matrix (Cheng and Higham, 1998), and the latter was used as the IBD genomic relationship matrix (G_{IBD}) in all analyses.

Statistical model. For each character a single trait animal model was fitted to the data using the gwaR package in R (https://github.com/steibelj/gwaR) as y = Xb + Zu + e. In this model, *y* is the vector containing the phenotypes, *X* is the incidence matrix relating records to the vector of fixed effects b, which includes the sex, slaughter group and the carcass weight as a covariate. Matrix Z relates the records in *y* to the random vector of additive genetic effects u, and e is the vector of random errors distributed as $e \sim N(0, I\sigma_e^2)$, where I is the identity matrix and σ_e^2 is the error variance. The distribution of BV was assumed to be $u \sim N(0, G\sigma_u^2)$, where G is one of the relationship matrices A_{22}, G_{IBS} or G_{IBD} and σ_u^2 is the additive genetic variance. As the models only differed in the relationship matrix used for specifying the covariance of BV, we will call these models as A_{22}, G_{IBS} and G_{IBD} hereafter.

Estimating predictive ability through cross-validation. For each trait the models A_{22}, G_{IBS} and G_{IBD} were compared by estimating the predictive ability as the correlation between predicted and observed phenotypes through a 5-fold cross-validation procedure with in-house developed functions written in R. The process consisted of randomly splitting the full dataset for each trait into 5 subsets; in each iteration, the animal's phenotypes were hidden from one subset ('test set') and the phenotypes of the 4 remaining ones were used to train the model ('training set'). As an outcome of each iteration, we obtained estimates of variance components and the solutions for all fixed and random effects of the model. We used these to predict the phenotypes of the animals in the test set as $y2 = \mathbf{x}'\hat{\boldsymbol{b}} + \hat{u}$. Once the 5 iterations were finished, the Pearson correlation coefficient ($r_{(y1,y2)}$) between the observed uncorrected (*y1*) and predicted (*y2*) phenotypes was calculated. The procedure was replicated 300 times, and all replicates (n=213) in which

the Newton-Raphson algorithm converged (to maximum of 50 iterations) with the three models were used for comparison.

Results

For all the models, the predictive ability was high for traits within the primal cut weight group (0.80-0.87), from low to moderate for the group's carcass measures, meat quality and laboratory analyses (0.34-0.70), and low for the traits in sensory panel group (0.21-0.40). In terms of selection models, the predictive ability of G_{IBS} and G_{IBD} were higher when compared to A_{22} in 18 out of 20 traits from carcass measures, meat quality, laboratory analyses and sensory panel. The increment in predictive ability obtained by using genomic relationship matrices with respect to A_{22} was more noticeable in the meat quality group, which ranged from 14 to 37%. It was of lesser importance in the primal cut weights, where the increment varied from 0.6 to 4%. Comparing the performance between G_{IBS} and G_{IBD}, a slight increment in the $r_{(y1,y2)}$ for 14 traits was obtained when fitting G_{IBS} with respect to G_{IBD}, ranging from 0.1 to 5.5%.

Discussion

Predictive ability was greater with genomic models when compared to A_{22}. The result agrees with those from stochastic simulations (Forneris *et al.*, 2016; Vela-Avitúa *et al.*, 2015) or real data (Jurcic *et al.*, 2021; Meuwissen *et al.*, 2014), though some of these investigations employed different IBS or IBD approaches to set up the relationship matrices. The improved performance of genomic matrices over A_{22} may be due to a better capture of the Mendelian sampling term (Daetwyler *et al.*, 2013). In addition, de los Campos *et al.* (2013) found that the predictive ability depends on the extent to which the marker genomic relationship describes the realized genetic relationship at the unobserved causal loci. Other authors have highlighted the importance of information of relatives on the prediction of genomic models. Habier *et al.* (2010) and Clark *et al.* (2012) have found that prediction accuracy is reduced when relationships between individuals from the training and test sets are distant or animals are unrelated. In our study, the choice of animals from the training and test sets was random and the range of relationships between individuals in generation F_2 ranged from 0 to 0.75, -0.276 to 0.9, and -0.26 to 0.92, for A_{22}, G_{IBS} and G_{IBD}, respectively. By randomly picking animal records for cross-validation, the predictive ability is higher than when the sampling is constrained using criteria such as family and relatedness (Daetwyler *et al.*, 2012). In about 70% of traits predictive abilities were higher with G_{IBS} than with G_{IBD}, while in the remaining traits they performed the

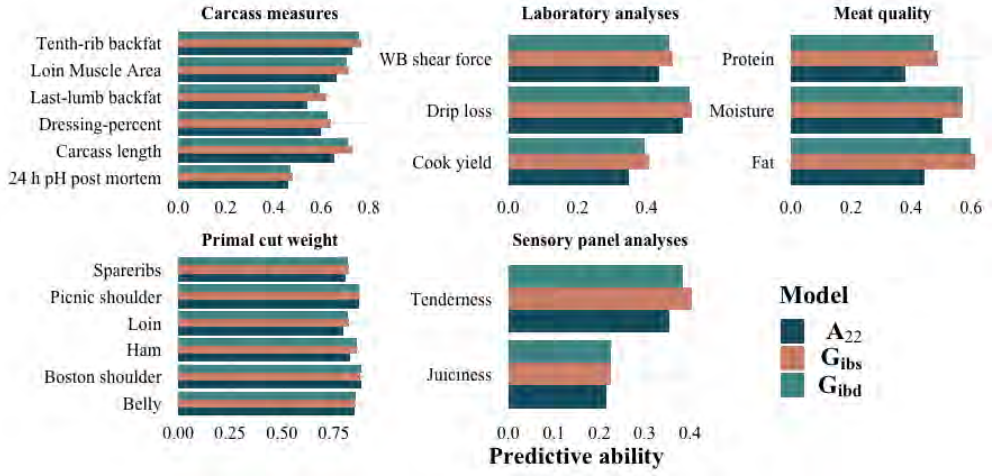


Figure 1. Predictive ability of meat quality and composition traits in F2 pigs using three linear mixed models.

same. The reason may be found in that G_{IBS} displays greater variation in genomic relationships, probably because relationships are referred to a more distant base population, which for an animal model would imply more sources of information with the consequent gain in precision to estimate the BV. Our results revealed sizeable relationships among F_0 animals and thus a deeper pedigree is important for more accurate predictions.

References

Casiró S., Velez-Irizarry D., Ernst C.W., Raney N.E., Bates R.O., *et al.* (2017). J. Anim. Sci. 95:545-558. https://doi.org/10.2527/jas2016.1003

Cheng S. H., and Higham N. J. (1998). SIAM 19, 1097–1110.

Clark S.A., Hickey J.M., Daetwyler H.D., and Van der Werf J.H.J. (2012). Genet. Sel. Evol. 44:4. https://doi.org/10.1186/1297-9686-44-4

Daetwyler H.D., Calus M.P.L., Pong-Wong R., de los Campos G., and Hickey, J.M. (2013). Genetics 193(2):347-365. https://doi.org/10.1534/genetics.112.147983

Daetwyler H.D., Swan A.A., Van Der Werf J.H., and Hayes B.J. (2012). Genet. Sel. Evol. 44(1), 1-11. https://doi.org/10.1186/1297-9686-44-33

De los Campos G., Vazquez A.I., Fernando R., Klimentidis Y.C., and Sorensen D. (2013). PLoS Genet. 9(7). https://doi.org/10.1371/journal.pgen.1003608

Forneris N.S., Steibel J.P., Legarra A., Vitezica Z.G., Bates R.O., *et al.* (2016). J. Anim. Breed. Genet. 133(6):452-462. https://doi.org/10.1111/jbg.12217

García-Baccino C.A., Munilla S., Legarra A., Vitezica Z.G., Forneris N.S., *et al.* (2016). J. Anim. Breed. Genet. 134(2), 109-118. https://doi.org/10.1111/jbg.12236

Habier D., Tetens J., Seefried F.R., Lichtner P., and Thaller G. (2010). Genet. Sel. Evol. 42(1), 1-12. https://doi.org/10.1186/1297-9686-42-5

Han L., and Abney M. (2011). Genet. Epidemiol. 35(1):557-567. https://doi.org/10.1002/gepi.20606.Identity

Han L., and Abney M. (2013). Eur. J. Hum. Genet. 21(2):205-211. https://doi.org/10.1038/ejhg.2012.148

Jurcic E.J., Villalba P.V., Pathauer P.S., Palazzini D.A., Oberschelp G.P.J., *et al.* (2021). Heredity. 127(2):176–189. https://doi.org/10.1038/s41437-021-00450-9

Meuwissen T.H.E., Odegard J., Andersen-Ranberg I., and Grindflek E. (2014). Genet. Sel. Evol. 46(1): 1-8. https://doi.org/10.1186/1297-9686-46-49

Sargolzaei M., Chesnais J.P., and Schenkel F.S. (2014). BMC Genom. 15(1). https://doi.org/10.1186/1471-2164-15-478

VanRaden P.M. (2008). J. Dairy Sci 91(11):4414-4423. https://doi.org/10.3168/jds.2007-0980

Vela-Avitúa S., Meuwissen T.H.E., Luan T., and Ødegård J. (2015). Genet. Sel. Evol 47(1). https://doi.org/10.1186/s12711-014-0084-2

Yang J., Benyamin B., McEvoy B.P., Scott G., Henders A.K., *et al.* (2010). Nat. Genet. 42(7):565-569. https://doi.org/10.1038/ng.608

318. Modelling missing pedigree with metafounders and validating ssGBLUP evaluation for Uruguayan Holstein dairy cattle

R.D. López-Correa[1,2*], A. Legarra[3] and I. Aguilar[4]

[1]Universidad de la República, Facultad de Agronomía, Montevideo, Uruguay; [2]Universidad de la República, Facultad de Veterinaria, Montevideo, Uruguay; [3]UMR GenPhySE, INRA Toulouse, Castanet Tolosan, France; [4]Instituto Nacional de Investigación Agropecuaria, Montevideo, Uruguay; rlopez@fagro.edu.uy

Abstract

The objective of this study was to check the quality of single-step GBLUP (ssGBLUP) predictions in Uruguayan dairy cattle, with special focus on the modelling of missing pedigree defined as metafounders (MF). A second objective was to use different estimators to compute relationship matrix among MF (Γ). Five strategies followed to fit unknown parent groups (UPGs) in ssGBLUP were: (1) MF using original Γ; (2) MF with modified Γ based on the median of elements in original Γ; (3) MF bounding base allelic frequencies of markers between 0 and 1 before computing Γ; (4) UPG in A^{-1}; and (5) UPG in H^{-1}. Estimated Γ with restricted allele frequencies gave less biased and more accurate genomic predictions. Future studies are warranted in the definition of Γ and UPGs to improve quality of genomic predictions and understand the origin of bias.

Introduction

Uruguayan dairy herds are creators of genetic progress through selection but are also heavy importers of foreign genetics. Use of genomic predictions is important as producers start using it, so it is important to ascertain the quality of these predictions. Genomic predictions can be done by ssGBLUP but it is challenging to include UPGs due to their specific structure. The use of UPGs in genomic model evaluations may lead to bias in genomic enhanced breeding values (GEBV) and reduce convergence speed in ssGBLUP (Misztal *et al.*, 2013; Tsuruta *et al.*, 2013). The application of MF theory (Legarra *et al.*, 2015; García-Bacino *et al.*, 2017) could be an alternative to fit missing pedigree in a genomic evaluation model and help to overcome bias and convergence problems. However, **Γ** may give estimation problems, for instance when genotyped individuals have a weak bond with UPGs (Lourenco *et al.*, 2020). Another important issue in the Uruguayan genetic evaluation is the restricted capacity to perform large progeny tests. Therefore, we need to define statistics that measure bias and accuracy of genomic predictions. Predictability is a major problem in dairy cattle genomic evaluations, because large differences between GEBV and EBV after progeny test were reported (Bradford *et al.*, 2019). Thus, the linear regression method (Legarra and Reverter, 2018), commonly known as LR, could be a useful tool to assess the predictive ability and accuracy of genomic predictions. The objective of this study was to check the quality of ssGBLUP predictions in Uruguayan dairy cattle, with special attention to the modelling of missing pedigree defined as MF. A second objective was to use different estimators to compute gamma relationship matrix among MF.

Materials & methods

Data. Data consisted of milk yield records from one to five lactations that were used for the Uruguayan Holstein genetic evaluation in April 2021. A total of 1,031,174 records were available for 367,423 cows. The pedigree file included 578,172 animals and was created by tracing back three generations of ancestors of either cows with lactation records or genotyped animals. In total, 18 UPGs were defined based on sex, origin (foreign, national) and birth year. Genotypes for 5,573 animals were available, 3,499 were generated using the ICBF_IDBv3 50k panel and 2,074 genotypes of bulls were provided by international cooperation. After quality control, 39,288 SNPs were analysed. Cows had the largest number of genotypes in the sample (3,275) and more than 95% had at least a single lactation record. They were born between 1993 and 2019, and 263 of the cows had both parents genotyped whereas 1,297 had one parent genotyped. Genotyped bulls

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_318, © R.D. López-Correa *et al.* 2022

(2,298) were born between 1973 and 2019, and 36 bulls had both parents genotyped and 1,181 bulls had one parent genotyped. Most of genotyped animals had both parents known.

Models. Milk adjusted 305-d yield was analyzed in a single-trait repeatability animal model to obtain GEBV using ssGBLUP (Aguilar *et al.*, 2010; Christensen and Lund, 2010). The basic model included fixed effects of herd year season (n=54,274) and age-parity-calving interval-dry period (n=118). The random effects were the additive genetic effect (n=578,190) and the permanent environmental effects (n=423,317). Data included up to five lactation records. Matrix H^{-1} was defined as in Aguilar *et al.* (2010). Genomic relationship matrix (G) was computed by the first method of VanRaden (2008); and blended with default parameters as $G^{*}=0.95\times G + 0.05\times A_{22}$. To make G and A_{22} compatible, we used default tuning in BLUPF90 (Misztal *et al.*, 2018). To fit UPG five methodologies were applied: ssGBLUP with MF using original Γ (MFO model), ssGBLUP with MF using gamma-robust estimator (MFGrob model), ssGBLUP with MF using gamma-complete estimator (MFBou model), ssGBLUP with UPG in A^{-1} (PED model), and ssGBLUP with UPG in H^{-1} (EXA model). For MF models, each UPG was regarded as a MF, i.e. a different ancestral population, and original Γ was computed by generalized least squares using observed genotypes (García-Bacino *et al.*, 2017) and for each MF we calculated the base allelic frequencies. Additionally, the gamma-robust estimator designed Γ with two different values based on the original Γ. The diagonal elements of Γ had the median of all self-relationships, whereas the off diagonals contained the median of all relationships across MF. The complete gamma estimator bounded base allelic frequencies of markers between 0 and 1 before computing Γ. Values in Γ^{-1} were included in the numerator relationship matrix $A(\Gamma)^{-1}$, as previously defined (Legarra *et al.* 2015). The PED model fitted UPGs as a fixed effect into ssGLBUP equations and they were included into the pedigree-based numerator relationship matrix A^{-1} by QP-transformation (Quaas and Pollak, 1981). The EXA model considered UPGs as a random effect into ssGLBUP equations and they were implemented in both A^{-1} and G^{-1} (Misztal *et al.*, 2013).

Validation of genomic predictions. To validate genomic predictions, we used 177 genotyped sires with at least 10 daughters with lactation records (valG). For those sires, we compared their GEBV predicted from the whole dataset ('w') with GEBV predicted from a partial dataset ('p'). The partial dataset removed phenotypes of cows recorded after 2014, so that records of daughters from those sires were not considered. GEBV of sires in valG with the whole dataset used 23,482 lactation records of 14,167 cows, 171 of which were genotyped. All GEBV were expressed in relation to the mean of GEBV for cows born in 2010 and with lactation records. GEBVs obtained with reduced data (GEBVp) and with whole data (GEBVw) were compared using statistics described in the 'LR' method (Legarra and Reverter, 2018). Predictions of GEBV and estimates of UPG effects were obtained by ssGBLUP using BLUP90iod2 program (Tsuruta *et al.*, 2001). The Γ was computed using gammaf90 program from BLUPF90 suite.

Results

Models. The original Γ showed higher extreme values than elements of Γ obtained with gamma complete estimator (Table 1). Using the gamma-robust estimator the self-relationship of MF (diagonal) was 0.776 and relationship between MF (off-diagonal) was 0.641.

Table 1. Statistics of estimated original Γ (MFO) and Γ with restricted allele frequencies (MFBou).

	MFO		MFBou	
	Diagonal values	Off-diagonal values	Diagonal values	Off-diagonal values
Minimum	0.661	-2.371	0.654	-0.259
Median	0776	0.641	0.762	0.5911
Maximum	6.302	1.354	1.478	0.709

Except for MFO, all models converged. However, EXA (average rounds: 86), MFBou and MFGrob (average rounds: 89) converged faster than PED (average rounds: 190).

Validation of genomic predictions. In Table 2, statistics of LR are described for each model, except for MFO. MBou model gave the lowest bias (108). All models had a slope that was far from one and it ranged from 0.672 to 0.699 across models. Correlations between GEBVw and GEBVp were similar for models and doubled the accuracy of GEBVs when using whole data instead of partial data.

Discussion

In this study, we assessed ssGBLUP for the genomic evaluation of Holstein in Uruguay and several methodologies were tested to include UPG in the genomic models. Bias was present in all models. MFBou, MFGrob and EXA models gave better genomic predictions than PED. In contrast, a simulated dairy cattle population, Bradford *et al.*,2019) obtained unbiased and accurate predictions with ssGBLUP and metafounders for the original Γ. Besides, Bradford *et al.* (2019) reported greatest bias and least accuracy for ssGBLUP with UPG for H in a moderate heritability trait. In our study a possible confounding among different UPGs could also happen because of a low number of animals contributing to estimate those UPGs (Misztal *et al.*, 2013). Except for MFO, all models were robust and converged, but overdispersion was observed. The low values of slope could be partly explained by the small number of bulls in the validation set (177 genotyped bulls). Our study showed a greater variation of values in the original Γ than those estimated for dairy cattle (Legarra *et al.*, 2015; Bradford *et al.*, 2019). This implies that in the Uruguayan Holstein genetic evaluation some MF represented very different base populations. The gamma complete estimator computes different values for elements of Γ. Thus, Γ would be a better proxy to reflect relationship between UPGs and we would be able to obtain less bias in genomic evaluations.

Conclusions

We found that ssGBLUP was successfully implemented in the Uruguayan Holstein genetic evaluation. Genomic models that used metafounders with the complete gamma estimator gave less biased and more accurate predictions. However, the origin of bias was not completely understood, so future studies are warranted in the definition of Γ and of UPGs to improve quality of genomic predictions.

Acknowledgements

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 772787.

Table 2. Bias ($\hat{\mu}_{w,p}$), slope ($\hat{b}_{w,p}$) and ratio of accuracies ($\hat{\rho}_{w,p}$) between GEBV estimated for each model using bulls in validation set (valG).

	ValG		
Model	($\hat{\mu}_{w,p}$)	($\hat{b}_{w,p}$)	($\hat{\rho}_{w,p}$)
MFGrob	118	0.672	0.520
MFBou	108	0.698	0.538
PED	296	0.629	0.487
EXA	122	0.699	0.514

References

Aguilar, I., I. Misztal, D. L. Johnson, A. Legarra, S. Tsuruta *et al.* (2010). J Dairy Sci 93(2): 743-752. https://doi.org/10.3168/jds.2009-2730.

Bradford, H.L., Y. Masuda, P. M. VanRaden, A. Legarra and I. Misztal (2019). Modeling missing pedigree in single-step genomic BLUP. J Dairy Sci 102(3):2336-2346. https://doi.org/10.3168/jds.2018-15434.

Garcia-Baccino, C.A., A. Legarra, O.F. Christensen, I. Misztal, I. Pocrnic *et al.*(2017). Genet Sel Evol 49(1): 34. https://doi.org/10.1186/s12711-017-0309-2.

Legarra, A., O.F. Christensen, Z.G. Vitezica, I. Aguilar and I. Misztal (2015). Genetics 200(2): 455-468. https://doi.org/10.1534/genetics.115.177014.

Legarra, A. and A. Reverter (2018). Genet Sel Evol 50(1): 53. https://doi.org/10.1186/s12711-018-0426-6.

Lourenco, D., Legarra,A., Tsuruta, S., Masuda, Y., Aguilar *et al.* (2020). Genes. 11, 790; https://doi.org/10.3390/genes11070790.

Misztal, I., S. Tsuruta, D.A. L. Lourenco, Y. Masuda, I. Aguilar, *et al.* 2018. Manual for BLUPF90 family of programs. Available at: http://nce.ads.uga.edu/wiki/

Misztal, I., Z.G. Vitezica, A. Legarra, I. Aguilar, and A.A. Swan (2013). J Anim Breed Genet 130(4): 252-258. https://doi.org/10.1111/jbg.12025.

Quaas, R. L. and E. J. Pollak (1981). J Dairy Sci 64(9): 1868-1872.

Tsuruta, S., I. Misztal, D.A.L. Lourenco, and T.J Lawlor. (2013). J. Dairy Sci. 97:5814–5821. http://dx.doi.org/10.3168/jds.2013-7821.

Tsuruta, S., I. Misztal, and I. Stranden (2001). J Anim Sci 79(5): 1166-1172. https://doi.org/10.2527/2001.7951166x.

VanRaden, P.M. J. Dairy Sci. 2008, 91, 4414–4423. https://doi.org/10.3168/jds.2007-0980.

319. Estimating metafounder relations for Finnish multibreed beef population

M. Taskinen*, A.-M. Leino, T.J. Pitkänen and E.A. Mäntysaari

Natural Resources Institute Finland, Genomics and breeding, Alimentum, Myllytie 1, 31600 Jokioinen, Finland; matti.taskinen@luke.fi

Abstract

The Metafounder (MF) concept was applied to Finnish beef cattle to harmonize the pedigree and genomic relationships. The multi-breed beef population was divided into MF groups according to available number of genotyped animals. A Gamma matrix (Γ) describing the relations between the MF groups was estimated from genomic information, augmenting the beef breeds with separately calculated allele frequencies from dairy breeds. To match the MF groups with larger number of unknown parent groups used in the breeding value evaluation, a covariance type polynomial fitting was constructed for the Γ matrix. We show that the pedigree relations can be adjusted using the Γ matrix of the MF concept to be, on average, close to the genomic relations.

Introduction

Finland is upgrading its breeding value evaluation scheme for beef cattle from traditional pedigree evaluations into genomic evaluation. As a part of this change, the current evaluation models have been changed from single to multi-breed evaluations since pure-breed populations are rather small in size and cross-breeds that link the populations are common. Further, practically all slaughter records from breeding and commercial herds can be utilized in the breeding value evaluation, improving the accuracy of the evaluations for weight traits through the genetic correlations (Pitkänen *et al.* 2021).

The project aims to build multi-breed, multi-trait single-step genomic evaluations for slaughter, weight and calving traits for pure and cross-bred Aberdeen Angus (aa), Charolais (ch), Hereford (hf), Limousin (li), and Simmental (si) animals. The pedigree structure is complicated and comprises a large number of unknown parent groups (UPG) to consider different genetic base of animals accurately. In such a setting, harmonizing pedigree (A) and genomic (G) relationship matrices according to assumptions of the same scale and equal base populations can be challenging (Christensen, 2012). We aim to obtain this goal by fitting the metafounder (MF) approach (Legarra *et al.*, 2015) to scale the A to fit G. The approach defines the base populations by specific MF groups with a certain level of self-relationships within MF and relationships across MF, contrary to the classical expectation of unknown, non-related base population in the mixed model equations (Schaeffer, 1994). Thereafter, a Gamma matrix (Γ) establishes both the relationships among the descendants of MF and the levels of inbreeding. Therefore, in addition to breed-wise differences, Γ elements should also describe possible changes with time in the MF base populations. This paper illustrates an idea of how to construct MF relations in Γ matrix for building A^{Γ} that is augmented with MF groups.

Materials & methods

In the MF concept, the MF groups replace the UPG in the pedigree. Metafounders fill-in missing parental relations in the pedigree as the MF groups are additionally related through the Γ matrix that is estimated from the genomic information.

Data and pedigree. As a starting pedigree for this work, we used a complete Finnish dairy and beef pedigree (n=10,416,246) from the Nordic dairy and beef routine evaluations because some suckler cows originated from dairy breeds. The pedigree was pruned for the multi-breed beef evaluation with RelaX2 software (Strandén

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_319, © M. Taskinen *et al.* 2022

and Vuori, 2006) by using the pedigree depth of five generations from the animals having either observations or genotypic information or both (n= 378,935). The pruned pedigree comprised 415,344 animals. The UPG (n=98) were created according to country of origin (domestic, foreign), breed group and birth year class. Breed groups of UPG were: aa, ch, hf, li, si, Blonde d'Aquitane, Highland Cattle, other beef breeds, Holstein, RDC and Finncattle, and Other that consisted, for example, of animals with unknown breed information. Animals whose main breed proportion were a minimum of 87.5% were considered pure-breds (aa 11%, ch 14%, hf 20%, li 11%, si 7%). Proportion of F1 animals in the data was 9% and other crosses and breeds 28%.

In total, 7,759 animals were genotyped (aa 1,739, ch 1,492, hf 1,495, li 1,293, si 902, other 838). The group other comprised animals that were not pure-bred. Animals were genotyped with the Euro G MD chip in which around 56 K SNP are accessible for the Nordic partners. The genotype data was cleaned with standard protocols and imputed with the FImpute software (Sargolzaei *et al.*, 2014). The final imputed genotype data comprised 39,608 SNP markers for the beef animals.

Metafounder groups. The grouping of the MFs depends on the number of available genotyped animals for each group. Most of the genotypes of this study were from the five most common pure beef breed animals (aa, ch, hf, li, and si). These breeds also had enough genotyped foreign animals to be able to build separate MF groups according to the country of origin. Because the pruned pedigree included some ancestors of the dairy breeds, three dairy MF groups were formed (RDC, Finncattle (FIC), and Holstein). All other pure- and cross-bred animals were placed to group 'Others'. In addition to breed wise differences, the genotypes should also describe the changes in time in the MF base populations. However, only the five domestic pure-bred MF groups had enough genotyped animals to be able to create more than one birth year classes. All other MF groups comprised a single year-class only.

Estimation of Γ matrix. In this study, information for building the Γ matrix was calculated using Bpop program (Strandén and Mäntysaari, 2020) that estimates allele frequencies (AF) of missing parent groups in a pedigree pruned to certain depth from the genotyped animals for the estimation of the Γ matrix. The pruning depth is here a crucial parameter in the estimation of the Γ matrix as the contribution of the genomic information to the MF groups depends on the distribution of the missing parents in the Γ estimation pedigree. With higher pruning depth most of the missing parents are located further away from the genotyped animals diluting the contribution and shifting the mean birth years of the missing parents of the groups further to the past. Here, the mean birth year of the missing parents determines the temporal location of the MF group.

Because of the low number of genotyped animals, AF were estimated in this study using pruning depth zero, defined as removing all but genotyped parents. This was done for all breed groups except for group Others. As the ancestors of cross-bred animals of the group Others were commonly pure-breds, the cross-bred genotyped animals contributed mostly to the pure-bred MF groups, not to the group Others. Therefore, AF of group Others were estimated from a pedigree that had no parental information at all. This is effectively the same as using the mean AF.

The Γ matrix was calculated from AF using the method of moments (Garcia-Baccino *et al.*, 2017), i.e. from $\Gamma=8 \times \text{var}(AF)$.

Augmenting Γ matrix with the dairy breeds. The Γ matrix was further augmented using blocks derived from AF of the dairy breeds (Kudinov *et al.*, 2020). Only the common SNPs of the Euro G MD chip were used for the off-diagonal blocks. Out of the 39,608 beef SNPs, 37,614 were common with RDC/FIC and 37,288 with Holstein. Furthermore, RDC/FIC and Holstein had 36,995 common SNPs.

Time trend fitting for Γ matrix. The multi-breed, multi-trait breeding value evaluation has more UPG than the Γ matrix constructed above. Especially since all breeds have more than one (birth) year class, whereas only the domestic pure-breds from the five common MF groups consisted of two year classes. In order to obtain matching MF groups for the other birth years, a covariance function fit (Kirkpatrick *et al.*, 1994) was constructed for the Γ matrix:

$$\Gamma = \phi K \phi^{\Gamma} \quad (1)$$

where each row of the ϕ matrix is

$$\phi_i = [0, \ldots, 0, y_i\text{-}d_i, 0, \ldots, 0, 1, 0, \ldots, 0] \quad (2)$$

having a linear time factor coefficient y_i-d_i with a mean birth year y_i of the MF group i and a 'base population year' d_i (described below), and a coefficient 1 for a constant term located on certain columns of the row-vector ϕ_i. The symmetric matrix K contains unknowns of linear time factors, as well as constant terms, and it can be solved using least-squares from:

$$K = (\phi^{\Gamma}\phi)^{-1}\phi^{\Gamma}\Gamma\phi(\phi^{\Gamma}\phi)^{-1} \quad (3)$$

Once the matrix K is solved, more Γ matrix values can be interpolated (and extrapolated) by replacing the years y_i in the Equation 1 with the mean birth years of the UPG of the evaluation. This effectively applies a bilinear function:

$$\Gamma_{i,j}(x,y) = a_{i,j}(x\text{-}d_i)(y\text{-}d_j) + b_{i,j}(x\text{-}d_i) + b_{j,i}(y\text{-}d_j) + c_{i,j} \quad (4)$$

as relation between breeds i and j where x and y are the mean birth years of the groups and $a_{i,j}$, $b_{i,j}$, $b_{j,i}$, and $c_{i,j}$ are corresponding solved unknowns of the K matrix.

A well-known problem in polynomial functions is their poor ability to extrapolate the predictions outside the data range. As the number of genotyped animals was quite small in this study, the polynomial fitting needed further restrictions for the within-breed group relation terms $\Gamma_{i,i}(y,y)$ to prevent them from growing too high in the past. For this, the time derivatives of the within-breed group relations $\partial_y \Gamma_{i,i}(y,y)$ were forced to be zero at the base population years, d_i, by setting the corresponding b terms to zero in the K matrix. In this study, d_i were set to 1988 for all breeds. Furthermore, the time factors of the domestic pure-bred MF groups were shared with the corresponding foreign MF groups in order to be able to fit the time trend of the domestic groups to the foreign groups as well.

Results

For the forthcoming beef cattle single-step evaluations, Γ matrix was interpolated / extrapolated to separate year-classes for all decades starting from 1980 except for the group Others that had only a single year class. Averages of Γ matrix adjusted pedigree relations (A^{Γ}) and genomic relationships (G) were then calculated for groups of genotyped animals. Figure 1 compares the former (left image) with the latter (right) by grouping the genotyped animals to separate year classes, given the breed group comprised enough genotyped animals. The A^{Γ} relations are on average quite close to the genomic relationship values.

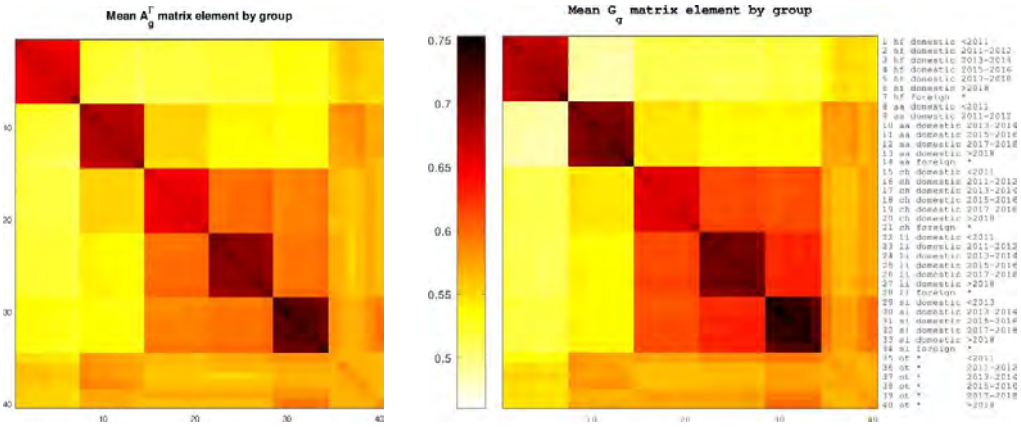


Figure 1. Averages of pedigree relations with metafounders (left) and genomic relationships (right) for genotyped animals by breed groups and year classes.

Discussion

It was shown that the modelling of complicated Γ matrix is possible, and the relationships using the MF concept become, on average, close to the genomic relations. The constructed Γ matrix can be further refined when the number of genotyped animals increases in the near future.

References

Christensen O.F. (2012). Genet. Sel. Evol. 44(1):37. https://doi.org/10.1186/1297-9686-44-37

Garcia-Baccino C.A., Legarra A., Christensen O.F., Misztal I., Pocrnic I., *et al.* (2017). Gen. Sel. Evolution, 49(1):1-14. https://doi.org/10.1186/s12711-017-0309-2

Kirkpatrick M., Hill W.G., and Thompson R. (1994). Genetics Research, 64(1):57-69. https://doi.org/10.1017/S0016672300032559

Kudinov A.A., Mäntysaari E.A., Aamand G.P., Uimari P., and Strandén I. (2020) J. Dairy Sci. 103(7):6299-6310. https://doi.org/10.3168/jds.2019-17483

Legarra A., Christensen O.F., Vitezica Z.G., Aguilar I., and Misztal I. (2015) Genetics 200(2): 455-468. https://doi.org/10.1534/genetics.115.177014

Pitkänen T.J., Taskinen M., Mäntysaari E.A., and Leino M. (2021) Interbull bulletin No. 56, Leeuwarden, the Netherlands.

Sargolzaei M., Chesnais J.P., and Schenkel F.S. (2014). BMC Genomics, 15:478. https://doi.org/10.1186/1471-2164-15-478

Schaeffer L.R. (1994). J. Dairy Sci. 77(9):2671-2678. https://doi.org/10.3168/jds.S0022-0302(94)77209-X

Strandén I. and Mäntysaari E.A. (2020). Agric. Food Sci. 29(3):166–176. https://doi.org/10.23986/afsci.90955

Strandén I., and Vuori K. (2006). Proc. of the 8th WCGALP, Belo Horizonte, MG, Brazil.

320. A deregression method for a single-step SNP BLUP model using all genotype data

Z. Liu[1*] and Y. Masuda[2]

[1]IT solutions for animal production (vit), Heinrich-Schroeder-Weg 1, 27283 Verden, Germany; [2]Rakuno Gakuen University, 584 Bunkyodai-Midorimachi, Ebetsu, Japan; zengting.liu@vit.de

Abstract

For various applications in dairy cattle evaluation, pseudo-phenotype data are needed for cows with own records or bulls with daughters. The purpose of this paper was to develop a GEBV deregression method for the single-step model using all genotype data. A special single-step SNP BLUP model was assumed in this study to deregress GEBV. All animals with own phenotype data and all genotyped animals including young candidates were considered in the GEBV deregression. The same pedigree file as well as the same genotype data were used in the deregression process as the original single-step evaluation. Thanks to the efficient single-step SNP BLUP model, the proposed GEBV deregression should be feasible for analysing millions of genotyped animals. Methodological and technical issues were addressed, and a validation procedure was proposed for the GEBV deregression method. Analysis of real data will be required to verify the developed GEBV deregression method.

Introduction

Pseudo-phenotype of animals are useful data for diverse genetic or genomic analyses. By definition, pseudo-phenotype data should ideally be corrected for all model effects, such as fixed and non-genetic random effects, so that only additive genetic effects and random errors remain in the pseudo-phenotype records. The special property of the pseudo-phenotype data makes genetic or genomic evaluations easier that will be based on the pseudo-phenotype records, because only the additive genetic effects need to be estimated. Daughter yield deviations (DYD, VanRaden and Wiggans 1990; Liu *et al.* 2004) and deregressed proofs (Jairath *et al.* 1998) are two commonly used forms of pseudo-phenotype data. Calculation of DYD for complex statistical models like the random regression test-day model (Liu *et al.* 2004) is technically challenging. In contrast, EBV or proof deregression can be applied to any linear function of estimated breeding values (EBV) to bypass complex evaluation models, such as multi-lactation random regression test-day model or maternal effects model. A single-trait single-effect model can be applied in most cases for the proof deregression. Therefore, the proof deregression provides an easier alternative than the DYD calculation for generating pseudo-phenotypic data. The main objective of this study was to derive deregressed GEBV of a single-step SNP BLUP model (ssSNPBLUP, Liu *et al.* 2014) for bulls with daughters or cows with own phenotypic records using all available genotype data.

Materials & methods

Estimation of breeding values of animals can be understood as a process of regressing phenotype data on additive genetic effects of the animals using pedigree information. The deregression of EBV is to reverse the regression process in the genetic evaluation process (Jairath *et al.* 1998) to obtain the pseudo-phenotype data for the animals with own phenotype records. In literature, many non-iterative deregression methods were proposed. However, the deregression methods on an animal-by-animal basis, termed as scalar deregression, were proven to be only sub-optimal by Calus *et al.* (2016). The iterative deregression method (Jairath *et al.* 1998), also called matrix deregression, uses the pedigree and was demonstrated to be accurate and the deregression process was proven to be reversible (Mark *et al.* 2002; Calus *et al.* 2016).

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_320, © Z. Liu and Y. Masuda 2022

A GEBV deregression method for the single-step SNP BLUP model. The deregression of conventional EBV (Jairath *et al.* 1998) makes use of the inverse of numerator relationship matrix A. Masuda *et al.* (2021) suggested using the inverse of the unified genomic relationship matrix H for deregressing GEBV of the single-step GBLUP model (Misztal *et al.* 2009). The idea of using H^{-1} for the GEBV deregression had also been independently proposed by Esa Mäntysaari and Zengting Liu of the Interbull working group Genomic-free EBV for MACE in 2019. A major concern of the GEBV deregression using the H^{-1} matrix was raised because of a fast-increasing, large number of genotyped cows seen in many countries like the United States and Germany. Therefore, we propose here a GEBV deregression method assuming the single-step SNP BLUP model (ssSNPBLUP, Liu *et al.* 2014) that can efficiently utilize all genotype data of up to millions of genotyped animals. For deregressing GEBV of the single-step genomic evaluation, we assume a special ssSNPBLUP model:

$$y = \mu 1 + u = e \tag{1}$$

where y is a vector of deregressed GEBV of animals with own phenotype data, 1 is a vector of 1s, μ is a general mean, u is a vector of GEBV for the animals with own phenotype data or for genotyped animals, and e is a vector of residuals. Denote σ_u^2 as additive genetic variance. The deregressed GEBV y are unknown and will be estimated. We assume that:

$$[var(\mathbf{e})]^{-1} = \mathbf{D}\sigma_e^{-2} = diag\{n_i\}\sigma_e^{-2} \tag{2}$$

where D is a diagonal matrix containing effective daughter contribution (EDC) of bulls with daughters or effective record contribution (ERC) of cows with own phenotype records on the animal-model basis, n_i is EDC or ERC for animal i, $i = 1, \ldots, n$, and n is the number of animals with phenotype data. σ_e^2 is residual variance. We use genotype data of all animals, including culled animals or young animals, for the GEBV deregression. In addition, all animals with any source of phenotype data are included in the deregression process. The definition of animals with own phenotype data depends on the applied single-step model. For the single-step evaluation with phenotype data coming exclusively from one country or population, the animals with own phenotype data are usually domestic cows with own phenotypic records. In the single-step evaluation without foreign phenotype data, bulls with daughters are not considered as animals with own phenotype data here rather as ancestors, because their domestic daughters are treated as animals with own records.

Mixed model equations for the GEBV deregression. As stated above for the single-step evaluation without foreign phenotype data, all cows with own phenotype data as well as all genotyped animals are treated as animals with data in the GEBV deregression. We denote the number of non-genotyped animals with own phenotype n_{NP}. Let n_T be the total number of genotyped or phenotyped animals. As usual, we define non-genotyped or genotyped animals as Group 1 or 2, respectively. Ancestors of the genotyped animals (Group 2) or phenotyped and non-genotyped animals (Group 1) are denoted as Group 0. Unknown parent groups (UPG) for the special ssSNPBLUP model (1) are set up via the Quaas-Pollak transformation according to Vandenplas *et al.* (2021a). We use matrix Q_2 to assign the genotyped animals to their UPG. Furthermore, we use a diagonal matrix for ERC for the non-genotyped animals with phenotype data (Group 1):

$$\mathbf{W}_1 = diag\{n_{11} \quad n_{21} \quad \cdots \quad n_{n_{NP}1}\} \tag{3}$$

Likewise, ERC for the genotyped animals of Group 2 are represented with a diagonal matrix as well:

$$\mathbf{W}_2 = diag\{n_{12} \quad n_{22} \quad \cdots \quad n_{n_g 2}\} \tag{4}$$

If a genotyped animal has no phenotype, e.g. a young calf, then its ERC is 0. Under the assumption of known SNP effects available, right hand side (RHS) of GEBV for all the genotyped animals can be modified prior to the deregression process, following the idea of Vandenplas *et al.* (2021b). Mixed model equations without the SNP effects are given in Equation 5 for the GEBV deregression.

$$\begin{bmatrix} \mathbf{1}'\mathbf{W}_1\mathbf{1} + \mathbf{1}'\mathbf{W}_2\mathbf{1} & \mathbf{0} & \mathbf{0} & \mathbf{1}'\mathbf{W}_1 & \mathbf{1}'\mathbf{W}_2 \\ & \lambda\mathbf{A}^{tt} + \lambda(\frac{1}{k} - 1)\mathbf{Q}_2{}'\mathbf{A}_{22}^{-1}\mathbf{Q}_2 & \lambda\mathbf{A}^{t0} & \lambda\mathbf{A}^{t1} & \lambda\mathbf{A}^{t2} - \lambda\left(\frac{1}{k} - 1\right)\mathbf{Q}_2{}'\mathbf{A}_{22}^{-1} \\ & & \lambda\mathbf{A}^{00} & \lambda\mathbf{A}^{01} & \lambda\mathbf{A}^{02} \\ & & & \mathbf{W}_1 + \lambda\mathbf{A}^{11} & \lambda\mathbf{A}^{12} \\ & symm. & & & \mathbf{W}_2 + \lambda(\mathbf{A}^{22} + \left(\frac{1}{k} - 1\right)\mathbf{A}_{22}^{-1}) \end{bmatrix} \times \begin{bmatrix} \hat{\mu} \\ \hat{\mathbf{t}} \\ \hat{\mathbf{u}}_0 \\ \hat{\mathbf{u}}_1 \\ \hat{\mathbf{u}}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{1}'\mathbf{W}_1\mathbf{y}_1 + \mathbf{1}'\mathbf{W}_2\mathbf{y}_2 \\ -\lambda\frac{1}{k}\mathbf{Q}_2{}'\mathbf{A}_{22}^{-1}\mathbf{Z}\hat{\mathbf{g}} \\ \mathbf{0} \\ \boldsymbol{\Delta}_1 \\ \boldsymbol{\Delta}_2 + \lambda\frac{1}{k}\mathbf{A}_{22}^{-1}\mathbf{Z}\hat{\mathbf{g}} \end{bmatrix} \tag{5}$$

where $\hat{\mathbf{g}}$ is a vector of SNP effect estimates assumed to be known without error, t denotes additive genetic effects of UPG, Z is a design matrix of order $n_g \times m$ containing all genotypes, n_g is the number of the genotyped animals of Group 2, m is the number of SNP markers fitted, and A_{22} is pedigree relationship for the genotyped animals, k is the proportion of residual polygenic variance not explained by the SNP markers, $\lambda = \sigma_e^2/\sigma_u^2$ is the variance ratio. RHS for the non-genotyped animals with own phenotypes (Group 1) or the genotyped animals (Group 2) are respectively:

$$\boldsymbol{\Delta}_1 = \mathbf{W}_1\mathbf{y}_1 = \{n_{i1}y_i\} \tag{6}$$

$$\boldsymbol{\Delta}_2 = \mathbf{W}_2\mathbf{y}_2 = \{n_{i2}y_i\} \tag{7}$$

where n_{i1} or n_{i2} represent EDC or ERC of the i-th animal of Group 1 or 2, respectively.

Following the idea of the iterative deregression method by Jairath *et al.* (1998), the input GEBV of the n_T animals with either own phenotype or genotype data must be kept unchanged when solving Equation 5, only GEBV of the ancestors (Group 0 animals), the unknown parent groups, and the general mean need to be estimated.

When the solutions of Equation 5 are converged, RHS for the non-genotyped animals with own phenotype data (Group 1) are calculated as:

$$\Delta_1 = \mathbf{W}_1\mathbf{1}\hat{\mu} + \lambda\mathbf{A}^{10}\hat{\mathbf{u}}_0 + (\mathbf{W}_1 + \lambda\mathbf{A}^{11})\hat{\mathbf{u}}_1 + \lambda\mathbf{A}^{12}\hat{\mathbf{u}}_2 \tag{8}$$

RHS for the genotyped animals (Group 2) are computed with:

$$\Delta_2 = \mathbf{W}_2\mathbf{1}\hat{\mu} + \lambda\mathbf{A}^{20}\hat{\mathbf{u}}_0 + \lambda\mathbf{A}^{21}\hat{\mathbf{u}}_1 + \left(\mathbf{W}_2 + \lambda\left(\mathbf{A}^{22} + \left(\tfrac{1}{k} - 1\right)\mathbf{A}_{22}^{-1}\right)\right)\hat{\mathbf{u}}_2 - \lambda\tfrac{1}{k}\mathbf{A}_{22}^{-1}\mathbf{Z}\hat{\mathbf{g}} \tag{9}$$

If animal *i* with own phenotype is not genotyped, its deregressed GEBV is:

$$\hat{y}_i = \Delta_{i1}/n_{i1} \tag{10}$$

Otherwise, if the animal *i* with phenotype data is genotyped, its deregressed GEBV is:

$$\hat{y}_i = \Delta_{i2}/n_{i2} \tag{11}$$

where Δ_{i1} or Δ_{i2} are the i-th element of Equation 8 or 9 corresponding to animal *i*, respectively. Because animals without own phenotype data have zero EDC or ERC, only animals with own phenotype data have a deregressed GEBV.

A validation of the GEBV deregression method for the single-step evaluation. A designated genomic evaluation based on the special single-step SNP BLUP model 1 needs to be conducted for validating the proposed GEBV deregression method. The input pseudo-phenotype data for the validation study are the deregressed GEBV for all animals that had own phenotype data in the original single-step evaluation. All genotyped animals that have been evaluated in the original single-step model must be included in this validation, too. For genotyped animals without own phenotypes, e.g. young candidates, their phenotype data are treated as missing in the validation study, because those genotyped animals without own phenotype records have no defined deregressed GEBV. The SNP effect estimates from the original single-step evaluation must be assumed to be known without error. Mixed model Equation 5 can be solved as a pure conventional model. After the solutions of Equation 5 are converged, GEBV of all the animals with phenotype or genotype data are compared to the original single-step genomic evaluation. If the GEBV from the validation evaluation are equal to those from the original single-step evaluation, then the GEBV deregression process is proven to be correct and the GEBV degression process is verified to be reversible.

Results & discussion

Generation of accurate pseudo-phenotype data is required by numerous statistical analyses in animal breeding. We have developed a GEBV deregression method for single-step evaluation based on a single-step SNP BLUP model (Liu *et al.* 2014). All animals with phenotype data and all animals with genotypes, including genotyped young candidates, must be considered in the GEBV deregression process as in the original single-step evaluation. We use the SNP effect estimates from the original single-step evaluation for the deregression process. Thanks to the high efficiency of the ssSNPBLUP model, this GEBV deregression method is feasible for handling millions of genotyped animals.

One major concern about the deregressed GEBV is that they may lead to an inflation of the conventional MACE evaluation. This inflated variance of evaluation can be avoided if we can exactly reverse-engineer the single-step evaluation in the GEBV deregression process. Obviously, the single-step evaluation and the following GEBV deregression must not include foreign bulls' phenotype data for the purpose of submitting national deregressed GEBV of bulls to the current conventional MACE evaluation.

In this study we made only a theoretical contribution to the topic of GEBV deregression for the single-step model. The proposed deregression method must be verified by applying a reversibility test to filed data.

References

Calus M. P. L., Vandenplas J., Ten Napel J., and Veerkamp R. F. (2016). J Dairy Sci. 99(8):6403-6419. http://dx.doi.org/10.3168/jds.2016-11028

Jairath L., Dekkers J.C.M., Schaeffer L.R., Liu Z., Burnside E.B., *et al.* (1998). J. Dairy Sci. 81(2):550-562. https://doi.org/10.3168/jds.S0022-0302(98)75607-3

Liu Z., Reinhardt F., Bünger A., and Reents R. (2004). J. Dairy Sci. 87(6):1896–1907. https://doi.org/10.3168/jds.S0022-0302(04)73348-2

Liu Z., Goddard M. E., Reinhardt F., and Reents R. (2014). J. Dairy Sci. 97(9):5833-5850. https://doi.org/10.3168/jds.2014-7924

Mark T., Fikse W. F., Emanuelson U., and Philipsson J. (2002). J. Dairy Sci. 85(9):2393–2395. https://doi.org/10.3168/jds.S0022-0302(02)74320-8

Masuda Y., Liu Z., and Sullivan P. (2021). Interbull Bulletin 56:142-146.

Misztal I., Legarra A., and Aguilar I. (2009). J. Dairy Sci. 92(9):4648-4655. https://doi.org/10.3168/jds.2009-2064

Vandenplas J., Eding H., and Calus M. P. L. (2021a). J. Dairy Sci. 104(3):3298-3303. https://doi.org/10.3168/jds.2020-19460

Vandenplas J., Eding H., and Calus M. P. L. (2021b). Interbull Bulletin 56:22-29.

VanRaden P.M., and Wiggans G.R. (1991). J. Dairy Sci. 74(8):2737-2746. https://doi.org/10.3168/jds.S0022-0302(91)78453-1

321. Variance component estimation for single-step genomic BLUP for Australian terminal sire sheep

P.M. Gurman[1*], L. Li[1], A.A. Swan[1], N. Moghaddar[2] and J.H.J. van der Werf[2]

[1]AGBU, a joint venture of NSW Department of Primary Industries and University of New England, 2351, Armidale, Australia; [2]University of New England, NSW 2351, Australia; pgurman@une.edu.au

Abstract

Many national genetic evaluation systems have transitioned to single-step genomic BLUP (SS-GBLUP) for breeding value estimation utilising variance components estimated from pedigree-based REML. The genomic and numerator relationship matrices are often weighted in SS-GBLUP by a parameter λ, that minimises bias and increases accuracies and is set to 0.5 in Australian sheep analyses. This study estimates variance components from genomic relationship matrices weighted by λ ranging from 0 to 1 by genomic REML. The impact of using these variances for each λ was then examined in SS-GBLUP models via cross-validation metrics. Data on terminal sire sheep in Australia were analysed using univariate models for carcase and live weight traits. We found that the maximum log-likelihood was estimated at lambda values between 0.5 and 0.875, while cross-validation results suggest that increasing accuracies can be achieved with increasing λ towards one and no significant change to bias using estimated variance components.

Introduction

National genetic evaluations have largely transitioned to single-step genomic BLUP where sufficient genotypes are available. This transition occurred for Australian sheep in 2017 (Brown *et al.*, 2018). Variance components for these analyses have often reused the same components that were derived from, and used in pedigree-based BLUP analyses. Further, it was identified that blending the genomic relationship matrix (G) with part of the numerator relationship matrix (A_{22} for genotyped animals) ($\lambda G + (1 - \lambda)A_{22}$) helped to increase accuracy and reduce the bias that was observed when including genomic data (McMillan and Swan, 2017; Zhang *et al.*, 2017), either correctly weighting the genomic and pedigree information, or underutilising the genomic information. Recent studies have investigated using either sampling-based methods or genomic REML (GREML) based methods to fit two effects simultaneously to examine how the variance partitions between the two effects (with the ratio between the two effects being λ), a pedigree-based genetic effect using the numerator relationship matrix and a genomic-based genetic effect using the single-step relationship matrix H (Samaraweera *et al.*, 2021; Torres-Vázquez *et al.*, 2021). This study investigates the impact of G constructed for a range of λ values on variance component estimates, log-likelihood values, and SS-GBLUP cross-validation metrics to determine what is the optimal value of λ for use in SS-GBLUP analyses.

Materials & methods

Data for 9,688 terminal sire sheep were extracted from the Sheep Genetics LAMBPLAN terminal sire analysis, along with pedigree information on their ancestors. These animals were selected for simultaneous phenotypic recording for carcase and live weight traits including intramuscular fat (IMF, %), shear force (SF5, newtons), carcase eye muscle depth (CEMD, mm), carcase c-site fat (CCFAT, mm), carcase weight (CWT, kg) and post-weaning weight (PWT, kg). These animals are mostly from the Sheep Cooperative Research Centre Information Nucleus Flock (van der Werf *et al.*, 2010) and the Meat and Livestock Australia Resource Flock populations which form the genomic reference population for Australian sheep, so all animals were genotyped and were recorded for all phenotypes Phenotypes were pre-adjusted as part of routine evaluation for combinations of birth type, rearing type, age of measurement, age of dam, and hot carcase weight, depending on the trait. Contemporary groups for IMF, SF5, CEMD, CCFAT, and CWT

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_321, © P.M. Gurman *et al.* 2022

were based on kill group, while PWT was based on breed, flock, management group and sex. A breed-adjusted genomic relationship matrix was constructed (Makgahlela *et al.*, 2013), which utilised multiple sets of allele frequencies to reduce breed level misalignment between pedigree and genomic relationship matrices.

Univariate variance component estimation was performed using MTG2 (Lee and Van der Werf, 2016). The model used was $y = X\beta + Zu + ZQg + Zm + \varepsilon$ where y is the vector of pre-adjusted phenotypes for fixed effects; X is the design matrix for the fixed effects (contemporary groups); β is the vector of contemporary group solutions, Z is the design matrix for the random effects associated with breeding values for individual animals (animals with records by breeding values), which in this case is I; u is the vector of random additive genetic effects with $N(\mathbf{0}, \boldsymbol{G}_w\sigma_a^2)$; G_w is the weighted G $G_w = \lambda G + (1 - \lambda)A_{22}$); Q is the matrix of genetic group proportions, defined by the breed of origin from the pedigree; g is the vector of random genetic group effects with $N(\mathbf{0}, \boldsymbol{QQ}'\sigma_g^2)$; M is the design matrix for the random maternal permanent environment effects; m is a vector of random maternal effects only fitted for two weight traits (CWT and PWT) with $N(\mathbf{0}, \boldsymbol{MM}'\sigma_m^2)$ and ε is the vector of residuals effects with $N(\mathbf{0}, \boldsymbol{I}\sigma_e^2)$. All random effects were fitted as an animals by animals matrix, to accommodate MTG2. For GREML, the value of λ was varied between 0 and 1 in steps of 0.05.

The single-step cross-validation was performed as a 5-fold analysis. The 9,688 animals were randomly allocated into one of the five folds, stratified by breed and sire family. This process was repeated five times resulting in five replicates of five-folds and 25 SS-GBLUP analyses, where the phenotypes for the validation animals were removed from the analysis and EBVs calculated. These cross-validation analyses followed the same model as the GREML analyses, except that the inverse of the appropriate joint relationship matrix, H^{-1}, was used instead. Five cross-validation metrics are presented, for which we define the following for the validation animals: y^* as the phenotypes adjusted for contemporary group solutions from the model with appropriate λ; $\hat{u}_p$ as the EBVs from the partial analyses (data removed for the validation animals); $\hat{u}_w$ as the EBVs from the whole analyses (data retained for the validation animals), K as the subblock of H for the validation animals, and $\sigma_{a,\infty}^2$ as the genetic variance at equilibrium in a population under selection, assumed for simplification to be σ_a^2 and h is the square root of the heritability. Traditional accuracies were calculated as $cor(y^*, \hat{u}_p)/h$ and the dispersion as the regression of phenotype on EBV. Metrics based on Legarra and Reverter (2018) using the method LR (from 'linear regression') were also calculated: LR accuracies as:

$$\sqrt{\frac{cov(\widehat{\mathbf{u}}_p, \widehat{\mathbf{u}}_w)}{(diag(\boldsymbol{K}) - \bar{\boldsymbol{K}}) \times \sigma_g^2}};$$

LR bias as $(\overline{\widehat{\boldsymbol{u}}_p} - \overline{\widehat{\boldsymbol{u}}_w})/\sigma_a$ and LR dispersion as $cov(\widehat{\boldsymbol{u}}_p, \widehat{\boldsymbol{u}}_w)/var(\widehat{\boldsymbol{u}}_p)$.

Results

The variance components estimated for each value of λ are presented in Figure 1, along with a vertical dotted line at the value of λ where the maximum log-likelihood value was observed, between 0.5 and 0.85 across the six traits. The genetic variance increased from a pedigree only model (i.e. λ=0), reaching a peak around λ=0.25 and decreasing as λ continued to increase. The residual variance followed an inverse pattern to the genetic variance. For most traits, the maximum log-likelihood was observed at the λ value where the genetic variance was similar to that observed at λ=0. It should also be noted that the genetic group variances were not, as is often assumed, equal to the genetic variances. For IMF, SF5 and CEMD, the ratio of genetic group variance to the genetic variance was between 0.06 and 0.6, while for the two weight traits (CWT and PWT) the ratio was between 1.97 and 4.4. The only trait where the ratio crossed one over the range of λ was CCFAT, which was between 0.76 and 1.52.

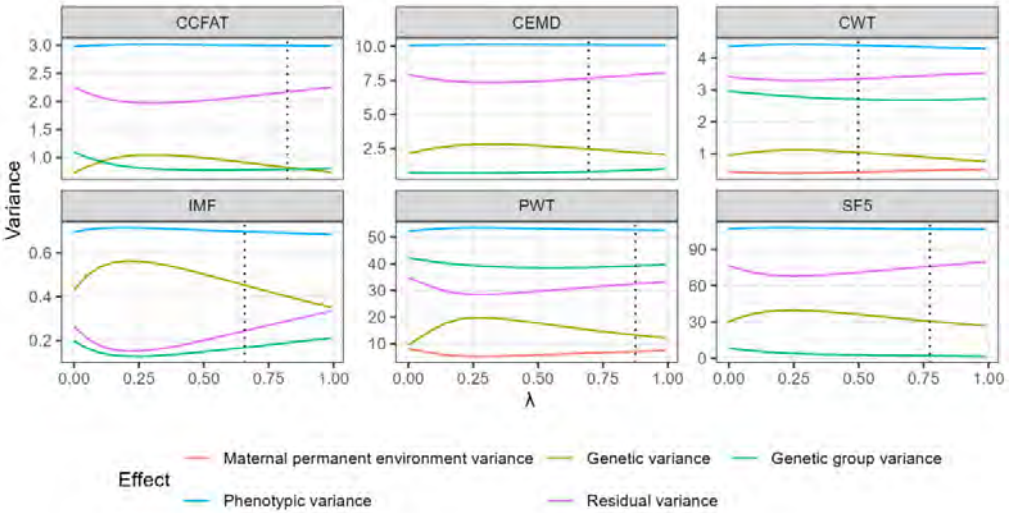


Figure 1. Variance components estimated for each value of λ. Vertical dotted lines indicate the λ value where the optimal log-likelihood value was estimated. Trait abbreviations are defined in the Materials & methods.

For the cross-validation analyses (Figure 2), both the phenotypic and LR accuracies show a mostly linear increase as the value of λ increases, although for λ values <0.3 for some traits a slight increase in accuracy is observed. For most traits except CWT, slight increases in traditional dispersion were observed as λ increased. This was not observed in the LR dispersions, with these dispersions being consistent over the range of λ. Finally, the LR biases were largest for CWT, CCFAT and PWT, though the latter decreased slightly at higher values of λ. No significant patterns concerning λ were observed.

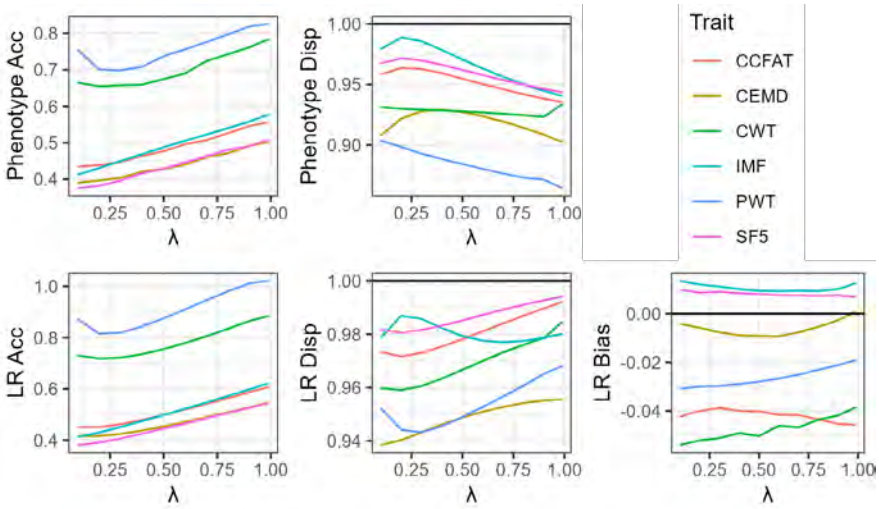


Figure 2. Cross-validation metrics calculated as the mean of the metric within each cross-validation group. Trait abbreviations are defined in the Materials & methods.

Discussion

Of note here is that the log-likelihood results and the cross-validation results provide different indications to the optimal model, due in part to the decreasing genetic variances as λ approaches one. Methods for accommodating individual λ values per trait optimally have been developed by Meyer in a paper submitted to this congress. It is unclear why the highest genetic variances were estimated around λ=0.25 which warrants further investigation. A more diverse set of traits would need to be studied to determine if the estimated heritabilities always decrease with high λ. These results would also need to be verified in models that utilise other methods for aligning A and G, e.g. metafounders and other scaling parameters (Christensen, 2012; Legarra *et al.*, 2015) and in other datasets, including for different traits and different genomic structures. The genetic group variances estimated here are of note, suggesting that where genetic groups are fitted as random, variances should be estimated for this effect and used in SS-GBLUP.

This paper finds that while log-likelihood values presented here suggest that individual λ values per trait are optimal, cross-validation suggests that using only genomic information is optimal if variance components are used that were estimated for the G used in SS-GBLUP. Further work is warranted to validate these results in other datasets.

Acknowledgements

Work was funded by Meat and Livestock Australia project L.GEN.1815.

References

Brown, D.J., Swan, A.A., Boerner, V., Li, L., Gurman, P.M., *et al.* (2018) Proc. of the 11th WCGALP, Auckland, New Zealand.

Christensen, O.F. (2012) Genet. Sel. Evol., 44:37. https://doi.org/10.1016/S0167-5877(00)00182-3.

Lee, S.H. and van der Werf, J.H.J. (2016) Bioinformatics, 32(9):1420–1422. https://doi.org/10.1093/bioinformatics/btw012

Legarra, A., Christensen, O.F., Vitezica, Z.G., Aguilar, I. and Misztal, I. (2015) Genetics, 200(2): 455–468. https://doi.org/10.1534/genetics.115.177014

Legarra, A. and Reverter, A. (2018) Genet. Sel. Evol., 50(1):53. https://doi.org/10.1186/s12711-018-0426-6

Makgahlela, M.L., Strandén, I., Nielsen, U.S., Sillanpää, M.J. and Mäntysaari, E.A. (2013). J. Dairy Sci., 96(8):5364–5375. https://doi.org/10.3168/jds.2012-6523

McMillan, A.J., and Swan, A.A. (2017) Proc of the 22nd AAABG, Townsville, Australia.

Samaraweera, A.M., Torres-Vázquez J.A., Jeyaruban M.G., Johnston, D.J. and Boerner, V. (2021) Proc of the 24th AAABG, Adelaide, Australia.

Torres-Vázquez, J.A., Samaraweera, A.M., Jeyaruban, M.G., Johnston, D.J., and Boerner, V. (2021) Proc of the 24th AAABG, Adelaide, Australia.

van der Werf, J.H.J., Kinghorn, B.P. and Banks, R.G. (2010) Anim. Prod. Sci., 50(12):998–1003. https://doi.org/10.1071/AN10151

Zhang, Y.D., Swan, A., Johnston, D.J., and Girard, C.J. (2017) Proc of the 22nd AAABG, Townsville, Australia.

322. Finetuning hyper-parameters increases the prediction accuracy in single-step genetic evaluation

M. Neshat[1,2,4*], Md.M. Momin[1,2,3,4], B. Truong[1], J.H.J. van der Werf [5], S. Lee[6] and S.H. Lee[1,2,4*]

[1]Australian Centre for Precision Health, Univ. of South Australia Cancer Research Institute, University of South Australia, Adelaide, SA, 5000, Australia; [2]UniSA Allied Health and Human Performance, Univ. of South Australia, Adelaide, SA, 5000, Australia; [3]Department of Genetics and Animal Breeding, Faculty of Veterinary Medicine, Chattogram Veterinary and Animal Sciences University (CVASU), Khulshi, Chattogram, 4225, Bangladesh; [4]South Australian Health and Medical Research Institute (SAHMRI), University of South Australia, Adelaide, SA, 5000, Australia; [5]School of Environm. and Rural Science, Univ. of New England, Armidale, NSW, Australia; [6]Div. of Animal Breeding and Genetics, National Institute of Animal Science (NIAS), South Korea; mehdi.neshat@unisa.edu.au; hong.lee@unisa.edu.au

Abstract

Single step genomic best linear unbiased prediction (HBLUP) has been widely used in livestock breeding. The HBLUP method (e.g. BLUPf90) requires hyper-parameters to combine genomic and pedigree relationships and these should be adequately initialised to maximise the accuracy of genomic prediction. In this study, we assess the performance of HBLUP, using various values of hyper-parameters in simulated genomic data. We show that the tuning parameter (tuning GRM relative to the pedigree-based numerator relationship matrix) considerably increases prediction accuracy, confirming previous studies. The scale factor, α, which scales the allele effect size by its frequency, also affects accuracy and the optimal scale factor can vary for each trait. In conclusion, fine-tuning the hyper-parameters of HBLUP is necessary to maximize prediction accuracy and the scale factor should be considered.

Introduction

Single step genomic best linear unbiased prediction uses a H-matrix that is a harmonised matrix of a pedigree-based numerator relationship matrix (NRM) and a genomic relationship matrix (GRM). The method is known as HBLUP. The H-matrix allows us to use the information of non-genotyped individuals in genomic prediction, using a data augmentation technique (see Legarra *et al.* 2009; Misztal *et al.* 2009). HBLUP has been widely used in the genetic evaluation of livestock (McMillan *et al.* 2017).

In HBLUP, there are several (hyper) parameters that can determine the performance of HBLUP.

Firstly, blending is important because it ensures GRM being a positive definite matrix (VanRaden 2008) thereby avoiding numerical problems in HBLUP (Legarra *et al.* 2009). Secondly, tuning is important because it adjusts the scale of GRM relative to that of NRMbefore inverting. Given that GRM is based on samples in the last few generations (genotyped individuals) whereas pedigree has been recorded from the founders, tuning can correct for this scale difference (Legarra *et al.* 2009; Miztal *et al.* 2009). Thirdly, parameters required to construct GRM may be important. A pairwise genomic relationship is the product of scaled genotypic coefficients of two random individuals (VanRaden 2008; Yang *et al.* 2010). Speed *et al.* (2012) generalised these forms, introducing a scale factor (α) that can determine the genetic architecture of a complex trait (aka heritability model).

In this study, we investigate the hyper-parameters required to estimate GRM to see how they affect the accuracy of HBLUP. First, various values of tuning and blending hyper-parameters were applied and compared to assess the performance of HBLUP, using simulated data with various scenarios of the historical population. In the analyses, we used the direct average information algorithm (Lee *et al.* 2006; Yang *et al.*

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_322, © M. Neshat *et al.* 2022

2011) that is robust to non-positive definite GRM so that we can assess all sorts of blending values including blending = 1. This is followed by exploring various values for the α of GRM that can determine the genetic architecture of a complex trait (heritability model).

Materials & methods

Simulated data. QMSim software (Sargolzaei *et al.* 2009) was used to simulate the historical population for 100 generations. In each generation, 50 males and 500 females were randomly selected and mated, generating 1000 offspring. Nine thousand biallelic markers in total were simulated, which were equally distributed across 30 chromosomes. Phenotypes were simulated for a complex trait, varying the scale factor (α=0 or -0.5) (Gowane *et al.* 2019). For the HBLUP analyses, the last five generations were recorded for phenotype and pedigree (n=5,000) and the last two generations were genotyped (n=2,000). To assess HBLUP accuracy, 1000 individuals were randomly selected from the last two generations as the target dataset, and the remaining 4,000 samples (1000 genotyped and 3,000 ungenotyped) were used as the discovery dataset.

Bending and tuning of GRM in HBLUP. The blending process can adjust the GRM to be a positive definite matrix (Legarra *et al.* 2009, Misztal *et al.* 2009). The adjusted GRM, referred to as $G_{blended}$, can be expressed as:

$$G_{blended} = \theta G + (1-\theta)A_{22} \quad \forall\ 0 \leq \theta \leq 1 \tag{1}$$

where θ is a positive coefficient to achieve a balance between GRM and the part of NRM (A_{22}) that is corresponding to the numerator relationships between the genotyped individuals.

Subsequently, the tuning process can be applied to adjust the scale of the GRM relative to that of the NRM. Following Legarra *et al.* (2014), the adjusted GRM is tuned as

$$G_{tuned} = \beta G_{blended} + \omega \mathbf{J} \tag{2}$$

where J is a matrix with the same size of GRM, and all elements are equal to one, and ω and β are computed as:

$$\omega = \frac{(I A_{22}\, I - I \hat{G} I)}{n^2} \qquad \beta = \frac{\frac{\left[\sum_{i=1}^{n} A_{22_{i,i}} - I' A_{22} I\right]}{n}}{\frac{\left[\sum_{i=1}^{n} G_{i,i} - I' G I\right]}{n}} \tag{3}$$

where I is an array with the size of $n \times 1$ and all values equal to one.

Genomic relationship matrix (GRM) and scale factor (α). It is assumed that the relationship between genetic variance and allele frequency can change depending on the evolutionary forces such as selections, mutation, migrations, and genetic drift. In the following equation, we can see that the variance of the ith genetic variant (v_i) is a function of the allele substation effect (β_i) and the allele frequency (p_i), which can be written as (Momin *et al.* 2021)

$$Var(v_i) = 2p_i(1-p_i)^{1+2\alpha} \times {\beta_i}^2 \tag{4}$$

where α is the scale factor that can determine the relationship between genetic variance and allele frequency, i.e. the genetic architecture of a complex trait. In the infinitesimal model (Falconer and Mackay 1996), α is assumed to be zero for all traits. An alpha value of -0.5, assumes that the genetic variance of the causal variant has a uniform distribution across tits minor allele frequency spectrum. However, it has been

reported that the value of α can vary, depending on trait and population (Momin *et al.* 2021; Speed *et al.* 2017; Speed *et al.* 2012). The generalised form of the GRM with the hyper-parameter of alpha (Speed *et al.* 2012) can be written as

$$G_{ij} = \frac{1}{d}\sum_{l=1}^{L}\left[(x_{il} - 2p_l)(x_{jl} - 2p_l)\right][2p_l(1-p_l)]^{2\alpha} \quad (5)$$

where G_{ij} is the genomic relationship between the i^{th} and j^{th} individuals, the number of SNPs is L and d is the expected diagonals and computed as:

$$d = L \cdot \mathbb{E}[(x_{il} - 2p_l)^2[2p_l(1-p_l)]^{2\alpha}] \quad (6)$$

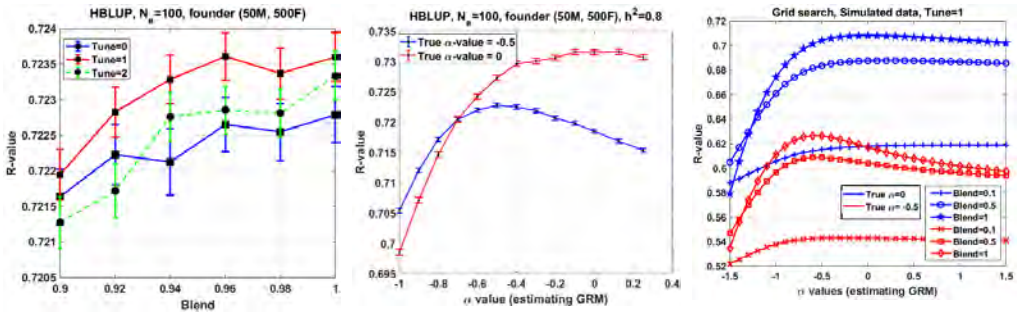
Statistical models. In this paper, we used genome-based restricted maximum likelihood (GREML) and HBLUP, based on linear mixed models, to predict individual breeding values (Gao *et al.* 2012).

Results

Figure 1a shows that the tuning process significantly improves the prediction accuracy (Pearson corelation coefficient (R-value) between true and estimated breeding value), confirming previous studies. Blending with θ<1 is not really improving the performance of HBLUP. The tuning parameter (tune=1, Eq. 3) based on Legarra *et al.* (2009) performed better than tune=2 (Eq. 3 except for β=1). Figure 1b shows that the choice of the α value (scale factor) is important and an optimal α can improve the prediction accuracy. As expected, the highest prediction accuracy is achieved when using the true α value. As shown in Figure 1c, the best configuration of the hyper-parameters could be obtained, using a grid search that considered blend, tune and alpha (scale factor) simultaneously.

Discussion

A blending strategy was not effective in improving the accuracy of HBLUP for various historical population scenarios (result not shown) although one study reported that blending could increase the prediction accuracy in dairy cattle (Gao *et al.* 2012). The second observation was that the tuning methods on the H-matrix could significantly increase the prediction accuracy. We compared three tuning methods



(a) **(b)** **(c)**

Figure 1. HBLUP accuracy and hyper-parameters. (a) The HBLUP accuracy (R-value) improves when using tune=1 (Equation 3) or tune=2 (Equation 3 except β=0 Vitezica *et al.* 2011). However, blending (θ<1) would not increase the accuracy for this simulated dataset. The error bars are 95% CI over 3,000 replications. (b) The HBLUP accuracy increases when using the true α values used in the simulation. The error bars are 95% CI over 3,000 replications. (c) The HBLUP accuracy for a single simulation replicate using a grid search method where the prediction accuracy was measured from 5-fold cross validation.

including tune=1 (based on Legarra's method (2009)), tune=2 (based on Vitezica *et al.* 2011), and tune=0 (i.e. without tuning). It is well known that alpha can vary across populations and traits (Momin *et al.* 2022). We show that the effect of α on the HBLUP accuracy was considerable. Our findings confirm that fine-tuning the hyper-parameters of HBLUP is necessary where the scale factor, a novel hyper-parameter in the context of HBLUP, should be considered. It is recommended that a grid search or similar optimisation algorithm should be used to find the best configuration for these hyper-parameters including blending, tuning and alpha.

Acknowledgements

This study is supported by Cooperative Research Program for Agriculture Science and Technology Development (PJ0160992022) from the Rural Development Administration, Republic of Korea.

References

Gao H., Christensen O.F., Madsen P., Nielsen U.S., Zhang Y., Lund M.S., and Su G., (2012). *Genet Sel Evol*, *44*(1): 1-8. https://doi.org/10.1186/1297-9686-44-8.

Gowane G.R., Lee S.H., Clark S., Moghaddar N., Al-Mamun H.A., and van der Werf, J.H. (2019). J Anim Breed Genet, 136(5): 390-407. https://doi.org/10.1111/jbg.12420

Lee S.H., and Van Der Werf J.H. (2006). *Genet. Sel. Evol*, *38*(1): 1-19. https://doi.org/10.1186/1297-9686-38-1-25

Legarra A., Aguilar I. and Misztal I. (2009). *J Dairy Sci*. 92(9): 4656-63. https://doi.org/10.3168/jds.2009-2061

McMillan A.J., and Swan A.A. (2017). In *Proc Assoc Advmt Anim Breed Genet*. Fitzroy, Australia

Misztal I., Legarra A., and Aguilar I. (2009). *J Dairy Sci*, 92(9): 4648-4655. https://doi.org/10.3168/jds.2009-2064

Momin M.M., Shin J., Lee, S., Truong B., Benyamin B., and Lee S.H. (2021). *bioRxiv*. https://doi.org/10.1101/2021.09.16.460619

Sargolzaei M. and Schenkel F.S. (2009) *Bioinformatics*, 25(5): 680–681. https://doi.org/10.1093/bioinformatics/btp045

Speed D., Cai N., Johnson M.R., Nejentsev S. and Balding D.J., (2017). *Nature Genetics*, 49(7): 986-992. https://doi.org/10.1038/ng.3865

Speed D., Hemani G., Johnson M.R. and Balding D.J., (2012) *Am J Hum Genet*. 91(6): 1011-1021. https://doi.org/10.1016/j.ajhg.2012.10.010

VanRaden P. M. (2008). *J Dairy Sci*, 91(11): 4414-4423. https://doi.org/10.3168/jds.2007-0980

Vitezica Z.G., Aguilar I., Misztal I., and Legarra, A. (2011). *Genet Res*, *93*(5): 357-366. https://doi.org/10.1017/S001667231100022X

Yang J., Benyamin B., McEvoy B.P., Gordon S., Henders A.K., *et al.*, 2010. *Nature Genetics*, *42*(7): 565-569. https://doi.org/10.1038/ng.608

323. Comparison of single-step and multi-step evaluations for U.S. milk, fat, and protein

R.R. Mota[1*], A. Cesarani[2] and P.M. VanRaden[3]

[1]Council on Dairy Cattle Breeding, 4201 Northview Dr Ste 302, Bowie, MD 20716, USA; [2]University of Georgia, 425 River Rd, Athens, GA 30602, USA; [3]USDA Animal Genomics and Improvement Lab, Bldg 5 BARC-West, Beltsville, MD 20705, USA; rodrigo.mota@uscdcb.com

Abstract

Predictive abilities of single-step (SS) and multi-step (MS) evaluations were compared using current and truncated domestic phenotypes for U.S. yield traits. Accuracy from SS was several percentage points higher than MS to predict either MS or SS current genomic estimated breeding values (GEBV) of 2,619 Holstein and 345 Jersey bulls from 4-year earlier data. Regressions for GEBV were close to 1.0 and similar for MS and SS for all traits and both breeds. Accuracy advantages of SS have increased since a previous study, perhaps indicating more bias from genomic preselection in MS. The MS methods put too much weight on pedigree evaluation, whereas the SS model had no polygenic effect and was more stable because new phenotypes have little influence on marker effects. Further testing will include more traits, foreign data from Interbull, and crossbreds to determine the potential to use SS methods.

Introduction

Most dairy cattle genomic evaluations still use MS methods, but recent research has improved both efficiency and accuracy of SS methods, and genomic preselection may increasingly bias MS methods. This study compares SS methods to official MS methods applied to the current or truncated domestic U.S. yield records. Previous reports (Masuda *et al.*, 2018a,b) included only Holstein records and less recent data than in the all-breed dataset now available. Masuda *et al.* (2018a) predicted 2015 protein daughter yield deviations from 2011 truncated Holstein phenotypes. They found little difference between SS and MS accuracy and regressions when unknown parent groups were applied only to pedigree but not to genomic relationships. Those earlier comparisons reported R^2 of 52% for SS and 51% for MS with corresponding regressions of 0.78 for SS and 0.81 for MS. Masuda *et al.* (2018b) compared SS vs MS trends and observed underestimation for the latter due to preselection in traditional PTA, especially for young bulls. The authors predicted that differences would further increase for genotyped bulls with the accumulation of preselection bias over the years, which would become more evident with shorter generation intervals and more recent data. Thus, comparisons were updated to determine if the theoretical advantages of SS (Aguilar *et al.*, 2010) now give larger actual advantages over MS when applied to national data.

Materials & methods

Genotypes were available from 5.1 million purebreds; genotypes of crossbreds were excluded from both SS and MS methods. The SS GEBV were estimated using APY method with 45,000 core animals: 15,000 each for Holstein (HO) and Jersey (JE) and 5,000 each for Ayrshire, Brown Swiss, Guernsey, with genotypes and phenotypes for all 5 breeds together (Cesarani *et al.*, 2022). The MS GEBV used current official methods with separate marker effects for each breed. During iteration, the MS used 0.6 million genotyped animals that had phenotypes or descendants with phenotypes, and the SS used 3.9 million genotyped animals that had phenotypes, descendants with phenotypes, or non-genotyped parents; for efficiency GEBVs of the remaining animals are computed only after convergence. Phenotypes were 305-d milk, fat, and protein lactation records solved in multi-trait models in SS and the first step of MS followed by single-trait marker effects in MS. The complete dataset included 99 million lactations beginning in 1960 for MS, whereas SS used a subset of lactations beginning in 2000. For both SS and MS, the complete dataset included calving

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_323, © R.R. Mota *et al.* 2022

dates up to June 2021, while the reduced dataset included calving dates only to June 2017 after truncating the last 4 years. The EBV from Interbull were not included. Thus, validation comparisons were limited to 2,619 U.S. HO and 345 U.S. JE bulls that had ≥100 daughters in the complete but no daughters in the reduced dataset. Other breeds had fewer than 25 such bulls and are not reported here. The unknown parent groups defined for MS were also used in SS but were not optimal after removing the first 40 and last four years of phenotypes; MS software automatically condenses groups outside the minimum and maximum year with phenotypes for each trait. The final GEBVs from both SS and MS were predicted from the earlier SS and MS GEBVs with simple regressions and with extra regressions (SS+ and MS+) on birth year and pedigree EBV from the respective SS or MS evaluation (VanRaden, 2021). Direct genomic values (DGV) from the MS evaluation were also examined because the SS did not include a polygenic effect and thus effectively only solved for a DGV. A MS evaluation with only 1% polygenic variance and no extra weight on pedigree EBV in the final blending step was computed for comparison.

Results

The pedigree models using SS or MS software had similar correlations of initial with later EBV as expected. The correlations ranged from 0.58 to 0.73 and were consistent with the large gain in reliability (REL) when bulls became proven. The GEBV correlations were much higher than for EBV as expected. Correlations from SS were higher than MS model for all traits and both breeds but correlations of the DGV from MS software were higher than from SS for traits of HO (Table 1). Regressions for pedigree models were all below the desired value of 1.0 and were similar using MS or SS software. Genomic preselection can reduce the regressions on pedigree EBVs because selection intensity may be much higher for animals with lower EBV, whereas GEBVs should be much less biased by preselection. Regressions for GEBV were close to 1.0 and similar for MS and SS for all traits and breeds. Regressions were also close to 1.0 for the MS DGV for HO but were lower for JE (Table 2). The standard deviations (SD) were similar for pedigree EBVs from MS and SS software, as expected. The SD were higher from final than truncated EBVs, also consistent with the large gains in EBV REL. The final GEBV had higher SD than the truncated GEBV in the SS and MS models, as expected.

Table 1. Squared correlations (×100) of final and truncated EBVs within the same evaluation system.

Breed	Trait	Genomic EBVs			Pedigree EBVs	
		MS	SS	MS DGV	MS	SS
Holstein	Milk	74	81	93	37	34
	Fat	78	83	95	42	34
	Protein	75	83	94	45	37
Jersey	Milk	78	83	86	53	45
	Fat	70	78	78	39	34
	Protein	73	83	80	49	46

Table 2. Regressions of final on truncated EBVs within the same evaluation system.

Breed	Trait	Genomic EBVs			Pedigree EBVs	
		MS	SS	MS DGV	MS	SS
Holstein	Milk	0.95	0.96	0.97	0.75	0.72
	Fat	0.96	0.96	0.95	0.85	0.76
	Protein	0.91	0.95	0.94	0.80	0.73
Jersey	Milk	1.05	0.95	0.94	0.91	0.84
	Fat	0.94	0.91	0.84	0.75	0.67
	Protein	1.00	0.98	0.88	0.86	0.82

Predictions of final MS or SS GEBV or DYD using truncated GEBV from the other software or with extra regressions provides more information about the benefits and biases, with results for milk shown as an example (Table 3). The reduced SS had higher R^2 than the reduced MS GEBV to predict either the MS or SS final GEBV for both HO and JE breeds; differences in R^2 were larger when predicting SS than MS. The extra regressions improved MS+ predictions compared to MS more than SS+ compared to SS, indicating smaller biases in SS as desired. The extra regressions were significant in SS but smaller in magnitude. The R^2 for MS+ and SS+ were very similar with the extra regressions included. The b_1 regression estimates were a little closer to 1.0 for MS than for SS in both breeds, but the MS+ estimates were very high for b_1 as previously shown (VanRaden, 2021) with the extra regressions included.

The SS and MS final GEBVs for milk were correlated by 0.98 and the reduced GEBVs were correlated by 0.96 for the 2,619 U.S. HO bulls that had ≥100 in 2021 but no daughters in 2017. The pedigree EBV from SS and MS software were correlated by 0.99 for final and 0.96 for the reduced EBV. Those differences are likely due to the known biases detected by the extra regressions, inclusion of polygenic effects, and slight differences in input records. The corresponding correlations for the 345 JE bulls were 0.98 for final GEBVs, 0.94 for reduced GEBVs, 0.98 for final EBVs, and 0.98 for reduced EBVs.

Genetic trend estimates were similar for MS and SS; correlations of birth year with GEBV for HO were 0.28 and 0.17 with reduced and final MS vs 0.29 and 0.16 with reduced and final SS. Correlations in the final data may be lower because the young bulls were highly selected. To predict later MS or SS GEBVs, the extra regressions showed that parent average (PA) should be subtracted from the reduced MS GEBVs and that the birth year trend was steeper in young bulls than proven bulls. The trend was also too steep and of same magnitude in the reduced SS GEBVs. With SS, the extra regression on PA was positive to predict MS GEBV but near 0 to predict SS GEBV. When comparing only bulls with 1000 daughters, 5 of the 10 largest differences of SS – MS were clones of famous bulls. Clones have identical DGVs, but their polygenic effects differ slightly due to being modelled as full sibs in the MS model. Thus, their GEBVs and daughter counts are copied to the clones in the official system. Different methods may be needed when using a genomic relationship matrix because clones make the matrix singular. The SS and MS GEBVs had small differences for the highest ranking HO bulls for net merit with >1000 daughters, indicating little bias after bulls are widely used. Biases may occur with early daughters from highly selected mates. The selection and mating patterns for HO may be more extreme and cause more biases than for JE or smaller breeds. The MS model with 1% polygenic variance and no extra weight on pedigree EBV gave GEBV more similar to SS with correlation increased to 0.97, nearly identical prediction R^2, and b_1 slightly closer to 1.0 than SS. Thus, previous revisions to the MS model to reduce overprediction of young animals (i.e. 10% polygenic variance and extra weight on PA) were a main difference between MS and SS predictions.

Table 3. Squared correlations (x100) and regressions to predict final MS and SS GEBV and MS DYD for milk from truncated MS and SS GEBV or with extra regressions on birth year and PA included (SS+ and MS+).

Breed	Predict:	Squared correlation from 2017:				Regression (b1) from:			
	2021	MS	SS	MS+	SS+	MS	SS	MS+	SS+
HO	MS	74	76	80	79	0.95	0.90	1.41	1.00
	SS	74	81	83	85	0.99	0.96	1.58	1.15
	DYD	60	63	69	68	0.93	0.90	1.52	1.08
JE	MS	78	80	80	81	1.05	0.96	1.47	0.96
	SS	74	83	79	84	0.99	0.95	1.58	1.09
	DYD	66	72	71	73	0.98	0.93	1.51	1.01

Discussion

The higher R^2 of SS to predict later MS or SS GEBV or DYD indicates better predictive ability than official MS for both HO and JE, but the b_1 regressions were a little closer to 1.0 for MS than for SS. The SS model without a polygenic effect may be too rigid and adapts less to new data of individual animals in large populations. Instead, new animals adjust the marker effects just a little, but their final GEBV mostly reflects the same historical marker effects they had when younger. For HO, the MS DGV were even more consistent than the SS GEBV. The earliest MS models in 2008 also used no polygenic effect but since 2010 have assumed 10% variance to avoid overprediction and obtain b_1 closer to 1.0. The MS model may now be assigning too much variance to the polygenic effect because recent enhancements and additions to the marker list should account for more of the genetic variance, but the SS model assigns too little polygenic variance (0%). Tests of German and Nordic SS models have assigned 30% polygenic variance to all traits (Alkhoder and Liu, 2021; Andersen *et al.*, 2021). The U.S. MS model has put extra emphasis on pedigree EBV in the final blending step; those weights were decreased from 15 to 10% for yield traits in August 2021. Other model changes may be needed to prevent overprediction of final genetic trend that occurred in both the MS and SS GEBVs of young bulls. Negative values for the b_0 intercept are also a problem but were not yet compared in this validation. Future comparisons will include foreign data and more traits. Comparison of MS and SS results should lead to better understanding and improvements to both systems. The SS software could simplify and improve multi-trait modeling. Switching the current run flow to SS software for all the traits will take some effort and more computation but could be worthwhile.

References

Alkhoder H. and Liu Z. (2021) Interbull Bull. 56:30-40.

Andersen T., Nielsen U., Mäntysaari E., Pösö J., Rius-Vilarrasa E., and Aamand G. P. (2021) Interbull Bull. 56:90-93.

Aguilar I., Misztal I., Johnson D.L., Legarra A., Tsuruta S., and Lawlor T.J. (2010) J. Dairy Sci. 93(2):743-752. https://doi.org/10.3168/jds.2009-2730

Cesarani A., Lourenco D., Tsuruta S., Masuda Y., Legarra A., Nicolazzi E.L., VanRaden P.M., and Misztal I. (2022) J. Dairy Sci. (submitted)

Masuda Y., Misztal I., VanRaden P., and Lawlor T. (2018a) J. Dairy Sci. 101(Suppl. 2):182(abstr. 171).

Masuda Y., VanRaden P.M., Misztal I., and Lawlor T.J. (2018b) J. Dairy Sci. 101(6):5194–5206. https://doi.org/10.3168/jds.2017-13310

VanRaden P.M. (2021) Improved genomic validation including extra regressions. Interbull Bull. 56:65–69.

324. Large-scale multiple-trait single-step marker model implementation

B.L. Harris*, R.G. Sherlock and M.A. Nilforooshan

Livestock Improvement Corp. LTD., Private Bag, 3016, Hamilton, New Zealand; bevin.harris@lic.co.nz

Abstract

An approach to solve large-scale multiple-trait single-step marker models is discussed. The approach uses a block-diagonal Jacobi preconditioned matrix within a preconditioned conjugate gradient solver. This approach has successfully been used for single-trait single-step marker models containing 10s of millions of animals in the pedigree and 100s of thousands of genotyped individuals. This study used data from the New Zealand national genetic evaluation for the milk production traits, combined with bovine SNP markers from Livestock Improvement Corporation. The multiple-trait models took approximately 2.5 times longer to converge than the single-trait first lactation models and required approximately two times the number of iterations. The true multiple-trait genomic breeding values had high correlations with approximated multiple-trait genomic breeding values calculated using selection index method that combined single-trait genomic breeding values. There was re-ranking of elite sires, even though the correlations between true and approximate multiple-trait genomic breeding values were high, suggesting that implementing multiple-trait single-step marker models would be worthwhile.

Introduction

In most livestock breeding programs, it is common to use genomic data consisting of thousands of single nucleotide polymorphisms (SNPs) within genetic evaluations to produce genomic breeding values (GBVs). More recent genomic evaluation methods allow the simultaneous modelling of genotyped and non-genotyped animals. There are two equivalent methods: single-step genomic BLUP (ssGBLUP), which directly estimates the genomic breeding values and single-step SNPBLUP (ssSNPBLUP) that directly estimates the random SNP effects. Both methods are computationally intensive for large-scale evaluations with 10s of millions of animals in the pedigree and 100s of thousands of genotyped individuals in the analysis. The computational demands and short turnaround times (<7 days) required by breeding programs to produce GBVs have limited the large-scale genomic evaluation to single-trait analyses.

It would be useful to obtain multiple-trait (MT) GBVs from large-scale genomic evaluations. Multiple-trait genomic evaluation of the animals enables the use of all available information on several traits to obtain more accurate and timely genomic evaluations. This is particularly useful for low heritable traits such as fertility measures in dairy cattle. Another advantage of MT genomic analysis is obtaining unbiased estimates for traits measured only on animals selected based on a correlated trait. For example, the evaluation of dairy cows' second lactation production yields occurs after the selection has been applied to the first lactation yields.

For this study, we used ssSNPBLUP despite ssGBLUP approach having better convergence properties (Vandenplas *et al.*, 2018). The size of the genomic component of the mixed model equations (MME) for ssSNPBLUP is determined by the number of SNP markers rather than the number of genotyped individuals as in ssGBLUP. In most large-scale applications, the number of genotyped individuals is far greater than the number of SNP markers used in the analysis, allowing ssSNPBLUP to have a lower memory requirement. The lower memory requirement was advantageous for our computing environment. We used a set of equations that include a specialised pre-conditioned conjugate gradient (PCG) (Vandenplas *et al.* 2018 and Mäntysaari *et al.*, 2020) for solving the MME, which overcomes the poorer convergence properties of ssSNPBLUP.

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_324, © B.L. Harris *et al.* 2022

The purpose of this study was first to evaluate the computational feasibility of MT ssSNPBLUP by extending a specialised PCG solver with a block-diagonal Jacobi pre-conditioned matrix to multiple traits. The specialised PCG had been successfully used to solve large-scale single-trait models in New Zealand multibreed dairy-cattle genomic evaluation. The second purpose was to determine the accuracy of selection index procedures used to combine single-trait GBV to approximate MT GBVs. This quantifies the accuracy loss using single-trait analyses compared to MT analyses.

Materials & methods

Data used in this study was from the New Zealand national dairy cattle genetic evaluation in December 2021 and was combined with bovine SNP markers from Livestock Improvement Corporation. The data included 32 million animals in the pedigree from multiple breeds and crosses. Most animals were Holstein-Friesian, Jersey, or their crosses, with Ayrshire's being the next largest breed group. Approximately 200k genotypes were included in the analysis, with males accounting for 5% and females 95% of the genotypes. Phenotypic records for milk, milk fat and protein yields from lactations 1- 6 were used in the milk production models. Lactations 1, 2 and 3-6 were considered as distinct traits for each yield phenotype.

The ssSNPBLUP generic model statement for all models and each trait was:

$$y = Xb + Z_g M_g m + Z_n u_n + Za + Zp + e \quad (1)$$

where y is the phenotype vector, X and Z are incidence matrices, Z_g and Z_n are rows of Z corresponding to genotyped (g) and non-genotyped (n) animals, M_g is the SNP marker matrix with each column centred to have a mean of zero, m is the vector of SNP marker effects with each column is centred to have mean of zero, b is the vector of fixed effects including hybrid vigour, genetic groups, and breed covariates, u is the vector of genomic breeding values for non-genotyped animals, a is the vector of polygenic effects, p is the vector of permanent environment effects, and e is the random residual effect. The model includes the polygenic effects because we assume that the SNPs do not capture all the genetic variance.

The models included fixed effects for contemporary-group (herd-year-age-season) and covariates for hybrid vigour, breed proportion, age at calving and the stage of lactation, and genetic groups for animals with missing parents. The genetic groups were assigned by sex of missing parent and birth year (in 5-year blocks). The genetic groups were fitted directly as fixed effects and not via the numerator relationship matrix. The MME were set up using the approach outlined by (Fernando *et al.*, 2016). The MME requires the inverse numerator relationship matrix for non-genotyped animals, containing >31 million animals. This inverse can be done efficiently (Fernando *et al.*, 2016) using sparse Cholesky decomposition and is suitable for large-scale parallel computing implementations. The Cholesky decomposition was pre-calculated before solving the MME.

The single-trait ssSNPBLUP models were solved using a specialised PCG solver with a block-diagonal Jacobi preconditioned matrix. The blocks in the pre-conditioner matrix were associated with the major factors fitted in the model: fixed effects, SNP marker solutions, animal genomic breeding values for non-genotyped animals, polygenic breeding values, and permanent environmental effects. The blocks contained the matrix inverses for each block. The strategies to obtain the matrix inverses involved Cholesky decomposition for the breeding value blocks and direct inverses for the other blocks with equation absorption. An extension of the block-diagonal Jacobi pre-conditioned matrix to multiple traits is straightforward. The blocks for the SNP marker solutions, non-genotyped animal genomic breeding values, polygenic breeding values and permanent environment effects included all the MME elements for all the traits. The multiple-trait block for the permanent environment effects has values only on

the diagonals of trait sub-blocks, allowing the entire block to be inverted directly and efficiently using a simple recursive Schur algorithm. The multiple-trait block for the fixed effects is a series of single-trait sub-blocks because of the zero error covariance among the traits associated with fixed effects. The block of equations for MT SNP marker random effects within the MME and its inverse was computed separately before the PCG solver.

The accuracy of MT GBV was compared with the accuracy of the approximate MT GBV derived from single-trait GBVs. A selection index approach was used to compute the approximate MT GBV. This approach is based on VanRanden (2001) to avoid double accounting of non-independent sources of information from each single-trait analysis in the approximate MT GBV. Comparisons were made by correlating the true and the approximate MT GBV and regressing the approximate MT GBV on the true MT GBV.

All the analyses were run on an Amazon r5d.24xlarge EC2 server with 48 Intel Xeon cores and 768 Gb of memory.

Results

The MT models all had ~300M equations and ~246M phenotypic records. The single-trait first lactation analyses all had ~99M equations and used ~60M phenotypic records. Table 2 provides the solving times in minutes and the number of iterations required to solve the equations using the specialised PCG solver with a block-diagonal Jacobi preconditioned matrix. The standard PCG convergence metric was used (Shewchuk 1994), with the convergence criterion set to $<10^{-11}$. For comparison, a traditional MT BLUP model for milk yield with no SNP markers took 504 iterations and 956 minutes to solve with the specialised PCG solver. The results from a comparison of true and approximate MT GBV for genotyped sires enrolled in the national genetic evaluation using linear regression across breeds are in Table 3. Figure 1 illustrates the convergence behaviour for multiple and single-trait (1st lactation) milk yield ssSNPBLUP.

Table 1. The number of iterations and time to solve the multiple (lactations treated as 3 traits) and single-trait (1st lactation) analyses.

	Multiple trait		Single trait	
Phenotype	**Number of iterations**	**Solving time (min)**	**Number of iterations**	**Solving time (min)**
Milk	1,391	4,320	924	1,435
Fat	1,120	4,143	746	1,269
Protein	1,221	3,972	801	1,275

Table 2. A comparison of true and approximate multiple-trait GBV for genotyped sires enrolled in the national genetic evaluation.

Phenotype	Trait	Correlation	Slope	Intercept
Milk	All[1]	0.987	0.97	-59.6
Fat	All[1]	0.985	1.00	-1.77
Protein	All[1]	0.989	0.96	-1.44

[1] Weighted average of all lactations.

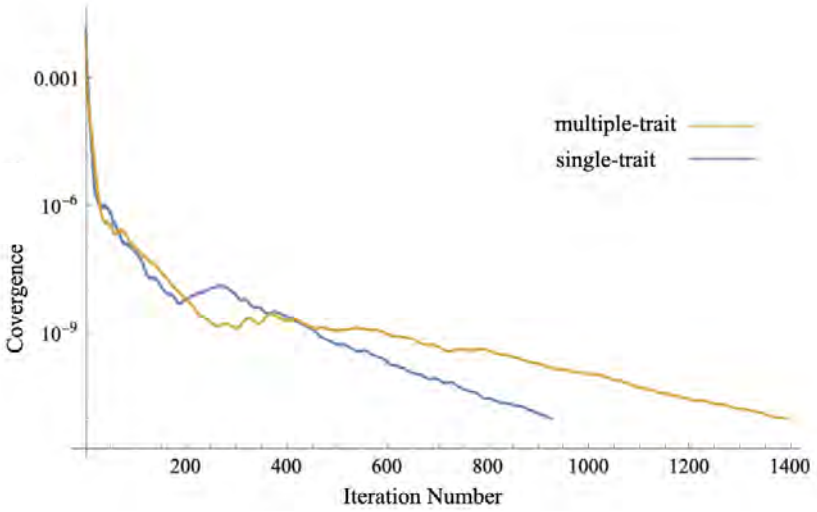


Figure 1. Convergence behaviour for multiple (lactations as 3 traits) and single-trait (1st lactation) milk yield ssSNPBLUP.

Discussion

Solving large-scale ssSNPBLUP models using PCG solvers is complex. A single trait model using a diagonal pre-conditioner matrix can take 10s of thousands of iterations and show poor or erratic convergence behaviour. The use of specialised PCG solvers can reduce the number of iterations and avoid erratic convergence behaviour. In this study, we extended the use of a block-diagonal Jacobi preconditioned matrix from a single trait to multiple traits. The convergence of the MT analyses achieved convergence with only ~50% increase in the number of iterations compared to the single-trait ssSNPBLUP models and ~2.5-fold increase in the number of iterations compared to the traditional MT BLUP model for the corresponding traits. The MT models took approximately 3 days to converge with no prior solutions. An elapsed time of 3 days is a good result. It is feasible to parallelise analyses for different phenotypes across several servers and achieve a 7-day turnaround for timely breeding decisions. More time reductions may be achieved by using prior solutions to reduce the number of iterations. Further work is required to test the effectiveness of prior solutions within ssSNPBLUP MT models.

The approximate MT GBVs were unbiased and highly correlated with the true MT GBVs. Even though the correlations were high, elite sires re-ranked. Thus, the use of true MT GBVs would be advantageous for genomic selection and genomic pre-screening.

References

Fernando R.L., Cheng H., Golden B.L., and Garrick D.J. (2016) Genet. Sel. Evol. 48:96 https://doi.org/10.1186/s12711-016-0273-2

E.A.Mäntysaari E.A., Koivulal M. and Strandén I. (2020) J. Dairy Sci. 103:5314 https://doi.org/10.3168/jds.2019-17754

An Introduction to the Conjugate Gradient Method Without the Agonizing Pain. Shewchuk J.R. (1994). https://www.cs.cmu.edu/~quake-papers/painless-conjugate-gradient.pdf

Vandenplas J., Eding E., Calus, M.P.L, and Vuik C. (2018) Genet. Sel. Evol 50:51 https://doi.org/10.1186/s12711-018-0429-3

VanRaden P.M. (2001) J. Dairy Sci. 84 (E. Suppl.) E47-E55

325. An iterative method for deregressed proof in single-step genomic BLUP and its application to a dairy cattle population

Y. Masuda[1*], Z. Liu[2] and P. Sullivan[3]

[1]Rakuno Gakuen University, Bunkyodai-Midorimachi 582, 069-8501 Ebetsu, Japan; [2]IT solutions for animal production (vit), Heinrich-Schroeder-Weg 1, 27283 Verden, Germany; [3]Lactanet, 555 Boul des Anciens-Combattants, Sainte-Anne-de-Bellevue, Quebec, Canada; yutaka@rakuno.ac.jp

Abstract

The objective of this study was to describe an iterative algorithm to deregress genomically-enhanced breeding values (GEBV) in single-step genomic BLUP (ssGBLUP). The method is an extension of the previous work by the first two authors, who gave a deregression method in singe-step SNP BLUP. The inverse of the unified relationship matrix (H^{-1}), which is a function of the inverse of genomic relationship matrix (G^{-1}) and the inverse of the additive relationship matrix (A^{-1}), was considered in the deregression method. Effective daughter contribution (ERC) for bulls or effective record contribution (ERC), which are functions of reliability of GEBV, should be used as weights on deregressed proof. The deregression algorithm consists of a series of matrix-vector multiplications, and the number of iterations is expected to be limited. The computational cost of this approach was discussed. Further research is needed to confirm the 'reversibility' of deregressed GEBV.

Introduction

The pedigree-based estimated breeding values (EBV) are often 'deregressed' to give pseudo-phenotypes for further genetic analyses, including international dairy sire evaluation, so-called multiple across country evaluation (MACE), genome-wide association study, and routine genomic evaluation using 'multi-step' methods. In dairy cattle, the multi-step methods optionally use MACE EBV as external information to improve the reliability of genomically-enhanced breeding values (GEBV) for genotyped animals. Accurate deregression methods are required to sustain the entire routine genetic and genomic evaluations system at the domestic and international levels.

Recently, some countries implemented a 'single-step' approach combining all available phenotypes, pedigree, and genotypes in a single system of mixed model equations, which provides GEBV for all genotyped and non-genotyped animals. Two classes of models have been suggested: single-step genomic BLUP (ssGBLUP; Aguilar *et al.*, 2010; Christensen and Lund, 2010) and single-step SNP BLUP (ssSNPBLUP; Liu *et al.*, 2014). The ssGBLUP model uses the inverse of unified relationship matrix (H^{-1}), which includes the inverse of genomic relationship matrix (G^{-1}), together with the inverse of additive relationship matrix (A^{-1}). Pedigree-based deregression methods (e.g. Jairath *et al.*, 1998) can be extended to account for the genomic relationships to give unbiased pseudo-phenotypes.

The objective of this study was to show an iterative algorithm to deregress GEBV in ssGBLUP. For ssSNPBLUP, Liu and Masuda (2021) comprehensively discussed how the deregression can be applied.

Materials & methods

Deregression model. For a particular trait, the deregression is based on the following model:

$$\mathbf{y} = \mathbf{1}\mu + \mathbf{u} + \mathbf{Q}\mathbf{t} + \mathbf{e} \quad (1)$$

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_325, © Y. Masuda *et al.* 2022

where y is a vector of deregressed GEBV (dGEBV), μ is the overall mean, u is a vector of GEBV, t is a vector of unknown parent groups (UPG), e is a vector of residuals, and 1 is the vector of ones, and Q is a matrix relating UPG to dGEBV. The additive genetic variance (σ^2_u), the residual variance (σ^2_e), and the variance ratio ($\lambda = \sigma^2_e/\sigma^2_u$) are assumed to be known.

Three groups for animals are assumed: group 1 is for non-genotyped animals with the phenotype(s), group 2 is for genotyped animals regardless of whether phenotyped or not, and group 0 for the ancestors of the animals in groups 1 and 2 (i.e. animals without genotypes and phenotypes).

A vector, u* = u + Qt, is split into three groups. Let

$$\theta = \begin{bmatrix} \mathbf{u}_0^* \\ \mathbf{u}_1^* \\ \mathbf{u}_2^* \\ \mathbf{t} \end{bmatrix} = \begin{bmatrix} \mathbf{u}_0 + \mathbf{Q}_0\mathbf{t} \\ \mathbf{u}_1 + \mathbf{Q}_1\mathbf{t} \\ \mathbf{u}_2 + \mathbf{Q}_2\mathbf{t} \\ \mathbf{t} \end{bmatrix},$$

where Q_0 is Q for group 0, Q_1 for group 1, and Q_2 for group 2. The inverse of additive relationship matrix for θ is

$$\mathbf{A}^* = \begin{bmatrix} \mathbf{A}^{00} & \mathbf{A}^{01} & \mathbf{A}^{02} & \mathbf{A}^{0t} \\ \mathbf{A}^{10} & \mathbf{A}^{11} & \mathbf{A}^{12} & \mathbf{A}^{1t} \\ \mathbf{A}^{20} & \mathbf{A}^{21} & \mathbf{A}^{22} & \mathbf{A}^{2t} \\ \mathbf{A}^{t0} & \mathbf{A}^{t1} & \mathbf{A}^{t2} & \mathbf{A}^{tt} \end{bmatrix}.$$

After applying QP-transformation (Misztal *et al.*, 2013), the inverse of unified relationship matrix with genotyped and non-genotyped animals is defined as

$$\mathbf{H}^* = \begin{bmatrix} \mathbf{H}^{00} & \mathbf{H}^{01} & \mathbf{H}^{02} & \mathbf{H}^{0t} \\ \mathbf{H}^{10} & \mathbf{H}^{11} & \mathbf{H}^{12} & \mathbf{H}^{1t} \\ \mathbf{H}^{20} & \mathbf{H}^{21} & \mathbf{H}^{22} & \mathbf{H}^{2t} \\ \mathbf{H}^{t0} & \mathbf{H}^{t1} & \mathbf{H}^{t2} & \mathbf{H}^{tt} \end{bmatrix} = \mathbf{A}^* + \begin{bmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & (\mathbf{G}^{-1} - \mathbf{A}_{22}^{-1}) & -(\mathbf{G}^{-1} - \mathbf{A}_{22}^{-1})\mathbf{Q}_2 \\ \mathbf{0} & \mathbf{0} & -\mathbf{Q}_2'(\mathbf{G}^{-1} - \mathbf{A}_{22}^{-1}) & \mathbf{Q}_2'(\mathbf{G}^{-1} - \mathbf{A}_{22}^{-1})\mathbf{Q}_2 \end{bmatrix} \quad (2)$$

where A^{-1}_{22} is the inverse of the additive relationship matrix for genotyped animals. The above matrix, H^*, can be modified to support random UPG, metafounders (Legarra *et al.*, 2015), alternative UPG models (Masuda *et al.*, 2022), and simplified G^{-1}, e.g. generated by the algorithm for proven and young (APY) or the Woodbury formula (Misztal *et al.*, 2014; Mäntysaari *et al.*, 2017).

Iterative algorithm. The system of mixed model equations under model (1) is

$$\begin{bmatrix} \mathbf{1}'\mathbf{W}_1\mathbf{1} + \mathbf{1}'\mathbf{W}_2\mathbf{1} & \mathbf{0} & \mathbf{1}'\mathbf{W}_1 & \mathbf{1}'\mathbf{W}_2 & \mathbf{0} \\ \mathbf{0} & \lambda\mathbf{H}^{00} & \lambda\mathbf{H}^{01} & \lambda\mathbf{H}^{02} & \lambda\mathbf{H}^{0t} \\ \mathbf{W}_1\mathbf{1} & \lambda\mathbf{H}^{10} & \mathbf{W}_1 + \lambda\mathbf{H}^{11} & \lambda\mathbf{H}^{12} & \lambda\mathbf{H}^{1t} \\ \mathbf{W}_2\mathbf{1} & \lambda\mathbf{H}^{20} & \lambda\mathbf{H}^{21} & \mathbf{W}_2 + \lambda\mathbf{H}^{22} & \lambda\mathbf{H}^{2t} \\ \mathbf{0} & \lambda\mathbf{H}^{t0} & \lambda\mathbf{H}^{t1} & \lambda\mathbf{H}^{t2} & \lambda\mathbf{H}^{tt} \end{bmatrix} \begin{bmatrix} \hat{\mu} \\ \widehat{\mathbf{u}_0^*} \\ \widehat{\mathbf{u}_1^*} \\ \widehat{\mathbf{u}_2^*} \\ \hat{\mathbf{t}} \end{bmatrix} = \begin{bmatrix} \mathbf{1}'\mathbf{W}_1\mathbf{y}_1 + \mathbf{1}'\mathbf{W}_2\mathbf{y}_2 \\ \mathbf{0} \\ \mathbf{W}_1\mathbf{y}_1 \\ \mathbf{W}_2\mathbf{y}_2 \\ \mathbf{0} \end{bmatrix} \quad (3)$$

where $\mathbf{y}_1$ is $\mathbf{y}$ for non-genotyped animals, and y_2 for genotyped animals. The weight matrices are W_1 = diag$\{n_{i1}\}$ and W_2 = diag$\{n_{i2}\}$, where n_{i1} (n_{i2}) is effective record contribution (ERC) of i-th non-genotyped (genotyped) animal. If an animal has no phenotype(s), the weight is set to 0.

This system of equations includes all genotyped animals considered in the original genetic evaluation because G^{-1} ties a genotyped animal to the other through the precise relationships among genotyped animals. The deregression method is based on the animal model, and y includes cows with phenotypes(s) only. For bulls, dGEBV are computed as Liu and Masuda (2021).

Following Jairath *et al.* (1998), dGEBV for cows (y_1 and y_2) can be available using an iterative method shown below.

1. Put $\hat{\mu} = 0$, $\widehat{\mathbf{u}_0^*} = \mathbf{0}$, $\widehat{\mathbf{u}_1^*} = \mathbf{0}$, and $\widehat{\mathbf{u}_2^*} = \mathbf{0}$.
2. Calculate $\widehat{\mathbf{u}_1^*} = \mathbf{u}_1 - \mathbf{1}\hat{\mu}$ and $\widehat{\mathbf{u}_2^*} = \mathbf{u}_2 - \mathbf{1}\hat{\mu}$, where u_1 (u_2) is a vector of GEBV, obtained by the original genetic evaluation, for non-genotyped (genotyped) animals.
3. Solve
$$\begin{bmatrix} \mathbf{H}^{00} & \mathbf{H}^{0t} \\ \mathbf{H}^{t0} & \mathbf{H}^{tt} \end{bmatrix} \begin{bmatrix} \widehat{\mathbf{u}_0^*} \\ \hat{\mathbf{t}} \end{bmatrix} = \begin{bmatrix} -\mathbf{H}^{01}\widehat{\mathbf{u}_1^*} - \mathbf{H}^{02}\widehat{\mathbf{u}_2^*} \\ -\mathbf{H}^{t1}\widehat{\mathbf{u}_1^*} - \mathbf{H}^{t2}\widehat{\mathbf{u}_2^*} \end{bmatrix}.$$
4. Calculate
$$\mathbf{z}_1 = \mathbf{W}_1\mathbf{y}_1 = \mathbf{W}_1\mathbf{1}\hat{\mu} + \lambda\mathbf{H}^{10}\widehat{\mathbf{u}_0^*} + (\mathbf{W}_1 + \lambda\mathbf{H}^{11})\widehat{\mathbf{u}_1^*} + \lambda\mathbf{H}^{12}\widehat{\mathbf{u}_2^*} + \lambda\mathbf{H}^{1t}\hat{\mathbf{t}}$$
$$\mathbf{z}_2 = \mathbf{W}_2\mathbf{y}_2 = \mathbf{W}_2\mathbf{1}\hat{\mu} + \lambda\mathbf{H}^{20}\widehat{\mathbf{u}_0^*} + \lambda\mathbf{H}^{21}\widehat{\mathbf{u}_1^*} + (\mathbf{W}_2 + \lambda\mathbf{H}^{22})\widehat{\mathbf{u}_2^*} + \lambda\mathbf{H}^{2t}\hat{\mathbf{t}}$$
And
$$m = \mathbf{1}'\mathbf{W}_1\mathbf{y}_1 + \mathbf{1}'\mathbf{W}_2\mathbf{y}_2 = \mathbf{1}'\mathbf{z}_1 + \mathbf{1}'\mathbf{z}_2.$$
5. Calculate $\hat{\mu} = m/(\mathbf{1}'\mathbf{W}_1\mathbf{1} + \mathbf{1}'\mathbf{W}_2\mathbf{1})$.
6. Go to Step 2 until convergence is reached.
7. After the convergence, calculate dGEBV as $\mathbf{y}_1 = \mathbf{W}_1^{-1}\mathbf{W}_1\mathbf{y}_1 = \mathbf{W}_1^{-1}\mathbf{z}_1$ and $\mathbf{y}_2 = \mathbf{W}_2^{-1}\mathbf{W}_2\mathbf{y}_2 = \mathbf{W}_2^{-1}\mathbf{z}_2$.

Evaluation of computing cost using real data. The first author evaluated the deregression cost for GEBV of the udder score in Japanese Holstein. Data included more than 750,000 phenotypes, 157,000 pedigree animals, and 120,000 genotypes with 45,724 SNP markers. A single-trait ssGBLUP model, similar to the traditional animal model used in national genetic evaluation, was applied. The QP-transformed $\mathbf{H}^*$ (2) with 29 UPG was formed. The system of mixed model equations was solved by PCG using BLUP90IOD2 (Lourenco *et al.*, 2020). The ERC for each cow (Liu *et al.*, 2017) was calculated based on the reliability of GEBV; the reliability was calculated the posterior variance of GEBV using a Gibbs sampler. A custom software was developed for deregression.

Results & discussion

The current method is useful for a population with the ssGBLUP evaluations because pre-calculated $\mathbf{G}^{-1}$ and the other part of $\mathbf{H}^*$ are immediately available. The algorithm consists of a series of matrix-vector multiplications. Therefore, the number of iterations should be limited to get convergence. Although step 3 needs to solve a system of sparse equations, it can be solved inexpensively by typical iterative solvers, e.g. preconditioned conjugate gradient (PCG). Alternatively, the matrix factorization is feasible to solve the equations if the order of equations is small enough.

In ssGBLUP with real data, PCG required 607 rounds in 30 minutes to get converged. The Gibbs sampler generated 3,000 samples to compute the variance of GEBV within 12 hours. The deregression algorithm took a few minutes to be converged. The total computing cost for deregression is expected to be much lower than the original single-step evaluation. There are two possible concerns in deregression in practice. First, the reliability computation is a possible bottleneck. A new algorithm to approximate the reliability of GEBV in ssGBLUP (Bermann *et al.*, 2021) should remove the issue. Second, some UPG models in ssGBLUP would be unstable in solving mixed model equations. The options include refinement of UPG definition, alternative UPG models, or random UPG.

The deregressed proof should be reversible (Mäntysaari *et al.*, 2011) in the sense that the recalculated evaluations using dGEBV as phenotypes are identical to the original genomic evaluations. To ensure reversibility, Liu and Masuda (2021) suggested that a deregression formula should include all animals with the phenotype(s), all genotyped animals, the same pedigree animals used in the original single-step evaluation.

A deregression approach was derived from ssSNPBLUP by Liu and Masuda (2021). In theory, their method is applicable to GEBV calculated from any single-step model. However, both methods have not been tested for reversibility with real data. Further research is needed to verify reversibility and computational feasibility for both ways.

Acknowledgements

The first author thanks the Holstein Cattle Association of Japan (HCAJ) for providing us with the phenotypes, pedigree, and genotypes. Toshimi Baba (HCAJ Hokkaido Branch) is acknowledged for the initial preparation of phenotypic data.

References

Aguilar I., Misztal I., Johnson D.L., Legarra A., Tsuruta S., *et al.* (2010) J. Dairy Sci. 93(2):743-752. https://doi.org/10.3168/jds.2009-2730

Bermann M., Lourenco D., Misztal I. (2021) J. Anim. Sci. skab353. https://doi.org/10.1093/jas/skab353

Christensen O.F., Lund M.S. (2010) Genet. Sel. Evol. 42:2. https://doi.org/10.1186/1297-9686-42-2

Jairath L., Dekkers J.C.M., Schaeffer L.R., Liu Z., Burnside E.B., *et al.* (1998) J. Dairy Sci. 81(2):550-562. https://doi.org/10.3168/jds.s0022-0302(98)75607-3

Legarra A., Christensen O.F., Vitezica Z.G., Aguilar I., Misztal I. (2015) Genetics. 200(2):455-468. https://doi.org/10.1534/genetics.115.177014

Liu Z., Goddard M.E., Reinhardt F., and Reents R. (2014) J. Dairy Sci. 97(9):5833-5850. https://doi.org/10.3168/jds.2014-7924

Liu Z., VanRaden P., Lidauer M., Calus M., Benhajali H. *et al.* (2017) Interbull Bull. 51:75–85.

Liu Z., Masuda Y. (2021) Interbull Bulletin 56:41-51.

Lourenco D.A.L., Legarra A., Tsuruta S., Masuda Y., Aguilar I., *et al.*, (2020). Genes. 11(7):790. https://doi.org/10.3390/genes11070790

Mäntysaari E.A., Koivula M., Strandén I., Pösö J., Aamand G.P. (2011) Interbull Bull. 44:19-24.

Mäntysaari E.A., Evans R.D., Strandén I. (2017) J. Anim. Sci. 95(11):4728-4737. https://doi.org/10.2527/jas2017.1912

Masuda Y., VanRaden P.M., Tsuruta S., Lourenco D.A.L., Misztal I. (2022) J. Dairy. Sci. In Press. https://doi.org/10.3168/jds.2021-20293

Misztal I, Vitezica Z.G., Legarra A., Aguilar I., Swan A.A. (2013) J. Anim. Breed. Genet. 130(4):252-258. https://doi.org/10.1111/jbg.12025

Misztal I., Legarra A. Aguilar I. (2014) J. Dairy Sci. 97(6):3943-3952. http://dx.doi.org/10.3168/jds.2013-7752

326. Efficient approximation of reliability for genomic breeding values in large systems

H. Eding

CRV Animal Evaluation Unit, Wassenaarweg 20, 6843 NW Arnhem, the Netherlands; herwin.eding@crv4all.com

Abstract

This paper presents a simplified procedure to obtain approximations of genomically enhanced breeding value (GEBV) reliabilities. This procedure assumes direct genomic reliabilities are known from genomic validation studies and are roughly equal for all genotyped animals. Direct genomic reliabilities form an upper limit to the information added by genotyped relatives to ungenotyped animals. The concept of 'genotype confidences' is introduced to limit the propagation of EDC from genotyped relatives, such that the genomic EDC do not exceed the number of EDC corresponding to the direct genomic reliabilities. The algorithm, if implemented in a routine evaluation, only adds marginally to the runtime of an evaluation. The method is illustrated in a comparison with GEBV reliabilities from a routine evaluation using pseudo records.

Introduction

The Interbull working group for genomic reliabilities published a bulletin in 2017 (Liu *et al.*, 2017) in which a protocol was proposed to estimate GEBV in a single step evaluation. One attractive facet of this protocol was the fact that it treated genomic data as an addition to conventional sources of data (own, offspring records, pedigree), potentially avoiding the necessity to completely redesign reliability estimation software.

The Interbull protocol encompassed six separate steps:

1. Calculate SNP reliabilities
2. Derive DGV reliabilities
3. Adjust DGV reliabilities
4. Calculate genomic gain => EDC_{add} = $EDC_{dgv} - EDC_{conv}$
5. Propagation => EDC^*_{add} = prop(EDC_{add})
6. Calculate final reliabilities => EDC_{final} = $EDC_{conv} + EDC^*_{add}$

Where EDC_{add} are the difference in expected daughter contributions (EDC), due to genomics, observed in genotyped animals. EDC^*_{add} contain these differences for all animals, after propagation (VanRaden and Wiggans, 1991).

Step 1 approximates the reliability of each SNP-allele of each SNP in a external program running a Bayesian analysis (*cf* Zaabza *et al.* 2021). These are then used as input in step 2 to derive DGV-reliabilities which are subsequently adjusted for (reference-)population means in step 3 after which the EDC added by the genomic information extracted in step 4 by subtracting the EDC_{dgv} from EDC_{conv}, where the latter is the EDC of a BV estimate without genomic information (a conventional EBV). In step 5 the vector with EDC_{add} (containing added EDC for genotyped animals only) is sent off to a propagation subroutine, which calculates the spread of genomic information in the wider population with non-genotyped animals. After this step we have EDC_{add} for all animals in the evaluation (EDC^*_{add}). In step 6 we can derive the final GEBV reliability by adding EDC^*_{add} to EDC_{conv} and back-transforming these EDC to a reliability.

Steps 1 to 4 result in EDC_{add} for genotyped animals. In the original proposal, steps 1 to 4 rely on external programs and procedures. Moreover, the Bayesian methods (Zaabza *et al.* 2021) implemented in the

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_326, © H. Eding 2022

software proved to be too time-consuming, making this protocol not very attractive for routine evaluations. However, information about EDC_{add} for traits are already available from our GEBV validation work flow (Stoop *et al.*, 2014). Using this information allows to skip these steps and save much time during evaluation. Hence a GEBV reliability approximation, corresponding to the original intent of the Interbull protocol, but relying on already available EDC_{add} data is proposed.

Materials & methods

It may be noted that the key steps in the Interbull protocol are step 4-6: Calculate EDC_{add} for genotyped animals, propagate to ungenotyped animals and calculate final reliabilities from (propagated) added EDC and conventional EDC. Hence, our implementation was built around these key steps.

Basic algorithm. The implementation consists of the following steps:

1. Calculate conventional reliabilities, transform to EDC.
2. Calculate SNP 'genotype confidences'.
   a. Set genotyped animals to rate of imputation;
   b. Propagate genotype confidence.
3. Calculate DGV reliabilities for all animals; transform to EDC.
4. Sum conventional and genomic EDC to obtain final EDC; transform to GEBV reliabilities.

Conventional reliabilities are calculated using standard algorithms such as the method for calculating multi-trait EDC (MT-EDC) and transforming these to reliabilities of traits in a multi-trait evaluation devised by Liu *et al.* (2001).

Direct genomic reliability. The reliability of a single genotype is derived from validation results in the following manner: GEBV validation is performed on data of a group of validation animals; typically candidate animals without progeny or observations, only a genotype. In the GEBV validation procedure deregressed proofs (DRP) for a trait are used as the phenotype which the breeding values (BV) are predictor:
DRP = b0 + b1 BV

The regression is performed twice: Once on conventional EBV, usually PA, and once on genomic breeding value (GBV). The realized R^2 of these regressions can be interpreted as reliabilities of PA en GBV in predicting the DRP. The difference in R^2 is a measure of the information added by the genotype. The exact amount of information added by the genotype can be calculated by transforming the R^2 into EDC, given the heritability of the trait under consideration:

$$DRP = b0 + b1\ PA \quad => R^2_{PA} \quad => EDC_{PA}$$

$$DRP = b0 + b1\ GBV \quad => R^2_{GBV} \quad => EDC_{GBV}$$

And

$$EDC_{GBV} - EDC_{PA} = EDC_{geno}$$

The variable EDC_{geno} is subsequently transformed back to a reliability R^2_{geno} which is the mean reliability of the direct genomic effect for a particular trait. It is assumed that R^2_{geno} for a trait is approximately equal for all genotyped animals.

Genotype confidence. Since propagation is aimed at modelling the information on unknown genotypes of ungenotyped animals, given genotyped relatives, the concept of 'SNP genotype confidence' (R^2_{SNP}) is introduced. The genotype confidence of genotyped animals is set to the 'rate of imputation' (typically nearly 1). For ungenotyped animals genotype confidences are initially set to zero. Genotype confidences are transformed to EDC using an arbitrary h^2 and propagation is performed. The resulting vector is back transformed to genotype confidences R^2_{SNP}, which are (very nearly) 1.0 for genotyped animals (particularly those with many genotyped offspring) and <1.0 for ungenotyped animals with few genotyped relatives. The reliability of the DGV for each (genotyped or ungenotyped) animal *i* for a trait *t* is: $R^2_{DGV,it} = R^2_{SNP,i} \times R^2_{geno,t}$

Transforming R^2_{DGV} to EDC for each animal and all traits gives EDC^*_{add} (propagated added EDC) which are added to the EDC from conventional information (EDC_{conv}): $EDC_{final} = EDC^*_{add} + EDC_{conv}$ => R^2_{final} (GEBV reliability)

Results

As an illustration of the quick approximation proposed the genomic reliabilities are presented from an evaluation of the traits Milking Speed and Temperament. The results are from two types of evaluation: (1) a routine genomic pseudo-record evaluation (rel PSR GEBV); and (2) a single step SNP BLUP evaluation using the method described above (rel SNPBLUP). Both evaluations consisted of 7,300,000 pedigreed animals, 4,770,000 with phenotypes, 624,000 of which were genotyped for a panel of 37,998 SNP.

Reliabilities showed good agreement from 2010 onwards. This coincides with the moment genotyping of the dairy population in the Netherlands became more common. Before 2010 we see higher reliabilities from the SNPBLUP evaluation. These were caused by propagation of genomic EDC from genotyped animals to ungenotyped relatives and agrees with the implicit imputation in a SNPBLUP evaluation, which means that for ungenotyped animals at least some genotype information is available. Total runtime for the reliability estimation was 19:25 minutes, of which 18:40 minutes were taken up by conventional reliability calculations.

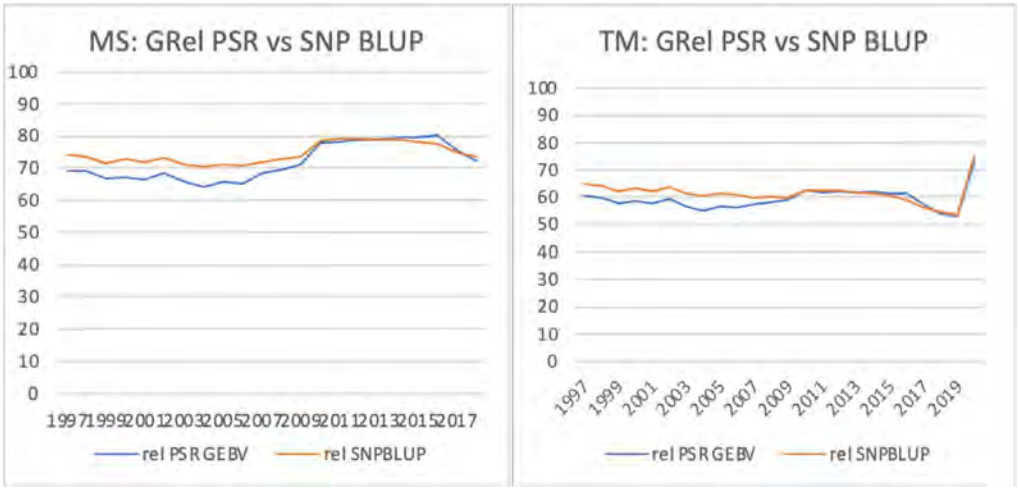


Figure 1. Genomic reliabilities for HF bulls (by birth year) with a conventional reliability >45% evaluated in a single step SNP BLUP run using the quick approximation and a genomic pseudo-record run.

Discussion

The algorithm described above is quick and adds marginally to the run time of conventional reliability calculations (in testing we found 1 to 7 minutes of extra runtime, where the total reliability estimation took 0:19 for the MS/TM evaluation shown above to 52:00 hours for our Fertility evaluation). It relies on few outside parameters: rate of imputation and direct genomic reliabilities for traits in the evaluation, obtained from genomic validation.

Direct propagation of EDC_{add} gives undesirable results in that a (ungenotyped) sire with many genotyped offspring receives added EDC >> EDC_{geno} after direct propagation, which is incorrect since EDC_{geno} is an upper limit of what a known genotype can contribute to the genomic reliability. Hence the necessity of the concept of genotype confidence, which can be interpreted as the reliability with which the genotype of an animal is known. Genotype confidences preclude the need to approximate the genomic reliability of each imputed animal explicitly, which would be costly in time and computer resources.

The use of genotype reliabilities to add to genomic reliabilities has precedent. From 2014 onwards CRV Animal Evaluation Unit estimates GEBV using a pseudo-record approach, in which extra traits are fitted, correlated to existing traits, whose observation is the DGV from the stand-alone GSEL estimation of DGV (Stoop *et al.* 2014). These pseudo-traits are fitted with a heritability of (nearly) 1.0 and a genetic correlation between existing and pseudo-trait that is the square root of the direct genomic reliability estimated from validations. This approach results in GEBV reliabilities with an increase in EDC proportional to the direct genomic reliability.

The algorithm was tested during prototyping to ensure results were as expected. Results of this approximation have not been formally validated yet. Though results show good agreement with other types of analysis, it must be realized this method is an approximation. Use of this method is recommended until efficient, more formal methods to derive genomic reliabilities (e.g. Bermann *et al.*; 2022) become operationally available.

References

Bermann M., Lourenco D., and Misztal I. (2022) J. of Anim. Sc. 100(1):1–7, https://doi.org/10.1093/jas/skab353

Liu Z., Reinhardt F., and Reents R. (2001) The effective daughter contribution concept applied to multiple trait models for approximating reliability of estimated breeding values. Interbull Bulletin 27: 41-47.

Liu, Z. *et al.* (2017) Approximating genomic reliabilities for national genomic evaluation. Interbull Bulletin 27: 75-85.

Stoop M. *et al.* (2014) Using Pseudo-observations to combine Genomic and Conventional Data in the Dutch National Evaluation, Proc. of 10th WCGALP, Vancouver, Canada

VanRaden P.M., and Wiggans G.R. (1991), J. Dairy Sci. 74(8):2737-2146, https://doi.org/10.3168/jds.s0022-0302(91)78453-1

Zaabza H. *et al.* (2021) JDS Communications 2(3):137 – 141, https://doi.org/10.3168/jdsc.2020-0058

327. Evaluating the suitability of subjectively defined base populations

J. ten Napel[1*], J. Vandenplas[1], M. Calus[1], I. Strandén[2], M. Lidauer[2] and R. Veerkamp[1]

[1]*Wageningen University & Research, P.O. Box 338, 6700 AH Wageningen, the Netherlands;*
[2]*Natural Resources Institute Finland (Luke), Jokioinen, Finland; jan.tennapel@wur.nl*

Abstract

Defining multiple base populations for a genetic evaluation is subjective and there is generally no feedback on the suitability of the defined base populations prior to model validation. We present a number of statistics that can be used to evaluate defined base populations. Application of these statistics to one simulated and two practical datasets showed that in practical datasets the number of base animals per base population was in one case very low. In many cases the available genotype information for different base populations was coming from exactly the same genotyped animals. Both issues are likely to complicate genomic evaluation. A tool to estimate these statistics is available in the MiXBLUP software suite.

Introduction

Any pedigree contains founder animals with unknown parents. There may be other animals with unknown parents because of incorrectly or not recording parents at birth, or uncertainty about the sire in case of unobserved matings or use of pooled semen. We refer to these animals as base animals. In genetic evaluations, we generally assume that these base animals are sampled from an infinite population, so we consider them to be genetically unrelated. If base animals originate from various populations or different generations within the same population, we may define several base populations, based on an expected difference in genetic level, due to origin or selection history. Criteria often used for defining base populations are breed, line within breed, sex, years of birth and selection path (Quaas and Pollak, 1981; Mrode, 2014: 55), but the result is subjective. Legarra *et al.* (2015) suggested to use genomic information to take into account that base populations are finite in reality, and that genetic relationships may exist within and across base populations. Such related base populations are referred to as metafounders (MF). In contrast to MF, base populations that are assumed to be unrelated are referred to as unknown parent groups (UPG). Defining multiple base populations is aimed at minimizing bias in genetic trend and estimated breeding values of selection candidates and reducing the mean square error of the evaluation (Foulley *et al.*, 1990). Whether the definition used is effective, depends on the true differences in genetic level and the precision with which genetic level of each base population is estimated. For MF, it also depends on the amount and quality of genomic data available to estimate genetic relationships within and between MF. In practice, the appropriateness of the defined base populations can only be assessed by extensive model validation. The aim of this paper is to present a number of base population statistics that can help to evaluate the suitability of defined base populations prior to any model validation.

Materials & methods

Description of datasets. The base population statistics are illustrated with three example datasets (Table 1). Dataset A is a simulated set of data, which was used in the study of Van Grevenhof *et al.* (2019) that recommended the use of MF. Dataset B is a subset of a routine genetic evaluation and consists of purebred pigs of a single line, their crossbred progeny and ancestors of dams of crossbred pigs., In dataset B, base populations were defined by line and one or more years of birth. Dataset B had a history of poor convergence when using UPG. Dataset C is a set of a routine evaluation of trout using UPG, defined by breeding population. Each founder animal was assigned to a single base population in dataset A and B, but to two in dataset C.

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_327, © J. ten Napel *et al.* 2022

Table 1. Descriptive statistics of example datasets.

	Dataset A	Dataset B	Dataset C
Lines + crosses, N	3 + 2	4 + 3	1 + 0
Animals in pedigree, N	51,500	421,511	385,200
Generations in pedigree, N	8	42	12
Animals genotyped, N	9,250	72,854	6,351

Base population statistics. We introduce three base population statistics, which all include the parameters c_{ij} and q_{ij}. We define a genomic base population which consists of all genotyped animals with a path of non-genotyped ancestors to a base population. Genotyped animals that are descendants of two genotyped parents are not included as they do not provide genotype information to a base population. Now c_{ij} quantifies the width of the path of non-genotyped ancestors to the base population and hence is the genomic contribution of the genomic base animal *j* to base population *i*. If c_{ij} is lower than 1, then animal *j* has genotyped ancestors. The parameter q_{ij} is the fraction relating the contribution of base population *i* to the total genetic value of the animal *j* and is an element of the Q matrix in QP transformation (Quaas and Pollak, 1981). If q_{ij} is lower than 1, then animal *j* is related to multiple base populations.

The first statistic quantifies the amount of genotype information available for each base population. Imputation of non-genotyped animals from genotyped descendants only using pedigree relationships causes loss of information, because the genotype information has to be distributed to two parents in every generation without genotypes. We therefore introduce equivalent number of base animals genotyped for a base population (Neq_i). It is calculated as:

$$Neq_i = \sum_{j=1}^{N} c_{ij} q_{ij} * \left(\frac{1}{2}\right)^{gener_j} \quad (1)$$

where *N* is the number of animals in the genomic base population and $gener_j$ is the maximum number of generations between animal *j* and base population *i*.

The second statistic quantifies uniqueness of genotype information and is the extent to which the same genotype information was used for a pair of base populations, which we call auto-similarity of one base population to the other. To illustrate the concept, imagine 24 balls, 8 red, 8 blue and 8 orange. The balls are placed in two bowls, so the first bowl contains 5 red and 4 blue balls. The second contains the remaining 3 red, 4 blue and 8 orange. The similarity of bowl one to bowl two is 3 red + 4 blue over 9 balls is 0.78. The similarity of the second bowl to the first one is 3 red + 4 blue over 15 balls is 0.47. The colours in the illustration are genotyped animals, the bowls are base populations and the number of balls of the same colour in the same bowl is *q*. So auto-similarity of base population *i* to *j* ($AS_{i\ to\ j}$) is calculated as:

$$AS_{i\ to\ j} = \frac{\sum_{k=1}^{N} c_{ik} \min(q_{ik}, q_{jk})}{\sum_{k=1}^{N} c_{ik} q_{ik}} \quad (2)$$

An $AS_{i\ to\ j}$ of 0 means that genotype information available to two base populations is independent. A value of 1 means that genotype information available is identical.

The third statistic quantifies the proximity of genotype information and is the weighted average number of generations between base animals in a base population and the genomic base population. It is calculated as:

$$\delta Gener_i = \frac{\sum_{j=1}^{N} c_{ij} q_{ij} gener_j}{\sum_{j=1}^{N} c_{ij} q_{ij}} \quad (3)$$

δGener differentiates between many remote genotyped descendants and fewer proximate ones for a given *Neq*.

Results

The number of base animals as a proportion of the total number of animals in the pedigree was similar across datasets (0.5-2.0%). The number of defined base populations varied enormously across datasets and ranged from 3 to 212 (Table 2). Also, the number of base animals per base population varied considerably between datasets and within Dataset B (Figure 1).

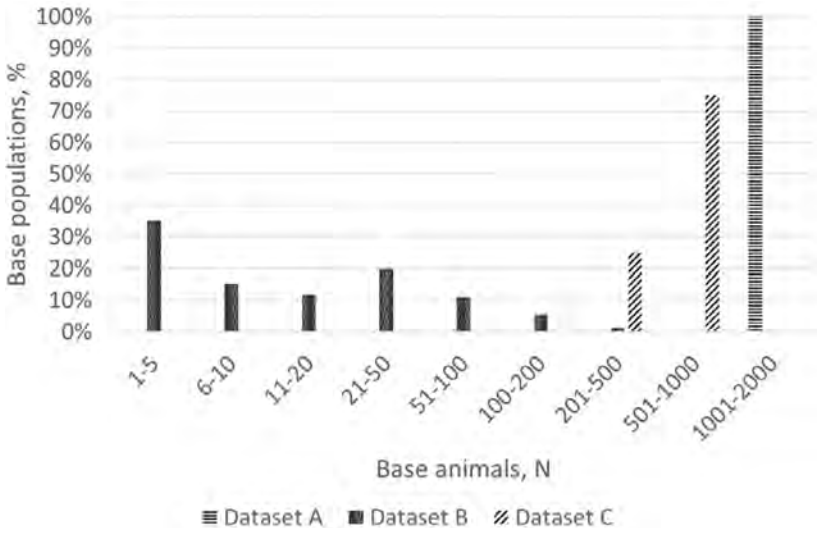


Figure 1. Histogram of number of base animals per base population.

Table 2. Statistics of base populations.

	Dataset A	Dataset B	Dataset C
Base populations, N	3	212	8
Base animals in total, N	1,030	5,702	2,028
Base populations without a link to genotypes, %	0%	47.5%	50.0%
Base populations Neq[1] <0.05	0%	43.5%	0%
Average non-zero Neq[1]	9.5	8.0	5.3
Range non-zero Neq[1]	9.4-9.6	1.0-27.0	1.0-9.5
Average AS[2] across pairs of base populations	0.06	0.07	0.14
Full AS[2] with at least one other base population, %	0.0%	43.9%	50.0%
Average size of genomic base population[3] per base population	2,247	12.2	507
Average δGener[4]	8.4	25.7	8.3

[1] Neq: amount of genotype information expressed as equivalent number of base animals genotyped (Equation 1).

[2] AS: auto-similarity of genotype information of a base population to another and is a measure of the extent to which the same genotype information was used for both base populations (Equation 2).

[3] Genomic base population: genotyped animals with a path of non-genotyped ancestors to the base population.

[4] δGener: number of generations between genomic base population and base animals (Equation 3).

In dataset A, all base populations were well-linked to genotyped animals (Table 2), whereas in datasets B and C, half the base populations had no genetic link to a genotyped animal. The average of non-zero Neq was similar across datasets, but variation was larger in datasets B and C. Of all base populations in datasets B and C, 43.9 and 50% had an AS of 1 with at least one other base population, meaning that their available genotype information was based on the same information (Table 2). In dataset A, there were no pairs of base populations with full AS. In datasets A and C, average δGener was 8.4 and 8.3, respectively. In dataset B, there were five clear clusters with a δGener of 0, 18, 24, 33 and 39 generations.

Discussion

The statistics presented quantify amount, uniqueness and proximity of genomic information for base populations and were useful for evaluating definitions of base populations to be fitted either as UPG or MF.

All studied datasets represented well-recorded populations with 98% or more of pedigree records being complete. Defining base populations based on an expected or *a priori* known difference in genetic level is only meaningful if these true differences can be estimated, so base populations should be defined to be sufficiently large.

The history of poor convergence when adding UPG to the model for dataset B is most likely caused by the large number of base populations with a small number of base animals (Figure 1). Poor connection of MF with genotyped animals, and pairs of MF with full auto-similarity (dataset B and C) will cause the gamma matrix of relationships within and between MF to be singular. If the gamma matrix (or Q'Q in QP transformation) is close to singularity, so is the coefficient matrix of the evaluation, which is known to slow down convergence when solving equations with a conjugate gradient method.

Another issue is the very deep pedigree of dataset B, in some cases, 30 generations of ancestors before the first animals with a genotype. The reasoning was not to lose any known pedigree relationships by removing generations of ancestors from the pedigree. For such data, the challenge is to reduce the number of generations in the pedigree without losing genetic relationships. A solution may be to define a set of base animals two or three generations before the first animals with data or genotypes and use pedigree records of ignored ancestors to construct relationships between these base animals, similar to MF.

Funding

This study was financially supported by the Dutch Ministry of Economic Affairs (TKI Agri and Food project 16022) and the Breed4Food partners Cobb Europe, CRV, Hendrix Genetics and Topigs Norsvin.

References

Foulley J.L., Bouix J., Goffinet B. and Elsen J.M. (1990) Connectedness in Genetic Evaluation. In: Gianola D. and Hammond K. (eds) Advances in Statistical Methods for Genetic Improvement of Livestock. Springer-Verlag, Heidelberg

Legarra A., Christensen O.F., Vitezica Z.G., Aguilar I. and Misztal I. (2015) Genetics 200(2): 455-468. https://doi.org/10.1534/genetics.115.177014/-/DC1

Mrode R.A. (2014) Linear Models for the Prediction of Animal Breeding Values, 3rd edition. CABI, Wallingford, UK

Quaas R.L. and Pollak E.J. (1981) Journal of Dairy Science 64(9): 1868-1872. https://doi.org/10.3168/jds.S0022-0302(81)82778-6

Van Grevenhof E.M., Vandenplas J. and Calus M.P.L. (2019) Journal of Animal Science 97(2): 548-558. https://doi.org/10.1093/jas/sky433

328. Maximum likelihood estimation of metafounder parameters for single and multiple breeds

O.F. Christensen[1*] and A. Legarra[2]

[1]Center for Quantitative Genetics and Genomics, Aarhus University, 8830 Tjele, Denmark,
[2]GenPhySE (Genetique, Physiologie et Systemes d'Elevage), INRAE, 31326 Castanet Tolosan, France;
olef.christensen@qgg.au.dk

Abstract

For single-step genomic best linear unbiased prediction, the use of so-called metafounders provides a coherent approach for combining pedigree and genomic relationships. Here, maximum likelihood estimation of metafounder parameters is presented. For one breed, the maximum likelihood estimate of metafounder parameter can be determined by solving a cubic equation. For multiple breeds, we provide an expression for the likelihood function, and for numerical maximisation of this function using gradient methods, we express the metafounder relationship matrix as a function of partial relationship matrices and metafounder parameters.

Introduction

Single-step genomic best linear unbiased prediction (ssGBLUP) is a methodology for genomic evaluation with partially genotyped populations, i.e. some individuals are genotyped and others are not, and phenotypes may be recorded in both subsets (Legarra *et al.*, 2014). The core idea in ssGBLUP is the construction of a relationship matrix across all individuals that combines genomic and pedigree relationship matrices. For the combined relationship matrix, the concept of metafounders (Legarra *et al.* 2015, Garcia-Baccino *et al.* 2017) provides a coherent approach to make genomic and pedigree relationships relative to the same base population and therefore compatible. With metafounders, the combined relationship matrix is

$$\mathbf{H}_{\Gamma}^{-1} = \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{G}^{-1} - (\mathbf{A}_{\Gamma})^{-1} \end{bmatrix} + (\mathbf{A}_{\Gamma}^{full})^{-1}$$

where G = MM^{T} / (k/2), with matrix M containing genotypes coded as -1, 0, 1, and k being the number of markers, $\mathbf{A}_{\Gamma}^{full}$ is the pedigree relationship matrix for all individuals with relationships among base individuals defined by metafounder parameters in matrix Γ, and A_{Γ} is the submatrix of $\mathbf{A}_{\Gamma}^{full}$ for genotyped individuals. Further details on metafounders and the computation of these matrices can be found in Legarra *et al.* (2015).

The matrix Γ contains parameters determining the relationships among base populations, within breeds (diagonal elements) and between breeds (off-diagonal element). The objective of this paper is to present methods for maximum likelihood estimation of these parameters, first for a single breed, and second for multiple breeds.

Maximum likelihood estimation of metafounder parameter for a single breed

For a single breed, the pedigree relationship matrix for genotyped individuals $\mathbf{A}_{\gamma} = \mathbf{A}(1-\gamma/2) + \gamma\mathbf{1}\mathbf{1}'$ where $\gamma \in [0;2[$ is the metafounder parameter, and A is the usual pedigree relationship matrix for genotyped individuals. Inference about g is from the observed marker genotypes M.

A model for observed marker genotypes is that $\mathbf{m}_j \sim N(2p_j - 1, 2p_j(1-p_j)\mathbf{A})$, $j = 1,\ldots,k$ are independent, where m_j are columns in M, and allele frequencies $p_j \sim N(0.5, \sigma_p^2)$, $j = 1,\ldots,k$ are independent. The metafounder parameter $\gamma = 8\sigma_p^2$. An expectation-maximisation (EM) algorithm can be derived for this

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_328, © O.F. Christensen and A. Legarra 2022

model, but the resulting algorithm involves numerical maximisation for the M-step and numerical integration for the E-step, and the algorithm as a whole becomes computationally troublesome to use in practice, and is not presented here. Computationally more feasible algorithms that approximate this EM-algorithm can be constructed, for example the one presented by Garcia-Baccino *et al.* (2017), but the statistical properties of these are not known.

Another approach, suggested by Christensen (2012), consists of studying the marginal distribution of m_j (integrating allele frequency p_j) and approximating this by a multivariate normal distribution with the correct marginal mean and marginal variance-covariance matrix, i.e. $\mathrm{m}_j \sim N(0,0.5\mathrm{A}_\gamma)$. Maximum likelihood estimation for this approach is presented here.

The joint density of $\mathrm{M} = [\mathrm{m}_1,\ldots\mathrm{m}_h]$ is

$$f(\mathbf{M}) \propto \prod_{j=1}^{k} \det(0.5\mathbf{A}_\gamma)^{-1/2} \exp(-\mathbf{m}_j'(0.5\mathbf{A}_\gamma)^{-1}\mathbf{m}_j / 2)$$

where a proportionality constant is ignored. Using that the product of exponential terms is the exponential of the sum, and $\sum_j \mathbf{m}_j'(0.5\mathbf{A}_\gamma)^{-1}\mathbf{m}_j = \mathrm{Tr}((0.5\mathbf{A}_\gamma)^{-1}\mathbf{M}\mathbf{M}')$, likelihood function is

$$L(\gamma) \propto \det(0.5\mathbf{A}_\gamma)^{-k/2} \exp(-(1/2)\mathrm{Tr}((0.5\mathbf{A}_\gamma)^{-1}\mathbf{M}\mathbf{M}'))$$

Taking logarithm of the likelihood function, introducing the notation $\mathbf{G} = \mathbf{M}\mathbf{M}' / (k/2)$, and ignoring constants, we obtain the log-likelihood function

$$\ell(\gamma) = -(k/2)\log(\det(\mathbf{A}_\gamma)) - (k/2)\mathrm{Tr}((\mathbf{A}_\gamma)^{-1}\mathbf{G})$$

Using formulas from Henderson and Searle (1981), we obtain:

$(\mathbf{A}_\gamma)^{-1} = \mathbf{A}^{-1} / (1-\gamma/2) - \mathbf{A}^{-1}\mathbf{1}\mathbf{1}'\mathbf{A}^{-1}\gamma / ((1-\gamma/2)(1-\gamma/2+\gamma\mathbf{1}'\mathbf{A}^{-1}\mathbf{1}))$ and
$\det(\mathbf{A}_\gamma) = \det(\mathbf{A})(1-\gamma/2)\ (1+\gamma\mathbf{1}\ \mathbf{A}\ \ \mathbf{1}/(1-\gamma/2)) = \det(\mathbf{A})(1-\gamma/2)\quad (1-\gamma/2+\gamma\mathbf{1}\ A\ \ \mathbf{1})$.

Inserting those into the log-likelihood function, and using short-hand notation $a = \mathbf{1}'\mathbf{A}^{-1}\mathbf{1}$, $b = \mathrm{Tr}(\mathbf{A}^{-1}\mathbf{G})$ and $c = \mathrm{Tr}(\mathbf{A}^{-1}\mathbf{1}\mathbf{1}'\mathbf{A}^{-1}\mathbf{G})$, we obtain

$$\ell(\gamma) = -\frac{k(n-1)}{2}\log(1-\gamma/2) - \frac{k}{2}\log(1-\gamma/2+\gamma a) - \frac{k}{2}\frac{b}{1-\gamma/2} + \frac{k}{2}\frac{c\gamma}{(1-\gamma/2)(1-\gamma/2+\gamma a)}$$

The maximum likelihood estimate is obtained by differentiating $\ell(\gamma)$ and setting equal to zero, where we note that the $k/2$ in front of each term in the expression can be ignored,

$$0 = \frac{(n-1)/2}{1-\gamma/2} - \frac{-1/2+a}{1-\gamma/2+\gamma a} - \frac{b/2}{(1-\gamma/2)^2}$$
$$+c\frac{(1-\gamma/2)(1-\gamma/2+\gamma a) - \gamma\left((1-\gamma/2)(-1/2+a) - (1-\gamma/2+\gamma a)/2\right)}{(1-\gamma/2)^2(1-\gamma/2+\gamma a)^2}$$

Multiplying by $(1-\gamma/2)^2(1-\gamma/2+\gamma a)^2$ on both sides of the equation followed by some algebraic manipulations we obtain a cubic equation

$$e_3\gamma^3 + e_2\gamma^2 + e_1\gamma + e_0 = 0$$

where $e_3 = -n(-1/2+a)^2/2, \quad e_2 = (n(-3/2+a)+a-b(-1/2+a)+c)(-1/2+a),$
$e_1 = (n-1)(-3/2+2a)-(-1+2a)(-3/2+a+b)$ and $e_0 = n-2a-b+2c.$

Solving a cubic equation, first requires the computation of the discriminant

$\Delta = 18e_3e_2e_1e_0 - 4e_2^3e_0 + e_2^2e_1^2 - 4e_3e_1^3 + 27e_3^2e_0^2$. If $\Delta>0$ then there are three distinct real roots, and if $\Delta<0$ then there is only one real root (and two complex roots). Following the computation of the discriminant, root(s) need to be computed, it needs to checked whether root(s) are within the interval [0;2[, and investigated which root (or possibly the boundary value 0) is the maximum.

Maximum likelihood estimation of metafounder parameters for multiple breeds

Here, we present maximum likelihood estimation for multiple breeds or genetic origins.

The columns of the genotype matrix M are assumed independent and distributed as $m_j \sim N(0,0.5A_\Gamma)$, where

$$\boldsymbol{\Gamma} = \begin{bmatrix} \gamma_{b_1} & \cdots & \gamma_{b_1,b_r} \\ \vdots & \vdots & \cdots \\ \gamma_{b_1,b_r} & \cdots & \gamma_{b_r} \end{bmatrix}$$

is a symmetric positive definite matrix with all $\gamma_b<2$, and A_Γ is the metafounder relationship matrix. Then the joint density of M = [m1,…m_p] is

$$f(\mathbf{M}) \propto \prod_{j=1}^{k} \det(0.5\mathbf{A}_\Gamma)^{-1/2} \exp(-\mathbf{m}_j'(0.5\mathbf{A}_\Gamma)^{-1}\mathbf{m}_j/2),$$

where proportionality constants have been ignored. Using that the product of exponential terms is the exponential of the sum, and that $\sum_j \mathbf{m}_j'(0.5\mathbf{A}_\Gamma)^{-1}\mathbf{m}_j = \mathrm{Tr}((0.5\mathbf{A}_\Gamma)^{-1}\mathbf{M}\mathbf{M}')$, the likelihood function is

$$L(\boldsymbol{\Gamma}) \propto \det(0.5\mathbf{A}_\Gamma)^{-k/2} \exp(-(1/2)\mathrm{Tr}((0.5\mathbf{A}_\Gamma)^{-1}\mathbf{M}\mathbf{M}'))$$

Taking logarithm of the likelihood function, noting that $\mathbf{G} = \mathbf{M}\mathbf{M}'/(k/2)$, and ignoring constants, we obtain the log-likelihood function

$$\ell(\boldsymbol{\Gamma}) = -(k/2)\log(\det(\mathbf{A}_\Gamma)) - (k/2)\mathrm{Tr}((\mathbf{A}_\Gamma)^{-1}\mathbf{G})$$

Maximum likelihood estimates of metafounder parameters can then be obtained by numerical maximization of this function.

For numerical maximization using gradient methods, derivatives of the log-likelihood function are needed. From the definition of A_Γ it can be shown using proof by induction that

$$\mathbf{A}_\Gamma = \sum_b \mathbf{A}^b(1-\gamma_b/2) + \sum_{b,b':b'>b} \mathbf{A}^{b,b'}((\gamma_{b'}+\gamma_b)/2) - \gamma_{b,b'})/4 + \sum_b \mathbf{C}^b\gamma_b + \sum_{b,b':b'>b} \mathbf{C}^{b,b'}\gamma_{b,b'}$$

With A^b being the breed b specific partial relationship matrix, $A^{b,b'}$ being the breeds b, b' segregation partial relationship matrix, matrix C^b having entries $C^b_{i,i'} = f_i^b f_{i'}^b$, and matrix $C^{b,b'}$ having entries $C^{b,b'}_{i,i'} = f_i^b f_{i'}^{b'} + f_i^{b'} f_{i'}^b$. Further details about partial relationship matrices can be found in Garcia-Cortes and Toro (2006). From this expression for matrix A_Γ derivatives with respect to metafounder parameters can be obtained.

Discussion

We have here presented maximum likelihood estimation of metafounder parameters. With genotypes on one breed, the parameter can be estimated by solving a cubic equation, whereas with genotypes on several breeds, numerical maximisation of the likelihood function is required for estimating parameters. Therefore, we have a theoretically based procedure for estimating metafounder parameters in simple cases with one or few breeds/ populations.

More challenging cases with large number of populations or genetic origins exist. In particular, for dairy cattle genetic evaluation, pedigree usually is incomplete with unknown parents of certain groups of individuals, e.g. due to imported bulls or incomplete pedigrees, and can be handled using so-called unknown parent groups. For metafounder approach this provides a challenge because number of unknown parent groups is large, making the number of parameters in Γ matrix prohibitively large, and either a reduction in number of unknown parent groups, as in Kudinov *et al.* (2020), or a parameterisation of Γ using fewer parameters, as in Koivula *et al.* (2021), needs to done.

References

Christensen O.F. (2012) Genet. Sel. Evol. 44:37. http://doi.org/10.1186/1297-9686-44-37

Garcia-Baccino C.A., Legarra A., Christensen O.F., Misztal I., Pocrnir I. *et al.* (2017) Genet. Sel. Evol. 49:34. http://doi.org/10.1186/s12711-017-0309-2

García-Cortés L.A. and Toro M.A. (2006) Genet. Sel. Evol. 38:601-615. http://doi.org/10.1051/gse:2006024

Henderson and Searle (1981) SIAM Review. 23(1):53-60. https://doi.org/10.1137/1023004

Koivula M., Stranden I., Aamand G.P. and Mantysaari E.A. (2021) Interbull Bulletin 56, Leeuwarden, the Netherlands, April 26 – 30, 2021

Kudinov A.A., Mantysaari E.A., Aamand G.P., Uimari P., and Stranden. I. (2020) J. Dairy Sci. 103(7):6299:6310. https://doi.org/10.3168/jds.2019-17483

Legarra A., Christensen O.F., Aguilar I., and Misztal I. (2014) Livest. Sci. 166:54-65. http://dx.doi.org/10.1016/j.livsci.2014.04.029

Legarra A., Christensen O.F., Vitezica Z., Aguilar I. and Misztal I. (2015) Genetics 200(2):455-468. http://doi.org/10.1534/genetics.115.177014

329. Exploring the statistical nature of independent chromosome segments

M.K. Hollifield*, M. Bermann, D. Lourenco and I. Misztal

University of Georgia, Department of Animal and Dairy Science, Athens, GA 30602, USA; marykate.hollifield@uga.edu

Abstract

Previous research in genomics developed the concept of independent chromosome segments and suggested that genomes are a composite of chromosome segments derived from preceding generations. The number of independent chromosome segments (M_e) was posited as $4N_eL$, a function of effective population size (N_e) and genome length in Morgans (L). The objective of this study was to investigate the statistical nature of independent chromosome segments. In a simulated population, N_a animals were selected randomly as founder animals. For all animals, the genome was split into N_s equal-sized segments. For each specific location, segments of non-founders were assigned the most similar segment of one founder. Genomic analyzes estimated the value of the segments, and breeding values were a sum of all segments for a specific animal. Accuracies were maximized at N_a=N_e, N_s=4L, but they were not as high as in GBLUP, suggesting that the current choice for segments is an approximation.

Introduction

The genome is inherited in large segments from each parent and is a mosaic of ancestral segments that are broken in each generation (MacLeod, 2005; Pääbo, 2003). Genomic selection works by estimating independent chromosome segment effects (Goddard, 2009). Hypothetically, for populations with small effective population size (N_e), the estimated number of independent chromosome segments (M_e) is $4N_eL$, where L is genome length in Morgans (Stam, 1980). Experimentally, M_e has been related to the eigenvalues of the genomic relationship matrix (G), as the dimensionality of genomic information is approximately $4N_eL$ (Pocrnic *et al.*, 2016). Predictions are accurate using computing algorithms that assume a limited dimensionality of G (Misztal *et al.*, 2014a).

Stam (1980) defined the number of chromosome segments as a random variable, i.e. of statistical nature, and showed that segments are of different sizes but with a mean of $1/4N_e$. Ignoring the physical properties of the segments and exploring their statistical nature, the genome could be possibly divided into equal-sized segments, and each unique segment could be assigned a value. If the additive value of each segment is known, can genomic prediction be based solely on those segments? Conceptually, chromosome segments effects are additive; thus, an animal's breeding value is composed of a sum of these effects (Misztal *et al.*, 2020). In this study, we assume that the segments are distributed in N_e animals for a population, and the number of segments is equal to the genome size divided by 4L. Then, we estimated the effects of individual chromosome segments to assess if all additive genetic information can be composed in $4N_eL$ independent chromosome segments of equal size.

Materials & methods

Data. Data were simulated using QMSim (Sargolzaei and Schenkel, 2009). The population contained 20,000 animals and ten non-overlapping generations. Each generation included 2,000 animals with N_e equal to 20. One polygenic trait was simulated with a heritability of 0.3. The last three generations were genotyped, and each genome contained ten 1 M long chromosomes, for a total genome length of 10 M and 50,000 SNP.

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_329, © M.K. Hollifield *et al.* 2022

Models and computations. A group of animals was selected to generate conceptual independent chromosome segments. Hereinafter, these animals will be referred to as 'founder' animals. The founder animals were selected from the genotyped animals in the last three generations. Each animal's genome composed of 50,000 SNP was split into discrete segments. To test the effects of the theoretical $4N_eL$ independent chromosome segments, N_e (20) founder animals and 4L (40) segments per animal were analyzed. This study conducted two main analyses: (1) founder animals were selected randomly or by the least related animals in the population; and (2) various numbers of segments and founder animals were compared.

The chromosome segments of each animal were denoted as p_{ij}, where p is the j^{th} segment of the i^{th} animal. To analyze the behavior around $4N_eL$ segment effects, the number of segments per animal (N_s) varied while the number of founder animals (N_a) remained fixed to N_e (20). Similarly, N_s was fixed to 4L (40) as N_a varied. Each segment of each non-founder animal was assigned to a founder animal segment that was the most similar. Only the effects of the founder animal's chromosome segments were tested to evaluate independent chromosome segment effects. Altogether, there were N_aN_s segment effects total, and the ij^{th} segment effect was numbered as $(i-1)N_a + j$. The similarities per non-founder segment were calculated by $min\ (\Sigma \mid p_{kj} - p_{lj} \mid)$, where k is the founder animal with the minimum sum of the SNP codes in segment j for non-founder animal l. The effects of the hypothetical independent chromosome segments were estimated using:

$$y = 1'\mu + Z_s s + e \quad (1)$$

where y is a vector of phenotypes, Z_s is an incidence matrix relating the assigned founder animal segments to all animals, s is a random vector of estimated segment effects, and e is a random vector of errors.

Accuracies were computed to investigate the theoretical number of $4N_eL$ independent chromosome segments containing all genetic information in a population. Accuracies were calculated for animals in the last generation by *cor* (TBV, $Z_s\hat{s}$), where $Z_s\ \hat{s}$ is a vector of breeding values based on segments and TBV is a vector of the true breeding values outputted by QMSim. Accuracies of segment effects were compared with the true accuracy (i.e. *cor* (TBV,GEBV)). GEBV were computed using GBLUP in the BLUPF90 software suite (Misztal *et al.*, 2014b), whereas $\hat{s}$ and $Z_s\hat{s}$ were computed using in-house software written in Fortran.

Results

Accuracies of estimated chromosome segment effects with random and the least related animals as founder animals are shown in Figures 1 and 2. In Figure 1, N_a is fixed to N_e (20), and in Figure 2, N_s is fixed to 4L (40). It was hypothesized that accuracy is maximized with $4N_eL$ chromosome segments, which aligns with the theory that there is no new genetic information if $4N_eL$ independent chromosome segments are explained (Pocrnic *et al.*, 2019).

The purple lines in Figure 1 and 2 indicate where $4N_eL$ chromosome segment effects are estimated and are close to where accuracy is maximized and begins to plateau or decrease. The maximum accuracy of segment effects was 0.69 for both random and least related founder animals, and the accuracy of GBLUP was 0.90. The results suggest that 4L segments of N_e founder animals contain all additive information in the population. However, the accuracy of GBLUP is greater than that of the chromosome segment effects, suggesting that the chromosome segments do not account for all genetic variation as they were calculated. It was presumed that if the least related animals in the population were used as the founder animals, more ancestral segments would be accounted for than randomly chosen founder animals, and accuracy would be greater. Similar to the initial proposition of APY, the core was defined as the 'proven' animals with their own phenotypes and progeny phenotypes, hence the animals with the most information and greatest genetic

impact on the population (Misztal *et al.*, 2014). Core animals have since been redefined, and random animals are appropriate to use as core and produce the same accuracy as the previous definition (Bradford *et al.*, 2017).

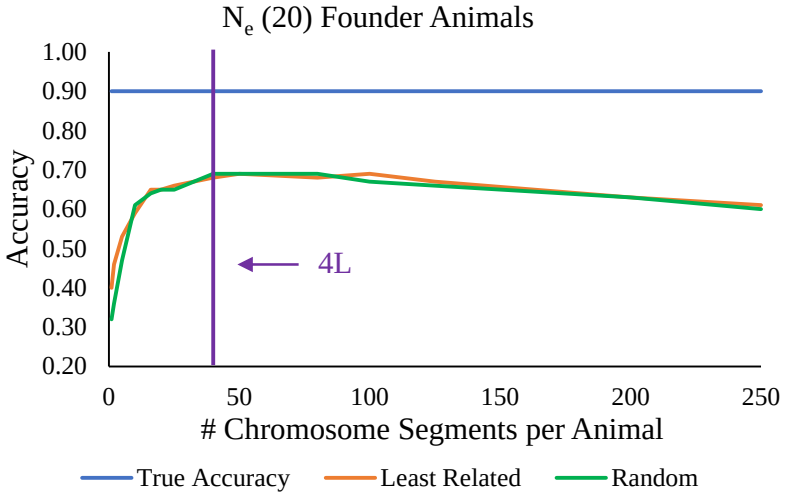


Figure 1. Accuracies of chromosome segment effects with random and least related founder animals. The number of founder animals is fixed to N_e (20). Accuracies of chromosome segment effects were calculated by correlating the TBV and the estimated segment effect. The true accuracy was calculated by correlating the TBV and GEBV. The purple line represents $4N_eL$ independent chromosome segments.

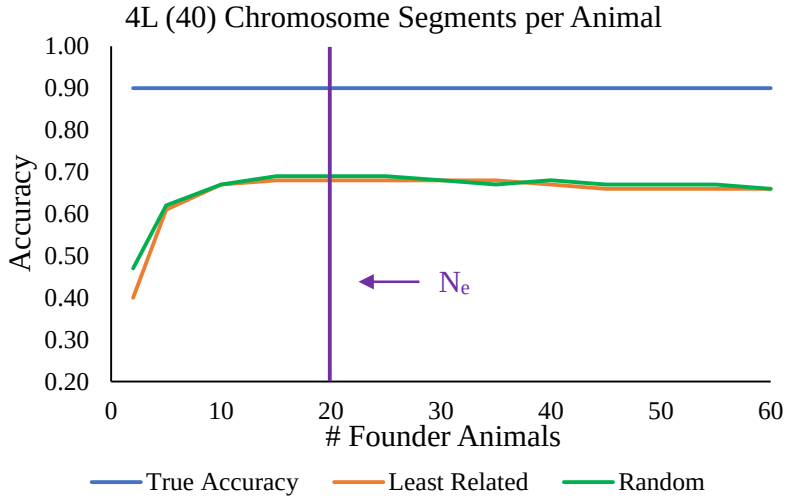


Figure 2. The number of chromosome segments per animal is fixed to 4L (40). The methods and components are the same as in Figure 1.

As aforementioned, DNA is inherited from ancestors as contiguous blocks formed by recombination throughout the evolution of the population (Pääbo, 2003). Biologically, In violation of present concepts of physical chromosome segments, this study used constant segment size, and recombination was not considered. A follow-up study will assume haplotypes following the physical properties of segments. The results presented indicate some agreement with expectation, but it is not complete, perhaps haplotypes better fit the definition.

Discussion

Approximately $4N_eL$ chromosome segments contain almost all the genetic information that can be explained in a population. It is important to notice that M_e is a statistical concept, and the way those M_e segments are physically organized in haplotypes depends on linkage disequilibrium. Further investigation is needed on the definition and characteristics of physical chromosome segments. If each segment has an additive value and can be ranked, genomic predictions can potentially be based on segments alone. By using the concepts of limited dimensionality and independent chromosome segments, overlapping information in a population can be removed, reducing computing time and data storage space for genomic evaluations.

References

Bradford, H., Pocrnić I., Fragomeni B., Lourenco D., and Misztal I. (2017) J. Anim. Breed. Genet.134(6):545-552. https://doi.org/10.1111/jbg.12276

Goddard, M. (2009) Genetica 136(2):245-257. https://doi.org/10.1007/s10709-008-9308-0

MacLeod, A. K. (2005) Detecting ancestral junctions in inbred populations. University of Edinburgh. Edinburgh, UK.

Misztal, I., Legarra A., and Aguilar I. (2014a) J. Dairy Sci. 97(6):3943-3952. https://doi.org/10.3168/jds.2013-7752

Misztal, I. Tsuruta, S., Lourenco, D., Masuda, Y., Aguilar, I., *et al.* (2014b) Manual for BLUPF90 family of programs. Available at: http://nce.ads.uga.edu/wiki/lib/exe/fetch.php?media=blupf90_all7.pdf

Misztal, I., Lourenco, D., and Legarra, A. (2020) J. Anim. Sci. 98(4):1-14. https://doi.org/10.1093/jas/skaa101

Pääbo, S. (2003) Nature 421(6921):409-412. https://doi.org/10.1038/nature01400

Pocrnic, I., Lourenco, D., Masuda, Y., Legarra, A., and Misztal, I. (2016) Genetics 203(1):573-581. https://doi.org/10.1534/genetics.116.187013

Pocrnic, I., Lourenco, D., Masuda, Y., and Misztal, I. (2019) Genet. Sel. Evol. 51(1):75. https://doi.org/10.1186/s12711-019-0516-0

Sargolzaei, M., and Schenkel, F. S. (2009) Bioinformatics 25(5):680-681. https://doi.org/10.1093/bioinformatics/btp045

Stam, P. (1980) Genet. Res. 35(2):131-155. https://doi.org/10.1017/S0016672300014002

330. Including local genomic breed proportions in genomic predictions for crossbred

J.H. Eiríksson*, G. Su and O.F. Christensen

Center for Quantitative Genetics and Genomics, Aarhus University, Blichers Allé 20, 8830 Tjele, Denmark; jonh@qgg.au.dk

Abstract

Genomic evaluations for crossbred animals introduce challenges not encountered for purebred animals. We propose to include effects of local breed proportions (LBP) based on detected breed of origin of alleles in a genomic model for crossbred dairy cows. We tested different segment length for modelling the LBP: 1, 20, 100 and 500 markers, for genomic prediction based on data of first lactation protein yields of 5,176 Danish crossbred dairy cows with contribution from three dairy breeds. The models that included LBP had slightly higher cross-validation predictive ability, 0.390 for all segment lengths compared with 0.385 for a model without LBP effects. The LBP models also had less dispersion bias. Inclusion of LBP in genomic predictions can therefore improve predictions even when considering the local effects in segments of hundreds of markers.

Introduction

Genomic evaluations for crossbred introduce challenges not encountered in genomic evaluation within the pure breeds. Among those challenges is the lower level of linkage disequilibrium between markers and quantitative trait loci across breeds than within breed levels (Ibánēz-Escriche *et al.*, 2009), genetic background (Falconer, 1989), and breed segregation, i.e. the additional genetic variation in offspring of crossbred parents (Christensen *et al.*, 2015).

Models that split the breeding values of crossbreds into breed specific terms and account for breed origin of alleles (BOA) can in principle account for the different linkage between breeds (Ibánēz-Escriche *et al.*, 2009). However, these models do not always result in more accurate predictions than models assuming common marker effects (Sevillano *et al.,* 2017). Applications of BOA models (e.g. Sevillano *et al.* 2017) have generally not accounted for the breed segregation term which Christensen *et al.* (2015) presented in their derivation of the models for three-way crosses.

For BOA models with training on purebred data, intercepts for correction of breed levels are necessary for the prediction of breeding values of crossbreds (Eiríksson *et al.*, 2021) and might not be readily available from the genomic evaluations of purebred. Phenotypes of crossbreds along with BOA information could give more precise estimates of the breed level intercepts than purebred information alone. Additionally, BOA could give information on the local breed proportions (LBP), i.e. the contribution of the breeds to different regions of the genome, which can vary in offspring of crossbred animals and similarly to the breed segregation term. The LBP can be considered at individual marker level, or for a number of consecutive markers because of the high correlation between BOA of proximate markers.

Therefore, the objectives of this study were to compare predictive ability of genomic models including global genomic breed proportions (GBP) with a model that additionally includes local breed proportions (LBP) for genomic prediction for crossbred cows. We considered different segment lengths for the LBP model (LBPM).

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_330, © J.H. Eiríksson *et al.* 2022

Materials & methods

Data. Single nucleotide polymorphism (SNP) genotypes from the EUROG MD custom chip and phenotypes of 305-day first lactation protein yields of 5,176 Danish crossbred dairy cows from 74 herds were used in this study. These cows were crosses of Holstein (H), Jersey (J) and Nordic Red Dairy Cattle (R) in various combinations. We imputed the genotypes to a set of 50,684 markers, to fill in missing genotypes, and phased them to two haplotypes using FImpute (Sargolzaei *et al.*, 2014), for the genotyped crossbred and 2,500 genotyped animals from each of the pure breeds. We detected BOA using the AllOr method (Eiríksson *et al.*, 2021) with a purebred reference genotypes of 2,500 cows from each of the breeds, a segment length of 100 SNPs and shifting five SNPs between rounds in the method.

For investigating crossbreed-specific genetic effects, we constructed a corrected phenotype, i.e. corrected for both fixed effects and genomic estimated breeding values computed from purebred information. We calculated the corrected phenotype as:

$$y^* = y - \hat{a}_{BOM},$$

where y is a vector of phenotypes of first lactation 305-day protein yields that were corrected for herd-year and calving age effects, and $\hat{a}_{BOM}$ is a vector of genomic estimated breeding values. The $\hat{a}_{BOM}$ were calculated with a breed of origin model using solutions from each of the three purebred genomic evaluations, as described in Eiríksson *et al.* (2021), but without the intercepts for breed level corrections. The estimated effects from purebred evaluations, i.e. estimated marker effects and residual polygenic effects, were from the February 2021 run of the genomic evaluations of the three pure breeds from the Nordic Cattle Genetic Evaluation.

Models. We compared models with and without including local breed proportion effects.

The GBP model (GBPM) included regressions on global breed proportions and genomic breed heterozygosity rates:

$$y^* = 1u + f_H c_H + f_J c_J + k_{H,J} h_{H,J} + k_{H,R} h_{H,R} + k_{J,R} h_{J,R} + e \quad (1)$$

where u is an intercept, 1 is a vector of ones, f_b contains global proportions of alleles assigned to breed b (b=H, J, R), c_b is a regression coefficient on breed proportions, $k_{b,b'}$ contains genomic breed heterozygosity rate, i.e. proportion of loci with alleles assigned to two different breeds, b and b', $h_{b,b'}$ is the regression coefficients on genomic breed heterozygosity, and e is the random residual $e \sim N(0, I\sigma^2_e)$. The values in f_b were calculated as $f_{i,b} = \Sigma_j (x_{b,i,1,j} + x_{b,i,2,j})/(2m)$, where $x_{b,i,g,j}$ is 1 for marker j with allele on haplotype g assigned to breed b, 0 if the allele is assigned to another breed, and a value between 0 and 1 if the allele was not assigned to definite breed and m is the number of markers. Note that since $f_H + f_J + f_R = 1$, one of the global breed proportion effects can be left out of the model, given that an intercept is included. The values in $k_{H,J}$ were calculated as $k_{i,b,b'} = \Sigma_j (x_{b,i,1,j} \circ x_{b',i,2,j} + x_{b',i,1,j} \circ x_{b,i,2,j})/m$ where $\circ$ denotes element wise multiplication.

The LBPM additionally included random LBP effects:

$$y^* = u + f_H c_H + f_J c_J + k_{H,J} h_{H,J} + k_{H,R} h_{H,R} + k_{J,R} h_{J,R} + Z^*_H p_H + Z^*_J p_J + Z^*_R p_R + e \quad (2)$$

where $Z^*_b = Z_b - f_b 1'$, with Z_b containing the local breed proportions, i.e. the proportion of alleles on a segment that were assigned to breed b. Furthermore, p_b is a vector with random effects of local proportion of breed b, with $p_b \sim N(0, I\sigma^2_{p,b})$, where $\sigma^2_{p,b}$ is the variance for segment LBP effect for breed b (similar to SNP variance for SNP-BLUP).

An equivalent model of the LBPM is to construct relationship matrices for the local effects, which is computationally more efficient for short segments. The LBPM with relationship matrix is then:

$$y^* = u + f_H c_H + f_J c_J + k_{H,J} h_{H,J} + k_{H,R} h_{H,R} + k_{J,R} h_{J,R} + v_H + v_J + v_R + e \quad (3)$$

where v_b is a vector of local breed proportion animal effects with, $v_b \sim N(0, Q_b \sigma^2_{v,b})$ where $Q_b = Z^*_b Z^{*\prime}_b / (\mathrm{tr}(Z^*_b Z^{*\prime}_b)/n)$ where tr stands for trace and n is number of animals.

Implementation. We estimated the variance components for LBPM with segment length of 500, 100, 20, and 1 SNPs. For the 1 SNP estimation we constructed a relationship matrix and added a low value, 0.0001, to the diagonal of Q_b to make it invertible. We used the AI-REML procedure from the DMU package (Madsen and Jensen, 2013) for variance component estimation.

We tested the predictive ability of the LBPMs and GBPM using five-fold cross validation, where we randomly split the herds into five groups, and made the prediction for animals in each group based on records from the other four groups. For all models, the predictions $\hat{y}$ included all the effects in the model plus $\hat{a}_{BOM}$. We compared the predictions with y, i.e. the phenotypes corrected for fixed herd-year and calving age effects.

Results

The genomic breed proportions according to the BOA assignment were 49% H, 15% J, and 36% RDC. According to the results from GBPM, a 10% increase in genomic breed heterozygosity increased protein yield by 1.6 kg (SE=1.0), 1.7 kg (SE=0.4), and 2.5 kg (SE=1.1), for H/J, H/R and J/R respectively. For the LBPMs, the local breed proportion effects together explained around 2% of the variance in y^*, with the highest proportion for the shortest segments (Table 1). The variance related to H proportions was the largest, ranged from 20.8 to 25.8, compared with the variances for J and R proportions, which ranged from 5.0 to 5.7 and 3.1 to 4.8, respectively (Table 1). The predictive ability of the models from the cross-validation was equal for all LBPM, and 0.005 higher than the correlation from GBPM (Table 1). The LBPM had less dispersion bias than GBPM, irrespective of segment length.

Discussion

The $\sigma^2_{v,H}$, $\sigma^2_{v,J}$, and $\sigma^2_{v,R}$ estimates are dependent on the breed composition in the datasets. Local breed proportions are for example constant throughout the genome of F1 animals, which were around 30% of our data, and they do therefore not contribute to this variance. Alternative standardization of the LPM relationship matrices, e.g. based on setting the expected diagonal values for a specific type of cross to be 1, could give estimated variances that do not suffer from this. However, in our study that would make

Table 1. Variance components for breed proportions as well as prediction accuracy and dispersion bias for first lactation protein yield.

Model[3]	Segm. length	Variance				Cross-validation	
		$\sigma^2_{v,H}$	$\sigma^2_{v,J}$	$\sigma^2_{v,R}$	σ^2_e	Cor($\hat{y}$,y)	b_1[2]
GBPM						0.385	0.880
LBPM	500	20.8[1]	5.0[1]	3.1[1]	1,578.7	0.390	0.901
LBPM	100	24.8[1]	5.6[1]	4.8[1]	1,572.8	0.390	0.901
LBPM	20	25.7[1]	5.7[1]	4.7[1]	1,571.9	0.390	0.901
LBPM	1	25.8	5.7	4.8	1,571.6	0.390	0.901

[1] Calculated as $\sigma^2_{v,b} = \sigma^2_{p,b} * \mathrm{tr}(Z^*_b{}' Z^*_b)/m$.

[2] Dispersion bias, estimated as the slope of the regression of y on $\hat{y}$.

[3] GBPM = global breed proportion model, LBPM = local breed proportion model.

comparison across segment lengths skewed. The estimated variances suggest that there is variation related to LBP in the data and the magnitude is about 1-2% of the phenotypic variance and around 5% of the additive genetic variance.

The improvement in predictive ability from LBPM compared to GBPM was marginal in our study (Table 1). However, improvement in prediction was not expected for all the animals in the data, LBP could be more important for groups with few or no F1 animals. Additionally, the training set for estimating LBP effects was not large, only around 4,000 animals for estimating LBP effects for the three breeds. The relatively small difference in variance explained and predictive ability between alternative segment lengths suggests that local breed proportion effects can be included for longer segments, e.g. 100 SNPs, and therefore reduce the number of parameters without compromising accuracy.

The Q_b matrices in Equation 3 have similar structure as the genomic breed-segregation partial relationship matrix presented by Christensen *et al.* (2015). Both have off-diagonal indicating if animals share origin in fewer or more loci than that is expected based on breed proportions. Christensen *et al.* (2015) only considered three-way crosses with the same terminal sire breed and the diagonal of the breed-segregation partial relationship matrix is therefore constant, equal to 0.5. In our case, with crossbred animals of varying breed compositions, the diagonal elements were highly variable, but with the average equal to one because we standardized the matrix with the trace.

Detected BOA has to our knowledge not before been applied to model heterozygosity in crossbreds. If BOA information is available, modelling heterozygosity in this way is simple, and can be simpler than having to rely on pedigree information.

The LBPM presented here is an add-on to the BOA model with training in purebred population (Eiríksson *et al.* 2021). A potential complication with the application of the BOA model in that paper is that the intercepts for correction for breed level differences may not be available from routine genomic evaluation. The global breed proportion effects could be used for accounting for breed levels in that model. Other authors have presented BOA models with training in both purebreds and crossbreds (Karaman *et al.*, 2021) without accounting for breed segregation. Whether the inclusion of local breed proportion effects in those models improves prediction is an interesting area for further development.

Inclusion of LBP effects can moderately increase the predictive ability and reduce bias even when the local effects are considered in segments of hundreds of markers. They should therefore be considered in future development of models for genomic prediction for crossbred.

References

Christensen O.F., Legarra A., Lund M.S., and Su G. (2015) Genet. Sel. Evol. 47:98. https://doi.org/10.1186/s12711-015-0177-6

Eiríksson J.H., Karaman E., Su G. and Christensen O.F. (2021) Genet. Sel. Evol. 53:84. https://doi.org/10.1186/s12711-021-00678-3

Falconer D.S. (1989) Introduction to quantitative genetics. Longman Scientific & Technical, Harlow, United Kingdom.

Ibánēz-Escriche N., Fernando R.L., Toosi A., and Dekkers J.C. (2009) Genet. Sel. Evol. 41:12. https://doi.org/10.1186/1297-9686-41-12

Karaman E., Su G., Croue I., and Lund M.S. (2021) Genet. Sel. Evol. 53:46. https://doi.org/10.1186/s12711-021-00637-y

Madsen P., and Jensen J. (2013) DMU, a package of analysing multivariate mixed models. Ver. 6, rel 5.2. Available at: https://dmu.ghpc.au.dk/dmu/DMU/Doc/Current/dmuv6_guide.5.2.pdf

Sargolzaei M., Chesnais J.P., and Schenkel F.S. (2014) BMC Gen. 15:478. https://doi.org/10.1186/1471-2164-15-478

Sevillano C.A., Vandenplas J., Bastiaansen J.W.M., Bergsma R., and Calus M.P.L. (2017) Genet. Sel. Evol. 49:75. https://doi.org/10.1186/s12711-016-0234-9

331. Investigating the impact of APY core size and definition in single-step GBLUP evaluations

R. Abdollahi-Arpanahi*, D. Lourenco and I. Misztal

Department of Animal and Dairy Science, University of Georgia, Athens, GA 30602, USA;
rostam.abdollahi@uga.edu

Abstract

Algorithm of proven and young (APY) uses a subset of genotyped individuals as a core and remaining genotyped individuals as noncore for creating a sparse approximation of the inverse of a large genomic relationship matrix. This study aimed to investigate eight core definitions based on relatedness and distribution across generations, with six different core sizes, to explore the prediction accuracy of APY. The dataset consisted of 357K Duroc pigs, with 111K genotypes, and phenotypes for three production traits. When the core size was equal to the number of largest eigenvalues explaining 50% of the variation of G (eigen50; n=160), the core based on most popular animals or evenly sampled across the pedigree delivered the highest prediction accuracy. As the core size increased to the number of eigenvalues explaining 99% of the variation in G (eigen99; n=7,320), the prediction accuracy of all core types was nearly identical to the regular ssGBLUP.

Introduction

Inverting the genomic relationship matrix for larger than 150K genotyped individuals is a barrier for application of Single-step genomic BLUP (ssGBLUP) (Fragomeni *et al.*, 2015). Misztal *et al.* (2014) proposed the 'algorithm for proven and young animals' (APY) based on genomic recursions to create a sparse approximation of the inverse of G to enable large-scale single-step genomic BLUP (ssGBLUP) evaluations. In APY, genotyped animals are partitioned into two groups of core and noncore. The GEBV of noncore animals is conditioned on the GEBV of core animals. The computing cost of APY G^{-1} is cubic with the number of core animals and linear with the number of noncore. The core size depends on the dimensionality of G, which has been addressed by the eigenvalue decomposition of G (Pocrnic *et al.*, 2016). However, the choice of core animals is still a question that requires further research. Therefore, this study aimed to assess the performance of eight core types in six different core sizes, based on the eigenvalue decomposition of G, on the prediction accuracy of ssGBLUP using 111k genotyped Duroc pigs.

Materials & methods

Data structure. We used a Duroc pig dataset consisted of 357,220 pedigreed animals, of which 111,583 were genotyped for 32,824 SNP markers after the quality control. Data were collected from 2012 to 2021. Studied traits included average daily gain at nursery (ADGn, n=221,955), average daily gain at finishing (ADGf, n=195,946), and backfat (BF, n=195,967). The descriptive statistics of studied traits can be found in Table 1.

Model and analysis. The statistical model for genetic evaluations was as follow:

$$y_t = Xb_t + W_1l_t + W_2c_t + W_3a_t + e_t \quad (1)$$

where y_t is the vector of observations for trait *t*; *t* refers to ADGn, ADGf, and BF; b_t is the vector of fixed effects; $\mathbf{l}_t \sim MVN(0, \mathbf{I}\sigma_l^2)$, $\mathbf{c}_t \sim MVN(0, \mathbf{I}\sigma_c^2)$ and $\mathbf{a}_t \sim MVN(0, \mathbf{H}\sigma_a^2)$ are the vectors of random effects for litter, pen, and direct additive genetic effects, respectively; e_t is the vector of residuals. The X, W_1, W_1, and W_1 are design matrices for the effects in l_t, c_t, and a_t, respectively. Here σ_l^2, σ_c^2, and σ_a^2 are variances of litter

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_331, © R. Abdollahi-Arpanahi *et al.* 2022

size, pen effects, and additive genetic affects; H is the (co)variance matrix between animals constructed by integrating pedigree and genomic relationship matrix (Aguilar *et al.*, 2010). We used ssGBLUP with APY to compute GEBV. For more detailed description of APY see Misztal (2016). The ssGBLUP was run in a multiple-trait animal model framework.

Direct additive genetic and litter effects were fit for three traits, pen effect was fit for ADGf and BF traits. The fixed effects were gender for three traits, heard-year-week for ADGn, barn for ADGn and BF. Age of weighting was considered as a covariate for ADGn and ADGf.

Core sizes. At first, the G matrix was constructed, then singular value decomposition has been applied. Then, six core sizes equal to 160, 700, 1,363, 2,344, 7,320, 10,000 individuals were chosen according to the number of largest eigenvalues explaining 50, 80, 90, 95, 99 and >99% of the variation in G.

Core definitions. Eight core types were defined including: (1) most popular animals (MPA): Animals with higher than specific number of progeny (>400, >125, >35, >30, >20), both parents known and own performance record; (2) least popular males (LPM): Males with both parents known, without progeny, and without own performance record; (3) least popular females (LPF): Females with both parents known, with neither progeny nor own performance record; (4) random set (Rnd): Core animals were randomly sampled from all genotyped animals; (5) evenly distributed across pedigree (Ped): Animals were ordered in genealogical order, then core animals were evenly sampled across pedigree; (6) within family selection (Fam): from each full-sib family only one progeny was randomly assigned to core group; (7) unrelated animals (Unrel): animals with genomic relationship less than a threshold (<0.10, <0.15, <0.20, <0.22, <0.27) were chosen; (8) QR decomposition: The genotype matrix (Z) was decomposed into a product Z= QR according to QR factorization (Golub and Van Loan, 1996).

The direct inverse of G using all genotyped animals in ssGBLUP was used as the baseline (Regular ssGBLUP).

Prediction accuracy. The prediction accuracy of each scenario was estimated for 17,594 young, genotyped pigs born in 2020 as follows:

$\widehat{acc}_{PA} = \frac{corr(\mathbf{y_{adj}}, GEBV)}{\sqrt{h^2}}$, where y_{adj} is the vector of adjusted phenotypes for all non-genetic effects.

Results & discussion

Core definitions. The prediction accuracy using ssGBLUP with direct inversion of G (i.e. the baseline) was 0.85, 0.80, and 0.77 for ADGn, ADGf and BF, respectively. When the number of animals in the core was equal to eigen99, all core types delivered similar prediction accuracy and slightly less than the baseline (Figure 1). When the core size was less than the number of largest eigenvalues explaining 90% of the variation of G, the MPA delivered the highest prediction accuracy, and the QR and Unrel attained the lowest compared to other core definitions (Figure 1). The results showed that by increasing the core size, the difference among various core definitions faded. They were nearly identical as the number of animals in the core increased to a value equal to the number of largest eigenvalues explaining 99% of the variation in G.

The correlation between GEBV obtained from the regular ssGBLUP and APY ssGBLUP with the core size corresponding to eigen99 was greater than 0.99. In Rnd, Ped, Fam, Unrel core definitions, genotyped animals are distributed well across the generations. Misztal (2016) pointed out that the prediction performance of APY depends on how well core animals represent the independent chromosome segments

(ICS). It has been indicated that core definitions representing all generations and maximizing the number of genotyped offspring would deliver the highest predictive ability (Ostersen *et al.*, 2016; Vandenplas *et al.*, 2018). Including animals from all generations in the core set was also proposed by Bradford *et al.* (2017).

QR core showed the lowest GEBV correlation with regular ssGBLUP (0.992) at eigen99, and it needed the largest number of iterations to reach convergence. Opposite to Rnd and Ped core definitions, the QR core included old animals born from 2012 to 2016. Hence, unequal distribution of core animals across years/generations would result in slightly less accurate GEBV (Vandenplas *et al.*, 2018). It is important to include animals from all generations in the core set.

Regardless of trait and core definition, the accuracy of APY increased 33% from eigen50 to eigen80, 12% from eigen80 to eigen90, and 8% from eigen90 to eigen99. The improvement in the prediction accuracy by increasing the core size from eigen99 to 10,000 animals was negligible (< 1%). When the core size was determined based on eigen99, the difference between prediction accuracy using APY and regular ssGBLUP for ADGn, ADGf, and BF was 0.007, 0.008, and 0.007, respectively.

Core size. We found that when the core size was less than the number of largest eigenvalues explaining 95% of the variation in G, the choice of core animals was critical to accurately predict GEBV using APY ssGBLUP. However, as the core size increased to the number of largest eigenvalues explaining 99% or greater than 99% of the variation in G, all core types converged to similar prediction accuracy. Pocrnic *et al.* (2016) showed that the core size in APY ssGBLUP required to predict GEBV as accurate as the regular ssGBLUP with direct G inverse is related to the dimensionality of G. The effective dimensionality of G is dependent on the minimum number of ICS in the population, number of SNP markers, and number of genotyped animals.

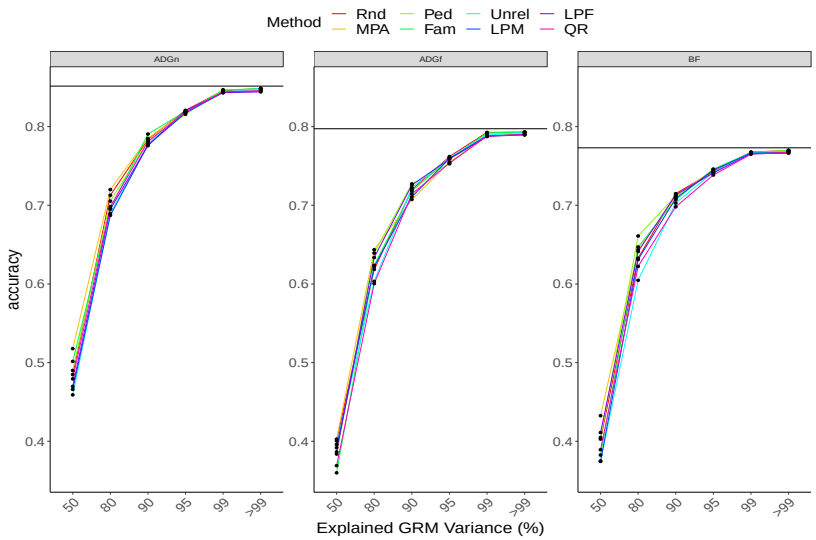


Figure 1. Prediction accuracy of genomic estimated breeding values (GEBV) from different core definitions and sizes in APY ssGBLUP for Average daily gain at nursery (ADGn), average daily gain at finishing (ADGf), and backfat (BF). The core sizes were determined based on the number of largest eigenvalues explained 50% (n=160), 80% (n=700), 90% (n=1,363), 95% (n=2,344), 99% (n=7,320) and >99% (n=10,000) of the variation in G. The solid black line is the prediction accuracy derived using the regular ssGBLUP as the baseline.

To conclude, the core size is more important than the core definition. Randomly sampling the core from the pool of genotyped animals or evenly selecting animals across generations with the size equal to the number of largest eigenvalues explaining 99% of the variation in G is adequate for accurately predicting GEBV using APY ssGBLUP.

References

Aguilar I., Misztal I., Johnson D.L., Legarra A., *et al.* (2010) *J Dairy Sci 93*(2):743-752. https://doi.org/10.3168/jds.2009-2730

Bradford H.L., Pocrnić I., Fragomeni B.O., Lourenco D., *et al.* (2017) *J Anim Breed Genet 134*(6):545-552. https://doi.org/10.1111/jbg.12276

Fragomeni B.O., Lourenco D., Tsuruta S., Masuda Y., *et al.* (2015) *J Dairy Sci 98*(6):4090-4094. https://doi.org/10.3168/jds.2014-9125

Misztal I. (2016) *Genetics 202*(2):401-409. https://doi.org/10.1534/genetics.115.182089

Misztal I., Legarra A., and Aguilar I. (2014) *J Dairy Sci 97*(6):3943-3952. https://doi.org/10.3168/jds.2013-7752

Ostersen T., Christensen O.F., Madsen P., and Henryon M. (2016) *Genet Sel Evol 48*(1):48. https://doi.org/10.1186/s12711-016-0227-8

Pocrnic I., Lourenco D., Masuda Y., and Misztal I. (2016) *Genet Sel Evol 48*(1):82. https://doi.org/10.1186/s12711-016-0261-6

Vandenplas J., Calus M.P.L., and Ten Napel J. (2018) *J Anim Sci 96*(6):2060-2073. https://doi.org/10.1093/jas/sky136

332. Strategies for choosing core animals in APY and their impact on the accuracy of single-step genomic predictions

N.P.P. Macciotta[1*], C. Dimauro[1], D. Lino Laurenco[2], A. Cesarani[12], L. Degano[3] and D. Vicario[3]

[1]Dipartimento di Agraria, University of Sassari, 07100 Sassari, Italy; [2]Department of Animal and Dairy Science, University of Georgia, Athens, GA 30602, USA; [3]ANAPRI, 33100 Udine, Italy; macciott@uniss.it

Abstract

Different strategies were compared for defining core animals in the APY algorithm using either simulated or real data: random; youngest; rank based on the contribution to the genomic relationship matrix. The selection of animals with the lowest rank according to their contribution to the G matrix provided slightly higher accuracies, especially in the simulated data. In real data, particularly for larger sizes of core animals, the criteria of choice appear less important, confirming results of earlier studies. Data considered in the present work were characterized by a complete pedigree. Thus, the key contributors approach for core definition could also be tested under these conditions and with large datasets.

Introduction

Single-step GBLUP (ssGBLUP) has enabled genomic selection (GS) application in farm animal populations where only a portion of the animals in the pedigree are genotyped. The standard ssGBLUP requires the inversion of the genomic relationship matrix (G), which is computationally feasible for up to 150,000 genotyped individuals (Bradford *et al.*, 2017). However, after more than ten years of GS, the number of genotyped animals has passed the one million mark in some cattle populations (https://queries.uscdcb.com/Genotype/counts.html). An option to face this challenge is using the Algorithm for Proven and Young (APY), where genotyped animals are divided into core and noncore, with only the portion of G corresponding to the core animals directly inverted (Misztal *et al.*, 2014). Pocrnic *et al.* (2016) showed that the APY method is able to provide the same GEBV accuracies of the standard ssGBLUP when the number of core animals corresponds to that of independent chromosome segments (M_e) segregating in the population. The M_e is defined as $4N_eL$ (Stam, 1980), where N_e is the effective population size and L is the genome length in Morgans. This number can be approached by the number of eigenvalues explaining 98% of the variance of G (Pocrnic *et al.*, 2016). In general conditions, a random choice of core animals seems to be the best option (Bradford *et al.*, 2017; Fragomeni *et al.*, 2015). Although the random choice of core animals is widely accepted and used in APY ssGBLUP evaluations, some studies have systematically investigated ways to identify core animals (Bradford *et al.*, 2017; Ostersen *et al.*, 2016). In this work, we applied the concept of key contributors (Neuditschko *et al.*, 2017) to identify animals that affect the (co) variance structure of G. This approach is based on the eigenvalue decomposition of G, and the resulting key contributors were used as core animals in APY ssGBLUP.

Materials & methods

Data. Two data sets were used. The first was a simulated data set generated for the XVII QTL MAS workshop (Usai *et al.*, 2012). It consisted of a population with 4,100 individuals (100 males), spread into five generations with a genome of 5 chromosomes, each carrying 4,000 equally spaced SNPs. A complete pedigree was available. Three phenotypes were also available for females, with heritabilities of 0.35, 0.35, and 0.50, respectively.

The second was a real data set of 136k milkability records of Italian Simmental cattle (Cesarani *et al.*, 2021). A total of 11,636 animals were genotyped with 42,141 SNPs. Genotyped cows born from 2017 to 2019 were considered as the validation population. Scenarios are represented in Table 1.

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_332, © N.P.P. Macciotta *et al.* 2022

Statistical analysis. Using the simulated data, two multivariate statistical analyses were performed to derive the contribution of each animal to G (G_{contr}). The first was a principal component analysis (PCA). G_{contr} was calculated as:

$$G_{contr} = \sum_{i=1}^{i=k} \left(\frac{a_{ij}}{\sqrt{\lambda_i}} \right)^2 \quad (1)$$

where a is the coefficient of the i-th eigenvector for the j-th animal; λ is the i-th eigenvalue; k is the number of retained eigenvectors. The second was a Multivariate Factor Analysis (MFA). In this case, G_{contr} was calculated as:

$$G_{contr} = \sum_{i=1}^{i=k} {l_{ij}}^2 \quad (2)$$

where l is the loading of the i-th rotated factor for the j-th animal; k was defined before. In factor terms, the G_{contr} is equivalent to the communality for each animal.

Four strategies were then compared for structuring the genetic (co)variance between individuals:

1. BLUP using the average relationship matrix (A);
2. ssGBLUP, using the inverse of the realized relationship matrix (H) (Aguilar *et al.*, 2010);
3. $ssGBLUP_{APY}$, with the APY algorithm and three core sizes (10, 20, 30%).

Four criteria for choosing the genotyped (ssGBLUP) and the core ($ssGBLUP_{APY}$) animals were tested: (1) random; (2) the best for G_{contr}; (3) the bottom for G_{contr}; (4) the youngest. For criteria 2 and 3, G_{contr} was calculated using either PCA or MFA. Thus, 1000 (random, best, bottom, youngest) animals were genotyped among the 4,100 available for ssGBLUP; among them, 10, 20, 30% were selected as core in $ssGBLUP_{APY}$.

For the ssGBLUP applied to real data, 10, 20, and 30%, of genotyped animals were considered (321 candidate cows plus other animals). In $ssGBLUP_{APY}$, all genotyped animals were considered and three different core sizes (10, 20, and 30%) were tested. Genotyped or core animals in the two methods were chosen according to the four criteria used in simulated data.(Table 1).

Validation was assessed by calculating the correlation between (G)EBV and true breeding values of the validation population (simulated data), or by the correlation between (G)EBV when proven and (G)PA of candidates when born (real data).

Results

Accuracies of (G)EBV for the simulated data are reported in Figure 1. For all the three traits, smaller values were showed by BLUP compared to genomic model. For ssGBLUP and $ssGBLUP_{APY}$, the choice of bottom G_{contr} animals yielded higher accuracies.

Table 1. Data structure for the different scenarios of GS.

Data	Scenario	Genotyped	Validation	Core size
Simulated	ssGBLUP	1000	1000	
	$ssGBLUP_{APY}$	1000		100, 200, 300
Real	ssGBLUP	1,164, 2,327, 3,491	321	
	$ssGBLUP_{APY}$	11,636		1,164, 2,327, 3,491

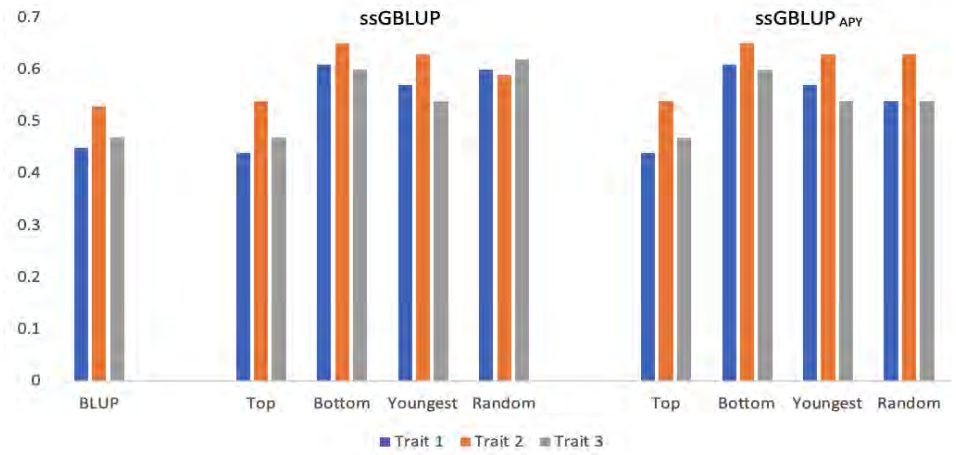


Figure 1. Accuracies of (G)EBV predictions for simulated data under different models and strategies of choice of core animals.

Figure 2 reports results for the real data. Also in this case, as expected, BLUP showed the lowest prediction accuracy. An improvement of accuracy can be observed as the proportion of genotyped animals increases. A clear pattern of the strategy used for choosing core animals cannot be detected. In the core animal size 10% scenario, slightly better accuracies can be observed for the random choice, whereas for larger core sizes (i.e. 20 and 30%), youngest or bottom G_{contr} animals provided larger values. As expected, using all genotyped animals (i.e. $ssGBLUP_{APY}$) provided better results than ssGBLUP.

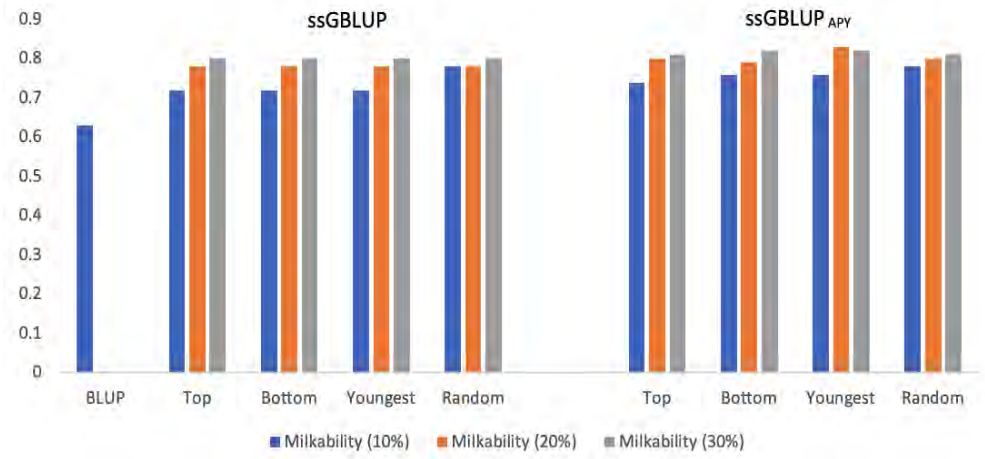


Figure 2. Accuracies of (G)EBV predictions for real data under different models and strategies of choice of core animals.

Discussion

The optimization of genotyping strategies and the definition of core animals have been widely investigated in livestock. Theoretically, to maximize the reliability of genomic predictions, genotyped animals should represent all the independent chromosome segments existing in the population (Misztal, 2016). Previous studies have found no basic differences among criteria used to define core animals, and random selection was the better option (Bradford *et al.*, 2017; Fragomeni *et al.*, 2015). In the present study, the use of different strategies either in ssGBLUP or in ssGBLUP$_{APY}$ resulted in differences, even if small, in genomic prediction accuracies. Among the various methods, the selection of animals with the lowest rank according to their contribution to G evaluated by PCA provided slightly higher accuracies, especially in the simulated data. However, the core sizes used in this study were smaller than the ideal number specified as M_e or the number of eigenvalues explaining 98% of the variance of G. In such a case, different core definitions may result in different accuracies. In real data, particularly for larger sizes of core animals, the criteria of choice appear less important, confirming results of earlier studies. Bradford *et al.* (2017) pointed out that the core definition may be important in the case of incomplete pedigrees. Data considered in the present work were characterized by a complete pedigree. Thus, the key contributors approach for core definition could also be tested under these conditions and with large datasets.

References

Aguilar I., Misztal I., Johnson D.L., Legarra A., Tsuruta S., *et al.* (2010) J. Dairy Sci. 93(2):743–752. https://doi.org/10.3168/jds.2009-2730

Bradford H.L., Pocrnic I., Fragomeni B.O., Lourenco D.A.L., and Misztal I. (2017) J. Anim. Breed. Genet. 134(6):545–552. https://doi.org/10.1111/jbg.12276

Cesarani A., Garcia A., Hidalgo J., Degano L., Vicario D., *et al.* (2021) J. Dairy Sci. 104(5):5719-5727. https://doi.org/10.3168/jds.2020-19838

Fragomeni B.O., Lourenco D.A.L., Tsuruta S., Masuda Y., Aguilar I., *et al.* (2015) J. Anim. Breed. Genet. 132(5):340-345. https://doi.org/10.1111/jbg.12161

Misztal I. (2016) Genetics. 202, 401–409. https://doi.org/10.1534/genetics.115.182089

Neuditschko M., Raadsma H.W., Khatkar M.S., Jonas E., Steinig E.J., *et al.* (2017) PLoS ONE 12(5):e0177638. https://doi.org/10.1371/journal.pone.0177638

Ostersen T., Christensen O.F., Madsen P., and Henryon M. (2016) Genet. Sel. Evol. 48(1):1-10. https://doi.org/10.1186/s12711-016-0227-8

Pocrnic I., Lourenco D.A., Masuda Y., Legarra A., and Misztal, I. (2016) Genetics 203(1):573-581. https://doi.org/10.1534/genetics.116.187013

Stam P. (1980). Genet. Res. 35(2):131-155. https://doi.org/10.1017/S0016672300014002

Usai M.G., Gaspa G., Macciotta N.P.P., Carta A., and Casu S. (2014) BMC Proceedings 2014, 8(Suppl 5):S1. https://doi.org/10.1186/1753-6561-8-S5-S1

333. Genomic prediction accuracy: how low can we go?

A. Reverter[1*], P.A. Alexandre[1], Y. Li[1], B.C. Hine[2], C.J. Duff[3], A.B. Ingham[1] and L.R. Porto-Neto[1]

[1]CSIRO, Agriculture and Food, Queensland Bioscience Precinct, 306 Carmody Rd., St Lucia, Brisbane, QLD 4067, Australia; [2]CSIRO, Agriculture and Food, F.D. McMaster Laboratory, Chiswick, New England Highway, Armidale, NSW 2350, Australia; [3]Angus Australia, 86 Glen Innes Rd., Armidale, NSW 2350, Australia; tony.reverter-gomez@csiro.au

Abstract

Accurate genomic predictions underpin the success of genomic selection-based breeding programs. Therefore, efforts to compute and improve prediction accuracy (ACC) are imperative. Here we use theoretical expectations and simulation to highlight the linear relationship that exists between ACC and phenotypic differences in pre-defined percentiles. We found that an ACC of 0.4 yielded a phenotype difference of one standard deviation (SD) units between the top and bottom quartiles. In addition, a seemingly moderate increase in ACC of 0.05 (absolute value) translates into an increase of 0.127 SD units between top and bottom quartiles and to a 2.54 SD units' difference when extrapolated to a perfect ACC. These theoretical expectations were validated with results from a dataset of 3,408 Australian Angus steers with genomic predictions for a range of feedlot and carcase traits. We conclude that moderate predictions accuracies of ~0.4 are sufficient to effectively draft animals into quartiles.

Introduction

In the last decade there has been a tremendous research effort aimed at boosting the accuracy (ACC) of genomic predictions (GEBVs), by exploring alternative analytical models, genomic relationship matrices, numerical approaches and uses of functional biology information. To name a few recent examples from livestock, Moghaddar *et al.* (2021) showed that using prioritized genetic markers and accounting for different allele frequencies in reference populations resulted in a slightly higher ACC. Similarly, Hidalgo *et al.* (2021) concluded that including genomic information in training sets provided an increase in ACC, while Brito Lopes *et al.* (2020) showed that an artificial neural network can provide superior genomic predictions.

However, one common finding among these studies is that the observed increase in prediction ACC is seemingly and deceptively small, ranging from 0.02 to 0.08 (absolute value). Clouding the impact of these small improvements is the difficulty in translating ACC values into tangible measures for seedstock and commercial producers.

Therefore, the objective of this study was to develop theoretical expectations, along with simulation studies and results from real data, to explore the relationship between prediction ACC, ACC changes, and phenotypic differences on the ability of predictions to distinguish extreme performing individuals.

Materials & methods

Expectations under standard normal distribution theory. We are interested in the average value beyond a nominal percentile threshold T:

$$\mathrm{E}[x/x > T] = \frac{1}{\sqrt{2\pi}} \int_T^{\infty} x e^{\frac{-x^2}{2}} dx = -\frac{1}{\sqrt{2\pi}} \int_T^{\infty} e^{\frac{-x^2}{2}} d\left(\frac{-x^2}{2}\right) = \frac{1}{\sqrt{2\pi}} e^{\frac{-T^2}{2}} \quad (1)$$

The above expression needs to be divided by the cumulative density of the percentile in question. Five percentile splits will be considered to divide the data into equal parts, including: 50% (halves), 33% (thirds),

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_333, © A. Reverter *et al.* 2022

25% (quarters), 20% (fifths) and 10% (tenths). For each percentile split, the value of interest is the expected difference, in standard deviation (SD) units, between the top and bottom percentile.

Simulation of correlated normal random variables. To explore the relationship between ACC and phenotypic differences in pre-defined percentile splits we simulated a series of bivariate standard normal variables with a given correlation and for n=10,000 animals. In this setting, the two variables were intended to represent the true and predicted genetic values while the correlation indicated ACC. Nineteen ACC values were explored from 0.05 to 0.95 using 0.05 increments. Each simulation scenario was repeated 100 times and at each time, animals were ranked based on the prediction and the difference in true value between the top and bottom percentile, for the five percentile splits, was recorded.

Validation with real data. To validate the findings yielded by the simulation, we interrogated the data from the study of Alexandre *et al.* (2021) which was based on a population of 3,408 Australian Angus steers with GEBVs for two feedlot and five carcase traits: average daily gain, average daily dry matter intake, carcase weight, eye muscle area, marbling score, ossification score and subcutaneous rib fat depth. Seven year of birth cohorts were represented in the data and each was used in turn as validation samples to generate seven average ACC values for each phenotype and where ACC was computed as the correlation between the GEBVs and the adjusted phenotype divided by the square root of heritability. The 49 ACC values (from seven traits by seven validation cohorts) averaged 0.43 and ranged from 0.08 to 0.76. In each case, adjusted phenotypes from the validation sample were sorted according to the GEBVs and the difference between the top and bottom percentile expressed in SD units.

Results

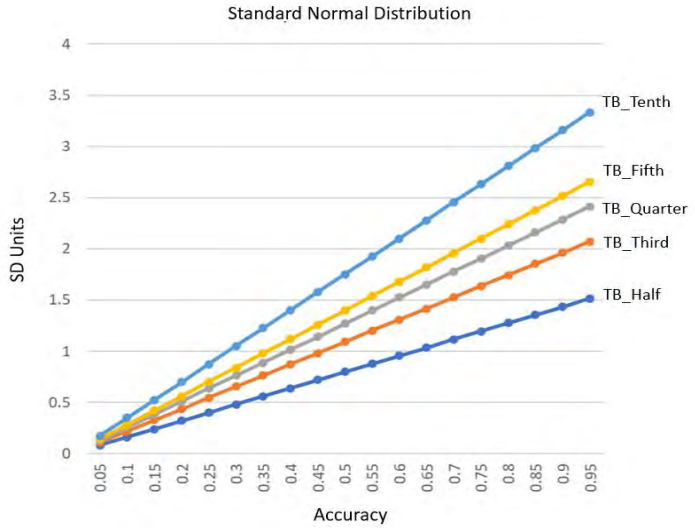
Simulation study. Table 1 presents the theoretical thresholds and expected values from a standard normal distribution and five percentile splits. At 50% percentile, corresponding to category 'Half', the threshold is zero and values above zero are expected to average 0.797 SD units. The difference between the top and bottom half is then twice 0.797 or 1.595 SD units. Similarly, at 10% percentile, corresponding to category 'Tenth', the threshold is 1.282 and values beyond this threshold average 1.754 and the expected difference between the top and bottom 10% equals 3.509 SD units. Figure 1A illustrates the difference in true value, averaged over 100 simulations, between the top and bottom (TB) for each percentile (TB_Half, TB_Third, TB_Quarter, TB_Fifth and TB_Tenth) and at each simulated value for the correlation between the true and the predicted value.

Real data validation. Figure 1B presents the equivalent results based on the real dataset for the five percentiles and 49 ACC values available. Overlaid in Figure 1B are the regression equations with a restriction of zero intercept and their corresponding goodness of fit based on the R^2 coefficient.

Table 1. Theoretical thresholds (T) and expected mean values ($E[x>T]$) for five percentile categories from a standard normal distribution.

Category	Percentile, %	Threshold, T	$E[x>T]$	Expected difference top – bottom
1 Half	50%	0.000	0.797	1.595
2 Third	33%	0.431	1.090	2.181
3 Quarter	25%	0.674	1.270	2.541
4 Fifth	20%	0.841	1.399	2.798
5 Tenth	10%	1.282	1.754	3.509

(A)



(B)

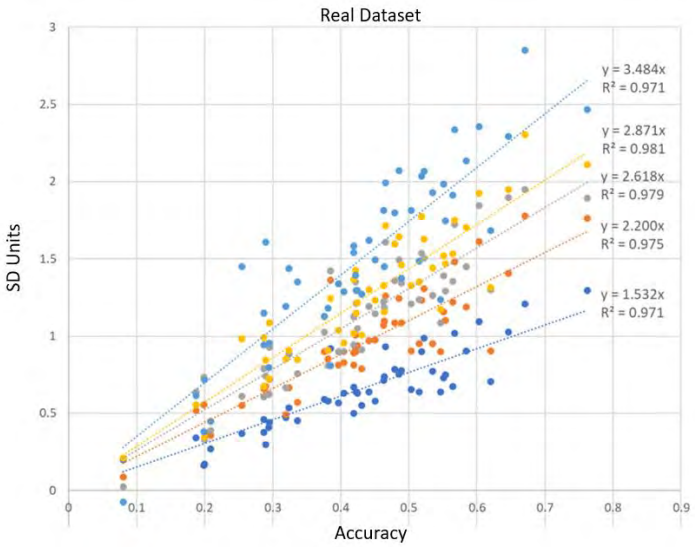


Figure 1. Simulation based on a standard normal distribution (A): difference in true value between the top and bottom for each percentile split with 19 values of accuracy, increasing stepwise, between the true and the predicted value. Real dataset (B): difference in adjusted phenotypes in standard deviation units between the top and bottom for each percentile for 49 accuracy values (from seven traits by seven validation cohorts).

Discussion

While efforts to increase the accuracy of genomic predictions should continue, genomic-based technologies are already allowing commercial beef producers to predict the genetic merit of individual animals of unknown pedigree with increased ease and accuracy (Hine *et al.* 2021).

The expected difference between the top and bottom percentile (Table 1, last column) indicates what could be achieved with perfect ACC. This expected maximum difference can be interpreted as the linear regression of phenotypic difference on ACC (Figure 1). Therefore, a seemingly moderate increase in ACC of 0.05 (absolute value) can translate into an increase of 0.05 × 2.5410 = 0.127 SD units between top and bottom quartiles and 0.05 × 3.5088 = 0.175 SD units between the top and bottom deciles.

The simulation study revealed the perfect linear relationship that is expected between prediction accuracy and phenotypic differences in pre-defined percentiles (Figure 1A). At ACC = 0, animals are ranked at random and the expected difference in any percentile is zero. A prediction accuracy of 0.4 can discriminate a difference of 0.64, 0.87, 1.01, 1.12, and 1.40 SD units respectively for 50, 33, 25, 20 and 10% percentiles. Because of linearity, these differences half and double for prediction accuracies of 0.2 and 0.8, respectively.

Using the real dataset (Figure 1B), the range of ACC values from 0.08 to 0.76 allowed us to confirm expectations with high precision. With an R^2>97% in all cases, the fitted regression models were very closely aligned with the expected values (last column of Table 1): 1.532 vs 1.595, 2.200 vs 2.181, 2.618 vs 2.541, 2.871 vs 2.798, and 3.484 vs 3.509 for 50, 33, 25, 20 and 10% percentiles, respectively.

In real units, when applied to the real data, and for the case of carcase weight with a SD of 31.09 kg (Alexandre *et al.* 2021), the average ACC value 0.43 is expected to identify a difference of 0.43 × 2.541 × 31.09 = 33.97 kg between the top and bottom quartiles, and a difference of 0.43 × 3.5088 × 31.09 = 46.91 kg between the top and bottom deciles. Similarly, an ACC increase of 0.05 translates in an increase in discrimination power between top and bottom quartile or decile of 3.95 and 5.44 kg of carcase weight, respectively.

In conclusion, we have shown that the linear relationship that exists between prediction accuracy and predefined differences in actual phenotype allows for the interpretation of accuracy values in a tangible form for scientists and producers. While only proven animals with lots of relatives will attain ACC>90% for a trait of interest, seemingly small ACC maybe sufficient to draft commercial animals in a cost-effective manner.

References

Alexandre P.A., Li Y., Hine B.C., Duff C.J., Ingham A.B. *et al.* (2021) Genet Sel Evol 53(1):77. https://doi.org/10.1186/s12711-021-00673-8

Brito Lopes F., Magnabosco C.U., Passafaro T.L., Brunes L.C., Costa M.F.O. *et al.* (2020) J Anim Breed Genet. 137(5):438-448. https://doi.org/10.1111/jbg.12468

Hidalgo J., Lourenco D., Tsuruta S., Masuda Y., Breen V. *et al.* (2021) J Anim Sci. 99(9):skab239. https://doi.org/10.1093/jas/skab239

Hine B.C., Duff C.J., Byrne A., Parnell P., Porto-Neto L. *et al.* (2021) Anim Prod Sci. (Online ahead of print). https://doi.org/10.1071/AN21051

Moghaddar N., Brown D.J., Swan A.A., Gruman P.M., Li L. *et al.* (2021) J Anim Breed Genet. (Online ahead of print). https://doi.org/10.1111/jbg.12638

334. Genomic prediction in smallholder tropical dairy herds: modelling milk yield in first and later lactations

E. Negussie[1*], M.H. Lidauer[1], S. Meseret[3], A. Tera[2], E. Tessema[2], B. Jufar[2], T. Negera[2], A. Hassen[2], G. Gebreyohanes[3], O.A. Mwai[3] and R. Mrode[3]

[1]Natural Resources Institute Finland (Luke), Myllytie 1, 31600 Jokioinen, Finland; [2]Livestock Development Institute, P.O. Box 22692, Addis Ababa, Ethiopia, [3]International Livestock Research Institute, Box 30709-01001 Nairobi, Kenya; enyew.negussie@luke.fi

Abstract

This study examined ways to correct for environmental effects and effectively utilize all the available performance records in Ethiopian smallholder dairy production systems. Modelling contemporary groups based on villages rather than the usual herd resulted in proper correction for environmental effects and gave higher (0.18 vs 0.12) heritability estimates for milk yield. Besides, in such production systems farmers tend to milk cows for several later lactations and performance records are a scarce resource. Thus, the use of a multiple-trait single-step genomic prediction model that considered 1st and 2nd plus later lactations as two distinct but correlated traits, gave the highest standard deviation of genomic enhanced breeding values (1.365 vs 0.840 kg/d) and better reliability of genomic prediction than univariate models.

Introduction

In the developing countries, most dairy production occurs in smallholder production systems, where herd sizes are small and are with little or no routine systems for capturing performance and pedigree data (Kosgey and Okeyo, 2007). With the advent of genomic selection and the recent successful incorporation of genomic information into breeding schemes, the reliance on very large populations of phenotyped animals has been somewhat relaxed. Therefore, the use of genomic relationship matrix for genotyped animals or the combination of genotypic information with pedigree information applying the single-step GBLUP (ssGBLUP) approach presents unique opportunities to overcome the limiting data and other infrastructural problems. The reliability of genomic prediction among others, depends on the heritability of the trait and the statistical method used to estimate the single nucleotide polymorphism (SNP) effects in the reference population. Further, in tropical dairy herds, farmers tend to milk their cows for up to five or more lactations. In such production systems, accurate estimation of the genetic merit of animals relies on our ability to effectively utilize all the available performance records fitting appropriate statistical models. The main objectives of this study were therefore to develop a single-step genomic prediction model where 1st and 2nd plus all later lactations were considered as distinct but correlated traits and to estimate the accuracy of predicting breeding values.

Materials & methods

Data. Data for this study was obtained from the Ethiopian national dairy cattle performance recording, extracted from the International Livestock Research Institute (ILRI) database (https://www.adgg.ilri.org/uat/auth/auth/login) for cows registered in the Luke ICI project and African Dairy Genetic Gain (ADGG) program from 1990 to 2020. The raw data consisted of 272,447 test-day (TD) milk observations for cows milking from 1 to 10 lactations. Prior to analyses, the data was edited to include cows with age at first calving ≥18 months, milk yield between 1 and 40 liters, and records from 5 to 365 days in milk (DIM). Calving seasons were divided into three, Long dry (October to February), Short rainy (March to May) and Long rainy seasons (June to September), respectively. Cow contemporary groups was tailored to dairy farming practice in the Ethiopian highlands. Thus, contemporary groups were specified by village × calving

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_334, © E. Negussie *et al.* 2022

year × calving season (VYS) classes. Although, the data included records of cows up to 10 lactations, to make effective use of all available TD records, the whole data was divided into two distinct traits including 1st and 2nd plus later lactations. The final edited data consisted of 152,133 TD milk records from 17,501 dairy cows in 6,140 herds located within 544 villages with more data came from recent years (after 2010). The pedigree included 21,501 animals. For estimation of variance components (VCE), a sample of the data was extracted. To avoid selection bias, cows with later lactation records were required to have valid first lactation records and sampled herds had to have at least 5 first calvers per calving year. There were 56,770 and 33,506 observations in 1st and 2nd plus later lactations in the data for VCE.

Genotypes. Genotypic data were available on 5,146 animals which were genotyped by the ADGG program with the GeneSeek Genomic Profiler (GGP) Bovine 50K chip. About 47,843 SNPs were returned from the laboratory and after the edits (Aliloo *et al.*, 2018), 40,581 SNPs were made available for analysis. These were imputed to the Illumina HD chip (686,052 SNPs) using a reference population consisting of crossbred cattle from a previous East Africa Dairy Genetics Gain project and several European Holstein-Friesian, Jersey, Guernsey, and Ayrshire purebred animals (Aliloo *et al.*, 2018). Of the 17,501 animals in the final data, 5,119 were genotyped and 2,140 cows were genotyped and phenotyped.

Model. For estimation of variance-covariance components and associated genetic parameters, a multiple-trait repeatability TD animal model, where 1st and 2nd plus later lactations (2+) were considered as two different traits, was fitted. The model in matrix form can be described as follows:

$$\begin{bmatrix}\mathbf{y}_{i1}\\ \mathbf{y}_{i2+}\end{bmatrix} = \begin{bmatrix}\mathbf{X}_{i1} & 0\\ 0 & \mathbf{X}_{i2+}\end{bmatrix}\begin{bmatrix}\boldsymbol{\beta}_{i1}\\ \boldsymbol{\beta}_{i2+}\end{bmatrix} + \begin{bmatrix}\boldsymbol{\Phi}_{i1} & 0\\ 0 & \boldsymbol{\Phi}_{i2+}\end{bmatrix}\begin{bmatrix}\boldsymbol{\lambda}_{i1}\\ \boldsymbol{\lambda}_{i2+}\end{bmatrix} + \begin{bmatrix}\mathbf{Z_1}_{i1} & 0\\ 0 & \mathbf{Z_1}_{i2+}\end{bmatrix}\begin{bmatrix}\mathbf{htd}_{i1}\\ \mathbf{htd}_{i2+}\end{bmatrix} + \begin{bmatrix}\mathbf{Z_2}_{i1} & 0\\ 0 & \mathbf{Z_2}_{i2+}\end{bmatrix}\begin{bmatrix}\mathbf{pe}_{i1}\\ \mathbf{pe}_{i2+}\end{bmatrix} + \mathbf{Z_3}_{i2}\mathbf{pe2L}_{i2+} + \begin{bmatrix}\mathbf{Z_4}_{i1} & 0\\ 0 & \mathbf{Z_4}_{i2+}\end{bmatrix}\begin{bmatrix}\mathbf{a}_{i1}\\ \mathbf{a}_{i2+}\end{bmatrix} + \begin{bmatrix}\boldsymbol{\varepsilon}_{i1}\\ \boldsymbol{\varepsilon}_{i2+}\end{bmatrix} \quad (1)$$

where y_{i1} and y_{i2+} are the vectors of TD milk yield observations, β_1 and β_{2+} are the vectors of fixed effects consisting of VYS classes, calving age classes, and breed group, λ_1 and λ_{2+} are vectors of the regression coefficients for the fixed lactation curves modelled by 4th order Legendre polynomials plus the Wilmink term, which were nested within calving season classes; htd_{i1} and htd_{i2+}, pe_{i1} and pe_{i2+}, $pe2L_{i2+}$, a_{i1} and a_{i2+}, and e_{i1} and e_{i2+} are vectors of random effects of herd-test-day, permanent environmental effects nested within lactation and across traits, permanent environmental effects within later lactations (2 to 10 lactations) to model repeatability across later lactations, additive animal genetic effects, and residual effects respectively. The design and coefficient matrices X, Φ, Z_1, Z_2, Z_3 and Z_4 related observations to fixed β and l effects and random htd, pe, pe2L and a effects, respectively. Variance components were estimated using BLUP animal model and the average information algorithm.

Genomic prediction. A multiple-trait single-step genomic prediction model (ssGBLUP_mt) where 1st and 2nd plus all later lactations were considered as two distinct traits as described in Equation 1 was developed. Relationships between animals were described by the H matrix that combines the marker-based (G) and pedigree-based (A) relationship matrices. The H matrix was built by blending the G matrix with 10% of the residual polygenic effect.

Model comparison and validation. For validation purposes, the ssGBLUP_mt model was compared against univariate BLUP (pBLUP_univ) and univariate single-step genomic prediction (ssGBLUP_univ) model. The univariate models considered all TD milk observations of the data as the same trait and included the same model effects as specified in Equation 1 for y_{i1}. Model comparisons were made in terms of correlations between and standard deviation (SD) of estimated breeding values (EBV) and genomic enhanced breeding values (GEBV). Predictions were carried out using the MiX99 software (Strandén *et*

al., 2018). In addition, forward validation was used to assess the accuracy of predicting breeding values for candidate animals with only genotype information. For this, milk records of cows born in 2015 and afterwards (2,872 cows with total 23,575 records) in the genomic prediction system were excluded. The reliability of genomic predictions for the candidates whose records were excluded was computed as squared correlation coefficient between the candidates' GEBV and their yield deviations (YD) divided by $\sqrt{h^2}$. The degree of under or over-prediction was estimated by regressing their GEBV on their YD.

Results

Average test-day milk production. The overall mean (±SD) for TD milk in 1st and 2nd plus later lactations in smallholder cows in Ethiopia were 10.6 (5.4) and 11.0 (4.6), respectively. The lactation curves for TD milk nested within calving seasons in Figure 1 shows that cows calving in the long rainy season peaks higher than those that calved in the two other seasons. In addition, Figure 1 towards the end of the lactation curve shows unrealistic upward trend when TD milk observations beyond 350 DIM were included in the analyses.

Variance components. The estimates of heritability from the multiple-trait model for the 1st and 2nd plus later lactations were 0.18 and 0.17, respectively. The repeatability of TD milk yield in 1st and 2nd plus later lactations were 0.67 and 0.57, respectively. Genetic correlation between 1st and 2nd plus later lactations was 0.87 and repeatability across later lactations was 0.34.

Animal evaluation and validation of genomic prediction. Table 1 presents the comparison between pBLUP_univ, ssGBLUP_univ and ssGBLUP_mt animal evaluations in terms of correlations between models and SD of (G)EBVs. For cows with at least 5 TD milk observations, the SD of breeding values was 0.97 liters for ssGBLUP_mt, whereas it was 0.59 liters when applying the ssGBLUP_univ and was as low as 0.57 liters when applying pBLUP_univ model. The SD of (G)EBVs in bulls were higher than those in cows and were 1.37 liters, 0.91 liters and 0.84 for the ssGBLUP_mt, ssGBLUP_univ and pBLUP_univ models, respectively. For cows with observations, the correlations between pBLUP_univ and ssGBLUP_mt, and between ssGBLUP_univ and ssGBLUP_mt models were 0.87, and 0.89, respectively. Whereas for bulls, the correlations between pBLUP_univ and ssGBLUP_mt, and between ssGBLUP_univ and ssGBLUP_mt models were 0.88, and 0.89, respectively. Validation reliabilities for 2,872 validation cows was 0.078 for

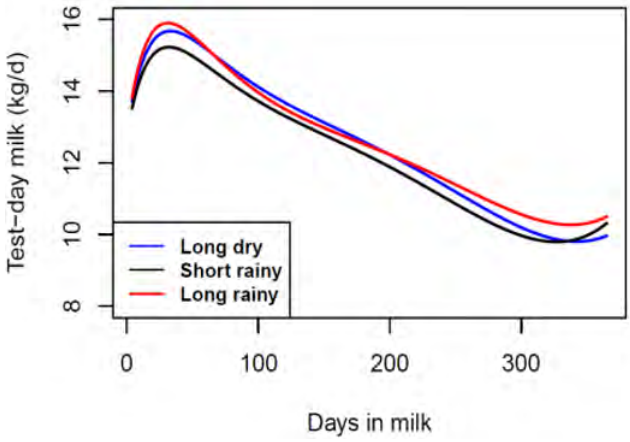


Figure 1. Lactation curves of cows nested within calving seasons in smallholder dairy herds for TD milk data from 5 to 365 days in milk.

Table 1. Correlations between pBLUP_univ, ssGBLUP_univ and ssGBLUP_mt models and their corresponding standard deviation of EBVs and GEBVs.

Models	Correlations[1]		Standard deviation of G(EBV)	
	Cows[2]	Bulls[3]	Cows[2]	Bulls[3]
pBLUP_univ	0.87	0.88	0.57	0.84
ssGBLUP_univ	0.89	0.89	0.59	0.91
ssGBLUP_mt			0.97	1.37

[1] Pearson correlations between ssGBLUP_mt GEBVs and pBLUP_univ EBVs, and between ssGBLUP_mt GEBVs and ssGBLUP_univ GEBVs.

[2] Cows = 10,277 cows >5 TD milk records.

[3] Bulls = 106 bulls >5 daughters.

pBLUP_univ model. Whereas it was 0.091 and 0.101 for the ssGBLUP_univ and ssGBLUP_mt models, respectively. The coefficient from the regression of predicted candidates (G)EBVs on candidates YD was 0.77, 0.87 and 0.87 for the pBLUP_univ, ssGBLUP_univ and ssGBLUP_mt models, respectively.

Discussion

The average daily milk yield from most tropical dairy production systems is reported to be about 6.4 litres (Makau *et al.* 2020). The estimates from this study for the 1st and 2nd plus later lactations is slightly higher but it is close to those reported by (Mrode *et al.* 2021) for the high-grade cows from Tanzania. Estimated heritability from the ssGBLUP_mt model was much higher than 0.12 reported earlier on part of same data. The genetic correlation estimated between 1st and 2nd plus later lactations was lower than unity indicating that milk yield in 1st and later lactations are indeed distinct traits. Thus modelling 1st and 2nd plus later lactations as different but correlated traits is one logical way forward. The ssGBLUP_mt model had the highest SD of GEBVs offering much more variability with greater options for identifying the best animals for breeding. The lower than unity correlations between the three models for both cows and bulls indicate a possible re-ranking of top cows and bulls. As expected, the validation reliabilities of ssGBLUP models were higher than pBLUP. Gains of 0.013 and 0.011 on validation reliability were achieved by using ssGBLUP_mt model. These values are slightly lower than those reported by Mrode *et al.* (2021) and could possibly be due to slightly lower number of genotyped animals with data included or due to the variance components used. Therefore, it is expected that as and when all the genotyped animals get sizable phenotypic records the situation will improve.

References

Aliloo H., Mrode R., Mwai A.O., Ni G., Goddard M.E., *et al.* (2018). J. Dairy Sci. 101, 9108–9127. https://doi.org/10.3168/jds.2018-14621

Kosgey I.S. and Okeyo A.O. (2007). Small Rumin. Res. 70, 76–88. https://doi.org/10.1016/j.smallrumres.2007.01.007

Makau D.N., VanLeeuwen J.A., Gitau G.K., McKenna S.L., Walton C., *et al.* (2020). Veterinary Medicine International. https://doi.org/10.1155/2020/3262370

Mrode R., Ojango J., Ekine-Dzivenu C., Aliloo H., Gibson J., *et al.* (2021) J. Dairy Sci. 104:11779–11789. https://doi.org/10.3168/jds.2020-20052

Strandén I., Taskinen M., Matilainen K., Lidauer M. and Mäntysaari E.A. (2018) Proc. of the 11th WCGALP, Auckland, New Zealand.

335. Genomic prediction of wool shedding in a composite sheep flock using pooled allele frequencies and individual genotypes

N. Vargas Jurado[1,3*], J.W. Keele[2], L.A. Kuehn[2] and R.M. Lewis[1]

[1]University of Nebraska – Lincoln, Lincoln, NE 68583, USA; [2]USDA, ARS, U.S. Meat Animal Research Center, Clay Center, NE 68933, USA; [3]Current address: Natural Resources Institute Finland (Luke), 31600, Jokioinen, Finland; napoleon.vargas@luke.fi

Abstract

Due to rising shearing costs and fluctuations in wool prices, developing composite sheep breeds that shed their wool annually has gained interest. Pooling DNA of progeny is a cost-efficient alternative to estimate genomic breeding values (GEBV) of sires. The aim of the study was to measure the correlation between sire GEBV obtained by pooling and mean sire progeny phenotype scaled by the heritability. Wool shedding records on lambs and adults from a composite flock were analysed, and a total of 48 sires and 48 pools were genotyped. Pools were constructed using partitioning around medoids. The correlation between genomic and pedigree relationships was 0.83±0.02. Accuracy of GEBV for sires was 0.41 (0.14-0.62) while the correlation between GEBV and EBV obtained from pedigree best linear unbiased prediction was 0.79 (0.65-0.88). In conclusion, GEBV of sires obtained by pooling allele frequency of their progeny and pedigree EBV were consistent.

Introduction

Rising shearing costs coupled with considerable variation in wool prices have resulted in an increased interest in the use of breeds and their crosses that shed their wool annually. Crossbred lines (composites) that partially or fully shed their wool have been developed by introducing hair genetics (Wiltshire Horn, Dorper, Katahdin) into wool flocks (Johnson *et al.*, 2007; Pollot, 2010; Vargas Jurado *et al.*, 2016).

Wool shedding (WS) is commonly assessed using discrete categories and animals are assigned to them based on their extent of shedding. The number of categories is usually not standard (Pollot, 2010; Matika *et al.*, 2013; Vargas Jurado *et al.*, 2020). Despite the ordinal nature of WS and differences in scoring systems, heritability of WS ranged from 0.26±0.06 (Pollot, 2010) to 0.83±0.06 (Matika *et al.*, 2013). Still, inclusion of WS in routine genetic evaluation is limited. Similarly, genomic analyses of WS either through marker-assisted selection or genome-wide association studies have not been extensive (Johnson *et al.*, 2007; Matikka *et al.*, 2013).

The cost of genotyping has been an impediment to uptake of genomics in sheep (Rupp *et al.*, 2016). Using pooled allele frequencies for the prediction of the genetic merit has been shown to be an economic and reliable alternative to the use of individual genotypes (Bell *et al.*, 2018; Alexandre *et al.*, 2019; Vargas Jurado *et al.*, 2021). In this context, the aim of this study was to estimate sire genomic breeding values (GEBV) using their individual genotypes along with pooled allele frequency and WS scores of relatives, and to estimate prediction accuracy.

Materials & methods

All animal use was approved by the U.S. Meat Animal Research Center (USMARC, Clay Center, Nebraska, USA) Institutional Animal Care and Use Committee.

Data. Phenotypic records were obtained from a crossbred flock composed of five lines resulting from crosses of Romanov, Katahdin, and White Dorper sheep (Vargas Jurado *et al.*, 2016). Wool shedding was

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_335, © N. Vargas Jurado *et al.* 2022

visually assessed using a 9-point scoring system, from zero to eight, where zero and one represented animals with less than 25% wool coverage, two to six represented animals between 25 and 75% wool coverage and seven and eight represented animals with more than 75% wool coverage. A total of 4,767 records from lambs at approximately 170 (SD 12.0) d of age and 3,345 repeated records from 1,527 adult ewes (1, 2, 3, or 4 yr old). While some sires had a lamb record, none had adult wool shedding records. The pedigree consisted of 9,682 individuals and eight genetic groups based on genetic line.

Pools and genotypes. Because of the ordinal nature of wool shedding, individuals were allocated to pools using partitioning around medoids by constructing a distance matrix based on pedigree relationship, contemporary group (line by year of birth), birth and rearing type, weaning weight, and age of dam. These variables were included because they significantly defined variation in WS (Vargas Jurado *et al.*, 2016). A total of 3,038 individuals were allocated to 48 pools. The average number of individuals in a pool was 63.3 (SD 14.8). A total of 48 sires were sampled from the population. Sires and pools were genotyped using the Illumina Ovine 50k chip. For sires and pooled genotypes, fixed markers were removed (i.e. minor allele frequency <0.05). The progeny of sires forming the pools were not individually genotyped.

Variance components and breeding values. Variance components and pedigree best linear unbiased prediction (BLUP) EBV were obtained by fitting a bivariate threshold model where:

$$y_{ij} = k \Leftrightarrow \tau_{j(k-1)} < u_{ij} < \tau_{jk}, \quad (1)$$

where y_{ij} was the j^{th} ordinal trait (j =1, 2 for the lamb and adult WS measurements, respectively) for the i^{th} individual taking values in one of K = 9 possible WS categories (k = {0, 1, …, 8}), τ_j was the vector of thresholds for the j^{th} trait, and u_{ij} was the underlying liability for wool shedding (latent Gaussian variable). In Equation 1 it was further assumed that:

$$\mathbf{u} = \begin{bmatrix} \mathbf{u}_1 \\ \mathbf{u}_2 \end{bmatrix} = \begin{matrix} \mathbf{X_1\beta_1} + \mathbf{Z_1a_1} \\ \mathbf{X_2\beta_2} + \mathbf{Z_2a_2} \end{matrix} + \mathbf{Wc} + \begin{matrix} \mathbf{e}_1 \\ \mathbf{e}_2 \end{matrix} \quad (2)$$

where β_j and a_j were vectors of fixed and additive genetic effects, respectively, and X_j and Z_j were incidence matrices relating to fixed and additive effects, respectively, for the j^{th} trait. W and c corresponded to the vector of permanent environmental random effects and its incidence matrix for adult WS. Fixed effects for lambs included sex and age of dam while for ewes included age at scoring. For both traits contemporary group and birth and rearing category were also included as fixed effects. Residuals were assumed to be Normally distributed and correlated between traits. Random effects in Equation 2 were assumed to follow a normal distribution with zero mean. Additive genetic effects accounted for pedigree relationships and were assumed to be correlated between traits. Permanent environmental effects were uncorrelated to additive genetics effects. The model described by equations 1 to 2 was fitted using the rjmc module in DMU (Madsen *et al.*, 2014). A total of 55,000 samples were produced with 5,000 used as burn-in and a thinning interval equal to five.

Sire and pool GEBV were obtained using genomic BLUP (GBLUP) with a hybrid genomic (G) matrix, containing individually genotyped sires and pool allele frequencies (Vargas Jurado *et al.,* 2021). Pedigree relationships among sires, between sires and individuals in pools, and among individuals within a pool were combined into an average relationship matrix (Ā). Accuracy of GEBV was determined as: (1) the correlation between sire (and pool) GEBV and EBV; and (2) the correlation between GEBV and mean progeny phenotype (for sire families). Lastly, a neighbour joining tree was constructed using sire and pool genotypes (separately) based on Euclidean distance. Nodes were labelled based on their GEBV (below or above the mean).

Results

Heritability of lamb and ewe WS were 0.51±0.03, 0.52±0.03, respectively. The repeatability of ewe WS was 0.61±0.02. The correlation between lamb and adult WS was 0.80±0.02. The overall correlation between elements of G and Ā was 0.85±0.02. Correlations between entries of G and Ā corresponding to relationships among sires, among animals within a pool, and between sires and pools were 0.91±0.01, 0.67±0.03, and 0.66±0.03, respectively. Moderately high correlations between sire and pool GEBV with lamb or adult EBV were found (Table 1).

The accuracy of GEBV measured as the correlation (scaled by $\sqrt{h^2}$) between mean progeny phenotypes (uncorrected) and sire GEBV was 0.41 (confidence interval: 0.14-0.62). Individual sires and pools clustered relatively well on the estimated phylogenetic tree based on whether their GEBV value was above or below the mean (Figure 1).

Discussion

With no genotypes available on individuals forming pools, GEBV and EBV in progeny could not be compared. Therefore, the focus was on the concordance between genomic and pedigree relationships: (1) among sires; (2) among progeny averages within a pool; and (3) between sires and progeny averages within pools. The high concordance in these comparisons (0.52-0.79) was likely due to genomic relationships between sires and pools representing the average of the true relationships in the population (Vargas Jurado *et al.*, 2021). Clusters of sires and pools based on genomic information distinctly showed groupings of individuals with better (or worse) than average performance (GEBV). Furthermore, the clustering methodology showed the value of combining pedigree and systematic environmental effect levels in constructing pools.

Assuming a Gaussian trait and 10 half-sib progeny per sire, accuracy is approximately 0.77. Correlations between EBV, derived from pedigree BLUP and a threshold model, and GEBV were moderately high (0.52 to 0.79) even though WS was a categorical trait with a markedly non-Gaussian distribution (Vargas Jurado *et al.*, 2020). Moreover, the accuracy of GEBV (scaled by $\sqrt{h^2}$) was moderate. Such was encouraging given that uncorrected progeny phenotypes were used to calculate accuracy due to challenges in adjusting the underlying Gaussian liability for WS. Fixed effects adjustments on liability will be considered in the future.

Based on these results, pooling provides an alternative for estimating the genetic merit of sires, even for non-Gaussian traits, especially when economic constraints do not allow for individually genotyping progeny. Moreover, the moderate accuracy of pool GEBV highlights the possibility of using 'average' GEBV as benchmark for comparing the genetic merit of different flocks. Furthermore, these accuracies are comparable to pedigree-based evaluations with a complete pedigree and performance records (Vargas Jurado *et al.*, 2021).

Table 1. Correlation (confidence interval[1]) between sire or pool genomic breeding values[2] with ewe or lamb estimated breeding values from pedigree[3].

	Trait EBV	
Category	**Ewe**	**Lamb**
Sire	0.79 (0.65, 0.88)	0.68 (0.49, 0.80)
Pool	0.52 (0.28, 0.70)	0.56 (0.33, 0.73)

[1] Asymptotic approximation given by Park and Lee (2001).

[2] Estimated from genomic BLUP (GBLUP) with a hybrid G matrix.

[3] Estimated from multiple trait pedigree BLUP.

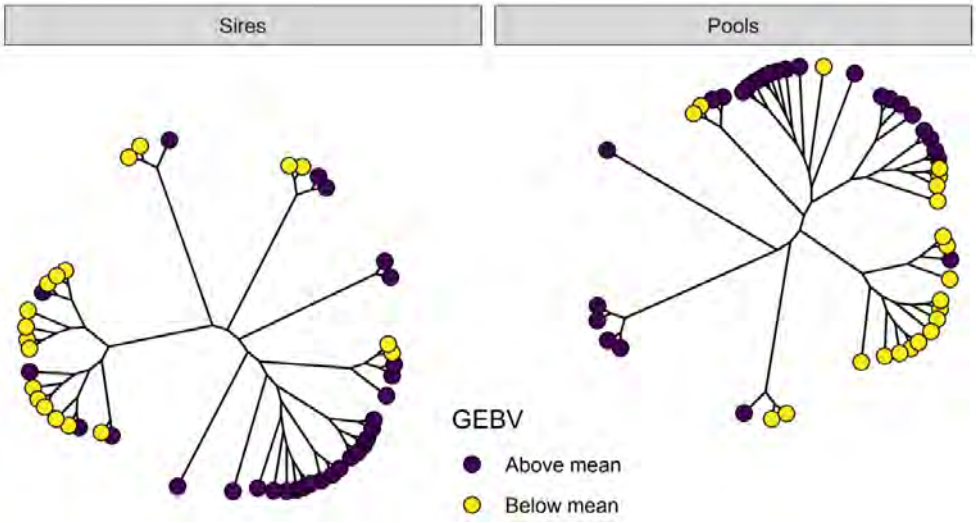


Figure 1. Neighbour joining tree using sire and pool genotypes and a Euclidean distance matrix. Colouring of the nodes was based on genomic breeding values obtained from genomic best linear unbiased prediction with a hybrid genomic relationship matrix.

A key application of pooling is that performance data lacking associated pedigree information can still be used. This may be particularly relevant to integrate data from commercial settings, such as that available from feedlots and abattoirs, into routine genomic evaluation.

References

Alexandre P.A., Porto-Neto L.R., Karaman E., Lehnert S.A., and Reverter A. (2019) J Anim Sci 97(12): 4761–4769. https://doi.org/10.1093/jas/skz344

Bell A.M., Henshall J.M., Porto-Neto L.R., Dominik S., McCulloch R. *et al.* Genet Sel Evol (2017) 49:1-7. https://doi.org/10.1186/s12711-017-0303-8

Johnson P.L., O'Connell D., Dodds K.G., Sumner R.M., McEwan J.C. *et al.* (2007) Proc. of the AAABG, Armidale, NSW, Australia.

Madsen P., Jensen J., Labouriau R., Christensen O.F., and Sahana G. (2014) Proc of the 10th WCGALP, Vancouver, BC, Canada.

Matika O., Bishop S.C., Pong-Wong R., Riggio V., and Headon D.J. (2013) Anim Genet 44(6):742-749. https://doi.org/10.1111/age.12070

Park E., and Lee Y.J. (2001) Commun Stat – Simul 30(1):129-142. https://doi.org/10.1081/SAC-100001863

Pollot G.E. (2011) J Anim Sci 89(8):2316-2325. https://doi.org/10.2527/jas.2010-3713

Rupp, R., Mucha S., Larroque H., McEwan J., and Conington J. (2016) Anim. Frontiers 6(1):39-44. https://doi.org/10.2527/af.2016-0006.

Vargas Jurado N., Leymaster K.A., Kuehn L.A., and Lewis R.M. (2016) J Anim Breed Genet 133(5):396-403. https://doi.org/10.1111/jbg.12215

Vargas Jurado N., Kuehn L.A., and Lewis R.M. (2020) J Anim Breed Genet 137(4):365–373. https://doi.org/10.1111/jbg.12449

Vargas Jurado N., Kuehn L.A., Keele J.W., and Lewis R.M. (2021) G3 11(11):1-16. https://doi.org/10.1093/g3journal/jkab231

336. Optimal core definitions for the APY model with an example in large-scale pig dataset

I. Pocrnic[1*], F. Lindgren[2], D. Tolhurst[1], W.O. Herring[3] and G. Gorjanc[1]

[1]The University of Edinburgh, The Roslin Institute, EH25 9RG, Edinburgh, United Kingdom; [2]The University of Edinburgh, School of Mathematics, EH9 3FD, Edinburgh, United Kingdom; [3]Genus PIC, Hendersonville, TN 37075, USA; ivan.pocrnic@roslin.ed.ac.uk

Abstract

Genetic evaluation databases accumulated vast amounts of genomic information that are skyrocketing computational needs for standard genomic evaluation models due to their cubic computational complexity. Several scalable approaches have been proposed, such as the Algorithm for Proven and Young (APY), where genotyped animals are usually randomly partitioned into core and noncore subsets to obtain sparse approximation of the inverse of genomic relationship matrix. Here we show a deterministic optimisation of the core set using a sequential sampling approach. We test this optimisation on a large pig dataset. Our results confirm that the APY is robust and that the size of the core set is critical. Our results further show that the stability of APY can be enhanced with an optimal spread of core animals across a given domain of genotyped animals. We discuss possibilities of an alternative optimisation based on the Nearest Neighbours Gaussian Process that also results in a sparse inverse.

Introduction

In the era of genomics, computational burden is inevitable with the standard genomic evaluation models, predominately due to the cubic cost of inverting the dense genomic relationship matrix. Several scalable genomic approaches have been proposed, such as reparametrizing the models with marker effects, inducing sparsity in the inverse, or dimensionality reduction. The Algorithm for Proven and Young (APY) is one of the approaches, where genotyped animals are usually randomly partitioned into core and noncore subsets. The core animals then form the dense part of the inverse, while the non-core animals are conditionally independent given the core animals and form the diagonal part of the inverse. While the APY provides a good approximation of the full standard model, random partitioning can make results unstable, possibly affecting accuracy or even reranking animals (Misztal *et al.*, 2020). In this contribution we: (1) show application of optimal core construction using conditional sequential sampling algorithm on a large pig dataset; and (2) discuss alternative core construction using the Nearest Neighbours Gaussian Process (NNGP).

Materials & methods

Phenotypic data on 42,868 animals collected from 1999 to 2021, for a moderately heritable trait measured on purebred (33,544), crossbred (114) and backcross (9,210) pigs were provided by PIC (a Genus company, Hendersonville, TN, USA). Genomic information was available for 49,788 pigs, including all animals with phenotypes, and consisted of 42,707 single-nucleotide polymorphism markers. Validation subset included 478 phenotyped and genotyped youngest animals born in 2021 (their phenotypes were removed from the analysis). We used GBLUP in the following model $y = Xb + Za + Wc + e$, where y is a vector of phenotypes, b is a vector of contemporary group fixed effects, a is a vector of random animal effects with $a \sim N(0, G\sigma_a^2)$, c is a vector of random litter effects with $c \sim N(0, I\sigma_c^2)$, e is a vector of random residuals with $e \sim N(0, I\sigma_e^2)$, X, Z, and W are corresponding design matrices, G is the genomic relationship matrix with added I0.01 to enable inversion, and σ_a^2, σ_c^2, and σ_e^2 are respectively known additive, litter and residual variances. The model was run using BLUP90IOD software (Tsuruta *et al.*, 2001), either with the full inverse of G or with the APY inverse of G as defined in Misztal *et al.* (2014). For the APY we applied two core selection strategies: (1) strategy where the core animals were randomly sampled among all the genotyped animals;

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_336, © I. Pocrnic *et al.* 2022

and (2) optimised core selection strategy where we iteratively selected core animals with the largest variance conditionally to previously selected core animals (Pocrnic *et al.*, 2021). For the random strategy, sampling was replicated five times to manifest potential instability of genetic evaluations and we present the mean ±95% CI. The number of core animals was the same in both strategies and was based on the number of eigenvalues that explained from 10 to 99% of the variation in G (Pocrnic *et al.*, 2016). We compared strategies based on the correlations between genomic estimated breeding values (GEBV) obtained with the full inverse and GEBV obtained by each APY strategy. Furthermore, we assessed the predictive ability for the validation animals as correlations between their GEBV and phenotypes adjusted for the fixed effects in the model. To gain insight into the behaviour or the algorithms, we have visualised population structure with a non-linear dimension reduction technique called Uniform Manifold Approximation and Projection (UMAP; McInnes *et al.*, 2020).

Results

The size of the core set in the APY ranged from 8 to 8,348 animals and corresponded to the number of eigenvalues that explained from 10 to 99% of the variation in G. Figure 1 shows that correlations between GEBV obtained with the full and APY inverse of G were almost a linear function of the percentage of explained variation. Both optimised and random core selection methods performed equally well when the number of core animals was equal to the number of eigenvalues explaining more than 98% of variation in G. The difference between the methods was larger in the validation population (Figure 1B), than in the entire genotyped population (Figure 1A). There was more uncertainty with random core sampling when the number of core animals was low, especially in the validation population (Figure 1B).

As expected from previous studies (Pocrnic *et al.*, 2016), the predictive ability was largely determined by the number of core animals (Figure 2), but with no difference between the APY and full inverse when the number of core animals was equal to the number of eigenvalues explaining more than 98% of variation in G. Also, differences between 90 to 99% were minimal. In this sense, the optimised core selection achieved

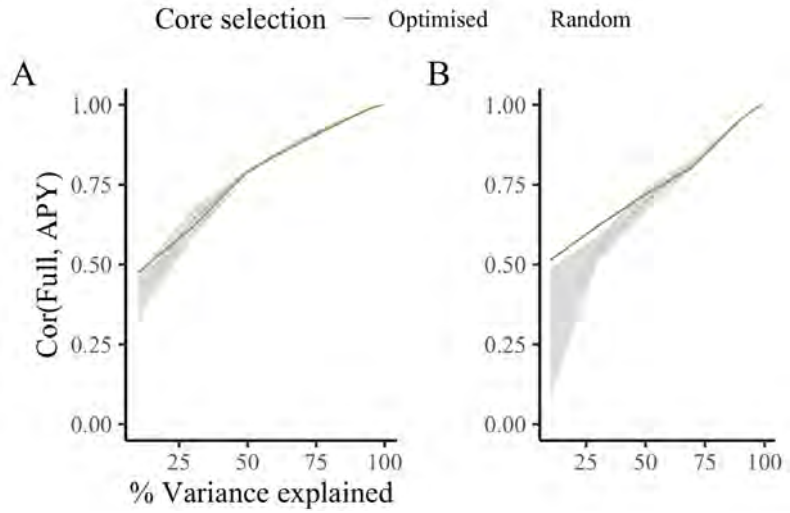


Figure 1. Correlation between GEBV obtained with the full and APY inverse of the genomic relationship matrix for all genotyped (A) and validation animals (B). APY is based on the optimised or random (mean ± 95% CI) core selection, with the number of core animals defined as the number of eigenvalues that explained certain percentage of variance.

the same satisfactory predictive ability as the random selection when the number of core animals was large, but also achieved notable improvements when the number of core animals was smaller. This improvement can be demonstrated by comparison with a large confidence interval in the random sampling strategy for a small number of core animals. This interval reduces as the core gets larger. The apparently higher predictive ability of random core selection with the larger core subsets (>95%) was due to rounding and was always less than 0.005, compared to optimised core selection.

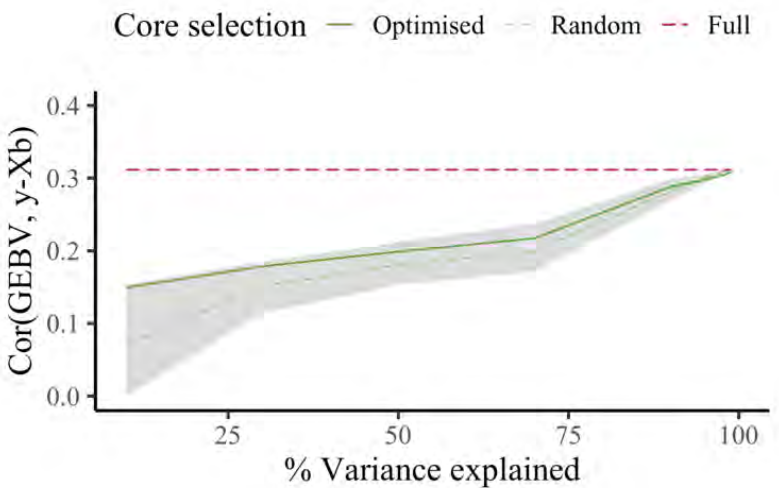


Figure 2. Predictive ability for validation animals. Predictive ability is the correlation between GEBV obtained with the full or APY inverse of the genomic relationship matrix and phenotypes adjusted for fixed effects. APY is based on the optimised or random (mean ± 95% CI) core selection, with the number of core animals defined as the number of eigenvalues that explained certain percentage of variance.

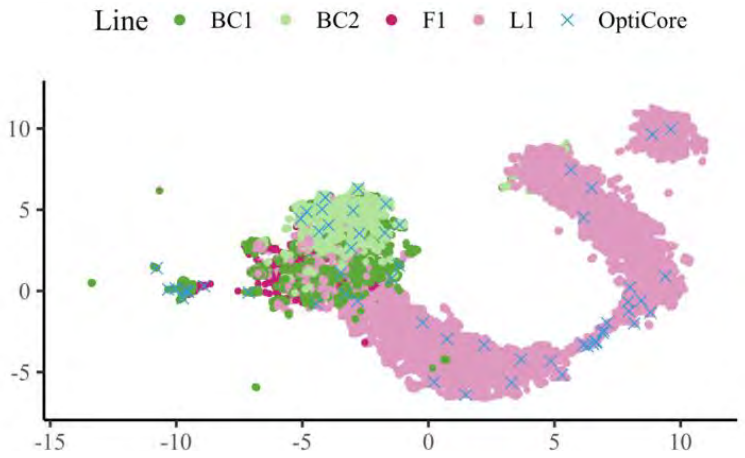


Figure 3. Uniform Manifold Approximation and Projection for purebred (L1), crossbred (F1), and backcross (BC1, BC2) pigs, with first 60 optimally selected core animals.

The UMAP demonstrates several clusters in the population. For example, Figure 3 shows the UMAP for the entire genotyped population, in which purebred animals (L1) cluster into three groups and the majority of crossbreed animals (F1, BC1, and BC2) are closely positioned to one of these clusters. Optimised core set spreads correspondingly.

Discussion

We have demonstrated that the APY is robust, largely dependent on the size of the core set and can be further enhanced by an optimal spread of core animals. While the optimised core selection strategy can introduce stability in the predictive ability (accuracy), especially when the core size is less than optimal, it also introduces additional computing cost compared to random sampling strategy. The optimised strategy builds upon sequential sampling algorithm and results in the spread of core animals far away from each other in a covariance sense. From that viewpoint, the additional computing cost could be justified in several cases, specifically when the genotyped population consists of multiple breeds, lines, and/or crossbred animals. In that sense, the optimised core selection strategy can be further extended towards building a unique core set for each animal based on the NNGP methodology, which recently gained popularity for its ability to deal with large-scale applications in spatial statistics (Datta *et al.*, 2016). While the strategy we presented is depended on application through the APY, the extension into NNGP leads to a sparse inverse through a sparse Cholesky factor (e.g. Faux *et al.* 2012; Finley *et al.*, 2019) and as such could substitute the APY-based inverse matrix when convenient. If n is the total number of genotyped animals, the computational complexity of the full inverse is $O(n^3)$ that is reduced to $O(m^3+2m^2(n-m))$ in the case of the APY or $O(nm^3)$ in the case of the NNGP, where m is either the number of the core animals in APY or the nearest neighbours in NNGP, and assuming $m << n$. On top of the computational complexity for the inversion itself, optimised strategy for APY and NNGP introduce additional computing cost of finding optimal core animals or nearest neighbours. Our future research will investigate faster alternatives for finding optimal set of core animals and nearest neighbours as well as assessing the practical benefit of proposed strategies within complex population structures. Furthermore, we will consider situations in which there is a benefit of using a set of nearest neighbours to form a sparse NNGP inverse relationship matrix rather than being fitted as a core in APY.

References

Datta A., Banerjee S., Finley A.O., and Gelfand A.E. (2016) J. Am. Stat. Assoc. 111:800-812. https://doi.org/10.1080/01621459.2015.1044091

Faux P., Gengler N., and Misztal I. (2012) J. Dairy Sci. 95:6093-6102. https://doi.org/10.3168/jds.2011-5249

Finley A.O., Datta A., Cook B.D., Morton D.C., Andersen H.E. *et al.* (2019) 28:401-414. https://doi.org/10.1080/10618600.2018.1537924

McInnes L., Healy J., and Melville J. (2020) arXiv preprint. https://arxiv.org/abs/1802.03426v3

Misztal I., Legarra A., and Aguilar I. (2014) J. Dairy Sci. 97:3943-3952. https://doi.org/10.3168/jds.2013-7752

Misztal I., Tsuruta S., Pocrnic I., and Lourenco D.A.L. (2020) J. Anim. Sci. 98:1-8. https://doi.org/10.1093/jas/skaa374

Pocrnic I., Lourenco D.A.L., Masuda Y., and Misztal I. (2016) Genet. Sel. Evol. 48:82. https://doi.org/10.1186/s12711-016-0261-6

Pocrnic I., Lindgren F., and Gorjanc G. (2021) Proc. of EAAP, Davos, Switzerland. https://doi.org/10.3920/978-90-8686-918-3

Tsuruta S., Misztal I., and Stranden I. (2001) J. Anim. Sci. 79:1166-1172. https://doi.org/10.2527/2001.7951166x

337. Updating the core animals in the algorithm for proven and young in the American Angus Association national evaluations

A.L.S. Garcia[1*], S. Miller[2], S. Tsuruta[3], D. Lourenco[3], I. Misztal[3], D. Lu[1] and K. Retallick[1]

[1]Angus Genetics Inc., American Angus Association, 3201 Frederick Av., St. Joseph, MO 64506, USA; [2]AGBU, a joint venture between NSW Department of Primary Industries and University of New England, 2351, Armidale, Australia; [3]University of Georgia, 425 River Road, Athens, GA 30602, USA; agarcia@angus.org

Abstract

With the widespread adoption of genotyping by Angus breeders, the American Angus Association has now more than one million animals genotyped. Genomic evaluations are conducted weekly with single-step genomic BLUP (ssGBLUP) using the algorithm for proven and young (APY). A set of high accuracy core animals has been used since the implementation of ssGBLUP in 2017. We investigated changes in genomic EBV (GEBV) and convergence behavior with different sets of core animals. Data from the routine growth genetic evaluation were used, and GEBV from different sets of core animals were compared with the original set defined in 2017. The rank correlation of breeding values was ≥0.99, and the number of rounds was 57% smaller with an updated random core set. Maximum changes in breeding values were at most 1.31 genetic standard deviations of the traits investigated. The core animals in the official evaluations were successfully updated in May of 2021.

Introduction

Single-step is the standard method for beef cattle genetic evaluations in the USA. The American Angus Association (AAA) first released ssGBLUP (Aguilar *et al.*, 2010; Lourenco *et al.*, 2015) genomically enhanced EPD for its national cattle evaluation in July 2017, for this paper we will describe the estimates of animals' genetic merit as GEBV. At the time of the implementation, a total of 350,000 genotyped animals were available, and APY (Misztal *et al.*, 2014a; Misztal, 2016) was used to obtain the inverse of the genomic relationship matrix (G). In APY, the genotyped animals are partitioned into core and non-core sets. A sparse representation of G^{-1} is then possible without directly inverting G, enabling ssGBLUP evaluations for many genotyped animals. For the AAA evaluations, the set of core animals was chosen in 2016/2017 based on trait accuracy and has not changed since.

The adoption of genomic selection has been very successful in the Angus breed in the USA since 2017, and AAA breeders have now genotyped more than one million animals. It took about ten years (2008-2018) to reach 500 thousand genotyped animals but only three more years to reach the one million mark (2018-2021). At such a fast-genotyping pace, one question is whether updating the set of core animals in APY overtime is necessary to accurately represent the whole genotyped population. The follow-up question is about the benefits and potential downsides of making such a change on a national scale. Therefore, our objective was to investigate the impact of updating the core animals in APY ssGBLUP on GEBV and convergence behavior for the AAA evaluations.

Materials & methods

Data and model. A dataset for animals born up to 2019 was extracted from the American Angus Association (AAA) database. Phenotypes were available for birth weight (BW; n= 9,202,115), weaning weight (WW; n= 9,710,750), and post-weaning gain (PWG; n=4,723,244). There were 11,843,831 animals in the pedigree of which, 924,968 were genotyped at 39,773 markers after quality control. We used the same growth model as the AAA official evaluations: $y_t = Xbt + W_1u_t + W_2mat_t + W_3mpe_t + e_t$. Where t refers to traits (BW, WW, and PWG); y is the vector of phenotypes; b_t is the vector of fixed contemporary groups; u_t,

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_337, © A.L.S. Garcia *et al.* 2022

mat_t, and mpe_t are the vectors of random additive direct, maternal, and maternal permanent environmental effects, respectively; e_t is the vector of residuals. The u_t and mat_t were present for BW, u_t, mat_t, and mpe_t for WW and, only u_t for PWG. The matrices X, W_1, W_2 and W_3 are the incidence matrices for the effects in u_t, mat_t, and mpe_t, respectively.

Genetic evaluations at the AAA are performed using ssGBLUP (Aguilar *et al.*, 2010) with APY (Misztal, 2014a; Misztal, 2016). The derivations and the final form of the APY G^{-1} can be found in the cited references. The number of core animals in APY is determined based on the number of eigenvalues explaining 98-99% of the variance in G (Pocrnic *et al.*, 2016).

Core animals in APY. The core animals chosen for the implementation of ssGBLUP in 2017 contain 19,019 high accuracy (HA) animals born up to 2016 (mostly up to 2015). Different sets of randomly sampled core animals were compared to the HA core. To investigate the impact of choosing core animals from different generations, we randomly sampled four sets of 19,019 genotyped animals born up to 2010 (Y1), 2015 (Y2), 2017 (Y3), and 2019 (Y4). It is worth noting that scenario Y2 is a random sample of contemporaries of those animals used in the HA core, and Y4 represents a core sample taken from the whole genotyped population (born up to 2019). To investigate the impact of the number of core animals on GEBV changes, four more scenarios were tested by randomly choosing 10,000 (S1), 25,000 (S2), 30,000 (S3) and 35,000 (S4) core animals from the whole genotyped population. All analyses were performed using the BLUPF90 family of programs (Misztal *et al.*, 2014b). Correlations and absolute changes in GEBV (i.e. changes are expressed in genetic standard deviation (SDa) units) were used as metrics to investigate the impact of changing the core animals. Additionally, the number of rounds to convergence were used to evaluate the convergence behavior with different core animals in APY.

Results

Correlations and changes in GEBV presented here were investigated by contrasting the tested scenarios to the current evaluation (HA). This is because, in practice, critical changes to the current system potentially affect breeders. The GEBV correlations for all animals were ≥0.99, whereas, for genotyped animals, they were ≥0.98 across all scenarios.

Changes in GEBV. Table 1 shows the mean and maximum absolute changes in GEBV across different scenarios. The changes are presented in SDa units to allow easy comparison across traits. The SDa for the traits investigated is 5.38 for BW lbs, 31.41 lbs for WW and 28.08 lbs for PWG. For BW, the biggest change was 1.31 SDa, whereas for WW and PWG the biggest changes were 1.29 and 1.21, respectively. Overall, the largest changes were identified for animals with lower accuracy, genotyped animals, animals that changed from core to noncore and vice versa, and animals with many young genotyped progeny in the core.

Convergence. Table 1 shows the number of rounds to solve the system of equations in ssGBLUP with preconditioned conjugate gradient and iteration on data. The number of rounds to convergence was the largest with the oldest set of core animals (Y1), and rounds decreased with core samples more representative of the whole population (Y4). Memory usage increased with the number of core animals (from 109 GB in S1 to 281 GB in S4), which indicates the availability of restricted memory is a limiting factor when expanding the number of core animals.

Sum of elements of G for core animals (G_{cc}). The sum of all G_{cc} elements is in Table 1. Values were greater for Y1 and HA but decreased as more recent genotyped animals were included in the core. A correlation of 0.81 was observed between the number of rounds to convergence and the sum of G_{cc} for the scenarios based on different generations.

Table 1. Mean and maximum absolute GEBV changes[1], number of rounds to convergence, and statistics of G_{cc} across all scenarios and traits.

Scenario	N	BW		WW		PWG		Rounds	Sum(G_{cc})
		Mean	Max	Mean	Max	Mean	Max		
HA	19,019	-	-	-	-	-	-	3,449	1,008,639
Y1	19,019	0.02	1.31	0.02	1.21	0.02	1.08	4,643	2,662,809
Y2	19,019	0.01	1.05	0.02	1.23	0.01	1.21	3,904	477,087
Y3	19,019	0.01	0.93	0.02	1.29	0.01	1.03	3,120	152,131
Y4	19,019	0.01	1.27	0.02	0.89	0.01	0.86	1,981	24,067
S1	10,000	0.02	1.31	0.03	1.18	0.02	1.07	1,820	11,185
S2	25,000	0.01	0.82	0.02	1.24	0.01	1.07	2,059	37,864
S3	30,000	0.01	1.09	0.02	0.90	0.01	0.97	2,124	43,758
S4	35,000	0.01	1.07	0.02	1.09	0.01	0.77	2,188	59,150

[1] Changes are calculated as the absolute difference in GEBV from the official evaluation (HA) and expressed in SDa units.

Discussion

Changes in the set of core animals only marginally affected the ranking of animals, as the correlations are ≥0.99 regardless of the core choice, as long as the number of core animals explains at least 98% of the variation in G (Misztal *et al.*, 2020; Hidalgo *et al.*, 2021). In beef cattle in the US, even if the ranking of animals remains unchanged, changes in GEBV or expected progeny difference (EPD) are critical because breeders pay close attention to the actual EPD, especially on popular sires.

Changes in GEBV. The largest changes in GEBV observed in this study were slightly over one SDa. Similar findings were reported by Misztal *et al.* (2020) when they investigated core-dependent changes in APY across different populations. The authors reported smaller maximum changes in GEBV compared to our study, but this is likely because in our study, the baseline for comparison is a core set that was created in 2017 (HA). Therefore, the differences between the core sets were larger. Increasing the core size did not reduce the changes in GEBV (Table 1). Misztal *et al.*, 2020 reported that differences are minimized as the core size increases due to the reduction in the Mendelian sampling error in the APY formula. However, this was shown comparing pairs of core sets randomly sampled from the whole population, which was not the comparison made in our study. Expanding the tests to other traits published by the AAA, and with repeated random core samples (results not shown), the magnitude of the changes in GEBV was always similar, with the average absolute changes close to zero and maximum changes around one SDa for any given trait. These changes are not negligible in the context of the AAA evaluation. Still, individual changes were often on lower-accuracy animals or animals with many recent genotyped progeny. This would indicate a set of core animals that is more representative of the whole genotyped population is beneficial for the evaluation.

Convergence. The convergence rate was much improved with an updated sample of core animals. It is clear from Table 1 that randomly sampling core animals from the whole genotyped population yields the best convergence results. For instance, comparing core animals born up to 2010 (Y1) to the set sampled from the whole population (Y4), the number of rounds needed to reach convergence was reduced by 57%, which can dramatically reduce computing time for routine evaluations. Scenarios S1-S4 indicate that a smaller number of core animals results in a smaller number of rounds, as well as less memory usage, but caution is needed as we need to guarantee that a minimum number of core animals is used to explain 98-99% of the variation in G.

The change in convergence behavior seems to be linked with the condition of the blocks that form G and APY G^{-1}. The results in Table 1 show an association between the sum of the elements of G_{cc} and the convergence rate.

The AAA has recently (May 2021) updated the set of core animals used in routine evaluations to better represent the whole genotyped population. Across all genetic evaluations conducted, correlations between GEBV for all animals were always ≥0.99. The core update reduced the number of rounds to convergence by 43%, on average, across all traits. This greatly reduced the computing times of routine genetic evaluations without compromising the quality of the predictions. Despite some large individual changes, the update was well received by the AAA membership, and future research will determine when the next update will be needed.

References

Aguilar I., Misztal I., Johnson D.L., Legarra A., Tsuruta S., *et al.* (2010). J Dairy Sci 93(2):743-752. https://doi.org/10.3168/jds.2009-2730

Hidalgo J., Lourenco D., Tsuruta S., Masuda Y., Miller S., *et al.* (2021). JAS 99(2):1-10. https://doi.org/10.1093/jas/skab004

Lourenco D., Tsuruta S., Fragomeni B.O., Masuda Y., Aguilar I., *et al.* (2015). JAS 93(6):2653-2662. https://doi.org/10.2527/jas.2014-8836

Misztal I., Tsuruta S., Pocrnic I., and Lourenco D. (2020). JAS 98(12):1-8. https://doi.org/10.1093/jas/skaa374

Misztal I. (2016). Genetics 202(2):401-408. https://doi.org/10.1534/genetics.115.182089

Misztal I., Legarra A., and Aguilar I. (2014a). J Dairy Sci. 97(6):3943-52. https://doi.org/10.3168/jds.2013-7752.

Misztal I., Tsuruta S., Lourenco D., Masuda Y., Aguilar I., Legarra A., *et al.* (2014b). Available at: http://nce.ads.uga.edu/wiki/lib/exe/fetch.php?media=blupf90_all2.pdf

Pocrnic I., Lourenco D., Masuda Y., Legarra A., and Misztal, I. (2016). Genetics 203(1):573-581. https://doi.org/10.1534/genetics.116.187013

338. A critical aspect when using APY inversion with Single-Step GBLUP

C. Edel*, E.C.G. Pimentel, R. Emmerling and K.-U. Götz

Bavarian State Research Center for Agriculture (LfL), Institute of Animal Breeding, Professor-Dürrwaechter-Platz 1, 85586 Poing-Grub, Germany; christian.edel@lfl.bayern.de

Abstract

If the algorithm for proven and young (APY) is used to invert the genomic relationship matrix within the framework of Single-Step genomic evaluations, critical side-effects on genomic EBV can occur. Here we demonstrate that if parent animals are not included in the APY-core, their Mendelian sampling deviations are missing in the EBV of their offspring. Additionally, a critical downward bias in genomic predictions for some animals can occur under specific but not uncommon conditions. We illustrate some of the mechanisms involved using a small example and define the circumstances under which a downward bias will occur. Based on this definition, we selectively omit a small group of genotyped animals from the APY-core of the routine evaluation for protein yield in dual-purpose Fleckvieh to demonstrate the resulting effects on their Single-Step predictions.

Introduction

APY was proposed as a method to approximate genomic predictions when the genomic relationship matrix G involved is too large for direct inversion (Misztal *et al.*, 2014). With APY, only a submatrix of G corresponding to a group of core animals is inverted explicitly and the remaining parts of the APY-inverse of G are then derived using recursions. Originally proposed as an 'algorithm for proven and young' with proven animals defined as core, several investigations have suggested that the choice of the core animals is of little relevance, if the core size is sufficiently large (Misztal, 2016; Pocrnic *et al.*, 2016). In April 2021 we established Single-Step methodology using APY as routine-evaluation for all traits in dual-purpose Fleckvieh and Brown-Swiss within the German-Austrian joint evaluation. During the implementation of the new system, we observed that Single-Step proofs of genomic young bulls (genotyped bulls with genotyped offspring but no daughter proofs, GYB) showed remarkable differences depending on whether they were included in the APY-core or not. Further investigation confirmed that proofs of GYB not included in the APY-core tend to be underestimated and that the degree of underestimation increased with the size of their respective offspring group. Since the consistency of proofs of the most widely used genomic young bulls is a key-parameter for the acceptance of genomic predictions in practical breeding programs, this aspect had to be investigated in more detail.

Methods

Example pedigree. To illustrate the mechanisms involved we will refer to a small example pedigree shown in Table 1. Scenarios A and B represent different situations with respect to animals being assumed to be genotyped (gt) and/or member of the APY-core. For each scenario the corresponding combined covariance-matrix H of true breeding values for gt and non-gt animals is calculated (Legarra *et al.*, 2009). To allow for a clearer interpretation, in all these calculations matrix G is replaced by the corresponding submatrix $A_{gt,gt}$ of the standard numerator relationship matrix A, applying APY-recursion depending on the respective scenario. Final matrix H for each scenario is decomposed to give H = TDT', where T is a lower triangular matrix and D is a diagonal matrix of individual Mendelian variances. In this context each row *i* of T illustrates how each breeding value *i* is a function of Mendelian sampling (MS) terms of animals 1:*i*. Likewise, each row *i* of T^{-1} illustrates how each MS-term *i* is a function of breeding values of animals 1:*i*. Multiplication TDT' can be done sequentially by calculating outer products of the respective row/column *i* of T times scalar $d_{i,i}$ and then summing up all partial matrices, thereby illustrating animal *i*'s contribution

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_338, © C. Edel *et al.* 2022

to the final matrix H. The same applies for the inverse $H^{-1} = (T')^{-1}D^{-1}T^{-1}$. In the case of A^{-1} this calculation of partial matrices provides the theoretical background of Henderson's rules for the inverse of the numerator relationship matrix.

Real data. Data from the August 2021 routine-run of the German-Austrian Single-Step evaluation for protein yield in Fleckvieh were used. Total number of animals in the pedigree was 16,864,007, from which 13,417,282 had at least one test-day observation for protein yield. The number of genotyped animals was 370,123. Matrix G^{-1} was set up using APY with 26,724 animals in the core set. Estimated breeding values (EBV) from the routine run were used as reference. We then selected 50 genotyped GYB (GYB50) and omitted them from the APY-core. GYB50 were selected by the following criteria:

- GYB with a large group of genotyped offspring in the evaluation system, that were produced from mating with non-gt females (23-267 matings, average is 66).
- Non-gt females mated to the GYB had to show on average a positive MS (3.0-5.2 kg, average is 3.7 kg). MS-terms used as criteria were calculated from conventional EBV (additional run) without the use of genomics.

To keep the size of the APY-core constant, we replaced the GYB50 by 50 other gt bulls of the same cohort (without genotyped offspring). Finally, we compared the resulting EBV (EBV_{drop}) to the reference (EBV_{ref}). All results shown are EBV on the natural scale (protein-yield, kg). Genetic standard deviation of the trait is 16.4 kg.

Results & discussion

Example pedigree. Rows of matrix T corresponding to animals 5, 7 and 8 (GYB, mate, offspring) for scenarios A and B are given in Table 2. Scenario 'Base' refers to a situation where all animals are gt and in the APY-core. Recall that for ease of interpretation we used submatrix $A_{gt,gt}$ instead of G, so in this case H=A. With sire 5 and dam 7 genotyped but omitted from the APY-core (scenario A), MS-terms of 5 and 7 no longer contribute to the breeding value of offspring 8. If there is no information available to estimate MS-terms for animals 5 and 7 this will have no effect on the estimates of their offspring. However, in a dairy cattle situation, cow 7 will most likely provide information to estimate her MS from her own performance. In this situation, her estimated MS-term will be missing in the parent-average part of EBV of offspring 8. If mate 7 is assumed to be non-gt (scenario B), it is no longer part of the APY-recursion. In scenario B GYB 5 will still not contribute to the BV of animal 8 while the MS of dam 7 will make a contribution, which is slightly higher than expected. Row 7 of Matrix T now additionally includes a negative coefficient of the MS of dam 7 to her mate (GYB 5), contributing to a negative correlation between the MS-terms of animals 5 and 7 in the final matrix. Since GYB 5 has no information to estimate his own MS but mate 7 provides her

Table 1. Example pedigree with two scenarios A and B.

Animal	sire	dam	sex	isGT (A)	isCore (A)	isGT (B)	isCore (B)
1	0	0	m	1	1	1	1
2	0	0	m	1	1	1	1
3	0	0	f	1	1	1	1
4	0	0	f	1	1	1	1
5 (GYB)	1	3	m	1	0	1	0
6	1	4	f	1	1	1	1
7 (mate)	2	4	f	1	0	0	0
8 (offspring)	5	7	m	1	0	1	0

own performance, parts of her MS-estimate will be transferred to GYB 5 but with an opposite sign. This will be the case with every additional non-gt dam mated to the non-core GYB. Assortative mating to a group of superior dams would then result in downward bias in the EBV of the GYB.

As mentioned in the introductory section, initial observations of this phenomenon suggested that there should be an additional effect that scales the bias produced by assortative mating to the number of offspring produced by such matings. Recalling Henderson's rules for setting up A^{-1}, outer products of the rows of matrix T^{-1} (Table 3) will provide the coefficients each animal will contribute to the respective elements of the final inverse. The most prominent aspect here is the deficiency found in the term that offspring 8 provides to the diagonal element of its sire 5 (-0.12^2 instead of -0.50^2). This deficiency is only partly compensated by the non-regular diagonal contribution of dam 7 to the diagonal element of her mate 5 (0.20^2). Since a similar deficiency will be present in every additional offspring's contribution to the final diagonal element of its sire, this will produce an increasing inconsistency with respect to the MS variance assigned to the sire in the final matrix. Thus, the bias in the MS estimate of the GYB, that was produced by assortative mating will increase with the number of such matings.

Real data. Table 4 summarizes the changes observed in routine Single-Step EBV when selectively omitting GYB50 from the APY-core. Results show a downward bias in EBV_{drop} for GYB50 as well as a reduced model fit. In contrast, a systematic bias is mostly absent in the results of all other groups shown. In GYB50 the downward bias in EBV_{drop} is highly correlated to the size of the offspring group of the GYB produced by matings to ngt females as illustrated in Figure 1.

Table 2. Rows of T corresponding to animals 5, 7 and 8 for scenarios Base, A and B.

Scenario	Row	Animal							
		1	2	3	4	5	6	7	8
Base	5	0.50		0.50		1.00			
	7		0.50		0.50			1.00	
	8	0.25	0.25	0.25	0.25	0.50		0.50	1.00
A	5	0.50		0.50		1.00			
	7		0.50		0.50			1.00	
	8	0.25	0.25	0.25	0.25				1.00
B	5	0.50		0.50		1.00			
	7		0.50		0.50	-0.20		1.00	
	8	0.25	0.25	0.25	0.25			0.58	1.00

Table 3. Rows of T^{-1} corresponding to animals 5, 7 and 8 for scenarios Base and B.

Scenario	Row of	Animal							
		1	2	3	4	5	6	7	8
Base	5	-0.50		-0.50		1.00			
	7		-0.50		-0.50			1.00	
	8					-0.50		-0.50	1.00
B	5	-0.50		-0.50		1.00			
	7	-0.10	-0.50	-0.10	-0.50	0.20		1.00	
	8	-0.19	0.04	-0.19	0.04	-0.12		-0.58	1.00

Table 4. Changes in single-step EBV when dropping GYB50 from APY-core.[1]

	N	avDiff	a	b	Fit
GYB50	50	-1.78	-2.79	1.03	0.97
others	370,073	-0.02	-0.01	1.00	1.00
mates	3,116	0.45	0.45	1.00	1.00
offspring	3,276	0.01	0.03	1.00	1.00

[1] Groupsize (N), average difference in EBV (avDiff: EBV_{drop}-EBV_{ref}), Intercept, slope and model-fit (a, b and Fit) of regression of EBV_{drop} on EBV_{ref} for GYB50, all other gt animals in the system (others), the ngt mates of GYB50 (mates) and the gt offspring of GYB50 produced by these matings (offspring).

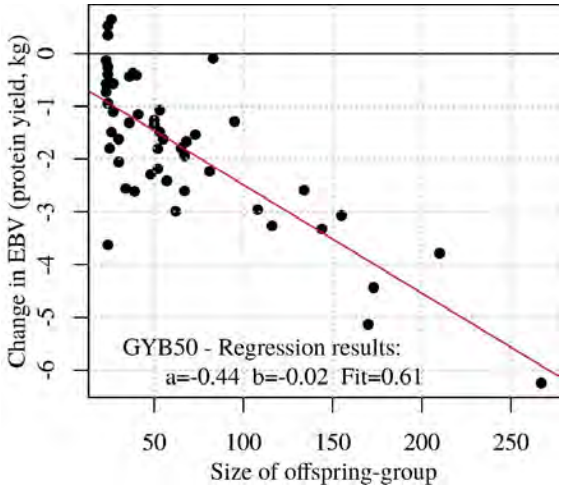


Figure 1. Change in EBV of GYB50.

Conclusions

We have demonstrated that by omitting a gt parent from the APY-core its MS deviation will be absent in the EBV of its offspring. By assigning an APY-core of sufficient size, this deficiency might be partly compensated by the additional information provided by the G matrix. Therefore, in genomic evaluation systems, it might not be as pronounced as illustrated here with the small pedigree example using $A_{gt,gt}$ instead of G. However, if a superior GYB that was systematically mated to superior but non-gt dams is not considered in the APY-core, a systematic downward bias in its EBV will be observed, that will increase with the number of such matings. Since offspring EBV will not be biased accordingly, an increasing discrepancy between average EBV of unselected offspring and their average mid-parent EBV will be the consequence. The situation described here is not uncommon in practical breeding programs where the consistency of proofs of the most popular genomic young bulls is of high importance. Therefore, a random choice of APY-core animals should be avoided, irrespective of additional considerations about the core-size. Although not 'proven' in a strict sense, GYB should be included in the APY-core to prevent strong effects on their genomic EBV.

References

Legarra A., Aguilar I., and Misztal I. (2009) J. Dairy Sci. 92(9):4656-4663. https://doi.org/10.3168/jds.2009-2061

Misztal, I., A. Legarra, and I. Aguilar (2014) J. Dairy Sci. 97(6):3943–3952. https://doi.org/10.3168/jds.2013-7752

Misztal I. (2016) Genetics. 202(2):401-409. https://doi.org/10.1534/genetics.115.182089

Pocrnic, I., Lourenco, D. A. L., Masuda, Y., Legarra, A., and Misztal, I. (2016) Genetics, 203(1):573-581. https://doi.org/10.1534/genetics.116.187013

339. Measures to quantify the accuracy and the erosion of genomic predicted breeding values

B.C.D. Cuyabano[1*], D. Boichard[1] and C. Gondro[2]

[1]*Université Paris Saclay, INRAE, AgroParisTech, GABI, Domaine de Vilvert, 78350 Jouy-en-Josas, France;*
[2]*Department of Animal Science, Michigan State University, 474 S Shaw Ln, East Lansing, MI 48824, USA;*
beatriz.castro-dias-cuyabano@inrae.fr

Abstract

Genomic predicted breeding values (GPBV) to select animals in breeding programs are nowadays routinely adopted by most commercial livestock production systems. Accuracy of the GPBV (defined here as their correlation with the true phenotypes) has a theoretical limit of $\sqrt{h^2}$ that is achieved when, among other factors, available SNPs are in sufficient LD with the QTL. However, even under such assumptions, realized accuracies of GPBV remain generally below $\sqrt{h^2}$, because allele frequencies and LD patterns differ between reference and target populations (particularly noticeable across generations), resulting in the so-called erosion of SNP effects and consequently, erosion of the accuracy of the GPBV. We present here a measure to quantify the erosion of the GPBV's accuracy through the genomic correlation between reference and target populations and validate this measure through simulations.

Introduction

Nowadays, GPBV are routinely used for the selection of animals in many commercial livestock breeding programs. The accuracy of GPBV is therefore a very important factor for the success of a breeding program. However, realized accuracies of GPBV remain below the theoretical maximum ($\sqrt{h^2}$), even when the reference population is sufficiently large, and SNPs included in the model are in sufficient LD with the QTL. That is particularly noticeable over generations, as we observe the so-called erosion of SNP effects accompanied by the erosion of the GPBV's accuracy. Erosion occurs mostly due to differences in LD patterns and allele frequencies between reference and target populations; for example, if in the reference population a SNP is in strong LD with the QTL, a large effect will be assigned to it. However, if due to segregation over the generations, the LD between this SNP and the QTL becomes weaker in the target population, an effect closer to zero should be assigned to this SNP. The decay in prediction accuracy due to differences in allele frequencies and LD patterns, specially over generations, is a topic widely known and discussed by animal breeders and quantitative geneticists (Daetwyler *et al.*, 2008; Habier *et al.* 2013; Wientjes *et al.*, 2015, 2016; Pszczola and Calus, 2016; van den Berg *et al.*, 2019; Dekkers *et al.*, 2021). Quantifying the erosion at the individual SNP level is in fact, a difficult and unresolved task. It is, however, more tractable to quantify the erosion of the accuracy of the GPBV through a metric based on the genomic correlation between reference and target populations, which we present in this paper and validate through simulations.

Materials & methods

Consider the genomic model $y_1 = X_1\beta + M_1\alpha + \varepsilon_1$, in which y_1 are the phenotypes measured in the reference population, β are the fixed effects and X_1 their design matrix in the reference population, M_1 is the (centred) SNP-genotypes matrix of the reference population, $\boldsymbol{\alpha} \sim N(\mathbf{0}, \boldsymbol{I}_m\sigma^2_\alpha)$ are the SNP effects, and $\boldsymbol{\varepsilon}_1 \sim N(\mathbf{0}, \boldsymbol{I}_n\sigma^2_\varepsilon)$ are the random residuals. Solving Henderson's mixed model equations (HMME) (Henderson *et al.*, 1959), we have the estimates $\widehat{\boldsymbol{\beta}} = [\boldsymbol{X}_1'(\boldsymbol{M}_1\boldsymbol{M}_1'\sigma^2_\alpha + \boldsymbol{I}_n\sigma^2_\varepsilon)^{-1}\boldsymbol{X}_1]^{-1}\boldsymbol{X}_1'(\boldsymbol{M}_1\boldsymbol{M}_1'\sigma^2_\alpha + \boldsymbol{I}_n\sigma^2_\varepsilon)^{-1}\boldsymbol{y}_1$ and $\widehat{\boldsymbol{\alpha}} = \boldsymbol{M}_1'(\boldsymbol{M}_1\boldsymbol{M}_1'\sigma^2_\alpha + \boldsymbol{I}_n\sigma^2_\varepsilon)^{-1}\boldsymbol{y}_1^*\sigma^2_\alpha$, such that $\boldsymbol{y}_1^* = \boldsymbol{y}_1 - \boldsymbol{X}_1\widehat{\boldsymbol{\beta}}$. Finally, the fitted breeding values for the n_1 animals in the reference population are $\widehat{\boldsymbol{g}}_1 = \boldsymbol{M}_1\widehat{\boldsymbol{\alpha}}$, and let M_2 be the (centred) SNP-genotypes matrix of the target population, the GPBV for the n_2 animals in the target population are $\widetilde{\boldsymbol{g}}_2 = \boldsymbol{M}_2\widehat{\boldsymbol{\alpha}}$.

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_339, © B.C.D. Cuyabano *et al.* 2022

Theoretical limit of GPBV's accuracy without accounting for erosion. Our interest lies on the accuracy of the GPBV, i.e. on $R = \widehat{cor}(\tilde{\boldsymbol{g}}_2, \boldsymbol{y}_2^*)$, such that $\boldsymbol{y}_2^* = \boldsymbol{y}_2 - \boldsymbol{X}_2\hat{\boldsymbol{\beta}}$. Using Fisher's z-transformation on R, we have that $Z = log\left(\frac{1+R}{1-R}\right) \sim N\left(log\left(\frac{1+\rho}{1-\rho}\right), \frac{4}{n_2-3}\right)$, with ρ being the true accuracy, i.e. $\rho = \sqrt{h^2}$.

By reversing Fisher's z-transformation $R = \varphi(Z) = \frac{e^Z-1}{e^Z+1}$ is a function of Z, and using Jensen's inequality we show that:

$$E[R|\text{no erosion}] = E[\varphi(Z)|\text{no erosion}] \leq \varphi(E[Z]|\text{no erosion}) = \frac{e^{E[Z]}-1}{e^{E[Z]}+1} =$$
$$\left[e^{log\left(\frac{1+\rho}{1-\rho}\right)} - 1\right] \Big/ \left[e^{log\left(\frac{1+\rho}{1-\rho}\right)} + 1\right] = \rho = \sqrt{h^2}. \quad .$$

Theoretical limit of GPBV's accuracy accounting for erosion. We hypothesized that the genomic correlation between reference and target populations (r) affects linearly the mean of the z-transformed correlations, resulting in $Z = log\left(\frac{1+R}{1-R}\right) \sim N\left(r\ log\left(\frac{1+\rho}{1-\rho}\right), \frac{4}{n_2-3}\right)$.

Using Jensen's inequality again we show that, under our hypothesis:

$$E[R|\text{erosion}] \leq \frac{\left(1+\sqrt{h^2}\right)^r - \left(1-\sqrt{h^2}\right)^r}{\left(1+\sqrt{h^2}\right)^r + \left(1-\sqrt{h^2}\right)^r} \underset{r\to 1}{\longrightarrow} \sqrt{h^2}.$$

Erosion of the GPBV. We consider the erosion of the GPBV (δ_{GPBV}) as the difference between $\sqrt{h^2}$ and the theoretical limit of GPBV's accuracy accounting for erosion, scaled by $\sqrt{h^2}$,

$$\delta_{GPBV} = 1 - \frac{E[R|\text{erosion}]}{\sqrt{h^2}} \quad (1)$$

Genomic correlation between populations. Let M be a (centred) SNP-genotypes matrix ($n \times m$), with singular-value decomposition (SVD) $\left(\sum_{j=1}^{m} 2p_j(1-p_j)\right)^{-1/2}\boldsymbol{M} = \boldsymbol{UDV'}$, such that p_j's are the alleles frequencies. In this SVD, D is a diagonal matrix of the $k = rank\ (M)$ singular-values, and $U_{n\times k}$ and $V_{m\times k}$ are matrices of unitary eigen-vectors relative to the individuals and to the SNPs, respectively. Knowing that components in D^2 explain the variation of the whole system M, each U_{ik} explains the contribution of the i-th individual to the k-th component, and each V_{jk} explains the contribution of the j-th SNP to the k-th component. To obtain the genomic correlation between reference and target populations (r), we need to quantify the different contribution of the SNPs to the system's variation in the two populations. To do so, we perform the aforementioned SVD on M_1 and M_2, then build a matrix $\boldsymbol{T} = \sqrt{(n_2/n_1)}\boldsymbol{V}_2'\boldsymbol{V}_1\boldsymbol{D}_1$ which correlates the contributions of the SNPs in both populations, while correcting for the different population sizes, weighting these correlations by the singular-values on the reference population. Next, we obtain the SVD $T = U_T D_T V_T$, and perform the linear regression $D_T \sim D_2$ with a quadratic term, i.e. we fit $d_{Ti} = a + bd_{2i} + cd_{2i}^2$. Finally, we obtain the genomic correlation between reference and target populations as $r = a + b + c$.

Simulation study. We simulated 50K SNPs and additive phenotypes for h^2 = 0.05,0.15,…,0.9,0.95, using 2K SNPs as QTL. A base reference population of 5,000 individuals was used to estimate variance components using REML (Patterson and Thompson, 1971) and SNP effects â (solving HMME). Using â we obtained the GPBV for three different target populations (1000 individuals each) with increasing generation distances (one, five, and ten) from the reference population. Then we compared our theoretical measures $E[R|\text{erosion}]$ and δ_{GPBV} with the realized prediction accuracies ($R = \widehat{cor}(\tilde{\boldsymbol{g}}_2, \boldsymbol{y}_2^*)$) and erosion ($\hat{\delta} = 1 - R/\sqrt{h^2}$).

Results

In Figure 1 we observe that as the generation distance between reference and target populations increases, the relationship $d_{2i} \times d_{Ti}$ moves further from the identity line and becomes more quadratic, resulting in

decreasing genomic correlations as the generation distances increase: r = 0.49, r = 0.42 and r = 0.33 for target populations respectively one, five and ten generations apart from the reference population.

In Figures 2 and 3 we observe that our theoretical curve $E[R|\text{erosion}]$ accurately fits the realized $\hat{h}^2$ and $R = \widehat{cor}(\tilde{\boldsymbol{g}}_2, \boldsymbol{y}_2^*)$, and that δ_{GPBV} is quite accurate to assess the erosion. For extreme h^2 (<0.1 and >0.85) however, there is still a small over expectation for R, which results in δ_{GPBV} being lower than the observed $\hat{\delta}$. Low heritability traits (h^2<0.2) also present a large $Var(\delta)$, indicating that precise assessment of erosion may be difficult for such traits.

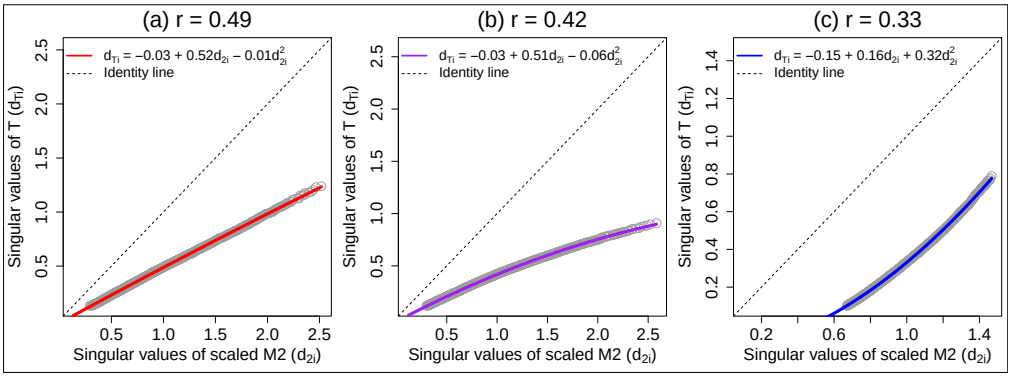


Figure 1. Singular values of the scaled M_2 (d_{2i}) vs singular values of $\boldsymbol{T} = \sqrt{(\boldsymbol{n_2}/\boldsymbol{n_1})}\boldsymbol{V_2'V_1D_1}$ (d_{Ti}), and estimated coefficients fitting the model $d_{Ti} = a + \text{bd}_{2i} + cd_{2i}^2$ for the target populations (a) one, (b) five, and (c) ten generations apart from the reference population.

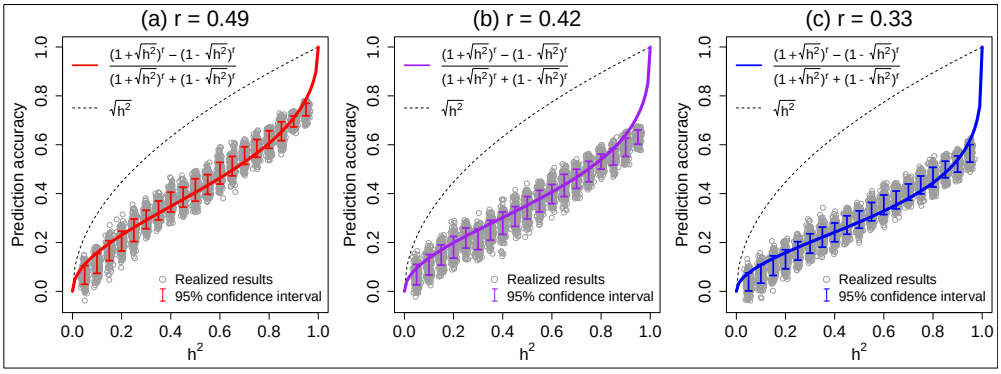


Figure 2. Theoretical $E[R|\text{erosion}]$ (full coloured line) and $E[R|\text{no erosion}]$ (dashed line), and the realized results for the target populations (a) one, (b) five, and (c) ten generations apart from the reference population.

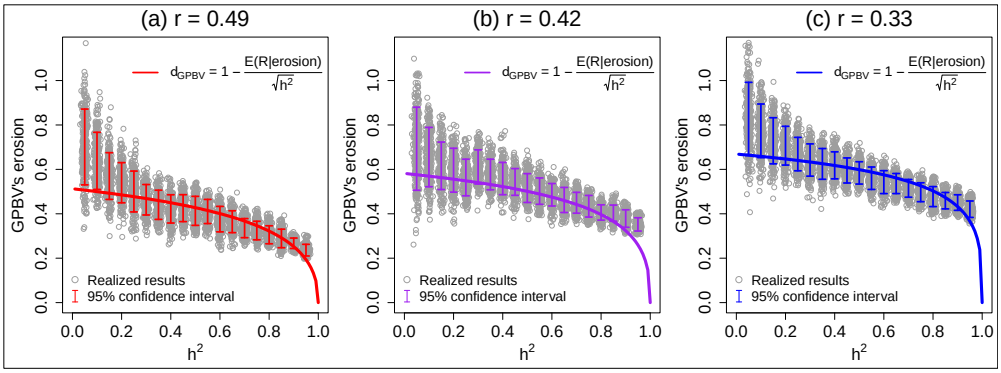


Figure 3. Theoretical erosion (δ_{GPBV}) and realized results for the target populations (a) one, (b) five, and (c) ten generations apart from the reference population.

Discussion

We hypothesized that once we know the genomic correlation between reference and target populations (r), we can define the maximum accuracy of the GPBV as

$$E[R|\text{erosion}] \leq \frac{(1+\sqrt{h^2})^r-(1-\sqrt{h^2})^r}{(1+\sqrt{h^2})^r+(1-\sqrt{h^2})^r} \underset{r\to 1}{\longrightarrow} \sqrt{h^2},$$

and the results obtained with our simulations support this hypothesis. The measure we defined to quantify the erosion of the accuracy of the GPBV ($\delta_{GPBV} = 1 - E[R|\text{erosion}]/\sqrt{h^2}$) was also quite accurate, although it may be imprecise or underestimated for low heritability traits (h^2<0.2). One important element for calculating $E[R|\text{erosion}]$ and δ_{GPBV} is the genomic correlation r, a single value capable to summarize all the information about the differences in allele frequencies and LD patterns observed in both the reference and target populations. Although computationally costly for large populations and dense genotype data, our results indicate that the genomic correlation r obtained with the SVDs of genomic matrices is trustworthy. This work focused on the calculus of $E[R|\text{erosion}]$ and δ_{GPBV}, and did not explore the underlying meaning of the values of the coefficients a, b, and c that compose r, i.e. how allele frequencies, LD patterns, number of SNPs and population sizes affect each of these coefficients, but we do understand that such a study is relevant.

References

Daetwyler H.D., Villanueva B. and Wooliams J.A. (2008) PloS One 3(10):e3395. https://doi.org/10.1371/journal.pone.0003395.

Dekkers J., Su H. and Cheng J. (2021) Genet Sel Evol 53:55. https://doi.org/10.1186/s12711-021-00647-w.

Habier D., Fernando R.L. and Garrick D.J. (2013) Genetics 194(3):597-607. https://doi.org/10.1534/genetics.113.152207.

Henderson C.R., Kempthorne O., Searle S.R. and von Krosigk C.M. (1959) Biometrics 15(2):192. https://doi.org/10.2307/2527669.

Patterson H. and Thompson R. (1971) Biometrika 58(3):545. https://doi.org/10.2307/2334389.

Pszczola M. and Calus M.P.L. (2016) Animal 10(6):1018-1024. https://doi.org/10.1017/S1751731115002785.

Van den Berg I., Meuwissen T.H.E., MacLeod I.M. and Goddard M.E. (2019) J Dairy Sci. 102(4):3155-3174. https://doi.org/10.3168/jds.2018-15231.

Wientjes Y.C.J., Veerkamp R.F., Bijma P., Bovenhuis H., Schrooten, C. *et al.* (2015) Gen Sel Evol 47(1):5. https://doi.org/10.1186/s12711-014-0086-0.

Wientjes Y.C.J., Bijma P., Veerkamp R.F. and Calus M.P.L. (2016) Genetics 202(2):799-823. https://doi.org/10.1534/genetics.115.183269.

340. Model building pipeline to maximize the accuracy of breeding values prediction

J. Jenko[1*] and Ø. Nordbø[2]

[1]Geno SA, Storhamargata 44, 2317 Hamar, Norway; [2]Norsvin R&D, Storhamargata 44, 2317 Hamar, Norway; janez.jenko@geno.no

Abstract

To maximize the accuracy of selection, models used for the prediction of breeding values need to correct phenotypes for all known important environmental effects. We have developed a pipeline where a fixed part of the model is built automatically, and the proposed models are validated in terms of accuracy and bias of their predictions. After the best model for a single trait is selected, genetic correlations are estimated between all the combination of traits and the optimal multitrait models are proposed. We have tested the model building pipeline on the development of single step genomic predictions for a conformation data set with repeated records on the Norwegian Red dairy cattle breed. The comparison between the single step genomic predictions using the old multitrait models and the new multitrait models averaged over nineteen traits showed a relative increase of 26% (from 0.62 to 0.78) in the accuracy of predictions.

Introduction

Accuracy of selection is among selection intensity, genetic variation, and generation interval one of the four components influencing genetic improvement of a population (Falconer, 1996). For a trait with a single record per individual the accuracy of prediction is a square root of heritability. When multiple measurements are taken on the animal for a single trait, the breeding value is predicted from the mean of these records. Repeated records enable estimation of resemblance between records due to the permanent environmental factors (Mrode, 2014). This is decreasing the variance within the individual which leads to an increase in the accuracy of prediction. The gain in accuracy of prediction is increasing with more records, but the marginal benefit in precision is decreasing.

Before variance component estimation and prediction of breeding values, phenotypes have to be corrected for all known important effects. The decision whether the effect is included into the model is often based on F-test where the simple and alternative models are compared. Other criteria like coefficient of determination and adjusted coefficient of determination can be used as well. Finally, models can be validated for their prediction accuracy.

Conformation traits have been part of the breeding goal from the very beginning of the breeding program for Norwegian Red dairy cattle in the nineteen sixties. The importance of conformation traits in breeding goal has increased from around 10% in the beginning to almost 30% in the 2020. Back in the history phenotypes for exterior traits in Norwegian Red dairy cattle were only collected on the first lactating cows. However, since 2014, repeated recording of conformation traits on cows in the first five lactations were collected and the starting point of the current work was to utilize the repeated records into the routine genomic evaluations. To increase the accuracy of predictions furthermore, new multitrait models will be built to utilize information from genetically correlated traits.

Materials & methods

Data. The data comprised of twenty-four traits collected between 1988 and 2021. There were 646,291 records collected on 585,306 cows, from which 16,924 were genotyped. Only cows in the first five lactations were recorded. At the time of recording the majority were in the first lactation (545,518) followed by

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_340, © J. Jenko and Ø. Nordbø 2022

the cows in the second (59,827), third (32,156), fourth (6,006), and fifth (2,784) lactation. The data were recorded on 22,036 herds. There were seven traits that were standardized since they have changed the scale of recording from 1-3 to 1-9 in 2014. To keep the phenotypic variance constant, phenotypes recorded in 2014 or later were standardized for heterogenous variance due to the lactation number. Except stature and supernumerary teats all the traits were also standardized for heterogenous variance due to the classifier and the year of taking the scoring.

Model development and estimation of variance components. To build the fixed part of the model and estimate the variance components, only data recorded after 2011 from herds with at least 100 recordings for stature were used. Furthermore, only data recorded between 14 and 305 days in lactation were kept. This gave 68,227 records collected from 440 herds. Fixed part of the model was built with an inductive approach. Here, the effects were included in the fixed part of the model only if they have passed the inclusion criterion for a specific scenario. Three different inclusion criteria were set for the four optimization parameters. The four optimization parameters were: (1) R square or the proportion of the variance explained; (2) Adjusted R square, which in comparison to R square penalizes for addition of extra variables; (3) *P*-value that compares the reference and alternative model using the F-test; and (4) Gain by loss which is the relative change in residual sum of squares between the reference and alternative model per change in degrees of freedom between the reference and alternative model. Each of the optimization criteria was tested with three inclusion criteria. R square and adjusted R square with 0.1, 0.01, and 0.001, *P*-value with 0.01, 1×10^{-5} and 1×10^{-10}, and Gain by loss with 0.01, 0.001, and 1×10^{-4}. If the tested effect passed the inclusion criterion for the optimization parameter and its effect was the biggest among all the effects, it was included in the model. By default, each model had already included the fixed effects of herd-five-years and year-month of calving. The inclusion of the other twelve candidate effects (number of days since calving, barn type, recording version, feeding system, milking system, classifier, lactation number, time since milking within time from calving, heterozygosity, year of recording, hours since milking, and age at calving) was tested in a sequential order so that the most important effect was included first. If a linear covariate was included in the model, the squared covariate was also tested. The same was done for cubic covariate if the squared covariate was included in the model. Once none of the effect was passing the inclusion criterion, the model building was finished. Finally, two and three-way interactions were tested and included in the model if passing the inclusion criterion for the optimization parameter. With this approach twelve models were developed for each of the twenty-four traits and variance components were estimated using DMU 6 (Madsen and Jensen, 2013).

Cross-validation of alternative models. The proposed models were then tested for the accuracy of prediction using a cross-validation approach. Here 269,114 records collected on 226,485 cows after the year 2005 were used. First estimated breeding values (EBV) using a single step genomic approach on this complete data set were calculated with the DMU 6 (Legarra *et al.*, 2014; Madsen and Jensen, 2013). Then, phenotypes of 6,841 cows that were genotyped and were not included in the variance component estimation were masked and EBV using the single step genomic approach were calculated again. Finally, yield deviations (YD) were estimated from the complete data set and correlated with the EBV from the data set with masked phenotypes. This was done on the animals with masked phenotypes only. The model with the highest accuracy and the lowest bias was then selected for the construction of multitrait models.

Estimation of variance-covariance parameters and construction of multitrait models. The variance-covariance parameters were estimated on the bivariate models across all the twenty-four traits. Traits with high genetic correlations were then combined into six multitrait models. Some traits were included in more than one multitrait model to increase the accuracy of prediction for the traits in a particular multitrait model.

Cross-validation of final multitrait models. The accuracy of predictions of the newly developed multitrait models were finally compared with the accuracy of predictions from the old multitrait models. This was done with masking of the phenotypes from the youngest 5,000 genotyped cows. The YD calculated from the complete data set and EBV from the masked data set were calculated using the MiX99 software (MiX99 Development Team, 2019). The obtained YD and EBV were correlated (Cor) and corrected for the heritability to estimate the accuracy of prediction (Acc) using the following equation:

$$Acc = Cor(YD,EBV) / (n /(n + (\sigma_e / \sigma_a)))^{0.5} \quad (1)$$

where n is the mean number of observations for each animal with masked phenotype, σ_e is the error variance and σ_a is the additive variance. Error and additive variances estimated from the new models were used when estimating the accuracy for both old and new models.

Results

The heritability and prediction accuracy from old and new multitrait models are presented in Table 1. Out of twenty-four traits breeding values for nineteen traits were predicted with both old and new models.

Most of the traits have moderate heritability spanning from 0.14 to 0.31. Only teat length and stature have high heritability of 0.41 and 0.58, respectively. The prediction accuracy has increased for all the traits when the new multitrait models were used. Average accuracy when the old multitrait models were used for predictions was 0.62 where a minimum accuracy of 0.45 was achieved for rear legs rear view and a maximum accuracy of 0.72 was achieved for stature. Average accuracy when the new multitrait models were used for predictions was 0.79 where a minimum accuracy of 0.60 was achieved for rear legs rear view

Table 1. Prediction accuracy for the conformation traits of Norwegian Red dairy cattle with previous and current multitrait models (old/new multitrait model).

Trait	Previous heritability	Current heritability	Accuracy previous model	Accuracy current model
Stature (1./1.)	0.41	0.58	0.72	0.89
Rump angle (1./2.)	0.19	0.28	0.64	0.87
Body depth (1./3.)	0.14	0.19	0.53	0.77
Chest width (1./3.)	0.11	0.15	0.53	0.74
Rump width (1./3.)	0.22	0.30	0.61	0.80
Rear legs rear view (2./1.)	0.08	0.14	0.45	0.60
Rear legs side view (2./1.)	0.09	0.17	0.62	0.70
Foot angle (2./1.)	0.09	0.14	0.64	0.71
Udder balance (3./4.)	0.11	0.17	0.68	0.85
Rear udder width (3./4.)	0.13	0.17	0.71	0.81
Rear udder height (3./4.)	0.11	0.16	0.70	0.83
Rear teat placement (4./5.)	0.22	0.26	0.69	0.83
Fore udder attachment (4./4.)	0.12	0.17	0.51	0.63
Front teat placement (4./5.)	0.18	0.27	0.70	0.79
Central ligament (5./5.)	0.10	0.14	0.54	0.64
Supernumerary teat (5./6.)	0.21	0.17	0.56	0.95
Teat thickness (5./6.)	0.20	0.29	0.64	0.82
Teat length (5./6.)	0.27	0.41	0.71	0.86
Udder depth (5./4.)	0.25	0.31	0.68	0.83

and a maximum accuracy of 0.95 was achieved for supernumerary teat. This means that on average the new multitrait models increase the prediction accuracy for 26% (from 0.62 to 0.79). Correlations between the predicted breeding values from old and new models were from 0.78 to 0.97.

Discussion

Three main reasons for the increase in the accuracy of predictions with the new models exist. The first one is the inclusion of phenotypes from multiple lactations instead of only the first one. As mentioned earlier this enables estimation of resemblance between records due to the permanent environmental factors and decreases the variance within the individual which leads to an increase in the accuracy of predictions (Mrode, 2014). The second reason is a better correction of phenotypes for all known environmental effects. Old models were all correcting traits for the same fixed effects. Model development pipeline corrected phenotypes only for the important known effects and different fixed effects were included in the different models. Using the cross-validations we have compared the proposed models and selected the model that is maximising prediction accuracy and minimising bias. The third reason is the optimal construction of multitrait models. The old models were built so that trait describing the same part of the cow exterior were combined into a multitrait model. The new multitrait models were constructed so that traits with high genetic correlations were combined into a single multitrait model.

The new multitrait models for the prediction of breeding values for exterior traits were sent for Interbull test in August 2021. After their approval they were successfully implemented in the routine genetic evaluations for the Norwegian Red dairy cattle in December 2021. Given the improved accuracy of genomic predictions with new, updated models, we expect to see an increased genetic gain in the coming years.

Acknowledgements

We want to thank the Research Council of Norway for funding this research through the project 309611, 'Large scale single step genomic selection in practice' and under the BIONÆR program, project number 282252.

References

Falconer D.S., and Mackay T.F.C. (1996) Introduction to quantitative genetics. Fourth edition. Longman, Harlow, UK.

Legarra A., Christensen O.F., Aguilar I., and Misztal I. (2014) Livestock Science 166:54-65. https://doi.org/10.1016/j.livsci.2014.04.029

Madsen P., and Jensen J. (2013) DMU Ver. 6, rel. 5.2. Available at: https://www.researchgate.net/publication/291444592_A_user's_guide_to_DMU.ls

MiX99 Development Team (2019) MiX99: A software package for solving large mixed model equations. Release XI/2019. Available at: http://www.luke.fi/mix99

Mrode R.A. (2014) Linear models for the prediction of animals breeding values. 3rd edition. CABI, Wallingford, UK.

341. Genomic prediction in multi-breed Irish beef cattle population using a breed origin of alleles model

A. Guillenea[1*], M.S. Lund[1], R. Evans[2], V. Boerner[1] and E. Karaman[1]

[1]*Center for Quantitative Genetics and Genomics, Aarhus University, 8830 Tjele, Denmark;*
[2]*Irish Cattle Breeding Federation, Bandon, Co. Cork, Ireland; ana.guillenea@qgg.au.dk*

Abstract

The objective of this study was to evaluate genomic prediction using a breed origin of alleles (BOA) model for a multi-breed Irish beef cattle population. Our BOA model uses one matrix with allele counts for each breed to predict breed-specific marker effects, enabling to use crossbred and purebred animals in the reference population. Pedigree-based best linear unbiased prediction (PBLUP), a genomic model assuming a homogeneous population (SNP-BLUP) and BOA model were compared. Accuracy was estimated as the correlation between breeding values and corrected phenotype divided by the square root of heritability. The accuracy of predictions using BOA increased by 82 and 42% for crossbred and purebred animals, respectively, compared with PBLUP, and improvements of BOA over SNP-BLUP were of 6% for crossbred animals.

Introduction

Genomic prediction (GP) has been rapidly implemented in cattle due to its ability to accurately predict the genomic values of selection candidates early in life. Accuracy of genomic prediction relies on the linkage disequilibrium (LD) between genetic markers, usually single nucleotide polymorphisms (SNPs), and causal mutations. The stronger the LD, the more accurately will be the prediction of the direct genomic breeding values (DGVs). The effects of causal mutations captured by the SNPs is not consistent across distantly related populations due to differences in LD patterns. Additionally, some causal mutations may not segregate in all breeds, or the allelic effects of causal mutations may differ in different breeds due to epistasis and differences in allele frequencies (Goddard *et al.*, 2018). Therefore, multiple breed evaluation might benefit from a common multi-breed reference population when the models consider the differences in LD and/or marker effects among the involved breeds. The breed origin of alleles BOA model (Karaman *et al.*, 2021) predicts breeding values based on breed-specific marker effects estimated with purebred and crossbred data, allowing for accounting for correlations between the SNP effects of the different breeds. The application of BOA model can be advantageous in multi-breed populations to the extent that the breeds are distantly related. Otherwise, a joint analysis of all available data can be highly competitive. The Irish beef cattle population is a multi-breed population involving around 40 breeds, with some of them being dairy breeds. Therefore, it is expected that accounting for breed-specific marker effect would be more accurate than assuming homogeneous marker effects across the breeds. The objective of this study was to compare accuracy of predictions from a pedigree-based model (PBLUP) or using genomic information in a joint analysis (SNP-BLUP) or BOA model for the Irish beef cattle population.

Materials & methods

Data. The data used in the present study belongs to an Irish cattle multi-breed population (ICBF, http://www.icbf.com). We kept as main breeds the four pure breeds which contributed the most to the crossbred genotyped animals in the population. The pure breeds were Limousin (LM), Charolais (CH), Angus (AA) and Holstein (HO). For BOA estimation, as a fifth group, we kept five breeds that followed the main and treated them as an 'others' (OT) group. We selected genotyped crossbred animals for estimating BOA with more than 85% of the four main breeds (166,925 animals). Carcass weight (kg) data were available on 10,159,493 animals born between 2000 and 2020 slaughtered between 300 and 1,277 days of age. Purebred

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_341, © A. Guillenea *et al.* 2022

animals from minor breeds and animals without registered sire were removed, resulting in 9,947,707 phenotypic records. After quality control, 46,983 SNPs remained available. The pedigree contained 16,120,121 animals born between 1950 and 2021.

Breed origin estimation. We used the AllOr method (Eiríksson *et al.* 2021) to infer breed origin to each SNP allele. We included a donor genotype library with purebred animals of main breeds and OT, and we also included a pedigree file obtained by tracing the genotyped crossbreds back. On average, 96.95% of the alleles were assigned a breed of origin. The alleles not assigned to a particular breed were defined as coming from the OT group.

Prediction of breeding values. The phenotypes for the whole population were pre-corrected for the following non-genetic effects (*fixed*: birth year; type of birth; factory of slaughter; and age at slaughter linear, quadratic and cubic; *random*: contemporary group of herd of slaughter (CG1) and prior to slaughter; and dam permanent environment). Then, we selected a subset (as explained later in the validation subsection) to run PBLUP, SNP-BLUP and BOA models. For PBLUP we traced three generations back the pedigree of animals in the subset.

Models. PBLUP and SNP-BLUP models

$$y^* = 1u + Xb + Zu + e \quad (1)$$

where y* is a vector of adjusted phenotypes of all animals, 1 is a vector of 1s, μ is the overall mean, X is the matrix of breed proportions computed from BOA assignments, b is the vector of fixed breed effects, and e is the vector of random residuals. Random residuals were assumed to follow a normal distribution, $e \sim N(0,\sigma^2_e)$, where σ^2_e is the residual variance. Z and u changes depending on the model. For PBLUP u is a vector of EBVs, distributed as $u \sim N(0,A\sigma^2_g)$, where σ^2_g is the additive variance, and A is the numerator relationship matrix, Z relates phenotypes to animals. For the SNP-BLUP model, u is the vector of SNP effects and Z is the matrix of centered genotypes based on current allele frequency.

BOA model

$$y^* = 1u + Xb + Z_{CH}u_{CH} + Z_{LM}u_{LM} + Z_{AA}u_{AA} + Z_{HO}u_{HO} + Z_{OT}u_{OT} + e \quad (2)$$

where y*,1, μ, X, b, and e were as described in Equation 1, Z_{CH}, Z_{LM}, Z_{AA}, Z_{HO}, and Z_{OT} are the matrices of breed-specific content of SNP for CH, LM, AA, HO and OT, u_{CH}, u_{LM}, u_{AA}, u_{HO} and u_{OT} are vectors of SNP effects for CH, LM, AA, HO, and OT, respectively. The Z matrices were formed by assigning each allele to a certain pure population for which the expected probability is the highest. The entry at a locus in an Z matrix, for instance, Z_{CH}, were the number of '*A*' alleles (0, 1 or 2) originating from CH for an animal. Consequently, if an animal had '*aa*' genotype or had no allele originating from CH, the corresponding entry in Z_{CH} was zero. The Z matrices were centered prior to analysis. The analysis assumed correlations between SNP effects of the different breeds as follows: a multivariate normal distribution was assigned for the vectors of SNP effects: $[u'_{CH}, u'_{LM}, u'_{AA}, u'_{HO}, u'_{OT}]' \mid B \sim N(0,B \otimes I)$, where I is an identity matrix, and B is as in (Karaman *et al.*, 2021). Briefly, the diagonals are the breed-specific SNP variances and off-diagonals are covariances.

The PBLUP analysis was performed in DMU (Madsen *et al.*, 2014). The SNP-BLUP and BOA analyses were written and carried out in Julia software (Bezanson *et al.*, 2017).

Validation. The validation animals were selected from the whole population, using paternal half-sib families in which more than half of the animals were born after 2018. For SNP effects estimation we used a subset of animals in CG1 with >10 animals (28,138 animals) and we removed animals in the validation that were included in this subset. We kept 30,234 animals for validation, of which 24,805 were crossbred and 5,618 were purebred. Estimated breeding values were calculated as the summation of the breeding value and breed effects × breed proportions for each animal. The prediction accuracies were calculated as the correlation between adjusted phenotype (y*) and DGV divided by the square root of the heritability. We obtained a bootstrap sample to obtain 10,000 samples that were used to calculate accuracies, standard errors and to make the comparisons between the models with a paired t-test.

Results

Figure 1 shows the prediction accuracies obtained with PBLUP, SNP-BLUP and BOA.

Based on the validation procedure, gains in prediction accuracy with SNP-BLUP over PBLUP were in 72 and 38% for crossbred and purebred animals, respectively, and gains in prediction accuracy with BOA over PBLUP were 82 and 42% for crossbred and purebred animals, respectively. Regarding SNP-BLUP and BOA, there was no significant improvement in accuracy of purebred animals, but improvements of 6% for crossbred animals. Additionally, posterior distribution of B from BOA were used to calculate the correlation of SNP effects among the breeds. Correlations ranged from 0.45 (HO – LM) to 0.89 (HO – OT).

Discussion

The increases in accuracies when including genomic data were observed for both crossbred and purebred animals. Further improvements were obtained when using BOA. In general, the correlations between SNP effects of the OT group and the four main breeds were higher than between the main breeds (results not given). The estimates of high correlations between the SNP effects of the breeds may occur for different reasons: (1) there were five breeds combined in the OT group (beef and dairy breeds) which may lead to problems in assignments of alleles; (2) unassigned alleles were included in the OT group; (3) the imputation of genotypes was carried out jointly in a multi-breed setting for the entire population which might lead to errors in genotypes, and therefore in predictions using BOA. It has been shown that errors in assignments

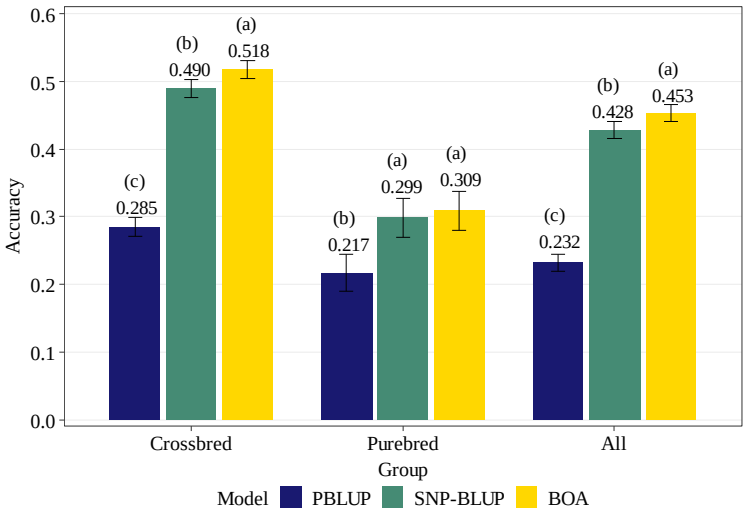


Figure 1. Accuracies and standard errors for validation animals.

of alleles lead to an increase in the correlation of SNP effects between the breeds (Guillenea *et al.*, 2022). Even with correlations higher than expected, the BOA model outperformed PBLUP but more importantly, also out-performed the SNP-BLUP model for animals of the Irish cattle population. Similar results were reported in a simulation study comparing BOA with a model that combines the genomic information of the different breeds (Karaman *et al.*, 2021). The Irish beef cattle evaluation is a very large diverse population in which the vast majority of the animals are crossbred and due to the levels of older phenotypes most are non-genotyped. This implies that a model needs to be developed that can handle both situations. We are currently working on an extension of the BOA model to implement it in populations where some animals are non-genotyped, in a combination of the BOA model with a single-step SNPBLUP model (Liu *et al.*, 2014).

Acknowledgements

We thank GentTORE and ANII Uruguay (POS_EXT_2018_1_154296) for financial support.

References

Bezanson J., Edelman A., Karpinski S., and Shah. V. (2017) SIAM rev. 59(1):65-98. https://doi.org/10.1137/141000671

Guillenea A., Su G., Lund M.S., Karaman E. (2022). J. Dairy Sci. (in press) https://doi.org/10.3168/jds.2021-21173

Eiríksson J.H., Karaman E., Su G., and Christensen O.F. (2021) GSE 53(1), 1-13. https://doi.org/10.1186/s12711-021-00678-3

Goddard M., MacLeod I., Kemper K., Xiang R., Van den Berg I., *et al.* (2018) Proc. of the 11th WCGALP, Auckland, New Zealand.

Karaman E., Su G., Croue I., and Lund M.S. (2021) GSE 53(1), 1-13. https://doi.org/10.1186/s12711-021-00637-y

Liu Z., Goddard M., Reinhardt F., and Reents R. (2014) J. Dairy Sci. 97, 5833-50. https://doi.org/10.3168/jds.2014-7924

Madsen P., Jensen J., Labouriau R., Christensen O., Sahana G., DMU. (2014) Proc. of the 12th WCGALP, Vancouver, Canada.

342. Breed-origin-of-alleles approach using summary statistics for multi-breed genomic prediction in dairy cattle

J.B. Clasen[1,2*], W.F. Fikse[3], G. Su[2] and E. Karaman[2]

[1]Institution of Animal Breeding and Genetics, Swedish University of Agricultural Sciences, Ulls väg 26, 756 51 Uppsala, Sweden; [2]Centre for Quantitative Genetics and Genomics, Aarhus University, Blichers Allé 20, 8830 Tjele, Denmark; [3]Växa, Swedish University of Agricultural Sciences, Ulls väg 26, 756 51 Uppsala, Sweden; julie.clasen@qgg.au.dk

Abstract

Systematic crossbreeding strategies between dairy cattle breeds in dairy herds are becoming more and more attractive to farmers, and this leads to a request for genomically enhanced breeding values for crossbred females in the dairy herds. Accurate genomic prediction of crossbred animals can be achieved if the genotypic and phenotypic data of the breeds involved in the crossbreeding are available to form reference populations, to estimate marker effects. However, sharing genotype and phenotype data between breed populations may be an issue due to privacy and competition. This study investigated genomic prediction of two-breed and three-breed rotational crossbred dairy cattle using summary statistics and a breed-origin of alleles model. The results indicate that the approach can yield almost as high prediction accuracies as having full information from the pure breeds.

Introduction

Crossbreeding between dairy cattle breeds and using genomic testing to predict genomically enhanced breeding values (GEBVs) on future replacement heifers in dairy herds are becoming more attractive to farmers (Magne and Quenon, 2021; Thomasen *et al.*, 2020), and this leads to a stronger request from dairy farmers for predicting GEBVs for crossbred dairy cattle (Clasen *et al.*, 2021).

Previous studies have shown that combining genotype and phenotype data from multiple breeds and crossbreds in a joint reference population can yield higher prediction accuracies for crossbred animals than prediction based on single breed or crossbred reference populations (e.g. Vandenplas *et al.*, 2018). Assuming homogeneous SNP effects across breeds and crossbreds may, in a multi-breed prediction model, not be the best approach if the structure of linkage disequilibrium (LD) and effects of quantitative trait loci (QTL) differs between the crossbred animals and their purebred ancestors (Vandenplas *et al.*, 2016). Instead, assuming heterogeneous SNP effects by tracing each allele back to its breed-origin (BOA) may potentially yield higher genomic prediction accuracies for crossbred animals (Karaman *et al.*, 2021).

Sharing genotype and phenotype data between breeding organisations or foreign countries is difficult due to privacy restrictions and competition (Tenopir *et al.*, 2011). This issue makes genomic prediction for crossbred animals less accurate if the prediction relies on data from reference breeds outside the organisation, that cannot be obtained. A possible solution is to perform genomic prediction based on a meta-analysis of summary statistics including predicted allele substitution effects of single nucleotide polymorphism (SNP) markers and their prediction error (co)variances (Vandenplas *et al.*, 2018). Such an approach is widely used in human genetics (e.g. Maier *et al.*, 2018).

The aim of this study was to investigate the efficiency of using summary statistics as alternative to having full data from purebreds for genomic prediction in two- and three-breed rotational crossbreeding systems using a BOA model.

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_342, © J.B. Clasen *et al.* 2022

Materials & methods

Simulated populations. Three base breed populations of 1,050 Danish Holstein (HOL), 1,050 Red Dairy Cattle (RDC), and 220 Danish Jersey (JER) were simulated based on real genotype data (Karaman *et al.*, 2021). Only the first five first chromosomes (12,664 SNPs) were used in the simulation to reduce the computational demand. Among them, 500 SNPs were randomly chosen as QTL. Twelve generations were simulated by randomly mating 1000 females to 50 males within the HOL and RDC population in each generation. Correspondingly in the JER population, 200 females were randomly mated to 20 males. The population sizes and gender distributions were kept constant in each generation. Additionally, two crossbred populations were simulated based on the purebred populations. A three-breed rotational crossbreeding system was used to produce a crossbred population consisting all three purebreds (MIX) by initially mating HOL females to JER males in the first generation, and then alternating between mating crossbred females to RDC, HOL, and JER males. A two-breed rotational crossbreeding system was used to produce a population consisting of crossbred animals between JER and HOL (JXH) starting by mating HOL females to JER males in the first generation, and then alternating between mating crossbred females to HOL and JER sires. Both crossbred populations were simulated for 12 generations resulting in 1,050 animals per generation and all animals being sired by a HOL male in the last generation. Thus, the average genomic breed proportions in generation 12 were 57.3% HOL, 28.2% RDC, and 14.5% JER for the MIX population, and 66.7% HOL and 33.3% JER for the JXH population.

Each breed population had breed-specific allele substitution at the 500 QTL. Computed from the genetic variances and co-variances in the base populations, the genetic correlations between the breeds were 0.87 between HOL and RDC, 0.62 between HOL and JER, and 0.68 between RDC and JER. Heritabilities of the simulated trait were around 0.4 for the three breeds.

The reference populations consisted of animals from generations 9, 10, and 11. Joining all pure breeds and crossbred populations in a combined reference gave a total of 13,260 animals. The average genomic breed proportion of the animals in the combined reference population were 42.2% HOL, 31.6% RDC, and 26.1% JER. The animals in the test populations were all from generation 12.

Breed-origin-of-alleles model. The SNP effects were estimated for each breed as follows:

$$y = 1\mu + Xb + M_1\beta_{HOL} + M_2\beta_{RDC} + M_3\beta_{JER} + e \quad (1)$$

where y is the vector of phenotypes, 1 is a vector of 1s, μ is the overall mean, X is the matrix of (centred) estimated genomic breed proportions computed from genomic data, b is the vectors of fixed breed effects, and M_1, M_2, and M_3 are the column-centred matrices of breed-specific allele content of SNPs for HOL, RDC, and JER, respectively, βs are vectors of SNP effects for HOL, RDC, and JER, respectively, and e is a vector of residuals. We used a Bayesian approach to estimate the dispersion and location parameters and assigned a normal distribution prior to the vectors of SNP effects for each breed: $\beta_i|\sigma^2_{\beta i} \sim N(0, I\sigma^2_{\beta i})$, where *i* stands for the breeds HOL, RDC and JER. Residuals were *a priori* assumed to follow a normal distribution, $e \sim N(0,R)$ where R was a diagonal matrix of residual variances.

The GEBVs were predicted using SNP effects estimated from the combined reference population. In predicting GEBVs for crossbred animals, alleles of each SNP were traced back to their (known) breed of origin, and the SNP effect from the respective breeds. The GEBVs for pure breeds were predicted by multiplying SNP effects with the corresponding allele counts. Additionally, for prediction of GEBVs in crossbred animals, the fixed breed effects were multiplied by genomic breed proportions. Prediction accuracies were calculated as the correlation between the simulated genetic value and the predicted GEBVs.

Summary statistics. If only summary statistics were available for a certain breed, the prior distribution assigned to the SNP effects of the breed was reformulated using the results from a separate analysis of that individual breed. The mean and variance of the prior distribution of the SNP effects was set to the estimated SNP effect and its posterior variance from the separate analysis. The number of animals and their mean phenotypes was used to update the prior for the fixed breed effects.

Scenarios. We investigated different combinations of information sources, including full genotype and phenotype data, only summary statistics, or no data available from the pure breed populations (HOL, JER, and RDC) using the BOA approach. The crossbred populations (MIX and JXH) always had phenotype and genotype information available. All of the analyses were carried out in scripts written in Julia and were replicated 25 times. The prediction accuracies of the different scenarios were compared by a paired t-test using a significance level of $P<0.05$ within test populations.

Results

Within all test populations, using summary statistics from the pure breeds yielded slightly (0.021-0.046) lower prediction accuracies compared with using full genotype and phenotype data, while having no information from the pure breeds yielded much (0.086-0.224) lower prediction accuracies as shown in Table 1. Replacing full data with summary statistics from JER had no or little effect on the prediction accuracies for any of the test populations except JER. When only summary statistics or no information was available from either JER or RDC reference populations or both, the prediction accuracies decreased more for the RDC test population than JER.

Discussion

Our results show that all breeds represented in a crossbreeding system should be included in the reference population to obtain higher prediction accuracies. In the situation when full genotype and phenotype data is not available from all or some of the breeds, using summary statistics can yield prediction accuracies almost as high as when full data is available. Furthermore, using summary statistics to run prediction models does not require a large amount of data, which makes it more practically feasible.

Table 1. Prediction accuracies for Danish Holstein (HOL), Danish Jersey (JER), Swedish Red (RDC), three-breed rotational crosses (MIX) and JER × HOL rotational crosses (JXH) in different scenarios including full phenotype and genotype data (F), summary statistics (S), or no information (-).[1]

Data from reference population					Test population				
HOL	RDC	JER	MIX	JXH	HOL	RDC	JER	MIX	JXH
F	F	F	F	F	0.798_{a}	0.761_{b}	0.743_{b}	0.753_{b}	0.789_{b}
S	S	S	F	F	0.752_{c}	0.731_{c}	0.710_{c}	0.720_{c}	0.768_{c}
F	S	S	F	F	0.795_{b}	0.554_{d}	0.673_{d}	0.718_{c}	0.782_{e}
F	S	F	F	F	0.795_{b}	0.553_{d}	0.748_{a}	0.718_{c}	0.785_{d}
F	F	S	F	F	0.799_{a}	0.762_{a}	0.713_{c}	0.755_{a}	0.791_{a}
-	-	-	F	F	0.605_{d}	0.537_{e}	0.657_{e}	0.590_{f}	0.676_{gh}
F	-	-	F	F	0.755_{c}	0.528_{f}	0.491_{g}	0.647_{e}	0.669_{h}
F	-	F	F	F	0.759_{c}	0.529_{f}	0.667_{de}	0.655_{d}	0.693_{f}
F	F	-	F	F	0.763_{c}	0.743_{c}	0.499_{f}	0.706_{c}	0.679_{g}

[1] Accuracies within test population without common subscripts differ significantly ($P<0.05$).

The sizes of the crossbred reference populations in this study were as large as for HOL and RDC, which is probably optimistic since number of (genotyped) crossbred dairy cattle is small in most dairy cattle populations. Using summary statistics from JER (which had a smaller population size) had little or no effect on the prediction accuracies, compared with having full information from JER, because there were relatively more animals involved with JER genes in the crossbred reference populations than in the purebred JER reference. Furthermore, all animals in the MIX and JXH test populations were HOL-sired, making the information from the RDC and JER reference populations less important than HOL in predicting GEBVs for the crossbred animals. Assuming smaller crossbred reference populations with various breed combinations in the real-life situation, information from purebreds will be more important than indicated in this study.

A main reason for crossbreeding in dairy cattle is utilizing the benefits of heterosis effects (Sørensen *et al.,* 2008), which were ignored in the simulation of phenotypes and estimation of GEBVs in this study. Heterosis is caused by non-additive effects: dominance effects between alleles at the same loci and epistatic effects between alleles at different loci. Ignoring the non-additive effects may cause some biasedness of the genomic prediction (Su *et al.,* 2012). Therefore, if the current model should be applied to routine genomic breeding evaluation, non-additive effects should be considered in the model.

References

Clasen J.B., Bengtsson C., Kallström H.N., Strandberg E., Fikse W.F. *et al.* (2021) Animal 15(12):100409. https://doi.org/10.1016/j.animal.2021.100409

Clasen J.B., Fikse W.F., Kargo M., Rydhmer L., Strandberg E. *et al.* (2020) J Dairy Sci 103(1):514-528. https://doi.org/10.3168/jds.2019-16958

Karaman E., Su G., Croue I., and Lund M.S. (2021) Genet Sel Evol 53(1):46. https://doi.org/10.1186/s12711-021-00637-y

Magne M.A. and Quenon J. (2021) Agron Sustain Dev 41(2):1-15. https://doi.org/10.1007/s13593-021-00683-2

Maier R.M., Zhu Z., Lee S.H., Trzaskowski M., Ruderfer D.M *et al.* (2018) Nat Commun 9(1):989. https://doi.org/10.1038/s41467-017-02769-6

Su G., Christensen O.F., Ostersen T., Henryon M., and Lund M.S. (2012) PLoS One 7(9):e45293. https://doi.org/10.1371/journal.pone.0045293

Sørensen M.K., Norberg E., Pedersen J., and Christensen L.G. (2008) J Dairy Sci 91(11):4116-4128. https://doi.org/10.3168/jds.2008-1273

Tenopir C., Allard S., Douglass K., Aydinoglu A.U., Wu L. *et al.* (2011) PLoS One 6(6): e21101. https://doi.org/10.1371/journal.pone.0021101

Thomasen J.R., Liu H., and Sørensen A.C. (2020) J Dairy Sci 103(1):597-606. https://doi.org/10.3168/jds.2019-16974

Vandenplas J., Calus M.P.L., and Gorjanc G. (2018) Genetics, 210(1):53-69. https://doi.org/10.1534/genetics.118.301109

Vandenplas, J., Calus, M.P.L., Sevillano C.A., Windig J.J., and Bastiaansen J.W.M. (2016) Genet Sel Evol 48:61. https://doi.org/10.1186/s12711-016-0240-y

Part 19

Methods and tools: software and computing strategies

343. Subsetted orthogonal data augmentation for fast parallel implementation of Bayesian models for whole-genome analyses

L.H. Chen and G.S. Plastow*

Department of Agricultural, Food and Nutritional Science, Livestock Gentec, University of Alberta, Edmonton, AB T6G 2H1, Canada; plastow@ualberta.ca

Abstract

A novel subsetted orthogonal data augmentation (SODA) approach was proposed for fast parallel implementation of Bayesian linear regression models for whole-genome analyses. Performance was evaluated using a simulated data set from the QTL-MAS XIV workshop. Computing speed, convergence, and genomic prediction accuracy and bias were compared to those using a regular Gibbs sampling algorithm and the original orthogonal data augmentation (ODA). Results showed that the SODA approach required less time and number of operations to compute per CPU than the ODA and regular algorithms. The SODA algorithm converged faster and had larger effective sample size than the ODA algorithm. Genomic prediction accuracy and bias were similar among different approaches. The proposed SODA approach will be especially useful when the number of markers is very high, where a regular algorithm or the ODA approach are infeasible or takes too long to complete.

Introduction

Bayesian linear regression models have been widely used for whole-genome analyses including genome-wide association studies and genomic prediction (Mewwissen *et al.*, 2001; Gianola, 2013). Estimates for the unknown model parameters are usually obtained via a specially designed MCMC sampling approach, such as a Gibbs sampler. The MCMC sampling procedure is computationally demanding and cannot be easily parallelized to take advantage of modern computers with multiple CPUs. Ghosh and Clyde (2011) proposed an orthogonal data augmentation (ODA) approach without changing the posterior distributions of the model parameters of interest. Zhao *et al.* (2020) adapted the ODA approach to implement a parallel sampling algorithm for genomic prediction using Bayesian linear regression models. Although computing time was shortened by several magnitudes with HPC of hundreds of CPUs, the ODA approach suffered from slow convergence issues. The ODA approach requires computation of a matrix product between the transposed and original design matrix, and Cholesky decomposition of the resulting matrix. It may become infeasible when a very high-density marker panel or whole-genome sequence is used.

One solution is to augment subsets of the design matrix instead of the full matrix. Such an approach has the potential to reduce computing time and improve convergence due to smaller number of augmented phenotypes. It is also feasible when using very high density or whole-genome sequence markers. The objective of this study was to evaluate the performance of a novel subsetted orthogonal data augmentation (SODA) approach and compare it with the ODA approach as well as the regular Gibbs sampling approach for genomic prediction using Bayesian linear regression models.

Materials & methods

Bayesian linear regression models. For simplicity and without loss of generality, the Bayesian linear regression model with a general mean, additive marker effects, and a residual term can be written as:

$$y = 1\mu + Xa + e \quad (1)$$

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_343, © L.H. Chen and G.S. Plastow 2022

where y is a vector of phenotypes for n individuals, μ is the general mean, X is an n × p matrix of marker genotype covariates (coded as 0,1,2), a is a vector of additive marker effects, and e is a vector of residuals. Elements of X are standardized as $\left(x_{ij} - 2p_j\right)/\sqrt{2p_j(1-p_j)}$, where p_j is the allele frequency at marker j. μ follows a flat prior distribution; $e|\sigma_e^2 \sim N(0, \sigma_e^2)$ with $\sigma_e^2|v_e, s_e^2 \sim \chi^{-1}(v_e, s_e^2)$. Prior distribution for marker effects varies with different Bayesian Alphabet models (Gianola, 2013). For a BayesR model as proposed by Erbe *et al.* (2012), $a_j|\pi, \sigma_a^2 \sim \pi_1\delta_0 + \pi_2 N(0, \lambda_2\sigma_a^2) + \pi_3 N(0, \lambda_3\sigma_a^2) + \pi_4 N(0, \lambda_4\sigma_a^2)$, where π is the probability following a Dirichlet distribution:

$$\frac{1}{Beta(\alpha)}\prod {\pi_k}^{\alpha_k - 1},$$

δ_0 is a point mass at zero, σ_a^2 is a common variance, and λ_k is a scaling factor taking values 0.0001, 0.001 or 0.01 as suggested by the author. σ_a^2 can be estimated from data by giving it a prior distribution, $\sigma_a^2|v_a, s_a^2 \sim \chi^{-1}(v_a, s_a^2)$. To facilitate computation, an indicator variable I_j is introduced to indicate which mixture distribution a_j is currently sampled from. I_j follows a categorical distribution $\prod \pi_k^{[I_j=k]}$, where $[I_j = k]$ evaluates to 1 if $I_j = k$, and 0 otherwise.

Regular Gibbs sampling algorithm. A regular Gibbs sampling algorithm can be used to construct a MCMC chain by sampling from the full conditional posterior distributions of the unknown model parameters. The most computationally expensive part of the Gibbs sampling algorithm is for the sampling of marker effects, which must be sampled one after another. As described by Erbe et. al. (2012), sampling of marker effect a_j can be proceeded by first sampling I_j from a categorical distribution with $p(I_j = k|\Theta_{-j}, y)$, followed by sampling a_j from $f(a_j|y, \Theta_{-j}, I_j)$, where Θ_{-j} denotes all parameters except a_j and I_j.

$$p(I_j = k|\Theta_{-j}, y) = \frac{\pi_k f(y|\Theta_{-j}, I_j = k)}{\sum \pi_k f(y|\Theta_{-j}, I_j = k)} \text{ with}$$

$$f(y|\Theta_{-j}, I_j) \propto [I_j = 1] + [I_j = k, k \neq 1](x'_j x_j \frac{\lambda_k\sigma_a^2}{\sigma_e^2} + 1)^{-\frac{1}{2}} exp\left\{\frac{\left(x'_j y^*\right)^2}{2\sigma_e^2\left(x'_j x_j + \frac{\sigma_e^2}{\lambda_k\sigma_a^2}\right)}\right\} \quad (2)$$

where y^* is the vector of phenotype corrected for all effects except effect of marker j.

$$f(a_j|y, \Theta_{-j}, I_j) = [I_j = 1]\delta_0 + [I_j = k, k \neq 1]N\left(\frac{x'_j y^*}{x'_j x_j + \sigma_e^2/\lambda_k\sigma_a^2}, \frac{\sigma_e^2}{x'_j x_j + \sigma_e^2/\lambda_k\sigma_a^2}\right) \quad (3)$$

Orthogonal data augmentation (ODA). Following Ghosh and Clyde (2011) and Zhao *et al.* (2020), the observed design matrix in (1) including the intercept, W_o, can be augmented with a $(p + 1) \times (p + 1)$ matrix W_a, so that:

$$[W_o' \quad W_a'] \begin{bmatrix} W_o \\ W_a \end{bmatrix} = D,$$

where D is a diagonal matrix. It was shown that by choosing the largest eigenvalue of $W_o'W_o$ as the diagonals of D, a real matrix, W_a, was guaranteed to exist. W_a can be obtained via singular value decomposition or Cholesky decomposition of $dI - W_o'W_o$. Zhao *et al.* (2020) applied the ODA approach to implement a parallel Gibbs sampling algorithm for genomic prediction. With orthogonal data augmentation, Equation 2 and 3 become:

$$f(y|\Theta_{-j}, I_j) \propto [I_j = 1] + [I_j = k, k \neq 1](\frac{d\lambda_k\sigma_a^2}{\sigma_e^2} + 1)^{-\frac{1}{2}} exp\left\{\frac{\left(x'_j y\right)^2}{2\sigma_e^2\left(d + \frac{\sigma_e^2}{\lambda_k\sigma_a^2}\right)}\right\} \quad (4)$$

$$f(a_j|y, \Theta_{-j}, I_j) = [I_j = 1]\delta_0 + [I_j = k, k \neq 1]N\left(\frac{x'_j y}{d + \sigma_e^2/\lambda_k\sigma_a^2}, \frac{\sigma_e^2}{d + \sigma_e^2/\lambda_k\sigma_a^2}\right) \quad (5)$$

Computation for (4) and (5) can be conducted simultaneously for all markers, and therefore, can be implemented in parallel on a modern computer with multiple CPUs.

Subsetted orthogonal data augmentation (SODA). The original genotype covariate matrix can be arbitrarily subsetted into equal or similar sized blocks, X_{o1}, X_{o2}, …, X_{om}. Each block X_{ot} can be augmented with an augmenting matrix X_{at} so that:

$$[X'_{ot} \quad X'_{at}] \begin{bmatrix} X_{ot} \\ X_{at} \end{bmatrix} = D_t,$$

where D_t is a diagonal matrix. With subsetted orthogonal data augmentation, Equation 2 and 3 become:

$$f(y|\Theta_{-j}, I_j) \propto [I_j = 1] + [I_j = k, k \neq 1]\left(\frac{d_j}{\sigma_e^2/\lambda_k\sigma_a^2} + 1\right)^{-\frac{1}{2}} exp\left\{\frac{\left(x_j' y^*\right)^2}{2\sigma_e^2\left(d_j+\sigma_e^2/\lambda_k\sigma_a^2\right)}\right\} \quad (6)$$

$$f(a_j|y, \Theta_{-j}, I_j) = [I_j = 1]\delta_0 + [I_j = k, k \neq 1]N\left(\frac{x_j' y^*}{d_j+\sigma_e^2/\lambda_k\sigma_a^2}, \frac{\sigma_e^2}{d_j+\sigma_e^2/\lambda_k\sigma_a^2}\right) \quad (7)$$

Computation for (6) and (7) can be done simultaneously for markers within the same block.

Data analysis. Simulated data from the 14th QTL-MAS workshop was used to evaluate performance of the three algorithms. Details of the data set was described by Szydlowski and Paczynska (2011). Briefly, 2,326 individuals with phenotypes and genotypes on 10,031 markers were used as training population, and 900 individuals with genotypes only were used for validation.

The number of blocks in the SODA approach ranged from 2 to 150. Number of computational operations, computing time, convergence of the MCMC chain, and accuracy and bias of genomic prediction were compared. A program was written in Fortran 2018 with OpenMP. The program was run on a dedicated HPC node with 24 Intel Platinum 8160F Skylake CPU cores @ 2.1Ghz and 187G RAM. For each scenario, the Gibbs sampling steps were run for a MCMC chain length of 50,000. Computing time was recorded as wall clock time from start to the end of running the program. Convergence of the MCMC chain was diagnosed using Geweke diagnostic in the R CODA package. This was done by dividing the first half of the chain into multiple bins and testing the equality of sample means from each bin and the last half of the chain. Lagged autocorrelations were also computed to evaluate the effective sample size of the MCMC chain. Accuracy of genomic prediction was calculated as correlation between genomic estimated breeding values and the true genetic values for the validation population. Bias of genomic prediction was evaluated by regressing true genetic value on genomic estimated breeding value for the validation population. The computer program used in this study can be downloaded for research use from https://github.com/liuhong-chen/SODABayes.

Results & discussion

The number of computational operations per CPU as a percentage of that required by the regular Gibbs sampling algorithm were determined (Figure 1). The ODA algorithm requires around 20% of the operations per CPU of the regular algorithm with 24 CPUs, while the number of operations for the SODA algorithm decreases with increasing number of blocks. The decreased number of operations for the SODA algorithm was due to smaller number of augmented phenotypes. Computing time decreased from 1,150 s for the regular algorithm to 478 s for the ODA algorithm and to a minimum of 168 s for the SODA algorithm with 16 blocks (Figure 1). The decreased computing time was due to reduced number of operations. Computing time increases for the SODA algorithm after reaching the bottom at around 16 blocks. The increase of computing time was due to increased overhead cost of the multi-threaded OpenMP program. The results suggest that the SODA algorithm with the studied data set can be 6.8 and 2.8 times faster compared to the regular algorithm and the ODA algorithm, respectively. To achieve an optimum computing performance,

the number of blocks needs to be chosen to balance the computational operations and overhead costs. The relative overhead cost is expected to decrease with increasing number of phenotypic observations and the number of marker loci as the threads will have more computational work.

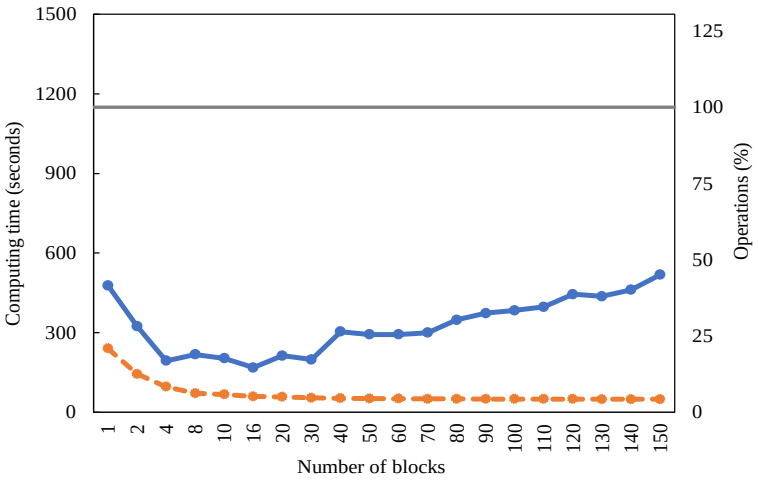


Figure 1. Comparison of number of operations per CPU (orange dotted line) and computing time (blue solid line). The flat grey line indicates number of operations and computing time for the regular Gibbs sampling algorithm.

Geweke convergence diagnostic results indicated that the regular Gibbs sampling algorithm took about 2,000 samples to converge, while the ODA and SODA with 16 blocks required around 24,000 and 12,000 samples, respectively, to converge. Lagged autocorrelation decayed fastest for the regular algorithm, and slowest for the ODA algorithm. Autocorrelation for SODA algorithm with 16 blocks was similar to that of the regular algorithm.

Accuracy and bias of the genomic estimated breeding values were similar for the regular, ODA, and SODA algorithms.

Conclusions

A subsetted orthogonal data augmentation approach was proposed for fast parallel implementation of Bayesian linear regression models. The proposed approach is superior to the original augmentation approach in terms of computing speed and convergence and should be feasible even when a very high-density marker panel or sequence is used.

Acknowledgements

This work was funded by Genome Alberta, Genome Canada, Alberta Innovates and the Government of Alberta through the EBS Competition. The work was enabled in part by support provided by WestGrid (www.westgrid.ca) and Compute Canada (www.computecanada.ca).

References

Brøndum, R.F., Bowman, P.J., Goddard, M.E., *et al.* (2012) BMC Genomics 13(1):1-11. https://doi.org/10.1186/1471-2164-13-543

Erbe, M., Hayes B. J., Bowman P. J., *et al.* J. of Dairy Sci. (2012) 95(7): 4114-4129. https://doi.org/10.3168/jds.2011-5019

Gianola, D. (2013) Genetics 194(3): 573-596. https://doi.org/10.1534/genetics.113.151753
Ghosh, J., and Clyde M.A. J. Am. Stat. Assoc. (2011) 106(495): 1041-1052. https://doi.org/10.1198/jasa.2011.tm10518
Meuwissen T.H.E, Hayes B.J., Goddard M.E. (2001) Genetics 157:1819–1829. https://doi.org/10.1093/genetics/157.4.1819
Szydlowski M, Paczynska P. (2011) BMC Proc. 5(Suppl 3): S4. https://doi.org/10.1186/1753-6561-5-S3-S1
Zhao, T., Fernando R., Garrick D., and Cheng H. Genet. Sel. Evol. (2020) 52(1): 1-11. https://doi.org/10.1186/s12711-020-00533-x

344. A simulator to evaluate gene editing assisted selection programs

X-J. Yu[1], B.S. Dagnachew[2] and T.H.E. Meuwissen[1]

[1]Norwegian University of Life Sciences, P.O. Box 5003, 1432, Ås, Norway; [2]Nofima, Norwegian Institute for Food, Fisheries and Aquaculture Research, 1433, Ås, Norway; xijiang.yu@nmbu.no

Abstract

A stochastic simulation package of breeding programs was developed in Julia to evaluate gene editing assisted selections (GEAS). Its generic design makes it to be a general-purpose breeding program simulator. The package can take real or simulated haplotypes of founder populations. It then drops the genotypes through arbitrary multi-trait breeding schemes. Traits can be continuous and binary. The number of genes and the distribution of their effects must be specified. Breeding value estimates (EBV) are obtained from SNP-BLUP by default. Known QTL may be emphasized by treating them as fixed effects in the EBV. GEAS can also incorporate other EBV estimation methods. Gene-editing of known QTL is an option for the breeding programs. Calculations are parallelized wherever possible for computational speed. Early simulation results show significantly genetic gain improvement by applying gene-editing in breeding programs. GEAS is easily maintained and extended, thanks to Julia features.

Introduction

Gene editing technology (e.g. Wargelius A., 2019) is becoming increasingly mature and accurate. It is a powerful tool to produce more productive, functional, and disease resistant farm animals. Increased disease resistance can greatly improve animal welfare. Whilst its technology and legislation are still under development, it is important to perform computer simulation studies to explore optimal strategies to make use of gene-editing.

Softwares that simulate gene-editing schemes, e.g. Jenko *et al.*, 2016, Gaynor *et al.*, 2020 are usually R based, which limits speed and ability to deal with large datasets. Moreover, simulation programs that can easily setup various scenarios and easily maintained and extended are needed, to deal with the many ways to apply gene editing techniques in breeding. Contrary to some other gene editing scenarios, we will assume here that embryos are edited instead of mature individuals.

The GEAS stochastic simulation package aims to evaluate a multitude of gene edit assisted selection programs. The application described below is adapted to fish breeding schemes, but the approach is general. The generic design made the package useful to evaluate conventional breeding schemes of sexually reproducing diploid organisms. The GEAS package, which is written in Julia (https://julialang.org), parallelizes calculations wherever possible. Traits may be continuous (e.g. a production trait) or binary (e.g. a disease resistance trait) in nature. The breeding population structure and gene makeups may be specified by users. Gene editing is applied to known genes. The interface of the package takes in a base population either from real data or a base population simulator. GEAS is easy to maintain and expand, thanks to Julia language features.

Materials & methods

GEAS was designed to evaluate gene editing assisted selection. It can, however, also simulate traditional (genomic selection) breeding schemes. A brief overview of the design and many features of the package is described below (Figure 1).

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_344, © X-J. Yu *et al.* 2022

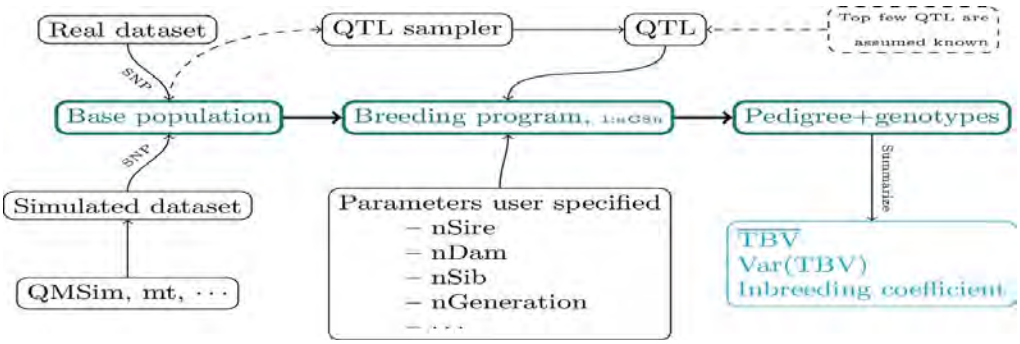


Figure 1. The design of the GEAS package.

The program starts from a founder population. QTL are sampled from the founder genotypes. The founder population genes flow then through a breeding program defined by a dictionary of parameters. The resulting pedigree and genotypes are stored for downstream evaluations.

Genotypes of a founder population can be obtained from real data. They can also come from a base population simulator. The program can automatically download and update base-population simulators such as QMSim, and MaCS. Other simulators can also be integrated into GEAS.

The number of crossovers that occur on a chromosome are sampled from a Poisson distribution with parameter the length of the chromosome in Morgan. It is by default 10^8 base pairs per Morgan. Offspring haplotypes are created in parallel over individuals. That is the number of individuals simultaneously generating equals the number of threads available to the program.

QTL are sampled from the available SNPs in the founder population. The number, and distribution of the QTL effects are specified by the users. QTL can be sampled independently for each trait. In this case, each trait can have different numbers of QTL with a different distribution. QTL can also be pleiotropic and have a multi-variate normal distribution. The QTL with the largest effects are recorded and are candidates for gene editing.

GEAS was designed primarily to mimic fish breeding schemes, with the potential to integrate gene-editing methods. Two traits are simulated by default. One of them is a continuous trait (e.g. production). The other one is a binary trait (e.g. disease resistance). The offspring in each generation may be split into two groups. Group I stays in the nucleus of the breeding program. Group II serves as test individuals (typically challenge tested for a disease). In case of the challenge test, the default survival rate is 50% to maximize the variance of the records. The genotypes of both groups, together with specified traits from group I and II, respectively, are recorded for genetic evaluation. More groups can be added together with a specification of their recorded traits.

SNP-BLUP is the default genomic breeding value (GEBV) prediction method. It uses single trait analyses and the genotypes of all generations. The model for the single trait analyses is:

$$y_{ij} = \mu + f_j + \sum_{k=1}^{n_{\mathrm{SNP}}} \mathrm{SNP}_{ijk} \cdot a_k + e_{ij}$$

That is, individual *i*'s phenotype in generation *j* is modeled by a generation effect f_j, sum of SNP genotypes (SNP_{ijk}) times their effects and a residual. The cyan part is the GEBV of this individual.

The SNP effects, $\widehat{a_k}$, estimated from the challenge group are used to calculate the GEBV of individuals in the production group. A combined index is then calculated for everyone in the production group using economic weights provided by the user. The individuals with top indices are selected as parents for the next generation. They are randomly mated.

For known QTL, users can choose whether to edit them in the breeding generations. The editing is one QTL a generation, in descending order of the alleles' absolute effect sizes. All the embryos/offspring born in the nucleus (group I) are assumed edited for the beneficial QTL allele with a default efficiency of 100%. A lower editing efficiency or fraction of edited offspring may also be specified.

As an alternative to editing, the known QTL may be emphasized in the SNP-BLUP breeding value estimation procedure. This is achieved by: (1) adding the known QTL to the SNP-set used for genetic evaluation; and (2) treating the known QTL as fixed effects instead of as random effects that are severely regressed back to 0. The latter is realized by not adding the λ term, i.e. the variance ratio of the error variance to the SNP variance, to the corresponding SNP's diagonal in the mixed model equations, which implies that the SNP is fitted as a fixed effect.

The GEAS package was written in the Julia language. It is easy to modify, maintain, and extend. This combined with its high computational speed makes the overall efficiency higher than e.g. that of C, or Fortran. Its speed is as fast as C, or Fortran implementations. The program also provides a web interface for end users to ease the design of breeding programs.

The software is freely available on https://github.com/xijiang/GEAS.jl. It is a standard Julia package which can be installed locally. Its documentation is also provided on the website.

Results of example simulations

In an example breeding scheme, three breeding strategies were compared using the same founder population. They are: (1) GEBV estimation with SNP-BLUP; (2) treat all known QTL as fixed effects; (3) edit known QTL (one in each generation). The QTL scenarios were: (1) small number of QTL, 500; and (2) large number OF QTL 5,000, which were combined with QTL effects distributed as either normal or Laplacian distribution. Each scenario had 30 repeats. In each generation, 100 sires and 200 dams were randomly mated to produce 40 sibs. 30 of them were challenged. The rest 10 were evaluated and selected for the next generation.

Figure 2 left-panel shows the results of 5,000 Laplacian QTL. Gene-editing gave the maximal TBV improvement. Traditional SNP-BLUP gave the least. Similar results were obtained with other scenarios. Figure 2 right-panel compared the genetic improvement with gene editing in four scenarios. It shows that the fewer QTL, the higher genetic improvement in the first 10 generations. Distributions with extreme QTL effects (Laplacian) showed more improvement for gene editing effects. A population with 5,000 Laplacian distributed QTL has similar traditional selection and gene editing effects as a population with 500 normally distributed QTL.

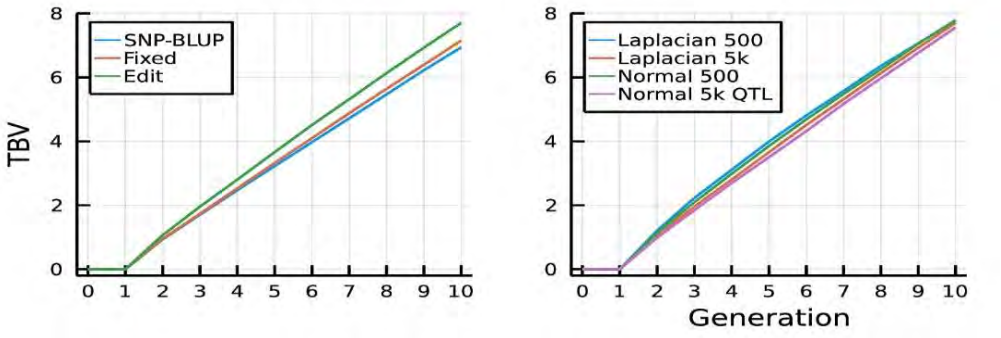


Figure 2. A comparison of different selection methods and genetic architectures.

Discussion

We have developed a new breeding program simulator. This package is fast, easy to maintain and extend. It also utilized many features of modern computer languages. More documentation and demonstrations can be found at https://github.com/xijiang/GEAS.jl.

A preliminary simulation study showed significantly enhanced breeding progress due to gene editing technology. Gene editing gave an extra genetic progress of about 8 to 14% over ordinary SNP-BLUP in our scenarios, even if only one QTL was modified per generation. Emphasizing known QTL also resulted in more genetic improvements of 2-3%. Gene-editing will require, however, precisely positioned causative mutations, which are therefore important.

The current version of GEAS provides many possibilities for alternative scenarios. In the future, we will simulate more scenarios of breeding related to gene editing with GEAS. Meanwhile, more functions will be integrated. Multi-allelic QTL may be a future option. The web interface will be more fledged for a better user experience. The breeding part will have more alternative breeding value estimation methods.

References

Chen G.K., Marjoram P., Wall J.D., Fast and flexible simulation of DNA sequence data., Genome Res, 2009.

Gaynor R.C., Gorjanc G., and Hickey J.M., AlphaSimR: an R package for breeding program simulations. G3, 2021, 11(2).

Jenko J., Gorjanc G., Cleveland M.A., *et al*., Potential of promotion of alleles by genome editing to improve quantitative traits in livestock breeding programs. GSE, 2015, 47:55.

Sargolzaei, M. and F. S. Schenkel. QMSim: a large-scale genome simulator for livestock. Bioinformatics, 2009. 25: 680-681.

Wargelius A., Application of genome editing in Aquatic farm animals: Atlantic salmon., Transgenic res., 2019, 28, 101:105.

345. A database structural improvement for efficient trait variation curation in animal QTLdb and CorrDB

Z-L. Hu, C.A. Park and J.M. Reecy*

Department of Animal Science, Iowa State University, 2255 Kildee Hall, Ames, IA, USA; jreecy@iastate.edu

Abstract

In livestock genetics studies to improve production traits, the importance of accurately identifying and recording traits of interest cannot be overemphasized. However, it has been a challenge to consistently and unambiguously name and compare traits of interest that span time and geographic locations and are evaluated by different people. While the Animal Trait Ontology for Livestock, Clinical Measurement Ontology, Livestock Product Trait Ontology, and Vertebrate Trait Ontology, among others, provide good frameworks for using controlled vocabularies, building concept hierarchies, and defining relationships between terms, gaps still remain as to what level of granularity should be documented. This is important to efficiently use available resources for effective data curation. In the Animal QTLdb and CorrDB, we have introduced an approach to extend livestock trait ontologies for practical data curation that allows effective capture of trait variations while keeping complexities to a manageable level.

Introduction

The volume of genotype-to-phenotype data from livestock animal genomics studies has been increasing explosively over the past 20+ years, owing to continued progress in sequencing and genotyping technologies. For example, the cattle QTL and association data curated into the Animal QTLdb has increased 330-fold over the past 15 years (Figure 1). Trait ontology development using Vertebrate Trait Ontology (Park *et al.* 2013), Livestock Product Trait Ontology (Park *et al.*, 2021), and Clinical Measurement Ontology (Shimoyama *et al.*, 2012) has been an ongoing part of Animal QTLdb and CorrDB developments (Hu *et al.*, 2007, 2013, 2016, 2019). The increase in incoming data presents challenges for Animal QTLdb and CorrDB data curation, not only due to an increase in workload, but also due to a need to accommodate more diverse data formats, methods of data detection and analysis, and degree of trait data granularity in each new experiment. For example, for the most common trait, Average Daily Gain (ADG), many use cases have been identified in different species, such as in cattle (ADG in cows, heifers, on test, pre-weaning, post-weaning), pigs (120-210d, 120-240d, 21-120d, 21-210d, 21-240d, 21-46d, 210-240d, 30-100kg, 46-120d, 46-210d, 46-240d, BirthTo30kg, pre-weaning, post-weaning), and sheep (3-6Mos, 6-12Mos, 6-9Mos, 9-12Mos, BirthTo3Mos, BirthTo6Mos, pre-weaning, WeaningTo6Mos, WeaningTo9Mos). Clearly, numerous additional variants are possible, depending on study criteria such as age (time), weight, and/or production management stage. The combined use of these factors can vastly increase the number of different ADG terms. Previously, we developed a scheme to accommodate 'sibling traits' (Götz *et al.*, 2012) modified from their base trait (e.g. ADG) in order for reported QTL/associations to have a proper trait name to link to (Hu *et al.*, 2019). This conceptual design was implemented in the QTLdb curator environment and it worked fine, except that the number of such sibling traits quickly exceeded what could be easily managed. The problem was more obvious when a curator needed to choose a trait from a daunting lengthy list of (sibling) traits. When it came to recording the sibling traits as part of a trait ontology hierarchy, concerns arose, such as continually growing trait lists, possibilities of sibling traits as children of base traits, etc. (Figure 2a). To solve these problems, we have envisioned a 'trait variant' structure for practical use in the Animal QTLdb and CorrDB curator tools environment. Here we report the initial steps toward this effort.

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_345, © Z-L. Hu *et al.* 2022

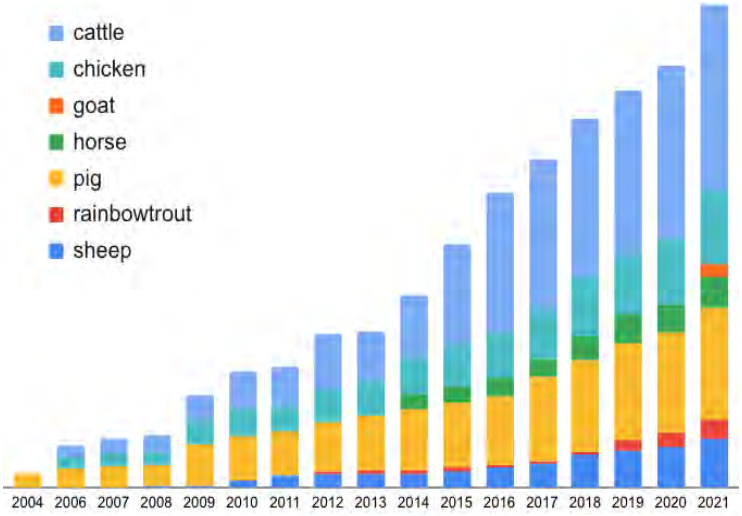


Figure 1. QTL/association data growth in the Animal QTLdb. Data counts were transformed for better visualization of scale. For the actual data counts please refer to https://www.animalgenome.org/QTLdb.

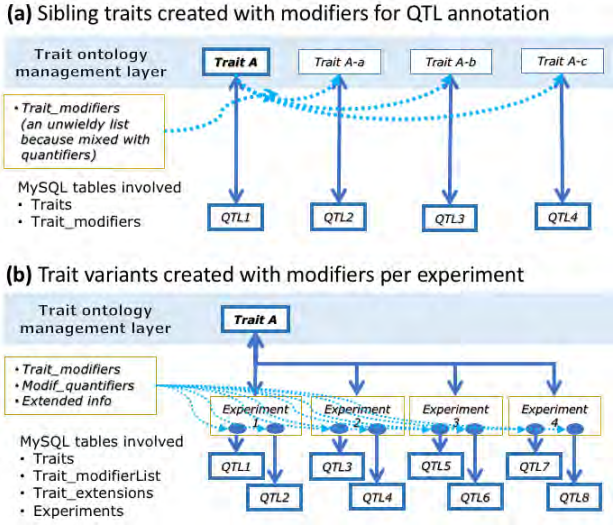


Figure 2. Conceptual data structure differences between 'sibling traits' ['modified'] as part of ontology extensions (a), and 'variant traits' [modified with extended info] created at the experiment level (b), and their use in QTL data annotations.

Results

Approach. A trait term may be 'modified' by another term, as a property, to produce an extension of the trait term as a new term. Examples of such extensions include muscle pH (pH) measured at 24hr or 48hr post-mortem, or in different muscles, such as biceps femoris or semimembranosus; milk yield (MY) measured daily or cumulatively up to 305 days of calf age, or MY compared between the first and second parities, etc. In these cases, if we call pH or MY the base trait, then base traits with additional properties

are modified traits. In order to bring the modifier terms to a manageable level in the trait data curation environment, we use additional terms to extend the context of a modifier. Examples of such terms can be number of days (time) in ADG measurements, parity number in MY measurements, etc. The extended names of traits incorporating the base trait in addition to the relevant modifiers are called trait variants (Figure 2b).

Implementation. By limiting modifiers to some major terms, we were able to create a relatively short list of modifiers (Analysis, Anatomy location, Environment, Herd, Instrument, Measurement, Parity, Pedigree, Stage, Time) for a pull-down menu in the curator tool (Figure 3). Each of the modifier terms has its own list of relational terms to further define its context (e.g. age, before, after, by, during, etc. for 'time'; above, anterior, at, below, dorsal, posterior, by, in, etc. for 'anatomy locations'; and so on). In the QTLdb data curation environment, multiple modifiers can be applied to a base trait. (An example with multiple modifiers is meat color lightness measured 24hr post-mortem on the semimembranosus using the CIELAB method.) This has helped to cover most, if not all, trait variants we have encountered so far. Under this management scheme, our goal is to use a set of terms over time for consistency. This implementation has improved ease of use for QTLdb/CorrDB curators by limiting the long list of sibling traits. Since we implemented the creation and management of trait variants at the experiment level (details entered once for each publication; Figure 3), it essentially moved the trait variant management out of the trait ontology term management realm (for the database as a whole; Figure 2b). The use of more MySQL tables in this implementation (Figure 2a versus Figure 2b) reflects our emphasis on the concept partitioning, compartmentalization, and relationship building, which directly result in a reduction in human effort while leaving the computer/database to handle the complexities.

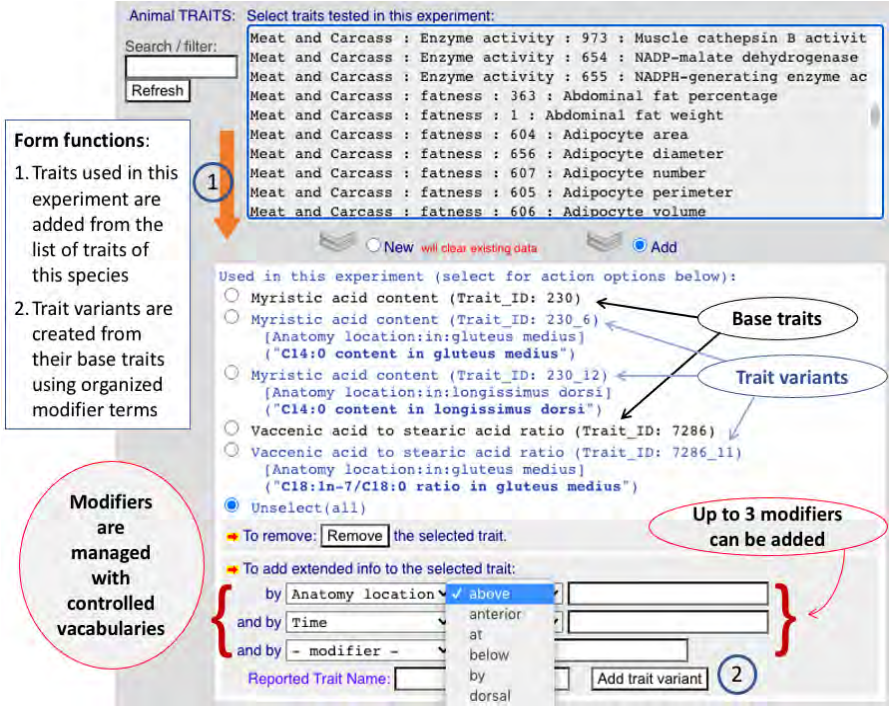


Figure 3. A web form from the implementation of the trait-variant concept showing how trait variants are created from their base traits using modifiers as controlled vocabularies.

Discussion

Our approach effectively helped reduce the lengthy list of 'modifiers' for practical use in the data curator environment, while increasing the trait term modifier system, which can be handled in the database backend once computer programs are developed. As such, this study provided a path for ontology developments by using proper concept partitioning approaches.

As with the development of trait ontologies, the trait variant system may also be subject to future developmental improvements. Because a trait term modifier system will be used well into the future and across species, its stability is important in order to ensure consistent and stable development down the road. For example, properly defining 'base traits' and 'modifiers' is important at the onset. For instance, 'MY' and '305-day MY' may be nearly interchangeable concepts, since 305-day MY is typically the default measurement standard in dairy production throughout the world. In situations like these, careful consideration is needed to weigh the advantages and disadvantages and in order to establish the most appropriate basic trait and its modifiers.

Because we implemented the creation of trait variants at the experiment level (Figure 3), curators must repeat the creation of a trait variant for each and every experiment in which it is used. This may necessitate future improvements regarding how trait variants created in one experiment may be reused in another experiment when traits are defined similarly in both experiments.

References

Götz F., Wächter T., and Schroeder M. (2012) Bioinformatics 28(12): i292–i300. https://doi.org/10.1093/bioinformatics/bts215.

Hu Z-L, Fritz E.R and Reecy J.M. (2007) Nucleic Acids Res 35(Database issue):D604-D609. https://doi.org/10.1093/nar/gkl946.

Hu Z-L, Park C.A., and Reecy J.M. (2016) Nucleic Acids Res 44(D1): D827-D833. https://doi.org/10.1093/nar/gkv1233.

Hu Z-L, Park C.A., and Reecy J.M. (2019) Nucleic Acids Res 47(D1):D701–D710. https://doi.org/10.1093/nar/gky1084.

Hu Z-L, Park C.A., and Reecy J.M. (2019) Nucleic Acids Res. 47(D1):D701–D710. https://doi.org/10.1093/nar/gky1084.

Hu Z-L, Park C.A., Wu X.L. and Reecy J.M (2013) Nucleic Acids Res 41(D1): D871-D879. https://doi.org/10.1093/nar/gks1150.

Park C.A., Bello S.M., Smith C.L., Hu Z-L., Munzenmaier D.H., *et al.* (2013) J Biomed Semantics 4(1):13. https://doi.org/10.1186/2041-1480-4-13.

Park C.A., Hu Z-L, and Reecy J.M. (2021) Livestock Product Trait (LPT) Ontology. Available at: https://www.animalgenome.org/bioinfo/projects/lpt/. Last updated: December 16, 2021.

Shimoyama M., Nigam R., McIntosh L.S., Nagarajan R., Rice T., *et al.* (2012) Front Genet 3:87. https://doi.org/10.3389/fgene.2012.00087.

346. Efficient algorithms to identify duplicated genotypes in large datasets

M.H. Ferdosi[1*], S. Masoodi[2] and M. Khansefid[3]

[1]AGBU, a joint venture of NSW Department of Primary Industries and University of New England, 2351, Armidale, Australia; [2]Computer Engineering Group, University of Payam Noor, Tehran, Iran; [3]Agriculture Victoria, AgriBio, Centre of AgriBioscience, 3083, Bundoora, Australia; mferdos3@une.edu.au

Abstract

In this paper, we introduced two novel algorithms to identify duplicated genotypes. The runtime of these algorithms was compared with the widely adopted Exhaustive Search algorithm using simulated data. We found that both new algorithms could significantly reduce the execution time. Further, the optimised Matrix Algebra Approach algorithm was faster than the Dis-Similarity lookup table and could improve the performance nearly 34 times compared to Exhaustive Search.

Introduction

Thousands of farm animals are routinely genotyped with Single Nucleotide Polymorphism (SNP) panels every month in Australia. These new genotypes are continuously added to the genomic datasets for use in genomic evaluation or parent identification. To ensure data integrity, efficient algorithms and programs are needed to assess the quality of the data in large genomic datasets. Checking for mislabelled and duplicated genotypes is the first quality control step when new genotypes are added to a dataset or combined with genomic datasets. Identifying and removing repeated genotypes is also critical before imputation and haplotype phasing. Previously developed algorithms often use fingerprinting methods to identify duplicate genotypes with an approximate approach (Jin *et al.*, 2017). In this paper, we proposed two efficient and deterministic algorithms to identify duplicated genotypes that could also handle missing and wrong genotypes.

Materials & methods

Simulated genotypes. Genotypes were simulated using QMSim V1.10 (Sargolzaei and Schenkel, 2009). Thirty chromosomes, each with 333 SNPs and a length of 100 centimorgans were simulated for 10 generations. Hence, in total 9,990 SNPs were simulated which was marginally more than Illumina BovineLD v2.0 BeadChip with 7,931 evenly spaced SNPs (recommended SNP-chip for cost-effective genotyping of cows in commercial herds). In each generation, 20 males were mated with 400 females. The genotypes across all generations (8,000 individuals) were used in this study. The genotypes of 16 individuals (every 500th animal) were duplicated. In the next step, two percent genotyping error and missing SNPs were added to the genotypes at random using R (R Core Team, 2021). The genotypes were stored in an integer matrix in which each element was 0, 1, 2 or 5, representing *aa*, *ab*, *bb* or a missing genotype, respectively.

We compared four methods for identification of duplicated genotypes.

Exhaustive search (ES). In this method three nested loops were used to check every pair of individuals. The algorithm stopped the computation, and the two individuals were considered different when the error threshold (10 percent) between the genotypes was reached. We considered this algorithm as the base method to identify duplicated genotypes.

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_346, © M.H. Ferdosi *et al.* 2022

Matrix algebra approach (MAA). This method is an extension and modification of the previously described method to identify opposing homozygotes by Ferdosi and Boerner (2014). Suppose M is an integer matrix of 0, 1, 2 and 5, for *aa*, *ab*, *bb* and missing genotypes, respectively. The rows contain the genotypes for the individuals and the columns are associated with the markers. Therefore, $M_{n\times m}$ is the matrix of genotypes, where *n* is the number of individuals and *m* is the number of markers. In the first step, a Boolean matrix for each genotype class (i.e. 0, 1, 2 and 5) should be constructed. For example, for *aa* (i.e. 0) genotypes, we need a matrix (Z matrix in Equation 1) with 1's and 0's; 1's for the presence of *aa* and 0's otherwise.

$$\begin{aligned}\mathbf{R} &= 1-(\mathbf{M}/5)\\ \mathbf{Z} &= -((\mathbf{M}-2\mathbf{J})/2)\circ\mathbf{R}\\ \mathbf{O} &= (\mathbf{J}-(\mathbf{M}-\mathbf{J})\circ(\mathbf{M}-\mathbf{J}))\circ\mathbf{R}\\ \mathbf{T} &= (\mathbf{M}/\mathbf{2})\circ\mathbf{R}\\ \mathbf{S} &= \mathbf{Z}\mathbf{Z}^{\mathrm{T}}+\mathbf{O}\mathbf{O}^{\mathrm{T}}+\mathbf{T}\mathbf{T}^{\mathrm{T}}\end{aligned} \quad (1)$$

where ∘ is the Hadamard product; R is a Boolean matrix in which each element is 0 or 1 corresponding to a missing or non-missing genotype, respectively. Z, O and T are incidence matrices containing 0's and 1's for presence of *aa* (i.e. 0), *ab* (i.e. 1) and *bb* (i.e. 2) genotypes in the M matrix, respectively. J is the matrix of ones and S is the similarity matrix that shows the number of common SNPs between two individuals. Individuals with high level of similarity (>90% of genotypes) were considered as duplicated samples. The key point for optimisation of the MAA algorithm is that Z, O and T matrices are Boolean and their cross products is symmetric. Finally, the $1-\left(\frac{\mathbf{S}}{\mathbf{R}\mathbf{R}^{\mathrm{T}}}\right)$ matrix shows the proportions of SNPs varying between individuals.

Dis-similarity lookup table (DSLT). Firstly, a Dis-Similarity Lookup Table (L) was calculated. This table of dimensions 256 × 256 consists on each axis all the combinations of four SNPs (0000, 0001, 0002, 0005, 0010, …, 5555) and the values stored are the numbers of elements different between the four combinations of SNPs (0, 1, 2, 5). For example, the dissimilarity between 0000 and 0012 is 2. The error per SNP was calculated by dividing one by the total number of SNPs, such that when accumulated across all SNPs the maximum value is one. For example, the table value when the marker matrix contains 20 SNPs between 00 and 12 is 0.05. For computational simplicity, the marker matrix was packed, P, such that each four adjacent SNPs for an animal were combined into a single number between 1 and 256. The proportion of dissimilarity between two genotypes was then calculated as

$$Mismatch_{i,j} = \sum_{n=1}^{n=N\vee m>T} \mathbf{L}\left(\mathbf{P}_{i,n},\mathbf{P}_{j,n}\right) \quad (2)$$

where *T* is the threshold for the number of mismatching genotypes, *N* is total number of segments and *m* is the number of mismatches up to the n^{th} segment. Finally, it is possible to exit this summation early if mismatch is greater than some threshold *T* (10% of segments).

Matrix algebra approach optimised (MAAO). This approach optimised the MAA algorithm further by firstly checking the 1/5 of the markers to identify the individuals which were different in at least 10% of total number markers (i.e. unduplicated genotypes). Subsequently, for the potentially duplicated individuals, all markers were checked to confirm the similarity between genotypes was greater than 90%. Hence, the algorithm did not assess all markers if the number of dissimilar markers was above the threshold in the subset of markers.

All algorithms were implemented in C++ and parallelised with OpenMP Version 5 (OpenMP Architecture Review Board, 2018).

Comparison of algorithms. The speed of the algorithms was compared using the simulated genotypes. The runtime of the different methods was measured for 500, 1000, 1,500, …, and 8,000 individuals. The codes were compiled with GNU Compiler Collection version 11 and tested on a Linux system with 64 GB of random-access memory (DDR4 2400) and Intel Central Processing Unit (CPU; Intel Core™ i7-7700K Processor, 4 cores × 4.2 GHz).

Results & discussion

Figure 1 shows the logarithm of elapsed time to identify duplicated genotypes. Both MAA and DMS algorithms were faster than the ES method, especially when the genotypes of 8,000 individuals were checked for duplicates. DMS was around 14 times, MAA was around 10 times and OMAA was around 34 times faster than ES method (all single thread). The MAA and DMS could extract information from the raw genotypes, allowing less computation during pairwise comparison. This optimisation was more obvious in DMS algorithm. Multi-threading with quad-core CPU has similar effects on MAA and DMS algorithms and increased their performances nearly 3.8 times but increased the ES method's performance almost 4.8 times because the ES method did not require any initial optimisations and could benefit from vectorisation more effectively.

Conclusions

In this article, we presented two novel deterministic algorithms for identification of duplicated genotypes. The performance of these methods can be improved further by identifying the duplicated genotypes with an approximate approach. We reported preliminary results in this paper, and further study is required to compare these algorithms to identify duplicated genotypes with the previously developed algorithms such as 'Genetic Relationship and Fingerprinting' as well (Jin *et al.*, 2017).

Acknowledgements

This study was supported by Meat and Livestock Australia project L.GEN.1704.

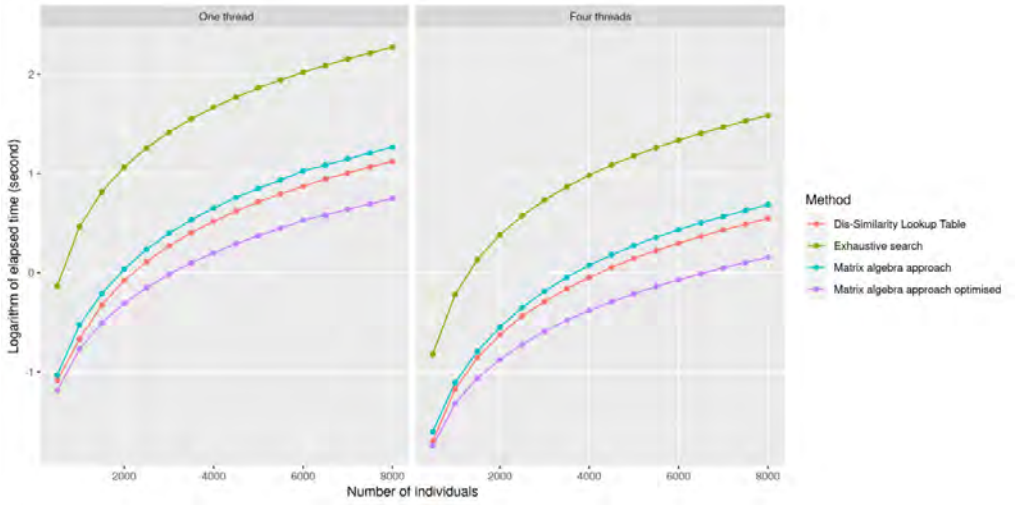


Figure 1. The figure compares the performance of four algorithms to identify duplicated genotypes.

References

Ferdosi, M. H. and V. Boerner (2014) Liv. Sci. 166:35–37. https://doi.org/10.1016/j.livsci.2014.05.026.

Jin, Y., A. A. Schäffer, S. T. Sherry and M. Feolo (2017) PloS one 12(6). https://doi.org/10.1371/journal.pone.0179106

OpenMP Architecture Review Board (2018) OpenMP application program interface version 5.0.

R Core Team (2021). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.

Sargolzaei, M. and F. S. Schenkel (2009) Bioinformatics 25(5):680–681. https://doi.org/10.1093/bioinformatics/btp045

347. Genotype error correction and the accurate identification of recombination sites

M.M. Hindle[1*] and A. Kranis[2]

[1]*The Roslin Institute and R(D)SVS, University of Edinburgh, Midlothian EH25 9RG, United Kingdom;* [2]*Aviagen Ltd, Edinburgh, United Kingdom;* matthew.hindle@rolsin.ed.ac.uk

Abstract

Measuring recombination rate and identifying hotspots requires reliable crossover detection, which is dependent on accurate phasing and genotyping. Given the low number of crossover events that occur at meiosis even low levels of genotyping and phasing errors can frustrate detection. We outline a pipeline for phasing and crossover detection and evaluate the performance across simulated populations, which contain from 0 to 9% genotyping errors. We present a novel algorithm for genotype correction in pedigreed populations, genoFix, which combines a Bayesian network estimation of Mendelian errors with an empirical LD approach. The algorithm typically achieved a precision of >90% and recall of >80% at 1% error for a chicken population with broiler-type haplotype diversity and was still effective even at atypically high rates of diversity, using populations simulated from randomized founders. Correction of genotyping errors improved both phasing accuracy and crossover detection.

Introduction

Genotyping errors (GEs) lead to downstream errors in phasing and crossover detection. Existing approaches like PedCheck can predict single Mendelian errors using likelihood odds ratios (O'Connell and Weeks, 1998) which can be masked or corrected. Bayesian Network models (BNs) are an attractive approach to modelling the Mendelian probability of genotype inheritance and SoftCSP (Sanchez *et al.*, 2007) describe a weighted constraint networks approach to the detection of GEs and their correction with BNs. GEs can also be identified and corrected based on linkage disequilibrium (LD) probabilities (Wang *et al.*, 2019), which could be a complementary approach to the Mendelian inheritance approach. We present the genoFix algorithm which combines both these approaches. We demonstrate that the subsequent reduction in GEs improved the accuracy of phasing and crossover detection.

Methods

There are four key stages in the workflow (Figure 1a). We simulated two genotype populations from two founder groups, the first set of founders had broiler type (BT) haplotype diversity and the second had high haplotype diversity. Each population was seeded from 405 founder haplotypes across 13 levels of genotyping error between 0-9%. The first 5,000 snps on chromosome 1 of simulated datasets were corrected using the novel algorithm genoFix and the performance evaluated. We then phased haplotypes for both uncorrected and genoFix corrected genotypes using AlphaImpute2 and evaluated the impact of GE and genoFix correction on phasing. Recombination was detected using the xOverDetect algorithm. We evaluate the impact of GE and genoFix on the accuracy of crossover prediction.

Simulating genotypes. All simulated populations were generated from phased founders using the qGsim package (github.com/andreaskranis/quickgsim). To simulate the BT population, we used a complete three generation pedigree of 3,902 chickens genotyped with a 50K SNP chip and phased with AlphaImpute2 (Whalen and Hickey, 2020). From the third generation 405 phased haplotypes were selected as founders. A population of 7,802 chickens was generated by dropping the founders' haplotypes onto a 5-generation real pedigree. The number of simulated crossovers was randomly sampled from a Poisson distribution with a mean equal to the size of chromosome in morgans and resampled if zero. Crossover locations were

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_347, © M.M. Hindle and A. Kranis 2022

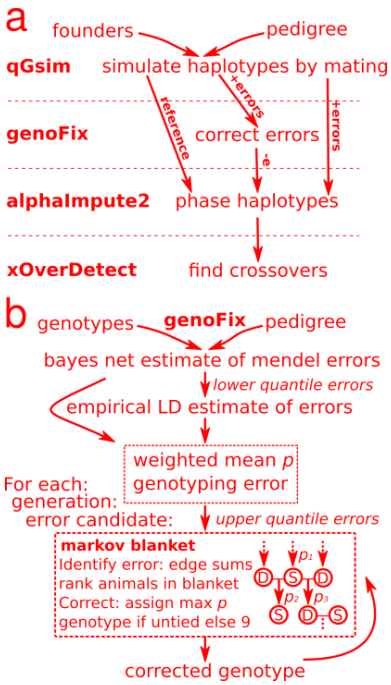


Figure 1. Overview of (a) workflow stages for crossover detection and (b) the main steps in the genoFix algorithm for identifying and correcting genotype errors.

uniformly and randomly distributed across the genome. Datasets with simulated errors were created for GE from 0-9%. Haplotypes, genotypes and crossovers were exported from qGSim before and after errors were introduced so that an exact reference for calculating performance were retained for each dataset. The simulation was repeated for a high haplotype diversity founder set by randomising the phasing of heterozygous SNPs among the founders.

Correcting genotype errors. We developed genoFix (github.com/andreaskranis/genofix) using a combined Mendelian and Empirical LD approach (Figure 1b) for estimating the probability of a genotype error P(E). All parameters listed below were used across all datasets and selected based on parameter grid searches to maximise f-score based on the 1% GE dataset. BNs were implemented using pgmpy (Ankan and Panda, 2015). We constructed Markov blankets for each animal from all breeding pairs and their offspring and include two back generations from the pedigree. For each blanket a BN was constructed for all ancestors where the conditional probability tables on nodes were calculated from Mendelian inheritance probabilities. Missing genotype values and upper terminal nodes in the blanket assume probabilities from the populations' allele frequency. This was then used to calculate the genotype state probabilities P(G) for a given SNP in an animal from the ancestors. An offspring derived P(G) was calculated from mean P(G) of Mendelian inheritance across all children. The final Mendelian SNP P(G) is the product of these ancestor and offspring estimates. We estimated P(E) as the difference between the maximum probability for a genotype and the observed probability. An empirical based LD was estimated for SNPs using a sliding windows of 9 SNPs around each SNP. Observations from the upper 20% quantile of predicted Mendelian errors were excluded from empirical calculations. For a given SNP the empirical observation were calculated as the weighted sum of genotypes combinations observed in the outermost window and inner nested windows. The weight function was defined for a window (W) where W_n is the observed set and W_o is the outermost

window. Observations are calculated for a SNP window by summing over the weighted observations of all symmetric inner windows down to a minimum window size of three (Equation 1).

$$empiricalobservations = \sum_{n=1}^{|W_o|} obs\,(W_n) \times \frac{1}{|W_o| - |W_n|} \quad (1)$$

We observed that true GEs indirectly increase the P(E) estimates on surrounding nodes in a blanket. In order to discriminate these direct and indirect effects we implemented a graph based ranking system where the rank of a node (SNP in an individual) is the sum of P(E) weights on incoming and outgoing edges (pedigree inheritance). We evaluate only the upper 40% quantile of P(E) estimates, and flag SNPs as errors if they rank first within a Markov blanket and have a P(E) value >0.6. If the most probable genotype state has a P(G) that is 1.4 times greater than the next state we correct the genotype, else we label it as unknown. Iterative correction proceeds per generation, such that each generation corrected includes genoFix corrections from the previous generations. BN calculations run in parallel on GPU and CPU and are scalable for high performance compute.

Phasing and crossover detection. For pedigree assisted phasing of haplotypes with AlphaImpute2 we used a conservative peeling threshold of 0.99, 100 SNP phasing consensus window, 20 phasing cycles, 400 phasing particles and specified % error. We developed xOverDetect to identify crossovers in phased data. A fault-tolerant mode evaluates support for SNP blocks flanking a crossover, requiring >2 heterozygous SNPs and >1% heterozygosity.

Results

We observed that at broiler type (BT) heterogeneity levels, as found in a typical chicken breeding line, genoFix can reduce errors in genotyping array data (Figure 2a,c). At 1% GE the algorithm has 97% precision and 84% recall. The high precision value results in <0.05% of new errors introduced by genoFix (Figure 2b). Figure 2b,d shows the same accuracy statistics for a simulated dataset with atypically high heterogeneity levels, designed to explore the upper tolerance of the algorithm, and its dependence on LD. At 1% GE high heterogeneity reduces the performance to 91% precision and 43% recall. In both dataset precision is more stable across higher error levels than recall. GEs below 1% suffer a small penalty in both precision and recall, which is due to fixed parameters that were optimized for 1% GE.

The reduction in GEs due to genoFix improves the downstream accuracy of AlphaImpute2 (Figure 2e,f). This was critical for datasets of >4% GE where phasing accuracy decreases rapidly. GenoFix was still able to reduce GE for these datasets. At all GE levels >1% genoFix improved phasing and reduced both false positives (fp) and false negatives (fn) crossovers (Figure 3a,b). Error tolerance in crossover detection reduced fp but with a resulting increase in fn. The fp rate in crossover detection after genoFix was slightly increased For GEs at 0.5 and 0.75% with marginal or little improvement in fn rate. This reflects the penalty to genoFix precision and recall when using parameters optimized for 1% GE.

Discussion

In these datasets accurate phasing with AlphaImpute2 required GEs below a critical 4% (Figure 2). GenoFix was able to improve the genotyping accuracy of datasets and rescue simulated genotype arrays with >4% error, even in a population with high heterogeneity. Crossovers are rare but strongly localized in the genome (Groenen *et al.*, 2009), with a median vertebrate rate of less than one crossover per Mb (Wilfert *et al.*, 2007). At these rates fp are highly problematic and even 3% GEs could add unacceptable noise (~one fp per 5Kb) to hotspot analysis. High GE rates require large sample populations to achieve adequate signal-to-noise ratios for lower intensity hot-spots. The genoFix was able to improve the accuracy of crossover detection for simulated datasets with >0.75% GE. Genotype correction was shown to be a valuable step within the phasing and crossover detection workflow.

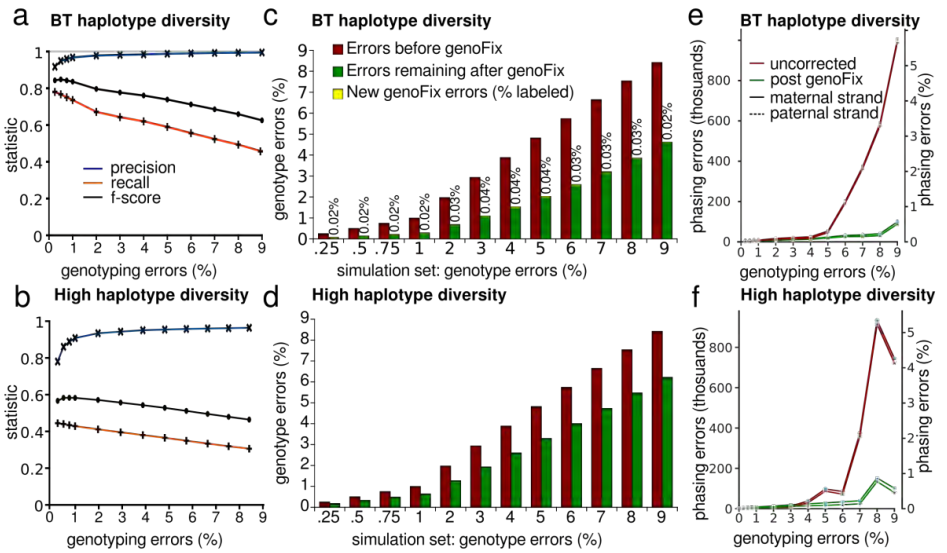


Figure 2. Evaluation of genotype genoFix correction of simulated datasets generated from BT (a,c) and high haplotype diversity (b,d) founders. For both dataset (a,b) precision and recall for genotype correction and (c,d) errors before and after genoFix. Finally, impact of genotype correction on alphaimpute2 phasing accuracy for simulated datasets for (e) BT founders and (f) high haplotype diversity founders.

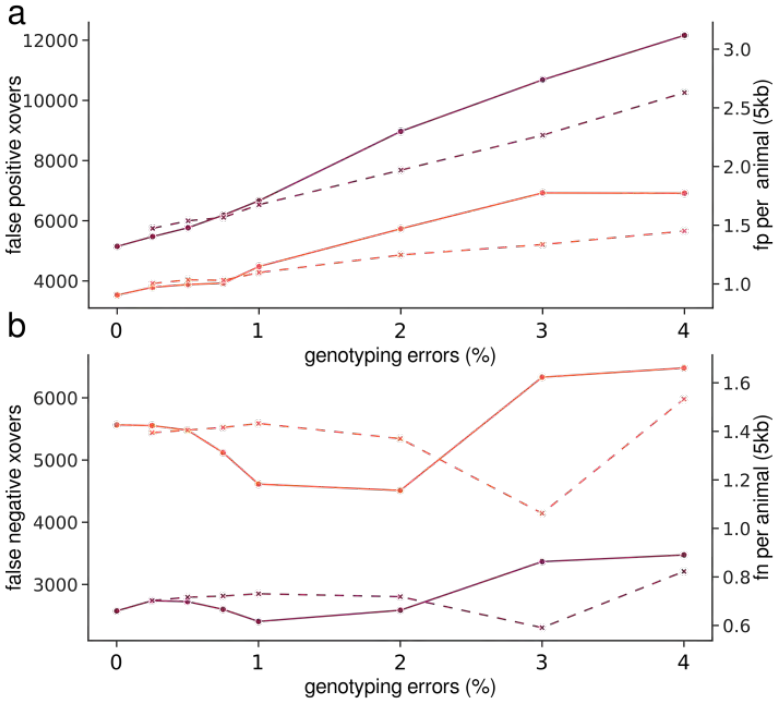


Figure 3. Crossover detection accuracy between 0-9% simulated GEs, before and after genoFix correction. Data is simulated from BT founders. (a) False positives (fp) are crossovers predicted that are not overlapping real crossovers. (b) False negatives (fn) are real crossovers that are not overlapping predicted crossovers.

References

Ankan, A., Panda, A., 2015. Proc. of SciPy, Austin, Texas, pp. 6–11. https://doi.org/10.25080/Majora-7b98e3ed-001

Groenen, M.A.M., Wahlberg, P., Foglio, M., Cheng, H.H., Andersson, L., *et al.* 2009. Genome Res 19, 510–519. https://doi.org/10.1101/gr.086538.108

O'Connell, J.R., Weeks, D.E., 1998. Am. J. Hum. Genet. 63, 259–266. https://doi.org/10.1086/301904

Sanchez, M., de Givry, S., Schiex, T., 2007. Proc. of ACIA. IOS Press, NLD, pp. 29–37.

Wang, X., Lund, M.S., Ma, P., Janss, L., Su, G., *et al.* 2019. J. Anim. Sci 10, 8. https://doi.org/10.1186/s40104-019-0315-z

Whalen, A., Hickey, J.M., 2020. bioRxiv https://doi.org/10.1101/2020.09.16.299677

Wilfert, L., Gadau, J., Schmid-Hempel, P., 2007. Heredity 98, 189–197. https://doi.org/10.1038/sj.hdy.6800950

348. A phasing algorithm targeted at dairy cattle populations having genotyped ancestors with many genotyped offspring

M. Dahl* and O.F. Christensen

Center for Quantitative Genetics and Genomics, Aarhus University, 8830 Tjele, Denmark; mdahl@qgg.au.dk

Abstract

In dairy cattle, the prevalence of genomic testing is now widespread, and commonly both sires and many offspring in the form of children, grandchildren, and great-grandchildren, are genotyped. Calling phases from genotypes is useful for several reasons, in particular for construction of haplotypes and imputation of missing genotypes. In this paper, we present a new computationally efficient method to call phases of grandfathers and great-grandfathers when enough genotyped offspring are present. The method consists of first identifying shared haplotypes between ancestor and offspring, followed by phasing segments based on homozygote genotypes in offspring, and finally merging haplotype segments to call entire ancestral haplotypes. The implementation is based on using bit vectors and fast binary computations. Results show that the method is effective in finding shared haplotypes to be used for phasing.

Introduction

Phasing using homozygotes in relatives and offspring has explored by others. Kong *et al.* (2008) introduced Long Range Phasing (LRP) that phases based on likely identity by descent (IDB) segments of the chromosome, using homozygote markers of fixed length genotype segments, to call paternal and maternal alleles using surrogate parents. The method is implemented in AlphaPhase (Hickey *et al.* 2011, Money *et al.* 2020). Using offspring for phasing was presented in Ferdosi *et al.* (2014) and Nettelblad (2012). They both show that very accurate phasing can be done on parents when enough genotyped offspring are present.

The aim of our paper is to present a computationally efficient method using likely IBD segments to call phases of grandfathers and great-grandfathers when enough genotyped offspring are present.

Materials & methods

Under the assumption of one recombination per meiosis, the average length of a recombination-free sequence of markers passed from an ancestor is half the size of the shared haplotypes. These recombination-free sequences can be used for phasing. The method presented here, is based on identifying parts of the inherited chromosome in offspring, where a haplotype segment from an ancestor is inherited and no recombination has taken place. After ancestor haplotypes have been computed, the offspring genotypes are used to do internal phasing of the called haplotypes. This is done by the following sequence of operations:

1. Identify likely shared haplotypes between ancestor and offspring, using opposing homozygotes, allowing for small percentage of errors.
2. Split the chromosome into segments of equal length.
3. For each segment extract genotype segment from all offspring where segment is fully contained in likely shared haplotype. Set minimum requirements for number of offspring.
4. Split offspring genotype segments into two recombination free groups.
5. Use these two groups to call phase on the ancestor haplotypes.
6. Do internal phasing of identified haplotype segments using heterozygotes in the called haplotype segments.

Test data was generated using QMSim (Sargolzaei and Schenkel 2009). A chromosome with 3,273 markers was created, i.e. as chromosome one on the Illumina BovineSNP50 BeadChip v3. A nucleus population

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_348, © M. Dahl and O.F. Christensen 2022

producing bulls and a production population of cows was created. Then 10 generations of dams in the production population were mated to 10 generations of sires from the nucleus population. Offspring in generation 10 was used for phasing, for grandfather phasing generation 9 sires were used, for great-grandfathers, sires in generation 8 were used. Random positions in genotypes were removed or changed to simulate missing genotypes and genotyping errors.

Step 1: identifying shared haplotypes. Genotypes coded as 0,1,2. Parameters: segment length 8; minimum length 24 SNPs; error window size 16; max errors in window 3; min no opposing length 80; max. error percentage 3.0%. Method: split genotypes in segments of a length that enables fast computation – in this example 8; compute opposing and comparisons; merge no opposing; remove too many opposing and too short; recursive computation of longest shared no opposing that has less than 3% errors.

Fast identification of shared haplotypes is important for obtaining acceptable runtimes. The method stores genotypes as three bit vectors aa, bb and ab, i.e. a bit set mark whether the genotype has AA, BB or AB SNP in the bit position. This enables use of binary instructions and, or, not or x or, when comparing genotypes. These instructions are very fast in most programming languages. Opposing homozygotes between two genotypes can be computed as the number of bits set in the expression: $((aa_1 \text{ and } bb_2)$ or $(aa_2 \text{ and } bb_1))$.

The method is as follows: (1) split the two genotypes into segments with fast binary instructions for fast computing; (2) create arrays of aa, bb, and ab bit segments for all genotypes; (3) do a pairwise comparison between segments by computing number of opposing and number of SNPs present in both genotypes; (4) merge adjacent segments with zero opposing homozygotes; (5) set an error window size and the maximum allowed number of errors in the window; (6) scan all adjacent segments with non-zero opposing homozygotes. If adjacent segments with the length of the window or less, contain more than the maximum allowed number of errors, then remove the adjacent segments from the array; (7) set a minimum length of pairwise shared no opposing haplotypes between the two genotypes and a maximum error percentage; (8) group segments into groups of adjacent segments. Remove all groups with a total length less than the minimum length; (9) for the remaining groups, compute the maximum length of no opposing homozygotes using elements in the group that have fewer errors than the maximum error percentage. This can be done by recursion. Return the longest identified. The method is illustrated in Figure 1.

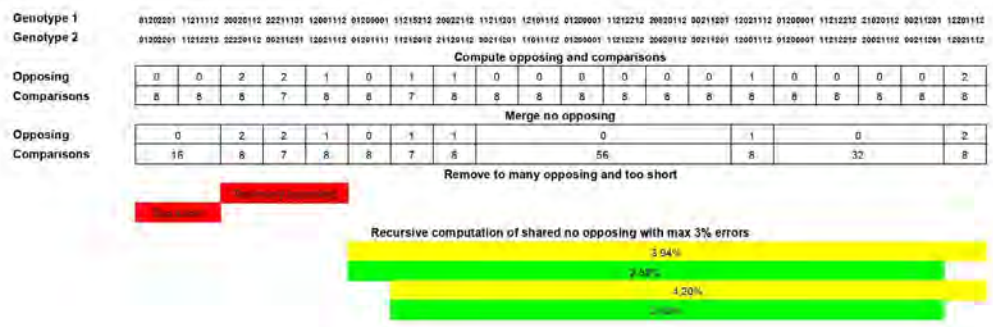


Figure 1. Identification of long shared no opposing SNPs between two genotypes.

Steps 2 to 5: phasing segments based on homozygotes in offspring. Following step 1, all pairs of offspring and sire have had their shared haplotypes computed. An offspring and a sire may share haplotypes from a shared common ancestor. These can be mistaken as the inherited haplotype. To decrease the possibility of this, we chose segments that are sufficiently long. In the experiments presented here, the simulated genotypes were split into 12 segments. Offspring where an ancestor is grandfather or great-grandfather more than once will not be used for phasing. To get variation, pedigrees of all the offspring must have the grandfather or great grandfather as the only ancestor common in offspring pedigrees.

For each of the segments the method does the following: (1) find all elements where the segment is entirely in the shared haplotype and enough offspring exist – store them; (2) split them into the two largest groups that split, around the heterozygotes in the ancestor's genotype. Allow for errors; (3) call phases using homozygotes in the two groups and homozygotes in the ancestor. Allow for errors. In steps 2 and 3 fast bit operations are used.

Step 6: internal phasing of the computed haplotype segments. The haplotype segments phased in step 2 to 5 has no internal phase. In this section a method to call internal phase is described.

The method consist of: (1) Ensure that all haplotype segment pairs have 3 heterozygotes, by doing the following: In all ancestor haplotype pairs with less than 3 heterozygotes, delete heterozygote calls; Merge all homozygote haplotype pairs to an adjacent segment pair. (2b) For each pair of adjacent haplotype segments, called left segment and right segment: Find the three rightmost heterozygote positions in left haplotype pair; Create two short haplotypes with three heterozygote positions; Similar for the right haplotype pair. (3) Find the internal phase of left and right haplotype pair making a vote between all offspring segments that contains all the positions in the two short haplotypes from the left and right haplotype pair.

The proposed method was implemented in a Python3.10 program where bit-vectors are stored in a byte array, that support fast binary computations.

A test was run on a HP notebook (Intel i7-8,565U CPU – 16GB RAM, cache size 8,192 KB). On that hardware on average 1,468 genotype comparisons per second were executed, during the test run. Parameter's settings in program test were as follows. Missing marker percent in genotypes: 3%; Genotype SNP count: 3,273; Error window size: 272: Max opposing count in window: 6; Min no opposing length: 648; Max errors percent: 1% for 0.1 SNP error pct, 2% for 1% error pct.

Results

It is clearly seen that the identified shared haplotypes can be used to call haplotype segments in grand- and great-grandfathers while correcting genotyping errors (Table 1 and 2). In most cases the shared haplotypes can be used for calling the entire haplotype.

Table 1. Phasing of grandfathers.

Simulated offspring	Random SNP call percentage	Phased great grandfathers	Haplotype calls	Haplotype call percent	Haplotype error percent	Haplotype segment calls	Haplotype segment call percent	Haplotype segment error percent	Missing haplotype segment phases
50	0.1	1000	1,988	99.569	0.077	24,000	99.572	0.051	6
200	0.1	100	200	99.552	0.211	2,400	99.557	0.073	0
50	1.0	999	1,954	99.241	0.471	23,976	99.241	0.342	22
200	1.0	100	196	99.197	0.603	2,400	99.203	0.319	2

Table 2. Phasing of great-grandfathers.

Simulated offspring	Random SNP call percentage	Phased great grandfathers	Haplotype calls	Haplotype call percent	Haplotype error percent	Haplotype segment calls	Haplotype segment call percent	Haplotype segment error percent	Missing haplotype segment phases
100	0.1	1000	1,982	99.562	0.137	24,000	99.564	0.051	9
400	0.1	100	196	99.549	0.078	2,400	99.553	0.088	2
100	1.0	998	1,960	99.207	0.633	23,952	99.202	0.330	18
400	1.0	100	198	99.209	0.309	2,400	99.217	0.313	1

Discussion

A new and computationally fast method for identifying shared haplotypes between genotype individuals is presented, and we have shown that the method is effective in finding shared haplotypes that can be used for phasing. Future work involves an extensive comparison with other software, in particular AlphaPhase (Hickey *et al.* 2011, Money *et al.* 2020) and Fimpute (Sargolzaei, Chesnais and Schenkel (2014). Preliminary investigations show that our algorithm is superior in terms of both speed and accuracy compared to AlphaPhase, but memory usage should also be a criterion for comparison.

References

Ferdosi, M. H., Kinghorn, B. P., van der Werf, J. H. J and Gondro, C. (2014) Genet. Sel. Evol 46:11 https://doi.org/10.1186/1297-9686-46-11

Hickey J. M., Kinghorn, B. P., Tier, B., Wilson, J. F., Dunstan, N. *et al.* (2011) Genet. Sel. Evol 43:12. https://doi.org/10.1186/1297-9686-43-12

Kong A., Masson G., Frigge M. L., *et al.* (2008) Nat Genet. 40(9):1068-75 https://doi.org/10.1038/ng.216

Money, D., Wilson, D., Jenko, J. Whalen, A. Thorn, S. *et al.* (2020) Genet. Sel. Evol 52:38. https://doi.org/10.1186/s12711-020-00558-2

Nettelblad, C. (2012) BMC Genetics 13:85 https://doi.org/10.1186/1471-2156-13-85

Sargolzaei, M., Chesnais J. P. and Schenkel F. S. (2014) BMC Genomics 15:478. https://doi.org/10.1186/1471-2164-15-478

Sargolzaei, M. and F. S. Schenkel (2009) Bioinformatics 25: 680-681. https://doi.org/10.1093/bioinformatics/btp045

349. Recent advances and future needs in genotype imputation

P.M. VanRaden*, D.J. Null and A.S. Al-Khudhair

USDA, Animal Genomics and Improvement Lab, Bldg 5 BARC-West, Beltsville, MD 20705, USA;
paul.vanraden@usda.gov

Abstract

Genotype imputation has enabled selecting and using many more loci in research and prediction for much less cost than genotyping all individuals with the same array density or technology. Genotyping arrays are updated often to include new QTLs or higher-effect markers and applied to new animals without re-genotyping all previous or reference animals. That may require imputing the new loci from less-dense to higher-density arrays and from descendant genotypes to ancestors in the opposite direction of normal reference and target populations. Routine expansion and recycling of haplotype libraries and updating only individuals expected to change can speed imputation for routine predictions. This study compared imputation strategies for 78,964 loci of 4.6 million Holsteins. Sequence imputation is more challenging due to higher error rates and more complex variants, but new techniques could make sequencing cost-competitive with arrays for routine predictions.

Introduction

Genotypes at thousands of loci are combined into genomic relationships, and genomic predictions are usually obtained by multiple regression of phenotypes on genotypes. Solutions require replacing any unknown genotypes using allele frequencies, imputed genotypes, allele content (dosage), or genotype probabilities (Zheng *et al.*, 2011). Unknown phenotypes usually are easy to exclude from models but can also be 'imputed' to reduce computation using the same canonical transformation as when all traits are measured (Ducrocq and Besbes, 1993). Unknown genotypes can be imputed using pedigrees and a linear model one locus at a time from observed genotypes (Gengler *et al.*, 2007). Similar algebra uses pedigree relationships to linearly impute genomic relationships for ungenotyped animals from genotyped animals in single-step predictions (Aguilar *et al.*, 2010) and can improve marker effects by including phenotypes of ungenotyped parents, for example. More accurate haplotype-based imputation uses allele patterns across linked loci such as in long-range phasing or hidden Markov models and high-quality genome reference assemblies. This nonlinear imputation restricts genotype dosage to the range of homozygous reference to homozygous alternate (usually 0 to 2), whereas dosages from linear imputation can exceed the valid range. Known genotypes, pedigrees, and phenotypes could be used to impute the missing genotypes; however, accuracy is usually high by imputing genotypes just once without using the phenotypes, which often have low heritability and low correlations with individual genotypes. Accuracy is sometimes higher with multi-breed than single-breed imputation or using two steps instead of directly imputing all variants from the lowest density (VanRaden *et al.*, 2013). Imputation strategies for large datasets must balance accuracy with computing costs and adapt to properties such as array densities, sequence depth, error rates, and population structure in the input data. Imputation methods developed in animal breeding are often hundreds of times faster than from human genetics but with similar accuracy, especially with pedigree available (Sargolzaei *et al.*, 2014; Miar *et al.*, 2017). By processing in birth date order, most progeny haplotypes can be quickly selected from the two haplotypes of each genotyped parent or grandparents instead of from a long list of population haplotypes. This study provides a speed and accuracy comparison and a brief overview of genotype imputation strategies.

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_349, © P.M. VanRaden *et al.* 2022

Materials & methods

Microarray genotypes for 78,964 SNPs of 4.6 million Holsteins were imputed using Findhap version 3 either with no prior haplotype library or by obtaining priors from genotypes of a subset population that included 369,063 bulls plus their 73,813 dams, followed by one iteration including all animals. About 30,000 newly genotyped animals are imputed weekly using the prior haplotype library from the previous monthly update. The weekly imputation is nearly as accurate as the full monthly, with prediction correlations near 1.0 in larger breeds (Wiggans *et al.*, 2015). Each month the whole file is reprocessed to add about 120,000 new genotypes, update about 5,000 previous animals whose genotype or pedigree changed and obtain priors for the next weekly and monthly updates. Only the new and changing animals get their haplotype numbers updated, but any missing alleles within each haplotype can be filled by the added data and improve the imputed genotypes of other animals. The maximum length of haplotypes processed was reduced recently from 700 to 250 SNPs to speed imputation and limit the memory needed. The list of usable SNPs is updated about once per year. Routine evaluations with expanding reference populations must decide how frequently to reimpute all genotypes, such as when new QTLs are added or which animals to update, for example, only the most recent or those whose pedigrees change. Such strategies were compared using Holstein genotypes from the Council on Dairy Cattle Breeding (CDCB) database that included 48 different arrays ranging from <3,000 to >600,000 usable markers.

Results

Imputation using no prior haplotype library and four iterations would take 24 days with 30 processors and up to 500 Gb memory. Time was reduced to nine days by first obtaining a prior haplotype library from only the bulls' and dams' genotypes. Each week, about one hour is required to impute genotypes of new animals using the previous month's haplotype library. About 10 hours are required once per month to reprocess the whole file and include the new animals. Input genotypes averaged 72% unknown, whereas only 0.03% were still unknown after imputation, and 1.7% were half known with one allele still missing. Allele frequencies are later substituted for the missing alleles when estimating SNP effects. The list of 78,964 SNPs includes about 80 known QTLs already genotyped plus locations of six more discovered QTLs expected to be available on future arrays to allow imputing those without starting the imputation from scratch. Some recently selected high-effect SNPs from the high-density (HD) array are genotyped for only about 5,000 animals (0.1% of the population) currently. Still, imputation fills nearly all (>99%) of the missing alleles. Other SNPs from the HD array, especially on the X chromosome, were added to fill gaps when converting to the ARS-UCD1.2 cattle reference map (Rosen *et al.*, 2020). Such SNPs and QTLs are provided to genotyping laboratories and included when designing their future arrays.

Discussion

Advances in imputation allow combining various datasets, but highly accurate prediction may require identifying more of the QTLs and genotyping those in both the reference and candidates for selection instead of relying on markers. Array genotypes are typically so accurate that animal and plant breeders rarely store their quality scores based on distance from cluster centroids, for example. Uncertain genotypes are instead set to missing, and markers or samples with high missing or error rates are not used so that a low, uniform error rate of 1% or less may be assumed for all array genotypes remaining. That strategy allows efficient computation and storage of input data, but output files are much larger after imputation and include many uncertain genotypes. Sequence genotypes often are less accurate due to more read errors and both alleles not being observed across the whole genome, such as in regions with low coverage or where alignment to the reference assembly is challenging. Imputation to sequence can discover better markers or QTL, but high accuracy is needed for imputed sequence variants to outcompete highly linked array genotypes directly measured in the reference population. Rare or less frequent alleles may have poor imputation accuracy and may thus not improve prediction even if their actual effects are larger than

nearby genotyped markers (Zhang *et al.*, 2018). For example, the best linked marker from an array may be correlated by 0.97 with the true QTL but have nearly 100% correct genotypes compared to only 95% correctly imputed QTL genotypes for animals not sequenced. Reference cows may contribute much less to the reliability of prediction or marker selection if their sequence genotypes are imputed from only 6,000 to 20,000 usable SNPs and their individual phenotypes have low heritability. To obtain the gains in prediction accuracy expected from larger reference populations, well-designed genotyping strategies and accurate imputation are needed (Judge *et al.*, 2017).

Variant calling, phasing, and imputation can be combined so that each sample has only two haplotypes per region, and sequence read errors are suppressed instead of called as additional variants. Only two alleles are reported with array genotypes, whereas multiple alleles up to seven often are reported at the same locus in sequence data. Using such variants may require software redesign to exclude the less common alleles or track allele definitions within locations within chromosomes. A location can have all four bases (A, C, G, T) as SNPs plus multiple insertions and deletions of varying lengths. Copy number variants are harder to define, call, and impute than SNPs or indels and have not yet contributed much to genomic prediction (Chen *et al.*, 2021) despite some large, known effects. Standard variant call format (vcf) can require much space and be difficult to read: 'Do not write home-brewed VCF parsing scripts – it never ends well' (GATK Team, 2021). Simpler file formats may be possible that more languages could read with code of the user's choice. When merging separate vcf files, information about less common alleles can be lost for variants not reported if all samples in a single file are homozygous for the reference allele or if many samples have insufficient coverage to detect the allele. Instead of merging all raw sequence data, vcf files could each include and report genotypes and read depth at an agreed list of common locations. This strategy is like all genotyping arrays that include a standard set of 6,000 or 50,000 markers to simplify later merging with other arrays or files. The genomic vcf (gvcf) format instead takes more space to store approximate read depth for all genomic locations.

Low-coverage sequencing could soon become less expensive than arrays, allowing for genotyping many more variants with high accuracy in populations with good sequence reference panels available (Rubinacci *et al.*, 2021) and can also be used in developing those reference panels (VanRaden *et al.*, 2015). Methods to account for read errors in low-coverage data are needed (Ros-Freixedes *et al.*, 2020), and some of the read error bias can be overcome by aligning to both the reference genome and the alternate genome simultaneously (VanRaden *et al.*, 2019). The posterior genotype probabilities reported in vcf usually assume prior probabilities of 0.25, 0.5, and 0.25 and could be recalculated with actual genotype frequency priors (p^2, 2pq, and q^2) to obtain better dosage estimates. Distributed processing works well for arrays with accurate genotypes at predefined locations, but centralized processing may be needed for low-coverage sequences instead of sending imputed genotypes and updating those as reference populations improve and more breeds are added. Storage is a much bigger issue with raw sequence data and many rare alleles detected as more samples are sequenced. Genotypes could be stored for only the alternate alleles instead of storing all the homozygous reference genotypes. Similarly, referential compression can produce cram rather than bam files by storing only the differences of the raw reads from the reference assembly (Shi *et al.*, 2019). Larger populations have been genotyped for humans than livestock, and >300 million sequence variants have been imputed for >20 million humans (TOPMed Imputation Server, 2021). In animal genetics, the goal is not to routinely store or impute more variants but to identify the most useful variants that improve prediction at a reasonable cost. Advances in genotyping and imputation will continue to make genomic selection more affordable and accurate in future livestock populations.

References

Aguilar I., Misztal I., Johnson D.L., Tsuruta S., and Lawlor T.J. (2010) J. Dairy Sci. 93(2):743-752. https://doi.org/10.3168/jds.2009-2730

Chen L., Pryce J.E., Hayes B.J., and Daetwyler H.D. (2021) Animals 11:541. https://doi.org/10.3390/ani11020541

Ducrocq V and Besbes B (1993) J. Anim. Breed. Genet. 110, 81-92. https://doi.org/10.1111/j.1439-0388.1993.tb00719.x

GATK Team. (2021) VCF variant call format. https://gatk.broadinstitute.org/hc/en-us/articles/360035531692-VCF-Variant-Call-Format

Gengler, N., Mayeres P., Szydlowski M, *et al.* (2007) Animal 1:21-28. https://doi.org/10.1017/S1751731107392628

Judge M.M., Purfield D.C., Sleator R.D., and Berry D.P. (2017) J. Anim. Sci. 95(4):1489–1501. https://doi.org/10.2527/jas.2016.1212

Miar Y., Sargolzaei M., and Schenkel F.S. (2017) J. Dairy Sci. 100(4): 2837-2849. https://doi.org/10.3168/jds.2016-11590

Ros-Freixedes R., Battagin M., Johnsson M., Gorjanc G., Mileham A.J., *et al.* (2020) Genet. Sel. Evol. 50:64. https://doi.org/10.1186/s12711-020-00536-8

Rosen B.D., Bickhart D.M., Schnabel R.D., Koren S., Elsik C.G., *et al.* (2020) GigaScience 9(3):giaa021. https://doi.org/10.1093/gigascience/giaa021

Rubinacci S., Ribeiro D.M., Hofmeister R.J., and Delaneau O. (2021) Nat Genet. 53:120–126. https://doi.org/10.1038/s41588-020-00756-0

Sargolzaei M., Chesnais J., and Schenkel F. (2014) BMC Genomics 15:478. https://doi.org/10.1186/1471-2164-15-478

Shi W., Chen J., Luo M., and Chen M. 2019. Bioinformatics 35(12):2058–2065. https://doi.org/10.1093/bioinformatics/bty934

TOPMed Imputation Server (2021) Nat. Inst. of Health https://imputation.biodatacatalyst.nhlbi.nih.gov/#!pages/home

VanRaden P.M., Bickhart D.M., and O'Connell J.R. (2019) J. Dairy Sci. 102(4):3216–3229. https://doi.org/10.3168/jds.2018-15172

VanRaden, P., Null, D., Sargolzaei, M., Wiggans, G., Tooker, M., Cole, J., *et al.* (2013). J. Dairy Sci. 96, 668–678. https://doi.org/10.3168/jds.2012-5702

VanRaden P.M., Sun C., and O'Connell J.R. (2015) BMC Genet. 16:82. https://doi.org/10.1186/s12863-015-0243-7

Wiggans G.R., VanRaden P.M., and Cooper T.A. (2015) J. Dairy Sci. 98(3):2039–2042. https://doi.org/10.3168/jds.2014-8868

Zhang Q., Sahana G., Su G., Guldbrandtsen B., Lund M.S., and Calus M.P.L. 2018. Genet. Sel. Evol. 50:62. https://doi.org/10.1186/s12711-018-0432-8

Zheng J., Li Y, Abecasis G.R., and Scheet P. (2011) Genet Epidemiol. 35(2):102–110. https://dx.doi.org/10.1002%2Fgepi.20552

350. Imputation of low coverage sequence data by utilization of pre-phased reads and HBimpute

T. Pook* and J. Geibel

University of Göttingen, Center for Integrated Breeding Research, Animal Breeding and Genetics Group, Albrecht-Thaer-Weg 3, 37075 Göttingen, Germany; torsten.pook@uni-goettingen.de

Abstract

The generation of high-quality genomic data for large populations is an incremental part of animal breeding. In this work, we are proposing an extension for the imputation software HBimpute, which was originally developed for low coverage sequence data from inbred lines in plant breeding. For this, sequencing reads of animals are phased based on genomic information from the sire to generate two 'pseudo'-inbred individuals that are separately processed in the imputation pipeline. The performance of the approach is demonstrated using a simulated data set of 520 chickens from a crossbreeding scheme. Absolute discordance rates to the real simulated genotype are much smaller (0.0303) than for state-of-the-art software using genotype calls (Beagle 5: 0.0887) or genotype likelihoods (0.0564). In particular, performance for rare variants is better than in traditional software.

Introduction

In the last decade, genotyping has been shown to be an incremental part of animal breeding. With the increasing availability of genotyping arrays, large-scale genotyping of breeding populations in various livestock species has become possible. Genotyping in animal breeding nowadays is usually done via SNP arrays and only selected animals are sequenced to use those animals as a reference panel for imputation to sequence level marker density (Bouwman *et al.*, 2018). Although the cost of sequencing substantially dropped in the last few years, costs and bioinformatic processing times are still too high for use in routine breeding programs. Cost reduction may be achieved by low-depth sequencing, which however is error-prone for heterozygous variants. Recently, Pook *et al.* (2021) suggested a new imputation approach for the imputation of inbred material in plant breeding that was sequenced at low coverage, yielding substantial improvements by the use of haplotype blocks from the software HaploBlocker (Pook *et al.*, 2019) and utilizing local read-depth in the imputation algorithm. This approach, however, is not directly applicable in animal breeding, as the identification of haplotype blocks requires accurate phase information, which is usually not given in livestock populations.

In this work, we propose an extension of the work of Pook *et al.* (2021) to perform read-phasing in advance of the variant calling. The key idea of the approach is to use genomic information from the sire animal that was sequenced with high sequencing depth, as the number of sires in animal breeding is usually limited. On the contrary, the number of dams in animal breeding is usually high, making approaches for pre-phasing in trios (Koren *et al.*, 2018) not applicable.

Materials & methods

Materials. Genotypic data was generated via simulations with the MoBPS software (Pook *et al.*, 2020) based on a real dataset of 40 broiler chickens (subset of Geibel *et al.* 2021) with 319,479 SNPs on chr20, simulated with a size of 2.32 Morgan. To mimic the genetic architecture of a crossbreeding scheme with different gene pools, two gene pools with 10 roosters and 250 hens each were generated. Next, 25 generations of mating within the respective pool were simulated. Subsequently, the actually used animals were obtained by mating a sire from the first gene pool with a dam from the second gene pool. Low coverage sequence data was simulated to use reads with a length of 2×150 base pairs with a target coverage of 1X. Additionally,

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_350, © T. Pook and J. Geibel 2022

calling errors with a frequency of 0.02 were simulated. Haplotypes of the sires were assumed to be perfectly known. The interested reader is referred to https://github.com/tpook92/MoBPS for scripts and details of the simulation.

Methods. HaploBlocker (Pook *et al.*, 2019) requires imputed and phased data. As in particular the phasing accuracy in traditional imputation software in low coverage sequence is far below what is needed for HaploBlocker, HBimpute cannot be directly applied. The core idea of our pipeline is the utilization of a typically small number of sires in animal breeding programs, allowing for the generation of high quality data of one of the parents of each animal. Based on this, each generated read for an animal is evaluated if the corresponding haplotype matches with one of the two haplotypes of the sire animal. Subsequently, the reads of an animal are split into two groups based on whether they correspond to a sire haplotype or not.

Instead of using the genotype of the animal itself in the imputation, two 'pseudo'-inbreds are generated by performing the SNP calling for the reads of the two haplotypes separately. Subsequently, the HBimpute pipeline as suggested in Pook *et al.* (2021) was applied.

Results of our imputation pipeline are compared against different versions of the imputation software Beagle (Browning *et al.*, 2018; Browning and Browning, 2013) using genotype hard calls (GT), genotyped likelihood (GL), or the same 'pseudo'-inbred panel as used in our suggested HBimpute pipeline. Other methods that are specifically designed for the imputation of low coverage sequence data are not presented in this manuscript as they were not suited for the available data structure. For example, QUILT (Davies *et al.*, 2021) requires a reference panel that includes haplotypes from both parents, and GLIMPSE (Rubinacci *et al.*, 2021) performed substantially worse in general in all considered tests.

Results & discussion

Quality of the pre-phasing. After pre-phasing, genotype calls for 55.1% of all cells of the genotype matrix were obtained. Note here that on average more genotype calls for the haplotype stemming from the sire were obtained as e.g. regions of identity-by-state between sire and dam haplotype are automatically assigned to the sire haplotype. The average read-depth for the sire haplotype is 1.13X compared to 0.66X for the dam haplotype with genotype calls for 65.4 / 44.8% respectively. 10.5% of all reads are not assigned to either parents. Of all assigned reads stemming for the sire, 93.7% are correctly assigned to the sire haplotype, while only 56.2% of the dam reads are assigned to the dam haplotype (overall 74.8% correct assignment).

Performance of the imputation methods. Overall, our suggested HBimpute pipeline yielded the best imputation performance both in terms of correction / imputation accuracy (0.397) and discordance rates between the imputed and the underlying real dataset (0.0303; Table 1). The use of GT led to substantially worse results, as the low number of reads per variant makes the calling of heterozygous variants challenging. Although this is slightly improved when using GL information in Beagle 4.1, overall imputation performance is still substantially worse than in HBimpute. The use of Beagle 5.0 GT with the use of our pre-phasing method and 'pseudo'-inbreds performed much better than all considered standard Beagle pipeline (0.378; 0.0359), thus indicating the potential of pre-phasing based on sire information.

Imputation quality based on the allele frequency. When analyzing the performance of the different imputation pipelines based on the frequency of the alleles, we observed that both HBimpute and Beagle 5.0 GT with pre-phasing perform well on all allele frequencies, while Beagle 4.1 GL seems to have a strong bias towards the more common variant, with even lower error rates for alleles with a frequency higher than 0.6 (Figure 1), thus indicating that pre-phasing is particularly useful for the imputation of rare variants.

Table 1. Performance of the different imputation pipelines based on imputation accuracy and discordance rates of the imputation.

Imputation method	Imputation accuracy	Discordance rate	Share heterozygosity
HBimpute	0.397	0.0303	0.190
Beagle 4.1 (GL)	0.304	0.0564	0.127
Beagle 4.1 (GT)	0.341	0.0812	0.082
Beagle 5.0 (GT)	0.317	0.0887	0.085
Beagle 5.0 (GT) + pre-phasing	0.378	0.0359	0.195

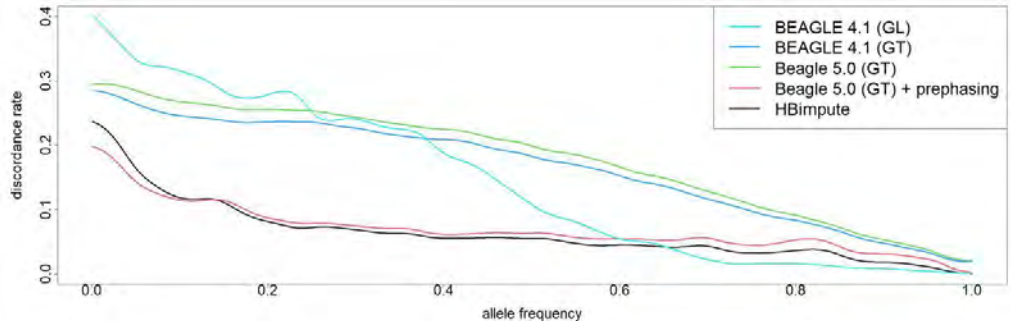


Figure 1. Discordance rates of the imputation depending on the used imputation pipeline.

When comparing the performance within HBimpute based on the number of reads available, we observed discordance rates of 0.0452 when having no reads, 0.0232 when having exactly one read available, 0.229 when having two reads, and 0.228 when having three or more supporting reads. When instead of comparing the discordance of the genotype, the qualities of the two haplotypes of an animal were analyzed separately, we observe discordance rates of 0.0199 for the sire haplotype and 0.0465 for the dam haplotype, indicating a much higher quality of the sire haplotype. A potential reason for this is that wrongly assigned reads to the sire do not negatively impact the pipeline as they are usually assigned based on IBS while bad-quality reads are preferentially assigned to the dam haplotype.

Imputation quality based on the number of used reads. In addition to the genotype call itself, HBimpute is also providing information on the number of reads used to make the call. This can both be seen as a quality indication and tool for the detection of copy number variation. In regions without available haplotype block or own read, discordance rates are 0.130 compared to 0.028 / 0.023 / 0.021 / 0.018 when having at least 1 / 10 / 50 / 250 supporting read. For the sire haplotype only 4.5% of all calls are using less than 10 reads while for the dam haplotype 17.1% of all calls are done with low read support. As only 10 different sires were used in the generation, haplotypes naturally occur in higher frequency.

Conclusions

Our suggested pipeline enables the use of HBimpute for non-inbred genetic material. By this, and in contrast to other imputation approaches, our approach is able to utilize the local read-depth and phase information from the sire animal. Thus, neither genotype information from the dam nor high sequencing depth for the individual itself is required to obtain high quality genomic data. By this, low coverage sequence could become a cost effective alternative to genotyping arrays with the additional benefits of a higher number of markers and detection of copy number variation when the number of sires in a breeding scheme is limited.

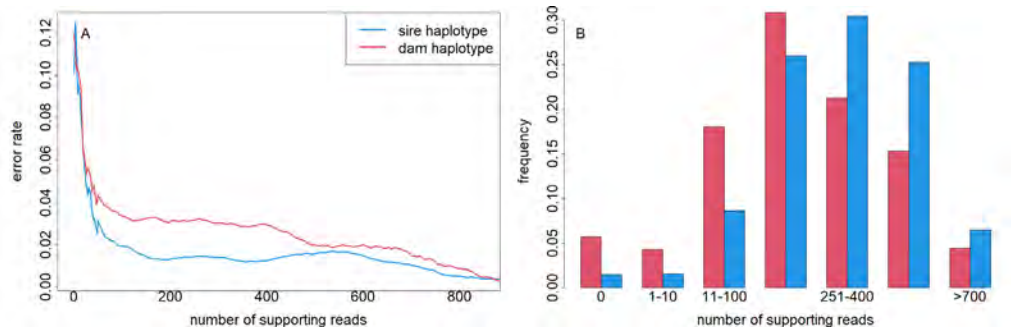


Figure 2. Discordance rate depending on the number of reads used in the HBimpute step (A) and frequency of the read-depth (B) depending on the sex.

It should still be noted that at this stage our method was only evaluated with simulated data and further development of the pre-phasing pipeline is needed to, e.g. run multiple iterations of our pipeline and use the imputed panel of the previous iteration to improve the pre-phasing. Thus, this should be seen as a proof-of-concept, but further evaluation on real data is required.

Acknowledgements

This study received financial support from the German Federal Ministry of Education and Research (BMBF) via the project MAZE (Funding ID: 031B0882).

References

Bouwman, A.C., Daetwyler, H.D., Chamberlain, A.J., Ponce, C.H., and Sargolzaei, M., *et al.* (2018) Nature Genetics 50(3): 362–367. https://doi.org/10.1038/s41588-018-0056-5

Browning, B.L., and Browning, S.R. (2013) Genetics 194(2): 459–471. https://doi.org/10.1534/genetics.113.150029

Browning, B.L., Zhou, Y., and Browning, S.R. (2018) The American Journal of Human Genetics 103(3): 338–348. https://doi.org/10.1016/j.ajhg.2018.07.015

Davies, R.W., Kucka, M., Su, D., Shi, S., and Flanagan, M., *et al.* (2021) Nature Genetics 53(7): 1104–1111.

Geibel, J., Reimer, C., Pook, T., Weigend, S., and Weigend, A., *et al.* (2021) BMC Genomics 22(1): 1–13. https://doi.org/10.1186/s12864-021-07663-6

Koren, S., Rhie, A., Walenz, B.P., Dilthey, A.T., and Bickhart, D.M., *et al.* (2018) Nature biotechnology 36(12): 1174–1182. https://doi.org/10.1038/nbt.4277

Pook, T., Schlather, M., de los Campos, G., Mayer, M., and Schoen, C.C., *et al.* (2019) Genetics1045-1061. https://doi.org/10.1534/genetics.119.302283

Pook, T., Schlather, M., and Simianer, H. (2020) G3: Genes, Genomes, Genetics 10(6): 1915-1918. https://doi.org/10.1534/g3.120.401193

Pook, T., Nemri, A., Gonzalez Segovia, E.G. Valle Torres, D., and Simianer, H., *et al.* (2021) PLOS Genetics 17(12): 1–22. https://doi.org/10.1371/journal.pgen.1009944

Rubinacci, S., Ribeiro, D.M., Hofmeister, R.J., and Delaneau, O. (2021) Nature Genetics 53(1): 120–126.

351. Quantitative trait simulation using MeSCoT software

V. Milkevych*, E. Karaman, G. Sahana, L. Janss, Z. Cai and M.S. Lund

Center for Quantitative Genetics and Genomics, Aarhus University, Tjele, Denmark; vimi@qgg.au.dk

Abstract

This work represents MeSCoT, a software dedicated to complex trait simulation through the mechanistic modelling of genes interactions. The software provides an innovative framework to study molecular mechanisms of gene-by-gene and gene-by-environment interactions. It closely imitates the basic *in vivo* mechanisms of complex trait realization and allows a detailed simulation of genes' regulatory interactions for variable genomic architectures. The MeSCoT software is scalable and well documented (https://genetics.ghpc.au.dk/vimi/mescot/).

Introduction

Quantitative (or complex) traits represent majority of traits that are of economic importance in agriculture (Goddard and Hayes 2009). Genetic variation of such traits is determined by a large number of loci and is related to a specific genomic architecture. The current approaches of breeding value prediction of a complex trait involve all genotyped single nucleotide polymorphisms (SNPs), regardless of their statistical significance (Meuwissen *et al.* 2001), and understanding of a genetic architecture that underlies quantitative traits could help improving statistical models for better predictive ability (Suravajhala *et al.* 2016).

This study represents a novel mechanistic approach to *in silico* quantitative trait simulation implemented within MeSCoT (Mechanistic Simulation of Complex Trait) software (Milkevych *et al.* 2021). The approach imitates core *in vivo* mechanisms of complex trait formation and quantitatively maps genotypic and phenotypic variation into the molecular mechanisms of gene expression. The software delivers a computational framework for $G\times G$ and $G\times E$ interaction studies as well as for verification of novel and existing statistical methods in quantitative genetics

Implemented models and algorithms allow detailed simulation of genes regulatory interactions for variable genomic architectures and generates extensive omics data. In addition to mechanistic models, the software includes the novel single-nucleotide polymorphism and omnigenic genetic models (Boyle *et al.* 2017). These models allow detailed studies of impact of genetic and phenotypic variance on gene regulation and, hence, can help reveal molecular mechanisms linking the heritability and variation in molecular traits.

Methods

Implemented models. A number of detailed mechanistic models constitute the computational framework implemented within MeSCoT software, in particular: the model of genes regulatory interactions, single-nucleotide polymorphism model, genotype by environment interaction model, and a model for genomic network geometry. The detailed mathematical formulation of these models as well as implemented numerical methods and algorithms can be reviewed in (Milkevych *et al.* 2021).

Simulation workflow & software interface. The schematic overview of MeSCoT simulation workflow and data depicted in Figure 1A; Figure 1B and 1C show the MeSCoT input and output interface respectively; adapted from (Milkevych *et al.* 2021).

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_351, © V. Milkevych *et al.* 2022

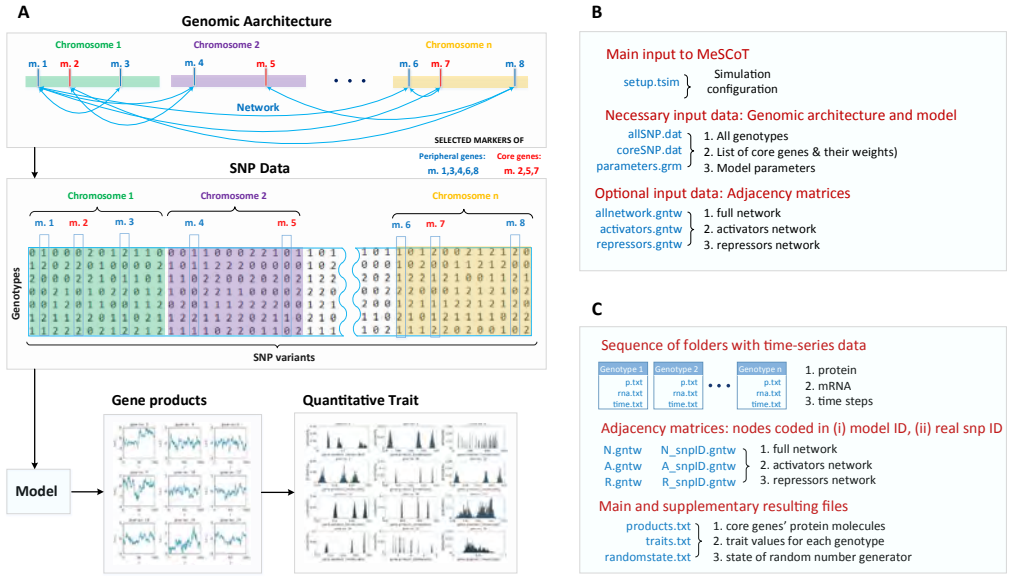


Figure 1. Simulation workflow & software interface. (A) The schematic overview of MeSCoT simulation workflow. The shaded areas depict distinct workflow components: data blocks and functional units. The black arrows represent information flows. The 'Genomic Architecture' is a data block of prior information regarding a modelled complex trait, such as peripheral and core genes (markers colored in blue and red respectively), their locations and network relations (blue arrows). The prior information is used to build a data file consisting the combined information for all genotypes (SNP variants) in population, the shaded area named 'SNP Data File'. The computational unit ('Model') utilizes the genomic and network information to produce 'Gene products' data that can be further used for the trait calculations, depicted within 'Quantitative Trait' area. (B) MeSCoT input interface. C) MeSCoT output interface. Blue color indicate different types of files and folders; black color describes a function and purpose of files; red color marks different interface groups.

Data used for the case studies. For the case studies we use a trait with a pure artificial genomic architecture determined by *in silico* generated genomic network. We simulated the genotypic data for 5,000 individuals completely at random, by sampling allele counts from {0,1,2} for 50 SNP loci. Among the simulated SNPs, 12 were set to be core genes (Boyle *et al.* 2017) with equal contributions to the trait phenotype. The rest of the SNPs were taken as peripheral genes.

Results & discussion

Genomic network. The MeSCoT software allows *in silico* networks simulation. The network data (data files representing adjacency matrices) was produced once and subsequently reused in all simulation studies. The resulting network geometry represented in terms of a directed graph is depicted in Figure 2A; adapted from (Milkevych *et al.* 2021).

Expression data. The detailed dynamics of mRNA and protein concentrations for all genes involved in the regulatory network is covered by MeSCoT functionality. As a result of the case study, the standardized expression level profiles for core genes (from the network depicted in Figure 2A) are depicted in Figure 2B.

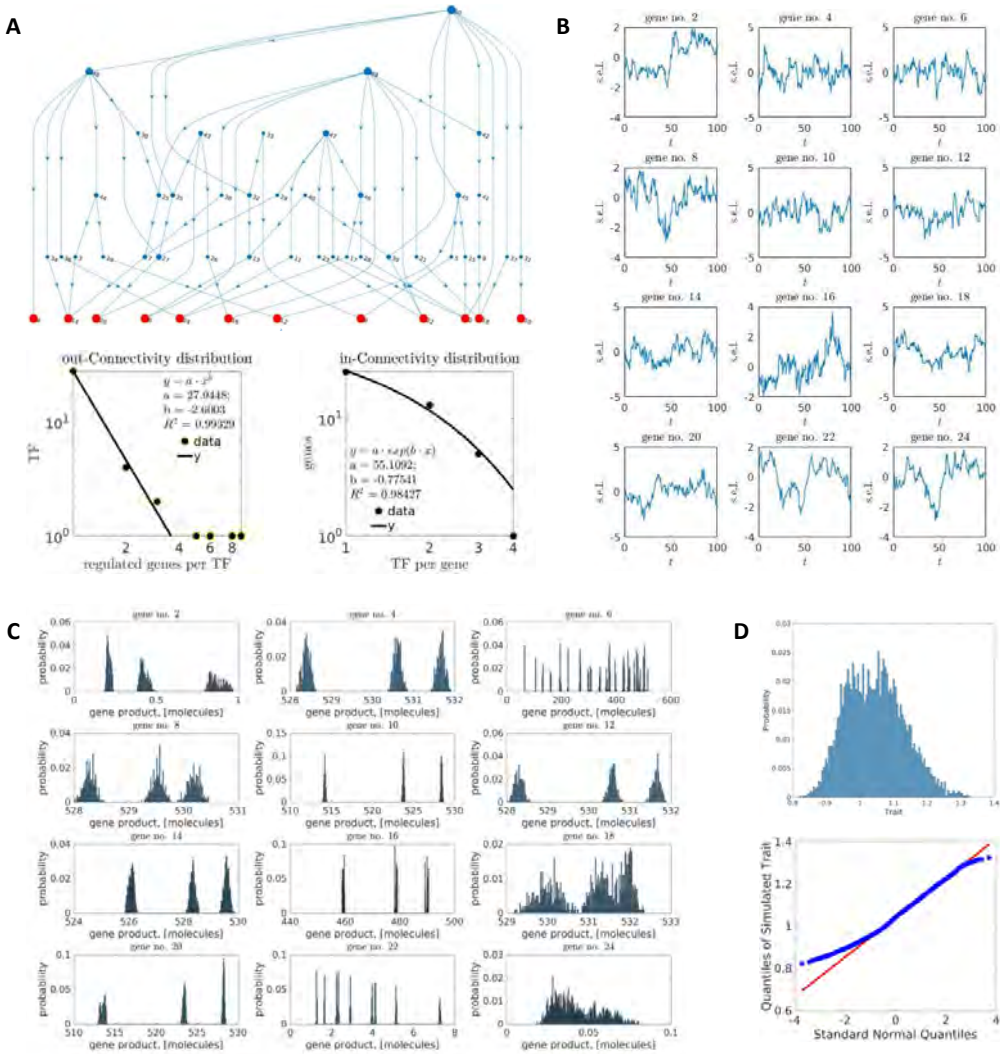


Figure 2. The results of the case study. (A) Simulated genomic network and its connectivity properties (log-log representation). The size of blue dots is proportional to the graph's nodes degrees; red dots indicate the core genes; nodes labels (numbers) correspond to SNP identity numbers in the genotypes data file; arrows indicate the directions of regulatory interactions; TF stands for transcription factor. (B) Standardized expression data for core genes. The expression level profiles were generated using mRNA concentration data of reference genotype; *s.e.l.* is standardized expression level; *t* is time in minutes. (C) Distributions of core genes' products (proteins concentrations); the distributions were generated for the population consisting of 5,000 distinct genotypes. (D) Distribution of simulated genotypic values (traits, expressed as normalized values). The distribution was generated for population consisting of 5,000 distinct genotypes; the trait values were calculated using the normalized values of protein concentrations.

Gene-by-gene interactions. The software allows extensive analysis of $G \times G$ interactions. Core genes' products form complex patterns, such as segregated distributions of protein molecules in Figure 2C. These patterns, as we believe, are due to the interplay between the genomic differences in population and the geometry of the genomic architecture (network, Figure 2A).

Cumulative contribution of all normalized core genes' products is visualized in Figure 2D where the result is represented as the probability distribution of simulated genotypic values upon 5,000 individuals. For the assumed genomic architecture and the study genotypes, the trait approach a normal distribution (Figure 2D). The trait distribution represented as a combination of minor distributions related to the individual core genes' products (Figure 2C).

Multi-trait case. While the current version of the software is able to simulate a multi-allelic QTL, the multi-trait functionality is under development.

Web resources. The software (executables for Linux and Windows platforms) as well as the necessary documentation with examples are freely accessible via the MeSCoT supporting web site: https://genetics.ghpc.au.dk/vimi/mescot.

References

Boyle, E.A., Y.I. Li, and J.K. Pritchard, 2017 An Expanded View of Complex Traits: From Polygenic to Omnigenic. *Cell* 169 (7):1177-1186. https://doi.org/10.1016/j.cell.2017.05.038.

Goddard, M.E., and B.J. Hayes, 2009 Mapping genes for complex traits in domestic animals and their use in breeding programmes. *Nature Reviews Genetics* 10 (6):381-391. https://doi.org/10.1038/nrg2575.

Meuwissen, T.H., B.J. Hayes, and M.E. Goddard, 2001 Prediction of total genetic value using genome-wide dense marker maps. *Genetics* 157 (4):1819-1829. https://doi.org/10.1093/genetics/157.4.1819.

Milkevych, V., E. Karaman, G. Sahana, L. Janss, Z. Cai *et al.*, 2021 MeSCoT: the tool for quantitative trait simulation through the mechanistic modeling of genes' regulatory interactions. *G3 Genes|Genomes|Genetics* 11 (7). https://doi.org/10.1093/g3journal/jkab133.

Suravajhala, P., L.J.A. Kogelman, and H.N. Kadarmideen, 2016 Multi-omic data integration and analysis using systems genomics approaches: Methods and applications In animal production, health and welfare. *Genetics Selection Evolution* 48 (1). https://doi.org/10.1186/s12711-016-0217-x.

352. SIRE 2.0: a novel software tool to estimate host genetic variation in infectious disease transmission

C.M. Pooley[1,2*], G. Marion[1] and A. Doeschl-Wilson[2]

[1]*Biomathematics and Statistics Scotland, The King's Buildings, Edinburgh, EH9 3FD, United Kingdom;*
[2]*The Roslin Institute, The University of Edinburgh, Easter Bush, EH25 9RG, United Kingdom;*
chris.pooley@bioss.ac.uk

Abstract

Selection of individuals that are genetically less susceptible to infection, less infectious once infected and who recover more quickly, offer an effective and long-lasting way to reduce the incidence and impact of infectious diseases in farmed animal populations. To date, however, genetic evaluations for these underlying host traits has been hampered by a lack of numerical tools capable of estimating them from available data. This paper introduces a new software tool called SIRE 2.0 which is able to estimate breeding values for polygenic susceptibility, infectivity and recoverability from diverse types of individual-level data. A simulation study is carried out to first validate SIRE 2.0, and then the dependency of prediction accuracies (PAs) on various factors is investigated. These, together with analytical expressions for PAs, indicate that SIRE 2.0 can reliably estimate genetic parameters for all three traits and provide insight into associated data requirements.

Introduction

The speed and severity with which a disease spreads in a population depends on three separate (but potentially correlated) epidemiological host traits: susceptibility, infectivity and recoverability (Pooley *et al.*, 2020, Bijma *et al.*, 2021). Recent research indicates that genetic improvement of all three traits simultaneously, rather than just susceptibility, would be substantially more effective at disease control (Tsairidou *et al.*, 2018, Bijma *et al.*, 2021, Doeschl-Wilson *et al.*, 2022). SIRE is an existing software tool (Pooley *et al.*, 2020) that estimates the contributions from fixed effects (e.g. SNP or vaccine effects) to these traits from various types of individuals' disease records (e.g. disease diagnostic test results or times of infection and/or recovery/death). This paper introduces SIRE 2.0, which incorporates additive genetic contributions into the model (and so builds on previous efforts, e.g. Anacleto *et al.*, 2015). These account for correlations in trait values between individuals through a relationship matrix (e.g. susceptible parents are more likely to have susceptible offspring), and between the traits themselves (e.g. more susceptible individuals may, on average, be more infectious).

Materials & methods

The genetic-epidemiological model. The system is assumed to consist of one or more closed 'contact groups'. The dynamics within each contact groups is represented by an SIR (or SI) compartmental model, with individuals classified as being either susceptible to infection (S), infected and infectious (I), or recovered/removed/dead (R). The force of infection (i.e. the probability per unit time of individual j becoming infected) is given by $\lambda_j = \beta e^{G_z} e^{g_j} \Sigma_i e^{f_i}$ and the time taken for individual j to recover after being infected is taken to be gamma distributed with mean $w_j = 1/(\gamma e^{r_j})$ and shape parameter k_{sh}. Here β and γ are population average transmission and recovery rates, g_j, f_j and r_j represent fractional deviations in the susceptibility, infectivity and recoverability for individual j, and the sum i goes over all infected individuals sharing the same contact group as j at any given point in time. Finally, G_z is a random effect (for group z with standard deviation σ_{GE}) which accounts for group-specific factors that influence the overall speed of an epidemic in one contact group relative to another (e.g. animals kept in different management conditions or environmental variation). The individual-based fractional deviations are parameterised in the following

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_352, © C.M. Pooley *et al.* 2022

way: $g=Xb_g+a_g+\varepsilon_g$, $f=Xb_f+a_f+\varepsilon_f$, $r=Xb_r+a_r+\varepsilon_r$, where g, f and r are vectors (containing information for all the individuals) that are decomposed into fixed effects b_g, b_f and b_r, where X is a design matrix (e.g. to account for sex differences in the traits or vaccination status), additive genetic contributions (a_g, a_f, a_r) that account for the correlations in trait values between different individuals (these are multivariate-normally distributed with zero mean and covariance matrix $A\otimes\Omega$, where A is the pedigree/genomic relationship matrix and Ω is a 3×3 covariance matrix that characterises potential correlations between traits). Finally, the residual contributions (ε_g,ε_f,ε_r) account for all other variation (these are also multivariate normal with zero mean and covariance matrix $I\otimes\Psi$, where I is the identity matrix reflecting the fact that residuals are assumed to be uncorrelated between individuals and Ψ is a 3×3 covariance matrix that characterises environmental correlations between traits). The total phenotypic covariance matrix is given by $\Theta=\Omega+\Psi$.

Software tool. The model from the previous section has been implemented in SIRE 2.0 (Pooley *et al.*, 2022) which uses an efficient Markov chain Monte Carlo (MCMC) algorithm to draw samples from the posterior. A graphical interface takes as input any combination of information about individuals' infection and recovery times, disease status measurements and disease diagnostic test results. The A matrix is specified (either from a pedigree or using a genomic relationship matrix), along with other fixed effects (if included), as well as details of which individuals belong in which contact group. A range of prior specifications can be made on different model parameters. The outputs consists of posterior trace plots for model parameters, distributions, visualisation of infection and recovery times, dynamic population estimates and summary statistics (means and credible intervals) as well as MCMC diagnostic statistics. Files containing posterior samples of parameters and events can be exported for further analysis.

Results

A simulation study was performed to first validate SIRE 2.0 and then explore factors affecting prediction accuracies for the additive genetic contributions to the traits. A simple half-sib population structure is assumed: Q unrelated sires are each randomly mated to P unrelated dams to produce P offspring (giving a total of $U=Q\times P$ individuals in the next generation). These individuals are randomly allocated across contact groups each containing N individuals. A randomly selected index case in each group is infected to instigate an epidemic. Simulated datasets are generated using a modified Doob-Gillespie algorithm (Pooley *et al.*, 2020).

Validation. We begin by looking at what might be considered a 'typical' scenario using the following parameters: $Q=100$, $P=20$, $N=10$, $\beta=0.05$ and $\gamma=0.1$ (corresponding to basic reproductive ratio $R_0=\beta N/\gamma=5$), $k_{sh}=3$, and $\sigma_{GE}=0.25$ (values for covariance matrices are shown in Figure 1). For simplicity no fixed effects are included. In total 20 exemplar datasets were generated (to account for system stochasticity) each of which consisted of the disease status of individuals periodically checked every $\Delta T=2$ time units up until $T_{max}=40$. Figure 1 shows tornado plots for posterior distributions inferred using SIRE 2.0. We find that, for the most part, the true parameter values lie within the 90% credible intervals indicating that inference is able to successfully capture these values. Based on the data from the $U=2000$ individuals used here we find the susceptibility Θ_{gg}, infectivity Θ_{ff}, and recoverability Θ_{rr} can be estimated with good precision, with slightly poorer estimates obtained for heritabilities h^2 and genetic correlations.

Prediction accuracies (PAs). The PAs are defined as correlations between the posterior means of the additive genetic components (a_g, a_f, a_r) and their true (usually unknown) values. Taking a best case scenario in which infection and recovery times are exactly known (along with simplifying assumptions including that traits are uncorrelated), analytical expressions for PAs can be derived using a maximum likelihood approach (not shown):

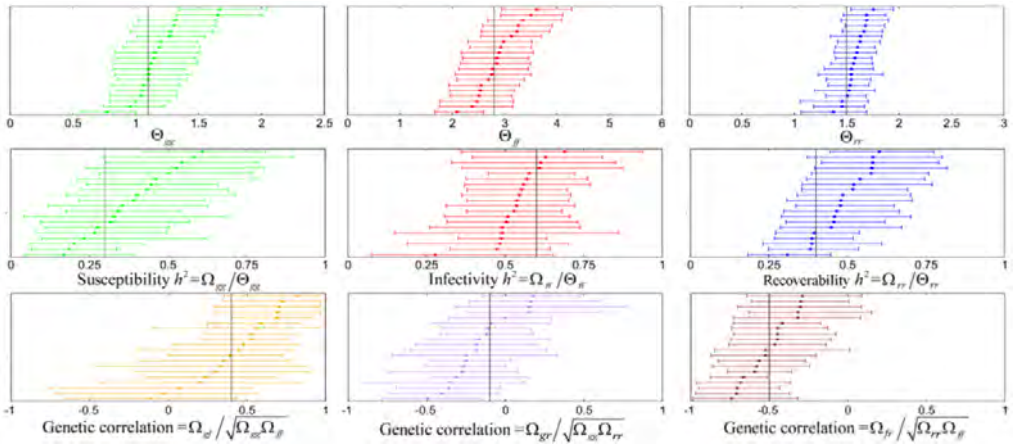


Figure 1. Tornado plots of posterior distributions for 20 replicated datasets (circles represent posterior means, error bars give 90% credible intervals and true parameter values indicated by the vertical black lines).

$$\alpha_g = \left[1+\frac{4N}{P\Omega_{gg}(N-1)}\left(1+\Theta_{gg}+\frac{1}{N}\Theta_{ff}\right)\right]^{-\frac{1}{2}},$$
$$\alpha_f = \left[1+\frac{4N}{P\Omega_{ff}\log(2N)}\left(1+2\Theta_{ff}+\frac{3}{\log(2N)}\Theta_{gg}\right)\right]^{-\frac{1}{2}},$$
$$\alpha_r = \left[1+\frac{4}{P\Omega_{rr}}\left(\frac{1}{k_{sh}}+\Theta_{rr}\right)\right]^{-\frac{1}{2}}. \quad (1)$$

Figure 2 shows results for these expressions under a variety of scenarios, and they are compared to posterior predictions using SIRE 2.0 on simulated data (the baseline set of parameters used were $\Omega=\Psi=I/2$, i.e. a heritability of 0.5 on all traits with no correlations, with others parameters as before): (a-b) shows that if the phenotypic variance in a trait is small then so is the corresponding PA (note for equal variance, recoverability has the highest PA, followed by susceptibility, with infectivity having the lowest). (c) Introducing correlations between any of the traits (in this case between infectivity and recoverability) increases PAs, especially for infectivity (hence Equation 1 represents a lower bound). (d) Varying the number of individuals per epidemic group N (for fixed overall number of individuals U) has little effect on the PA for susceptibility and recoverability, but has a marked effect on reducing the PA for infectivity as N becomes large (i.e. many small contract groups provide more information). (e) Increasing the number of sire offspring P (for fixed U) increases the PAs. In this case the simulated data includes only recovery times, and this leads to a shift to the right in the symbols with respect to the analytical results reflecting a reduction in statistical power (e.g. P would need to be ~10 times larger for an equivalent infectivity PA, although this factor is smaller for the other two traits). (f) This looks at varying the time-step ΔT at which disease diagnostic tests are made. We find that PAs are only marginally reduced even when ΔT is relatively large, indicating that precise measurements of infection and recovery time are not required.

Discussion

This paper introduces the software tool SIRE 2.0 which allows, for the first time, estimation of additive genetic contributions to susceptibility, infectivity and recoverability (as well as the genetic covariance matrix) from different types of real-world epidemiological data. Analytical expressions reveal that large additive genetic variation in these traits correspond to high PAs (supporting the effective use of selective breeding). Based on the empirical observation of super-spreaders (relatively few individuals causing most infections) it is known that phenotypic variation in infectivity is high, but as yet it is not clear to what extent

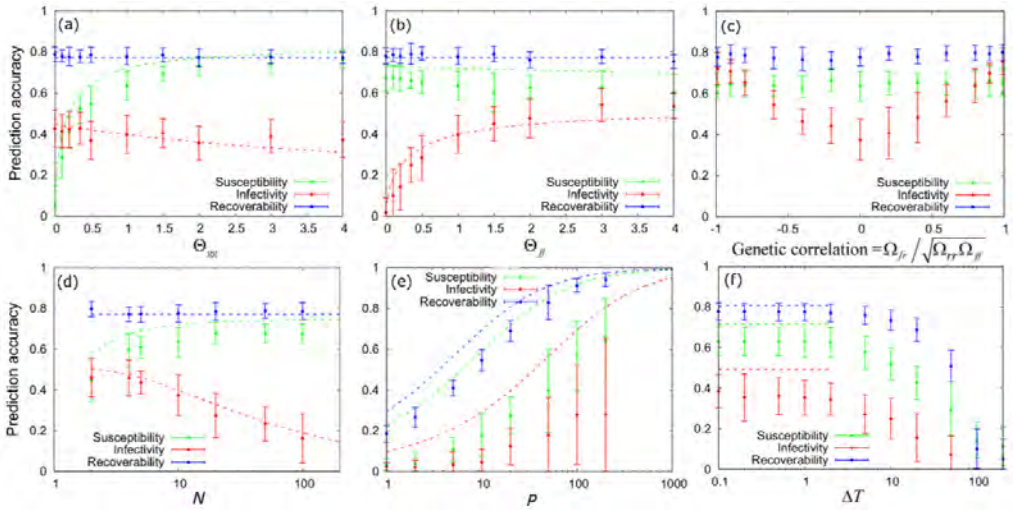


Figure 2. Prediction accuracies for susceptibility, infectivity and recoverability (circles with error bars give the mean and standard deviation of numerical estimates obtained from 20 simulated datasets, dashed lines from Equation 1).

that variation is environmental or genetic-based. The next version of the software SIRE 2.1 is currently under development to allow for arbitrary specification of the compartmental model, e.g. to allow for a latent state, or the possibility of accumulation of pathogens in the environment. These extensions will open up the possibility of applying this tool to a wider range of experimental or field data, as exemplified in Doeschl-Wilson *et al.* (2022) and Prentice *et al.* (2022).

References

Anacleto O., Garcia-Cortés L.A., Lipschutz-Powell D., Woolliams J.A., and Doeschl-Wilson A., (2015) Genetics 201:871-884. https://doi.org/10.1534/genetics.115.179853

Bijma P., Hulst A. D., and de Jong M. (2021) Genetics 220.1 (2022): iyab141. https://doi.org/10.1093/genetics/iyab141

Doeschl-Wilson A., Marion G., C.M. Pooley C.M., (2022) Proc. of the 12th WCGALP, Rotterdam, the Netherlands.

Pooley C.M., Marion G., Bishop S.C., Bailey R.I., and Doeschl-Wilson A.B. (2020) PLOS Comput Biol 16(12):e1008447. https://doi.org/10.1371/journal.pcbi.1008447

Pooley C.M., Marion G., and Doeschl-Wilson A, Susceptibility, Infectivity and Recoverability Estimator (SIRE 2.0) 2022. Available at: https://github.com/theITEAM/SIRE2.0

Prentice J.C, Pooley C.M., Tsairidou S, Pong Wong R., Anacleto O., *et al.* (2022) Proc. of the 12th WCGALP, Rotterdam, the Netherlands.

Tsairidou S., Allen A., Banos G., Coffey M., Anacleto O., *et. al*, (2018), Front. Vet. Sci.: 5:310. https://doi.org/10.3389/fvets.2018.00310

353. iVPSV: an intuitive visualisation platform for structural variation data

H. Jung[1*], B. Jeon[2], B. Hayes[1] and E. Ross[1]

[1]Queensland Alliance for Agriculture and Food Innovation, The University of Queensland, Australia; [2]Deaprtment of Communication and Arts, The University of Queensland, Australia; hyungtaek.jung@uq.edu.au

Abstract

Visual validation is an essential step to minimise false-positive predictions from structural variants (SVs) detection but a difficult task in life sciences. We present a Graphical User Interface (GUI)-based intuitive Visualisation Platform for Structural Variation (iVPSV), a free standalone desktop application software with Python support, that makes manipulating commonly used SV files more accessible to biologists with a simple mouse click. It enables them to perform end-to-end reproducible visual validation by facilitating the rapid discovery and identifying/navigating target regions/sequences (or user's interest) in SVs datasets for novice users. iVPSV can also create images (including summary reports) that display the read depth/contexts and sequence alignments necessary to adjudicate purported SVs across samples and the short and long-read sequencing technologies with multiple vcf and bam files. This specific benefit allows iVPSV to seamlessly integrate visual and interactive representations of SV files, supporting a wider scope of bioinformatics applications in the life sciences.

Introduction

Structural variation (SV), which includes mobile elements, is a fundamental genomic alteration that usually refers to a change of over 50 bp nucleotide fragments, including insertion, deletion, inversion, translocation and duplication (Sudmant *et al.*, 2015). While a considerable number of studies have provided state-of-the-art SV discovery tools, only limited investigations have been made for visualisation (Belyeu *et al.*, 2021; Nattestad *et al.*, 2021).

With the broad implementation of Next Generation Sequencing (NGS) technologies in the life sciences, the importance of visualisation for understanding biological phenomena is an increasingly important aspect in the genomic revolution. However, many of the available tools for SV visualisation are targeted for short-read alignments which cannot cover large compound SVs on several genomic loci. However, advances in long-read sequencing (Sedlazeck *et al.*, 2018; Jung *et al.*, 2019) have shed light on examples where visualising the portions of reads that maps to different parts of the reference is crucial. When a read is over 10 kb long, it can cover a wide-range of complicated SVs on the genome, but that context is easily lost using IGV (Robinson *et al.*, 2001), or other existing reference-centric viewers without adding a read-based perspective and showing all relevant reference loci (Nattestad *et al.*, 2021).

In recent years, other visualization tools have tackled some challenging aspects of validating structural variant calls. In particular, svviz (Spies *et al.*, 2015) has used realignment against a putative structural variant allele to better evaluate the evidence, SV-plaudit (Belyeu *et al.*, 2018) and SVCurator (Chapman *et al.*, 2019) have enabled crowd-sourced curation of structural variants. In addition, Ribbon (Nattestad *et al.*, 2021) has offered an interactive online SV visualisation tool along with other genomic features, and Samplot (Belyeu *et al.*, 2021) has provided a platform for structural variant visual validation and automated filtering including a machine learning package that dramatically decreases the number of false-positive without human review. Even with these very useful approaches, more complex patterns of structural variation remain hard to represent due to the lack of embedded visualisation step requires other third-party

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_353, © H. Jung *et al.* 2022

visualisation tools such as IGV (Robinson *et al.*, 2011) which is primarily built for short reads as well as for multiple vcf and bam files. Unfortunately, the true power of long reads is hidden when the tools we use are built for short reads. Furthermore, many existing SV visualisation tools require higher-order computational infrastructure (high-performance computing facilities and Cloud computing), lack GUIs and minimise the true power of long reads in visualisation, making them inaccessible to many researchers, and cumbersome for even experienced biologists. A more flexible approach including the acceptance of multiple vcf and bam files can be helpful by showing the full picture of alignments from both the reference and read perspective, allowing researchers to explore and communicate about complex variation in a more intuitive way. Thus, we have addressed this problem by creating iVPSV an interactive and intuitive visualization tool that displays alignments along with both reference and query sequences, along with variant calls, genes and other genomic features.

Materials & methods

iVPSV can be used by sophisticated data scientists and non-technical users who need an intuitive interface. The original intent for producing iVPSV was to reduce reliance on any command lines/scripts or web-based platforms, by creating a standalone lightweight program with substantially reduced computation demands. iVPSV provides key benefits in convenience, accessibility, and reproducibility because it does not include any heavyweight genome/transcriptome assembly, mapping and clustering workflows. iVPSV can execute and pre-analysed SV VCF files by filtering parameters on a user's desktop with the help of many packages (Table 1). While it mainly focuses on point-and-click analysis for less technical users, the Help function could provide an interactive experience for monitoring and iterating on an executed code.

Please note that iVPSV, a self-contained program, includes all necessary dependencies and executable packages. Simply unzip the folder and double-click iVPSV.exe after downloading the program.

Results

For user-friendliness and convenience, iVPSV was developed with five work modules covering different aspects of data visualisations, including post-processing, parsing, filtering, generating results, and help (Figure 1).

In addition, further assistance and information can be obtained via Help to improve processes and performance.

Table 1. Software packages integrated into iVPSV and their applications.

Software	Application
Python packages	
PyQt5 (https://pypi.org/project/PyQt5/)	PyQt is a Python binding of the cross-platform GUI toolkit Qt.
gffutils (https://github.com/daler/gffutils)	gffutils is for working with and manipulating the GFF and GTF format files typically used for genomic annotations.
PyVCF (https://pyvcf.readthedocs.io/en/latest/#)	PyVCF is for parsing variant call format by mimicking the csv module in the Python stdlib, as opposed to more flexible serialisation formats like JSON or YAML.
JavaScript Library	
D3.js (https://d3js.org/)	D3.js is for producing dynamic, interactive data visualisation with HTML, SVG and CSS.
jQuery	jQuery is for simplifying HTML DOM tree traversal, manipulation, event handling, CSS animation, and Ajax.

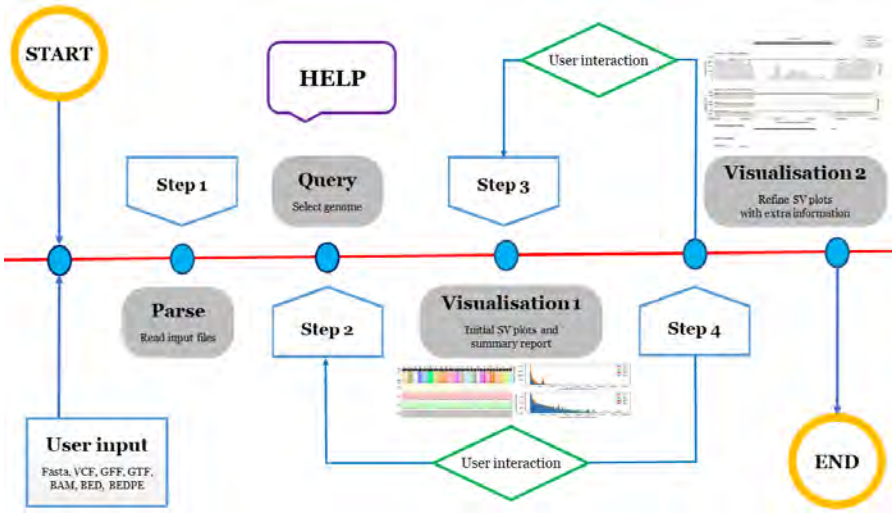


Figure 1. iVPSV workflow and data structures.

Discussion

iVPSV is implemented in Python and available under the MIT license and works on Windows, Linux and Mac systems. This package is also available on the PyPI python package manager. The current code runs under Python 3.7+ and JavaScript Library. Other dependency includes gffutils, PyVCF, PyQt5, D3.js, and jQuery (Table 1). We aimed to create a flexible, educational and user-friendly tool combining multiple features to visualise structural variant data. In the future, we will continue to update with new fast and easy GUI support including new embedding methods such as pan-SV visualisation, with low resource requirements and both multithread and cluster computing support, making these methods suitable for running on standard desktops in laptops. The future version of iVPSV will also include additional integration points allowing us to intersect, merge, count, complement, and shuffle genomic intervals from multiple files in widely-used genomic file formats.

References

Belyeu, J.R., Nicholas, T.J., Pedersen, B.S., Sasani, T.A., Havrilla, J.M., *et al.* (2018) Gigascience 7(7): giy064. https://doi.org/10.1093/gigascience/giy064

Belyeu, J.R., Chowdhury, M., Brown, J., Pedersen, B.S., Cormier, M.J., *et al.* (2021) Genome Biol. 22(1): 161. https://doi.org/10.1186/s13059-021-02380-5

Chapman, L.M., Spies, N., Pai, P., Lim, C.S., Carroll, A., *et al.* (2020) PLoS Comput. Biol. 16(6): e1007933. https://doi.org/10.1371/journal.pcbi.1007933

Jung, H., Winefield, C., Bombarely, A., Prentis, P., Waterhouse, P. (2019) Trends Plant Sci. 24(8): 700–724. https://doi.org/10.1016/j.tplants.2019.05.003

Nattestad, M., Aboukhalil, R., Chin, C.S., Schatz, M.C. (2021) Bioinformatics 37(3): 75–81. https://doi.org/10.1093/bioinformatics/btaa680

Robinson J.T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E.S., *et al.* (2011) Nat. Biotechnol. 29(1): 24–26. https://doi.org/10.1038/nbt.1754

Sedlazeck, F.J., Lee, H., Darby, C.A., Schatz, M.C. (2018) Nat. Rev. Genet. 19(6): 329–346. https://doi.org/10.1038/s41576-018-0003-4

Spies, N., Zook, J.M., Salit, M., Sidow, A. (2015) Bioinformatics 31(24): 3994–3996. https://doi.org/10.1093/bioinformatics/btv478

Sudmant, P.H., Rausch, T., Gardner, E.J., Handsaker, R.E., Abyzov, A., *et al.* (2015) Nature 526(7571): 75–81. https://doi.org/10.1038/nature15394

354. A variational Bayes method for genomic prediction increases accuracy and computing speed

J. Cheng[1*], H. Cheng[2], R. Fernando[3], C. Maltecca[1], L. Ma[4], J. Dekkers[3] and J. Jiang[1*]

[1]Department of Animal Science, North Carolina State University, Raleigh, NC 27695, USA; [2]Department of Animal Science, University of California, Davis, CA 95616, USA;[3]Department of Animal Science, Iowa State University, Ames, IA 50011, USA; [4]Department of Animal and Avian Sciences, University of Maryland, College Park, MD 20742, USA; jcheng29@ncsu.edu; jicai_jiang@ncsu.edu

Abstract

Various Bayesian methods have been developed to improve genetic prediction of complex traits. Although Bayesian methods such as Bayes-B have been proven to outperform genomic best linear unbiased prediction for prediction of complex traits, especially those that have major genes, they are usually implemented using Markov chain Monte Carlo, which is time consuming and has convergence issues when the number of markers is greater than the number of individuals. To address this issue, we developed a computationally efficient variational Bayes (Bayes-VB) method that can flexibly partition the whole genome markers into many groups. Our approach can improve the accuracy of genomic predictions compared to Bayes-B and is much faster, as illustrated here for growth rate of pigs under a disease challenge.

Introduction

Genomic prediction using high-density single nucleotide polymorphism (SNP) markers has been widely used in livestock animals. The standard model used for genomic prediction is the genomic best linear unbiased prediction model (GBLUP), which assumes that all SNP effects are drawn from the same normal distribution (VanRaden, 2008; Meuwissen *et al.*, 2001). Various Bayesian methods for genomic prediction have also been developed based on other prior distributions of SNP effects. One of these methods is Bayes-B, which assumes SNP effects follow a mixture of a heavy-tailed distribution and a spike distribution at zero (Meuwissen *et al.*, 2001). Although methods such as Bayes-B have been proven to outperform GBLUP for prediction of phenotypes with major QTL (Moser *et al.*, 2015; Yin *et al.*, 2020), it has several computational drawbacks. First, the Markov chain Monte Carlo (MCMC) used to implement these methods are time consuming. Second, the Bayesian model has difficulty converging when the number of markers is larger than the number of animals (Yin *et al.*, 2020). Last, the SNP effects are assumed to be independently and identically distributed in these Bayesian models. However, because of linkage disequilibrium (LD) between SNP markers, effects of markers that are close to each other may not be independent and, thus, tend to have similar effects and are more likely to follow the same distribution (Yang and Tempelman, 2012). Zeng *et al.* (2018) developed a method called Bayes-N to account for the dependence of effects for SNPs that are closely linked by grouping SNPs, which improved prediction accuracy compared with Bayes-B in simulated data. However, this method is still computationally demanding. To address these issues, we developed a Bayesian method (Jiang, 2018, available at https://sites.google.com/view/ssgp/home?authuser=0) that can account for the dependence of markers by flexibly partitioning the whole genome markers into a large number of groups based on proximity. The method was evaluated using data from a natural polymicrobial disease challenge model in pigs.

Materials & methods

Data. Data on 3,205 wean-to-finish pigs were available from a natural disease challenge model (see Cheng *et al.*, 2020, for details). Pigs were provided by seven companies, and each company was represented once each cycle (one batch per company per cycle) for a total of seven cycles. All animals were genotyped with a 650 k SNP chip. In total, 435,172 SNPs that passed quality control for all seven cycles, as described by

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_354, © J. Cheng *et al.* 2022

Cheng *et al.* (2020), were utilized for analysis. Average daily gain in the challenge nursery (cNurADG) and finisher (FinADG) were used to compare the accuracies of the different genomic predictions. Phenotypes for pigs that died were imputed and expanded as described by Cheng *et al.* (2020). These two traits reflect different underlying genetic architectures, as a major QTL was identified for cNurADG in the major histocompatibility (MHC) region that explained about 12.8% of genetic variance. In contrast, a relatively larger polygenic contribution to genetic variance was observed for FinADG, with the MHC region explaining only 2.7% of genetic variance (Cheng *et al.*, 2021).

Statistical model. The following statistical model was used:

$$\mathbf{y} = \mathbf{Xb} + \sum_{h=1}^{p} \mathbf{K}_h \mathbf{u}_h + \mathbf{e} \quad (1)$$

where y is phenotype; X is the design matrix for fixed effects; b is a vector of fixed effects, including batch (50 batches), mortality (1 for died; 0 for survived), and entry age, $\mathbf{b} \sim N(\mathbf{0}, \mathbf{I}\sigma_b^2)$, where σ_b^2 is the variance for the 'fixed effects', which was assigned a large value (e.g. 1e8); p indicates the number of groups of random marker effects; K_h is the design matrix, i.e. the genotypes for markers in group h; u_h is a vector of random marker effect for the h^{th} group, assumed to follow a normal distribution, with each group having its own variance component ($\sigma_{u_h}^2$), $\mathbf{u}_h \sim N(\mathbf{0}, \boldsymbol{I}\sigma_{u_h}^2)$; e is the vector of residual effect, assumed to follow a normal distribution, $\mathbf{e} \sim N(\mathbf{0}, \boldsymbol{I}\sigma_e^2)$. Residual variance is further assumed to follow an inverse-gamma distribution with two parameters (shape parameter c_0 and scale parameter d_0) and marker effect variance is assumed to follow a half-Cauchy distribution with group-specific parameters A_h, following Polson and Scott, (2012). In this study, the random effects of pen (437 unique pens in the challenge nursery and 262 in the finisher) and litter (1,119 litters) were also fitted by treating them as pseudo markers that were fitted in two extra groups.

Algorithms for model fitting. A variational Bayes (VB) (Beal, 2003) algorithm that approximates target distribution with a factored distribution that omits dependencies present in the target distribution was used to fit the model. To speed up the convergence, the parameter expanded method was applied to VB iterations (Jaakkola and Qi, 2007) by expanding the model by introducing auxiliary variables (c_h) to each group of SNPs: $\tilde{\mathbf{u}}_h = \mathbf{u}_h / c_h,\ and\ \tilde{\sigma}_{u_h}^2 = \sigma_{u_h}^2 / c_h^2$. After all parameters are updated in each iteration, the variational lower bound is maximized with respect to c_h, and u_h and $\sigma_{u_h}^2$ are then updated again using $\mathbf{u}_h = c_h \tilde{\mathbf{u}}_h, and\ \sigma_{u_h}^2 = c_h^2 \tilde{\sigma}_{u_h}^2$. In such a way, the strong coupling between variables in the original model was reduced.

Genomic predictions. Predictions from Bayes-VB were compared with those from GBLUP and Bayes-B. For Bayes-VB, the whole-genome markers were split into groups by an equal number of 200, 400, 600, 800, 1000, or 2,000 SNPs. The GBLUP and Bayes-B methods were implemented as described by Cheng *et al.* (2020) and Cheng *et al.* (2021), respectively, with the latter using an MCMC of length 50,000, with the first 5,000 iterations discarded as burn-in. Samples were saved at every 100th iteration of the chain. A prior probability of 0.1% of a SNP having a non-zero effect was used for the Bayes-B model. A seven-fold (each cycle was used as one fold) cross validation strategy was employed for genomic predictions of cNurADG and FinADG using the three methods, with data from six cycles used for training and the remaining cycle for prediction. The prediction accuracy was evaluated using the Pearson's correlation in the validation population between the genomic predictions and phenotypes adjusted by estimates of the fixed effects of batch, mortality, and entry age and the random effects of pen and litter from a GBLUP analysis of the full data set. Unbiasedness was assessed as the regression coefficient of adjusted phenotypes on genomic predictions in the validation population. The processor used for all analyses was a 3.00 GHz Intel Core i7 with 64 GB of memory.

Results

In general, for both cNurADG and FinADG, accuracies increased and then decreased with an increase in partition size regarding the number of SNPs. The optimal group size was 600 SNPs for cNurADG and 800 SNPs for FinADG, based on prediction accuracy. Comparison of prediction accuracies using Bayes-VB vs using GBLUP and Bayes-B are shown in Figure 1. For cNurADG, accuracies were higher for Bayes-VB and Bayes-B than for GBLUP for most cycles, except for cycle 2. Bayes-VB outperformed Bayes-B for all cycles. On average, the accuracy for Bayes-VB was 15 and 38% higher than for Bayes-B and GBLUP, respectively. For FinADG, similar to cNurADG, both Bayes-VB and Bayes-B outperformed GBLUP for most cycles (except for cycles 1 and 7). However, Bayes-VB only outperformed Bayes-B for cycles 2 and 5. On average, the accuracy for Bayes-VB was 4% lower than for Bayes-B but 10% higher than for GBLUP. Additionally, Bayes-VB resulted in slightly less bias than Bayes-B. Regarding computational time, Bayes-VB (using the optimal group size) was 76 times faster than Bayes-B (5 vs 380 min) for cNurADG and 54 times faster (5 vs 270 min) for FinADG.

Discussion

The optimal partition of the whole genome markers may be affected by two main factors, the LD decay pattern and the distribution of QTL. In this study, the optimal group size by number of SNPs was 600 SNPs for cNurADG and 800 SNPs for FinADG. This is more or less in agreement with the LD decay pattern in this pig population. The distribution of QTL may also play a role, i.e. a major QTL was found for cNurADG but was not so obvious for FinADG, which resulted in the difference in optimal group size (600 vs 800 SNPs) for cNurADG and FinADG. Even though an appropriate group size that matches LD length was used, these markers in the same group were not strongly associated with a major QTL for FinADG, thus not contributing to prediction accuracy. Previously, Zeng *et al.* (2018) used simulated data to demonstrate that modeling the dependence of effects of SNPs closely linked in the prior will improve the ability of SNPs close to a QTL to jointly capture its effect. The effect of model choice on prediction accuracy also depends on genetic architecture of a trait studied. Using simulation, Coster *et al.* (2010) reported that variable selection methods using Bayesian regression, such as Bayes-B, increased in accuracy when the number of QTL decreased. Our Bayes-VB method was also affected by the genetic architecture of a trait, as the increase in accuracy from Bayes-VB over GBLUP was greater for cNurADG (38%), which has a major QTL (MHC), than for FinADG (10%), which is not controlled by a major QTL. Interestingly, Bayes-B showed similar improvements in accuracy over GBLUP for cNurADG (20%) and FinADG (16%). This suggests that Bayes-VB is more affected by the genetic architecture of the trait analyzed than Bayes-B but further validation study is needed. In practice, a cross-validation strategy can be used to find the optimal group size using training population. The whole genome can also be split by haplotypes (unequal group size).

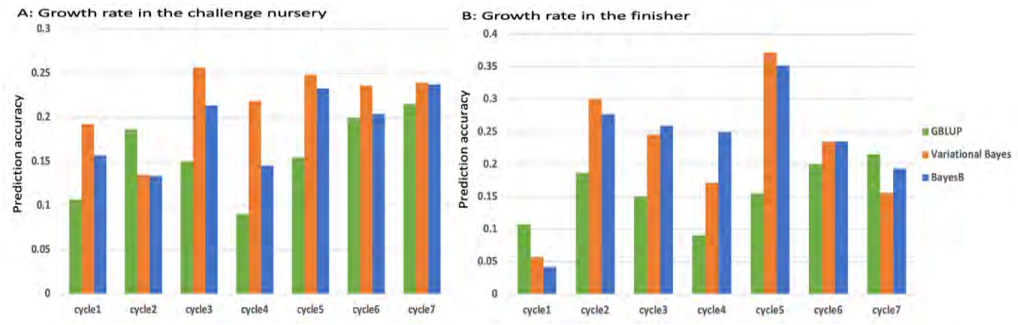


Figure 1. Comparison of genomic prediction accuracies using GBLUP, variational Bayes, and Bayes-B for growth rate in the challenge nursery and finisher for each cycle.

Conclusions

A computationally efficient method that can account for dependence of markers by flexibly partitioning the whole genome markers into a large number of groups was successfully developed.

Acknowledgements

Members of PigGen Canada are acknowledged for providing the pigs, John Harding, Michael Dyck, Frederic Fortin, Graham Plastow designed and developed protocols for animal sourcing, management, and phenotype recording, with funding from Genome Canada and Genome Alberta

References

Beal M.J. (2003). Variational algorithms for approximate Bayesian inference. PhD thesis. University College London, London.

Coster, A., Bastiaansen J.W., Calus M.P., van Arendonk J.A., and Bovenhuis H. (2010) Genet. Sel. Evol. 42: 9. https://doi.org/10.1186/1297-9686-42-9

Cheng J., Putz, A.M., Harding, J.C.S., Dyck, M.K., Fortin, F. *et al.* (2020) J. Anim. Sci. 98, 8. https://doi.org/10.1093/jas/skaa244

Cheng J., Fernando R., Kachman S., Cheng H., Lim K. *et al.* (2021) G3-Genes Genom. Genet. jkab441, https://doi.org/10.1093/g3journal/jkab441

Jaakkola T.S., Qi Y. (2007). Advances in Neural Information Processing Systems, pp 1097-1104

Jiang J. (2018). Genetic architecture of complex traits and accuracy of genomic selection in dairy cattle. PhD thesis. University of Maryland. https://doi.org/10.13016/xsqu-xjmp

Meuwissen T.H.E., Hayes B.J., and Goddard M.E. (2001) Genetics, 157(4), 1819-1829. https://doi.org/10.1093/genetics/157.4.1819

Moser G., Lee S.H., Hayes B.J., Goddard M.E., Wray N.R. *et al.* (2015) PLoS Genet. 11:e1004969. https://doi.org/10.1371/journal.pgen.1004969

Polson N.G., and Scott J.G. (2012). Bayesian Anal. 7 887–902. https://doi.org/10.1214/12-BA730

VanRaden P.M. (2008) J Dairy Sci. 91:4414–23. https://doi.org/10.3168/jds.2007-0980

Yang W. and Tempelman R.J. (2012) Genetics, 190(4):1491–1501. https://doi.org/10.1534/genetics.111.131540

Yin L., Zhang H., Zhou X., Yuan X., Zhao S. *et al.* (2020) Genome Biol. 21, 146. https://doi.org/10.1186/s13059-020-02052-w

Zeng J., Garrick D., Dekkers J., Fernando R. (2018) PLoS ONE 13(3): e0194683. https://doi.org/10.1371/journal.pone.0194683

355. Additions to the BGLR R-package: a new function for biobank size data and Bayesian multivariate models

P. Pérez-Rodríguez[1*] and G. de los Campos[2,3,4*]

[1]Colegio de Postgraduados, CP 56230, Montecillos, Estado de México, México; [2]Department of Epidemiology & Biostatistics, Michigan State University, USA; [3]Department of Statistics & Probability, Michigan State University, USA; [4]IQ – Institute for Quantitative Health Science and Engineering, Michigan State University, USA; perpdgo@gmail.com; gustavoc@msu.edu

Abstract

The BGLR R-package implements single-trait shrinkage and variable selection models that are routinely used for genomic prediction. The software was first released in 2014 and has been successfully adopted by researchers, with over 1000 monthly downloads from CRAN mirrors. We have recently implemented two additional functions to the package: BLRXy() and Multitrait(). The first function is optimized for genomic prediction problems involving a very large sample size (biobank-size data). The second function implements shrinkage and variable selection models for multi-trait models with structured or unstructured (co)variance matrices. In this note we describe the new functionality added to BGLR, present examples, and benchmarks.

Introduction

The BGLR function of the homonymous R-package (Pérez and de los Campos, 2014) implements Bayesian shrinkage and variable selection models for parametric and semi-parametric regressions (e.g. de los Campos *et al.*, 2013; Meuwissen *et al.*, 2001). The software was originally tailored for single-trait regressions involving many more covariates (p) than sample size (n, i.e. $p>>n$). Many modern data sets, including data from human biobanks and large-scale genomic evaluations, can have a very large sample size (hundreds of thousands of individuals with phenotype and genotype records). In some of these data sets sample size can vastly exceed the number of predictors. In such settings computational performance and speed can be improved by implementing algorithms using summary statistics ($y'y$, $X'X$, and $X'y$) as opposed to algorithms that use phenotypes (y) and incidence matrices (X) directly for computations as it is the case of the BGLR() R-function. Thus, to offer efficient software for biobank-sized data we added to the BGLR package the BLRXy() function which generates posterior samples using summary statistics as inputs; the function has a much faster performance than BGLR when $n>>p$. The second major addition we recently made was the extension of BGLR to multi-trait models. The function Multitrait() implements the same models implemented in the BGLR() function for systems of traits with structured and un-structured (co) variance matrices. In this paper we present new updates for the BGLR package for fitting univariate models for large-sample-size problems and multiple trait models.

Materials & methods

New algorithms for single-trait models when n>>p. Consider a linear regression model of the form: $y = X\beta + \varepsilon$, where $y = (y_1, y_2, \ldots, y_n)'$ is a vector of phenotypes, $\boldsymbol{X} = \{x_{ij}\}, i = 1, \ldots, n; j = 1, \ldots, p$ is a matrix of genotypes, $\boldsymbol{\beta} = (\beta_1, \beta_2, \ldots, \beta_p)'$ is a vector of model effects, and $\boldsymbol{\varepsilon} = (\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n)'$ is a vector of error terms. Under Gaussian assumptions, $\varepsilon_i \sim NIID(0, \sigma_\varepsilon^2)$, the likelihood function is $p(\boldsymbol{y}|\boldsymbol{\theta}) = (2\pi\sigma_\varepsilon^2)^{-\frac{n}{2}} exp\left\{-\frac{1}{2\sigma_\varepsilon^2}(\boldsymbol{y}-\boldsymbol{X\beta})'(\boldsymbol{y}-\boldsymbol{X\beta})\right\}$. Expanding the quadratic form, we can express the likelihood as a function of summary statistics $p(\boldsymbol{y}|\boldsymbol{\theta}) = (2\pi\sigma_\varepsilon^2)^{-\frac{n}{2}} exp\{\boldsymbol{y'y} + \boldsymbol{\beta X'X\beta} - 2\boldsymbol{\beta' X'y}\}$. Since data enters in the posterior distribution only through the likelihood, one can generate posterior samples using computations that are based on the summary statistics ($y'y$, $X'X$, and $X'y$. Sampling from the posterior distribution using algorithms that use the original data {y, X} requires using operators with complexity $O(n)$

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_355, © P. Pérez-Rodríguez and G. de los Campos *et al.* 2022

(here n represent sample size); these computations must be repeated for each effect in the model, leading to an algorithm with complexity $O(np)$ per cycle of the sampler. This approach, used in the BGLR() function, does not scale well for problems with $n >> p$. The BLRXy() function generate samples from the posterior distributions using the sufficient statistics {$y'y$, $X'y$, and $X'X$} as inputs; the algorithm has a complexity $O(p)$.

Multitrait models. The other addition we made was the implementation of shrinkage and variable selection methods for multi-trait regressions. The Multitrait() function implements models similar to those implemented in BGLR() for multi-traits models. The function has an user interface and a design similar to that of BGLR(). The response is specified using a matrix (observations in rows, traits in columns, missing values allowed). Predictors are specified using a list, as in BGLR, the user can include more than on term in the linear predictor. Each term can be either a multivariate linear regression (XB) or a matrix (U) of random intercepts (aka 'random effects'). For regression coefficients users can use flat priors ('FIXED'), IID Gaussian priors ('BRR'), or a prior from the spike-slab family (with a points of mass at zero and a multivariate gaussian slab, 'SpikeSlab'). All the priors, and distribution of the error terms, have (co)variance matrices as hyper-parameters. The Multitrait() function offers three specifications for priors on (co) variance matrices: unstructured (in which case and Inverse Whishart prior is used), recursive model, and factor-analytic structure (examples are provided below).

Results

Fitting genomic regressions to biobank-sized data using BLRXy(). We benchmarked the BLRXy() function against BGLR() for a linear regression model of the form $y = X\beta + \varepsilon$, with a Gaussian prior for effects, over a grid of values of sample size (from 10,000 to 250,000) and number of predictors (from 500, to 10,000). Since BLRXy() generates samples using summary statistics {$X'X$, $X'y$, $y'y$}, once these statistics are computed, the time it takes to collect a sample for all the parameters, does not depend on sample size. The left panel in Figure 1 shows the time it takes for BLRXy() to collect 1000 samples from the posterior distribution. The time grows exponentially with the number of predictors, for 10,000 predictors (e.g. SNPs), BLRXy() generates 1000 posterior samples in less than 1.5 minutes. Panel B in Figure 1 shows the time required by BLRXy() to collect 1000 samples relative to the time used by BGLR(). For problems involving $n \sim p$, the algorithm implemented in BLRXy() is approximately twice as fast than BGLR. However, for problems involving $n>p$ the proposed algorithm is one (e.g. 32 times faster, 1/0.032) or two orders of magnitude faster (e.g. 333 times faster, 1/0.003). The code is efficient when $n>p$ and this is the scenario that we want to emphasize in the benchmark.

Multi-trait genomic regression using the Multitrait() function. The scripts presented below shows three of the simplest models that can be fitted using Multirtait(): (1) Panel A shows how to fit a linear regression with a Gaussian prior; (2) Panel B presents an example of a multi-trait pedigree model, finally; (3) Panel C shows a model combining SNPs and pedigree information. By default, since the prior on (co)variance matrices were not specified, the software uses un-structured Inverse-Wishart priors. The Multitrait() function saves samples from the posterior distribution and returns estimates (posterior means and SDs). Samples from the posterior distribution are saved in files with a pre-fix specified in the argument 'save_at'. The estimated error (co)variance matrices, the corresponding posterior SD as well as the prior hyper-parameters used can be retrieved using (exemplifying with fmG) fmG$resCov. Predictions can be retrieved using fmG$ETAHat (this returns a matrix with predictions for observations in rows and traits in columns). Estimated effects can be retrieved using fmG$ETA[[1]]$beta and fmP$ETA[[1]]$u. The entries of fmG$ETA also include the estimated (co)variance matrices and the prior hyper-parameters used.

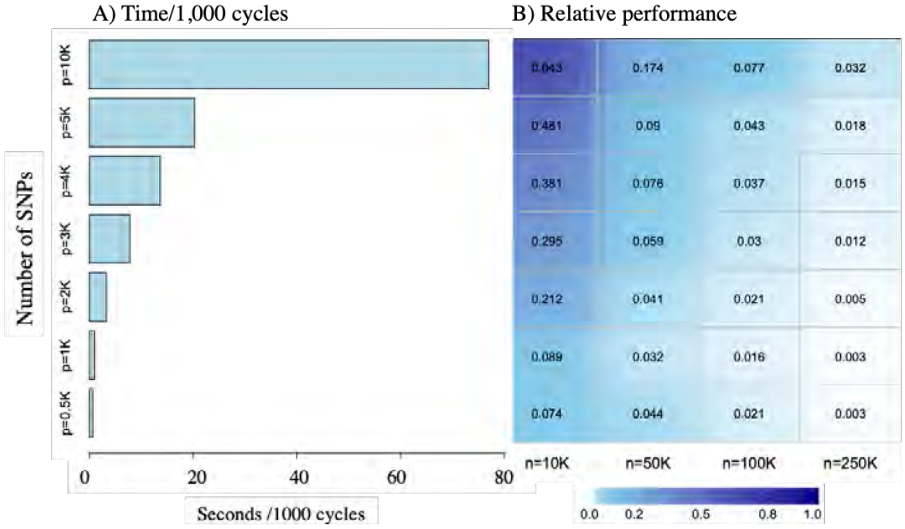


Figure 1. Computational performance BLRXy() and BGLR(). (A) Seconds needed for BLRXy() to generate 1000 posterior samples for a linear model of the form $y = X\beta + \varepsilon$; (B) Time used by BLRXy() to generate 1000 posterior samples, relative to the time used by BGLR() by sample size and number of predictors.

```
library(BGLR); data(wheat); A=wheat.A; X=scale(wheat.X); Y=wheat.Y
#A: Genomic regression
LP1<-list(list(X=X,model='BRR'))
fmG<-Multitrait(y=Y,ETA=LP1,
    nIter=10,000,burnIn=5,000,
    saveAt='snps_')
#B: Pedigree regression
LP2<-list(list(K=A,model='RKHS'))
fmP<-Multitrait(y=Y,ETA=LP2,
      nIter=10,000,burnIn=5,000,
      saveAt='ped_')
#C: Pedigree+SNPs
LP3<-list(list(X=X,model='BRR'),list(K=A,model='RKHS'))
fmG<-Multitrait(y=Y,ETA=LP3, nIter=10,000, burnIn=5,000, saveAt='pg_')
```

We benchmarked the Multitrait() function using a simple model of the form $y = X\beta + \varepsilon$ with a Gaussian prior for marker effects over a grid of values of the number of SNPs (1000 to 10,000), sample size (10,000 to 200,000) and traits (either two or four). The results are summarized in Figure 2.

Discussion

We have presented two important updates that we recently made to the BGLR package. We recommend using the BLRXy() (instead of BGLR()) for problems where sample size exceeds the number of predictors ($n > p$). In the case ($p > n$), it is better to use BGLR. One disadvantage of BLRXy() is that it requires more RAM memory than BGLR() because it requires $X'X$. The Multitrait() function fits for multiple-trait analyses using shrinkage and variable selection priors. Due to space restrictions, we presented only a very small set of the options that these functions offer. The Github repository at https://github.com/gdlc/bglr includes additional examples not covered here, such as models for variable selection.

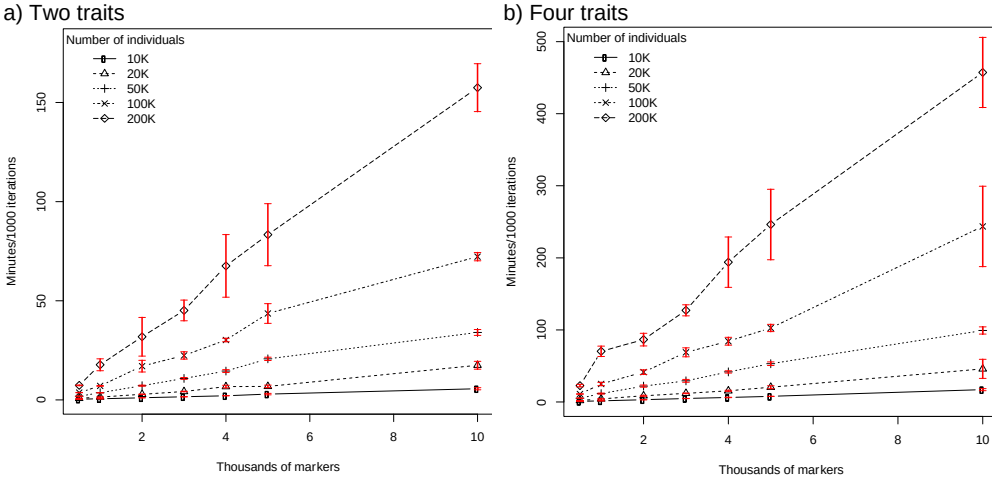


Figure 2. Average time (in minutes) to generate 1000, posterior samples by thousands of SNPs (x-axis) and sample size (K=1000), and number of traits (two or four, in the left and right panels). Intervals correspond to mean ± SD of computing times.

References

de los Campos G., Hickey J.M., Pong-Wong R., Daetwyler H.D., and Calus M.P.L. (2013) Genetics 193(2): 327-345. https://doi.org/10.1534/genetics.112.143313

Meuwissen T.H.E., Hayes B.J., and Goddard M.E. (2001) Genetics 157(4):1819-1829. https://doi.org/10.1093/genetics/157.4.1819

Pérez P., and de los Campos G. (2014) Genetics 198(2):483-495. https://doi.org/10.1534/genetics.114.164442

356. SFSI: a package to solve penalized selection indices and sparse genomic prediction

M. Lopez-Cruz[1,2*], G. de los Campos[1,2,3], B. Valente[4], C.Y. Chen[4], W. Herring[4] and A. Grueneberg[1]

[1]Department of Epidemiology and Biostatistics, Michigan State University, East Lansing, MI 48824, USA; [2]Institute for Quantitative Health Science and Engineering, Michigan State University, East Lansing, MI 48824, USA; [3]Department of Statistics and Probability, Michigan State University, East Lansing, MI 48824, USA; [4]The Pig Improvement Company, P.O. Box 55010, Winnipeg, MB, Canada; lopezcru@msu.edu

Abstract

We introduce the *SFSI* (Sparse Family and Selection Indices) R-package that offers tools for the solutions to penalized selection indices. We demonstrate how the package can be used to solve three selection index problems: (1) derivation of a selection index using high-throughput phenotype data (crop imaging); (2) single-trait; and (3) multi-trait sparse genomic prediction.

Introduction

The *SFSI* R-package solves penalized regression problems, including LASSO (Tibshirani, 1996), Ridge-Regression (Hoerl and Kennard, 1970), and Elastic Net (Zou and Hastie, 2005). There are several well-established packages available to fit penalized regressions including *glmnet* (Friedman *et al.*, 2010), *lars* (Hastie *et al.*, 2013), and *penalized* (Goeman, 2010). However, unlike other packages, *SFSI* takes as inputs covariance matrices and covariance vectors as opposed to a response vector and a matrix of predictors, which are the standard inputs of other packages. This seemingly small difference enables *SFSI* to solve a variety of problems that cannot be readily solved using standard packages for penalized regressions.

We developed early versions of the package to solve penalized selection indices for indirect selection for a breeding objective using high-dimensional traits (e.g. high-throughput phenotypes) as predictors. We first introduced these methods in Lopez-Cruz *et al.* (2020) and labeled this methodology as sparse selection index (SSI). It turns out that an objective function with the same structure can be used to obtain a sparse solution for genomic prediction. In Lopez-Cruz and De Los Campos (2021), we showed how this approach can be used to produce genomic predictions that rely on information from subsets of the data available for model training.

In the 'Methods' section of this study, we describe the methods implemented in *SFSI* and show how they can be used to derive penalized selection indices for mass selection and sparse single- and multi-trait genomic prediction. Subsequently, in the 'Results' section, we present applications based on real data. We close the article with a discussion on the advantages and limitations of SFSI.

Materials & methods

Consider a linear regression where a response vector, $y = \{y_1, \ldots, y_n\}$, is regressed on p predictors, $\mathrm{X} = \{x_{ij}\}$ (an $n \times p$ matrix), using a model of the form $y = \mathrm{X}\beta + \varepsilon$, where $\beta = \{\beta_1, \ldots, \beta_p\}$ is a p-vector of regression coefficients and $\varepsilon = \{\varepsilon_1, \ldots, \varepsilon_n\}$ is an n-vector of random error terms. In penalized regression, the coefficients are derived by minimizing an objective function that includes the residual sum of squares, $RSS = (y - \mathrm{X}\beta)'(y - \mathrm{X}\beta)$, plus a penalty function which often is the L1-norm: $\sum_{j=1}^{p} |\beta_j|$ (LASSO), the squared L2-norm: $\sum_{j=1}^{p} \beta_j^2$ (Ridge Regression), or a linear combination of the two (Elastic Net). The Elastic Net penalized RSS problem can be expressed as:

$$\hat{\beta} = arg\ \underset{\beta}{min}\left\{\tfrac{1}{2}\beta'\Sigma\beta - \Gamma'\beta + \lambda \mathrm{F}(\beta)\right\} \quad (1)$$

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_356, © M. Lopez-Cruz *et al.* 2022

where $\Sigma = X'X$, $\Gamma = X'y$, λ is a penalization parameter, and $F(\beta)$ is a penalty function on β given by $F(\beta) = \alpha \sum_{j=1}^{p} |\beta_j| + \frac{1}{2}(1-\alpha)\sum_{j=1}^{p} \beta_j^2$, with α being a number between 0 and 1 controlling the weights of the L1 and L2 penalties. Upon standardization, matrices Σ and Γ can be interpreted as method-of-moments estimates of the covariance matrix of the predictors, and the covariance matrix between predictors and the response, respectively.

In the SFSI package the functions lars2 and solveEN can be used to obtain estimates by solving the optimization problem in Equation 1 taking Σ and Γ as inputs via the Least Angle Regression (Efron *et al.*, 2004) and the coordinate descent algorithm (Friedman *et al.*, 2010), respectively. The fact that we use Σ and Γ as inputs gives *SFSI* great flexibility in terms of the problems that can be solved with it. We provide examples in the following sections.

Application 1: Sparse selection indices for mass selection. A selection index (SI, Hazel, 1943) predicts the genetic value (g_i) of a selection candidate for a breeding objective (y_i) as a weighted sum of p indicator (i.e. measured) traits, $\boldsymbol{x}_i = \{x_{i1}, \ldots, x_{ip}\}$, as $\hat{g}_i = \boldsymbol{w}'\boldsymbol{x}_i$. In a standard SI the weights, $\boldsymbol{w} = \{w_1, \ldots, w_p\}$, are derived by minimizing the mean squared prediction error, $MSPE = \frac{1}{2}\mathbb{E}(g_i - \boldsymbol{w}'\boldsymbol{x}_i)^2 = \frac{1}{2}\boldsymbol{w}'\mathbf{P}_x\boldsymbol{w} - \mathbf{G}'_{xy}\boldsymbol{w}$, wh, where P_x is the phenotypic (co)variance matrix of the measured traits, and G_{xy} is a vector with genetic covariances between the selection objective and each measured trait. The SI methodology was originally developed for problems involving a small number of measured traits. High-throughput technologies can produce thousands of indicator traits (e.g. imaging, sensors, gene expression). This can lead to over-fitting and thus poor accuracy of indirect selection. To address this problem, we (Lopez-Cruz *et al.,* 2020) developed a sparse selection index (SSI) which derives the weights using:

$$\hat{\boldsymbol{w}} = arg\ \underset{\boldsymbol{w}}{min}\left\{\frac{1}{2}\boldsymbol{w}'\mathbf{P}_x\boldsymbol{w} - \mathbf{G}'_{xy}\boldsymbol{w} + \lambda F(\boldsymbol{w})\right\} \quad (2)$$

Above, $F(_w)$ is a penalty function such as the one in Equation 1. To derive an SSI, one needs to: (1) obtain estimates of P_x and G_{xy}; (2) derive the weights $\hat{\boldsymbol{w}}$ by solving Equation 2 (this can be done using the SFSI); (3) evaluate the SSI for the selection candidates, $\hat{\boldsymbol{g}} = \mathbf{X}\hat{\boldsymbol{w}}$, where $X = \{x_{ij}\}$ is the $n \times p$ matrix containing all the measured traits. Once P_x and G_{xy} have been estimated, steps (2) and (3) can be implemented using the following script:

```
fm = solveEN(Px, Gxy, lambda=lambda)
gHat = fitted(fm, X)
```

Application 2: Sparse genomic prediction. In genomic prediction (Meuwissen *et al.*, 2001) via GBLUP the genetic merit of selection candidates (TST=testing) is predicted as a weighted sum of the values of individuals in a training set (TRN=training, y_{TRN}). Indeed, $\hat{\boldsymbol{g}}_{\text{TST}} = \mathbf{W}\boldsymbol{y}_{\text{TRN}}$,, where $\mathbf{W} = \mathbf{G}_{\text{TST,TRN}}(\mathbf{G} + \theta\mathbf{I})^{-1}_{\text{TRN}}$ is an $n_{\text{TST}} \times n_{\text{TRN}}$ matrix of weights with $G_{\text{TST,TRN}}$ being the genetic relationships matrix (GRM) between TST and TRN individuals, $(\mathbf{G} + \theta\mathbf{I})_{\text{TRN}}$ a diagonal-displaced GRM of the TRN set, and $\theta = \sigma^2_\varepsilon / \sigma^2_g$ the residual/genetic variances ratio. GBLUP predictions can be shown to be equivalent to a standard SI where the 'measured traits' are the phenotypes measured in the TRN individuals (Lopez-Cruz and de los Campos, 2021).

The presence of imperfect LD between markers and QTL, non-additive effects, and population structure can make a GBLUP sub-optimal. To deal with this, Lopez-Cruz and de los Campos (2021) proposed an SSI in which the prediction of the i_{th} TST individual is a weighted sum of a subset of the data from the TRN set. Sparsity was achieved by deriving weights using:

$$\widehat{\boldsymbol{w}}_i = arg\ \underset{\boldsymbol{w}_i}{min}\left\{\frac{1}{2}\boldsymbol{w}_i'(\mathbf{G}+\theta\mathbf{I})_{\text{TRN}}\boldsymbol{w}_i - \mathbf{G}_i'\boldsymbol{w}_i + \lambda F(\boldsymbol{w}_i)\right\} \quad (3)$$

where $\mathbf{G}_i = \mathbf{G}'_{\text{TST}(i),\text{TRN}}$ is the i^{th} row of $G_{\text{TST,TRN}}$. These sparse genomic predictions are obtained following the steps: (1) derive the matrix G; (2) estimate the variances ratio θ; (3) determine $G_{\text{TST,TRN}}$, $(G + \theta I)_{\text{TRN}}$, and y_{TRN} from data; (4) derive the weights $\widehat{\boldsymbol{w}}_i$ for each TST subject by solving n_{TST} times the problem in Equation 3 to get $\widehat{\mathbf{W}} = \{\widehat{w}_{ij}\}$; (5) obtain the predictions $\widehat{\boldsymbol{g}}_{\text{TST}} = \widehat{\mathbf{W}}\boldsymbol{y}_{\text{TRN}}$. After obtaining G and θ, steps (3)-(5) can be implemented (for a TST and TRN set indexed with integers vectors) using the *SSI* function as

```
fm = SSI(y,K=G,theta=theta,trn=TRN,tst=TST,lambda=lambda)
gHat = fitted(fm)
```

Application 3: Multi-trait sparse genomic prediction. In a multi-trait genetic model for s traits, $\boldsymbol{y} = \{\boldsymbol{y}_1, \ldots, \boldsymbol{y}_s\}$, the (co)variance of the genetic, $\boldsymbol{g} = \{\boldsymbol{g}_1, \ldots, \boldsymbol{g}_s\}$, and residual, $\boldsymbol{\varepsilon} = \{\boldsymbol{\varepsilon}_1, \ldots, \boldsymbol{\varepsilon}_s\}$, terms take the form var(g) = K ⊗ G and var(ε) = R ⊗ I, respectively. Here, K and R are $s \times s$ within-subject genetic and error (co)variances matrices, and ⊗ denotes the Kronecker product. Thus, a small modification of the script presented above can be used derive SSIs for multi-trait problems. After computing G, K and R, step (4) in the single-trait case is performed by replacing G by K ⊗ G, I by R ⊗ I, and making $\theta = 1$. For three traits, y_1, y_2, y_3, this is implemented as:

```
y = c(y1,y2,y3)
KG = kronecker(K,G)
RI = kronecker(R,diag(nrow(G)))
fm = SSI(y,K=KG,theta=1,D=RI,trn=TRN,tst=TST,lambda=lambda)
```

Results & discussion

In the following section we present an application using real data generated by the Pig Improvement Company (PIC, https://www.pic.com). We use this data to illustrate the implementation of the sparse genomic prediction presented in Application 2 in *Methods* section.

The data set comprises weight gain records for n=22,872 pig genotypes and a GRM G constructed from 34,373 SNP markers. We performed sparse predictions for a random 20% (n_{TST}=4,574) of the data using the remaining 80% (n_{TRN}=18,298) for model training. We used 100 different values of λ as in Lopez-Cruz and de los Campos (2021). The accuracy of the predictions $\widehat{\boldsymbol{g}}_{\text{TST}} = \widehat{\mathbf{W}}\boldsymbol{y}_{\text{TRN}}$ was evaluated as the correlation between $\boldsymbol{y}_{\text{TST}}$ and $\widehat{\boldsymbol{g}}_{\text{TST}}$. We also derived the GBLUP (i.e. the SSI with λ=0) using sub-samples consisting of 10, 20, …, 100% of the TRN set. We performed a total of 100 TRN-TST random partitions.

The results in Figure 1A show the prediction accuracy achieved by the SSI as a function of the sparsity of the index (i.e. the number of support points, that is the number of training data points with non-zero weight in the SI). As a reference, the boxplots in the same plot show the prediction accuracy achieved by the GBLUP for the same number of support points selected at random. The results show that including a sparsity-inducing penalty into the genomic prediction framework can provide sparse solutions with the same accuracy while using a reduced number of support points relative to the non-sparse approach. However, the best prediction accuracy achieved in this application was comparable to that of a GBLUP using the maximum sample size available. For other data sets, we have reported results where the SSI gave moderate gains in prediction accuracy (5-10%) over the GBLUP (Lopez-Cruz and De Los Campos, 2021).

Computational considerations: computational cost of implementing these applications depends on the type of the problem. For Application 1, the computational burden is due to the calculation of genetic covariances

G_{xy} (in Equation 2) because this requires fitting a sequence of bi-variate mixed models (as many measured traits). Computing the weights in this problem is not computationally demanding because the weights are solved only once. For Applications 2 and 3, the computational cost of computing an SSI for n_{TST} subjects become expensive as it involves solving n_{TST} penalized problems. However, this type of SSI allows solving each penalized problem independently and, therefore, it can be fully parallelized. The computational time of solving one penalized regression depends on the TRN set size and the value of λ. In a Linux-based system (2.4 GHz Intel Xeon Gold processor, 32 GB of RAM), solving one penalized problem in Equation 1 using n_{TRN}=10,000 training individuals takes on average 2.1 seconds for a very sparse (large λ) model, and around 8.01 seconds for a less sparse (smaller λ) model (Figure 1B).

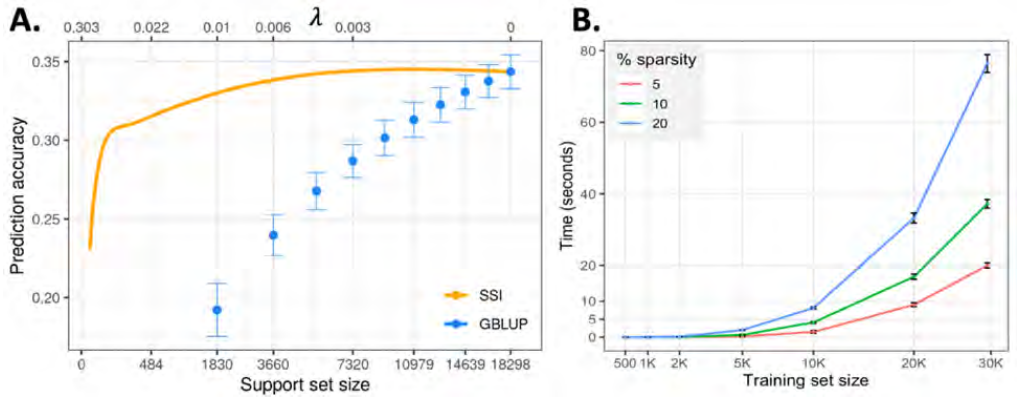


Figure 1. (A) Prediction accuracy of the SSI (orange line) versus the average number of support points given by the value of λ (top x-axis). Blue boxplots correspond to the accuracy of the GBLUP trained with 10, 20, ..., 100% of the TRN set (n_{TRN}=18,298). (B) Average computational time of solving one penalized problem (Equation 1) with the solveEN function by TRN set size and value of λ. Σ ($p \times p$) matrix) and Γ(p-vector) were sampled 100 times from the G matrix for 29,473 subjects used in Lopez-Cruz and de los Campos (2021), for p=500, 1K, 2K, 5K, 10K, 20K, and 29,472. The value of λ was chosen such that the model produced 5, 10 or 20% of sparsity (ratio between number of support points and p).

References

Efron B., Hastie T., Johnstone I., and Tibshirani R. (2004) Ann. Stat. 32(2): 407–99. https://doi.org/10.1214/009053604000000067

Friedman J., Hastie T., and Tibshirani R. (2010) J. Stat. Softw. 33(1): 1–22. https://doi.org/10.18637/jss.v033.i01

Goeman J.J. (2010) Biom. J. 52(1): 70-84. https://doi.org/10.1002/bimj.200900028

Hastie T., and Efron B. (2013). Lars: Least Angle Regression, Lasso and Forward Stagewise. Available at https://CRAN.R-project.org/package=lars

Hazel L.N. (1943) Genetics 28(6): 476–90. https://doi.org/10.1093/genetics/28.6.476

Hoerl A.E., and Kennard R.W. (1970) Technometrics 12(1): 55–67. https://doi.org/10.1080/00401706.1970.10488634

Lopez-Cruz M., and de los Campos G. (2021). Genetics 218(1): 1–10. https://doi.org/10.1093/genetics/iyab030

Lopez-Cruz M., Olson E., Rovere G., Crossa J., Dreisigacker S., et al. (2020) Sci. Rep. 10(8195). https://doi.org/10.1038/s41598-020-65011-2

Meuwissen T.H.E., Hayes B.J., and Goddard M.E. (2001) Genetics 157(4): 1819–1829. https://doi.org/10.1093/genetics/157.4.1819

Tibshirani R. (1996) J. R. Stat. Soc. B 58(1): 267–288. https://doi.org/10.1111/j.2517-6161.1996.tb02080.x

Zou H., and Hastie T. (2005) J. R. Stat. Soc. B 67(2): 301–20. https://doi.org/10.1111/j.1467-9868.2005.00503.x

357. INTERGEN: an efficient and flexible tool for large-scale genetic evaluation of complex animal and plant populations

V.S. Junqueira[1], G.S. Campos[2,3] and F.F. Cardoso[3*]

[1]Bayer Crop Science, Research and Development, BR 452, 38407-049 Uberlândia, MG, Brazil; [2]Departamento de Zootecnia, Universidade Estadual Paulista 'Júlio de Mesquita Filho', Via de Acesso Prof. Paulo Castellane s/n, 14884-900 Jaboticabal, SP, Brasil; [3]Embrapa Pecuária Sul, CP242, 96401-970 Bagé, RS, Brazil; fernando.cardoso@embrapa.br

Abstract

The INTERGEN package contains a wide range of statistical analyses useful for routine genetic evaluations. The package has been under constant development for enabling new models, increasing speed, and improving the efficiency of solving mixed model equations for genetic evaluations. The package was initially developed to perform Bayesian analyses, reaction norm models, multi-breed, and animal models with parentage uncertainty. More recently, it was expanded in version 1.4 to include Best Linear Unbiased Prediction (BLUP) with (or not) genomic information, approximate accuracies, and Iteration on Data (IOD) methods.

Introduction

The INTERGEN package is a collection of Fortran 90/95 routines useful for breeding and genetics analyses. This package contains a wide range of relevant statistical analyses and is currently available in three binaries. The first binary contains functionalities for conducting Bayesian analyses such as parentage uncertainty under multiple sire mating, multi-breed analysis, genotype-by-environment interaction, heterogeneity of residual variances, robustness analysis, genetic merit estimation including (or not) genomic information using mixed model equations, reduced animal model, reaction norm models, and normal prior for fixed effects. The second binary contains routines for approximate accuracies with (or not) genomic information for direct and maternal effects. A third binary is focused on estimating the genetic merit using Iteration on Data (IOD) method.

Materials & methods

The initial development of INTERGEN targeted a novel implementation of a series of analyses for studying genotype-by-environment using reaction norm models with unknown covariates in multi-breed populations with uncertain paternity (Cardoso, 2008). The initial software used several modules (SPARSEM, SPARSEOP, IOUNF, DENSEOP, and Gibbs) from Misztal (1999) useful for handling mixed models (e.g. building, in-memory storage, and solving) using Fortran 77/90 (Misztal 1999, 2002) as starting point. Subsequently, a set of non-public modules, subroutines, and functions were developed to enable below statistical models useful to handle the diversity of management situations found worldwide in livestock genetic evaluations, including:

- animal model with parentage uncertainty and multiple sire mating schemes;
- multi-breed populations;
- heavy-tailed structural models for robust and heterogeneous residual variances;
- reaction norm models with unknown covariates; and
- reduced animal model.

These implementations were made using the MCMC method via Gibbs and Metropolis-Hastings sampling. To our knowledge, no other software combines all these implementations with flexible model specification. Over the past decade, the INTERGEN package has been under constant development for handling

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_357, © V.S. Junqueira *et al.* 2022

additional models commonly used in routine genetic evaluations worldwide. These implementations are listed below:

- use of normal priors for fixed effects;
- mixed-model solutions (BLUE/BLUP) using in-memory equations (Misztal *et al.* 2002);
- threshold models (Gianola 1982);
- genomic module for single step GBLUP (ssGBLUP) models (Aguilar *et al.* 2010);
- routines for calculation of approximate accuracies (Misztal *et al.* 1993, 2013);
- solutions of mixed-models with Iteration-On-Data (IOD) using Preconditioned Conjugate Gradient (PCG) (Misztal 1999; Strandén and Lidauer 1999).

Currently, the INTERGEN package supports analysis using phenotypes, genotypes, and pedigree information. The software is flexible enough to handle complex models including multi-trait (non)correlated traits with additive, maternal, and non-additive effects. Three binaries that made up the software toolkit are: intergen1.4, the most complete option as it contains the majority of the functionalities for running Bayesian analysis and mixed-models considering a wide range of model effects; intergenacc, focused on estimating approximate accuracies; and, finally, intergeniod, relies on IOD methods to solve mixed model equations.

Genomic module. The genomic module contains a mix of sparse and dense matrices depending on the operation step aiming to optimize computing time and performance. The most time-consuming steps of matrices operations (multiplication and inversion) of this module were developed for using parallel programming in a combination of Intel MKL routines and OpenMP.

Computing environment. Intergen is available as a binary compiled under Unix (Linux and Mac) and Windows systems. The compilation of Fortran modules and subroutines are automated through Makefiles (GNU Make 4.2.1). Compilations are currently being done using Intel Ifort (Version 19.1.3.304).

Results

The statistical analyses implemented in INTERGEN have been used routinely for genetic evaluations in Brazil, being the official tool for statistical inference, variance component and genetic merit estimation in Pampaplus (Hereford and Braford) and PROMEBO (Angus, Brangus, Braford, Charolais, Devon and Hereford) programs over the past 6 years. In addition to this commercial use, several research publications have been published over the years in a wide range of species using the package.

As an example, a summary of the performance of the software using phenotypes, genotypes, and pedigree from an Angus dataset for estimating mixed model solutions, exact accuracies (direct inversion of the coefficient matrix), and approximate accuracies are presented in Table 1. Correlation between matrix solving and IOD solutions was 0.99.

The overall correlation between exact and approximate accuracies using Angus weaning weight phenotypes, genotypes, and pedigree was 0.97 and distributions can be seen in Figure 1. Note that approximate accuracies follow a similar trend as exact accuracies.

Discussion

INTERGEN package is freely available for research purposes and as an alternative for commercial genetic evaluations in animals and plants. The list of functionalities of the package is long and detailed information is fully presented in its manual (https://ainfo.cnptia.embrapa.br/digital/bitstream/item/202597/1/DT-153-online.pdf) to enable complex modeling required for running routine genetic evaluations and for research statistical-genetics analyses. The approximate accuracies (intergenacc) and IOD solutions (intergeniod)

Table 1. Description of the dataset used for evaluating the performance for mixed model solutions (GEBV), exact accuracy and approximate accuracy estimation for different beef cattle traits. A total of 5,588 genotyped samples and 106,121 SNPs were used in the analysis.

Trait[1]	Type of analysis	Pedigree depth	Number of phenotypes	Rounds[2]	Timing[5]		
					GEBV estimation	Exact accuracy[3]	Approximate accuracy[4]
WW	continuous	406,101	268,684	132	0:02:43	2:32:54	0:05:30
BW	continuous	405,267	140,437	202	0:03:39	1:45:58	0:05:29
PWG	continuous	404,236	182,029	190	0:01:01	0:36:35	0:05:29
WHC	discrete	403,365	97,237	176	0:00:44	0:22:54	0:05:28
YHC	discrete	397,949	79,319	83	0:00:21	0:17:38	0:05:23

[1] All beef cattle traits. WW: weaning weight, BW: birth weight, PWG: post-weaning weight, WHC: weaning hair coat, YHC: yearling hair coat.

[2] Number of rounds during PCG iterative convergence.

[3] Exact accuracies were estimated as the diagonal of the inverse of coefficient matrix from mixed model equations.

[4] Approximate accuracies were estimated following an iterative algorithm implemented in-house.

[5] Timing formatted as hours:minutes:seconds.

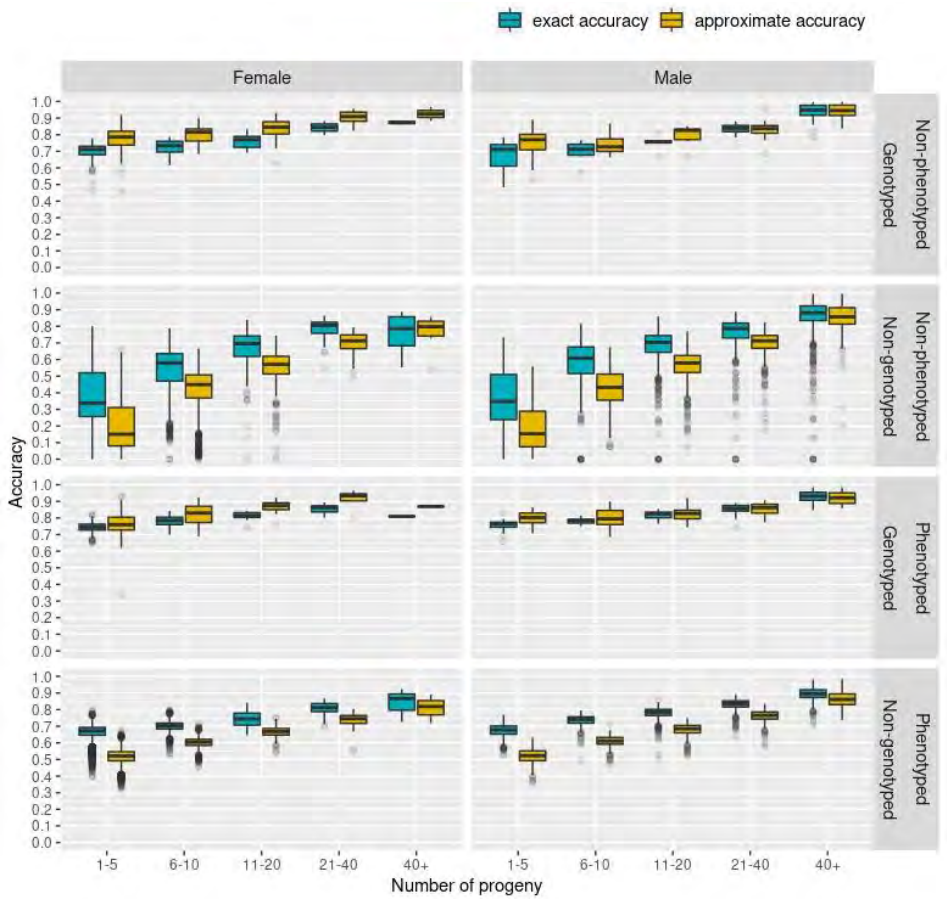


Figure 1. Distribution of exact and approximate accuracies for Angus male and female animals with different number of progeny, and different information available for estimating the genetic merit. Approximate accuracies estimated using intergenacc.

have very high correspondence to accuracies and solutions obtained by inverting the mixed model equations, being a suitable alternative for large-scale genetic evaluations involving millions of animals and equations.

Software distribution

INTERGEN package is distributed for Linux, Mac, Windows platforms (https://arquivos.cppsul.embrapa.br/downloads/intergen/). The package is free for research purposes but should be cited in publications. The intergenacc and intergeniod are available upon research or commercial agreement. The terms of conditions for commercial use (e.g. routine genetic evaluations) can be obtained from Labegen, Embrapa Pecuária Sul, Bagé, Rio Grande do Sul, Brazil (fernando.cardoso@embrapa.br).

Acknowledgements

To Dr Ignacy Misztal for making the initial modules available and to Dr Ignacy Aguilar and Dr Rodrigo Mota for contribution in the early phases of the software development. V.S. Junqueira acknowledges Dr Daniela Lourenço for hosting him during visiting research period at the University of Georgia. Research partially funded by Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), Grant/Award Number 305102/2018-4.

References

Aguilar, I., I. Misztal, D. L. Johnson, A. Legarra *et al.*, (2010) J. Dairy Sci. 93**:** 743-752. https://doi.org/10.3168/jds.2009-2730.

Cardoso, F. F. (2008) Embrapa Pecuária Sul: Bagé. 64p. Available at: https://ainfo.cnptia.embrapa.br/digital/bitstream/item/55814/1/DT74.pdf.

Gianola, D., (1982) J. Anim. Sci. 54(5)**:** 1079-1096. https://doi.org/10.2527/jas1982.5451079x.

Misztal, I., (1999) Interbull Bulletin (20)**:** 33-42. Available at: https://journal.interbull.org/index.php/ib/article/view/470.

Misztal, I., T. J. Lawlor and T. H. Short, (1993) J. Dairy Sci. 76**:** 1421-1432. https://doi.org/10.3168/jds.S0022-0302(93)77473-1.

Misztal, I., S. Tsuruta, I. Aguilar, A. Legarra *et al.*, (2013) J. Dairy Sci. 96**:** 647-654. http://dx.doi.org/10.3168/jds.2012-6272.

Misztal, I., S. Tsuruta, T. Strabel, B. Auvray *et al.* (2002). Proc. of the 7th WCGALP, Montpellier, France.

Strandén, I., and M. Lidauer, (1999) J. Dairy Sci. 82**:** 2779-2787. https://doi.org/10.3168/jds.S0022-0302(99)75536-0.

358. MiXBLUP 3.0 – software for large genomic evaluations in animal breeding programs

J. Vandenplas[1*], R.F. Veerkamp[1], M.P.L. Calus[1], M.H. Lidauer[2], I. Strandén[2], M. Taskinen[2], M. Schrauf[1] and J. ten Napel[1]

[1]*Wageningen University & Research, P.O. Box 338, 6700 AH Wageningen, the Netherlands;* [2]*Natural Resources Institute Finland (Luke), Jokioinen, Finland;* *jeremie.vandenplas@wur.nl*

Abstract

The software package MiXBLUP 3.0 allows to efficiently estimate (genomic enhanced) breeding values in livestock using various linear mixed effect models. MiXBLUP aims to be user-friendly while supporting efficient algorithms for solving large genomic evaluations. Originally developed for pedigree-based models, MiXBLUP has been extended to four different approaches for single-step genomic evaluations, allowing simultaneous analyses of all information for both genotyped and ungenotyped individuals. For all approaches, shared-memory parallelism is supported using tailor-made procedures and parallel libraries. It has been tested on some of the largest genomic datasets in the world. The software is developed by Wageningen University and Research in collaboration with the Natural Resources Institute Finland (Luke).

Introduction

During the last decades, the mixed model equations have been used for estimating breeding values in livestock using phenotypes at the traits of interest, together with pedigree. With the advent of genomic selection, the traditional mixed model has been extended to analyse simultaneously phenotypic, pedigree, and genomic information (Legarra *et al.*, 2014). Such models are commonly called single-step genomic models. Currently, genomic information used in routine evaluations contains the genotypes of several (hundred) thousand animals for around 50 thousand single nucleotide polymorphisms (SNP). The integration of such genomic datasets with large numbers of genotyped animals led to new challenges for solving the mixed model equations. Therefore, several approaches for solving the single-step models have been proposed in the literature and implemented in software for large genomic evaluations in livestock (e.g. Mäntysaari *et al.*, 2020).

Since 2006, the MiXBLUP software has been developed for breeding value estimation in commercial genetic programmes. It supports modern developments, such as complex models (e.g. random regression models) or linear-threshold models. With the advent of genomic selection, MiXBLUP has been continuously developed since 2014 to support the use of genomic information through four different single-step genomic approaches. This work aims to present the state-of-the-art of MiXBLUP for large genomic evaluations in animal breeding.

MiXBLUP and its components

MiXBLUP philosophy. The aim of developing MiXBLUP is to utilize efficient computing strategies for solving mixed model equations, while being easy to install and easy to use, even for sophisticated models. For example, MiXBLUP supports a wide range of input files and generates files in the required format for the solvers. So, various formats for the SNP genotype files are supported. When the genotype files are text files, they may contain marker alleles or marker genotypes. The marker alleles or genotypes may be in space-separated or in dense format. MiXBLUP now supports the Plink 1 binary format that requires less memory than text files by taking advantage that only two bits (instead of one byte) are needed to store a biallelic SNP genotype (Chang *et al.*, 2015). Due to this advantage, the Plink 1 binary format is also supported by other programs of the software package MiXBLUP described hereafter.

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_358, © J. Vandenplas *et al.* 2022

Being a versatile software, MiXBLUP is used in breeding programmes for cattle, pigs, poultry, sheep, horses, fish, goats, and dogs (ten Napel *et al.*, 2021).

Models supported for breeding value estimation. The statistical method commonly used for routine genetic evaluation is the so-called Best Linear Unbiased Prediction (BLUP). Briefly, a univariate animal model is as follows:

$$y = Xb + Wa + e,$$

where y is the vector of phenotypes, b is the vector of fixed effects, a is the vector of additive genetic effects, and e is the vector of residuals. The incidence matrices X and W relate the phenotypes to the fixed effects and additive genetic effects, respectively.

It is assumed that e follows a multivariate normal distribution, $MVN(0, R\ \sigma_e^2)$ with R being a diagonal matrix, and σ_e^2 being the residual variance, and that a follows a multivariate normal distribution $MVN(0, H\ \sigma_a^2)$ with σ_a^2 being the additive genetic variance. Traditionally, the H matrix was equal to the pedigree-based relationship matrix (A). In single-step genomic BLUP, the H matrix is a relationship matrix that includes the genomic relationship matrix (G) for genotyped animals and modified pedigree relationships for other animals.

Due to limitations related to the computation of G and its inverse (G^{-1}) when the number of genotyped animals is large, some methods were proposed to approximate G^{-1}, such as the algorithm for proven and young animals (APY; Misztal *et al.*, 2014), or to compute G^{-1} implicitly using the Woodbury matrix identity (Mäntysaari *et al.*, 2017). Such models that rely on G will hereafter be referred to as single-step genomic BLUP (ssGBLUP). Another proposed approach to avoid the computation of G or G^{-1}, is to augment the solution vector of the mixed model equations with SNP effects. The SNP effects can be fitted explicitly or implicitly in the underlying model as random effects (e.g. Liu *et al.*, 2014). Such models will hereafter be referred to as single-step SNPBLUP (ssSNPBLUP).

In addition to the original ssGBLUP for which G is explicitly inverted, the software MiXBLUP supports two ssGBLUP-based and one ssSNPBLUP approaches. For all approaches, a residual polygenic (RPG) effect that considers the part of the additive genetic effects not explained by the genomic data (Mäntysaari *et al.*, 2020) can be fitted.

First, MiXBLUP supports the algorithm for proven and young animals (APY; Misztal *et al.*, 2014). This algorithm approximates G^{-1} by dividing the genotyped animals into a limited set of so-called core animals, and the remaining genotyped animals are noncore. The computation of $\mathbf{G}_{APY}^{-1}$ involves the inversion of a genomic relationship submatrix among the core animals and the recursive computation of other coefficients for noncore animals.

Second, MiXBLUP supports the computation of the so-called T matrix, needed for an implicit inversion of G using the Woodbury matrix identity (ssGTABLUP; Mäntysaari *et al.*, 2017). Briefly, when an RPG effect is fitted, G has the form $G = ZBZ' + wA_{gg}$, where Z is a matrix of centered SNP genotypes, B is a diagonal matrix with diagonal entries scaling the genetic variance, w is the proportion of the RPG effect, and A_{gg} is the pedigree-based relationship matrix among genotyped animals. According to the Woodbury matrix identity:

$$G^{-1} = 1/w\ A_{gg}^{-1} - TT' \quad (1)$$

where $T = 1/w\ L^{-1}\ Z'A_{gg}^{-1}$, with L being the Cholesky factor of $K = (1/w\ Z'\ A_{gg}^{-1}Z + B^{-1})$.

In practice, G^{-1} is not computed explicitly, because the MiXBLUP iterative solvers only require its multiplication by a vector, which can be computed efficiently in multiple parts by using the matrices A_{gg}^{-1} and T (Mäntysaari *et al.*, 2017).

Third, MiXBLUP supports the ssSNPBLUP system proposed by Liu *et al.* (2014). This system has the specificity that the random SNP effects are not explicitly fitted in the model. Indeed, the solution vector that includes traditionally fixed and random effects is augmented with the SNP effects, and an inverse of the (co)variance matrix associated with the random additive genetic and SNP effects is derived. It has been shown that the absorption of the equations associated with the SNP effects results in the ssGTABLUP system (Mäntysaari *et al.*, 2020).

The MiXBLUP software provides several options for modelling base populations, such as the concept of metafounders (Legarra *et al.*, 2015), and genetic groups. Different options for scaling genomic relationships to the same base as pedigree-based relationships are also available.

Data processing. Since the users can provide a wide range of data file formats, MiXBLUP will first convert all data files in a format that its two solvers and other programs can read. Indeed, once converted, MiXBLUP will call different parallel programs for computing the various matrices required by the solvers. For example, MiXBLUP will call the program calc_grm, which has an important role in single-step genomic evaluations because it computes relationship matrices for single or multiple, related or unrelated base populations.

Two different solvers. The MiXBLUP software supports two different solvers. The default solver is derived from the MiX99 software (Lidauer *et al.*, 2019). This software was initially developed for pedigree-based genetic evaluations by the National Resources Institute Finland (Luke). With the advent of genomic selection, specifically the single-step models, both software MiX99 and MiXBLUP were adapted for using genomic information simultaneously with pedigree and phenotypic information. The MiX99 software is optimized for solving pedigree-BLUP, ssGTABLUP, and ssGAPYBLUP.

Recently, a solver called hpblup (Vandenplas *et al.*, 2020) was developed within the consortium Breed4Food and added to MiXBLUP. This program was specifically developed and optimized for solving the ssSNPBLUP system of Liu *et al.* (2014).

Computations. For efficient computations, both solvers support shared-memory parallelism through the OpenMP API (www.openmp.org) and various parallelized libraries. Furthermore, both solvers use the preconditioned conjugate gradient (PCG) algorithm as an iterative method. The program hpblup also supports a two-level PCG algorithm for efficiently solving the ssSNPBLUP linear system (Vandenplas *et al.*, 2020).

The main computational cost of a PCG iteration is the multiplication of the coefficient matrix of a system of equations by the so-called direction vector. For both solvers, this multiplication is performed in three main steps (in different orders for each solver). The first step is the multiplication of the least square part of the coefficient matrix by a vector; the second step involves all multiplications of matrices related to pedigree information, and the third step involves all multiplications of matrices related to genomic information. This third step is usually the most computationally expensive for large evaluations, and shared-memory parallel computing has been implemented in both solvers to improve their efficiency. For ssGAPYBLUP and ssGTABLUP, the computations related to genomic information mainly rely on parallel libraries, such

as the Intel Math Kernel Library. For ssSNPBLUP, a tailor-made parallel procedure has been developed for the multiplication of the genotype matrix by another matrix, while keeping the genotype matrix in the Plink 1 binary compressed format (Vandenplas *et al.*, 2020). More details in the different computations of both solvers can be found in Strandén *et al.* (1999), Mäntysaari *et al.* (2017), Lidauer *et al.* (2019), and Vandenplas *et al.* (2020).

Other programs. In addition to the two solvers, MiXBLUP also includes other programs such as calc_grm, MiXPRED, and reliabilities.exe. The program calc_grm is a program for computing pedigree and genomic relationship matrices, their inverses, or some related matrices, such as $\mathbf{G}_{APY}^{-1}$ and T. This program is also used for converting a text genotype file into Plink 1 binary files, and *vice versa*. The program MiXPRED allows the computation of interim GEBV of newly genotyped animals based a decomposition of the GEBVs obtained from a previous single-step evaluation into its different components, namely the direct genomic breeding values, the residual polygenic effects, and if fitted, other effects such as the genetic groups. The program reliabilities.exe allows the computation of approximated reliabilities for all animals in an evaluation, based on the algorithm of Tier and Meyer (2004).

System requirements and availability. The MiXBLUP software is available for Microsoft Windows and Linux/Unix environments and can be downloaded from the website www.mixblup.eu. MiXBLUP works with a license, which can be ordered through its website. More information can be found on www.mixblup.eu.

Acknowledgements

This study was financially supported by the Dutch Ministry of Economic Affairs (TKI Agri & Food Project 16022) and the Breed4Food partners Cobb Europe, CRV, Hendrix Genetics, and Topigs Norsvin.

References

Chang C.C., Chow C.C., Tellier L.C., Vattikuti S., Purcell S.M., *et al.* (2015). Gigasci. 4:7. https://doi.org/10.1186/s13742-015-0047-8

Legarra, A., Christensen, O.F., Aguilar, I. and Misztal, I. (2014). Livest. Sci. 166:54-65. https://doi.org/10.1016/j.livsci.2014.04.029.

Legarra, A., Christensen, O.F., Vitezica, Z.G., Aguilar, I. and Misztal, I. (2015) Genetics 200:455–468. https://doi.org/10.1534/genetics.115.177014.

Lidauer, M., Matilainen, K., Mäntysaari, E., Pitkänen, T., Taskinen, M., *et al.* (2019) MiX99: Technical reference guide for MiX99 solver. Release XI/2019. Available at: https://jukuri.luke.fi/bitstream/handle/10024/545540/refguide_solver.pdf?sequence=1&isAllowed=y

Liu, Z., Goddard, M., Reinhardt, F., and Reents, R. (2014) J. Dairy Sci. 97:5833–5850. https://doi.org/10.3168/jds.2014-7924.

Mäntysaari, E.A., Evans, R.D., and Strandén, I. (2017) J. Anim. Sci. 95:4728–4737. https://doi.org/10.2527/jas2017.1912.

Mäntysaari, E.A., Koivula, M., and Strandén, I. (2020) J. Dairy Sci. 103:5314–5326. https://doi.org/10.3168/jds.2019-17754.

Misztal, I., Legarra, A., and Aguilar., I. (2014) J. Dairy Sci. 97:3943–3952. https://doi.org/10.3168/jds.2013-7752.

Strandén, I., and Lidauer, M. (1999) J. Dairy Sci. 82:2779–2787. https://doi.org/10.3168/jds.s0022-0302(99)75535-9.

ten Napel, J., Vandenplas, J., Lidauer, M., Stranden, I., *et al.* (2021) Manual – MiXBLUP 3.0.1 manual, V3.0 – 2021-11. Available at: https://www.mixblup.eu/documents/8412108060_ASG_WLR_MixBlup%20Manual_LR.pdf

Tier, B., and Meyer, K. (2004) J. Anim. Breed. Genet. 121:77–89. https://doi.org/10.1111/j.1439-0388.2003.00444.x.

Vandenplas, J., Eding, H., Bosmans, M., and Calus, M.P.L. (2020) Genet. Sel. Evol. 52:24. https://doi.org/10.1186/s12711-020-00543-9.

359. From raw sensor and automated data to genetic evaluation and validation in the cloud

D. Schokker[1*], I.N. Athanasiadis[2], M. Poppe[1], J. ten Napel[1], C. Kamphuis[1] and R.F. Veerkamp[1]

[1]Wageningen University & Research, P.O. Box 338, 6700 AH Wageningen, the Netherlands; [2]Data Competence Centre, Wageningen University & Research, 6708PB Wageningen, the Netherlands; dirkjan.schokker@wur.nl

Abstract

Precision livestock farming tools, such as sensor and automation techniques, enable animal breeders to define novel traits. New challenges are data storage and handling and the development of genetic evaluation. Traditionally, the development of genetic evaluation for novel traits requires several steps, i.e. data curation, trait definition, variance component estimation (ASReml), genetic evaluation (MiXBLUP), and validation of the estimated breeding values (EBVs), requiring many iterations to optimize the genetic evaluations. We combined these steps in a cloud solution to make this entire process more efficient. An experiment was run with 1,782,373,113 daily milk yield records of 1,120,550 cows, all the way from data curation to trait definition and validation of the EBV. The resulting wall-time, i.e. elapsed actual time from start to finish of the entire process, was ~23 hours. The flexible cloud solution can be easily modified or adapted to develop novel traits.

Introduction

The increasing use of sensor and automation technology in livestock farming, for example, to monitor animal health and welfare in cows (Smith *et al.* 2006; Matthews *et al.* 2016; Ouweltjes *et al.* 2021), enable the development of new phenotypes and (complex) traits. The data these sensors generate can be both unstructured non-relational data, including camera or video images, and (non-standardized) structured data, like relational databases (Structured Query Language-SQL). Moreover, sensors are recording in real-time and generate large volumes of data, particularly when recording is done over a long period of time. Big data lakes have emerged as cloud infrastructures that effectively manage these data types, where it is needed to store, manage, access, and process large volumes of a wide range of structured, semi-structured, or unstructured data sources in raw format. In addition, different tools and resources are available in the cloud environment that can interact with such a data lake storage. Such tools and resources include platforms to transform, filter, and structure the data sources, making the cloud environment in combination with a data lake storage both flexible and scalable. Data lake storage can utilise the potential of large volumes of sensor data (Schokker *et al.* 2020). In animal breeding programmes, the challenge lies in utilizing these (novel sensor) data sources to identify the best parents for the next generation. Going from raw (sensor) data to genetic evaluation involves many steps. These pre-processing steps include reading in the data, followed by filtering, transforming, joining and aggregating the data, and writing the output (Gengler 2019). The genetic evaluation includes steps like generating subsets of the data to estimate genetic parameters, and predict breeding values (EBV) and their reliabilities. These downstream steps are performed with different software packages, often in different computing infrastructures. Also, after the final EBV validation, the whole process often has to start again with model adjustments or different data curation. In this study, we combined the entire process in a single cloud environment, linking all the different steps from pre-processing up to the genetic evaluation. As a feasibility experiment, we applied the pipeline to 1,782,373,113 daily milk yield records of 1,120,550 cows to define and evaluate a resilience trait.

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_359, © D. Schokker *et al.* 2022

Materials & methods

Design of cloud environment. We designed and built a custom-made cloud solution that suited our needs to perform all required steps to get from raw (sensor) data to genetic evaluation in a single cloud environment (Figure 1). The focus of this paper is on the infrastructure of this cloud solution.

Within the cloud-environment we employed a BLOB (binary large object) storage because this has tiered flexible storage and scale-up on performance-driven computing. This BLOB storage was linked to other data analytics resources, i.e. Databricks and our customised Docker containers for variance component estimation, genetic evaluation, and validation of the estimated breeding values. Databricks is an Apache Spark-based analytics service designed for data science and data engineering (Zaharia *et al.* 2016), and therefore, well-equipped to handle the nearly two billion milk yield records. Within the Databricks platform, we generated a pre-processing script for the raw (milk yield) data. Such an Extract, Transform, and Load (ETL) procedure is crucial for an easy and flexible scale-up of the process. This whole ETL procedure was written in Python (v3.7.3) and uses Apache Spark (v3.1.1) in the background for efficiency. For the subsequent genetic evaluation, we have used container technology, i.e. Docker.

Experiment. We intended to build the cloud solution for an existing genetic evaluation, so we used the dairy cattle dataset described by Poppe *et al.* (Poppe *et al.* 2020) as a case study. This dataset contained 1,782,373,113 milk yield records from 1,120,550 cows. Additionally, we used the pedigree and date of birth for those animals, which were necessary to curate a dataset for breeding value estimation. We calculated a potential resilience indicator trait from the milk yield records: natural log of the variance of deviations from the lactation curve, as defined by Poppe *et al.* (Poppe *et al.* 2020). This resilience indicator described the fluctuations in a frequently measured trait, here milk yield. This highly frequent sensor and automation data correspond to the developments in data collection. The ETL procedure was set-up in such a way to easily modify the criteria for filtering the data, including: (1) the milking system (automated or conventional); (2) parity (1, 2, >3); (3) breed; and (4) breed percentage.

Pre-processing. The whole pre-processing, i.e. ETL procedure, was performed within Databricks (https://azure.microsoft.com/en-us/services/databricks/), which we deployed on a computer cluster. We set the cluster with the following specification: the driver and workers all had 56 GB Memory and 16 cores, where

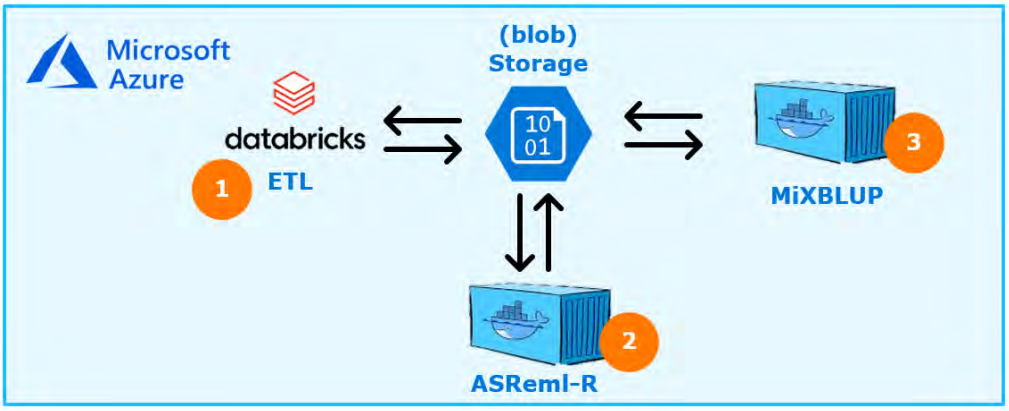


Figure 1. Schematic overview of the infrastructure and resources used in our cloud environment, i.e. Microsoft Azure. Four different modules could interact with the central BLOB (binary large object) storage: (1) pre-processing via Databricks; (2) Docker container with ASReml-R software; and (3) Docker container with MiXBLUP software.

the number of workers may vary from a minimum of 2 to a maximum of 8. Applying the data editing steps using the ETL procedure resulted in 352,871 cows with a resilience trait for at least one of the three lactations in the experiment.

Estimation of genetic parameters. For the estimation of the variance components, first a dedicated Docker (https://www.docker.com/) container was setup within our environment that could communicate via the central (blob) storage. This first Docker container, with Linux operating system, had installed ASReml-R (v4.1.0.110) (Butler *et al.* 2017), and for calculations we used a virtual machine with 2 GPU and 4 CPU cores with 112 GB Memory. Variance estimation is known to be computational intensive, therefore the pre-processed dataset, i.e. 352,871 cows, was split into five random subsets. We obtained five estimates for the heritability and genetic correlations among the three lactations.

Genetic evaluation. In the second Docker container, also with Linux operating system, we installed MiXBLUP v2.2 (ten Napel *et al.* 2017). For the calculations, we used a virtual machine with 4 CPU cores and 16 GB Memory. The model used in MiXBLUP was the same as the model used in ASReml-R. Now the full dataset, i.e. 1,120,500 cows, could be used for the final EBV calculation. From the full pedigree, we build a tailored pedigree starting with the cows present in the pre-processed data and adding five generations of male and female ancestors. We analysed the data using a statistical model with herd-year-season and lactation length as fixed effects, and cow as a random effect (Poppe *et al.* 2020).

Results

For the pre-processing step (ETL procedure) the parameters were set to include automated milking system, parities 1, 2, >3, and the breed Holstein-Friesian with a breed percentage of >75%. The wall-time, the elapsed time from the beginning to the end of the entire process, was approximately 23 hours. Table 1 shows the wall-time for each specific step.

Discussion

We demonstrated that by using a cloud solution we can go from (raw) sensor data to a trait of interest and perform the accompanying genetic evaluation and validation in approximately 23 hours, when processing a very large dataset of more than 1.7 billion milk records from more than 1,000,000 cows. Our results are promising, especially because the expanded use of sensors in the livestock sector will generate an ever-increasing volume of data, either unstructured or structured. This procedure will allow quick alterations in the pre-processing of the data to be tested and validated, therefore, optimizing the time to the final implementation of genetic evaluations using sensor data. The gain is in the generic tools for the data pre-processing and then the 'easy push of a button' for variance component estimation and breeding value estimation, something that often requires a lot of manual work. Our cloud solution enables scalability and flexibility in the cluster computing environment. Examples for scalability include upscaling and downscaling of IT requirements when needed. The flexibility allows easily sharing data, scripts, and results with your collaborators. Implementing the genetic evaluation and validation software in our cloud infrastructure, while avoiding data transferring from one system to another was challenging. Moreover, due

Table 1. Wall-time for each step.

Step	Wall-time (h)
Pre-processing	6
Estimation of genetic parameters	16.5
Genetic evaluation	0.5

to the nature of the cloud infrastructure, it becomes possible to run various genetic evaluations in parallel. For example, estimating breeding values with different models and calculating approximate reliabilities and yield deviations. Another possibility is to run the pre-processing procedure on different clusters, simultaneously, where small but relevant changes in the script can change the outcome of the feature of interest, e.g. specifying the number of parities to include, or what genotype to focus on, or specifying a period in time. These proposed workflows could decrease the wall-time, thus making this process more efficient.

Funding

This study was financially supported by the Dutch Ministry of Economic Affairs (TKI Agri and Food project 16022) and the Breed4Food partners Cobb Europe, CRV, Hendrix Genetics and Topigs Norsvin.

References

Butler D.G., Cullis B.R., Gilmour A.R., Gogel B.G. and Thompson R. (2017) ASReml-R Reference Manual Version 4, pp. VSN International Ltd, Hemel Hempstead, HP1 1ES, UK.

Gengler N. (2019) J Dairy Sci 102(6)**:** 5756-5763. https://doi.org/10.3168/jds.2018-15711

Matthews S.G., Miller A.L., Clapp J., Plotz T. and Kyriazakis I. (2016) Veterinary journal 217(43-51. https://doi.org/10.1016/j.tvjl.2016.09.005

Ouweltjes W., Spoelstra M., Ducro B., De Haas Y. and Kamphuis C. (2021) J Dairy Sci 104(11)**:** 11759-11769. https://doi.org/10.3168/jds.2021-20413

Poppe M., Veerkamp R.F., Van Pelt M.L. and Mulder H.A. (2020) J Dairy Sci 103(2)**:** 1667-1684. https://doi.org/10.3168/jds.2019-17290

Schokker D., Athanasiadis I.N., Visser B., Veerkamp R.F. and Kamphuis C. (2020) Animal 14(11)**:** 2397-2403. https://doi.org/10.1017/S175173112000155X

Smith K., Martinez A., Craddolph R., Erickson H., Andresen D. *et al.* (2006) An integrated cattle health monitoring system, pp. 4659-4662 in *Conference proceedings: Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Conference*, New York, NY, USA.

Ten Napel J., Vandenplas J., Lidauer M., Stranden I., Taskinen M. *et al.* (2017) MiXBLUP: A user-friendly softwarevfor large genetic evaluation systems, pp.

Zaharia M., Xin R.S., Wendell P., Das T., Armbrust M. *et al.* (2016) Communications of the ACM 59(11)**:** 56-65.

360. Indirect genomic prediction reduces computational costs in large-scale single-step evaluations

I. Strandén[1*], J. ten Napel[2], R.F. Veerkamp[2], R. Evans[3], S. Naderi[3], E.A. Mäntysaari[1] and J. Vandenplas[2]

[1]Natural Resources Institute Finland (Luke), Jokioinen, Finland; [2]Wageningen University & Research, P.O. Box 338, 6700 AH Wageningen, the Netherlands; [3]Irish Cattle Breeding Federation, Highfield House, Newcestown Road, Bandon, Cork, Ireland; ismo.stranden@luke.fi

Abstract

Computing time and memory requirements increase with single-step methods to estimate genomic breeding values when the number of genotyped animals increases. Computational costs can be reduced by omitting genotypes of animals without phenotype and progeny, often the candidate animals for selection. Indirect prediction of a candidate animal GEBV can be based on the animal's genotype and SNP marker solutions (DGV). Alternatively, the sum of DGV and residual polygenic (RPG) effect can be computed, denoted GRV (Genomic and Residual polygenic Value). We applied indirect genomic prediction for a 6 trait calving difficulty evaluation. There were 1.50 million genotyped animals of which 36% were considered candidate animals. Based on our results, DGV showed high accuracy but also bias due to omitting the RPG effect. The GRV prediction had high accuracy and low bias. Computing time was reduced by 33%.

Introduction

Practical implementation of single-step genomic BLUP (ssGBLUP; Aguilar *et al.* 2010; Christensen and Lund 2010) has met many challenges (e.g. Mäntysaari *et al.* 2020; Misztal *et al.* 2020). Continual increases in volumes of genotyped animals also increase the computational costs. A practical aspect seldom considered is the efficient calculation of genomic estimated breeding values (GEBV) for genotyped animals without own and progeny information, so called candidate animals. Excluding these genotypes in the evaluation and estimating their GEBVs afterwards may reduce computational costs (Tsuruta *et al.* 2021). Including these genotypes in the routine evaluation may even be undesirable (Koivula *et al.* 2018). Liu *et al.* (2016) presented formulas for predicting GEBVs of genotyped candidate animals when the model has a residual polygenic (RPG) effect. The aim of this study was to compare the performance of approaches for indirect estimation of GEBV for genotyped candidate animals.

Materials & methods

The ssGTABLUP approach. Mäntysaari *et al.* (2017) presented an efficient computational approach for ssGBLUP named ssGTABLUP where the genomic relationship matrix has the form $\mathbf{G}_C = \mathbf{G}_0 + \mathbf{C}$. Here, $\mathbf{C} = w\mathbf{A}_{gg}$, w is the RPG proportion and $\mathbf{A}_{gg}$ is the pedigree-based relationship matrix among the genotyped animals. The genomic part is $\mathbf{G}_0 = \mathbf{ZBZ}'$ where Z is an $n \times m$ matrix of centered marker genotypes, B is an $m \times m$ diagonal scaling matrix, n is the number of genotyped animals and m is the number of SNP markers. In the VanRaden (2008) method 1, the centering uses base population allele frequencies $p_i, i = 1, \ldots, m$ and the scaling matrix is $\mathbf{B} = \mathbf{I}\frac{1-w}{k}$ with the scaling constant $k = 2\sum_{i=1}^{m} p_i(1-p_i)$.

The $\mathbf{G}_C^{-1}$ matrix in the model equations (MME) can be calculated as $\mathbf{G}_C^{-1} = \frac{1}{w}\mathbf{A}_{gg}^{-1} - \frac{1}{w^2}\mathbf{A}_{gg}^{-1}\mathbf{ZK}^{-1}\mathbf{Z}'\mathbf{A}_{gg}^{-1}$ where $\mathbf{K} = \frac{1}{w}\mathbf{Z}'\mathbf{A}_{gg}^{-1}\mathbf{Z} + \mathbf{B}^{-1}$ is a symmetric positive definite matrix. In practice, iterative methods can be used to solve the MME where the $\mathbf{G}_C^{-1}$ matrix is not explicitly computed and the needed product $\mathbf{G}_C^{-1}\mathbf{v}$ is computed efficiently (Mäntysaari *et al.* 2020).

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_360, © I. Strandén *et al.* 2022

Let vector $\mathbf{u}_g$ ($\mathbf{u}_n$) have the single-step estimates for the genotyped (non-genotyped) animals. Assuming an RPG effect, the GEBVs for non-genotyped and genotyped animals can be decomposed as (Fernando *et al.* 2014; Fernando *et al.* 2016; Liu *et al.* 2016):

$$\mathbf{u} = \begin{bmatrix}\mathbf{u}_n \\ \mathbf{u}_g\end{bmatrix} = \begin{bmatrix}\mathbf{A}_{ng}\mathbf{A}_{gg}^{-1}\mathbf{Zg} + \boldsymbol{\epsilon} + \mathbf{d}_n \\ \mathbf{Zg} + \mathbf{d}_g\end{bmatrix} = \begin{bmatrix}\mathbf{A}_{ng}\mathbf{A}_{gg}^{-1}\mathbf{Zg} \\ \mathbf{Zg}\end{bmatrix} + \begin{bmatrix}\boldsymbol{\epsilon} \\ \mathbf{0}\end{bmatrix} + \begin{bmatrix}\mathbf{d}_n \\ \mathbf{d}_g\end{bmatrix} \quad (1)$$

where $\mathbf{A}_{ng}$ is the pedigree-based relationship matrix between the non-genotyped and genotyped animals, ε is the imputation residual, vector: $\mathbf{d} = \begin{bmatrix}\mathbf{d}_n \\ \mathbf{d}_g\end{bmatrix}$ corresponds to RPG, the part of the genetic effects not explained by the genomic data (Christensen and Lund 2012). Thus, genotyped animal GEBV has two components: (1) Zg, that is the direct genetic value (DGV) due to the marker effects; (2) d_g, that is the breeding value due to the RPG effect.

Indirect GEBVs of genotyped animals without own and progeny information. Computational costs in any ssGBLUP can be reduced using a two-step algorithm: (1) calculate GEBV without the candidate animals; (2) predict GEBV of the candidate animals. According to Equation 1, we need to compute $\hat{\mathbf{u}}_c = \mathbf{Z}_c\hat{\mathbf{g}} + \hat{\mathbf{d}}_c$ where subscript *c* refers to the candidate animals. The marker solutions can be easily calculated in ssGTABLUP after solving the MME using the formula $\hat{\mathbf{g}} = \frac{1}{w}\mathbf{K}^{-1}\mathbf{Z}'\mathbf{A}_{gg}^{-1}\hat{\mathbf{u}}_g$ (Liu *et al.* 2014).

Calculation of DGV for a candidate animal is straightforward but the RPG effects $\hat{\mathbf{d}}_c$ are unknown. According to Liu *et al.* (2016), $\hat{\mathbf{d}}_c = \mathbf{A}_{cg}\mathbf{A}_{gg}^{-1}\hat{\mathbf{d}}_g$ where $\hat{\mathbf{d}}_g = \hat{\mathbf{u}}_g - \mathbf{Z}\hat{\mathbf{g}}$ and $\mathbf{A}_{cg}$ is the pedigree-based relationship matrix between the genotyped candidate animals and the genotyped animals already included in the single-step evaluation. Calculation of $\hat{\mathbf{d}}_c$ can be done in two steps. First, $\mathbf{x} = \mathbf{A}_{gg}^{-1}\hat{\mathbf{d}}_g$ which can be done efficiently using sparse matrices (Strandén *et al.* 2017). Second, $\hat{\mathbf{d}}_c = \mathbf{A}_{cg}\mathbf{x}$ which can be computed efficiently (Colleau 2002) using the full pedigree-based relationship matrix times vector product:

$$\begin{bmatrix}\mathbf{b}_n \\ \mathbf{b}_g \\ \hat{\mathbf{d}}_c\end{bmatrix} = \begin{bmatrix}\mathbf{A}_{nn} & \mathbf{A}_{ng} & \mathbf{A}_{nc} \\ \mathbf{A}_{gn} & \mathbf{A}_{gg} & \mathbf{A}_{gc} \\ \mathbf{A}_{cn} & \mathbf{A}_{cg} & \mathbf{A}_{cc}\end{bmatrix}\begin{bmatrix}\mathbf{0} \\ \mathbf{x} \\ \mathbf{0}\end{bmatrix}.$$

Data and models. The data and associated variance components were extracted from the 6 trait calving difficulty evaluation for Irish beef cattle performed by ICBF. There were 9.54 million animals in the pedigree file of which 5.76 million had a data record. There were 50,240 SNP markers available from 1.50 million genotyped animals. The animals had been genotyped using the Illumina Bovine SNP50 Bead Chip (Illumina, San Diego, CA). The MME had 123.32 million unknowns of which 114.50 million were additive and maternal genetic effects.

Study design. GEBVs were estimated with two data sets. The full data had all genomic information. In the reduced data, genomic information from the candidate animals were removed. There were 533,116 candidate animals, i.e. the reduced data had 965,868 genotyped animals. For both evaluations, the genomic relationship matrix was based on VanRaden method 1 with a 20% RPG proportion. The Z matrix was centered using base population allele frequencies estimated by the Bpop program (Strandén and Mäntysaari 2020).

Candidate animal GEBV was estimated by three approaches: (1) PA: mean of parent GEBVs; (2) DGV: $\mathbf{Z}_c\hat{\mathbf{g}}$ using estimated SNP marker effects $\hat{\mathbf{g}}$ from the reduced data analysis; and (3) GRV: formula $\mathbf{Z}_c\hat{\mathbf{g}} + \mathbf{A}_{cg}\mathbf{A}_{gg}^{-1}\hat{\mathbf{d}}_g$. Comparison of these candidate animal GEBV with the full data GEBV for the direct and maternal effects was based on a Pearson correlation, linear regression, and level bias for each trait separately. Regression was from the full data GEBV on the reduced data GEBV. Level bias for trait *j* was defined as average of $(\hat{\mathbf{u}}_c - \hat{\mathbf{u}}_{c,full})/\boldsymbol{\sigma}_j$ where $\hat{\mathbf{u}}_c$ and $\hat{\mathbf{u}}_{c,full}$ are GEBV from the reduced and full data analysis, respectively, and σ_j is genetic standard deviation of the trait.

Solver. The ssGTABLUP MME were solved using preconditioned conjugate gradient (PCG) iteration by MiX99 software with 10 CPU threads (Strandén *et al.* 2018). The PCG method was assumed to be converged when $\sqrt{(\mathbf{Cs}^{[k]}-\mathbf{r})'(\mathbf{Cs}^{[k]}-\mathbf{r})/\mathbf{r}'\mathbf{r}} < 10^{-7}$ where C is the coefficient matrix of MME, $\mathbf{s}^{[k]}$ is vector of solutions at round *k*, and r is the MME right-hand side vector.

Results

Preprocessing and solver computing times were reduced effectively in the same proportion as the number of genotyped animals used in the full vs reduced data (Table 1). For the non-candidate animals and both the direct and maternal effects, correlations in GEBV between the full and the reduced evaluations were between 98.9 and 99.8% for the genotyped animals and between 99.0 and 99.6% for the non-genotyped animals across traits. The indirect GEBVs approximated by GRV gave the highest correlations, the lowest level biases, and no over- or under-dispersion (i.e. regression coefficient close to: (1), across all traits and genetic effects (Table 2 and 3). In comparison, DGV had correlations close to 1 (on average 0.97), but the regression coefficients indicated under-dispersion and level biases were higher.

Discussion

Computing time was reduced effectively by omitting genotypes of candidate animals because their number was large. In the indirect GEBV prediction, DGV showed high accuracy but consistent bias due to omitting the RPG contribution. In contrast, GRV gave high correlations and low bias. When the RPG weight is smaller than in our study (e.g. 5%), DGV and GRV are likely to give similar values. The GRV approach can also be used in intermediate GEBV prediction of newly genotyped animals when the full data single-step evaluations are performed at longer intervals such as once every 4 months.

Table 1. Statistics of solving ssGTABLUP using the full and reduced data.

Data	Pre time[1]	RAM[2]	N[3]	Solver time	Indirect time	Total time
Full	9.6h	592	665	12.2h	-	21.8h
Reduced (GRV)	5.5h	390	664	7.5h	0.7h	13.7h

[1] Preprocessing computing time.
[2] Peak RAM in GB.
[3] Number of PCG iterations.

Table 2. Correlations, regression coefficients, and level bias for direct GEBV of genotyped candidate animals computed from the full data vs reduced data.

		Trait 1	2	3	4	5	6
Correlation	PA	0.737	0.775	0.687	0.656	0.671	0.713
	DGV	0.978	0.981	0.971	0.966	0.969	0.982
	GRV	0.995	0.995	0.996	0.996	0.997	0.997
Regression coefficient	PA	0.968	0.964	0.942	0.938	0.972	0.970
	DGV	1.036	1.045	1.042	1.033	1.038	1.043
	GRV	0.994	0.993	1.002	1.003	1.004	1.000
Level bias	PA	0.021	0.022	0.027	0.026	0.022	0.014
	DGV	0.006	-0.004	0.026	0.038	0.038	0.001
	GRV	0.004	0.003	-0.002	-0.004	0.001	-0.005

Table 3. Correlations, regression coefficients, and level bias for maternal GEBV of genotyped candidate animals from the full data vs reduced data.

		Trait 1	2	3	4	5	6
Correlation	PA	0.675	0.662	0.659	0.647	0.725	0.720
	DGV	0.976	0.976	0.979	0.978	0.983	0.983
	GRV	0.988	0.989	0.996	0.995	0.997	0.994
Regression coefficient	PA	0.977	0.961	0.959	0.952	0.973	0.985
	DGV	1.043	1.033	1.037	1.028	1.036	1.050
	GRV	1.004	0.997	1.003	1.002	0.997	0.999
Level bias	PA	0.011	0.005	0.016	0.015	0.005	0.008
	DGV	0.026	0.028	0.017	0.021	0.052	0.035
	GRV	0.006	-0.001	0.004	0.002	0.000	0.005

References

Aguilar I., Misztal I., Johnson D.-L., Legarra A., Tsuruta S, and Lawlor T.J. (2010) J. Dairy Sci. 93:743-752. https://doi.org/10.3168/jds.2009-2730

Christensen O.F., and Lund M.S. (2010) Genet. Sel. Evol. 42:2. https://doi.org/10.1186/1297-9686-42-2

Colleau J.J. (2002) Genet. Sel. Evol. 34:409-421. https://doi.org/10.1186/1297-9686-34-4-409

Koivula M., Strandén I., Aamand G.P., and Mäntysaari E.A. (2018) J. Animal Breed. Genet. 135:107–115. https://doi.org/10.1111/jbg.12318

Liu Z., Goddard M.E., Reinhardt F., and Reents R. (2014) J. Dairy Sci. 97:5833-5850. https://doi.org/10.3168/jds.2014-7924.

Liu Z., Goddard M.E., Hayes B.J., Reinhardt F., and Reents R. (2016) J Dairy Sci 99:2016-2025. https://doi.org/10.3168/jds.2015-10394

Misztal I., Lourenco D., and Legarra A. (2020) J. Animal Sci. 98: 1-14. https://doi.org/10.1093/jas/skaa101

Misztal I., Lourenco D., Legarra A. (2020) J. Animal Sci. 98: 1-14. https://doi.org/10.1093/jas/skaa101

Mäntysaari E.A., Evans R.D., and Strandén I. (2017) J. Animal Sci. 95:4728-4737. https://doi.org/10.2527/jas2017.1912.

Mäntysaari E.A., Koivula M., and Strandén I. (2020) J. Dairy Sci. 103:5314-5326. https://doi.org/10.3168/jds.2019-17754

Strandén I., Matilainen K., Aamand G., and Mäntysaari E.A. (2017) J. Animal Breed. Genet., 134:264-274. https://doi.org/10.1111/jbg.12257

Strandén I., Taskinen M., Matilainen K., Lidauer M., and Mäntysaari E.A. (2018) Proc. of the 11th WCGALP, Auckland, New Zealand.

Strandén I., Mäntysaari E.A. (2020) Agr. Food Sci. 29:166-176. https://doi.org/10.23986/afsci.90955

Tsuruta S., Lourenco D.A.L., Masuda Y., Lawlor T.J., and Misztal I. (2021) J. Dairy Sci. Comm. 2:356-360. https://doi.org/10.3168/jdsc.2021-0097

VanRaden P.M. (2008) J. Dairy Sci. 91: 4414-4423. https://doi.org/10.3168/jds.2007-0980.

361. Accounting for trait-specific genomic and residual polygenic covariances in multivariate single-step genomic evaluation

K. Meyer

AGBU, a joint venture of NSW Department of Primary Industries and University of New England, Armidale, 2351, Australia; kmeyer@une.edu.au

Abstract

For multivariate, single-step genomic best linear unbiased prediction analyses fitting a breeding value model, it is often assumed that the proportions of total genetic variance accounted for by genomic markers and residual polygenic effects are the same for all traits. Different covariance matrices for the two types of genetic effects are readily taken into account by fitting them separately. However, this can lead to slow convergence rates in iterative solution schemes. We propose an alternative computing strategy which – exploiting a canonical transformation – allows for trait-specific covariances whilst directly fitting total genetic effects only. Its effects on convergence rates and gains in accuracy and bias of genomic evaluation compared to analyses assuming proportionality of covariance matrices are examined using a small simulation study. Results show comparatively little improvement in accuracies but worthwhile reductions in overdispersion of predicted genetic merits for genotyped individuals without phenotypes.

Introduction

With the advent of genomic evaluation, it was quickly recognised that the single nucleotide polymorphism (SNP) markers available do not necessarily explain all genetic variation. This can be accounted for by adding a 'residual polygenic' (RPG) effect to the model of analysis fitted (e.g. Jensen *et al.*, 2012). For univariate single-step analyses fitting a breeding value model (ssGBLUP) an equivalent alternative is to replace the genomic relationship matrix, G, with λG + (1 – λ)A_{22}, its weighted average with A_{22}, the submatrix of the numerator relationship, A, for genotyped animals and $\lambda \in [0,1]$ the proportion of the genetic variance attributed to SNP effects. This is readily extended to multivariate analyses if it can be assumed that Σ_G and Σ_A, the covariance matrices due SNP and RPG effects, are proportional.

This paper describes a computing strategy for ssGBLUP exploiting a transformation to canonical scale to avoid fitting a separate RPG effect when this does not hold, i.e. when $\Sigma_G \neq \Sigma_A$ (for c a constant). In addition, a small simulation study is presented to evaluate the potential advantages of accounting for trait-specific weighting of the two sources of genetic variation.

Computing strategy

Consider a multivariate linear mixed model for t traits:

$$y = Xb + Zu + e \tag{1}$$

with y, b, u and e denoting the vectors of observations, fixed effects, additive genetic effects and residuals and design matrices X and Z. Let genetic effects be ordered by traits within individuals so that u can be partitioned into parts for n_1 non-genotyped (u_1) and n_2 genotyped (u_2) individuals. This gives the mixed model equations (MME)

$$\begin{bmatrix} X'R^{-1}X & X'R^{-1}Z \\ Z'R^{-1}X & Z'R^{-1}Z + A^{-1} \otimes \Sigma_U^{-1} + \Delta \end{bmatrix} \begin{bmatrix} \hat{b} \\ \hat{u} \end{bmatrix} = \begin{bmatrix} X'R^{-1}y \\ Z'R^{-1}y \end{bmatrix} \tag{2}$$

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_361, © K. Meyer 2022

with $\Sigma_U = \Sigma_G + \Sigma_A$ the matrix of (total) genetic covariances between traits and the non-zero block of Δ containing the 'add-on' for genotyped individuals, arising from the inverse of the joint relationship matrix for genotyped and non-genotyped individuals, H (Aguilar *et al.*, 2010).

$$\Delta = \begin{bmatrix} 0 & 0 \\ 0 & \Delta_G - \mathrm{A}_{22}^{-1} \otimes \Sigma_U^{-1} \end{bmatrix} \tag{3}$$

For $\Sigma_G = c\Sigma_A$,

$$\Delta_G = (\lambda \mathrm{G} + (1-\lambda)\mathrm{A}_{22})^{-1} \otimes \Sigma_U^{-1} \tag{4}$$

i.e. a single matrix inverse of size $n_2 \times n_2$ suffices to obtain Δ_G. More generally, however, when $\Sigma_G \neq c\Sigma_A$, Δ_G is not separable and

$$\Delta_G = (\mathrm{G} \otimes \Sigma_G + \mathrm{A}_{22} \otimes \Sigma_A)^{-1} \tag{5}$$

so that the inverse of a matrix of size $tn_2 \times tn_2$ is needed. This can impose a considerable computational burden and become prohibitive for larger numbers of traits or genotyped animals.

Canonical transformation. For two conforming, symmetric matrices, Σ_1 positive definite and Σ_1 positive semi-definite, there exist a matrix Q so that $\mathrm{Q}\Sigma_1\mathrm{Q}' = \mathrm{I}$ and $\mathrm{Q}\Sigma_2\mathrm{Q}' = \mathrm{D}$ with D a diagonal matrix with elements, d_k, equal to the eigenvalues of $\boldsymbol{\Sigma}_1^{-1}\boldsymbol{\Sigma}_2$ (Anderson, 1958) and I an identity matrix. Q defines the so-called transformation to canonical scale which has seen considerable use to simplify multivariate linear mixed model analyses (e.g. Ducrocq and Besbes, 1993; Meyer, 1985). For $\mathrm{Q}\Sigma_G\mathrm{Q}' = 1$ and $\mathrm{Q}\Sigma_A\mathrm{Q}' = \mathrm{D}$, we can factor Δ_G into three terms:

$$\boldsymbol{\Delta}_G = [\mathbf{I} \otimes \mathbf{Q}'](\mathbf{G} \otimes \mathbf{I} + \mathbf{A}_{22} \otimes \mathbf{D})^{-1}[\mathbf{I} \otimes \mathbf{Q}] \tag{6}$$

with the middle term in Equation 6 the inverse of a matrix comprised of n_2^2 blocks of size $t \times t$ which are diagonal. Reordering the levels of u_2 and Δ_G according to animals within traits, denoted by superscript '*', results in:

$$\boldsymbol{\Delta}_G^{\star} = [\mathbf{Q}' \otimes \mathbf{I}](\mathbf{I} \otimes \mathbf{G} + \mathbf{D} \otimes \mathbf{A}_{22})^{-1}[\mathbf{Q} \otimes \mathbf{I}] \tag{7}$$

In this formulation, the middle term is block-diagonal with t blocks $(\mathbf{G} + d_k\mathbf{A}_{22})^{-1}$. Hence the $t(t+1)/2$ submatrices of $\boldsymbol{\Delta}_G^{\star}$ ($\boldsymbol{\Delta}_{G,ij}^{\star}$ for traits i and j) can be obtained as linear combinations of $r \leq t$ inverses of the individual blocks of size $n_2 \times n_2$, with r the number of distinct values d_k. For q_{ij} denoting the ij-$^{\text{th}}$ element of Q,

$$\boldsymbol{\Delta}_{G,ij}^{\star} = \sum_{k=1}^{t} q_{ki}\, q_{kj} (\mathbf{G} + d_k\mathbf{A}_{22})^{-1} \tag{8}$$

Following Mäntysaari *et al.* (2017), we can use the Woodbury matrix identity to compute the inverses required. For G = MM'/s with M the matrix of centred marker counts and s a scalar,

$$\begin{aligned} (\mathbf{G} + d_k\mathbf{A}_{22})^{-1} &= [\mathbf{A}_{22}^{-1} - \mathbf{A}_{22}^{-1}\,\mathbf{M}\,(s\,d_k\mathbf{I} + \mathbf{M}'\,\mathbf{A}_{22}^{-1}\,\mathbf{M})^{-1}\mathbf{M}'\,\mathbf{A}_{22}^{-1}]/d_k \\ &= [\mathbf{A}_{22}^{-1} - \mathbf{A}_{22}^{-1}\,\mathbf{M}\,\mathbf{E}\,(s\,d_k\mathbf{I} + \mathbf{T})^{-1}\,\mathbf{E}'\,\mathbf{M}'\,\mathbf{A}_{22}^{-1}]\,/d_k \end{aligned} \tag{9}$$

with T and E the matrices of eigenvalues and eigenvectors of $\mathbf{M}'\mathbf{A}_{22}^{-1}\mathbf{M}$, respectively. Matrices $(s\,d_k\mathbf{I} + \mathbf{T})$ are diagonal and thus trivial to invert. This facilitates efficient evaluation of the r diagonal blocks required through a single eigen-decomposition of a matrix of size $m \times m$, with m the number of markers.

Calculations can be implemented using either ordering of u_2. We assume animals within traits in the following. This allows for the direct use of highly optimised library routines for dense matrix calculations to evaluate the inverses required. For analyses forming the MME in core, multiplication with Q⊗I can be integrated into the design matrix Z, in the same way as for a transformation to principal components (Meyer *et al.*, 2015). Alternatively, using an iterative scheme to solve the MME, a product $\mathbf{\Delta}_G^\star \mathbf{r}_2^\star$ is required for each iterate (with $\mathbf{r}_2^\star$ the part of the vector of directions corresponding to $\mathbf{u}_2^\star$). This product is readily obtained by first forming $\mathbf{s}_2^\star = (\mathbf{Q} \otimes \mathbf{I})\mathbf{r}_2^\star$, then multiplying with the inverse diagonal blocks, yielding $\mathbf{t}_2^\star = (\mathbf{I} \otimes \mathbf{G} + \mathbf{D} \otimes \mathbf{A}_{22})^{-1}\mathbf{s}_2^\star$ and finally calculating $(\mathbf{Q}' \otimes \mathbf{I})\mathbf{t}_2^\star$.

Materials & methods

Data for three traits recorded on 21,000 animals in eight generations with heritabilities of 0.30, 0.25 and 0.10, respectively, and genetic correlations of 0.70 were generated using AlphaSim, version 1.05 (Faux *et al.*, 2016). Genotypes were constructed sampling 125 quantitative trait loci and 32,000 SNP markers. Subsequently, RPG effects were sampled for the given pedigree structure and added to the respective records and breeding values. Variances for RPG effects were chosen to yield variance ratios of magnitude 0.90, 0.85 and 0.20. Only marker information for 10, 30, 40 and 50% of randomly chosen individuals in generations five to eight was retained.

Genomic relationship matrices were build using Method 1 of Van Raden (2008), eliminating SNP with minor allele frequencies less than 2% and centering allele counts using 'observed' frequencies. To align G with A_{22} these were augmented by αJ with $\alpha = \mathbf{1}'(\mathbf{A}_{22} - \mathbf{G})\mathbf{1}/n_2^2$ for 1 and J a vector and a matrix, respectively, with all elements equal to unity (Vitezica *et al.* 2011). Estimates of variance components were obtained by restricted maximum likelihood (REML), fitting overall means as the only fixed effect. Trait-specific covariance matrices Σ_G and Σ_A were estimated by splitting u into parts explained by markers, g, and RPG effects, a and fitting these separately, assuming they were distributed as H⊗Σ_G and A⊗Σ_A, respectively, with Cov(g,a')=0. This was contrasted with analyses assuming $\Sigma_G = c\Sigma_A$ which fitted u with $\mathbf{\Delta}_G = (\lambda\mathbf{G} + (1-\lambda)\mathbf{A}_{22})^{-1} \otimes \mathbf{\Sigma}_U^{-1}$ and estimated λ by quadratic approximation of its profile likelihood. Analyses were carried out both in- and excluding records for the last generation.

To evaluate the impact on convergence behaviour, MME were also solved using iteration on data. This was done fitting either both g and a (FULL) or u (JOINT), and considering Σ_G and Σ_A to be either proportional (SINGLE λ) or trait specific (VARIED λ). Summary statistics calculated were means and standard deviations across 30 replicates for estimates of λ, i.e. the ratios of diagonal elements of Σ_G to Σ_U, and correlations (ρ) between and linear regressions (β) of true on predicted total breeding values (EBV), for genotyped individuals in the last generation.

Results & discussion

Means and standard deviations of summary statistics for VARIED λ and SINGLE λ together with corresponding values for analyses fitting single, fixed values of λ=0.50 (FIX λ=0.50) or λ=0.95 (FIX λ=0.95) are contrasted in Table 1 (analyses including all phenotypes). In spite of substantial differences in individual λ, improvements in ρ allowing for trait-specific values over those estimating a single value were modest. Similarly, deviations of β from the expected value of unity were only slightly reduced. Only for trait three and FIX λ=0.95 was a sizeable impact on ρ evident. Omitting phenotypes for generation eight (detailed results not shown) resulted in lower levels of ρ overall but yielded only slightly bigger improvements for VARIED λ over SINGLE λ than above. However this was accompanied by a drop of about 0.1 (from 0.984 to 0.874) in mean β for SINGLE λ for trait three, i.e. a marked increase in over-dispersion of EBVs.

Table 1. Means ($\bar{x}$) and standard deviations (sd) for estimates of variance ratios (λ), correlations (ρ) between and regressions (β) of true on predicted breeding values.

		VARIED λ			SINGLE λ			FIX λ=0.95			FIX λ=0.50		
Trait		1	2	3	1	2	3	1	2	3	1	2	3
$\hat{\lambda}$	$\bar{x}$	0.909	0.833	0.180	0.574	0.574	0.574	0.950	0.950	0.950	0.500	0.500	0.500
	sd	0.035	0.032	0.040	0.055	0.055	0.055	-	-	-	-	-	-
$\hat{\rho}$	$\bar{x}$	0.855	0.833	0.735	0.832	0.817	0.710	0.852	0.824	0.609	0.826	0.812	0.718
	sd	0.011	0.014	0.023	0.016	0.016	0.029	0.012	0.016	0.025	0.014	0.015	0.024
$\hat{\beta}$	$\bar{x}$	0.966	0.978	0.992	0.951	0.968	1.017	0.970	0.981	0.985	0.947	0.965	1.013
	sd	0.022	0.027	0.034	0.027	0.030	0.035	0.022	0.029	0.037	0.026	0.030	0.035

Additional univariate analyses (not shown) evaluating the profile for ρ and β over the whole range of λ showed that values were rather robust against considerable deviations of λ from their estimated, 'optimal' values. Similarly, literature results emphasised the insensitivity in ranking of ssGBLUP predictions of genetic merit against choices of λ (e.g. McMillan and Swan, 2017).

Whilst resulting in the same predicted breeding values, attempts to split u into its parts, g and a, (FULL) were clearly detrimental for convergence rates of iterative solution schemes. Means and ranges (in brackets) for the number of iterates required to solve the MME for VARIED λ were 1,036 (631-6,573) and 187 (141-263) for FULL and JOINT, respectively, i.e. JOINT reduced numbers of iterates required on average by about 80%. The highest numbers of iterates for the combination of FULL and VARIED λ were required for samples with individual λ close to unity (or zero), reflecting their effects on the condition number of the coefficient matrix in the MME. Corresponding numbers for SINGLE λ were 663 (563-818) and 152 (122-190).

Conclusions

Use of a canonical transformation facilitates trait-specific weighting of genetic covariances explained by genomic markers and RPG effects in ssGBLUP analyses, directly fitting total breeding values only. However, simulation results suggest that the resulting gains in accuracy of prediction compared to estimating a single, overall value of λ tend to be limited. Accommodating individual λ appeared to be most beneficial for a scenario with the values spanning almost the whole range of $[0,1]$ and in predicting breeding values for genotyped individuals without phenotypes. In practical applications, the trade-off between potential gains and additional computational demands need to be carefully assessed.

Acknowledgements

Work was supported by Meat and Livestock Australia grants L.GEN.1704 and L.GEN.2204.

References

Aguilar I., Misztal I., Johnson D.L., Legarra A., Tsuruta S., *et al.* (2010) J. Dairy Sci. 93 (2): 743–752. https://doi.org/10.3168/jds.2009-2730.

Anderson T. W. (1958) An Introduction to Multivariate Statistical Analysis. Wiley New York.

Ducrocq V., and Besbes B. (1993) J. Anim. Breed. Genet. 110: 81–92. https://doi.org/10.1111/j.1439-0388.1993.tb00719.x

Faux A-M, Gorjanc G., Gaynor R.C., Battagin M., Edwards S.M., *et al.* (2016) Plant Genome 9 (3): 1–14. https://doi.org/10.3835/plantgenome2016.02.0013.

Jensen J., Su G., and Madsen P. (2012) BMC Genetics 13: 44. https://doi.org/10.1186/1471-2156-13-44.

Mäntysaari E.A., Evans R.D., and Strandén I. (2017) J. Anim. Sci. 95 (11): 4728–4737. https://doi.org/10.2527/jas2017.1912.

McMillan A.J., and Swan A.A. (2017). Proc. Ass. Advan. Anim. Breed. Genet. 22: Paper no.130.
Meyer K. (1985) Biometrics 41 (1): 153–66. https://doi.org/10.2307/2530651.
Meyer K., Swan A.A., and Tier B. (2015) J. Anim. Sci. 93 (10): 4624–4628. https://doi.org/10.2527/jas.2015-9333.
Van Raden P. M. (2008) J. Dairy Sci. 91 (11): 4414–4423. https://doi.org/10.3168/jds.2007-0980.
Vitezica Z. G., Aguilar I., Misztal I., and Legarra. A. (2011). Genet. Res. 93 (5): 357–366. https://doi.org/10.1017/S001667231100022X.

362. Avoiding double counting in genomic reliability calculations

H. Ben Zaabza[1,3*], M. Taskinen[1], E.A. Mäntysaari[1], T. Pitkänen[1], G.P. Aamand[2] and I. Strandén[1]

[1]Natural Resources Institute Finland (Luke), 31600 Jokioinen, Finland; [2]NAV Nordic Cattle Genetic Evaluation, 8200 Aarhus N, Denmark; [3]Department of Animal and Veterinary Sciences, University of Vermont, Burlington, VT 05405, USA; hafedh.ben-zaabza@uvm.edu

Abstract

In this study, the performance of weights for multi-step reliability calculation was examined by 4 approaches: two Interbull methods, single-trait reverse reliability (ST-REV), and multiple-trait reverse reliability (MT-REV). We concentrated on weights for genotyped animals and used 305-d milk yield data in the first 3 lactations in Nordic (Danish, Finnish, and Swedish) Holstein dairy cattle. We compared reliabilities in a classical pedigree-based animal model to the back-calculated reliabilities. Correlations between the pedigree-based animal model reliabilities and the back-calculated reliabilities were high for all approaches (>0.90). However, gains from 0.02 to 0.11 on correlations between the classical and the back-calculated reliabilities were achieved using MT-REV compared with ST-REV and the Interbull methods. The MT-REV outperformed the ST-REV (higher correlations and smaller MSE) mainly for traits with many missing records. However, ST-REV and MT-REV yielded similar results for traits with few missing phenotype records.

Introduction

The total number of daughters of a bull has been used as a weighting factor in some routine genetic evaluations of dairy sires. However, the effective number of daughters does not account for factors such as the number of lactations of each daughter, daughter dam information, and contemporary group (CG). Interbull proposed a procedure to calculate weighting factors to use in international genetic evaluations (Interbull, July 2000 mimeo). The Interbull procedure consists of 2 steps: (1) calculating animal model reliability based on the animal's own performance; and (2) using the reliability of progeny to calculate a weight for each bull. Fikse and Banos (2001) developed new weighting factors based on daughter information for the application in MACE. In their study, they used the number of effective daughter contributions (EDC) instead of the number of daughters. Effective record contributions (ERC) can be obtained as a linear transformation of EDC. Taskinen *et al.* (2014) developed an iterative method for calculating ERC: so-called reverse reliability approximation. The authors applied their method for single-trait estimation. The types of weights used for de-regressed proofs (DRP) are important when either an approximate procedure is used to regress EBV or when individual reliabilities of final EBV are computed based on the prediction error variance of the model (Calus *et al.*, 2016). In this study, we present and test a multiple-trait reverse reliability approximation method to calculate ERCs. Our objective was to compare the performance of the different methods for calculating EDC/ERC.

Materials & methods

Interbull method 1. Interbull method 1 (IB1) uses a sire model and contains 2 steps. The first step consists of calculating reliability based on the animal's own performance records. This step accounts for the most important factors affecting reliability, including the proportion of genetic variance, repeatability of the trait, and the CG information. Step 2 calculates a weight for each bull using the calculated reliability of its progeny. EDC provides a measure of the precision of the daughter used to compute the bull's DRP. The formula for the computation of EDC, which includes the performance of the dam of daughter k of bull i, is: $\text{EDC}_i = \sum_k \frac{\lambda r^2_{k(o)}}{4 - r_{k(o)}(1 + r^2_{dam(o)})}$, where the summation is over all k daughters of the bull, $\lambda = \frac{4-h^2}{h^2}$, $r^2_{dam(o)}$, is the

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_362, © H. Ben Zaabza *et al.* 2022

reliability of the dam's own performance, and $r^2_{k(o)}$ is the reliability of animal *k*'s own performance computed as:

$$r^2_{k(o)} = \frac{n_k h^2}{1+(n_k-1)r^2_k}$$

where r^2_k is the reliability of the animal's records and n_k is the number of lactations of daughter *k* of the sire, adjusted for the CG size.

Interbull method 2. Interbull method 2 (IB2) is an extension of Interbull method 1 (IB1), where the animal model is used instead of the sire model. Thus, $\text{EDC}_i = \sum_k \frac{\lambda r^2_{k(o)}}{1-r_{k(o)}}$ and $\lambda = \frac{1-h^2}{h^2}$.

Single-trait reverse reliability estimation. Single-trait ERC calculation is based on so-called reverse reliability estimation (Taskinen *et al.* 2014), approximating ERC from a given reliability value by reversing the Harris and Johnson (1998) method. The original algorithm estimates animal model reliability using ERC. The reverse reliability estimation estimates $\widehat{\text{ERC}}$ that gives the same prediction error variance PEV as the original PEV*, where the asterisk stands for the initial value of prediction error variance calculated by the Tier and Meyer (2004) method when only a subset of (genotyped) animal reliabilities and their full pedigree are available.

Multi-trait reverse reliability estimation. This approach reverses the Tier and Meyer (2004) method. In the original Tier and Meyer (2004) algorithm, animal model reliabilities are approximated using ERC. In the reverse estimation approach, approximate $\widehat{\text{ERC}}$ are calculated that give the same prediction error variances PEV as the original PEV*. When the PEV* is from the full data with records on all animals, but ERC is solved for only a subset of (genotyped) animals, the approach automatically includes the information from the non-genotyped relatives of the genotyped animals. The calculation of $\widehat{\text{ERC}}$ involves the accumulation of information from progeny and parents using pedigree information as in the original Tier and Meyer approach. In the Tier and Meyer reliability estimation, each animal will receive a n_t by n_t prediction covariance matrix having the diagonals (PEV) used in the estimation of the reliabilities. However, in the reverse reliability estimation, with a given ERC^k for all animals, we estimate PEV^k, and if it does not correspond to PEV*, we search for a new ERC^{k+1} that will give PEV^{k+1} closer to PEV*. The approach needs to be repeated iteratively. The new $\text{ERC}^{[k+1]}$ was solved animal-wise from a non-linear equation:

$$\left\{\left(\mathbf{G}_i^{-1} + \mathbf{E}(\mathbf{c})\mathbf{R}_0^{-1}\mathbf{E}(\mathbf{c})\right)^{-1}\right\}_{jj} = \{\mathbf{y}\}_j$$

where $\{y\}$ are the original PEV* of animal *i* and trait *j*, $\mathbf{G}_i^{-1}$ contains $\mathbf{G}_0^{-1}$ and contributions from animal's offspring and ancestors, and R_0 is the between-trait residual variance-covariance matrix recorded for animal *i*. Animals may have missing observations for some traits, leading to missing ERC and a missing ERC pattern. For these ERC 'patterns', submatrices of R_0 were used by omitting rows and columns corresponding to the missing observations. The ERC values are in the diagonal E(c) matrix, $\{\mathbf{E}(\mathbf{c})\}_{jj} = \sqrt{\mathbf{c}_j}$, where c_j is the ERC for trait *j* to be solved. We implemented a Newton-Raphson based algorithm for solving the ERC_{rev}. See Ben Zaabza *et al.* (2022) for more details.

Performance of different methods for calculating EDC/ERC. The obtained ERC can be checked by back-calculating the reliability for animals by using an animal model. The model only contains an intercept and a genetic effect. The observations can have any value, and ERC are used as weights. The reliabilities for the model should be approximately the same as those used as input in ERC calculation.

Data. We applied the methods in Nordic (Danish, Finnish, and Swedish) Holstein Dairy cattle data, which had 8.10 million first, 5.97 million second, and 3.72 million third lactation 305-d milk yield records from 8.28 million Nordic Holstein dairy cows. The pedigree comprised up to 11 million animals, of which 274,145 were genotyped for 46,342 SNP markers.

Results & discussion

The correlations between approximate animal model reliabilities and back-calculated reliabilities are given in Table 1.

For Interbull method 1, correlations ranged from 0.969 for the first lactation to 0.937 for the third lactation. For Interbull method 2, however, they were between 0.921 for the first lactation and 0.883 for the third lactation. The correlations between reliabilities from ST-REV and back-calculated reliabilities ranged from 0.977 for first lactation to 0.958 for third lactation. An improvement of 0.02 was observed in the correlations between approximate animal model reliabilities and back-calculated reliabilities from MT-REV as compared to the ST-REV method. With Interbull method 1(2), MSE ranged from 0.099(0.0187) for the first lactation to 0.0214(0.0321) for the third lactation. With ST-REV (MT-REV), MSE ranged from 0.0081(0.0012) to 0.0107(0.0018), indicating that MT-REV outperformed ST-REV, especially for lactations 2 and 3. This could be expected because there were more missing observations in second and third lactations. The MT-REV also outperformed the Interbull method 1 and 2. This can be explained by the fact that the MT-REV method allows to account for information from non-genotyped animals using the pedigree relationship structure and the multi-trait covariance structure. Quite contrary, when both the sire and its daughter are genotyped, the use of the Interbull methods will lead to a double counting of the daughter record information. The MT-REV attempts to account for that double counting of information.

Conclusions

The performances of four methods for calculating EDC/ERC were compared in 305-d milk yield data in the first 3 lactations in Nordic Holstein dairy cattle. Our results indicate that ERC can be approximated satisfactorily for genotyped animals using a multiple-trait reverse reliability method, which outperformed the single-trait reverse reliability method mainly for traits with many missing records.

Table 1. Correlation (r), maximum difference (max), and mean-squared error (MSE) between the original reliabilities and the back-calculated reliabilities using the Interbull method 1 (IB1), Interbull method 2 (IB2), single-trait reverse reliability estimation (ST_REV), and multi-trait reverse reliability estimation (MT_REV).[1]

Methods	305-d milk lactation	r	Max	MSE	b_1	b_0
IB1	1	0.969	0.87	0.0990	0.814	-0.002
	2	0.949	0.86	0.0194	0.737	-0.006
	3	0.937	0.87	0.0214	0.714	-0.008
IB2	1	0.921	0.99	0.0187	0.766	0.011
	2	0.900	0.99	0.0298	0.689	0.006
	3	0.883	0.99	0.0321	0.665	0.004
ST_REV	1	0.977	0.59	0.0081	0.849	0.029
	2	0.965	0.63	0.0089	0.791	0.031
	3	0.958	0.63	0.0107	0.775	0.027
MT_REV	1	0.993	0.32	0.0012	0.935	0.005
	2	0.993	0.34	0.0018	0.910	0.007
	3	0.993	0.33	0.0018	0.909	0.005

[1] Intercept (b0) and slope (b1) of regression of original reliabilities on back-calculated reliabilities.

References

Ben Zaabza H., Taskinen M., Mäntysaari E.A., Pitkänen T., Aamand G.P., *et al.* (2022) J. Dairy Sci. In press.

Calus M.P.L., Vandenplas J., ten Napel J., and Veerkamp R.F. (2016) J. Dairy Sci. 99(8): 6403-6419. https://doi.org/10.3168/jds.2016-11028

Fikse W.F., and Banos G. (2001) J. Dairy Sci. 84(7): 1759-1767. https://doi.org/10.3168/jds.S0022-0302(01)74611-5

Harris B., and Johnson D. (1998) J. Dairy Sci. 81: 2723-2728. https://doi.org/10.3168/jds.S0022-0302(98)75829-1.

Interbull, July 2000 mimeo. Available at https://wiki.interbull.org/public/CoP_AppendixIV?action=SlideShow

Taskinen M., Mäntysaari E.A., Aamand G.P., and Strandén I. (2014) Proc. of the 10th WCGALP, Vancouver, Canada.

Tier B., and Meyer K. (2004) J: Animal Breed. Genet. 121:77-89.

363. JWAS version 2: leveraging biological information and high-throughput phenotypes into genomic prediction and association

H. Cheng[1*], R. Fernando[2], D. Garrick[3], T. Zhao[1] and J. Qu[1]

[1]Department of Animal Science, University of California Davis, Davis, CA 95616, USA; [2]Department of Animal Science, Iowa State University, Ames, IA 50010, USA; [3]School of Agriculture, Massey University, 4442, Palmerston North, New Zealand; qtlcheng@ucdavis.edu

Abstract

The initial release of JWAS was developed as an open-source software tool for single-trait or multi-trait linear mixed models including those used for genome-enabled prediction or genome-wide association studies. We have now made extensive updates to JWAS to incorporate new features to accommodate high-throughput phenotypes and biological information useful in bridging the gap between DNA sequence and phenotypic information. The new features include but are not limited to: (1) incorporating high-throughput phenotypes by simultaneous modeling of thousands of traits or through time-series (longitudinal trait) models; (2) using functional annotation information as a-priori biological knowledge; (3) extending mixed models to multi-layer neural networks that can accommodate intermediate omics data. This latest version of JWAS (version 2) is developed for researchers in both industry and academia to take full advantage of new information such as high-throughput phenotypes, functional annotations, and multi-omics data as they become increasingly available in this decade.

Introduction

The release of the USDA blueprint for animal genomics (2008-2017) triggered the development of many statistical methods to improve genetic progress as well as the implementation of these methods in software tools such as JWAS. The initial version of JWAS (Cheng *et al.*, 2018) was an open-source software tool for single-trait or multi-trait analyses accommodating those models used for genome-enabled prediction and genome-wide association studies, with either complete or incomplete genomic data, i.e. 'single-step' methods (Legarra *et al.,* 2009; Fernando *et al.,* 2014; Fernando *et al.,* 2016). A portfolio of Bayesian regression models including variable selection and shrinkage estimation methods were implemented in JWAS to provide broad scope of analyses.

With the development of high-throughput sequencing and phenotyping technology, the USDA blueprint in animal genomics for the next decade (2018-2027) (Rexroad *et al.*, 2019) has identified several areas requiring more advanced usage of genomic information for animal production. Thus, there was a need to develop statistical methods and tools that could incorporate high-throughput phenotyping, multi-omics data, and biological information such as functional annotations for genomic analyses. Extensive additional updates to our JWAS package were required to accommodate the issues identified in the new blueprint for improved utilization of genomic information. The availability of powerful and efficient statistical methods implemented in extensible software tools that can accommodate the advancing state of the art would have impacts on a wide range of breeding programs. To fill this gap, we extensively upgraded JWAS and released JWAS version 2. The main objective of this paper is to introduce new features in JWAS for researchers in industry or academia, so they can take full advantage of new information such as high-throughput phenotypes, multi-omics data, and functional annotation of animal genomes as they become increasingly available in this decade.

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_363, © H. Cheng *et al.* 2022

Materials & methods

A list of features in JWAS is shown in Table 1, and highlighted new features in JWAS version 2 are described below.

Mixed models extended to multi-layer neural networks including intermediate omics data. With the growing amount and diversity of intermediate omics data complementary to genomics (e.g. measures of DNA methylation, gene expression, and protein abundance), there is a need to develop methods to incorporate intermediate omics data into genome-enabled prediction and association studies to enhance conventional genomic evaluation. We have implemented a new method NN-LMM (Zhao *et al.*, 2021a; Zhao *et al.*, 2021b) to model the multiple layers of regulation from genotypes to intermediate omics features, then to phenotypes, by extending conventional linear mixed models to multi-layer artificial neural networks. NN-LMM incorporates intermediate omics features by adding middle layers between genotypes and phenotypes. Linear mixed models (e.g. pedigree-based BLUP, GBLUP, Bayesian Alphabet, single-step GBLUP, or single-step Bayesian Alphabet) can be used to sample marker effects or genetic values on intermediate omics features, and activation functions in neural networks are used to capture nonlinear possibly unknown relationships between intermediate omics features and phenotypes.

Using functional annotation information as a-priori biological knowledge. In addition to phenotypic, genomic, and pedigree data, massive amounts of biological information are being generated by the scientific community. For example, the Functional Annotation of Animal Genome (FAANG) consortium is functionally annotating genomic regulatory elements of domesticated animal species. However, such biological information is not generally considered in genomic prediction that routinely uses phenotypic information along with pedigree and SNP genotypes. We have implemented multi-class models (Wang *et al.*, 2021) in JWAS version 2 to allocate markers into different classes based on biological information and assigning separate priors to markers in these different classes. This is implemented for single-trait or multi-trait analyses.

Table 1. A list of features in JWAS for single-trait or multi-trait analyses.

Fixed class or covariate effects (e.g. age, sex)	
Non-genomic random effects (e.g. litter, pen)	
Use of pedigree information	
Joint fitting of direct and maternal genetic effects permanent environmental effects	
Use of genomic information	
Complete genomic data[1]	Genomic BLUP
	Bayesian Alphabet
Incomplete genomic data[2]	Genomic BLUP
	Bayesian Alphabet
Biological information	multi-layer neural networks[3]
	multi-class models[3]
High-throughput phenotypes	sparse factor models[3]
	random regression models[3]

[1] 'Complete genomic data' indicates that genotypes are available on all individuals.

[2] 'Incomplete genomic data' indicates that genotypes are not available on all individuals ('single-step' analyses).

[3] These new features are described in the Materials & methods section.

Incorporating high-throughput phenotypes by simultaneous modeling of thousands of traits through sparse factor models. Large-scale phenotypic data are becoming increasingly accessible due to advances in high-throughput phenotyping platforms and technologies for multi-omics profiling. Although the incorporation of large-scale phenotypic data into genome-enabled analysis can enhance the power of prediction and association inference, genomic analyses of high-dimensional, highly correlated data are challenging. We have implemented a Bayesian sparse factor model (Runcie *et al.*, 2021) with different prior assumptions on marker effects to simultaneously analyze hundreds to thousands of traits for genomic prediction or GWAS. This Bayesian sparse factor model can effectively reduce a large multi-trait model into a set of parallel single-trait models by introducing the concept of latent traits, i.e. traits that are not directly observed, but can be inferred based on their effects on the observed traits. This model is based on some biologically reasonable assumptions including: (1) a limited number of the latent traits control the majority of variation; (2) each latent trait controls only a subset of the observed traits; (3) each latent trait is controlled by genetics and the environment.

Incorporating high-throughput phenotypes through time-series (longitudinal trait) models. In animal and plant improvement, longitudinal traits may be recorded across several time points on a physiological cycle, and might be jointly analyzed to improve the performance relative to an analysis at one time point. The recent advancement in phenotyping platforms enables acquisition of large-scale non-destructive phenotypes measured at frequent intervals. In pedigree-based analyses, the use of random regression models (RRM) has been a typical approach employed for such analyses of longitudinal traits (Henderson, 1982); Laird and Ware, 1982). We have implemented Bayesian random regression models for marker effects that can accommodate variable selection for the analysis of longitudinal data that would accrue through high-throughput phenotyping.

Results

The JWAS software tool was and is being developed as a single-language software tool that is easy for both champions and novice community members to use, maintain, modify, or extend. In terms of speed, for the same BayesC analysis, JWAS is now faster than the C++ program, GenSel. Further, JWAS is relatively easy to extend: to accommodate categorical traits in JWAS required adding only about 40 lines of code; to extend conventional Bayesian regression methods to single-step analyses required adding one file composed of a few hundred of lines of code and a few minor modifications to the original code.

Discussion

Whole-genome analyses are usually computationally intensive even for data sets of moderate size. Although these analyses can be implemented with most dynamic languages such as R or Python with very readable code that is easy to understand, modify and extend, they are often too slow for real data analyses. Therefore, for computational efficiency, compiled languages such as C, C++, or Fortran have historically been used to implement methods for genomic analyses. In order to make these more accessible, a dynamic language such as R may be used as a user interface to the underlying methods that are written in the compiled language. The problem of using compiled languages, however, is that they are usually hard to understand, modify, extend and maintain, in addition to being difficult to readily deploy across different operating systems. Julia (Bezanson *et al.*, 2017), a relatively new scientific programming language, approaches the computing speed of compiled languages, but retains the benefits of dynamic language.

References

Bezanson J., Edelman A., Karpinski S., and Shah V. B. (2017) SIAM Review 59(1):65-98. https://doi.org/10.1137/141000671

Cheng H., Garrick D.J., and Fernando R.L. (2018) Proc. of the 11th WCGALP, Auckland, New Zealand. Fernando R.L., Dekkers J.C., and Garrick D.J. (2014) Genet. Sel. Evol. 46(1):1-13. https://doi.org/10.1186/1297-9686-46-50

Fernando R.L., Cheng H., Golden B.L., and Garrick D.J. (2016) Genet. Sel. Evol. 48(1):1-8. https://doi.org/10.1186/s12711-016-0273-2

Henderson C.R. (1982) Biometrics 38(3):623-640. https://doi.org/10.2307/2530044

Laird N.M., and Ware J.H. (1982) Biometrics 38(4):963-974. https://doi.org/10.2307/2529876

Legarra A., Aguilar I., and Misztal I. (2009) J. Dairy. Sci. 92(9):4656-4663. https://doi.org/10.3168/jds.2009-2061

Rexroad C., Vallet J., Matukumalli L.K., Reecy J., *et al.* (2019) Front. Genet. 10:327. https://doi.org/10.3389/fgene.2019.00327

Runcie D.E., Qu J., Cheng H., and Crawford L. (2021) Genome Biol. 22(1):1-25. https://doi.org/10.1186/s13059-021-02416-w

Wang Z., and Cheng H. (2021) Front. Genet. 12. https://doi.org/10.3389/fgene.2021.717457

Zhao T., Fernando R., and Cheng H. (2021a) G3.11(10). https://doi.org/10.1093/g3journal/jkab228

Zhao T., Zeng J., and Cheng H. (2021b) BioRxiv preprint https://doi.org/10.1101/2021.12.10.472186

364. ACCF90GS2: software for fast approximation of reliabilities of estimated breeding values in single-step GBLUP

M. Bermann[1*], D. Lourenco[1], A. Cesarani[2] and I. Misztal[1]

[1]*Department of Animal and Dairy Science, University of Georgia, 425 River Road, Athens, GA 30602, USA;*
[2]*Dipartimento di Agraria, Università degli Studi di Sassari, 07100 Sassari, Italy;* mbermann@uga.edu

Abstract

Inverting the mixed model equations (MMEs) in single-step GBLUP for calculating reliabilities is not feasible for large MMEs. We developed a program named ACCF90GS2 for approximating reliabilities of breeding values in single or multiple trait contexts. Two datasets were used to test the software. A small dataset was used to assess the accuracy of the method, whereas a large dataset including 335,325 genotyped animals was used to test the software performance. Correlations between exact and estimated reliabilities were 0.98 for the direct effect and over 0.94 for the maternal.Intercepts and slopes of regression of correct reliabilities from ssGBLUP on approximation reliabilities ranged from 0.01 to 0.04 and from 0.92 to 0.96, respectively. For the large dataset, the total computing time was 20 minutes. The computing cost of ACCF90GS2 increases linearly with the number of genotyped animals and traits. ACCF90GS2 is the newest tool in the BLUPF90 software suite to efficiently approximate reliabilities of breeding values with large-scale genomic data.

Introduction

Single-step GBLUP (ssGBLUP) allows estimating the breeding values jointly for genotyped and non-genotyped animals (Aguilar *et al.*, 2010). For that, ssGBLUP combines the inverses of the numerator relationship matrix (A), the genomic relationship matrix (G), and the numerator relationship matrix for the genotyped animals (A_{22}). When the number of genotyped animals is large, inverting G is not feasible. To overcome this problem, Misztal *et al.* (2014a) introduced the Algorithm for Proven and Young (APY). Although APY makes solving the mixed model equations (MMEs) possible for many genotyped animals, the inversion of the MMEs is still not feasible. Therefore, with extensive data, it is still impossible to obtain the prediction error variances (PEV) for reliability calculation from the diagonal of the inverse of the MMEs. Bermann *et al.* (2022) presented a method based on Liu *et al.* (2017) for approximating the reliabilities of estimated breeding values when APY is used. Bermann *et al.* (2022) showed the efficiency of this method with beef data for various models, but without considering maternal effect, which is common for some beef traits. We developed a program named ACCF90GS2 for approximating reliabilities of breeding values in single or multiple trait contexts that can accommodate models used in genetic evaluations.

Materials & methods

Approximation of reliabilities. We used the method of Bermann *et al.* (2022) for approximating the reliabilities of GEBV. The core of this method is the calculation of the reliabilities for a GBLUP model with APY, hereafter called genomic reliabilities. Their calculation is done by the block-sparse inversion of:

$$\mathbf{G}_{APY}^{-1}\,\sigma_{\mathrm{u}}^{-2} + \mathbf{D}\left(\sigma_{\mathrm{p}}^{2} - \sigma_{\mathrm{u}}^{2}\right),^{-1} \qquad (1)$$

where $\mathbf{G}_{APY}^{-1}$ is the inverse of the genomic relationship matrix calculated with APY, D is a diagonal matrix of effective record contributions (ERC), σ_{u}^{2} is the genetic variance, and σ_{p}^{2} is the phenotypic variance. Note that in the presence of other random effects besides the direct genetic effect, $\sigma_{\mathrm{p}}^{2} - \sigma_{\mathrm{u}}^{2}$ is greater than the error variance. If a maternal genetic effect is present, it can be treated as a direct genetic effect. This involves replacing σ_{u}^{2} in (1) by the maternal genetic variance (σ_{m}^{2}) and calculating ERC with the records linked to

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_364, © M. Bermann *et al.* 2022

the maternal genetic effect. The inclusion of a residual polygenic effect (RPG) in the model is done when constructing $\mathbf{G}_{APY}^{-1}$ as in Masuda *et al.* (2015). If so, the estimated reliability includes the residual polygenic effect.

The implementation of the method of Bermann *et al.* (2021) and the adjustment for maternal effects were as follows. First, we calculated the pedigree reliabilities (rel_{ped}) for direct and maternal effects (Misztal and Wiggans, 1988; Strabel *et al.*, 2001). Second, the pedigree reliabilities were back-solved using a root-finding technique to obtain ERC. Third, we calculated the genomic reliabilities (rel_{gen}) by a sparse inversion of Equation 1; two sparse inversions are needed when a trait has an associated maternal effect. Fourth, we calculated a second set of pedigree reliabilities $\left(rel_{\mathbf{A}_{22}}\right)$ with the method of Harris and Johnson (1998). The ERC for calculating these reliabilities were the same as in the second step but zeroing the ERC for non-genotyped animals. Then, we obtained the ERC for ssGBLUP (ERC_{ss}) as (Liu *et al.*, 2017):

$$ERC_{ss} = \lambda\left(\frac{rel_{ped}}{1 - rel_{ped}} + \frac{rel_{gen}}{1 - rel_{gen}} - \frac{rel_{\mathbf{A}_{22}}}{1 - rel_{\mathbf{A}_{22}}}\right) \quad (2)$$

where $\lambda = \frac{1-h^2}{h^2}$ and h^2 is the trait heritability calculated as $h^2 = \frac{\sigma_u^2}{\sigma_p^2}$ for direct effects and as $h^2 = \frac{\sigma_m^2}{\sigma_p^2}$ for maternal effects. Finally, the ssGBLUP reliabilities (rel_{ss}) were calculated as:

$$rel_{ss} = \frac{ERC_{ss}}{ERC_{ss}+\lambda} \quad (3)$$

The above procedure was repeated for each trait. Afterward, the adjustment proposed by Strabel *et al.* (2001) was used for accounting for multiple traits.

Software implementation. The method described in Bermann *et al.* (2022) and the adjustment for maternal effects were programmed in Fortran 90. We developed a software named ACCF90GS2, which will be included in the BLUPF90 software suite (Misztal *et al.*, 2014b). The software is optimized for parallel computations with OpenMP (http://www.openmp.org) and efficient matrix operations with the Intel Math Kernel Library (MKL; Inter Corporation).

Data. The data for this study was provided by the American Angus Association (St. Joseph, MO). Two datasets were created. A small dataset (Data1) was used to assess the accuracy of the approximation method, whereas a large dataset (Data2) was used to test the computational efficiency of the software. Table 1 shows a summary of each dataset.

Records for birth weight (BW), weaning weight (WW), and post-weaning gain (PWG) were available. In the model, the three traits had herd-year-season as a fixed effect and the direct genetic effect as random. BW and WW had maternal effect, whereas WW also included the maternal permanent environmental effect.

Table 1. Number of animals, animals with records, and genotyped animals for the two datasets.

		Data1	Data2
Animals in the pedigree		172,089	10,213,401
Animals with records		78,641	8,681,659
Genotyped animals	Core	10,523	10,523
	Non-core	9,477	324,802

Results

Statistics for the approximation method are shown in Table 2. The elapsed time for approximating the reliabilities was 0.9 and 20.8 minutes for Data1 and Data2. Note that Data2 is roughly 60 times bigger than Data1. However, the elapsed time for Data2 was 23 times the elapsed time for Data1.

Table 2. Correlation, intercept, slope, and mean absolute change (MAC) between the reliabilities calculated from the inverse of the MMEs and the approximated reliabilities for genotyped animals in Data1.

Trait	Effect	Correlation	Intercept	Slope	MAC
BW	Direct	0.98	0.04	0.93	0.01
	Maternal	0.94	0.03	0.92	0.01
WW	Direct	0.98	0.02	0.94	0.01
	Maternal	0.95	0.03	0.93	0.01
PWG	Direct	0.98	0.02	0.96	0.01

Discussion

The results from Table 2 show that the approximation works well, while the computing time is reasonable. The approximation quality depends on the statistical model, population structure, and genotyping strategy (Erbe *et al.*, 2018). The proposed method assumes that the direct and maternal effects are uncorrelated, which may impact the approximation quality. In this study, the correlations between direct and maternal effects were low (-0.24 for BW and zero for WW). Intuitively, the impact would be more significant when traits are highly correlated.

ACCF90GS2 also allows replacing $\mathbf{G}_{APY}^{-1}$ with a used defined G^{-1} in (1). In such a case, a dense inversion increases the computing cost with the number of genotyped animals and the inclusion of a RPG depends on how the user defined G^{-1}. Further, the program has an option to enhance reliabilities for non-genotyped animals using relatives' genomic information. The procedure derives weights from the single-step reliabilities of genotyped animals and applies them as ERC for non-genotyped animals. Overestimation of reliabilities for genotyped animals is expected, especially when the number of genotyped animals is large (Liu *et al.*, 2017). Consequently, ACCF90GS2 includes tools for adjusting the reliabilities of the genotyped animals. This adjustment is a simple linear regression with parameters given by the user, which optimally should be calculated from cross-validation studies. Other possible adjustments, most likely using relatives' contributions, would be explored for further implementation in ACCF90GS2. As future developments, ACCF90GS2 will include a method for approximating reliabilities of indirect predictions and random regression models.

References

Aguilar I., Misztal I., Johnson D.L., Legarra A. Tsuruta S. *et al.* (2010) J Dairy Sci. 93(2):743-52. https://doi.org/10.3168/jds.2009-2730

Bermann M., Lourenco D., Misztal I. (2022) Journal of Animal Science, skab353, https://doi.org/10.1093/jas/skab353

Erbe M., Edel C., Pimentel ECG., Dodenhoff J., Gotz KU. (2018) Interbull Bull 54:1–8.

Harris B., Johnson D. (1998) J Dairy Sci. 81(10):2723-8. https://doi.org/10.3168/jds.S0022-0302(98)75829-1

Liu, Z., VanRaden P.M., Lidauer M.H., Calus M.P.L., Benhajali H. *et al.* (2017) Interbull Bull 51:75–85.

Masuda Y., Misztal I., Tsuruta S., Legarra A., Aguilar I. *et al.* (2015) J Dairy Sci. 99(3):1968–1974. https://doi.org/10.3168/jds.2015-10540

Misztal I., Wiggans G.R. (1988) J Dairy Sci. 71(2):27-32. https://doi.org/10.1016/S0022-0302(88)79976-2.

Misztal I., Legarra A., Aguilar I. (2014a) J Dairy Sci. 97(6):3943-52. https://doi.org/10.3168/jds.2013-7752

Misztal I., Tsuruta S., Lourenco D.A.L., Masuda Y., Aguilar I., *et al.* (2014b) - Manual for BLUPF90 family of programs – Available at: http://nce.ads.uga.edu/wiki/doku.php?id=application_programs

Strabel T., Misztal I., Bertrand J. K. (2001) Journal of Animal Science 79(4):833–839. https://doi.org/10.2527/2001.794833x

365. One for all: LMT – the linear models toolbox

V. Boerner

Centre for Quantitative Genetics and Genomics (QGG), Aarhus University, Denmark; vinzent.boerner@gmx.de

Abstract

Developments in hardware resources, external computing devices, third party libraries and languages capacities must be considered if big data applications aim for cutting-edge performance. Legacy code structure can make such considerations difficult or impossible. For addressing these issues, LMT ('linear models toolbox') has been developed as a successor for DMU. LMT is a stand-alone, single-executable, instruction file driven program for genetic evaluation and variance component estimation in multivariate data sets with hundreds of millions of equations. LMT supports a wide range of models including ssSNPBLUP random regression test-day models, metafounders, and explicitly modelled genetic groups. Tested on large scale industry data, LMT decreased the necessary processing time by 93% when compared to DMU, and needed 1.7 hours to process a random regression test-day model with 350 million equations and 250 million observations.

Introduction

Over the last two decades the computing and programming environment where quantitative genetics operates in has seen a dramatic change. Hardware resources are abundant, fast third-party libraries and JIT compilers underpin powerful scripting languages, modern language features are available for easier maintenance and extension, and external computing devices enable unprecedented speed. In some cases, the structure of legacy code may not allow to address these developments efficiently, and only a re-development may enable sufficient performance. This paper describes LMT ('linear models toolbox'), which has been developed from scratch as a successor for DMU (Madsen *et al.* 2014) to address maintenance, interface, layout, and performance issues while maintaining a versatility equivalent to DMU.

Modelling capacity

Linear mixed models consist of a combination of factor structures and variance structures, and the modeling capacity of a linear mixed model software is a function of the supported operations applicable to supported factor and variance structures. Below, LMT's modeling capacity is described using this concept.

Supported variance structures. The most general variance structure supported by LMT is $\Omega_L(I \otimes \Gamma)\Omega_L'$, where Ω_L a block-diagonal matrix of lower Cholesky factors of within factor level co-variance matrices (e.g. an animal specific a genetic co-variance matrix), Γ is a matrix of between factor level co-variances (e.g. the numerator relationship matrix), and I is an identity matrix. Therefore, LMT supports assigning a unique co-variance matrix to every single modelled random factor level. Note that under common assumptions this structure simplifies to $\Gamma \otimes \Sigma$ where Γ is usually set to identity (I), the pedigree derived numerator relationship matrices (A), or a single step relationship matrix (H), and Σ is a co-variance matrix between correlated factors across or within trait.

For Γ, LMT supports I, auto-regressive matrices, externally supplied matrices in sparse or dense format, A, and H matrices for ssGBLUP (H_G Christensen and Lund 2010), ssGTBLUP (H_T, Mäntysaari *et al.* 2017), and ssSNPBLUP (H_S, Liu *et al.* 2014). Pedigree structures supported for building A and H are ordinary pedigrees and probabilistic pedigrees with more than one pair of parents per individual (Perez-Enciso and Fernando 1992), where the latter is a key feature for genetic evaluation in plants, fish and extensive livestock

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_365, © V. Boerner 2022

production systems. Further, diagonal elements of A can be derived in an iterative manner accounting for relationships between founders (VanRaden 1992).

For ssSNPBLUP, LMT supports co-variances between and within markers following structure $\Omega_L(I \otimes \Gamma)\Omega_L'$, where the supported Γ types are those pointed out above. For H_G, H_Γ, and H_S a non-zero polygenic weight can be provided for all H structures, and a single keyword in the instruction file invokes the adjustment of G to A (Christensen *et al.* 2012) for all types of H. By providing a pedigree containing phantom parents for the construction of A, for Γ of type A, H_G, H_Γ, and H_S LMT will automatically absorb the genetic group regression matrix into the respective Γ (Westell *et al.* 1988). Further, if provided with a metafounder co-variance matrix LMT will automatically build a metafounder enhanced numerator relationship matrix (A_{mf}) for all Γ which involve A (Legarra *et al.* 2015).

Supported factor structures. LMT supports the following factor structures: classification variable, real co-variables (e.g. 'weight'), integer co-variables (e.g. 'days in milk'), co-variables nested within classification variables (e.g. 'test day models'), genetic group co-variables, and arbitrary real number design matrices. If co-variables are nominated as integer LMT automatically generates regression tables. Further, regression matrices for genetic group co-variables are generated directly from the assigned pedigree during the model building process. The ability to model genetic groups as a separate factor allows to arbitrarily mix metafounders and genetic groups in the same model. For co-variables LMT understands complex polynomials (e.g. $\sin(x)/\mathrm{sqrt}(x)^3$) and will automatically conduct all necessary expansion to fit a random co-variable to its assigned variance structure.

Supported operations. LMT supports solving linear mixed models for obtaining BLUP and BLUE solutions for the modelled factor levels. Two different solvers are available: a direct solver which factorises the mixed model equation system, and an iterative solver which uses the preconditioned conjugate gradient (PCG) algorithm. Both solvers can be applied to equation systems involving any factor or variance structure, although computational feasibility and resource availability limit the application of the direct solver to medium sized data sets and genomic data included via H_G. The PCG solver can be applied to an unlimited number of equations and genotyped individuals and all types of single step models. LMT also supports the exact calculation of prediction error co-variances. For this operation both the above solvers can be used, calculating exact prediction error co-variances for target factor levels even for large multi-variate ssSNPBLUP models. Approximations of prediction error co-variances can be obtained using LMT's multi-chain prediction error co-variance sampler. Implementing a method described in (Harville 1999), the sampler runs a user-defined number of parallel chains on shared resources. That is, all factor and variance structures are shared between the chains which is a key feature to keep the computing resource requirements manageable when genetic markers are included in the model. Sampling is supported for all factor structures and all variance structures, except H_T and H_S. For estimating variance components, LMT provides a MC-EM-REML module (Harville 2004) and two different Gibbs samplers: a blocked sampler which samples all correlated factors within factor level from their joint conditional posterior distribution, and a single pass sampler which samples all factor levels jointly from their posterior distribution.

Additional features. 'On the fly of everything' is the core concept of LMT, meaning that absolutely everything necessary to run a particular model or job stream is calculated as required without the necessity to utilize or call any additional executable. LMT can accommodate for uploading an arbitrary number of pedigrees, genotype files, external Gs, external Γs, etc. and calculates derivatives of that input on the fly, whether it is diagonal elements of A, A_{mf} or H, or building an arbitrary number of Gs from genotypes.

Model communication

LMT's multi-variate model syntax (LMT-mms) is very closely related to the mathematical linear model notation found in the scientific notation. For putting the maximum amount of information into the equation the concept of specifiers has been developed which extends the model y=Xb+Zu to y()=X()b()+Z()u(), where y, X and Z are column names retrieved from the data file header, b and u are arbitrary user-defined factor names, and the parenthesis are optional specifiers which may contain further information about the response variable, model variable, or factor. Contrarily to other model syntax, LMT-mms maintains an acronym for the factor which is a core feature to allow for unambiguous cross-classification of factors over response variables. Variable specifiers contain all information necessary to build a variable and can be described as a tree diagram. A univariate model fitting a fixed classification variable, a nested continuous co-variables with polynomial expansion, and a genetic group co-variable can be written in LMT-mms as:

weight = herdyear×b+age(t(co(p(1,2,3);n(sex))))×c+id(t(gg(myped)))×g,

where 'weight', 'herdyear', 'age', 'sex' and 'id' are column names in the data file, 'b', 'c', and 'g' are user-defined factor names, '1,2,3' in 'age(t(co(p(1,2,3);n(sex))))' are instruction file locations of user-defined polynomials, and 'myped' is a the user-defined name of a user-supplied pedigree. A similar concept applies to the factor specifier which communicates information whether the factor is random and which variance is assigned to it. A bi-variate random regression test-day model using three different polynomials and an automatically generated regression table can be written in LMT-mms as:

milk=hy×b1+dim((t(co(t(i);p(1,2,3);n(id)))))×u1(v(g(1,3,5)))

prot=hy×b2+dim((t(co(t(i);p(1,2,3);n(id)))))×u2(v(g(2,4,6))),

where 'dim' is a column name in the data file, and the specifier in 'u1(v(g(1,3,5)))' assigns the Σ diagonal elements 1,3 and 5 of variance structure with the user-defined name 'g' to factor 'u1'.

Performance

The challenge of developing LMT's software frame was to achieve both, versatility as well as speed, all combined in a single executable. For bench-marking the speed LMT's PCG solver performance was tested with several different industry data sets used for genetic evaluation. Three examples are provided below where the data structures differ with regard to the computationally most demanding operations of the workflow. All runtime parameters provided are indicative.

DanBred ssGTBLUP evaluation for pigs. The data set consisted of 4 traits, about 3 million observations, 1 million animals in the pedigree, and 150,000 animals with 40k marker genotypes. This resulted in an equation system with 4.9 million equations, 4 correlated random genetic effects, 3 other random factors with correlated effects (e.g. birth-litter, service sire, etc.), and an H_T matrix constructed with a polygenic weight of 0.05 and an adjustment of G to A. Transitioning from the DMU tool chain to LMT decreased the total processing time from about 15 hours to less than 1 hour (-93%), and the seconds per solver iteration from 13.18 to 1.8, of which 90% where spent on T-matrix multiplications. Both programs ran on the same hardware and threading resources and used the same convergence criterion.

DanBred ssSNPBLUP evaluation for pigs. The data set consisted of 5 traits, 1.2 million animals in the pedigree, 220,000 animals with 40k marker genotypes, and H was constructed with the polygenic weight of 0.25 and an adjustment of G to A, which resulted in a system with 19 million equations. On an Intel(R) Core(TM) i9-10980XE CPU @ 3.00GHz with 18 cores using a ssSNPBLUP model instead of a ssGTBLUP

model reduced the total runtime from 4.6 hours to 2.1 hours with a convergence threshold $\frac{||Cx_i - b||}{||b||}$<10e-11, and reduced the peak memory requirement from 134GB to 17GB, making the ssSNPBLUP model the most viable solution for a time-critical and memory-aware single step genetic evaluation.

DataGene random regression test-day evaluation for Australian Holstein. The data set consisted of 9 traits (milk, fat and protein in 3 lactations) with a total of 236 million observations and 6.9 million animals in the pedigree. The data were modelled using a random regression test-day model, which resulted in 27 correlated genetic effects, 27 correlated permanent environmental effects, and a system of 355 million equations. On an Intel(R) Core(TM) i9-10980XE CPU @ 3.00GHz with 18 cores the total processing time for the dataset was 1.7 hours with a convergence threshold $\frac{||x_i - x_{i+1}||}{||x_i||}$<10e-5.

Conclusions

With only the AIREML functionality yet to be integrated, nearly all DMU functionalities, spread over 5 different executables, have been successfully integrated into LMT and further functionalities have been added. The achieved performance gains show that generalisation does not contradict performance.

References

Christensen O.F. and Lund M.S. (2010) Genet. Sel. Evol. 42:2. https://doi.org/10.1186/1297-9686-42-2.

Christensen O.F., Madsen P., Nielsen B., Ostersen T. and Su G. (2012) Animal 6(10):1565-1571. https://doi.org/10.1017/S1751731112000742

Harville D.A. (1999) Linear Algebra Appl. 289:203–24. https://doi.org/10.1016/S0024-3795(98)10183-0.

Harville D.A. (2004. J. Stat. Comput. Sim. 74:135–53. https://doi.org/10.1080/0094965031000110588.

Legarra A., Christensen O.F., Vitezica Z.G., Aguilar I., and Misztal I. (2015) Genetics 200 (2):455–68. https://doi.org/10.1534/genetics.115.177014.

Liu Z., Goddard M.E., Reinhardt F., and Reents R. (2014) J. Dairy Sci. 97(9):5833–50. https://doi.org/10.3168/jds.2014-7924.

Madsen, P., Jensen J., Labouriau R., Christensen O.F., and Sahana G. (2014) Proc. of the 10th WCGALP: Vancouver, Canada.

Mäntysaari E.A., Evans R.D., and Strandén I. (2017) J. Anim. Sci. 95(11):4728–37. https://doi.org/10.2527/jas2017.1912.

Perez-Enciso M. and Fernando R.L. (1992) Theor. Appl. Genet. 84(1):173–79. https://doi.org/10.1007/BF00223997.

VanRaden P.M. (1992) J. Dairy Sci. 75(11):3136–44. https://doi.org/10.3168/jds.S0022-0302(92)78077-1.

Westell R.A., Quaas R.L., and Van Vleck L.D. (1988) J. Dairy Sci. 71: 1310–18. https://doi.org/10.3168/jds.S0022-0302(88)79688-5.

366. Recent updates in the BLUPF90 software suite

D. Lourenco[1*], S. Tsuruta[1], I. Aguilar[2], Y. Masuda[3], M. Bermann[1], A. Legarra[4] and I. Misztal[1]

[1]University of Georgia, Department of Animal and Dairy Science, Athens, GA 30602, USA; [2]Instituto Nacional de Investigación Agropecuaria, 11500 Montevideo, Uruguay; [3]Rakuno Gakuen University, 069-8501, Ebetsu, Hokkaido, Japan; [4]Institut national de recherche pour l'agriculture, l'alimentation et l'environnement, UMR1388 GenPhySE, 31326 Castanet Tolosan, France; danilino@uga.edu

Abstract

BLUPF90 is a flexible software suite that estimates variance components, cross-classified and covariate effects, breeding values, accuracies, and SNP effects, among others. It is based on mixed models and focuses mainly on animal breeding and genetics applications. This collection of software is continually under development to incorporate new methods and algorithms, improve efficiency, and broaden usage. A significant change undergoing BLUPF90 is the merge of variance components and breeding values estimation software. Some of the recent developments include the implementation of *P*-values for SNP effects in single-step GWAS, a new algorithm to approximate reliability or accuracy of GEBV, the inclusion of metafounders or unknown parent groups in ssGBLUP to set the base populations, and several optimizations to speed up computing time with many genotyped animals. BLUPF90 is a resourceful collection of efficient software that handles millions of genotyped animals and dozens of millions of records or individuals in pedigree.

Introduction

The ever-increasing amount of data used in breeding and genetics applications requires efficient software to solve the mixed model equations (MME). In 1997, Ignacy Misztal developed a simple BLUP program to compute solutions for the MME in Fortran 90/95. The simplicity and flexibility of the architecture allowed extensions and improvements by several generations of scholars. Twenty-five years and many contributors later, BLUPF90 became a collection of software that handles virtually all models used in genetic evaluations in animal breeding. This collection includes software for variance components estimation in linear and linear-threshold models, large-scale genetic evaluations using linear and linear-threshold models, exact and approximate computations of accuracies, and general genomic computations. The software supports different models per trait for virtually any number of traits, multiple effects, random correlated and non-correlated effects, random regressions, missing observations, different covariance structures supplied by the user, and extensive genomic data.

Using genomic data stimulated the most comprehensive changes to BLUPF90. This is because the number of opportunities with genomics is endless. A genomic library with well over 200 options was created and encapsulated by nearly all BLUPF90 programs. Those options are mainly related to quality control of genomic data, statistics and visualization, construction and inversion of genomic and pedigree relationship matrices for genotyped animals, and matrix scaling to ensure compatibility in ssGBLUP. Although the primary genomic method implemented in BLUPF90 is ssGBLUP, GBLUP is also supported. Because of the equivalence between GBLUP-based methods and SNP-BLUP, SNP effects are easily computed through backsolving.

BLUPF90 has been in constant progress for 25 years. Some of the most recent developments in this software collection include the consolidation of variance components and breeding values estimation software, implementation of *P*-values for SNP effects in single-step GWAS, a new algorithm to approximate reliability or accuracy of GEBV, the inclusion of metafounders or unknown parent groups in ssGBLUP, and

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_366, © D. Lourenco *et al.* 2022

optimizations to speed up computations with many genotyped animals. This work will present the most critical innovations recently brought to BLUPF90.

Materials & methods

Software consolidation. There are currently 11 free programs in the BLUPF90 software suite (http://nce.ads.uga.edu/html/projects/programs/). Some estimate variance components, and some estimate breeding values. Some are frequentist (REML), and some use Bayesian inference through Gibbs Sampling. This generates questions on which software one should use. To avoid this issue, we recently consolidated the linear, frequentist software into one piece called blupf90plus and the most used Bayesian, Gibbs Sampling into gibbsf90plus. Thus, blupf90plus combines blupf90, remlf90, and airemlf90, whereas gibbsf90plus integrates gibbs2f90, gibbs3f90, thrgibbs1f90, and thrgibbs3f90. In blupf90plus, the default option is breeding values estimation, and variance components estimation (VCE) is turned on by adding OPTION method VCE to the parameter file. For gibbsf90plus, the same software is used for linear and threshold models. These consolidated programs were thoroughly tested with real and simulated populations.

Unknown Parent Groups (UPG) in ssGBLUP. In the BLUP animal model, the mixed model equations (MME) account for UPG using the QP transformation (Quaas and Pollak, 1981). Besides the default QP transformation for A^{-1}, in ssGBLUP, UPG can be assigned to A^{-1}, the inverse of the genomic relationship matrix (G^{-1}), and A^{-1} for genotyped animals (A_{22}^{-1}). This is called the QP transformation for H^{-1} (Misztal *et al.*, 2013). Because pedigree missingness does not affect genomic relationships, UPG can be assigned to A^{-1} and A_{22}^{-1}; this is known as the altered QP transformation for H^{-1} (Tsuruta *et al.*, 2019). The transformations were implemented in the BLUPF90 programs. QP transformation for A^{-1} is the default, whereas the QP for H^{-1} is activated by OPTION exact_upg and the altered QP for H^{-1} needs the previous option along with OPTION TauOmegaQ2 0.0 1.0. All three transformations were tested using US Holstein data for type and production traits with up to 861k genotyped animals.

Metafounders (MF) in ssGBLUP. MF act as proxies for the base animals. Metafounders and UPG are similar concepts; however, metafounders can be related and describe the buildup of coancestry, which is not possible for UPG. With MF, A_{22} is adjusted to G computed with 0.5 allele frequency. Both A^{-1} and A_{22}^{-1} are then modified by a matrix of relationships within and across metafounders (Γ). We developed a software called gammaf90 that computes Γ, and we changed BLUPF90 software to account for MF. When using MF for GEBV estimation, the RANDOM_TYPE in renf90.par should be add_an_meta. Among the free software, only blupf90plus and gibbsf90plus work with MF. The implementation of MF was tested in US Holstein data (861k genotypes), chicken data (107k genotypes), and dairy sheep data (6k genotypes).

Optimizations for large-scale ssGBLUP evaluations. The algorithm for proven and young (APY) was implemented in the BLUPF90 suite in 2015. APY allows the construction of a sparse representation of G^{-1}, making computations with millions of genotyped animals very efficient (Misztal *et al.*, 2014). An efficient algorithm for A_{22}^{-1} was also implemented (Masuda *et al.*, 2017). For blending of G and A_{22}, the initial implementation was inefficient. We recently changed the implementation of Colleau (Colleau, 2002) in BLUPF90 to compute coefficients of A_{22} separately for core and noncore animals. This new implementation was tested with US Holstein data containing 3.4M genotyped animals. Extensive changes in the iteration on data (i.e. iod) programs were also made to improve the computational efficiency in evaluations with extensive data. This included splitting data and pedigree and using different computing cores for parallel i/o.

An efficient algorithm to approximate reliability or accuracy of GEBV. Reliabilities of GEBV can be calculated as a function of prediction error variances (PEV) obtained from the diagonal of the inverse of the

coefficient matrix of MME. This inverse is not feasible with large datasets, and therefore, reliabilities need to be approximated. An algorithm based on Liu *et al.* (2017) was developed that considers the sparsity of APY G^{-1} (Bermann *et al.*, 2021). The algorithm computes weights for pedigree and phenotype contributions, adds the weights to the diagonals of APY G^{-1}, and uses block sparse inversion to compute coefficients for core and noncore animals separately. A brand-new program, accf90GS2, was created to accommodate the new reliability calculations. It is based on accf90 and considers maternal, multi-trait, and repeatability models. Tests included single- and three-trait models using American Angus datasets with up to 335k genotyped animals. An algorithm for the exact reliability of indirect predictions was also recently implemented.

***P*-values for SNP effects in single-step GWAS.** Because of the equivalence between ssGBLUP and ssSNP-BLUP, GEBV from the former can be backsolved to SNP effects. GWAS may use SNP effects or proportion of variance explained by SNP; however, it is usual to present *P*-values to test the statistical significance of SNP on traits of interest. We implemented *P*-values in ssGWAS based on PEV for SNP effects (Aguilar *et al.*, 2019). These GWAS are mathematically equivalent to single-marker regression as implemented in the EMMAX software. Obtaining *P*-values requires OPTION snp_p_value in blupf90 and postGSf90. Testing included US Holstein and American Angus data with 10k and 1.4k genotyped bulls, respectively.

Results

Software consolidation. The consolidation was done successfully, and all the tests resulted in identical solutions between blupf90plus and either blupf90 or remlf90 and airemlf90. The same was observed when comparing gibbsf90plus and its component software. Therefore, the consolidated software is ready for official deployment.

Unknown Parent Groups (UPG) in ssGBLUP. Using the altered QP transformation for H^{-1} resulted in greater reliability and lower dispersion bias than the QP for H^{-1} and A^{-1}. Although the three transformations are implemented, we recommend using the altered QP transformation when applying UPG in ssGBLUP. This may hold in most cases; however, testing is suggested.

Metafounders (MF) in ssGBLUP. The application of MF required new software, gammaf90, to estimate the gamma matrix. Gammaf90 is efficient with virtually any number of genotyped animals and MF; however, an accurate estimation required enough genotyped animals linked to MF. Overall, reliabilities or accuracies and dispersion bias were similar between models with the altered QP transformation in UPG and MF.

Optimizations for large-scale ssGBLUP evaluations. Time to compute the coefficients of A_{22} needed for blending in APY drastically reduced with the new algorithm for core and noncore animals separately. With 3.4M genotyped Holsteins, it took 12 minutes, compared to 41 hours with the previous method. The changes for parallel reading in the iod programs significantly reduced the time to solve the MME. Both modifications were necessary steps towards more efficient software for large-scale evaluations.

Efficient algorithm to approximate reliability or accuracy of GEBV. The new algorithm provided higher correlations (~0.98) with reliabilities based on PEV than the default algorithm (~0.90). The increase in computing time was linear with the number of genotyped animals and traits. It took 11 minutes for a three-trait model with 10M animals in the pedigree and 335k genotypes.

***P*-values for SNP effects in single-step GWAS.** We successfully implemented *P*-values for ssGWAS in BLUPF90. As expected, results were numerically identical to single-marker regression (i.e. EMMAX). Therefore, ssGWAS became a general and efficient QTL detection and test approach. This method can be

used in complex datasets such as those used in animal breeding, where only a proportion of pedigreed animals are genotyped.

Discussion

After 25 years, BLUPF90 became a resourceful collection of efficient software that handles ssGBLUP evaluations with millions of genotyped animals. Breed associations and breeding companies worldwide use this software suite for genomic evaluations in nearly all farm animal species. Academia uses extensively public versions for research. The subsequent updates will focus on further improving efficiency and convergence for large datasets and expanding the methods to increase the usability of the BLUPF90 software suite in animal and plant breeding. Constructive feedback is always welcome.

References

Aguilar I., Legarra A., Cardoso F., Masuda Y., Lourenco D. *et al.* (2019) Genet. Sel. Evol. 51:28. https://doi.org/10.1186/s12711-019-0469-3

Bermann M., Lourenco D., Misztal I. (2021) J. Anim. Sci. skab353 https://doi.org/10.1093/jas/skab353

Colleau J.J. (2002) Genet. Sel. Evol. 34:409-421 https://doi.org /10.1186/1297-9686-34-4-409

Liu Z., VanRaden P., Lidauer M., Calus M., Benhajali H. *et al.* (2017) Interbull Bull. 51:75–85.

Masuda Y., Misztal I., Legarra A., Tsuruta S., Lourenco D.A.L. *et al.* (2017) J. Anim. Sci. 95:49-52. https://doi.org/10.2527/jas.2016.0699

Misztal I., Vitezica Z. G., Legarra A., Aguilar I., and Swan A.A. (2013) J. Anim. Breed. Genet. 130:252-258. https://doi.org/10.1111/jbg.12025

Misztal I., Legarra A., and Aguilar I. (2014) J. Dairy Sci. 97:3943-3952. https://doi.org/10.3168/jds.2013-7752.

Quaas R. L. and Pollak E. J. (1981) J. Dairy Sci. 64:1868-1872. https://doi.org/10.3168/jds.S0022-0302(81)82778-6

Tsuruta S., Lourenco D.A.L., Masuda Y., Misztal I., and Lawlor T.J. (2019) J. Dairy Sci. 102:9956-9970. https://doi.org/10.3168/jds.2019-16789

367. From data to genomic breeding values with the MiX99 software suite

T.J. Pitkänen*, H. Gao, A. Kudinov, M. Taskinen, E.A. Mäntysaari, M.H. Lidauer and I. Strandén

Natural Resources Institute Finland (Luke), Myllytie 1, 31600 Jokioinen, Finland; timo.j.pitkanen@luke.fi

Abstract

Efficient computational solutions are a continuing concern for genomic predictions in animal and plant breeding. MiX99 was developed for fast solving of large-scale mixed model equations by preconditioned conjugate gradient with iteration on data. MiX99 offers stand-alone tools for pruning pedigree, computing inbreeding coefficients and allele frequencies, constructing and inverting various covariance matrices, calculating reliabilities, etc. In this paper, we describe in a workflow manner the most important programs and features of MiX99 that can be used for carrying out a single-step genomic best linear unbiased prediction based on the T matrix. The workflow is shown with a real dataset, and for each step we illustrate the applied optimization techniques, present the corresponding tools, and describe options in MiX99. Our example demonstrates also the computational load involved in the various steps.

Introduction

The MiX99 software (www.luke.fi/mix99) was developed for fast solving of large mixed model equations (MME) by a preconditioned conjugate gradient (PCG) iteration on data implementation (Strandén and Lidauer, 1999) and evolved to software capable of solving genomic prediction models with a large number of genotyped animals (Strandén *et al.*, 2018). Numerous utilities have been developed that are linked to the core programs. In general, MiX99 has two core parts: the pre-processor *mix99i* and the solver *mix99s* and its parallel version *mix99p*, enabling the users from pre-processing different types of raw data to obtaining the final estimated breeding values (EBV). In addition, MiX99 offers stand-alone tools for users to conduct related data analysis activities such as pruning pedigree, computing inbreeding coefficients and allele frequencies, constructing and inverting covariance matrices, calculating reliabilities, etc.

For the genomic prediction, MiX99 allows fitting the two basic equivalent models, i.e. genomic best linear unbiased prediction (GBLUP) (VanRaden, 2008) and single nucleotide polymorphism BLUP (SNPBLUP) (Strandén and Garrick, 2009). For single-step approaches, MiX99 can be instructed to perform the regular single-step GBLUP (ssGBLUP) (Legarra *et al.*, 2009, Christensen and Lund, 2010), an algorithm for proven and young (APY) (Misztal *et al.*, 2014), a single-step SNPBLUP (ssSNPBLUP) (Liu *et al.*, 2014), or a ssGBLUP implementation without the need to invert the genomic relationship matrix (G) (ssGTBLUP, Mäntysaari *et al.*, 2017) called ssGTABLUP when residual polygenic part is included. Thus, MiX99 allows the use of various linear MME and datasets for genomic prediction, such as single-step multiple trait random regression test-day models (Koivula *et al.*, 2015).

The objectives of this paper were: (1) to describe the most important programs of the MiX99 suite in a chronological workflow order applied for a particular ssGTABLUP evaluation; and (2) to demonstrate some new features of MiX99 within the process.

From raw data to solutions

To illustrate the workflow of running a ssGTABLUP model, a Red Dairy Cattle dataset from Nordic Cattle Genetic Evaluation (NAV) was employed and analyzed. Data comprised of 26.7 million records with 305-day milk, protein, and fat yields from first, second, and third lactations of 4.3 million cows. Pedigree included 5.9 million animals of which 169,779 were genotyped. For each work step, we described the

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_367, © T.J. Pitkänen *et al.* 2022

applied optimization techniques along with the corresponding tools we used in MiX99. In Table 1 we present the summary of the wall-clock time of each computation step. In the example, the pedigree-based relationship matrix A_{22}^{-1} for the genotyped animals was computed to get its diagonals for the preconditioner information. The A_{22} itself was not formed, nor the full inverse stored. Alternatively, A_{22}^{-1} diagonals can be attained faster using Monte-Carlo methods directly on pedigree (see option 2 below).

Estimation allele frequencies of the base population(s) using Bpop. In regular GBLUP, allele frequencies (AF) from an arbitrary base population have been recommended to be used in constructing the G matrix (VanRaden, 2008). The AF can be estimated for multiple base populations using the generalized least squares (GLS) method proposed by (McPeek *et al.*, 2004) and implemented in the *Bpop* program (Strandén *et al.*, 2017, Strandén and Mäntysaari, 2020). For a regular single-step, AF in chosen (earliest) base population can be used in constructing the G matrix. In our example here, however, we used multiple base populations as unknown parent groups (UPG) which were replaced by metafounders (MF, Legarra *et al.*, 2015).

Align G and A matrices and construct the gamma matrix (Γ). The compatibility of G and the numerator relationship (A) matrices is important in single-step methods. One commonly adopted approach is to scale the diagonal and off-diagonal elements of G to have the same average values as those of A (Christensen, 2012). The genomic information processing program *hgTinv* can be used to build various G matrices such as VanRaden 1 and 2 (VanRaden, 2008). Using MF, the UPG are replaced by related and inbred MF individuals. The MF will adjust the A and its part that includes the genotyped animals (A_{22}) matrices to be consistent with G where all AF are equal to 0.5. A covariance matrix of Γ needs to be built to reflect the 'relationships' among MF. The base population AF estimated with *Bpop* can be used in building the Γ matrix as described e.g. in Garcia-Baccino *et al.* (2017).

Compute inbreeding coefficients with RelaX2. Similar to the ordinary A matrix, in the MF approach the inverse of A augmented by the Γ matrix (A^{Γ}) can be constructed when the corresponding inbreeding coefficients are known. Given the input of the original pedigree and the Γ matrix, inbreeding coefficients can be calculated for all animals using the *RelaX2* program (Strandén and Vuori, 2006). It will also provide a new pedigree file with MF.

Predict genomic estimated breeding values (GEBV) with MiX99. Like single-step models with UPG, the MF model can be solved using ssGBLUP (with G_{APY} inverse if requested) or ssGTBLUP. The MiX99 software has two options to compute GEBV in ssGTBLUP:

Table 1. Runtime and CPU use of each program when solving a multiple trait single-step model with 5.9 million animals of which 169,779 were genotyped.

Step	Program	CPU[1] cores	Runtime (hrs:min)	Step	Program	CPU cores	Runtime (hrs:min)
1	*RelaX2*	1	00:21	4	*mix99i*	4	00:10
2	*Bpop*	4	00:09	5	*mix99s*	4	11:09
3	*hgTinv*	4	22:52	6	*apax99* (AM)	4	00:04
	3a) invert A_{22}		16:10	7	*apax99* (20p)	1	00:01
	3b) build T		06:43	8	*snp_blup_rel*	1	00:53[2]

[1] All programs executed in an Intel Xeon 2.8 GHz machine.

[2] Time for one trait.

- **Option 1: Compute matrix T and diagonal of A_{22}^{Γ} matrices with program hgTinv.** Given the input of pedigree, inbreeding coefficients, genotypes, and Γ^{-1}, the *hgTinv* program outputs T, as well as diagonal elements A_{22}^{Γ} (if requested). Once Γ^{-1} and inbreeding are given the rules of Henderson (1976) are used to compute the inverse of numerator relationship matrix for all the animals (A^{Γ}) and for the genotyped animals A_{22}^{Γ}. The advance of ssGTBLUP is that there is no need to compute G^{-1}, but instead an m×n T matrix, where m is the number of markers and n is the number of genotyped animals. If required, *hgTinv* can also generate G^{-1}, $\mathbf{G}_{APY}^{-1}$ *and* $\mathbf{H}_{Liu}^{22}$ from (Liu *et al.*, 2014).
  Compute GEBVs with mix99i and mix99s. To solve the MME and obtain the GEBVs, the files of pedigree, phenotypes, inbreeding coefficients, variance components, and T, are provided to *mix99i*. In the optimal use of ssGTBLUP, the *mix99s* or *mix99p* is asked to perform the computations involving the A_{22}^{Γ}, i.e. the inverse of relationship matrix of genotyped animals through using the pedigree file only and the metafounder information in Γ. While doing this, the iteration convergence can be improved by providing a file with the diagonals of A_{22}^{Γ} for the preconditioner. If high performance is required, this diagonal can be computed by program *iA22diag_make* using Monte Carlo sampling. With the MF approach, the file having the Γ^{-1} matrix must be supplied to *mix99i* using 'iGammaFILE' command. The *mix99i* (pre-processor) and *mix99s* or *mix99p* (solver) need to run sequentially.
- **Option 2: Compute Z and C^{-1} matrices with program T48eig_make.** Given the input of pedigree, genotypes, and Γ^{-1}, the *T48eig_make* program can output the centred marker covariate matrix Z and the C^{-1} matrix, where $\mathbf{C} = \mathbf{Z}'\mathbf{A}_{22}^{-1}\mathbf{Z}/\mathrm{w} + \mathbf{I}\mathrm{d}$, in which d = k/(1-w), w is the proportion of residual polygenic effect and $k = \sum_{l=1}^{m} 2p_l(1-p_l)$ with p_l the AF.
  Compute diagonal of A_{22}^{Γ} with program iA22diag_make. Given the input of pedigree and genotyped IDs, the *iA22diag_make* program outputs diagonal elements of A_{22}^{Γ}.
  Compute GEBVs with mix99i and mix99s. Given the input files of pedigree, phenotypes, inbreeding coefficients, variance components, C^{-1}, Z, Γ^{-1}, and $diag(\mathbf{A}_{22}^{\Gamma})$, similar as in Option 1, GEBVs can be achieved. In this option some of the computations to make the T matrix in option 1 are done within each PCG iteration in *mix99s*. This tends to reduce the total computing time when the model includes more SNPs than the expected number of iterations. Moreover, this approach directly produces the solutions for SNP effects from *mix99s*.

Calculate EBV reliabilities for all animals with mix99i and apax99. Calculation of reliabilities of EBV requires the prediction error variances (PEV). This can be done using exa99. However, for large systems, it is impossible to invert the coefficient matrix of the MME and approximation approaches have to be used. The *apax99* program supports different approaches to approximate EBV reliabilities and provides basic approximation options for GEBV reliabilities. For the given example, we used the method of Jamrozik *et al.* (2000).

Calculate GEBV reliabilities for genotyped animals with snp_blup_rel. In dairy cattle breeding, effective daughter contributions (EDC) or effective record contributions (ERC) has long been adopted as the input to reliability estimation. For single-step models, the ERC for genotyped animals can be computed by *apax99* and thereby be used in the calculation of GEBV reliabilities from SNPBLUP. The a*pax99* program offers options for so-called reverse reliability estimation that give reliability equivalent ERC (Ben Zaabza *et al.*, 2022). Then, given the ERC as weights in SNPBLUP, program *snp_blup_rel* can be used to calculate the GEBV reliabilities for the genotyped animals (Ben Zaabza *et al.*, 2020). EBV reliabilities for the non-genotyped animals can be obtained by returning the genomic added ERC to *apax99* (Ben Zaabza *et al.*, 2022).

Acknowledgements

Nordic Cattle Genetic Evaluation (Aarhus, Denmark) is acknowledged for providing the data.

References

Aguilar I., Misztal I., Johnson D.L., Legarra A., Tsuruta S., *et al.* (2010) J. Dairy Sci. 93(2):743–752. https://doi.org/10.3168/jds.2009-2730

Ben Zaabza H., Mäntysaari E.A., and Strandén I. (2020) Agr. Food Sci. 29:297-306. https://doi.org/10.23986/afsci.95617

Ben Zaabza H., Taskinen M., Mäntysaari E.A., Pitkänen T., Aamand G.P. *et al.* (2022) J. Dairy Sci. In press.

Christensen O. F. (2012). Genet. Sel. Evol. 44(1):37. https://doi.org/10.1186/1297-9686-44-37

Christensen O. F. and Lund M.S. (2010). Genet. Sel. Evol. 42:2. https://doi.org/10.1186/1297-9686-42-2

Garcia-Baccino C. A., Legarra A., Christensen O.F., Misztal I., Pocrnic I., *et al.* (2017). Genet. Sel. Evol. 49:34. https://doi.org/10.1186/s12711-017-0309-2

Henderson C. R. (1976). Biometrics:69-83.

Jamrozik J., Schaeffer L. and Jansen G. (2000). Livest. Prod. Sci. 66(1):85-92. https://doi.org/10.1016/S0301-6226(00)00158-5

Koivula M., Stranden I., Pösö J., Aamand G.P. and Mäntysaari E.A. (2015). J. Dairy Sci. 98(4):2775-2784. https://doi.org/10.3168/jds.2014-8975

Legarra A., Christensen O.F., Vitezica Z.G., Aguilar I. and Misztal I. (2015). Genetics 200(2):455-468. https://doi.org/10.1534/genetics.115.177014

Liu Z., Goddard M.E., Reinhardt F. and Reents R. (2014). J. Dairy Sci. 97(9):5833-5850. https://doi.org/10.3168/jds.2014-7924

Mäntysaari E.A., Evans R.D. and Strandén I. (2017). J. Anim. Sci. 95(11):4728-4737. https://doi.org/10.2527/jas2017.1912

McPeek M.S., Wu X. and Ober C. (2004). Biometrics 60(2):359-367. https://doi.org/10.1111/j.0006-341X.2004.00180.x

Meuwissen T.H., Hayes B.J. and Goddard M.E. (2001). Genetics 157(4):1819-1829. https://doi.org/10.1093/genetics/157.4.1819

Misztal I., Legarra A. and Aguilar I. (2014). J. Dairy Sci. 97(6):3943-3952. https://doi.org/10.3168/jds.2013-7752

Strandén I. and Garrick D.J. (2009). J. Dairy Sci 92(6):2971-2975. https://doi.org/10.3168/jds.2008-1929

Strandén I. and Lidauer M. (1999). J. Dairy Sci. 82(12):2779-2787. https://doi.org/10.3168/jds.S0022-0302(99)75535-9

Strandén I. and Mäntysaari E.A. (2020). Agric. Food Sci. 29(3):166–176. https://doi.org/10.23986/afsci.90955

Part 20

Statistical genetics: GWAS

368. Haplotypes affecting pulmonary arterial pressure in Angus cattle

D. Lu[1*], A. Garcia[1], S. Miller[2] and K. Retallick[1]

[1]Angus Genetics Inc., American Angus Association, 3102 Frederick Ave., Saint Joseph, MO 64506, USA;
[2]Animal Genetics & Breeding Unit, University of New England, Armidale NSW 2351, Australia; dlu@angus.org

Abstract

Cattle that have high pulmonary arterial pressure (PAP) and live at high elevations are at risk of high mountain disease, which causes economic loss to Angus breeders. A 2.3 kb haplotype, HapA hereafter, with frequency of 26.99% in the genotyped Angus cattle, was identified at chr7: 90824804- 90827178 and proven to be associated with lower PAP measures at weaning and yearling. Average PAP scores among HapA heterozygotes, HapA homozygotes, and alternate homozygotes were 41.97±7.52, 40.66±6.56, and 44.11±9.25, respectively. Carrying 1 copy and 2 copies of HapA helped lower PAP score by 1.9120±0.1938 (P<0.001) and 2.6620±0.3193 (P<0.001) compared to non-carriers, respectively. While efforts to pin down causal variants for PAP are underway, HapA might be useful for improving the accuracy of PAP breeding value estimates.

Introduction

Pulmonary hypertension, measured by mean pulmonary arterial pressure (PAP), is defined as high blood pressure in the lungs and the right side of the heart, which in cattle is often associated with elevations above 1,524 meters above sea level and indicative of high mountain disease (HMD) (Holt and Callan, 2007). Holt and Callan (2007) estimated that cattle losses due to HMD alone can be averaged around 3-5% of the calf crop, resulting in a financial loss of about US$60 million (McCormick 2011). In May 2020, the American Angus Association and Angus Genetics Inc. (AGI) officially released expected progeny differences (EPD) for PAP in efforts to help Angus breeders make better selection decisions with regard to the trait. With the rapid advancement of DNA and single nucleotide polymorphism (SNP) chip technology, researchers at AGI have been able to explore various approaches to improve the accuracy of PAP prediction, and herein reported a new discovery in our efforts to estimate breeding value for PAP. The objective of this exercise was to estimate effects of haplotypes potentially impacting PAP score in Angus beef cattle.

Materials & methods

Genotypes and imputation. This analysis involved 982,769 animals genotyped with various SNP arrays, namely i50K, HD50K, GGPLD, GGPHD, GGPUHD, AngusGS (hereafter GSv.1 and GSv.2). SNP in GSv.1 fully overlapped with part of GSv.2, thus all animals genotyped with GS.v1 were imputed to GSv.2 for all autosomal SNP. There were various versions of HD50K, of which ZBUv.2 was used as reference panel to impute those genotyped with i50K and other HD50K versions to ZBUv.2 autosomal SNP. SNP in GSv.2 and ZBUv.2 were combined to create a set of 85,134 SNP on 29 autosomes (hereafter 85K). Animals with original and imputed GSv.2, as well as animals with original and imputed ZBUv.2, were imputed to the 85K, which were then used as reference animals to impute GGPLD, GGPHD, GGPUHD animals up to the 85K. Imputation was performed using FImpute v.3 (Sargolzaei *et al.*, 2014). Imputation accuracy in terms of genotype concordance was 99.22-99.85%.

Animals and phenotypes. There were 19,175 animals measured for PAP born between 1993 and 2020, with 59.42% of them born between 2015 and 2020. Pulmonary arterial pressure was measured as described by Holt and Callan (2007). Tests for PAP were categorized into two levels of elevation and different age groups as described in Table 1. Those PAP measured at yearling and at moderate elevation were referred to as moderate PAP (mPAP hereafter). Similarly, PAP measured at high elevation when animals were

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_368, © D. Lu *et al.* 2022

Table 1. Description of PAP data.[1]

Elevation	n	AOT	PAP	h^2
Moderate (1,219-1,676 m) – mPAP	5,984	440.52±98.17	42.76±8.35	0.29
High (1,676 m and above) – wPAP	4,572	227.70±39.67	41.45±7.77	0.27
High (1,676 m and above) – hPAP	10,032	424.48±85.62	42.87±7.78	0.40

[1] n = number of records; AOT = age on test; PAP = pulmonary arterial pressure; h^2 = heritability.

424.48±85.62 and 227.70±39.67 days old were referred to as high PAP (hPAP hereafter) and wean PAP (wPAP hereafter), respectively. There were 1,413 animals tested for both wPAP and hPAP. Genetic parameters for PAP were estimated using airemlf90 (Misztal *et al.*, 2014). There were 8,497 animals with both genotypes and PAP measures, including 1,686, 4,032, and 3,266 animals with wPAP, mPAP, and hPAP, respectively. Of the 1,686 animals tested for wPAP, 487 of them were also tested for hPAP.

Methodology. Genomic regions of interest were identified by sliding a six-SNP window across 29 autosomes one SNP at a time, with SNP effect being the average of the window. The process started with estimating genomic breeding values (GEBV) with single-step genomic BLUP (ssGBLUP). The following statistical model was used y = Xb + Zu + e, where y was the vector of phenotypes recorded for mPAP, wPAP, and hPAP; b was the vector of fixed effects of contemporary group, and both the age of the animals, and the elevation fitted as covariates; u was the vector of additive genetic effects and e was the vector of residuals. The matrices X and Z were the incidence matrices for the effects in b and u respectively.

The genomic relationship matrix (G) was constructed as in VanRaden (2008) using allele frequencies from the current genotyped population. To avoid singularity, G was blended with 10% of the pedigree relationship matrix for genotyped animals (A_{22}). Because of the large number of genotyped animals, the algorithm for proven and young (APY) (Misztal *et al.*, 2014) was used to invert G. The number of core animals was based on the number of eigenvalues explaining 99% of the variation in G and it was the same core set as used in the official Angus weekly evaluation. Once GEBV were available, SNP effects were back-solved as described by Wang *et al.* (2012) and used for subsequent analyses. Calculations of GEBV and SNP effects were done using software from the Blupf90 family of programs (Misztal *et al.*, 2014).

A seven-SNP region was selected out of the postGS analysis based on their estimated SNP effect and the extent of linkage disequilibrium (LD) in the region. Effects of all possible haplotypes in the region, including haplotypes of seven, six, five, four, or three consecutive SNP, were estimated in univariate models similar to Equation 1. However, y was now the vector of either mPAP, or wPAP, or hPAP. Fixed effects were extended to include the effect of haplotype coded as categorical zero copy, one copy, or two copies of each haplotype investigated. One haplotype was investigated at a time. The relationship matrix G was replaced with the additive numerator relationship derived from pedigree. Because several haplotypes were tested, running the models through ASReml 4.1 (Gilmour *et al.*, 2015) helped expedite the analyses. Haplotype with the most significant impact on the wPAP, mPAP, and hPAP, as well as displaying the strongest LD among individual SNP, was chosen as a candidate haplotype.

Another statistical model was tested, where wPAP, mPAP, and hPAP were combined into only one trait namely PAP, thus one individual could have more than one PAP measures. The fixed effects included contemporary group, sex, age at test, elevation, and haplotype. Random effects included animal effect, permanent environmental effect to capture variance due to repeated records where animals were tested at both weaning and yearling, and error term. This model was run with ASReml 4.1 (Gilmour *et al.*, 2015).

Results & discussion

Genetic correlations between mPAP and wPAP, mPAP and hPAP, wPAP and hPAP were 0.83, 0.92, and 0.98, respectively. Results from the postGS suggested a 7-SNP segment, chr7: 90689839-90827178, be considered for haplotype analysis. Mean pairwise LD (r^2) in this segment was 0.91±0.08. After testing the effects of all possible haplotypes in the region, the focus has landed on a three SNP segment, chr7: 90824804- 90827178, within the 7 SNP region. Linkage disequilibrium in this three SNP segment was nearly 1.0. There were two main haplotypes in this region, namely HapA and HapB with frequency of 26.99 and 72.99%, respectively, out of 982,769 genotyped animals. Their frequency among the genotyped and PAP tested animals was 32.49 and 67.50%, respectively. HapA with a lower frequency in the population was used as reference haplotype in subsequent analyses. Table 2 presents average PAP score for animals carrying zero copy, one copy, or two copies of HapA. PAP score at weaning was lower than the score at yearling. Animals that carried one copy or two copies of HapA had PAP score consistently lower than those not carrying the haplotype. Genetic correlations of 0.83-0.98 between pairs of PAP traits suggested that they could be collapsed into one trait, whose mean among animals not carrying HapA was 44.11±9.25, decreasing to 41.97±7.52 and 40.66±6.56 in those carrying one copy and two copies of HapA, respectively. According to Holt and Callan (2007), cattle older than 12 months, living at elevation above 1,524 m, are considered normal if PAP score is below 41, however at risk of HMD if PAP score above 49. In the current study, there were 4,578 animals older than 12 months and being raised at altitude above 1,524 m. Of them, 2004, 2020, and 554 animals carried zero copy, one copy, and two copies of HapA, respectively. Among the non-carriers, 418 (20.86%) had PAP above 49. This percentage in the one copy and two copy carriers was 13.02 and 10.83%, respectively. If PAP score greater than 49 is indicative of risk of HMD, non-carriers of HapA were two times more likely to be at risk than HapA homozygotes.

Estimated effects of HapA are presented in Table 3. Carrying HapA significantly decreased PAP score compared to non-carriers. The effect of HapA appeared to be similar among the mPAP and hPAP animals, though slightly smaller in the wPAP animals. Contrasts of PAP score between HapA homozygotes and HapA heterozygotes were also tested and found to be not statistically significant among the wPAP and hPAP groups of animals; However, they were significant among the mPAP animals, as well as when the wPAP, mPAP, and hPAP were collapsed into one trait. In the current study, PAP score spread from a low of 25 to a high of 70, and if a score of 49 is the threshold to determine whether an animal is at risk or not at elevation above 1,524 m, a drop by approximately 2 points might be beneficial to the animal and thus the owner. Specifically, there were 741 individuals being raised at elevation above 1,524 m, older than 12 months of age, and PAP score greater than 49 in this study, and thus were at risk of HMD according to Holt and Callan (2007). Had the non-carriers of HapA carried one copy of the haplotype, the number of animals at risk of HMD would have dropped to 678, a save of 8.5% of the animals.

HapA can be used as a tool to screen animals potentially at risk of HMD. Alternately, SNP surrounding this haplotype region, or the haplotypes could be modelled appropriately when performing the genetic evaluation for PAP. Researchers at AGI are working to find causal variants for PAP specifically in the genomic region identified in this study. Once identified, causal variants can be put on Angus SNP chips for

Table 2. Summary of PAP score in the research population (mean ± SD).

Number of copies of HapA	Wean PAP	Moderate PAP	High PAP	Overall
0	42.08±8.21	44.45±9.55	44.73±9.20	44.11±9.25
1	40.25±6.43	42.16±7.76	42.62±7.62	41.97±7.52
2	38.65±4.38	40.89±7.19	41.39±6.61	40.66±6.56

Table 3. Estimated effects of one or two copies of HapA (Effect ± SE).[1]

HapA	Wean PAP	Moderate PAP	High PAP	Overall
0	0	0	0	0
1	-1.35±0.37*	-2.01±0.29*	-1.89±0.30*	-1.91±0.19*
2	-2.42±0.58*	-2.59±0.50*	-2.75±0.46*	-2.66±0.31*

[1] * = $P<0.001$; HapA 0, 1, 2 – Carrying 0, 1, 2 copies of HapA, respectively.

estimating breeding value for PAP. Yet, the haplotypes need to be validated. More animals need to be PAP tested and genotyped.

Conclusions

Haplotypes affecting PAP score in Angus beef cattle were discovered and found to have significant impact on the research animals. HapA has potential to help increase the accuracy of PAP EPD as a selection tool for Angus breeders, especially those who raise their cattle at high elevations.

References

Gilmour A., Gogel J., Cullis R., Welham J., and Thompson R. (2015). ASReml

Holt T. and Callan R. (2007). Vet Clin Food Anim 23:575-596. https://doi.org/10.1016/j.cvfa.2007.08.001

McCormick R. (2011) Brisket disease in cattle. Available at: http://www.uwyo.edu/uwexpstn/publications/impact-statements/2011/2011-6-richardmccormick.pdf.

Misztal I., Legarra A., Aguilar I. (2014) J Dairy Sci. 97(6):3943-52. https://doi.org/10.3168/jds.2013-7752.

Misztal I., Tsuruta S., Lourenco D., Masuda Y., Aguilar I. *et al.* (2014) Manual for BLUPF90 family of programs. Available at: http://nce.ads.uga.edu/wiki/lib/exe/fetch.php?media=blupf90_all2.pdf

Sargolzaei M., Chesnais J., and Schenkel F. (2014) BMC Genomics, 15:478. https://doi.org/10.1186/1471-2164-15-478.

User Guide Release 4.1 Functional Specification. VSN International Ltd., Hemel Hempstead, UK.

Wang H., Misztal I., Aguilar I., Legarra A, Muir W.M. (2012) Genetics Research 94(2):73-83. https://doi.org/10.1017/S0016672312000274.

369. Detection of shared genomic regions between production and health traits in German Holstein cattle

H. Schneider[1*], D. Segelke[2], J. Tetens[3], G. Thaller[4] and J. Bennewitz[1]

[1]Institute of Animal Science, University of Hohenheim, 70599 Stuttgart, Germany; [2]Vereinigte Informationssysteme Tierhaltung w.V. (VIT), 27283 Verden, Germany; [3]Department of Animal Sciences, University of Göttingen, 37077 Göttingen, Germany; [4]Institute of Animal Breeding and Husbandry, Christian-Albrechts University of Kiel, 24098 Kiel, Germany; helen.schneider@uni-hohenheim.de

Abstract

Global genetic correlations as common estimates of pleiotropy across the genome fail to dissect the absolute level of shared genomic regions and their effects on various traits. Methods in human genetics that try to transfer global estimates to the local scale showed up recently but have not been applied to livestock yet. One of them is LOGODetect, scanning summary statistics for shared genomic regions between two traits. We applied LOGODetect on a sample of German Holstein cattle to identify these regions for milk production, mastitis and claw disease traits and found a certain number of segments with concordant effects and enriched findings on chromosome 14 in the area of the *DGAT1* gene.

Introduction

Dissection of genetic correlation from GWAS summary statistics (Bulik-Sullivan *et al.* 2015) had become a common procedure to quantify the level of shared genetic influence on several traits to quantify pleiotropy (Guo *et al.* 2021). Nevertheless, it does not only fail to unveil the underlying causal mechanisms and exact position of pleiotropic variants in the genome (Visscher und Yang 2016). Moreover, global genetic correlations might also be veiled in their overall dimension as some local correlations can appear in antagonistic directions and global genetic correlations give average estimates across the genome, caching strong local correlations in the absence of any global correlation (Shi *et al.* 2017, van Rheenen *et al.* 2019). SNP-based estimation of global genetic correlation suffers from substantial LD between markers, because non-causal mutations might be raised in their signal even though their trait specific effect is merely LD based (van Rheenen *et al.* 2019). Therefore, common global estimates of genetic correlation cannot be directly transferred to the local scale (Werme *et al.* 2021) and different methods have been developed to deal with this issue. One of them is LOGODetect, a tool to detect shared genomic segments between various traits, that is detached from prior assumptions about the localization of these segments in the genome (Guo *et al.* 2021). So far, this has only been applied in human genetics.

In dairy cattle breeding, several novel health traits are recorded and entered the total merit index in many countries. An in-depth analysis of the genetic correlation might shed some light on the role of pleiotropy for these novel traits and existing production traits. This is peculiar not only because pleiotropy had been detected on a widespread level also in cattle (Xiang *et al.* 2017), but also because of economical, ecological and ethical concerns about livestock production. Therefore, the aim of the present study was to dissect local genetic correlations between milk productions, mastitis and claw diseases applying LOGODetect on a large sample of German Holstein cows.

Materials & methods

We analysed genotypes and phenotypes of 34,497 primiparous Holstein cows with lactation records of milk and health traits between 2015 and 2020, both provided by the national computing centre VIT (Vereinigte Informationssysteme Tierhaltung w.V., Verden). Health traits consisted of mastitis (MAS), claw ulcers (CCU), white line disease (CWL), laminitis (CLA), interdigital hyperplasia (CIH) and digital phlegmon

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_369, © H. Schneider *et al.* 2022

(CPH). With the formula of Nostitz and Mielke (1995) milk energy yield (MEY) was calculated to obtain a measure of the cow's energetic output via milk production. The cow's farm, year and season of lactation were combined to heard-year-season (hys), where each season contained three months according to the calendric partitioning of seasons. Cows were mostly genotyped with the Illumina Bovine Eurogenomics 10 K chip and then imputed to the Illumina Bovine 50 K SNP BeadChip V2 (Illumina Inc., San Diego, CA) according to the procedure applied in the national genetic evaluation (Segelke *et al.* 2012). 44,126 SNPs remained after filtering out SNPs on gonosomes and those with a MAF below 0.01. We performed a GWAS for all traits with the following mixed linear model that is implemented in GCTA (Yang *et al.* 2011), applying the 'leave-one-chromosome-out' (LOCO) approach according to Yang *et al.* (2014) as

$$y = \mu\,\mathrm{Xb} + g + e$$

Here, y is the vector of raw phenotypes that were standardized for the gaussian traits to facilitate the interpretation of SNP effects but remained coded as 0/1 for the disease traits. The vector b contains the fixed effects including the SNPs effect, the hys, age at first calving or days in milk. X as the incidence matrix for b inherits the SNPs amount of allele copies and further links the trait records to the covariate's effects. g is the polygenic term, accounting for relatedness in the sample with variance $\boldsymbol{g} \sim N(0,\mathbf{G}\sigma_g^2)$ implementing the LOCO-adjusted GRM G and e the residual term with variance $\mathbf{e} \sim N(0,\mathbf{I}\sigma_e^2)$. To decide whether the covariates age at first calving and days in milk had to be included for each trait we performed an ANOVA with the R package car (Fox 2019) and declared a covariate's effect to be significant if it's P-value was below 0.05. Additionally, 20 principal components computed with GCTA (Yang *et al.* 2011) were included as covariates in two different ways, either on a genomewide or chromosomal level. The latter involved only the principal components for the chromosome on which the candidate SNP was located. For subsequent usage, we chose the method where the inflation factor λ_{GC} was closest to unity. A threshold of 0.0000011 (corresponds to a genome wide significance) and 0.0005 (nominal significance) was used to declare a SNP as trait-associated. The GWAS summary data were used as input-data to detect local genetic correlations with LOGODetect, whose application can be taken from Guo *et al.* (2021), following the tutorial on Github (https://github.com/ghm17/LOGODetect). The key point and central output of this method is the scan statistic $Q(R)$. It consists of the SNP effects for both traits within a specified region, multiplied with each other and divided through the sum of LD scores for each of these SNPs. Negative values indicate regions with negative local genetic correlations and vice versa. As it can be seen from this calculation, LOGODetect doesn't give absolute estimates of genetic correlation but identifies regions with significant shared effects on multiple traits (Guo *et al.* 2021).

Results

The number of identified regions between MEY and a health trait varied between two and 20 was highest between MEY and CPH (20). These regions harboured between five (CLA) and 25 (CPH) SNPs, which did not necessarily reach significance in the single trait's GWAS. Two regions were detected in each case for CIH and CCU and four regions for CLA, CWL and MAS respectively. The average segment length was 1,027,368 bp. Regarding the effect directions, it has to be mentioned, that only seven of all 36 identified regions showed a negative sign. Five of them where for CPH and one for CWL and CCU, respectively. $Q(R)$ varied between -5.03 (MEY-CWL) and 14.31 (MEY-CIH), showing that the positive correlation of rising MEY and rising disease cases was about three times higher than the preventive effect of rising MEY on health traits. Figure 1 shows the distribution of concordant effects across chromosomes. Most chromosomes harbour either negative or positive effects except of chromosome 13 and 22, where bi-directional effects on various traits were detected. Located on chromosomes 7, 13 and 14 are regions that show concordant effects on MEY and multiple disease traits. What sticks out is chromosome 14, where positive effects of MEY and all disease traits except of CCU are around the well-known *DGAT1* mutation.

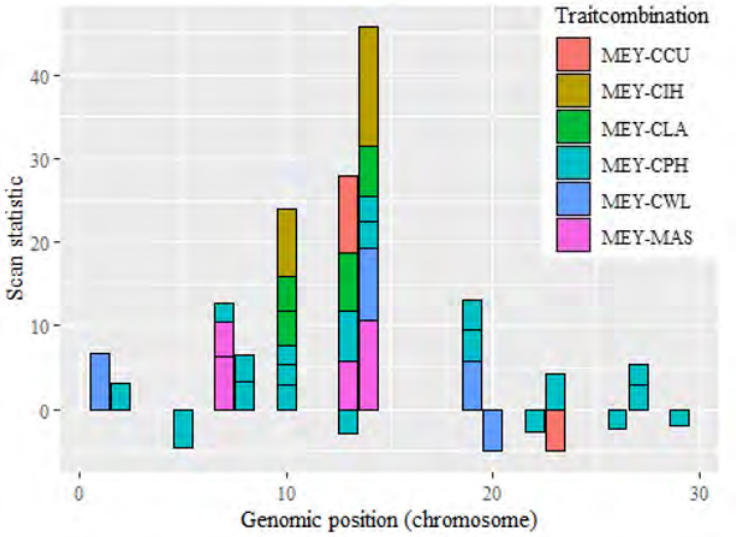


Figure 1. Distribution of genomic regions, which were identified with LOGODetect. Shown is their genomic position and scan statistic.

Discussion

By applying LOGODetect to different trait combinations it was possible to identify segments on the genome with significantly correlated effects on health traits and milk production in German Holstein. These findings are supported by other results that have for example shown pleiotropic effects of *DGAT1* not only on milk production but also on fertility (Oikonomou *et al.* 2009). Yet, the approach is only applicable to two traits simultaneously, making the analysis of multiple trait combinations rather protracted. Nevertheless, because of LD as a weighting factor for the marker effects on both traits, LOGODetect avoids unjustified preference of a sequence that is merely because of linkage in this region and outperforms other tools due to its unbiased and accurate detection of segments inheriting a larger proportion of genetic covariance (Guo *et al.* 2021). Moreover, unlike in other tools that are single SNP based there is no limitation to distinguish between the true SNPs signal and the signal of any neighbouring SNPs (Werme *et al.* 2021). Originally LOGODetect had been developed in human genetics, exhibiting substantially lower LD than livestock. Probably, that's why the choice of the weighting factor might need to be adapted in livestock. Suggestions by Guo *et al.* (2021) were applied here as their optimal choice is computationally expensive. Next to that, LOGODetect compares the null distribution of SNP effects under the assumption of no correlation applying a Monte Carlo approach with currently no existing computationally more efficient method to derive this distribution (Guo *et al.* 2021). What has to be considered is that the present study is limited by the number of markers, as most common variants explain only a small proportion of heritability (Visscher and Yang 2016). In the original methods, the authors suggested a splitting of the genome into independent LD blocks to receive more accuracy in the detection of significant regions (Guo *et al.* 2021). The limited number of markers hindered us to do so as each block would either have consisted of only few markers or would have covered a large part of the genome and neither of these opportunities fits the splitting's intention.

Acknowledgements

This study is part of the project 'QTCC: From Quantitative Trait Correlation to Causation in dairy cattle' and was supported by the German Research Foundation, DFG.

References

Bulik-Sullivan B., Finucane H., Anttila, V. *et al.* (2015) Nat Genet 47, 1236–1241. https://doi.org/10.1038/ng.3406

Fox J, Weisberg S. (2019): An R Companion to Applied Regression. Sage, Thousand Oaks CA. Available at https://socialsciences.mcmaster.ca/jfox/Books/Companion/.

Guo H., Li J. J., Lu Q.; Hou L. (2021) Nat Commun https://doi.org/10.1038/s41467-021-22334-6

Nostitz B. and Mielke H. (1995) J Anim Physiol An N 73 (1-5), S. 9–18. https://doi.org/10.1111/j.1439-0396.1995.tb00398.x

Oikonomou G., Angelopoulou K., Arsenos G., Zygoyiannis D., Banos G. (2009) Anim Genet 40 (1), S. 10–17. https://doi.org/10.1111/j.1365-2052.2008.01789.x

Segelke D., Chen J., Liu Z., Reinhardt F., Thaller G., Reents R. (2012) J Dairy Sci 95 (9), S. 5403–5411. https://doi.org/10.3168/jds.2012-5466

Shi H., Mancuso N., Spendlove S., Pasaniuc B. (2017) Am J Hum Gen 101 (5), S. 737–751. https://doi.org/10.1016/j.ajhg.2017.09.022

van Rheenen W., Peyrot W. J., Schork A. J., Lee S. H.; Wray N. R. (2019) Nat Rev Genet 20 (10), S. 567–581. https://doi.org/10.1038/s41576-019-0137-z

Visscher P. and Yang J. A plethora of pleiotropy across complex traits. Nat Genet 48, 707–708 (2016). https://doi.org/10.1038/ng.3604

Werme J., van der Sluis S., Posthuma D., Leeuw C. A. de (2021) in press https://doi.org/10.1101/2020.12.31.424652

Wu X-L, Heringstad B., Chang Y-M., Los Campos G. de, Gianola D. (2007) J Dairy Sci 90 (7), S. 3508–3521. https://doi.org/10.3168/jds.2006-762

Xiang R., MacLeod I. M., Bolormaa S., Goddard M. E. (2017) Sci Rep-UK 7 (1), S. 9248. https://doi.org/10.1038/s41598-017-09788-9

Yang J., Lee S. H., Goddard M. E., Visscher P. M. (2011) Am J Hum Gen 88 (1), S. 76–82. https://doi.org/10.1016/j.ajhg.2010.11.011

Yang J., Zaitlen N. A., Goddard M. E., Visscher P. M., Price A. L. (2014) Nat Genet 46 (2), S. 100–106. https://doi.org/10.1038/ng.2876

370. Detecting the presence of genomic imprinting for carcass traits in cattle using imputed high-density genotypes

D. Kenny[1], R.D. Sleator[2], C.P. Murphy[2], R.D. Evans[3] and D.P. Berry[1*]

[1]Teagasc, Moorepark, Fermoy, Co. Cork, P61 C996, Ireland; [2]Munster Technological University, Bishopstown, Co. Cork, T12 P928, Ireland; [3]Irish Cattle Breeding Federation, Highfield House, Bandon, Co. Cork, P72 X050, Ireland; donagh.berry@teagasc.ie

Abstract

Genomic imprinting is defined as the silencing of an allele at a given locus, at least partially, based on the sex of the transmitting parent. The objective of the present study was to detect single nucleotide polymorphism (SNP)-phenotype imprinting associations for carcass traits using imputed, high-density genotype data from 23,687 cattle. In total, 51, 226 and 192 SNPs located on six different chromosomes had significant imprinting associations with carcass weight, carcass conformation and carcass fat, respectively. Furthermore, an additional 36, 130 and 100 SNPs demonstrated suggestive imprinting associations with carcass weight, carcass conformation and carcass fat, respectively. Of the SNPs detected herein to have imprinting associations, no SNP was located within 0.5 Mb of a known imprinted gene.

Introduction

Traditional approaches to genetic evaluations assumes that inter-animal genetic and performance differences are due to differences in the DNA sequence of the animals' genomes. The epigenome, which manifests itself as changes in an animal's gene expression and phenotypic performance without an alteration in the underlying DNA sequence, deviates from this assumption (González-Recio *et al.* 2015). Genomic imprinting is a epigenetic phenomenon defined as the differing expression of an allele depending on the sex of the transmitting parent (Reik and Walter 2001). Knowledge of the presence of genomic imprinting has previously been shown to be beneficial in mate allocation programs (González-Recio *et al.* 2015), with the consideration of imprinting effects in genomic evaluations been reported to improve the accuracy of additive genomic breeding values (Nishio and Satoh 2015). The objective of the present study was to detect the presence of genomic imprinting for carcass weight (CW), carcass conformation (CC) and carcass fat (CF) in cattle using high-density genotype data, which, to the best of our knowledge, has not been reported previously.

Materials & methods

Data. After edits, carcass data, along with imputed, high-density genotype data, comprising 618,837 autosomal single nucleotide polymorphisms (SNPs), from 23,687 prime Irish beef cattle were available. The carcass data included carcass weight (CW), carcass conformation (CC; scored 1 [poor] to 15 [excellent]) and carcass fat (CF; scored 1 [thin] to 15 [fat]) information, measured in accordance with the EUROP grading system. All 618,837 SNPs had minor allele frequencies >0.05 and call rates >90%. All 23,687 animals also had call rates >90%. The parental origin of each allele was inferred by phasing the animals' genotypes in FImpute2 (Sargolzaei *et al.* 2014) using pedigree and genotype information from 130,178 animals; these animals included at least one genotyped parent of each animal of interest. Following phasing, four unique genotypes were distinguishable at each locus (e.g. AA, AB, BA and BB; where the first allele of the heterozygotes is paternally inherited and the second is maternally inherited). The four genotypes were then coded to represent additive (i.e. 0 (AA), 1 (AB and BA) and 2 (BB)), dominance (i.e. 0 (AA and BB) and 1 (AB and BA)) and imprinting (0 (AB), 1 (AA and BB) and 2 (BA)) effects.

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_370, © D. Kenny *et al.* 2022

Genome-wide association analysis. Using the GRAMMAR method outlined by Aulchenko *et al.* (2007a), all carcass phenotypes were pre-adjusted for nuisance variables, fitted as fixed effects, as well as for the direct polygenic effect of the animals *via* a genomic relationship matrix. All phenotypes were adjusted in the GenABEL package in R (Aulchenko *et al.* 2007b) using the following model:

$$y = 1\mu + X\beta + Z\alpha + e \quad (1)$$

where y was the vector of CW, CC scores or CF scores; 1 was a vector of ones; μ was the population mean; β was a vector of fixed effects that included contemporary group, birth herd type (i.e. beef or dairy), dam parity, recombination loss class, whether the animal was born a singleton or a twin and observed heterozygosity; α was a vector of random polygenic effects ($N(0,\mathbf{G}\sigma^2_a)$), where G and σ^2_a are a genomic relationship matrix and the genetic variance, respectively; e was a vector of random residual effects ($N(0,\mathbf{I}\sigma^2_e)$) where I and σ^2_e are an identity matrix and the residual variance, respectively; X and Z were incidence matrices for the fixed and random effects, respectively. Association analyses were then conducted separately for each locus using the following model:

$$e = 1\mu + b_1a_k + b_2d_k + b_3i_k + e_k \quad (2)$$

where e is the vector of corrected phenotypes from Equation 1; 1 was a vector of ones; μ was the population mean; a_k, d_k and i_k were the vectors of additive, dominance and imprinting genotype codes for locus k, respectively; b_1, b_2, and b_3 were the regression coefficients associated with the additive, dominance and imprinting genotype codes, respectively, and e_k was a vector of residuals. To test the significance of the b_3 regression coefficients and to correct for multiple testing, all *P*-values were transformed to q values (Storey and Tibshirani 2003). All SNPs with a q value ≤0.01 were considered significant, while all SNPs with a q value >0.01 and ≤0.05 were considered suggestive. Quantitative trait loci (QTL) regions were defined based on the flanking linkage disequilibrium patterns around the SNPs with significant or suggestive associations. To estimate the start and end positions of the QTL regions, all SNPs within a 0.5 Mb window that were in linkage disequilibrium (r^2 of ≥0.5) with another significantly or suggestively associated SNP on the same chromosome were considered part of a single QTL region. Overlapping regions were merged together and considered single regions.

Results

The associations between the imprinting genotype codes of each SNP and CW, CC and CF are in Figure 1. Fifty-one SNPs, 50 of which were located on *Bos taurus* (BTA) chromosome 2 with one (i.e. rs109356127) located on BTA 19, demonstrated significant (q≤0.01) imprinting associations with CW. In addition, a further 36 SNPs located on BTA 2 (29 SNPs), BTA 10 (one SNP), BTA 11 (one SNPs), BTA 16 (three SNPs), BTA 21 (one SNPs) and BTA 23 (one SNPs) demonstrated suggestive imprinting associations (0.01<q≤0.05) with CW. These SNPs collapsed into five distinct QTL regions located on BTA 2, ranging in length from 8.5 to 387.5 Kb, as well as a QTL region on BTA 16. All lead SNPs (i.e. the SNP with the strongest association) of these regions were intergenic. These intergenic QTL regions detected for CW stretched from 2.590 to 2.826 Mb, from 3.022 to 3.038 Mb, from 3.300 to 3.312 Mb, from 3.326 to 3.713 Mb, and from 4.100 to 4.104 Mb on BTA 2, as well as from 49.938 to 49.943 Mb on BTA 16. In addition to the imprinting associations, 23 SNPs and 41 SNPs on BTA 2 and BTA 6, respectively, demonstrated additive associations with CW, with three SNPs on BTA 2 demonstrating dominance associations with CW.

A total of 226 and 130 SNPs had significant and suggestive imprinting associations with CC, respectively. On BTA 2, 320 SNPs demonstrated either significant or suggestive imprinting associations, from which 23 imprinted QTL regions, ranging from 0.9 to 390.6 Kb in length, were identified. Based on the remaining

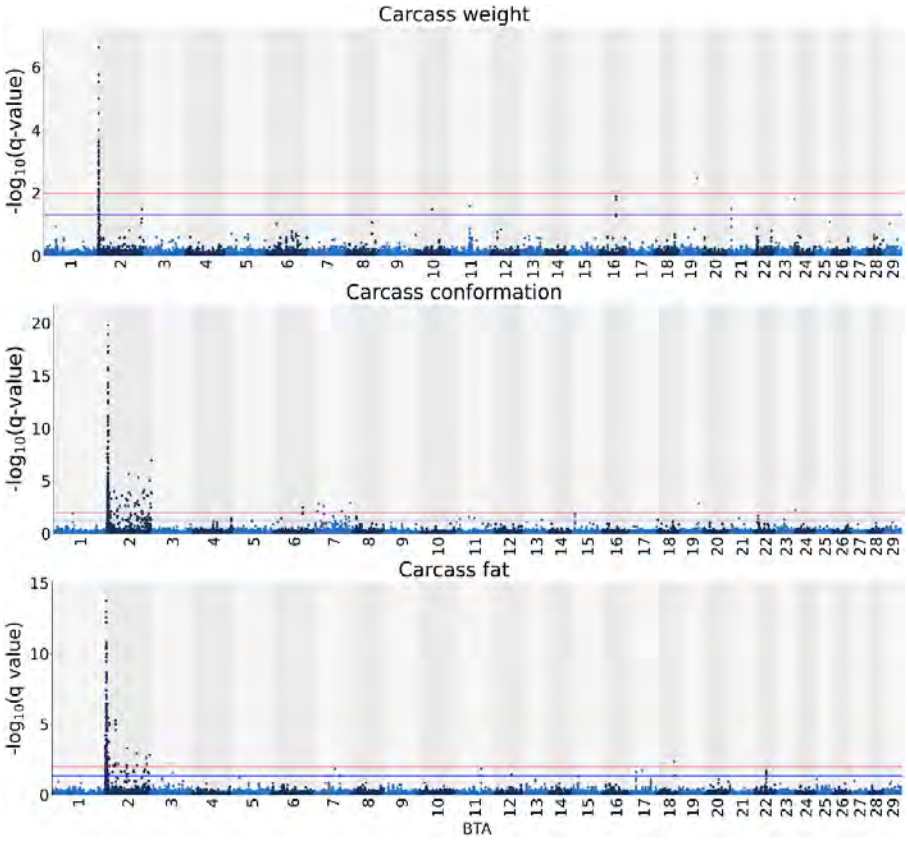


Figure 1 Manhattan plots of the $-\log_{10}$(q values) for imprinting associations with carcass weight, conformation and fat from each *Bos taurus* (BTA) chromosome. Blue and red lines are the thresholds for suggestive and significant associations, respectively.

SNPs demonstrating imprinting associations with CC, additional QTL regions were identified on BTA 4, BTA 6, BTA 8 and BTA 22. The associative lead SNPs of the QTL regions identified for CC included intronic variants of *NIPA1*, *NIAP2*, *IMP4*, *FAM168B*, *UGGT1*, *WDR33*, *BIN1*, *DDX58* and *STAC*, as well as intergenic variants. Furthermore, 12 SNPs on BTA 2 and one SNP on BTA 6 demonstrated additive associations with CC, and 45 SNPs, two SNPs, seven SNPs, one SNP and one SNP on BTA 2, 5, 8, 18 and 19, respectively, demonstrated dominance associations with CC.

Only 191 SNPs on BTA 2, along with a single SNP on BTA 18 (i.e. rs135579718) demonstrated significant imprinting associations with CF, with an additional 100 SNPs, located on eight different chromosomes, demonstrating suggestive associations with CF. The 292 SNPs demonstrating imprinting associations with CF collapsed into 26 distinct QTL regions on BTA 2, as well as into single QTL regions on BTA 7, BTA 18 and BTA 22. The lead SNPs of these regions included intronic variants of *OCA2*, *NIPA1*, *IMP4* and *WDR33*, *MAT2B* and *TAFA4*. Additionally, 161 and two SNPs on BTA 2 and 5, respectively, demonstrated additive associations with CF, with 179 SNPs on BTA 2, two SNPs on BTA 5 and a SNP on BTA 7 demonstrating dominance associations with CF. For all traits, no known imprinting gene was detected within 0.5 Mb of a SNP of interest.

Discussion

Knowledge of the presence of imprinting associations for a trait of interest and consideration of such associations in genomic evaluations can improve the accuracy associated with additive breeding values (Nishio and Satoh 2015). Furthermore, knowledge of the loci at which imprinting associations exist for carcass metrics could be implemented into mating advice programs for genotyped populations to exploit the effects of genomic imprinting in subsequent offspring (González-Recio *et al.* 2015). The presence of SNPs demonstrating imprinting association with carcass traits have been previously reported in cattle (Magee *et al.* 2010; Imumorin *et al.* 2011), although no such analysis has yet been conducted using high-density genotype data. The vast majority of the SNPs detected to have imprinting associations in the present study were located on BTA 2; this is despite the fact that imprinted genes have yet to be validated on BTA 2. Nonetheless, the presence of novel imprinting regions for beef carcass traits, namely CW and kidney, pelvic and heart fat, on BTA 2 was previously reported by Imumorin *et al.* (2011); Imumorin *et al.* (2011) advocates the need for further exploration on this chromosome. Additionally, an ortholog of an imprinted gene (i.e. IWS1) in humans and mice (Gregg *et al.* 2010) was located within 0.5 Mb of the QTL regions in the present study. Further to the fact that most SNPs demonstrating imprinting associations herein were located on BTA 2, genes widely acknowledged to having a key role is the expression of carcass phenotypes, namely *MSTN (Purfield et al. 2019)*, which was located near QTL regions detected herein, are also located on BTA 2.

References

Aulchenko Y.S., De Koning D.-J. and Haley C. (2007a) Genetics 177(1)**:** 577-585. https://doi.org/10.1534/genetics.107.075614

Aulchenko Y.S., Ripke S., Isaacs A. and Van Duijn C.M. (2007b) Bioinformatics 23(10)**:** 1294-1296. https://doi.org/10.1093/bioinformatics/btm108

González-Recio O., Toro M.A. and Bach A. (2015) Front Genet 6(305. https://doi.org/10.3389/fgene.2015.00305

Gregg C., Zhang J., Weissbourd B., Luo S., Schroth G.P. *et al.* (2010) Science 329(5992)**:** 643-648. https://doi.org/10.1126/science.1190830

Imumorin I.G., Kim E.-H., Lee Y.-M., De Koning D.-J., Van Arendonk J. *et al.* (2011) Front Genet 2(44. https://doi.org/10.3389/fgene.2011.00044

Magee D.A., Sikora K.M., Berkowicz E.W., Berry D.P., Howard D.J. *et al.* (2010) BMC Genet 11(1)**:** 1-15. https://doi.org/10.1186/1471-2156-11-93

Nishio M., and Satoh M. (2015) Genet Sel Evol 47(1)**:** 1-10. https://doi.org/10.1186/s12711-015-0091-y

Purfield D., Evans R. and Berry D. (2019) BMC Genom 20(1)**:** 1-17.

Reik W., and Walter J. (2001) Nat Rev Genet 2(1)**:** 21-32. https://doi.org/10.1038/35047554

Sargolzaei M., Chesnais J.P. and Schenkel F.S. (2014) BMC Genom 15(1)**:** 1-12. https://doi.org/10.1186/1471-2164-15-478

Storey J.D., and Tibshirani R. (2003) PNAS 100(16)**:** 9440-9445. https://doi.org/10.1073/pnas.1530509100

371. Evidence of genetic links between rumen bacteria and milk somatic cell score in Lacaune dairy sheep

G. Martinez Boggio[1], A. Meynadier[1], R. Rupp[1], C. Allain[2], A.J. Buitenhuis[3] and C. Marie-Etancelin[1*]

[1]GenPhySE – INRAE – ENVT – Université de Toulouse, 24 Chemin de Borde Rouge, 31326 Castanet-Tolosan, France; [2]INRAE – Unité Expérimentale de La Fage, 12250 Saint-Jean et Saint-Paul, France; [3]Center for Quantitative Genetics and Genomics, Aarhus University, Blichers Allee 20, 8830 Tjele, Denmark; christel.marie-etancelin@inrae.fr

Abstract

Mastitis is a high prevalence infectious disease in dairy sheep. Even if selection by milk somatic cell score (SCS) has been implemented to increase mastitis resistance, we wonder if rumen bacteria could influence udder inflammatory response. The objectives are to estimate genetic correlations between rumen bacteria and average SCS over lactation (LSCS), and to identify underlying mechanism between traits. 795 Lacaune ewes were studied using metabarcoding (16S rRNA gene) for rumen bacteria characterization and annual SCS records. Two-trait genomic models were used for the parameter estimation and GWAS. Results showed significant genetic correlations, and colocalized QTLs for LSCS and nine bacterial operational taxonomic units (OTUs) on chromosomes 3, 17, and 20. We found a genetic association between resistance/susceptibility to mastitis, using LSCS, and the ruminal bacteria abundance. These preliminary results pave the way for better understanding the biological basis and causalities of an intra-mammary inflammations in sheep.

Introduction

As in cattle, mastitis is one of the most important infectious diseases in small dairy ruminants, due to its high frequency and the price reduction paid to the farmer of up to 10% (95€ per 1000 litres of ewes' milk). The selection was focused on the milk SCS, as a tool to decrease subclinical intra-mammary infections, which is the main problem in dairy sheep. The existence of a genetic basis for mastitis resistance has been demonstrated, and the underlying mechanisms involved such as the *suppressor of cytokine signalling 2* (Socs2) gene (Rupp *et al.*, 2015). Recently, the influence of the rumen microbiome on subclinical ketosis has been demonstrated in cows (Gebreyesus *et al.*, 2020), but no work is available in sheep. Therefore, could rumen bacteria impact on intra-mammary inflammations in sheep? To address this, and after quantifying host control over rumen bacteria we: (1) estimated genetic correlations between rumen bacteria and LSCS; and (2) identified possible underlying mechanisms with genome-wide association analysis (GWAS).

Materials & methods

Animals phenotyping. A total of 795 Lacaune dairy ewes were reared at INRAE Experimental Unit of La Fage (UE 321 agreement A312031, Roquefort, France), between 2015 to 2019. All ewes (weighing 77 kg on average) raised indoors, and fed 93% meadow hay and silage plus 7% of concentrates (on dry matter basis). The daily somatic cell count (SCC) was quantified with a Fossomatic cell counter (Nanterre, France), as part of the official milk recordings of the flock. Then, the average LSCS was calculated as log of somatic cell count, including at least four milk test-days. Rumen sampling of adult ewes between 28 to 133 days in milk was done once per ewe using a vacuum pump and a medical gastric tube. The protocol received approval from the French Ministry of Higher Education, Research and Innovation – Animal Ethics Committee (approval number: APAFIS#6292-2016080214271984 v8). The RNA sequences of the 795 samples were processed with the FROGS 3.0 pipeline, and OTUs with abundances below 0.005% were removed (Bokulich *et al.*, 2013). The resulting abundance table included 2,059 OTUs, but we focused on the 306 OTUs with

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_371, © G. Martinez Boggio *et al.* 2022

heritability significantly different from zero (empirical significant threshold of 0.10 obtained after running 10,000 analyses where OTUs abundancies were shuffling randomly across individuals) using an additive relationship matrix.

Animals genotyping. DNA extraction from blood samples and genotyping were performed for 795 Lacaune ewes. From those ewes, 743 were genotyped using a medium-density SNP chip (Illumina Ovine SNP50 BeadChip, 54,241 SNPs). The remaining 52 ewes were genotyped with a low-density SNP chip (Illumina Ovine SNP15, 16,681 SNPs) and imputed to a medium-density SNP chip genotypes in the framework of the Lacaune dairy sheep genomic selection program (Larroque *et al.*, 2017). SNPs were removed using a call rate <0.99 for SNPs, and <0.95 for individuals, and minor allele frequency <0.05. In total, for the 26 *Ovis aries* autosomal chromosomes (OAR), 35,496 SNPs remained for the analyses, plus one SNP corresponding to the Socs2 gene mutation (OAR3).

Parameter estimation. To estimate the heritability of the 306 rumen OTUs and LSCS, and the genetic correlations between each OTU and LSCS, a two-trait animal model was used:

$$\begin{bmatrix}\mathbf{y_1}\\ \mathbf{y_2}\end{bmatrix} = \begin{bmatrix}\mathbf{X_1} & \mathbf{0}\\ \mathbf{0} & \mathbf{X_2}\end{bmatrix}\begin{bmatrix}\mathbf{b_1}\\ \mathbf{b_2}\end{bmatrix} + \begin{bmatrix}\mathbf{Z_1} & \mathbf{0}\\ \mathbf{0} & \mathbf{Z_2}\end{bmatrix}\begin{bmatrix}\mathbf{a_1}\\ \mathbf{a_2}\end{bmatrix} + \begin{bmatrix}\mathbf{e_1}\\ \mathbf{e_2}\end{bmatrix} \quad (1)$$

where y_1 and y_2 are the vectors of observations for LSCS and one OTU at a time, respectively; b_1 and b_2 are the vectors of fixed effects, b_1 including for LSCS: litter size, number of milking test-day, and b_2 for the OTUs: sampling year, total number of DNA sequences per sample, and three effects nested on year sampling: lactation number, sequencing run, and order-time of sampling; a_1 and a_2 are the vectors of random additive effects; and e_1 and e_2 are the vectors of residual effects. X_1 and X_2 are incidence matrices relating fixed effects to vector y_1 and y_2, respectively; Z_1 and Z_2 are incidences matrices relating the additive effects to vector y_1 and y_2, respectively. The distributional assumptions are $\mathbf{a} \sim N(0, \mathbf{G} \otimes \mathbf{W})$, and $\mathbf{e} \sim N(0, \mathbf{I} \otimes \mathbf{R})$; G is a genomic relationships matrix as $\mathbf{G} = \mathbf{ZZ}'/2\sum p_i(1-p_i)$ (VanRaden, 2008); I is an identity matrix; W and R are the variance-covariance matrices for the random additive and residual effects, respectively. The heritability was estimated as $h^2 = \sigma_a^2/\sigma_y^2$, and the genetic correlations as $r_g = \sigma_{a12}/\sqrt{(\sigma_{a1}^2\sigma_{a2}^2)}$. The variance components were estimated using AIREMLF90 (Misztal *et al.*, 2002). The significance was considered as genetic correlation value greater than the standard error.

Genome-wide association. The GWAS was performed using the two-trait model presented in Equation 1. After solving the single step GBLUP model, all SNP effects were estimated back solving the breeding values estimates as proposed by Aguilar *et al.* (2019) using POSTGSF90 (Misztal *et al.*, 2002). To correct for multiple testing the false discovery rate (FDR) was used as implemented in the p.adjust package in R software. A SNP was considered significant at an FDR of $P<0.10$.

Results

The average heritability estimated for LSCS was 0.39±0.07. Genetic parameters of the 14 rumen bacteria having significant genetic correlations with LSCS are presented in Table 1.

The GWAS for LSCS detected on OAR3 four significant QTLs: in decreasing order of significance, the SNP for Socs gene (129,722,200 bp), SNP rs412514556 (129,423,378 bp), SNP rs425386363 (23,336,791 bp) and SNP rs416770789 (133,977,235 bp). The microbiome GWAS detected 387 significant SNPs for 67 OTUs among the 306 OTUs with heritability significantly different from zero. SNPs were scattered on all chromosomes, with the highest *P*-values and highest number of signals on OAR1, OAR2, OAR3, and OAR5. But, in this study we focus on the OTUs with significant signals located close to the LSCS QTLs on OAR3 (Table 2).

Table 1. Heritability (h^2) and genetic correlations (r_g) between lactation somatic cell score and ruminal bacteria abundance.

Genus	OTU name	h^2 (SE)	r_g (SE)
Lachnospiraceae_NK3A20_group	OTU506	0.10 (0.06)	0.68 (0.63)
Prevotella	OTU871	0.17 (0.07)	0.58 (0.43)
Monoglobus	OTU1066	0.13 (0.06)	0.51 (0.40)
Leifsonia	OTU1479	0.14 (0.07)	0.48 (0.43)
Lachnospiraceae/unknown genus	OTU309	0.15 (0.07)	0.47 (0.41)
Mailhella	OTU2331	0.15 (0.06)	0.39 (0.34)
F082/unknown genus	OTU1009	0.14 (0.06)	0.35 (0.29)
Acetitomaculum	OTU612	0.23 (0.07)	0.31 (0.21)
Lachnospiraceae NK3A20 group	OTU278	0.20 (0.07)	0.30 (0.24)
Lachnospiraceae/unknown genus	OTU1666	0.19 (0.07)	-0.21 (0.20)
Prevotella	OTU1496	0.22 (0.07)	-0.32 (0.24)
Prevotella	OTU196	0.16 (0.06)	-0.35 (0.28)
Prevotella	OTU136	0.13 (0.06)	-0.36 (0.34)
F082/unknown_genus	OTU2431	0.20 (0.07)	-0.37 (0.22)

SE= standard error.

Table 2. Summary of the significant results obtained from the microbiome GWAS.

Genus	OTU name	SNP	Position (bp)	-log10 (*P*-value)
Prevotella	OTU1419	rs425386363	23,336,791	5.59
Lachnospiraceae/unknown genus	OTU1666[1]	rs425386363	23,336,791	6.90
Ruminococcus	OTU602	Socs	129,722,200	5.70

[1] OTUs with significant genetic correlations with LSCS.

Discussion

The LSCS heritability was greater than the 0.15 reported by Rupp *et al.* (2009), using the whole Lacaune population. Even if the genetic correlations have a high standard error which can lead to spurious ones, most genetic correlations between LSCS and OTUs were positive, meaning that higher bacterial abundance is genetically associated with susceptibility to mastitis. There is also a higher variety of genera within the positive correlations than for the negative correlations where the most frequently associated is *Prevotella*, a bacteria belonging to the *Bacteroidetes* phylum. Zhong *et al.* (2018) obtained *Bacteroidetes* as the most abundant in the group with low SCC, even if they showed that the rumen microbiome in cows was globally stable at different SCC levels. The link between gut bacteria and mastitis is probably not direct but would be through bacterial metabolites absorbed by the host, and which would have pro- or anti-inflammatory properties (Hu *et al.*, 2019). A QTL on OAR3 was detected for *Ruminococcus* OTU602 at the Socs gene position, and the Socs2 genotype known to have more LSCS in milk is associated with highest abundance of OTU602 (results not shown). In contrast, Chuang *et al.* (2021) found that *Ruminococcus* genus was phenotypically reduced in the rumen of cows with clinical mastitis. As well on OAR3, two QTLs were detected, one for *Prevotella* OTU1419 and one for *Lachnospiraceae* OTU1666. The SNP genotype associated with more LSCS in milk is linked with more *Prevotella* OTU1419 abundance, which is contrary with the main genetic correlations for this genus, but on others OTUs. The same SNP genotype associated with mastitis susceptibility is also linked with less *Lachnospiraceae* OTU1666 abundance, which is consistent

with the negative but weak genetic correlation of -0.21 with LSCS. This result on OTU1666 is in agreement with Wang *et al.* (2021) finding that this family is phenotypically associated with anti-inflammatory molecules such as 2-Phenylbutiric acid. But, for the three others OTUs belonging to the *Lachnospiraceae* family, an increase of their abundances is genetically linked with more LSCS. The identification in this study of QTLs on OAR3 common for LSCS and OTUs in addition to the results obtained for genetic correlations, contribute to the hypothesis of a genetic link between rumen bacteria and mastitis. We can conclude that host genetic may influence ruminal bacteria abundances and udder inflammation in sheep via mainly positive genetic correlations, but the underlying metabolic pathways are still difficult to identify with only the identification of bacterial genera. The genetic study of these bacteria's functions is underway in order to unravel the various results obtained at the OTU level.

References

Aguilar I., Legarra A., Cardoso F., Masuda Y., Lourenco D., *et al.* (2019). Genet Sel Evol 51:28. https://doi.org/10.1186/s12711-019-0469-3

Bokulich N.A., Subramanian S., Faith J.J., Gevers D., Gordon J.I., *et al.* (2013). Nat Methods 10:57–59. https://doi.org/10.1038/nmeth.2276

Chuang S.-T., Li K.-Y., Tu P.-W., Ho S.-T., Hsu C.-C., *et al.* (2021). Animals 11:3108. https://doi.org/10.3390/ani11113108

Gebreyesus G., Difford G.F., Buitenhuis B., Lassen J., Noel S.J., *et al.* (2020). J. Dairy Sci. 103:4557–4569. https://doi.org/10.3168/jds.2019-17824

Hu X., Li S., Fu Y., Zhang N., (2019). Eur. J. Clin. Microbiol. Infect. Dis. 38:1409–1423. https://doi.org/10.1007/s10096-019-03549-4

Larroque H., Chassier M., Saintilan R., Astruc J.M. (2017). Proc. of EAAP, Tallinn, Estonia.

Misztal I., Tsuruta S., Strabel T., Auvray B., Druet T., *et al.* (2002). Proc. of 7th WCGALP, Montpellier, France.

Rupp R., Bergonier D., Dion S., Hygonenq M.C., Aurel M.R., *et al.* (2009). J. Dairy Sci. 92:1203–1219. https://doi.org/10.3168/jds.2008-1435

Rupp R., Senin P., Sarry J., Allain C., Tasca C., *et al.* (2015). PLoS Genet 11:e1005629. https://doi.org/10.1371/journal.pgen.1005629

VanRaden P.M., (2008). J. Dairy Sci. 91:4414–4423. https://doi.org/10.3168/jds.2007-0980

Wang Y., Nan X., Zhao Y., Jiang L., Wang M., *et al.* (2021). J Animal Sci Biotechnol 12:36. https://doi.org/10.1186/s40104-020-00543-1

Zhong Y., Xue M., Liu J. (2018). Front. Microbiol. 9:3217. https://doi.org/10.3389/fmicb.2018.03217

372. Leveraging GWAS models to map selection on complex traits in massive datasets

T.N. Rowan[1*], R.D. Schnabel[2] and J.E. Decker[2]

[1]*University of Tennessee Institute of Agriculture, 2506 River Drive, 37996, Knoxville, TN, USA;*
[2]*University of Missouri, 920 East Campus Drive, 65211, Columbia, MO, USA; trowan@utk.edu*

Abstract

Understanding the biology that underlies artificial selection in animal breeding programs is of significant interest. Recent and ongoing polygenic selection is difficult to map because it does not leave well-defined signatures on the genome. When populations have temporally-stratified genomic data, we can use genome-wide association (GWAS) models to map selection during the sampled period using Generation Proxy Selection Mapping (GPSM). Here, we demonstrate the power of GPSM in two large datasets from American beef cattle populations with 11,759,568 imputed SNPs. We identify ongoing polygenic selection at 294 unique loci. We observe that GPSM has high power to detect minuscule allele frequency shifts over short timescales (<10 years). GPSM loci localized near genes, in annotated epigenetic marks, or in predicted functional genomic regions. We also observed minimal overlap in loci identified by GPSM and other traditional sweep detection methods.

Introduction

While selection occurs on phenotypes or estimated breeding values, the changes observed in a population result from changes in the frequencies of causal variants that underlie the selected trait. Occasionally, a large-effect causal variant is rapidly driven to high frequency in the population, leading to 'selective sweeps' where surrounding neutral genetic variation is reduced (Smith and Haigh 1974). Studies have mapped numerous selective sweeps in livestock species caused by selection on Mendelian traits or large-effect quantitative trait loci (QTL) (Gutiérrez-Gil *et al.* 2015). Selective sweeps at these loci were likely essential components of domestication and breed formation. That said, it is likely that the majority of selection has been on traits that are highly polygenic (Bolormaa *et al.* 2014). Polygenic selection results in subtle allele frequency shifts at hundreds or thousands of causal variants of small effect, which does little to alter surrounding neutral diversity. This makes mapping polygenic selection difficult.

In contemporary livestock populations, we have accumulated large genotypic datasets containing thousands or millions of animals, including influential founder animals. We can leverage these large populations with high-density, temporally-stratified genotypes to detect small directional changes to allele frequency caused by polygenic selection. We use Generation Proxy Selection Mapping (GPSM) to map loci actively responding to polygenic selection. GPSM uses a proxy for the generations that separate an individual from the beginning of a population's pedigree as the dependent variable in a GWAS model. The goal is to detect statistically significant associations between allele frequency and an animal's generation. These associations represent directional allele frequency changes caused by selection (Decker *et al.* 2012; Rowan *et al.* 2021). Simulations in previous work showed that by using a mixed model, GPSM can effectively distinguish selection from drift regardless of effective population size, skewedness of data, family or population structure, and other confounders. We apply GPSM to datasets from two modern American beef cattle breeds, Red Angus and Simmental. Using 11,759,568 imputed genotypes and conditional joint analysis of GPSM results, we can map putative sites underlying ongoing selection for complex traits, allowing us to understand better the biology that underlies selection for production traits.

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_372, © T.N. Rowan *et al.* 2022

Materials & methods

Genotype data and imputation. We used assay genotypes from two breed associations, Red Angus (n=46,454) and Simmental (n=90,580), ranging in density from 25K to 770K markers. Genotypes were imputed to a set of 811K SNPs, as described in Rowan *et al.* 2019 (Rowan *et al.* 2019), followed by imputation to 43,214,290 SNPs using the 4,931 animals in Run8 of the 1000 Bulls Project (Hayes and Daetwyler 2019), phased with Eagle (v2.4) (Loh *et al.* 2016) as a haplotype reference. We retained only high quality variants (VQSR Tranche 90) and removed poorly imputed (internal imputation R^2<0.4), or rare (minor allele frequency <0.01) to arrive at a set of 11,759,568 variants. We used genomic coordinates from the ARS-UCD1.2 assembly (Rosen *et al.* 2020) for all analyses.

Generation proxy selection mapping (GPSM). We used the breeder-reported birthdate to calculate the continuous years between an animal's birth and the day we began analyzing data (19 October 2020). We used this value as the dependent variable in GPSM analyses. We used the 811K SNP sets to construct genomic relationship matrices (GRM) for each breed. We used the '--mlma' function in GCTA (v 1.92.3) (Yang *et al.* 2010; Yang *et al.* 2011) to conduct our GPSM analysis using the following model:

$$y = \mu + bx + a + e \quad (1)$$

Here, y is a vector of generation proxies, μ is the sample mean, b is a SNP's regression coefficient, on a vector of animal genotypes x, a is a random vector of polygenic terms, and e is a random error term. We performed this analysis on the full datasets of both breeds, an age-truncated Red Angus dataset (animals born <10 years ago), and a purebred-only Simmental dataset. To refine GPSM signals, we performed a conditional and joint analysis (COJO) (Yang *et al.* 2012), conditioning associations on SNPs with GPSM P-values $<10^{-5}$. Significant COJO SNPs were ones with conditional and genome-wide P-values $<5\times10^{-8}$.

Results

In Red Angus, GPSM detected 2,914 and 9,065 genome-wide significant SNPs (Bonferroni-corrected P-values $<4.29\times10^{-9}$) in the full and birth date-truncated datasets, respectively (Figure 1A & 1B). Our COJO analysis of these GPSM results identified 72 unique lead SNPs in the full dataset and 96 in the truncated dataset (COJO SNPs). Sixteen of these SNPs were identified by both Red Angus datasets. The most striking example of the truncated dataset uncovering unique GPSM associations was on BTA2 near the gene *ARHGAP15*, a major immune function gene in cattle (Noyes *et al.* 2011; Álvarez *et al.* 2016). Both GPSM analyses in Red Angus identified a known pleiotropic QTL for body weight, carcass weight, and sexual precocity on BTA14. Many significant GPSM loci had not been previously identified by other sweep mapping studies (Gutiérrez-Gil *et al.* 2015). The increased power in these datasets uncovered dozens of novel signatures that were not identified by previous GPSM analyses in the same populations (Rowan *et al.* 2021).

In the American Simmental dataset, we performed two GPSM analyses to compare and contrast selection within the purebred Simmental part of the dataset and the entire open herdbook (Animals 5-100% pedigree estimated Simmental). GPSM identified 513 genome-wide significant SNPs in the purebred dataset driven by 18 unique COJO SNPs (Figure 1D). The strongest GPSM signature in purebred Simmental resided on BTA5, near the genes *PMEL* and *ERBB3*. This locus is a known modulator of coat color in European Simmental populations (Mészáros *et al.* 2015). Another selected locus was identified near the gene *KIT,* which contains alleles that control color-sidedness in many mammalian populations (Durkin *et al.* 2012). GPSM in the full Simmental dataset identified significantly more selected loci, 1,008 SNPs and 108 unique associations (Figure 1C). This analysis also identified strong ongoing selection at the same coat color and pattern loci and identified selection on other loci known to affect economically-important quantitative traits.

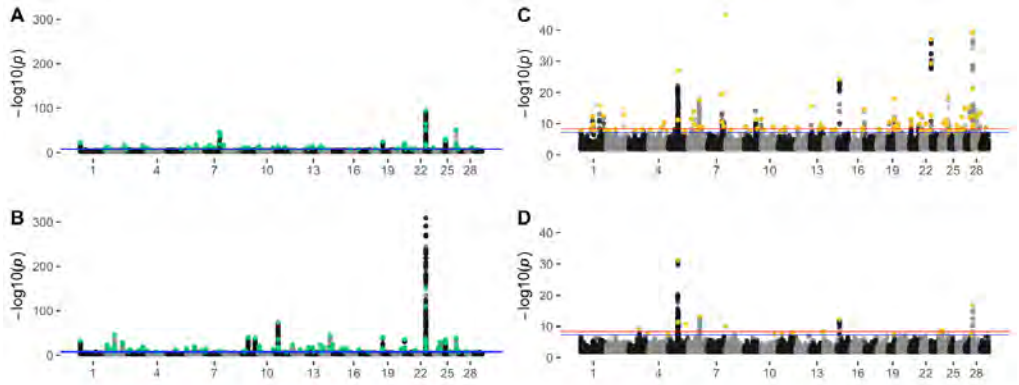


Figure 1. GPSM Manhattan plots for (A) full and (B) young animal truncated Red Angus datasets, and (C) full and (D) purebred Simmental datasets. Colored SNPs were identified as independent associations through COJO analysis. The red line represents Bonferroni significance, and the blue line represents a 5×10^{-8} significance threshold.

We observed only a handful of loci that were under selection in both the Red Angus and Simmental breeds. Red Angus's most significant GPSM signature resided on BTA23 (1-2 Mb), immediately upstream of the gene KHDRBS2. This area was also identified in the Simmental GPSM. However, the selected loci in the region were different, suggesting that depending on the population, selection may be acting on different regulatory variations for the same gene. Across datasets, most GPSM COJO SNPs had a clear candidate gene (Table 1). Many of the selected SNPs with a clear positional candidate gene were immediately upstream of the transcription start sites, suggesting possible selection on *cis*-regulatory variation.

Discussion

Our results show that GPSM, coupled with high-powered livestock datasets can detect very subtle shifts in allele frequency over short time periods, even in the absence of recorded phenotypes. Using a linear mixed model to test allelic associations with a generation, we can explicitly control the population and family structure, which often obscures selection mapping studies. Simulations in previous studies (Rowan *et al.* 2021) indicate that GPSM can effectively distinguish selection from drift and is robust to genotype sampling. The combination of GPSM and sequence-density imputed genotypes lets us map this selection at an extremely high resolution.

Our results suggested that since the initial import of Simmental to America, much selection pressure within the purebred animals has made them appear more similar to Angus animals (solid black-hided). The large sample sizes of our two datasets allowed us to divide them into informative subsets to understand

Table 1. Significant GPSM (COJO and genome-wide $P<5\times10^{-8}$) SNPs and their proximity to annotated genes.

Breed	Subset	COJO SNPs	In gene	Near gene (<50 kb to nearest)	Intergenic
Red Angus	Full	72	17	20	35
Red Angus	Young	96	36	24	36
Simmental	Full	108	32	36	40
Simmental	Purebred	18	6	7	3

how selection operates at very short timescales (Red Angus 'young' animals) or within a specific subset of an open herdbook breed (purebred Simmental). With large or more structured datasets, this subsetted approach to performing GPSM could be informative in understanding how changes to breeding decision-making have impacted the genome over time in a line or family-specific manner (e.g. genomic changes in terminal vs maternal lines). Breeders might also use GPSM results as 'biological priors' in Bayesian genomic prediction approaches. Further, we believe that using putatively selected loci as 'functional annotations' to the genome can add important context to other mapping studies (GWAS, eQTL, etc.).

References

Alachiotis N., and Pavlidis P. (2018) Communications Biology 1: 79. https://doi.org/10.1038/s42003-018-0085-8

Bolormaa S., Pryce J.E., Reverter A., Zhang Y., Barendse, *et al.* (2014) PLoS Genetics 10: e1004198. https://doi.org/10.1371/journal.pgen.1004198

Decker J.E., Vasco D.A., McKay S.D., McClure M.C., Rolf M.M., *et al.* (2012) BMC Genomics 13: 606. https://doi.org/10.1186/1471-2164-13-606

Ferrer-Admetlla A., Liang M., Korneliussen T., and Nielsen R. (2014) Molecular Biology Evolution 31: 1275–1291. https://doi.org/10.1093/molbev/msu077

Gutiérrez-Gil B., Arranz J.J., and Wiener P. (2015) Frontiers Genetics 6: 167. https://doi.org/10.3389/fgene.2015.00167

Hayes B.J. and Daetwyler H.D. (2019) Annual Reviews Animal Bioscience 7: 89–102. https://doi.org/10.1146/annurev-animal-020518-115024

Loh, P. R., Danecek, P., Palamara, P. F., Fuchsberger, C., A Reshef, Y., *et al.* (2016). Nature Genetics, 48(11), 1443–1448. https://doi.org/10.1038/ng.3679

Rowan T.N., Hoff J.L., Crum T.E., Taylor J.F., Schnabel R.D., *et al.* (2019) Genetics Selection Evolution 51: 77. https://doi.org/10.1186/s12711-019-0519-x

Rowan T.N., Durbin H.J., Seabury C.M., Schnabel R.D., and Decker J.E. (2021) PLoS Genetics 17: e1009652. https://doi.org/10.1371/journal.pgen.1009652

Rosen B.D., Bickhart D.M., Schnabel R.D., Koren S., Elsik C.G, *et al.* (2020) Gigascience 9. https://doi.org/10.1093/gigascience/giaa021

Smith J.M., Haigh J. (1974) Genetics Research 23: 23–35. https://doi.org/10.1017/S0016672300014634

Szpiech Z.A. and Hernandez R.D.(2014) Molecular Biology Evolution 31: 2824–2827. https://doi.org/10.1093/molbev/msu211

Yang J., Benyamin B., McEvoy B.P., Gordon S., Henders A.K., *et al.* (2010) Nature Genetics 2010;42: 565–569. https://doi.org/10.1038/ng.608

Yang J., Lee S.H., Goddard M.E., and Visscher P.M. (2011) American Journal of Human Genetics 88: 76–82. https://doi.org/10.1016/j.ajhg.2010.11.011

Yang J, Ferreira T, Morris AP, Medland SE, *et al.* (2012) 44: 369–75, S1–3. https://doi.org/10.1038/ng.2213

373. Fine-mapping young-stock survival QTL on chromosome 6 in Nordic Red Dairy Cattle

Z. Cai[1], X. Wu[1], B. Thomsen[2] and G. Sahana[1*]

[1]*Aarhus University, Center for Quantitative Genetics and Genomics, Blichers Allé 20, 8830 Tjele, Denmark;* [2]*Aarhus University, Department of Molecular Biology and Genetics, C.F. Møllers Allé 3, 8000 Aarhus, Denmark; goutam.sahana@qgg.au.dk*

Abstract

A genome-wide association study was carried out for young-stock survival (YSS) in Nordic Red Dairy cattle. The de-regressed proofs for 3,078 bulls for YSS index and four component traits were analyzed. The 50k genotypes were imputed to whole genome sequence level using *run7* 1000 bull genome sequence data. A major QTL on chromosome 6 was identified with rs133470751 as the lead SNP with $-\log_{10}$(*P*-value) equal to 47.02. The lead SNP is located in a gene desert. The post-GWAS analysis for functional annotations of the variants and presence for structural variants in the QTL peak was evaluated for being causal factor.

Introduction

Identification of the genetic variants associated with calf survival in dairy cattle will aid in the elimination of harmful mutations from the cattle population and the reduction of calf and young stock mortality rates (Wu *et al.* 2017). Mortality of cattle causes economic loss for dairy farmers. The cost at rearing period is relatively high compare with early embryonic loss, abortion and stillbirth. Calves and young stock that die during the rearing period result in lost revenue, fewer heifers for replacement, higher veterinarian costs and deteriorating animal welfare. Therefore, it is important for both economy and welfare, to breed for reducing mortality of calves and young stock.

An index for young stock survival (YSS) in calves was included in the Nordic total merit index by the Nordic Cattle Genetic Evaluation (NAV; www.nordicebv.info). The YSS index was calculated based on 4 survival traits: survival from 2 to 30 d for bull calves (BP1) and heifer calves (HP1), from 31 to 184 d for young bull calves (BP2), and from 31 to 458 d for young heifer calves (HP2). Calf death and survival during this period were recorded as 0 and 1, respectively. The YSS index was calculated by combining EBV for BP1, BP2, HP1, and HP2 by NAV (Denmark), which were weighted by their relative economic values and standardized (Carlen 2016).

The combination of genomic information and breeding values for YSS creates an opportunity to identify genomic variants with harmful effects on YSS. Once these variants have been identified, it will be possible to select against them and prevent at-risk matings between carriers. This selection will improve calf survival and reduce the cost per live cattle produced. In this study, we focus to identify the causal variant underlying a major QTL on chromosome 6 in Nordic Red Dairy cattle (RDC).

Materials & methods

Phenotype and genotype data. For genome scan, we used data from 3,078 RDC bulls. The phenotype used was de-regressed proofs for 4 component traits of YSS (i.e. BP1, BP2, HP1 and HP2) and YSS index.

All these RDC bulls were genotyped with Illumina 50k SNP chip; followed by imputed to HD level (777k), and then imputed to whole genome sequence level using *Run7* reference animals from the 1000 Bull Genomes project (BGP). The whole genome sequencing (WGS) reference panel was phased with Beagle4 (Browning *et al.* 2018) and recalled by shapeit2 (Delaneau *et al.* 2011). HD panel was phased using Beagle5

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_373, © Z. Cai *et al.* 2022

(Browning *et al.* 2018). Both steps of imputation were performed by Minimac4 (Fuchsberger *et al.* 2015). The average imputation accuracy, R^2 from Minimac4, was 91.2% across all minor allele frequency classes. The SNP with a minor allele frequency less than 1% or those deviating from Hardy-Weinberg proportions ($P<10^{-6}$) were removed. Association analysis was carried our using a linear-mixed model approach (Kang *et al.* 2010) using GCTA software (Yang *et al.* 2011). A SNP was declared significant if its *P*-value was less than multiple Bonferroni correction threshold (see Wu *et al.* 2017 for details).

Post-GWAS. To reveal the possible consequence of the variants within the QTL interval, we performed variants annotation using variant effect predictor (VEP) v104 (McLaren *et al.* 2016). The prediction impact with 'HIGH' was retained to check the segregation in the WGS animals from 1000 BGP Run7. The Gene Ontology (GO) (Ashburner *et al.*, 2000) was used to explore the function of identified candidate genes. The precomputed Repeatmasker (Chen 2004) profile was downloaded from NCBI, and checked the presence of the repeat sequence around the lead SNP. Variation in repeats can alter the expression of genes, and changes in the number of repeats have been linked to certain human diseases (de Bustos *et al.* 2016). The flanking 1 Mb region of was compared to the regulatory elements from previous studies (Kern *et al.* 2021). We also examined if chromosomal deletions (Mesbah-Uddin *et al.* 2018) located in or close of the QTL regions could be the causal variants.

Results

Association with component traits of YSS. We detected genome-wide associated SNP, including 9,162 SNP for BP1, 12,173 SNP for BP2, 8,709 SNP for HP1, and 14,096 SNP for HP2. Of the 14,567 SNP that were significant for YSS index, 8,856 SNP showed significant association with BP1, 11,688 SNP for BP2, 8,339 SNP for HP1, and 13,375 SNP for HP2. In addition, there were 8,268 SNP on BTA6, 3 SNP on BTA17, and 266 SNP on BTA23 were associated with both BP1 and HP1. There were 182 SNP on BTA4, 11,655 SNP on BTA6, and 2 SNP on BTA14 associated with both BP2 and HP2. There were 8,400 SNP on BTA6 were associated with both BP1 and BP2; and 8,243 SNP on BTA6 associated with both HP1 and HP2.

Association with YSS. We found a major QTL for YSS at chromosome 6 (Figure 1). The lead SNP of this QTL is 6:38065928 (rs133470751) with $-\log_{10}$(*P*-value) equal to 47.02. The annotation of this variant is intergenic variant. The lead SNP is located at the gene desert (Figure 2). There are several genes in a far distance which are worth to investigate.

Checking variants annotation of the flanking area of lead SNP. We checked the flanking 1 Mb region for high-impact (given by variants effect predictor) of the lead SNP. Two genes with several functionally annotated SNPs showed potential to be causal genes and/or causal variants. *FAM184B* have a stop gained

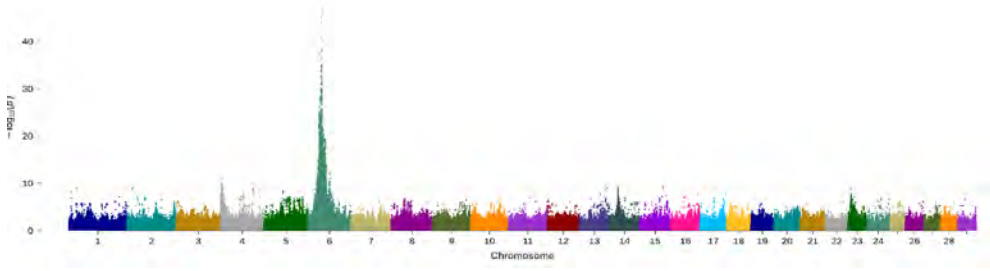


Figure 1. Manhattan plot for association of SNP with young stock survivor in Nordic Red cattle. Red horizontal line indicates genome-wide significance level [$-\log_{10}$(P)=8.5].

variants at 6:37201641. *LCORL* have three frameshift variants at 6:37401110, 6:37401770, 6:37402602 and 6:37403627 and one splice donor variant at 6:37412364. The segregation of these variants in non-Nordic Red WGS animals and Nordic Red WGS animals were listed in Table 1. *LCORL* belong to GO term 'regulation of transcription by RNA polymerase II'. Both genes do not show any clear functional meaning to YSS from GO, KEGG or mammalian phenotype database.

Unveiling the gene desert region around lead SNP. The Repeatmasker profile was downloaded from NCBI, there is a LINE element very close to the lead SNP, which is L2c. The copy in the reference genome is not a full length element, which means it is not active. However, we currently don't have the right dataset, long read sequences, to further investigate the repeat content of this region in our RDC animals. The sequence reads obtained with next generation sequencing platforms are short and simply do not span long repetitive sequences. We checked for the presence of regulatory elements in the flanking 1 Mb region of the lead SNP, and the result showed the region from 38 Mb to 39 Mb contains several regulatory elements even though the region is devoid of genes.

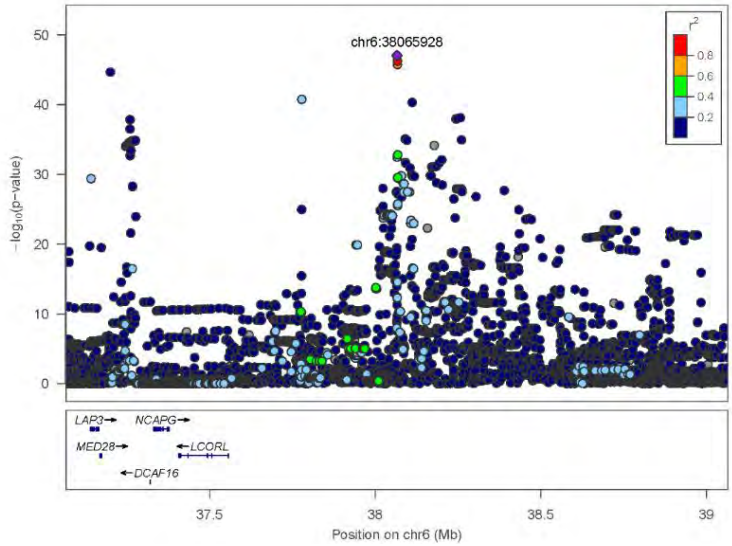


Figure 2. LocusZoom plot (Pruim *et al.* 2010) showing the QTL region on chromosome 6 in Nordic Dairy cattle; colour code is to show the linkage disequilibrium with the lead SNP.

Table 1. The segregation of the high-impact variants located within two candidate causal genes in 1000 bulls genome sequence without and within the Nordic Red cattle (RDC) sequences.

Variants	Gene	All except RDC			RDC		
		HOM-A1	HET	HOM-A2	HOM-A1	HET	HOM-A2
6:37201641	*FAM184B*	0	0	2,913	1	9	153
6:37401110	*LCORL*	2	28	2,882	0	0	164
6:37401770	*LCORL*	643	253	1,996	4	24	136
6:37402602	*LCORL*	0	2	2,916	0	32	129
6:37403627	*LCORL*	2	12	2,892	0	0	162
6:37412364	*LCORL*	1	39	2,866	0	2	160

We fine-mapped a major QTL affecting young-stock survival in Nordic Red dairy cattle. The lead SNP is located in a gene desert, indicating a regulatory variant may be the causal factor for this QTL. The high-impact variants in neighbouring genes were also studied. These variants are included in a custom made SNP array used in routine genotyping.

Acknowledgements

We are grateful to the Nordic Cattle Genetic Evaluation (NAV, Aarhus, Denmark) for providing the phenotypic data used in this study and Viking Genetics (Randers, Denmark) for providing samples for genotyping. This work was supported by research projects funded by the Milk Levy Fund (Mælkeafgiftsfonden, Skejby, Aarhus), Green Development and Demonstration Programme (GUDP) project from the Ministry of Environment and Food of Denmark, Viking Genetics, and NAV. The 1000 Bull Genomes Project is acknowledged for sharing WGS data.

References

Ashburner, M., C. A. Ball, J. A. Blake, D. Botstein, H. Butler, *et al.* (2000) Nat Genet 25(1):25-29. https://doi.org/10.1038/75556

Browning, B.L., Zhou Y., and Browning S.R. (2018) AJHG 103(3):338-348. https://doi.org/10.1016/j.ajhg.2018.07.015

Carlen E., Pedersen J., Pösö J., Eriksson J-Å., Nielsen U.S., *et al.* (2016) Interbull Bulletine No. 50. Puerto Varas, Chile. https://journal.interbull.org/index.php/ib/article/view/1408/1471

Chen, N. (2004) Curr Proto Bioinformatics 5(1):4.10.11-14.10.14. https://doi.org/10.1002/0471250953.bi0410s05

De Bustos, A., Cuadrado, A. and Jouve, N. (2016) Sci Rep 6, 36665. https://doi.org/10.1038/srep36665

Delaneau, O., Marchini J., and Zagury J. F. (2011) Nat Methods 9(2):179-181. https://doi.org/10.1038/nmeth.1785

Fuchsberger, C., Abecasis G.R., and Hinds D.A.. (2015) Bioinformatics 31(5):782-784. https://doi.org/10.1093/bioinformatics/btu704

Kang, H.M., Sul J.H., Service S.K., Zaitlen N.A., Kong S.Y., *et al.* (2010) Nat Genet 42(4):348-354. https://doi.org/10.1038/ng.548

Kern, C., Wang Y., Xu X., Pan Z., Halstead M., *et al.* (2021) Nat Commun 12(1):1821. https://doi.org/10.1038/s41467-021-22100-8

McLaren, W., Gil L., Hunt S.E., Riat H.S., Ritchie G.R., A. *et al.* (2016) Genome Biol 17(1):122. https://doi.org/10.1186/s13059-016-0974-4

Mesbah-Uddin, M., Guldbrandtsen B., Iso-Touru T., Vilkki J., De Koning D.J., *et al.* (2018) DNA Res 25(1):49-59. https://doi.org/10.1093/dnares/dsx037

Pruim, R.J., Welch R.P., Sanna S., Teslovich T.M., Chines P.S., *et al.* (2010) Bioinformatics 26(18):2336-2337. https://doi.org/10.1093/bioinformatics/btq419

Wu, X., Guldbrandtsen B., Nielsen U.S., Lund M.S., and Sahana G. (2017) JDS 100(8):6356-6370. https://doi.org/10.3168/jds.2017-12688

Yang, J., Lee S.H., Goddard M.E., and Visscher P.M. (2011) AJHG 88(1):76-82. https://doi.org/10.1016/j.ajhg.2010.11.011

374. Multibreed sequence level genome-wide association study of semen traits in tropical Australian cattle

W.L.A Tan[1*], L.R. Porto-Neto[2], A. Reverter[2] and M.R.S. Fortes[1]

[1]The University of Queensland, School of Chemistry and Molecular Biosciences, 68 Cooper Rd, The University of Queensland, Brisbane, QLD 4072, Australia; [2]CSIRO Agriculture and Food, 306 Carmody Road, Brisbane, QLD 4067, Australia; weiliangandre.tan@uq.net.au

Abstract

Male fertility affects the productivity and profitability of the beef cattle industry. Semen traits such as Percent Normal Sperm (PNS) are heritable indicators of bull fertility. These traits are collected in routine fertility assessments. Genome-Wide Association Studies (GWAS) were conducted in this study, using sequence level data to ascertain regions associated with PNS and two sperm abnormality measures in a multibreed population of 6,422 bulls. We report regions on chromosome 5 that are associated with PNS and Proximal Droplets (PD). Additionally, the X chromosome consistently contains important regions for all traits examined. Downstream analyses are required to identify candidate genes with a causal effect on semen traits.

Introduction

Bull fertility is an important trait for the beef cattle industry globally. However, it has gathered less consideration than its cow counterpart, which has benefited from advancements in biotechnology and selective breeding (Butler *et al.* 2019; Hayes *et al.* 2018). Semen traits are heritable indicators of bull fertility and can be used for genetic improvement (Butler *et al.* 2019). Previous GWAS have identified candidate genes for semen traits (Fortes *et al.* 2013). Genetic variants located through GWAS are often not causal mutations but Single Nucleotide Polymorphisms (SNP) in linkage disequilibrium to causal variants (Tenghe *et al.* 2016). Markers utilised in SNP panels consist of a small subset of common SNPs that segregate in the population. Advancement in next-generation sequencing and imputation methodologies allow animals genotyped with lower density SNP panels to be accurately imputed to sequence level variants (Browning and Browning 2009; Das *et al.* 2016). This preliminary GWAS uses imputed sequence level genotypes to identify regions associated to three semen traits, which may be incorporated in genomic predictions to improve accuracy.

Materials & methods

Animals and phenotypes. A multibreed dataset of 6,422 animals was used in this study. These animals are from two research populations and four stud herds. The two research populations consisted of animal data obtained from the Cooperative Research Centre for Beef Genetic Technologies (Beef CRC), consisting of 1,051 Brahman and 1,819 Tropical Composite (TC) bulls. The other four stud herds consisted of 1,288 Santa Gertrudis bulls, 760 Droughtmasters, 844 Ultra blacks, and 660 Belmont Red Composites. The three semen phenotypes used in this study are PNS (%) and two sperm abnormality measures: Mid-Piece abnormalities (MP, %) and PD (%).

Genotypes, genomic relationships, and GWAS. Animals were genotyped using a commercial 50K SNP chip and imputed to high density (~700K) using a phased reference generated by Eagle 2 (v2.4.1) and imputed using Minimac3 for autosomes and Minimac4 for the X chromosome (Das *et al.* 2016; Loh *et al.* 2016). The same method was used to impute to sequence level. SNPs with imputation r^2>0.8 in Minimac, a call rate >0.9 and a minor allele frequency >0.05 were kept for further analysis, leaving 21,804,882 SNPs, including 233,066 SNPs mapped onto the X chromosome after quality control. A high-density SNP panel (657,563 SNPs for autosomes and 22,775 SNPs for X) was used to build a genomic relationship matrices

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_374, © W.L.A Tan *et al.* 2022

(GRM) in GCTA (Yang *et al.* 2011), for the autosomes, which includes the Pseudo – Autosomal Region (PAR), whereas the X chromosome GRM was constructed separately using the non-PAR, as described by Yang *et al.* (2010). The Principal Components Analysis (PCA) was calculated using PLINK1.9 (Chang *et al.* 2015), and GRMs were used to account for genetic structure in the data. GWAS was performed using the Leave One Chromosome Out (LOCO) method in GCTA (Yang *et al.* 2011), as described by Yang *et al.* (2014). The phenotypes were pre-adjusted for their fixed effects (year of birth, breed, property) and covariates (age at measurement, PC1 and PC2). A genome-wide significance threshold of 5×10^{-8} was used, which is a conservative Bonferroni correction. Significant markers were merged into a single QTL region using bedtools (Quinlan and Hall 2010), and genes near each trait's strongest associated region were reported using data from Ensemble (Howe *et al.* 2021).

Results

The summary statistics of each semen trait and heritability is shown in Table 1. The results of the GWAS for each trait are represented in (Figure 1). A total of 100 genes for PNS, 84 genes for PD and 16 genes for MP were located in associated windows.

Table 1. Summary statistics and summary of GWAS results.

Phenotypes[1]	n	Mean	SD	LOCO[2]	h^2 (SE)[3]
PNS	6,055	61.76	27.53	9,677	0.18 (0.02)
PD	6,052	13.50	19.96	2,874	0.15 (0.02)
MP	6,052	11.39	11.04	603	0.13 (0.02)

[1] PNS % of normal sperm, PD % sperm with proximal droplets, MP % sperms with mid-piece abnormalities.

[2] Number of SNPs associated at $P<5\times10^{-8}$ in GWAS.

[3] SNP-derived heritability and standard error.

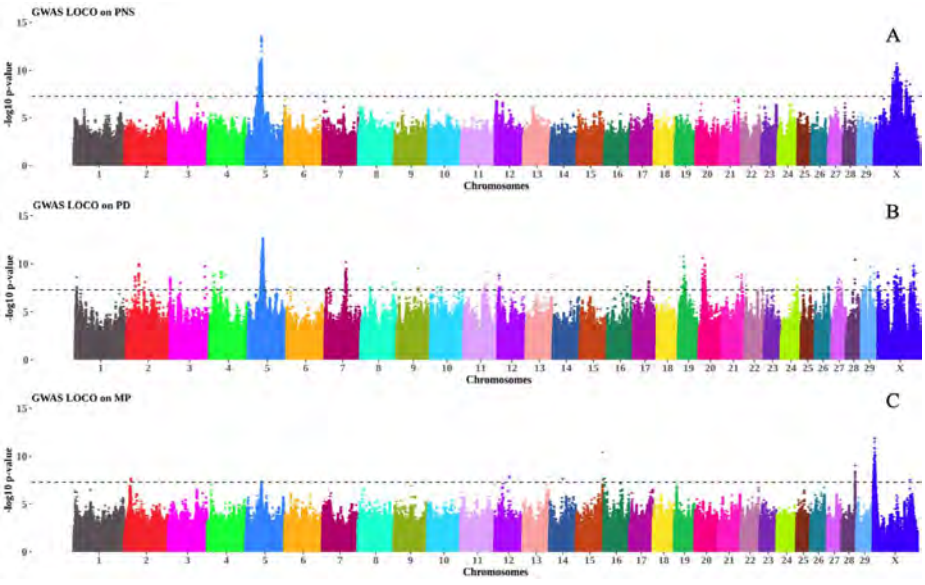


Figure 1. Manhattan plots showing associations identified for all three semen traits. Manhattan plots for PNS, PD and MP are labelled as A B and C, respectively. The inverse log *P*–values for each SNP are plotted along the y-axis for each chromosome on the x-axis. The dotted line represents the genome-wide significance threshold of 5×10^{-8}.

GWAS on PNS. A total of 9,677 significant associations were detected with PNS (Table 1). The largest concentration of SNPs associated with PNS was found between 35.1 Mb to 49.3 Mb at chromosome 5. The strongest genomic association (P=3.18×10^{-14}) for PNS was located at 46.4 Mb at chromosome 5. This highly associated region coded for DYRK2 and CAND1. The largest concentration of SNPs in the X chromosome was located between 58.2 Mb and 111.9 Mb, with the strongest genomic association (P=1.8×10^{-11}) located at 70.5 Mb. A significant SNP was also found at chromosome 12 at 3.83 Mb.

GWAS on PD. A total of 2,874 significant associations were detected for PD (Table 1). As seen in Figure 1, many regions throughout the genome were found to be associated with PD. The largest concentration of SNPs associated with PD was found between 38.7 Mb and 88.9 Mb of chromosome 5. The strongest genomic association (P=1.91×10^{-13}) was located at 46 Mb at chromosome 5. This highly associated region coded for DYRK2 and CAND1. In the X chromosome, several peaks were found at 1.5 Mb, 62.1 Mb and 111.9 Mb, as shown in Figure 1.

GWAS on MP. A total of 603 significant associations were detected for MP (Table 1). The largest concentration of SNPs was located between 2.7 Mb to 8.6 Mb of the X chromosome (Figure 1). The strongest genomic association (P=1.34×10^{-12}) was located at 6.27 Mb of the X chromosome. We did not find any genes in close proximity to this marker.

Discussion

The heritability estimates for PNS reported by Fortes *et al.* (2020) for TC (0.29) and Brahman (0.35) were higher than the 0.18 reported here. Similarly, the heritability estimates for PD reported by Fortes *et al.* (2020) for Brahman (0.35) and TC (0.21) were higher than the 0.15 reported here. Lastly, the heritability estimate for MP was higher than those reported for Brahman (0.07) but lower than those reported for TC (0.21) (Fortes *et al.* 2020). There were two genes in the regions of chromosome 5 that were strongly associated with PNS and PD. The DYRK family of kinases has been associated with rat spermatogenesis and is highly expressed from steps 6 to 11 in spermatids (Sacher *et al.* 2007). The CAND1 protein is highly expressed in the spermatozoa of fertile men (Park and Pang 2021). These regions may affect bull spermatogenesis and morphology. Several regions in the X chromosomes have been consistently associated with all three semen traits in this study. In a GWAS with TC bulls, Fortes *et al.* (2013) reported significant SNPs in the X chromosome, which overlaps with the region reported in this study. The highest association with PNS was located at 39.9 Mb (Fortes *et al.* 2013). Fortes *et al.* (2020) reported Quantitative Trait Loci (QTL) for PNS between 45.2 Mb and 49.7 Mb of the X chromosome in Brahman bulls. However, this QTL did not overlap with the regions reported in this study. Given that this is an across-breed GWAS, certain results that are significant in within-breed GWAS can remain hidden (Fortes *et al.* 2020). Further studies can examine GWAS results within-breeds that could locate breed-specific QTLs. For PD, Fortes *et al.* (2020) reported a QTL between 1.3 Mb and 8.5 Mb of the X chromosome of TC bulls that overlaps with the peak located at 1.5 Mb identified in this study. For MP, the most significant region in the X chromosome located at 6.27 Mb overlaps with the QTL between 4 Mb and 7.3 Mb in Brahman bulls reported by Fortes *et al.* (2020). Mueller *et al.* (2013) previously highlighted that species divergence genes in the X chromosome in mice and humans are highly expressed in the testis. The large number of significant SNPs that map to the X chromosome is consistent with the hypothesis by Mueller *et al.* (2013), which suggested the X chromosome as an intriguing region to look for genes associated with spermatogenesis and male fertility. Fortes *et al.* (2020) highlighted the theories of sex-chromosome meiotic drive or sexual antagonism as a possible explanation for the large number of genes associated with germ-cell development traits, with the latter being the more probable hypothesis in their study. While this hypothesis may plausible in populations with relatively low Minor Allele Frequency (MAF) of QTL SNPs, some markers identified as most significant for each QTL in our study had relatively higher MAF of more than 0.4 and thus this theory would be unlikely

in our context. Further research is still required to ascertain which theory is more probable. Our research uncovered important QTLs for semen traits which warrant further investigation in downstream analyses. If candidate genes with causal effect can be located, incorporation of these genes in genomic predictions may help improve accuracy.

References

Browning B. L., and Browning S. R. (2009). Am. J. Hum. Genet. 84(2):210-223. https://doi.org/10.1016/j.ajhg.2009.01.005

Butler M. L., Bormann J. M., Weaber R. L., Grieger D. M., and Rolf M. M. (2019) Transl. Anim. Sci. 4(1):423-441. https://doi.org/10.1093/tas/txz174

Chang C. C., Chow C. C., Tellier L C.A.M., Vattikuti S. *et al.* (2015) GigaScience 4(1). https://doi.org/10.1186/s13742-015-0047-8

Das S., Forer L., Schönherr S., Sidore C. *et al.* (2016) Nat. Genet. 48:1284-1287. https://doi.org/10.1038/ng.3656

Fortes M. R. S., Reverter A., Kelly M., McCulloch R., and Lehnert S. A. (2013) Andrology 1(4):644-650. https://doi.org/10.1111/j.2047-2927.2013.00101.x

Fortes M. R. S., Porto-Neto L. R., Satake N., Nguyen L. T. *et al.* (2020) Genet. Sel. Evol. 52:46. https://doi.org/10.1186/s12711-020-00563-5

Hayes B. J., Corbet N. J., Allen J. M., Laing A. R. *et al.* (2018) J. Anim. Sci. 97(1):55-62. https://doi.org/10.1093/jas/sky417

Howe K. L., Achuthan P., Allen J., Allen J. *et al.* (2021) Nucleic Acids Res. 49(1):D884-D891
https://doi.org/10.1093/nar/gkaa942

Loh P., Danecek P., Palamara P. F., Fuchsberger C. *et al.* (2016) Nat. Genet. 48:1443-1448. https://doi.org/10.1038/ng.3679

Mueller J. L., Skaletsky H., Brown L. G., Zaghlul S. *et al.* (2013) Nat. Genet. 45:1083-1087. https://doi.org/10.1038/ng.2705

Park Y., and Pang M. (2021) Antioxidants 10(1):98 https://doi.org/10.3390/antiox10010098

Quinlan A. R., and Hall I. M. (2010) Bioinformatics 26(6):841-842. https://doi.org/10.1093/bioinformatics/btq033

Sacher F., Möller C., Bone W., Gottwald U., Fritsch M. (2007) Mol. Cell. Endocrinol. 267:80-88. https://doi.org/10.1016/j.mce.2006.12.041

Tenghe A. M. M., Bouwman A. C., Berglund B., Strandberg E. *et al.* (2016) J. Dairy Sci. 99(7):5470-5485. https://doi.org/10.3168/jds.2015-10533

Yang J., Benyamin B., McEvoy B. P., Gordon S. *et al.* (2010) Nat. Genet. 42:565-569. https://doi.org/10.1038/ng.608

Yang J., Lee S. H., Goddard M. E., and Visscher P. M. (2011) Am. J. Hum. Genet. 88(1):76-82. https://doi.org/10.1016/j.ajhg.2010.11.011

Yang J., Zaitlen N. A., Goddard M. E., Visscher P. M., and Price A. L. (2014) Nat. Genet. 46:100-106. https://doi.org/10.1038/ng.2876

375. Improved genome-wide associations using a breed-specific 200K SNP chip for German Black Pied (DSN) cattle

P. Korkuć[1*], G.B. Neumann[1], D. Arends[1], M.J. Wolf[2], K. May[2], S. König[2] and G.A. Brockmann[2]

[1]Humboldt University Berlin, Albrecht Daniel Thaer- Institute for Agricultural and Horticultural Sciences, Animal Breeding Biology and Molecular Genetics, Invalidenstr. 42, 10115 Berlin, Germany; [2]Justus-Liebig-University of Gießen, Institute of Animal Breeding and Genetics, Ludwigstr. 21, 35390 Gießen, Germany; paula.korkuc@hu-berlin.de

Abstract

The DSN200K chip was specifically designed for German Black Pied cattle (DSN 'Deutsches Schwarzbuntes Niederungsrind') to improve genome-wide association analyses (GWAS) in DSN. Here, we performed GWAS with 2,099 DSN cows and 20 milk performance traits using 131,273 SNPs from the DSN200K chip. We compared those GWAS results to results obtained with the Illumina BovineSNP50 BeadChip (Illumina 50K) and *in silico* randomly generated SNP chips. Those random SNP chips had the same number of SNPs as on the DSN200K chip selected from a prefiltered set of 2.6 million sequence variants. We identified 13-15 times more associated SNPs with the DSN200K chip than with the Illumina 50K and 2.7-3.2 times more than with random SNP chips. With the DSN200K chip, we associated DSN-unique SNPs and SNPs with impact on gene transcripts which could potentially be causal mutations. These SNPs were seldom identified with the Illumina 50K or the random SNP chips.

Introduction

German Black Pied cattle are an endangered dual-purpose cattle breed, which is considered to be one of the ancestor populations of the modern Holstein breed. In order to track DSN-specific sequence variants that could help to conserve the specific genetic diversity of DSN and to allow for genetic improvement, we designed a breed-specific chip for DSN cattle (DSN200K chip) with 182,154 sequence variants (Neumann *et al.*, 2021).

Previous genome-wide association analyses (GWAS) in DSN were performed using genotypes from the Illumina BovineSNP50 BeadChip (Illumina 50K) (Korkuć *et al.*, 2021; May *et al.*, 2019; Meier *et al.*, 2020). The Illumina 50K chip was designed based on commercial breeds (Matukumalli *et al.*, 2009). Only 37K out of all 54K SNPs on the Illumina 50K chip were informative in DSN considering a SNP call rate>0.95 and a minor allele frequency (MAF)>0.05 and thus suitable for GWAS. Due to the ascertainment bias of the chip design, DSN-unique SNPs are missing on this chip. The designed DSN200K chip provides not only a higher density of SNPs but includes also DSN-unique SNPs and SNPs from previous GWAS in DSN. Moreover, it contains also SNPs of general interest such as SNPs with impact on gene functions and SNPs from the Illumina 50K chip. Additionally, SNPs were included that track most haplotype blocks found in DSN.

In this study, we performed GWAS with 2,099 DSN cattle and 20 milk performance traits using 131,273 SNPs from the DSN200K chip. We compared those results to results obtained with the lower-density Illumina 50K chip and *in silico* randomly generated SNP chips having the same number of SNPs as the original DSN200K chip in order to test whether only the number of SNPs or also the selection of SNPs on the DSN200K chip improve GWAS results.

Materials & methods

Genotypes and traits. Genotypes of 2,099 DSN cows from 8 farms in Germany were available from either Illumina 50K (1,473 cows) Korkuć *et al.*, 2021), DSN200K chip (453 cows) or whole-genome sequencing

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_375, © P. Korkuć *et al.* 2022

(173 cows) (Neumann *et al.*, 2021). For GWAS, genotypes were imputed to whole-genome sequencing (WGS) level using 304 sequenced DSN cattle (Neumann *et al.*, 2021) with BEAGLE v5.1 and then reduced to the SNP positions of the DSN200K and the Illumina 50K chip that were also available on the DSN200K chip. SNPs with MAF<0.05 and indels were removed, leading to 131,273 and 33,546 SNPs on the autosomes and chromosome X suitable for analyses on the DSN200K and the Illumina 50K chips, respectively. Milk performance data at 305 days was obtained from 'Vereinigte Informationssysteme Tierhaltung w.V.' as of February 2021. Traits included milk, fat, and protein yield in kilogram and fat and protein content in percent for the first three lactations and the lactation mean when data for all three lactations was available. Only full lactations with at least 270 days in milk were considered.

Genome-wide association analysis. Multiple linear regression models implemented in R were used to test the additive effect of each SNP. For each trait *Y*, the model included covariates for population stratification *ps,* farm *f*, sire *s*, birth year *by*, birth season *bs*, calving year *cy*, calving season *cs*, and age at first calving in days *ac*, together with the SNP genotype *gt* and the residual error *e:*

$$Y = ps + f^{*} + s^{*} + by^{*} + bs^{*} + cy^{*} + cs^{*} + gt + e \quad (1)$$

Population stratification *ps* was estimated using the pairwise population concordance test in PLINK v1.9 which is based on a pairwise identity-by-state matrix. The *P*-value cut-off of 0.0001 resulted in 65 clusters of relatedness. Covariates marked with an asterisk '*' were only included in the model when the difference in Akaike information criterion was ≤-10 between the null model ($Y = ps$) and the null model extended with one of the covariates ($Y = ps + covariate\ x$). To reduce the number of false positives in GWAS, *P*-values of each trait were corrected to an inflation factor of λ=1.3 on autosomes and chromosome X, separately. The number of independent SNPs (n=95,140) was estimated using LD-pruning SNPs with r^2>0.8 using PLINK v1.9 and used to adjust *P*-values for multiple testing with Bonferroni correction. After correction, SNPs were considered as significant when *P*<0.05 or suggestive when *P*<0.1. Associated SNPs within ±5 Mb across all traits were combined to quantitative trait loci (QTLs).

***In silico* random SNP chips.** For the design of the DSN200K chip, a set of 3.1 million quality-controlled sequence variants was used (Neumann *et al.*, 2021) which were filtered for MAF<0.05 and indels. From the left 2.6 million SNPs, 100 random SNP chips with 131,273 SNP positions were generated *in silico* by keeping all 33,546 informative SNPs from the Illumina 50K chip and selecting additional 97,727 SNPs randomly and uniformly across all chromosomes. The SNP distribution across chromosomes on the original and random SNP chips was similar. The mean distance between adjacent SNPs was 19,990±25,439 SD for the original and 19,972±26,148 SD for the randomly generated SNP chips. GWAS was repeated with those 100 random SNP chips as described above.

SNP annotation and haplotype blocks. The SNP selection categories for the design of the DSN200K chip and the estimation of haplotype blocks were thoroughly described in Neumann *et al.* (2021). SNP categories included SNPs from previous GWAS in DSN (Korkuć *et al.*, 2021; May *et al.*, 2019; Meier *et al.*, 2020; Wolf *et al.*, 2021), SNPs with high, moderate or low impact on gene transcripts (e.g. missense mutations) as estimated with Ensembl Variant Effect Predictor, and DSN-unique SNPs.

Results

GWAS with the DSN200K chip identified a total of 163 (131 significantly and 32 suggestively) associated SNPs (Table 1). Those SNPs covered 81 out of possible 87,266 haplotype blocks and could be grouped into 25 QTLs (61 haplotype blocks; 13 QTLs when considering only significant SNPs). The identified associated SNPs were mainly SNPs that were added to the DSN200K chip as significant SNPs from previous GWAS

results. But additional SNPs with high, moderate or low impact on the function of genes and DSN-unique SNPs that are important to identify breed-specific QTLs in DSN were also found. Many of the associated SNPs that had been found already in previous GWAS in DSN, had a potential impact on the gene transcript and/or were DSN-unique.

Only 13 (9 significantly and 4 suggestively) associated SNPs were identified when using the 33,546 SNPs from the Illumina 50K chip which were also present on the DSN200K chip. These were 13-15 times less associated SNPs than with the DSN200K chip while using only 4 times less SNPs. With regard to QTLs, 11 out of 25 would not have been detected when considering only SNPs from the Illumina 50K chip (5 out of 13 QTLs when considering only significant SNPs). Only 12 haplotype blocks were covered by those SNPs (8 when considering only significant SNPs).

GWAS with the randomly generated SNPs chips identified on average 60±8 SD associated SNPs (41±7 SD when considering only significant SNPs) (Table 1). This was around 2.7-3.2 times less SNPs than with the breed-specific DSN200K chip. Also, on average 1.6-1.7 times less haplotype blocks were covered. With regard to QTLs, with the randomly generated SNP chips on average 17±1 SD out of 25 QTLs that were found with the original DSN200K chip could be identified as well (11±1 SD out of the 13 QTLs when considering only significant SNPs). Almost no SNPs with impact on the gene transcripts or unique to DSN were identified with the randomly generated SNP chips.

The QTL with the highest significance using the DSN200K chip was found on chromosome 5 for milk fat content in lactation 1 (Figure 1). The top SNP of this QTL (5:93,516,066, rs211210569) is located in the gene *MGST1* (Microsomal Glutathione S-Transferase 1). This region was found to be as well associated with milk fat content in Holstein (Littlejohn *et al.*, 2016). This QTL would have been missed when considering only significant SNPs from the Illumina 50K chip, but would be likely found with any SNP chip of higher density.

Discussion

Higher density SNP chips (DSN200K and random SNP chips with 137K informative SNPs) identified as expected remarkably more associated SNPs that could be grouped also into more QTLs compared to the Illumina 50K chip. The higher number of associated SNPs in each QTL region improved also their

Table 1. Comparison of GWAS results using different SNP chips.

Chip	Threshold	Number of SNPs[1]				Number of haplotype blocks
		Previous GWAS	VEP impact	DSN-unique	Total[2]	Total[2]
DSN200K	MAF>0.05	1,898	40,120	7,295	131,273	87,266
	P<0.1	99	16	8	163	81
	P<0.05	83	10	8	131	61
Illumina 50K	MAF>0.05	134	1,016	0	33,546	26,615
	P<0.1	8	0	0	13	12
	P<0.05	6	0	0	9	8
Random SNP chips[3]	MAF>0.05	19±8	2,491±36	294±17	131,273±0	56,382±98
	P<0.1	12±2	1±1	0±1	60±8	49±6
	P<0.05	9±2	0±1	0±1	41±7	35±5

[1] SNPs can be assigned to multiple categories.

[2] Each SNP or haplotype block was counted only once.

[3] Values are listed including their ± standard deviations.

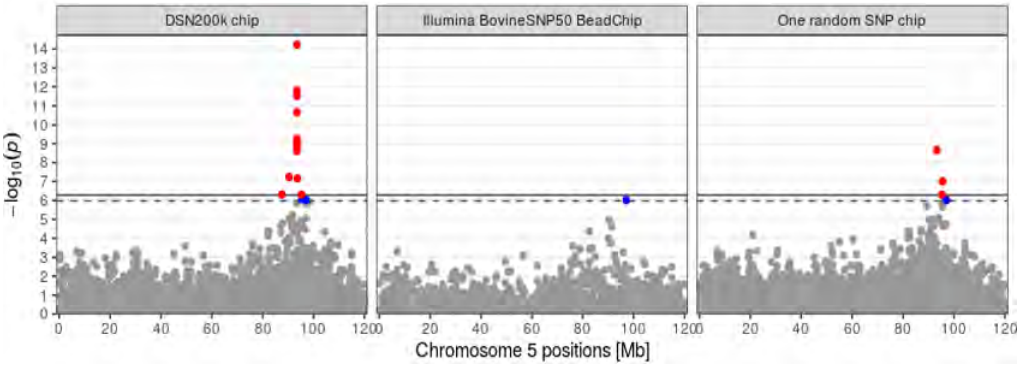


Figure 1. QTL on chromosome 5 for milk fat content in lactation 1 using different SNP chips (red=significant SNPs, blue=suggestive SNPs).

certainty. Furthermore, the carefully designed breed-specific DSN200K chip was better at identifying associated SNPs and QTLs with milk performance in DSN than the randomly generated SNP chips. This was expected, as this chip already contained previously identified SNPs associated with milk performance in a smaller population of 1,490 DSN cows (Korkuć *et al.*, 2021). With regard to SNP categories, SNPs having impact on gene transcripts or breed-specific SNPs were underrepresented on the randomly generated SNPs chips. Because these functional SNPs have a high potential for being causal for economically important phenotypes, they should be included on breed-specific SNP chips. Our findings support the fact that adding significant SNPs from GWAS and SNPs with functional impact or unique to the investigated breed to a custom SNP chip is beneficial for the target breed for future GWAS and genomic breeding. In conclusion, we obtained not only quantitatively but also qualitatively best results with the breed-specific DSN200K chip compared to the lower density Illumina 50K chip and to the randomly generated SNP chips.

References

Korkuć, P., Arends, D., May, K., König, S., Brockmann, G. A. (2021). Front. Genet. 12:275, https://doi.org/10.3389/fgene.2021.640039

Littlejohn, M. D., Tiplady, K., Fink, T. A., Lehnert, K., Lopdell, T., *et al.* (2016). Sci. Rep. 6(1):1-14, https://doi.org/10.1038/srep25376

Matukumalli, L. K., Lawley, C. T., Schnabel, R. D., Taylor, J. F., Allan, M. F., *et al.* (2009). PLoS One 4(4):e5350. https://doi.org/10.1371/journal.pone.0005350.

May, K., Scheper, C., Brügemann, K., Yin, T., Strube, C., *et al.* (2019). BMC Genomics 20(1):1-15. https://doi.org/10.1186/s12864-019-5659-4

Meier, S., Arends, D., Korkuć, P., Neumann, G. B., Brockmann, G. A. (2020). J. Dairy Sci. 103(11):10289-10298, https://doi.org/10.3168/jds.2020-18209

Neumann, G. B., Korkuć, P., Arends, D., Wolf, M. J., May, K., *et al.* (2021). BMC Genomics 22(1):1-13. https://doi.org/10.1186/s12864-021-08237-2

Wolf, M. J., Yin, T., Neumann, G. B., Korkuć, P., Brockmann, G. A., *et al.* (2021). Genes, 12(8):1163. https://doi.org/10.3390/genes12081163

376. Haplotype-based single-step GWAS for yearling temperament in American Angus

A.C. Araujo[1,2*], P.L.S. Carneiro[3], A.B. Alvarenga[2], H.R. Oliveira[2,4], S.P. Miller[5], K. Retallick[5] and L.F. Brito[2]

[1]State University of Southwestern Bahia, Praça Primavera 40, 45700-000 Itapetinga, Brazil; [2]Purdue University, 270 S. Russel Street, West Lafayette, IN 47907, USA; [3]State University of Southwestern Bahia, Rua Jose Moreira Sobrinho s/n, 45206-190 Jequie, Brazil; [4]University of Guelph, 50 Stone Road East, N1G2W1, Guelph, Canada; [5]American Angus Association, Angus Genetics Inc., 3201 Frederick Ave, St. Joseph, MO 64506, USA; araujoa@purdue.edu

Abstract

Understanding the genetic architecture of behavioral traits is of great interest for the livestock sector due to its economic and welfare implications. Therefore, the objective of this study was to perform traditional and weighted single-step genome-wide association studies (GWAS) for yearling temperament (YT) in North American Angus cattle using haplotypes. Approximately 266K YT records and 70K animals genotyped using a 50K single nucleotide polymorphisms (SNPs) panel were used. Pseudo-SNPs from the unique haplotype alleles generated from haplotype blocks with low (0.15), moderate (0.50), and high (0.80) linkage disequilibrium thresholds were used with and without fitting the non-clustered SNPs. Weighted single-step GWAS did not perform better than the traditional single-step GWAS. YT is a highly polygenic trait. The main candidate genes identified are involved in important biological processes and metabolic pathways related to behavioral traits, social interactions, and aggressiveness in cattle.

Introduction

Yearling temperament (YT) in North American Angus is a categorical trait measured on a scale from 1 to 6 for docile and very aggressive animals, respectively, and scored subjectively by farmers/handlers when a one-year-old calf is being processed through the chute and is a measure of cattle behaviour (Northcutt and Bowman, 2010). YT is heritable (heritability ~0.37) and improving cattle behaviour is paramount to the beef cattle industry, due to its relationship with other economically important traits (e.g. carcass end product) as well as animal handling and welfare (Northcutt and Bowman, 2010).

Haplotypes can better capture the linkage disequilibrium (LD) between markers and the quantitative trait loci (QTL) than SNPs and also account for epistatic effects (Hess *et al.*, 2017), which can improve QTL mapping. We have not found reports of haplotype-based GWAS for behavioural traits in cattle, and hypothesized that this approach could provide complementary information to dissect the genetic basis of cattle behaviour.

The traditional single-step GWAS (ssGWAS) consist in using phenotypes from genotyped and non-genotyped individuals to find QTL, however, it assumes that the markers explain similar and small proportion of the total additive genetic variance (Wang *et al.*, 2012). This limitation resulted in proposing the weighted ssGWAS (WssGWAS), in which priorities can be given to regions that explain greater proportions of the additive genetic variance (Wang *et al.*, 2012). Therefore, the objective of this study was to perform ssGWAS and WssGWAS for YT in North American Angus cattle using haplotypes.

Materials & methods

Datasets. No ethical review and approval were needed for this study because all datasets used were provided by commercial breeding operations. The American Angus Association (through Angus Genetics

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_376, © A.C. Araujo *et al.* 2022

Inc.; St Joseph, MO, USA) provided YT records for 266,029 animals born between 2001 and 2018 (147,671 bulls, 3,332 steers, and 115,026 females), in which 69,559 of them also had genotypes for 54,609 SNPs. The pedigree dataset included 4,410,551 animals. We have used the YT measured in a scale from 1 to 4 for docile and aggressive animals, respectively, by grouping the classes 4 to 6 (Northcutt and Bowman, 2010). The genotypic dataset was filtered to remove SNPs with minor allele frequency <0.01, call rate <0.90, difference between observed and expected heterozygosity >0.15, located in the non-pseudo-autosomal region of the X chromosome, and animals with call rate <0.90, remaining 42,633 markers and 69,437 animals for further analyses.

Haplotype block construction. First, the SNP genotypes were phased using the FImpute software v.3.0 (Sargolzaei *et al.*, 2014). Then haploblocks were created with LD measured as r^2 using the 'gpart' R package (Kim *et al.*, 2019). The LD thresholds of 0.15, 0.50, and 0.80 were used to create the haploblocks. The total number of blocks was 9,634, 5,617, 2,721 and the total number of SNPs within blocks was 21,784, 12,132, and 5,822 for the LD thresholds of 0.15, 0.50, and 0.80, respectively.

Traditional and weighted single-step GWAS. The ssGWAS consisted in predicting the genomic estimated breeding values (GEBVs) using the single-step genomic BLUP, back solving the GEBVs to the marker effects, and then calculating the percentages of the total additive genetic variance (PTGV) explained by each one of them. The THRGIBBS1f90 software (Misztal *et al.*, 2018) was used to predict the GEBVs for YT considering it as a categorical trait under a Bayesian approach, using the following mixed model:

$$y = Xb + Ww + Zu + e \quad (1)$$

where y is the vector of phenotypic records, b is the vector of systematic effects (age of dam, conception type, and calf age deviation from 365 days as linear covariate), w is the vector of contemporary group (CG) effects with $w \sim N(0, I\sigma_w^2)$, u is the vector of additive genetic effects with $u \sim N(0, H\sigma_u^2)$, and e is the vector of residuals $e \sim N(0, I\sigma_e^2)$. The CG were formed by concatenating birth month and year, birth herd, birth sex, weaning date, weaning herd, weaning sex, creep feeding offered or not, date of temperament measurement, YT measurement herd, sex at the YT measurement, temperament group age deviation, and presence of ultrasound records. X, W, and Z are the known design matrices for the systematic, CG, and additive genetic effects, respectively. H is the matrix that combines the pedigree (A) and genomic (G) relationships, which was computed directly as H^{-1} as suggested by Aguilar *et al.* (2010). A chain of 10,000 iterations with 1000 discarded as burn-in and 10 as thin was used to predict the GEBVs. The POSTGSf90 software (Misztal *et al.*, 2018) was used for post-GWAS analyses. The WssGWAS was performed following the algorithm proposed by Wang *et al.* (2012) using the non-linear A weights proposed by VanRaden (2008). We performed two additional iterations in the WssGWAS beyond the ssGWAS because this value was enough to achieve the convergence of the GEBV accuracy (Wang *et al.*, 2012).

Scenarios evaluated. Eighteen scenarios were defined by combining: (1) ssGWAS or WssGWAS in the second and third iterations; (2) G calculated considering only pseudo-SNPs from unique haplotype alleles derived from haploblocks with LD threshold of 0.15, 0.50, and 0.80; and (3) same as (2) including the non-clustered SNPs (NCSNP) with the respective LD thresholds.

Empirical selection of the candidate regions and functional analysis. First, a probability distribution (Beta or Gamma) was fitted to the PTGV in each scenario, then a quantile with the tail containing the markers that explained the biggest PTGV (top) was defined for further investigation. The candidate regions were selected as the top 0.001% genomic regions and the Biomart tool (www.ensembl.org/biomart/martview/ad1112a783c0e0ae22e6572189d5bead) was used to identify candidate genes overlapping. Gene

ontology biological processes and metabolic pathways were annotated for the genes found using the DAVID platform (https://david.ncifcrf.gov/home.jsp) by the Functional Annotation Tables for each gene.

Results

The WssGWAS did not perform better than ssGWAS, which found the largest number of top 0.001% genomic regions for YT regardless of the LD threshold used, with the PTGV summing up to 4.21% of the total additive genetic variance. As all the top genomic regions found by the WssGWAS scenarios were present in the ssGWAS scenarios, only the results for the latter are presented (Figure 1) and were used to find overlapping genes (Table 1).

Discussion

The fact that WssGWAS did not perform better than ssGWAS indicates the highly polygenic architecture of YT, as ssGWAS assumes the infinitesimal model for the QTL (Wang *et al.*, 2012). A total of 76 genes were identified across all scenarios. The biological processes associated with some of the genes overlapped with the top 0.001% genomic regions are related to neuron projection development (GO:0031175); axonogenesis, visual perception, forebrain development, embryonic eye morphogenesis, and retina development in camera-type eye (GO:0007409, GO:0007601, GO:0030900, GO:0048048, and GO:0060041,

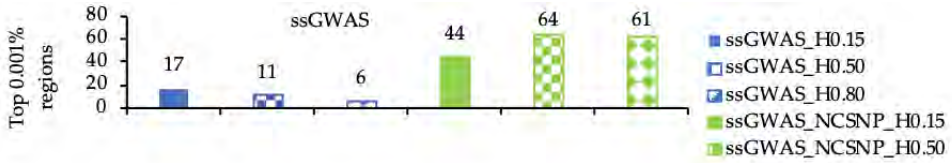


Figure 1. Number of top 0.001% genomic regions for yearling temperament in American Angus cattle found by traditional single-step GWAS (ssGWAS). H0.15, H0.50, and H0.80: only haplotypes from blocks with linkage disequilibrium (LD) thresholds of 0.15, 0.50, and 0.80, respectively; NCSNP_H0.15, NCSNP_H0.50, and NCSNP_H0.80: non-clustered SNPs (NCSNP) from blocks with LD thresholds of 0.15, 0.50, and 0.80, respectively.

Table 1. Gene ontology biological terms (GO_BP) and metabolic pathways from the Kyoto Encyclopedia of Genes and Genomes (KEGG) access of genes overlapped by top 0.001% markers for docility in North American Angus cattle[1].

Chromosome	Gene	S_P (Mb)[2]	E_P (Mb)[3]	GO_BP	KEGG
BTA5	*ATXN10*	116.029	116.169	GO:0031175	-
BTA10	*ADAM10*	51.536	51.679	-	bta05010
BTA11	*VAX2*	13.483	13.509	GO:0007409, GO:0007601, GO:0030900, GO:0048048, GO:0060041	-
BTA11	*ATP6V1B1*	13.454	13.480	GO:00076605, GO:0042472	-
BTA14	*CRISPLD1*	38.295	38.346	GO:0060325	-
BTA15	*CAPRIN1*	64.662	64.697	GO:0050775, G00061003	
BTA18	*FA2H*	2.151	2.206	GO:0032286, GO:0032287	-
BTA20	*SPEF2*	38.369	38.573	GO:0048702, GO:0048854, GO:0069541	-
BTA22	*PLXNA1*	60.240	60.280	GO:0021785, GO: 0048841, GO:1902287, GO:1990138	bta04360
BTA22	*CACNA2D3*	45.925	46.819	-	bta04921

[1] Only gene with a GO_BP or metabolic pathway related to behavioral traits are presented.

[2] Start position.

[3] End position.

respectively); sensory perception of sound and inner ear morphogenesis (GO:00076605 and GO:0042472, respectively); face morphogenesis (GO:0060325); positive regulation of dendrite and dendrite spine morphogenesis (GO:0050775 and GO0061003, respectively); central and peripheral nervous system myelin maintenance (GO:0032286 and GO:0032287, respectively); embryonic neurocranium, brain morphogenesis, and respiratory system development (GO:0048702, GO:0048854, and GO:0069541, respectively); branchiomotor neuron axon guidance, regulation of axon extensions involved in guidance, the semaphorin-plexin signaling pathway involved in axon guidance, and neuron projection extension (GO:0021785, GO: 0048841, GO:1902287, and GO:1990138, respectively). The three main metabolic pathways related to YT are related to Alzheimer disease, axon guidance, and oxytocin signalling pathway (bta05010, bta04360, and bta04921, respectively). In summary, they were mainly related to central and peripherical nervous system and sense organs, which affect social interactions between individuals because they are involved in capturing, decodifying, and reacting to external stimulus (Grandin and Deesing, 2014). Including the non-clustered SNPs and using different LD thresholds to create the haploblocks in the ssGWAS is recommended because it increased the likelihood of finding important genomic regions. The main candidate genes identified are involved in biological processes and metabolic pathways related to behavioural traits, social interactions, and aggressiveness in cattle.

References

Aguilar I., Misztal I., Johnson D.L., Legarra A., Tsuruta S., *et al.* (2010) J. Dairy Sci. 93(2):743–752. https://doi.org/10.3168/jds.2009-2730

Grandin T., and Deesing M.J. (2014) Behavioral genetics and animal science. In: Grandin T., Deesing M.J. (eds) Genetics and the Behavior of Domestic Animals, 2nd ed.; Academic Press: San Diego, CA, USA, 2014; pp. 1–40.

Hess M., Druet T., Hess A., Garrick D. (2017) Genet. Sel. Evol. 49:54. https://doi.org/10.1186/s12711-017-0329-y

Kim S.A., Brossard M., Roshandel D., Paterson A.D., Bull S.B., *et al.* (2019) Bioinformatics 35(21):4419–4421. https://doi.org/10.1093/bioinformatics/btz308

Misztal I., Tsuruta S., Lourenco D.A.L., Masuda Y., Aguilar I., *et al.* (2018). Manual for BLUPF90 Family Programs; University of Georgia, Athens, USA: Available at: http://nce.ads.uga.edu/897wiki/doku.php?id=documentation

Northcutt S., and Bowman B. (2010). By the Numbers: Docility Genetic Evaluation Research. Available at: http://www.angus.org/nce/documents/bythenumbersdocility.pdf

Sargolzaei M., Chesnais J.P., and Schenkel F.S. (2014). BMC Genomics 15:478. https://doi.org/10.1186/1471-2164-15-478

Vanraden P.M. (2008). J. Dairy Sci. 91(11):4414–4423. https://doi.org/10.3168/jds.2007-0980

Wang H., Misztal I., Aguilar I., Legarra A., Muir W.M. (2012). Genet. Res. 94(2):73–83. https://doi.org/10.1017/S0016672312000274

377. Genome-wide association analyses for various thermotolerance indicators measured during lactation in maternal-line pigs

P.H.F. Freitas[1], J.S. Johnson[2], H. Wen[1], J.M. Maskal[1], M.K. Byrd[1], F. Tiezzi[3], Y. Huang[4], A.P. Schinckel[1] and L.F. Brito[1*]

[1]Department of Animal Sciences, Purdue University, West Lafayette, IN 47906, USA; [2]USDA-ARS Livestock Behavior Research Unit, West Lafayette, IN, 47907, USA; [3]Department of Agriculture, Food, Environment and Forestry (DAGRI), University of Florence, 50144 Firenze, Italy; [4]Smithfield Premium Genetic, Rose Hill, NC 28458, USA; britol@purdue.edu

Abstract

Heat stress is a major welfare and economic issue in the swine industry. Genome-wide association studies (GWAS) represent a great tool to investigate the association of candidate genes and quantitative trait loci (QTL) with phenotypes of interest. Therefore, this study aimed to perform GWAS for six indicators of thermotolerance (automatically-measured vaginal temperature; ear, shoulder, rump, and tail temperatures; and respiration rate) in lactating sows to identify genomic regions and candidate genes influencing these traits in commercial Landrace × Large White sows during lactation stage under heat stress conditions. Sixteen genomic regions were identified on chromosomes 1, 2, 3, 5, 6, 7, 8, 10, 12, 13, 14, and 16, with ten of these genomic regions related with known QTL associated with health and production traits. These findings highlight the importance of thermoregulation indicators on the performance and well-being of pigs, and consequently adding such traits to the current selection indexes.

Introduction

Climate change-induced heat stress is considered a major welfare and production issue in the swine industry. Advances in genetic and genomic selection, nutrition, and management have increased pig performance through larger litter sizes (Kemp *et al.*, 2018), with a subsequent higher metabolic heat production (Stinn and Xin, 2014; Cabezón *et al.*, 2017). In general, swine breeding programs rely on a wide number of traits (e.g. litter characteristics, slaughter, and reproduction) integrated into selection indexes. This approach has led to the identification of traits that may assist in the selection of more heat tolerant animals.

Various efforts have been made in the identification of novel phenotypes under non-commercial conditions using simulated heat stress (Graves *et al.*, 2018; Kim *et al.*, 2018). In addition of the development of novel phenotypes, the identification of candidate genes and genomic regions related to thermotolerance traits based on genome-wide association studies (GWAS) is an important step to understand the genetic mechanisms underlying their phenotypic expression. However, there is still a need to further investigate the association of genomic regions and candidate genes with thermotolerance using larger datasets. Therefore, the study objective was to perform GWAS for novel indicators of thermotolerance aiming to identify key genomic regions and candidate genes influencing these traits in commercial Landrace × Large White lactating sows measured under heat stress conditions.

Materials & methods

Animals, phenotypic and genotypic dataset. A total of 1,639 multiparous lactating sows (Landrace × Large White; parity 4.27±1.53) and their litters (11.10±2.33 piglets/litter) were tested. Sows and their litters were housed under commercial production conditions within individual farrowing crates at a commercial sow farm in Maple Hill, North Carolina, USA (34.70738, -77.73653), and novel phenotypes associated with heat tolerance were measured between June 05, 2021 and July 30, 2021. Ear skin temperature (T_{ES}), shoulder skin temperature (T_{SS}), rump skin temperature (T_{RS}), and tail skin temperature (T_{TS}) were collected in

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_377, © P.H.F. Freitas *et al.* 2022

all sows throughout the length of the study, four times a day (at 08:00, 12:00, 16:00, and 20:00 hours) using an infrared temperature gun (Raytek model ST61; accuracy = ±1%; emissivity = 0.98; resolution = 0.10 °C; Raytek Corporation, Fluke Process Instruments, Everett, WA, USA). Vaginal temperature (T_V) was automatically measured in 10 min intervals using calibrated thermochron temperature recorders (iButton model DS1921H, calibrated accuracy ±0.01 °C; resolution = 0.125 °C; Dallas Semi-conductor, Maxim, Irving, TX, USA). Respiration rate (RR) was collected by counting flank movements for 15 s and multiplying by 4 to calculate breaths per minute at 08:00, 12:00, 16:00, and 20:00 hours daily (Johnson *et al.*, 2016; Kpodo *et al.*, 2019). In-barn environmental data were collected throughout the length of the trial using data loggers in each farrowing room (Hobo model #MX1101; data logger temperature/RH; accuracy ±0.20 °C and ±2% RH; Onset; Bourne, MA, USA). Each data logger was placed in 1 of 4 quadrants mounted at sow height to record dry bulb temperature and relative humidity, which were used to calculate dew point (Buck, 1981).

A total of 1,639 animals were genotyped using the PorcineSNP50K (50,703 SNPs) Bead Chip (Illumina, San Diego, CA, USA). Quality control (QC) of genotype data consisted of removing SNPs and animals with call rate below 0.90, and minor allele frequency lower than 0.01, using the BLUPF90 family software (Misztal *et al.*, 2018). After QC, 49,547 SNPs for 1,625 animals were kept for subsequent analyses.

Genome-wide association studies. Variance components and genomic estimated breeding values (GEBV) were estimated implementing repeatability models using the BLUPF90 family programs (Misztal *et al.*, 2018). The model fitted can be described as:

$$y = X\beta + Z_a a + Z_{pe} pe + e \quad (1)$$

where, y is the vector of phenotypic records for T_V, T_{ES}, T_{SS}, T_{TS}, T_{RS}, and RR; β is a vector of fixed effects, including parity, location, time of measurement, body size, scorer of body size, hair density, scorer of hair density; a is the random additive genetic effect assumed to follow $\mathbf{a} \sim N(0, \mathbf{G}\sigma_a^2)$. The G matrix was created based on the first method proposed by VanRaden (2008); p is the vector of random permanent environmental effects; X, Z_a, and Z_{pe} are the incidence matrices; and e is the residual effect assumed to follow $\mathbf{e} \sim N(0, \mathbf{I}\sigma_e^2)$.

After estimating the GEBVs, GWAS analyses were performed using the postGSf90 software (Aguilar *et al.*, 2014), considering the variance explained by moving windows of five adjacent SNPs. The GEBVs were back-solved to obtain the SNP effects following the equation:

$$\hat{\mathbf{u}} = \mathbf{Z}'[\mathbf{Z}\mathbf{Z}']^{-1}\hat{\mathbf{a}} \quad (2)$$

where û is the vector of estimated SNP effects for the additive genetic effect; Z is the matrix with genotypes for each SNP, and â is the vector of GEBVs. Genomic windows (±100 kb) were considered relevant if they explained ≥0.5% of the total additive genetic variance of the trait. Positional candidate genes were mapped using the Biomart tool (Kinsella *et al.*, 2011; http://useast.ensembl.org/index.html). Based on the start and end chromosomal positions, important genomic regions were further investigated to understand the biological processes related to the studied traits and to define the most likely functional candidate genes. All candidate genomic regions were searched in the Pig QTL Database (PigQTLdb; Hu *et al.*, 2018) to investigate whether they contained previously reported quantitative trait loci (QTL).

Results

Sixteen genomic windows were associated with the six different studied traits. These genomic regions were in the *Sus scrofa* (SSC) chromosomes SSC1, SSC2, SSC3, SSC5, SSC6, SSC7, SSC8, SSC10, SSC12, SSC13, SSC14, and SSC16, and the maximum proportion of the additive genetic variance explained was 1.16% (T_{ES}). Previously reported QTL within the relevant genomic windows are shown in Table 1. The main QTL are related to health (e.g. immune capacity) and production (e.g. feed conversion). Key candidate genes were found on these genomic regions. Tail skin temperature had the highest number of candidate genes (26 genes in 4 genomic regions), while T_{RS} had the lowest number of candidate genes (one gene in one genomic region). Some genes were commonly found between T_{SS}, T_{ES}, and T_{TS} (*CD209*, *CD320*, *ELAVL1*, *FBN3*, *FCER2*, *KANK3*, *PCP2*, *PRR36*, *TIMM44*, and *XAB2*). Genomic regions were commonly found between T_{ES} and T_{SS} (on SSC2), T_{RS} and T_{SS} (on SSC14), and T_{TS} and RR (on SSC8). =

Discussion

In this study, most of the candidate genomic regions (4) were found to be associated with T_{TS} and T_{SS}. The candidate genes located in these regions have been previously associated with health, feed efficiency, and production QTL, highlighting the importance of thermoregulation on the performance and well-being of animals raised under heat stress conditions. For instance, the gene *CD320* identified for T_{SS}, T_{ES}, and T_{TS} play an important role on reproduction. Mice lacking *CD320* were found to suffer from reproductive loss (Bernard *et al.*, 2018). Another interesting candidate gene (*ELAVL1*) is associated with immune response and was found for T_{SS}, T_{ES}, and T_{TS} (Karginov, 2019). Furthermore, the gene *TIMM44* (T_{SS}, T_{ES}, and T_{TS}) has key functions related to oxidative stress (Steibel *et al.*, 2011). From the sixteen relevant genomic regions, ten are associated with QTL related with health parameters, which shows the importance and relevance of the studied traits on a healthy and productive animal. The amount of candidate genes and genomic windows found to be relevant in this study indicates the complex nature of the studied traits.

Table 1. Previously reported quantitative trait loci (QTL) for related traits within candidate genomic regions.

Trait[1]	Genomic region (Mb)[2]	Related QTL
T_{ES}	SSC2: 70.61-71.56	Immune capacity, disease measurement.
T_{RS}	SSC14: 108.41-108.72	Mean corpuscular hemoglobin concentration, skeletal muscle morphology, response to bacterial infection.
T_{SS}	SSC1: 17.31-17.37	Red blood cell distribution width, blood glucose level.
	SSC2: 70.82-71.56	Blood immunoglobulin amount, disease process measurement, shoulder weight, liver weight.
	SSC3: 137.70-138.15	Disease process measurement, serum creatinine level.
	SSC14: 108.41-108.72	Average daily food intake, mean corpuscular hemoglobin concentration, skeletal muscle morphology, response to bacterial infection.
T_{TS}	SSC5: 62.12-62.20	Blood urea level, meat pH, outer ear area, outer ear prominence, outer ear size.
	SSC6: 149.37-149.50	Feed conversion ratio.
	SSC7: 38.82-39.02	Viable offspring quantity, hemoglobin concentration, blood potassium level.
	SSC8: 69.15-69.62	Disease process measurement, corpus luteum quantity.
T_V	SSC1: 58.45-58.58	Internal adipose amount, food intake rate.
	SSC14: 48.32-48.56	Abdominal fat weight, internal adipose amount, viable offspring quantity.
	SSC16: 21.58-21.70	Belly muscle percentage, outer ear area, immune response.
RR	SSC6: 15.37-15.41	Litter size, social interaction, hair pigmentation, meat pH, body weight, leg conformation.
	SSC8: 69.15-69.62	Corpus luteum number, feet weight, average daily gain, red blood cell count, uterine capacity.
	SSC12: 26.60-26.87	Response to viral infection, disease measurement, stress-related behaviour.

[1] T_{ES}: ear skin temperature; T_{SS}: shoulder skin temperature; T_{RS}: rump skin temperature; T_{TS}: tail skin temperature; T_V: vaginal temperature; and RR: respiration rate.
[2] bp: mega base pairs; SSC: *Sus scrofa* chromosome.

These findings are an important first step to understand the genetic background of close-to-biology indicators of thermotolerance in pigs. Further studies are needed to validate the relevant candidate genes and genomic regions found in this study.

Acknowledgements

This research was supported by the Agriculture and Food Research Initiative (AFRI) Competitive Grant number 2020-67015-31575 from the USDA National Institute of Food and Agriculture.

References

Aguilar I., Misztal I., Tsuruta S., Legarra A., and Wang H. (2014). Proc. 10th World Congr. Genet. Appl. Livest. Prod., Vancouver, Canada.

Bernard D. J., Pangilinan F. J., Cheng J., Molloy A. M., and Brody L. C. (2018). Human Molecular Genetics 27, 3627–3640. https://doi.org/10.1093/hmg/ddy267.

Buck A. L. (1981). J. Appl. Meteorol 20:1527-1532. https://doi.org/10.1175/1520-0450

Cabezón F. A., Schinckel A. P., Richert B. T., Peralta W. A., and Gandarillas M. (2017) Prof. Anim. Sci. 33:357–362. https://doi.org/10.15232/pas.2016-0158.

Graves K. L., Seibert J. T., Keating A. F., Baumgard L. H., and Ross J. W. (2018). Journal of Animal Science 96, 2419–2426. https://doi.org/10.1093/jas/skx037.

Hu Z.-L., Park C. A., and Reecy J. M. (2018). Nucleic Acids Research 47. https://doi.org/10.1093/nar/gky1084.

Johnson J. S., Sapkota A., and Lay D. C. (2016). Journal of Applied Physiology 120, 1249–1259. https://doi.org/10.1152/japplphysiol.00685.2015.

Karginov F. V. (2019). RNA Biology 16, 686–695. https://doi.org/10.1080/15476286.2019.1582954.

Kemp B., Da Silva C. L. A., and Soede N. M. (2018). Reprod. Dom Anim. 53, 28–36. https://doi.org/10.1111/rda.13264

Kim K.-S., Seibert J. T., Edea Z., Graves K. L., Kim E.-S., *et al.* (2018). Journal of Animal Science 96, 2074–2085. https://doi.org/10.1093/jas/sky131.

Kpodo K. R., Duttlinger A. W., and Johnson J. S. (2019). Translational Animal Science 3, 1375–1382. https://doi.org/10.1093/tas/txz033.

Misztal I., Tsuruta S., Lourenco D. A. L., Masuda Y., Aguilar I., *et al.* (2018) University of Georgia. Available at: http://nce.ads.uga.edu/wiki/doku.php?id=documentation

Steibel J. P., Bates R. O., Rosa G. J., Tempelman R. J., Rilington *et al.* (2011). PLoS ONE 6. https://doi.org/10.1371/journal.pone.0016766.

Stinn J. P., and Xin H. (2014). Trans. ASABE 57, 1517–1528. https://doi.org/10.13031/trans.57.10711

VanRaden P. M. (2008). Journal of Dairy Science 91, 4414–4423. https://doi.org/10.3168/jds.2007-0980.

378. Investigation of GWAS variants associated with loin depth in commercial pigs

S. Desire[1*], M. Johnsson[1,2], R. Ros-Freixedes[3], C. Chen[4], W. Herring[4], G. Gorjanc[1], R.J. Mellanby[5], J. Hickey[1] and M. Jungnickel[1]

[1]The Roslin Institute, The University of Edinburgh, Midlothian, United Kingdom; [2]Swedish University of Agricultural Sciences, Uppsala, Sweden; [3]Departament de Ciència Animal, Universitat de Lleida, Agrotecnio (CERCA Center), Lleida, Spain; [4]The Pig Improvement Company, Genus plc, Hendersonville, TN, USA; [5]The Royal (Dick) School of Veterinary Studies, The University of Edinburgh, Midlothian, United Kingdom; sdesire@ed.ac.uk

Abstract

Genome wide associate studies (GWAS) aim to identify genomic regions implicated in phenotype expression, but identifying true causative variants is difficult. Pig combined annotation dependent depletion (pCADD) scores may provide a straight-forward method of prioritising causal variants identified by GWAS. This study aimed to test using pCADD to identify candidate variants using loin depth in four pig lines. Three main regions on SSC7, 9, and 16 were identified in both maternal and terminal lines and have been found to be associated with carcass traits in previous studies. Two genomic regions were identified in maternal lines only, with SNPs within *MC4R* and *PIK3C2A* identified as the most likely causative variants. Several of the candidate SNPs have plausible biological links to muscle-related phenotypes, however further work is required to determine whether these are true causal variants.

Introduction

In genome-wide association studies (GWAS), genetic variants are independently tested for statistical association with a phenotype, with the aim of identifying the underlying causal genetic basis of the trait. Identifying true causal variants is difficult, due to high linkage disequilibrium (LD) between variants, a dilemma that is exacerbated in livestock species that have undergone selection for specific traits. While fine-mapping aims to solve this problem, there are many methods to choose from, with varying degrees of complexity. Pig combined annotation dependent depletion (pCADD) is an algorithm that estimates the deleteriousness of a given variant in the pig genome. As demonstrated by Derks *et al.* (2021), incorporating pCADD scores in the GWAS pipeline may provide a simple method for identification of candidate causative SNPs. The objective of this study was to use GWAS and pCADD to computationally identify SNPs that are likely to contribute to variance in pig loin depth, and to explore their biological plausibility by examining the genes in which these SNPs are located.

Materials & methods

Animals and phenotypes. Analyses were performed on 232,057 pigs from the Pig Improvement Company (PIC; Hendersonville, TN, USA). Pigs from two maternal lines (recoded as A and B) and two terminal lines (recoded as C and D) were used in this analysis. Breeding values for loin depth measured via ultrasound were calculated as part of routine genetic evaluations. Breeding values were de-regressed (dEBV) as per VanRaden and Wiggans (1991), and used as phenotypes in GWAS. To account for variation in the reliability of dEBVs, a weighting value based on Garrick *et al.* (2009) was calculated for each individual, and the reciprocal of each weight was used as a residual weighting factor in GWAS.

Genotypes and quality control. Tissue samples were acquired from tails, which were docked within 48 hours of birth, and samples were genotyped using a customized Illumina single-nucleotide polymorphism (SNP) chip, mapped to susScr11.1. SNPs with MAF<0.01 or a call rate <0.95 were removed from further

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_378, © S. Desire *et al.* 2022

analyses. After quality control, an average of 42,370 (SD 1,514) SNPs were available for each line and used in GWAS analysis. Variants from whole genome sequence datasets were available, imputed via hybrid peeling (Ros-Freixedes *et al.,* 2020), and these SNPs were used to identify candidate SNPs, post-GWAS.

Association analyses. GEMMA software (Zhou and Stephens, 2012) was used to estimate centered genomic relatedness matrices. GEMMAs was also used to apply a series of univariate linear mixed models:

$$y = Xb + u + e, \quad (1)$$

where y is the vector of de-regressed EBVs, the vector β, represents the vector of effect size of each SNP, and X is the incidence matrix relating genotypes with β. The vector $u \sim N(0, G_c\sigma_u^2)$ represents polygenic additive effect with covariance equal to the centred genomic relationship matrix G_c and variance σ_u^2. The vector *e* represents residual error, weighted to account for the variable reliability of de-regressed breeding values. To conservatively account for multiple comparisons in the GWAS, *P*-values were adjusted by Bonferroni correction. A SNP was significant at the genome level if -log10(*P*-value) was greater than -log10(0.05/n), where n represents the number of markers across the genome.

Identifying candidate SNPs. A region spanning 0.5 Mbp up and downstream of each significant SNP was defined as a distinct genomic region. Significant SNPs within 0.5 Mbp of each other were merged into a single region. The GWAS SNP with the highest significance from each genomic region (lead SNP) was identified, and all SNPs in high LD (^{3}0.7) with lead GWAS SNPs were identified from sequence data. We obtained pCADD scores (Groβ *et al.,* 2020) for each of these SNPs, along with their predicted effect, using Ensembl Variant Predictor (VEP). The variant with the highest pCADD score found to be in high LD with a lead SNP is referred to as the seqSNP. The genes harbouring seqSNPs were identified from the Ensembl database.

Results & discussion

The aim of this study was to identify regions associated with loin depth in four pig lines, and to interrogate the candidate causative SNPs, as suggested by pCADD scores. Table 1 lists the strongest associations found within each genetic line, along with a corresponding seqSNP for each region. While SNPs with strong genome-wide associations with loin depth were found for all genetic lines, no common regions were found across all four lines (Figure 1). Of the exonic seqSNPs identified, 11 were missense variants, 8 were synonymous, and one stop-gain. Six of the seqSNPs were intronic and 5 were intergenic.

Table 1. The most highly significant GWAS associations (lead SNPs) for de-regressed EBVs for loin depth in each of four purebred pig lines, and candidate causative SNPs (seqSNPs), identified using pCADD scores.

line	SSC	lead SNP		seqSNP		
		Position (Mbp)	GWAS -log10pvalue	Position (MbP)	pCADD score	Predicted consequence
A	6	45.74	11.14	52.0	34.95	missense
B	2	0.40	18.83	0.8	41.97	synonymous
C	16	34.10	21.55	34.0	43.19	missense
D	5	66.10	27.4	66.2	15.82	intergenic

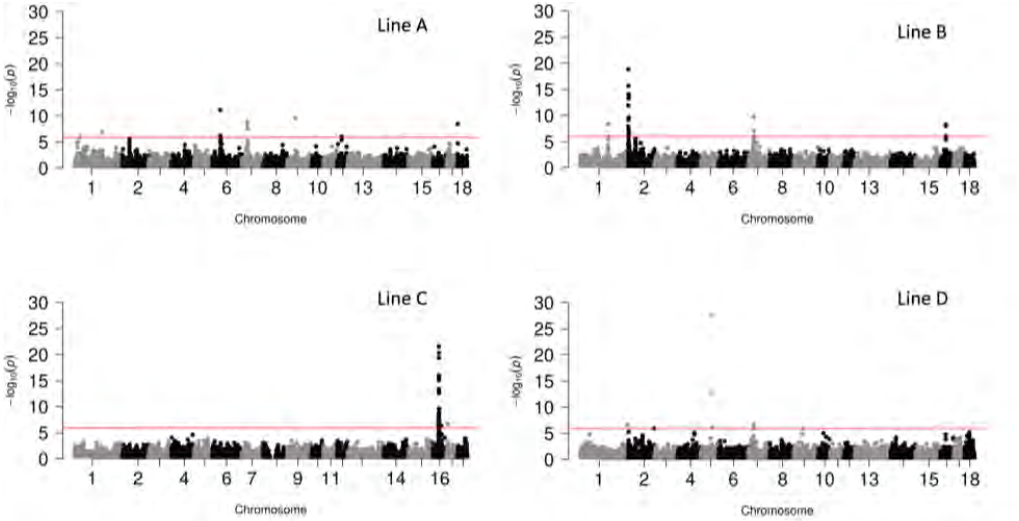


Figure 1. Genome wide plot of −log10 *P*-values (y axis) for association of SNPs with loin depth in four pig lines. The horizontal line indicates the threshold for genome wide significance.

One genomic region on SSC7 (~30 Mbp) was found to be statistically associated with loin depth in both maternal lines, and one terminal line (D). This region has been previously associated with loin muscle area and other body traits in pigs (Zhuang *et al.*, 2019; Bian *et al.*, 2021). Three separate seqSNPs were identified by pCADD scores for this region. Two were missense SNPs located in *GRM4* and *FANCE,* and one was a synonymous variant located in *CPNES*.

One maternal line and both terminal lines shared a large genomic region located on SSC16 (~32-34 Mbp) in common. Two of the seqSNPs identified for this region were located within known genes *ITGA1* and *ARL15*, while the third was an intron variant located in *MAST4*. Several studies have found this SSC16 region to be associated with loin muscle (Zhuang *et al.*, 2019; Bergamaschi *et al.*, 2020) and other body and growth-related traits (Gozalo-Marcilla *et al.*, 2021). In addition, *ITGA* has been shown to be up-regulated in steers with low marbling in the *Longissimus Dorsi* muscle (de las Heras-Saldana *et al.*, 2020). The gene *GRIK4* located on SSC9 (~47.7 Mbp) was implicated in one maternal and one terminal line and has been previously associated with loin muscle depth in pigs (Bian *et al.,* 2021). *MAST4* has not previously been implicated in pig production.

The maternal lines shared two significant genomic regions on SSC1 (~160.7 Mbp) and SSC2 (~41.5 Mbp). The most likely seqSNP identified for these regions within both breeds were missense variants located within the *MC4R* (SSC1), and *PIK3C2A* (SSC2). *MC4R* is a known regulator of feed intake, growth and meat quality traits (Kim *et al.*, 2000). Derks *et al.*, 2021 predicted a missense mutation within *MC4R* to be associated with fat deposition and growth in pigs, while (Reyer *et al.,* 2017) found a SNP near *MC4R* to be statistically linked to lean mass percentage and growth. On SSC2, *PIK3C2A* has been linked to lean meat percentage in pigs (Zhou *et al.*, 2021). Cánovas *et al.* (2010) showed *PIK3C2A* to be downregulated in pigs in a high lipid deposition group, while *PIK3C2A* is up-regulated in the high-marbled Hereford cattle (Roudbari *et al.*, 2020).

The strongest signals in the present study have been replicated elsewhere, supporting the evidence that these regions are important to the genomic control of muscle traits. pCADD identified several candidate

SNPs, some of which were in genes that have plausible biological links to muscle phenotypes, although some were located a considerable distance away from the lead SNP. It is notable that a large proportion (35%) of variants flagged by pCADD in this analysis were missense mutations. CADD was developed to estimate the pathogenicity of SNPs in humans (Kircher *et al.,* 2014), and as such prioritises potentially deleterious variants with protein-coding consequences. Generally, GWAS variants are expected to have small effect sizes, therefore the genes they impact are likely to have a small effect on the phenotype. Further work is required to determine whether the missense variants prioritised in this study have a functional effect on muscle phenotypes. Several intergenic, intronic, or synonymous variants were predicted to be likely candidate variants by pCADD scores. The next step is to perform eQTL analysis on porcine muscle sampled from the genetic lines used in the present study. These data will aid the interrogation of candidate causal SNPs.

References

Bergamaschi, M., Maltecca, C., Fix, J., Schwab, C., & Tiezzi, F. (2020). J. Anim. Sci. 98(1), 1–11. https://doi.org/10.1093/jas/skz360

Bian, C., Prakapenka, D., Tan, C., Yang, R., Zhu, D. *et al.* (2021). Genet. Sel. Evol.

53(1). https://doi.org/10.1186/s12711-021-00661-y

Cánovas, A., Quintanilla, R., Amills, M., & Pena, R.N. (2010). In BMC Genom. (Vol. 11). http://www.biomedcentral.com/1471-2164/11/372

de las Heras-Saldana, S., Chung, K. Y., Kim, H., Lim, D., Gondro, C. *et al.* (2020). Genes, 11(11), 1–17. https://doi.org/10.3390/genes11111381

Derks, M. F. L., Gross, C., Lopes, M. S., Reinders, M. J. T., Bosse, M. *et al. (*2021). Genomics, 113(August 2020), 2229–2239.

Garrick, D. J., Taylor, J. F., & Fernando, R. L. (2009). Genet. Sel. Evol. 41(1), 55. https://doi.org/10.1186/1297-9686-41-55

Gozalo-Marcilla, M., Buntjer, J., Johnsson, M., Batista, L., Diez, F., *et al.* (2021). Genet. Sel. Evol. 53(76).

Groβ, C., Derks, M., Megens, H.J., Bosse, M., Groenen, M.A.M. *et al.* (2020). Genet. Sel. Evol. 52(1). https://doi.org/10.1186/s12711-020-0528-9

Kim, K.S., Larsen, N., Short, T., Plastow, G., & Rothschild, M.F. (2000). Mamm. Genome. 11(2), 131–135. https://doi.org/10.1007/s003350010025

Kircher, M., Witten, D. M., Jain, P., O'Roak, B. J., Cooper, G. M. *et al.* (2014). Nat. Genet. 46(3), 310–315. https://doi.org/10.1038/ng.2892

Reyer, H., Varley, P. F., Murani, E., Ponsuksili, S., & Wimmers, K. (2017). Sci. Rep. 7(1). https://doi.org/10.1038/s41598-017-08961-4

Roudbari, Z., Coort, S. L., Kutmon, M., Eijssen, L., Melius, J., *et al.* (2020). Fron. Genet. 10. https://doi.org/10.3389/fgene.2019.01370

VanRaden, P.M., & Wiggans, G.R. (1991). J. Dairy Sci. 74(8), 2737–2746. https://doi.org/10.3168/jds.S0022-0302(91)78453-1

Zhou, S., Ding, R., Meng, F., Wang, X., Zhuang, Z., *et al.* (2021). BMC Genom. 22(1). https://doi.org/10.1186/s12864-020-07288-1

Zhou, X., & Stephens, M. (2012). Nat. Genet, 44, 821–824.

Zhuang, Z., Li, S., Ding, R., Yang, M., Zheng, E., Yang, H., *et al.* (2019). PLoS one., 14(6), 1–21. https://doi.org/10.1371/journal.pone.0218263

379. Multivariate GWAS revealed a critical region on chromosome 4 for developmental and feather growth traits in ducks

E. Tarsani[1*], S. Desire[1], O. Matika[1], K. McIntosh[1,2], A. Rae[2], A. Kranis[1] and K.A. Watson[1]

[1]The Roslin Institute and Royal (Dick) School of Veterinary Studies, University of Edinburgh, Easter Bush, EH25 9RG, Midlothian, United Kingdom; [2]Cherry Valley Farms Ltd., Grimsby Road., DN37 7DP, Laceby, United Kingdom; etarsani@exseed.ed.ac.uk

Abstract

Aim of this study was to discover the genetic variants for average daily gain (ADG), body weight (BW) and primary feather length (PRF) traits in Pekin ducks. Using 1,215 ducks and 45 k autosomal SNPs, multivariate and univariate genome-wide association analyses were performed. A total number of 50 markers were identified as common between all analyses and these SNPs were co-localized on chromosome 4. Current findings confirm previously reported findings for growth in ducks while proposing novel genetic loci associated with feather development and growth in ducks.

Introduction

Pekin duck has become the predominant breed used for meat, egg and feather production in the global duck industry. Current literature on duck genome-wide association studies (GWAS) is limited to univariate models conducted on quantitative traits such as growth and carcass traits (Deng *et al.* 2019, Zhou *et al.* 2018, Zhu *et al.* 2019), feeding behaviour (Zhu *et al.* 2019) and egg production (Liu *et al.* 2021) with no multivariate GWAS reported so far in ducks.

The objectives of the present study were to: (1) identify the genetic variants with pleiotropic effects on growth and primary feather length in ducks; and (2) investigate the most plausible pleiotropic genes for the examined traits.

Materials & methods

Data and quality control. Genotypic and phenotypic records were provided by Cherry Valley Farm (CVF). The available data consisted of 1,215 Pekin ducks (n= 298 males and n=917 females) with phenotypic records on average daily gain (ADG) at 27 to 39 days of age, body weight (BW) at 39 days and primary feather length (PRF) at 39 days of age. Animals were genotyped using a 60K Affymetrix SNP array. After quality control (samples: call rate: 0.95, markers: call rate: 0.95, MAF: 0.05 and HWE Fisher exact test *P*-value:10^{-6}) a total number of 1,215 samples and 45,080 autosomal SNPs were retained for further analyses.

Univariate and multivariate association analyses. Data were initially analyzed using linear mixed models fitting effects for sex and hatch as fixed effects and animal as random effects using ASReml (Gilmour *et al.*2015) software. Univariate association analyses using GEMMA (Zhou and Stephens 2012) software were conducted fitting the same models as those used in the variance component analyses. The traits analysed were ADG, BW and PRF.

The multivariate analysis, fitting a multivariate linear mixed model three traits (ADG, BW and PRF), subsequently performed in GEMMA software using same fixed effects as used before. For all analyses, SNPs with FDR *P*-value<0.05 were considered as genome-wide significant. Estimation of the genomic heritability was further implemented via the realized marker information. Finally, the genetic correlation (r_g) was estimated between bivariates i.e. ADG and BW, ADG and PRF, BW and PRF.

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_379, © E. Tarsani *et al.* 2022

Positional candidate genes. We used the SNPs that reached the significance threshold (FDR *P*-value) identified across all analyses to search for positional candidate genes. The search was restricted to SNPs within 50 kb downstream and upstream flanking regions of each significant marker for positional candidate genes using the NCBI database and the ZJU1.0 assembly. Due to the absence of a duck QTL database, we used the orthologues of chicken genes to search the documented role of the genes.

Results

Estimation of genetic parameters. The heritability estimates for additive effects were 0.28±0.04, 0.47±0.04 and 0.17±0.03 respectively for ADG, BW and PRF. The estimates of genetic correlation (r_g) between the ADG and BW was high (0.77) with negative and moderate estimates (-0.21) between ADG and PRF, and low and negative estimates (-0.09) between BW and PRF.

Comparison of genome-wide significant SNPs found by multivariate and univariate analyses. Figure 1 shows the profiles of the SNP *P*-values (expressed as $-\log_{10}$ values) across the three GWAS. Specifically, a total number of 112 genome-wide significant SNPs (FDR *P*<0.05) were detected by ADG univariate analysis and these SNPs were dispersed across nine autosomes (1-4, 10, 14, 15, 19 and 28). For BW univariate analysis, a total number of 138 SNPs across seven autosomes (1, 4, 15, 16, 20, 24 and 28) were found to reach genome-wide significance (FDR *P*-value<0.05). For PRF univariate analysis, a total number of 62 genome-wide significant SNPs (FDR *P*-value) were detected and these markers were located on two autosomes (4 and 18). The multivariate analysis resulted in 119 genome-wide significant (FDR *P*-value<0.05) SNPs across 10 autosomes (1-4, 11, 14, 15, 18, 19 and 28).

Comparison of SNP signals across the four association analyses resulted in fifty common significant markers on chromosome 4. Note that the 50 significant SNPs residing on chromosome 4 were co-localized

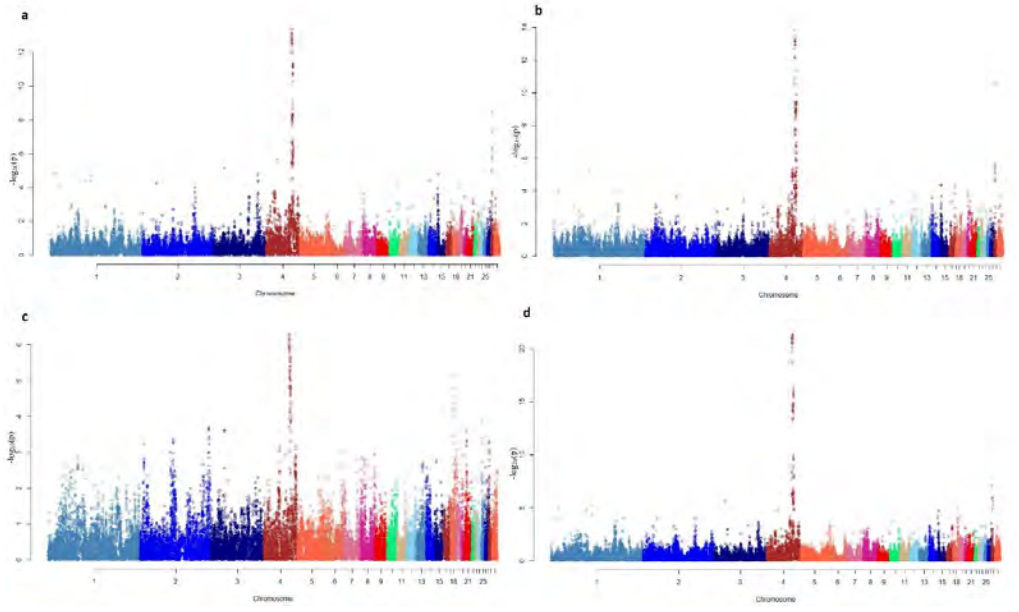


Figure 1. Manhattan plots showing the -$\log_{10}$(*P*-values) of SNPs across the 29 autosomes for average daily gain (ADG) (a), body weight (BW) (b), primary feather length (PRF) and all traits, respectively. Plots were constructed using the gaston R package.

in a region spanning 3,498,371 bp (57,648,845-61,147,216 bp) with high LD (r^2) levels (r^2:0.40-1). From the 50 SNPs, n=28 SNPs were located within 8 genes (*PPARGC1A, KCNIP4, PACRGL, SLIT2, LDB2, FBXL5, CC2D2A and CPEB2*) while the total number of positional candidate genes was 23 (*DHX15, PPARGC1A, ADGRA3, KCNIP4, PACRGL, SLIT2, LCORL, NCAPG, FAM184B, MED28, LAP3, CLRN2, QDPR, LDB2, TAPT1, PROM1, FGFBP2, FGFBP1, BST1, FBXL5, CC2D2A, C1QTNF7 and CPEB2*).

Discussion

Multivariate genotype-phenotype methods are known to increase the statistical power to detect marker-trait associations in the case of shared genetic basis between traits. The present study exploited a multivariate GWAS to identify candidate pleiotropic genetic loci for the ADG, BW and PRF traits in ducks. Our results identified two genes (*FGFBP1* and *FGFBP2*) which are fibroblast growth factors that have been reported to play critical role in embryogenesis and positive regulation of hair and feather development (Mandler and Neubüser 2004). Apart from these candidate pleiotropic genes, our study also supports findings by previous studies for growth traits in ducks. An interesting example is the *PPARGC1A* gene reported to play a role in promoting fatty acid oxidation of breast muscle in ducks (Fan *et al.* 2020). The other gene, *LDB2* has previously been reported as a candidate gene for body size and carcass traits in ducks (Deng *et al.* 2019). The *KCNIP4* gene, a member of potassium channel-interacting proteins has extensive physiological functions and insulin secretion was identified as an important QTL region associated with growth traits in chickens (Cha *et al.* 2021). *SLIT2* has also been reported to be strongly associated with back yield, feet weight, fat and muscle weight traits in chickens (Cha *et al.* 2021).

Several previous GWAS in chickens reported that chromosome 4 is a critical region influencing growth traits (Xu *et al.* 2013). We identified large QTL on chromosome 4 that has markers significantly associated with growth in ducks. Current findings by both univariate and multivariate GWA analyses could dissect the duck genetic architecture and have potential use in breeding programmes.

References

Deng MT, Zhu F, Yang YZ, Yang FX, Hao JP, *et al.* (2019). BMC Genomics. 20: 1–13. https://doi.org/10.1186/s12864-018-5379-1.

Zhu F, Cheng SR, Yang YZ, Hao JP, Yang FX, *et al.* (2019). Front Genet. 10: 702. https://doi.org/10.3389/fgene.2019.00702.

Zhou Z, Li M, Cheng H, Fan W, Yuan Z, *et al.* (2018). Nat Commun 2018 91. 9: 1–10. https://doi.org/10.1038/s41467-018-04868-4.

Liu H, Zhou Z, Hu J, Guo Z, Xu Y, *et al.* (2021). Anim Genet. 52: 536–541. https://doi.org/10.1111/age.13063.

Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, *et al.* (2007). Am J Hum Genet. 81: 559–75. https://doi.org/10.1086/519795.

Gilmour, A. R., Gogel, B. J., Cullis, B. R., Welham, S. J. and Thompson, R. (2015). Structural Specification, VSN International Ltd, Hemel Hempstead,HP1 1ES, UK www.vsni.co.uk

Zhou X, Stephens M. (2012). Nat Genet. 44: 821–824. https://doi.org/10.1038/ng.2310.

Mandler M, Neubüser A. (2004). Development. 131: 3333–3343. https://doi.org/10.1242/dev.01203.

Fan W, Liu W, Liu H, Meng Q, Xu Y, *et al.* (2020). BMC Genomics. 21: 1–15. https://doi.org/10.1186/s12864-020-6482-7.

Deng MT, Zhu F, Yang YZ, Yang FX, Hao JP, *et al.* (2019). BMC Genomics. 2019;20. https://doi.org/10.1186/s12864-018-5379-1.

Cha J, Choo H, Srikanth K, Lee SH, Son JW, *et al.* (2021). Genes (Basel). 12: 1170. https://doi.org/10.3390/genes12081170.

Xu Z, Nie Q, Zhang X. (2013). Curr Genomics. 14: 137–146. https://doi.org/10.2174/1389202911314020006.

380. Investigation of genes associated with maternal weaning weight in South African Bonsmara beef cattle

J.J. Reding[1,2*], R.R. Van der Westhuizen[1,2], D.P. Berry[2,3] and E. van Marle-Köster[2]

[1]SA Stud Book and Animal Improvement Association, 118 Henry Street, Westdene, 9300, Bloemfontein, South Africa; [2]University of Pretoria, Department of Animal Science, Lynnwood Road, 0002 Pretoria, South Africa; [3]TEAGASC, Moorepark, Fermoy, Ireland; jason@studbook.co.za

Abstract

Maternal weaning weight (WW_{MAT}), the dam's genetic and permanent environmental components that contribute to the weaning weight of her calf, is a trait of economic importance in extensive beef production systems. Positive genetic correlations between milk yield and WW_{MAT} have been reported. In this study, a genome wide association was performed to search for genes associated with WW_{MAT}. A total of twenty-three significant single nucleotide polymorphisms (SNPs) was identified with a further 103 suggestive SNPs. These SNPs were collapsed into 12 quantitative trait loci (QTL) across eight autosomes with an additional QTL residing on the X-chromosome. Ten different genes were co-located with the detected QTL. Genes identified were either essential in protein ubiquination pathways or were related to mammary morphology and ability to respond to infection by the presence of bacterial species.

Introduction

Milk production in beef cattle is an essential component of calf growth (Mulliniks *et al.*, 2020), especially in extensive systems due to seasonal fluctuations in natural veld quality and low forage availability (Cortés-Lacruz *et al.*, 2017, Mulliniks *et al.*, 2020). Milk yield is influenced by both the calf, through its demand for milk in conjunction with nursing frequency, along with the nutritional, genetic and health status of the cow. Other maternal effects include the maternal ability of the cow to protect the calf, the gestational environment and the transfer of antibodies through colostrum which contribute to calf growth up to weaning (Mulliniks *et al.*, 2020).

Positive genetic correlations between milk yield and maternal weaning weight (WW_{MAT}) have been well documented (Nesengani *et al.*, 2018) and thus warrant investigation into the underlying genetic mechanisms of these traits (Meyer *et al.*, 1994). The majority of single nucleotide polymorphism (SNP) analyses in cattle have focused on traits such as measured milk production, growth and carcass characteristics. Fewer studies have attempted to locate genomic regions associated with WW_{MAT}. Significant SNPs on multiple autosomes were linked to WW_{MAT} in Brahman cattle (Martínez *et al.*, 2017), while Hay and Roberts, (2019) identified a region on *Bos taurus* autosome (BTA) 24 that contributed significantly to the genetic variance of WW_{MAT}.

The South African Bonsmara, classified as a Sanga type, is a composite breed of 5/8 Afrikaner and 3/8 exotic (Milk Shorthorn, Hereford) established through a well-documented crossbreeding program. The breed is the most numerous in South Africa (SA) and is farmed under extensive production systems. In this study, a genome wide association was performed to search for genes associated with WW_{MAT}.

Materials & methods

Ethical approval for the use of external data were granted by the Research/Ethics Committee (EC-180000127), Faculty of Natural and Agricultural Sciences, University of Pretoria.

Genetic evaluations for WW_{MAT} in SA are generated from a univariate model that includes the fixed effects of calf sex, if the dam was primiparous or multiparous, the age of the calf and the dam when the calf was

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_380, © J.J. Reding *et al.* 2022

weighed and the contemporary group in which the calf was weaned. Random effects included are Sire × Herd × Year (SHY), the permanent environmental effect of the dam as well as the direct genetic effect of both the animal and the dam. A total of 1,242,158 weaning weight records are used in the genetic evaluation with an average weaning weight of 216.82 kg. Direct and maternal heritability estimates used are 0.27 and 0.17, respectively.

Estimated breeding values (EBVs) of 3,253 genotyped animals for WW_{MAT} were available, which consisted of 1,498 males and 1,755 females. The genotypes originated from three genotype panels, namely, 1,932 animals on the GeneSeek Genomic Profiler (GGP) 150K (140,113 SNPs), 589 animals on the GGP 80K (76,883 SNPs) and 732 animals on the Irish Cattle Breeding Federation (ICBF) International Beef and Dairy (IDB, 55,445 SNPs) platforms. All SNP locations were based on the ARS-UCD1.2, INSDC Assembly (GCA_002263795.2, Low *et al.*, 2019). Only autosomal and X-chromosomal SNPs with a known position and a call rate of ≥0.90 were retained. Non pseudo-autosomal SNPs on the X-chromosome were assumed homozygous for all male genotypes. All genotypes were subsequently imputed in FImpute v.3 to 134,421 SNPs.

Effective record contributions (ERCs) for each genotyped animal were generated using the reversed reliability approximation method in APaX99 (Lidauer *et al.*, 2017). The estimated breeding values for WW_{MAT} were subsequently deregressed using the Secant method in MiX99 (Strande, 1999) with the generated ERC used as a weighting factor. Only animals with an ERC of ≥0.5 were retained, which resulted in 1,454 animals for WW_{MAT} (median ERC=7.618) being available for the association analysis.

The weighted (by ERCs) deregressed WW_{MAT} EBVs were regressed on each SNP individually using a linear mixed model in WOMBAT (Meyer and Tier, 2012). An intercept term and the allele count per SNP were both included as fixed effects while relationships among animals was accounted using a genomic relationship matrix. The t-test statistic for all SNPs was obtained and SNPs with a $P \leq 1 \times 10^{-7}$ were considered to be genome-wide significant, while SNPs with a $P \leq 5 \times 10^{-6}$ were deemed to be suggestively significant.

The start and end of each quantitative trait loci (QTL) was defined by SNPs 0.5 Mb up and downstream that had an $r^2 > 0.5$ with the significant SNP. Overlapping QTL were consolidated into one QTL. Identified QTL were then explored using ENSEMBL in order to detect candidate genes residing within and Panther (Mi *et al.*, 2017) was used to list the biological and metabolic functions and/or processes of possible genes.

Results & discussion

Twenty-three significant SNPs were identified with a further 103 suggestive SNPs being identified. Fourteen of the significant SNPs were all located on BTA 8, with the remaining SNPs being on BTA 2, 6, 6, 9, 10, 16, 22 and *B. taurus* X-chromosome (BTX). Eleven different genes were co-located with the detected QTL (Table 1). The amalgamated QTL on BTA 8 contains nine significant SNPs that are in high linkage disequilibrium, with most of these being intron SNPs located in the *PCSK5* gene coding region.

DEGS1 was identified by Xu *et al.* (2018) to show differential gene expression in mammary tissue of dairy cattle after infection with *Escherichia coli* or *Streptococcus uberis*. *PCSK5* was reported by Hoac *et al.* (2018) to play a role in embryonic development, hormone regulation as well as bone mineralisation. *PCSK5* was associated with teat scores in Canadian Angus Cows (Devani *et al.*, 2021). These genes may play a role in the morphology of the udder as well as affect milk yield through mediating resistance to infection which would lead to mastitis.

Table 1. Detected quantitative trait loci with the most significant single nucleotide polymorphism and co-located genes.

BTA	Start	End	Distance (kbp)	SNPs in QTL	Genes
2	718140	882990	164.85	5	*HERC2*
6	374487838	37516229	28.39	3	*LCORL*
7	60414842	60452720	37.88	2	*SH3TC2*
8	51745358	51899568	154.21	5	
8	52180527	52333334	152.81	9	*PCSK5*
8	78145059	78145059	0	1	*NTRK2*
9	28257531	28257531	0	1	
9	56836009	56836009	0	1	*EPHA7*
10	2575344	2575344	0	1	
16	27298728	27407724	108.99		*FBXO28, DEGS1*
22	22033915	22033915		1	
X	59554341	59554341	0	1	DCX

Some of the detected genes (i.e. *DCX*, *EPHA7*, *FBXO28* and *SH3TC2)* have not been previously reported to be associated with any traits in beef cattle breeds. The number of genes linked to the posttranslational modification of proteins through ubiquination (i.e. *EPHA7*, *FBXO28* and *HERC2)* is interesting. This process essentially mediates the quantity and quality of various proteins that contribute to cellular homeostasis and influence life activities.

The *LCORL* gene has been identified as a candidate QTL for birth weight, growth and length in beef cattle breeds (McClure *et al.*, 2010; Lindholm-Perry *et al.*, 2011). Lindholm-Perry *et al.* (2013) reported 14 SNPs in the *LCORL* gene region to be associated with muscle and adipose tissue related *LCORL* transcript in crossbred beef cattle. In cows and heifers, an up-regulated expression of *LCORL* in adipose tissue is significantly correlated with average daily feed intake.

HERC2 is involved in protein modification through ubiquination and was only previously reported to be associated with back-fat thickness in the Hanwoo beef cattle breed (Naserkheil *et al.*, 2020). *NTRK2* is a transmembrane signal receptor associated protein. A study on longevity indicated that *NTRK2* plays an important role in the development of reproductive tissues in both female and male beef cattle (Mészáros *et al.*, 2014). The occurrence of multiple adipose and reproductive tissue related genes indicates the complexity and polygenic effect of WW_{MAT}.

References

Cortés-Lacruz, X., Casasús, I., Revilla, R., Sanz, A., Blanco, M., & Villalba, D. (2017) Livest. Sci. 202:143–149. https://doi.org/10.1016/j.livsci.2017.05.025

Devani, K., Crowley, J. J., Plastow, G., Orsel, K., & Valente, T. S. (2021) J. Anim. Sci. 99. https://doi.org/10.1093/jas/skab087

Gonzalez, M., Villa, R., Villa, C., Gonzalez, V., Montano, M., *et al.* (2020) J. Adv. Vet. Anim. Res. 7:234–241 https://doi.org/10.5455/JAVAR.2020.G415

Gregory, K. E., Cundiff, L. V., & Koch, R. M. (1992) J. Anim. Sci. 70:2366–2372. https://doi.org/10.2527/1992.7082366x

Hay, E. H., & Roberts, A. (2019) Livest. Sci. 229:118–125. https://doi.org/10.1016/j.livsci.2019.09.022

Hoac, B., Susan-Resiga, D., Essalmani, R., Marcinkiweicz, E., Seidah, N. G., & McKee, M. D. (2018) Bone 107:45–55. https://doi.org/10.1016/j.bone.2017.11.002

Lidauer, M., Matilainen, K., Mäntysaari, E., Pitkänen, T., Taskinen, M. *et al.* (2017) Technical Reference Guide for MiX99 Pre-Processor. XII.

Lindholm-Perry, A. K., Kuehn, L. A., Oliver, W. T., Sexten, A. K., Miles, J. R. *et al.* (2013) PLoS One 8. https://doi.org/10.1371/journal.pone.0080882

Lindholm-Perry, A. K., Sexten, A. K., Kuehn, L. A., Smith, T. P. L., King, D. A., *et al.* (2011) BMC Genet. 12. https://doi.org/10.1186/1471-2156-12-103

Low, W. Y., Tearle, R., Liu, R., Koren, S., Rhie, A. *et al.* (2019) bioRxiv. https://doi.org/10.1101/720797

Martínez, R., Bejarano, D., Gómez, Y., Dasoneville, R., Jiménez, A., *et al.* (2017) Genet. Mol. Biol. 40:453–459.

McClure, M. C., Morsci, N. S., Schnabel, R. D., Kim, J. W., Yao, *et al. (*2010) Anim. Genet. 41:597-607.

Mészáros, G., Eaglen, S., Waldmann, P., & Sölkner, J. (2014) Open J. Genet. 04:46–55. https://doi.org/10.4236/ojgen.2014.41007

Meyer, K., Carrick, M. J., & Donnelly, B. J. (1994) J. Anim. Sci. 72:1155–1165. https://doi.org/10.2527/1994.7251155x

Meyer, K., & Tier, B. (2012) Genetics 190:275–277. https://doi.org/10.1534/genetics.111.134841

Mi, H., Huang, X., Muruganujan, A., Tang, H., Mills, C., Kang, D. *et al.* (2017) Nucleic Acids Res. 45:D183–D189. https://doi.org/10.1093/nar/gkw1138

Mulliniks, J. T., Beard, J. K., & King, T. M. (2020) Appl. Anim. Sci. 36:70–77. https://doi.org/10.15232/aas.2019-01883

Naserkheil, M., Bahrami, A., Lee, D., & Mehrban, H. (2020) Animals 10:1–24. https://doi.org/10.3390/ani10101836

Nesengani, L. T., Nephawe, K. A., Sebei, J., Norris, D., & Maiwashe, A. (2018) Animal 12:199–204. https://doi.org/10.1017/S1751731117001483

Strande, I. (1999) J. Dairy Sci. 82(12):2779–2787. https://doi.org/10.3168/jds.S0022-0302(99)75535-9

Xu, L., Zhang, W. G., Shen, H. X., Zhang, Y., Zhao, Y. M., *et al.* (2018) Livest. Sci. 216:100–108. https://doi.org/10.1016/J.LIVSCI.2018.08.005

381. Genome-wide genomic and functional association study for workability and calving traits in Holstein cattle

T. Suchocki[1,2], B. Kosińska-Selbi[1], A. Żarnecki[2], M. Jakimowicz[1], W. Jagusiak[2,3], M. Morek-Kopeć[2,3] and J. Szyda[1,2*]

[1]Biostatistics Group, Department of Genetics, Wroclaw University of Environmental and Life Sciences, Kozuchowska 7, 51-631 Wroclaw, Poland; [2]National Research Institute of Animal Production, Krakowska 1, 32-083 Balice, Poland; [3]Faculty of Animal Science, University of Agriculture in Krakow, Mickiewicza 24/28, 30-059 Krakow, Poland; joanna.szyda@upwr.edu.pl

Abstract

Traits related to the calving abilities and workability of cows are essential for the dairy economy and the welfare of cows. The main goal of our study was to identify SNPs significantly associated with calving and workability traits, but having in mind the low heritability of those traits and then consequently the lack of genes with very strong individual effects, we also identified significant metabolic pathways and gene ontology terms. The data set originated from the EuroGenomics Holstein-Friesian population. All animals were genotyped or imputed to the Illumina BovineSNP50 BeadChip Version 2. The first step of the analysis was to identify significant SNPs and the second step comprised the estimation of pathway effects using a mixed effect model. The largest number of significant SNPs was detected for direct calving ease and for direct stillbirth Significant KEGG pathways and GO terms were identified.

Introduction

Traits related to the calving abilities and workability of cows are essential both for dairy economics and for the welfare of cows. According to Philipsson *et al.* (1994), in 1994 only five countries – the Nordic countries and Slovenia, considered calving traits in their national selection indices. Nowadays 19 countries carry out routine genetic evaluations of calving and workability. Despite generally low heritabilities, during the last decade, it has been possible to conduct successful selection, resulting in an improvement in calving ease and decrease in stillbirth rate, as demonstrated by Ma *et al.* (2019). Numerous Genome-wide Association Studies (GWAS) for calving performance have been conducted so far. Traits related to handling cows on farms have only recently been subjected to genetic evaluation, which is currently carried out by 14 countries on the international level. So far, only three GWAS results for workability traits have been published by Jardim *et al.* (2018), Marete *et al.* (2018), and Chen *et al.* (2020), indicating that further studies on various populations are needed for a better understanding of the genetic components of workability phenotypes. Consequently, the primary goal of our study was to identify SNPs significantly associated with the direct and maternal components of calving ease, direct and maternal components of stillbirth, cow temperament and milking speed. Having in mind the low heritabilities of all those traits we expect no genes with very strong individual effects, so we defined a secondary goal of the study which was the identification of metabolic pathways and gene ontology terms significantly associated with workability and calving that express a cumulated effect of many genes on a functional scope.

Materials & methods

Material. The analysed data originated from the EuroGenomics Holstein-Friesian dairy cattle population. For each bull, direct calving ease (DCE), maternal calving ease (MCE), direct stillbirth (DSB), maternal stillbirth (MSB), milking speed (MSP), and temperament (TEM), were available in the form of de-regressed breeding values (DRP) corresponding to the Interbull evaluation from April 2020. The majority of individuals (87%) was genotyped by the Illumina BovineSNP50 BeadChip Version 2 containing 54,609 SNPs. Bulls genotyped with other platforms were imputed to the above using the Beagle software. SNP

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_381, © T. Suchocki *et al.* 2022

preselection criteria comprised a minor allele frequency of at least 0.01 and a minimum call rate of 99%. After editing, 46,216 SNPs remained for further analysis.

Genome-wide association study. The following mixed model was used to estimate the additive effects of SNPs:

$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{Z_1 g} + \boldsymbol{Z_2 a} + \boldsymbol{\varepsilon} \quad (1)$$

where y represents a vector of DRPs for considered traits; β is a vector of fixed effects representing a general mean with a design matrix X; Z_1 is a design matrix for SNP genotypes, which is parameterized as -1, 0, or 1 for a homozygous, heterozygous and an alternative homozygous SNP genotype, respectively; g is a vector of random additive SNP effects; Z_2 is a design matrix for a residual polygenic effect; α is a vector of random bulls' residual additive polygenic effects, representing the part of additive polygenic variance which is not explained by SNPs; and ε is a vector of residuals with $\boldsymbol{\varepsilon} \sim N\left(0, \boldsymbol{D}\hat{\sigma}_e^2\right)$, where D is a diagonal matrix containing the reciprocal of effective daughter contribution per bull on the diagonal and $\hat{\sigma}_e^2$ representing the residual variance. The covariance structure of g was assumed to be $\boldsymbol{g} \sim N\left(0, \boldsymbol{I}\frac{\hat{\sigma}_a^2}{N_{SNP}}\right)$, with I being an identity matrix and $\hat{\sigma}_a^2$ representing the additive genetic variance of a given trait and $\mathrm{N}_{\mathrm{SNP}}$ is the number of SNPs. $\boldsymbol{a} \sim N(0, \boldsymbol{A}\hat{\sigma}_{a*}^2)$, where A is the numerator relationship matrix and $\hat{\sigma}_{a*}^2$ is a predetermined ratio of genetic variance for each trait. In this study $\hat{\sigma}_{a*}^2 = 0.4 \cdot \hat{\sigma}_a^2$. For testing the hypotheses ($H_0: g_i = 0$ vs. $H_1: g_i \neq 0$), we used the Wald test with the Bonferroni correction for multiple testing.

Functional analysis. To associate SNPs to closest genes, all polymorphisms' coordinates from the Illumina BovineSNP50 BeadChip were remapped to the ARS_UCD1.2 reference genome. SNPs were genomically annotated with the Variant Effect Predictor, with 5,000 bp as the maximum distance between a SNP and its closest gene. The functional information was expressed by pathways defined by the Kyoto Encyclopedia of Genes and Genomes (KEGG) and by Gene Ontology (GO) terms corresponding to the genes marked by SNPs. DAVID was used for the annotation. The estimation of functional effects was based on the following model:

$$\boldsymbol{y}^* = \boldsymbol{\mu} + \boldsymbol{Z}^* \boldsymbol{p} + \boldsymbol{\varepsilon}^* \quad (2)$$

where y^* is the absolute value of the SNP effect for a given trait; p is a random KEGG pathway or GO term effect and ε^* represents the residual term. Z^* is the incidence matrix for KEGG/GO term effects p. It is assumed that $\boldsymbol{p} \sim N\left(0, \boldsymbol{P}\hat{\sigma}_p^2\right)$ and $\boldsymbol{\varepsilon}^* \sim N\left(0, \boldsymbol{L}\hat{\sigma}_{\varepsilon^*}^2\right)$, where P is a covariance matrix of KEGG/GO terms, consisting of the percentage of common genes between them and $\hat{\sigma}_p^2$ represents the KEGG/GO term variance. L is a diagonal matrix with elements equal to 1 (if a SNP was assigned a KEGG/GO term) or 10 (if a SNP was not assigned any functional element), $\hat{\sigma}_{\varepsilon^*}^2$ is the residual variance. For testing the hypotheses $H_0: p_i = 0$ vs. $H_1: p_i \neq 0$, the Wald test with the Bonferroni correction was used.

Results & discussion

Direct calving ease. The largest number of significant SNPs was detected for DCE. GWAS results demonstrated an accumulation of 22 highly significant SNPs on BTA18, distributed across a region spanning 10,225,739 bp. Nine of those SNPs were also reported as associated with direct calving ease by Müller *et al.* (2017) and by Abo-Ismail *et al.* (2017). Significant SNPs were also identified on BTA5, BTA7, BTA17, BTA19, BTA21, BTA25 and BTA29. On the functional scope, we identified seven significant biological processes: cartilage condensation, cell aggregation, artery development, nucleus localization, cardiac myofibril assembly, myofibril assembly, and actin filament-based movement. Microfilament motor

activity was the significant molecular function, while two significant cell components were myosin complex and neuromuscular junction. No significant KEGG pathway was identified.

Maternal calving ease. For the maternal component of calving ease, much fewer significant SNPs (17) were identified than for the direct component (59). The most significant SNPs (8) were scattered over BTAX, from 863,652 bp up to 127,125,130 bp. Among those SNPs, two have already been previously reported by Cole *et al.* (2011) as associated with maternal calving ease and direct calving ease. A significant SNP on BTA6 and BTA20 were reported by Sahana *et al.* (2011). Many of the significant SNPs were reported to be associated with conformation traits, which may indirectly impact calving ease, i.e. body depth, dairy form, rump width, stature, and type. No significantly associated KEGG pathways were identified. Significant GOs represented processes of positive regulation of protein maturation and modulation of chemical synaptic transmission with synapses being associated with cellular components.

Direct stillbirth. 19 out of the 25 significant SNPs overlapped with those that were also significant for DCE. Both traits were characterised by many significant SNPs located mainly toward the end of BTA18, all the other 18 polymorphisms were pleiotropic for DSB and DCE. Five SNPs from BTA18 have already been reported by Höglund *et al.* (2012) and by Müller *et al.* (2017). Another chromosome with multiple significant SNPs is the BTAX, but each of the four polymorphisms points out at a very different location. The similarity of the genetic architecture underlying DSB and DCE revealed by GWAS was also reflected on the functional annotation level, since all eight GO terms significantly associated with DSB were also significant for DCE. There was no significant association of stillbirth with KEGG.

Maternal stillbirth. The largest number (seven) of SNPs significant for MSB was located on BTA9. Six of them also exhibited high effects on MSB. In the literature, none of those significant SNPs was reported. Still, three SNPs located on BTA18 were also significant for the DSB and DCE, five SNPs located on BTA9, BTA18 and BTAX were also associated with MCE. Signalling mediated by cyclic nucleotides was the biological process associated with MSB. The significant molecular function is binding of 1-phosphatidylinositol. MSB was the only calving related trait for which significant KEGG pathways were assigned: PD-L1 expression and PD-1 checkpoint pathway in cancer and Th1 and Th2 cell differentiation. Although at the first glance both pathways seem unrelated to each other – a common feature is their relation to immune response, especially through T-type lymphocytes.

Temperament. All SNPs significant for temperament were located on BTAX but across a long span of 130,498,495 bp. Interestingly, the ATPase copper transporting alpha gene (*ATP7A*) harbouring the SNP with the highest effect on temperament is associated with temperament-related phenotypes of lethargy and hypoactivity in the mouse. Eight GO terms were estimated as significant for temperament, which are related to the processes of gamma-aminobutyric acid secretion and transport, the transmembrane import of calcium ions and glucose into cells as well as the early endosome membrane cell component. The significant KEGG pathways are related to the metabolism of porphyrin, chlorophyll, glyoxylate, and dicarboxylate.

Milking speed. With one exception, all 24 SNPs significantly associated with MSP did not show pleiotropy with other traits considered in this study. The most SNPs (ten) were located on BTA19 with the most significant of them forming two groups. One, close to the beginning of the chromosome, between 7,250,802 bp and 7,717,717 bp harbouring two genes – *ANKFN1* and a novel gene *ENSBTAG00000038823*. The other, close to the end of the chromosome, between 59,364,966 bp and 59,547,890 bp, harbouring a novel gene *ENSBTAG00000048685*. One of the SNPs in the latter group was reported as associated with several udder conformation traits, such as front and rear teat placement, udder attachment, udder cleft, udder depth and height by Cole *et al.* (2011). A group of five highly significant SNPs was also detected on BTA6, three

of them within a short region, spanning 68,984 intergenic bps, close to the *MSX1* gene. A polymorphism on BTA5 was previously reported by Lund *et al.* (2018) as associated with milking speed. Among the 11 GO terms significantly influencing milking speed the biological processes related to hippo signalling and the catalysis of ceramides and glycosphingolipids. Significant GO terms representing molecular functions comprised ligand-gated ion channel activity, phosphoric diester hydrolase activity and voltage-gated calcium channel activity, and binding of calcium ions and 1-phosphatidylinositol. The only significant KEGG pathway comprised the metabolism of degradation of valine, leucine and isoleucine.

Conclusions

Calving traits are low heritable. Their additive genetic component does not typically exceed 10%. Therefore, the presence of single genes with marked effects is not to be anticipated for the traits. However, a surprisingly large overlap was observed between genomic regions, so that many polymorphisms detected as significant in our study have already been indicated by previous analyses listed in the Results & discussion section. This observation is important for carrying out the genomic selection on the traits, especially by incorporating targeted polymorphisms into the evaluation model. Still, from the basic perspective of a better understanding of the genomic background of the phenotypes, traits' low heritabilities indicate that the focus should be moved from single genes to metabolic pathways or gene ontologies significant for the manifestation of calving problems. Such a functionally based approach has not so far been considered neither in the relation to calving nor workability traits.

References

Abo-Ismail M.K., Brito L.F., Miller S.P., Sargolzaei M., Grossi D.A., *et al.* (2017) Genet. Sel. Evol. 49:82. https://doi.org/10.1186/s12711-017-0356-8

Chen S.Y., Oliveira H.R., Schenkel F.S., Pedrosa V.B., Melka M.G., *et al.* (2020) J. Dairy Sci. 103:10383–10398. https://doi.org/10.3168/jds.2020-18897

Cole J.B., Wiggans G.R., Li M., Sonstegard T.S., Lawlor Jr T.J., *et al.* (2011) BMC Genom. 12:408. https://doi.org/10.1186/1471-2164-12-408

Höglund J.K., Guldbrandtsen B., Lund M.S., Sahana G. (2012) BMC Genet. 13:71. https://doi.org/10.1186/1471-2156-13-71

Jardim J.G., Guldbrandtsen B., Lund M.S., Sahan G. (2018) J. Dairy Sci. 101:2199–2212. https://doi.org/10.3168/jds.2017-12982

Ma L., Cole J.B., Da Y., VanRaden P.M. (2019) J. Dairy Sci. 102(4): 3735–3743. https://doi.org/10.3168/jds.2018-15269

Marete A., Sahana G., Fritz S., Lefebvre R., Barbat A., *et al.* (2018) J. Dairy Sci. 101(7): 6205–6219. https://doi.org/10.3168/jds.2017-14067

Müller P., Rothammer S., Seichter D., Russ I., Hinrichs D., *et al.* (2017) J. Dairy Sci. 100:1987–2006. https://doi.org/10.3168/jds.2016-11506

Philipsson J., Banos G., Arnason T. (1994) J. Dairy Sci. 77:3252–3261. https://doi.org/10.3168/jds.s0022-0302(94)77266-0

Sahana G., Guldbrandtsen B., Lund M.S. (2011) J. Dairy Sci. 94:479–486. https://doi.org/10.3168/jds.2010-3381

382. Genomic region-specific association test identifies new loci for carcass traits in Hanwoo (Korean native cattle)

D.J. Lee, D.H. Lee, J.M. Kang, Y.K. Kim, Y.J. Chung and S.H. Lee*

Division of Animal & Dairy Science, Chungnam National University, Daejeon, 34134, Republic of Korea; slee46@cnu.ac.kr

Abstract

Genetic variations in quantitative traits are remarkably complex. Moreover, it is more challenging to detect causative genetic variation because small effects mainly influence quantitative traits. The objective of this study is to search for markers that are significantly involved in carcass traits using missense mutations in which single nucleotide polymorphisms (SNPs) change amino acids. Genomic region-specific association test used missense variants and carcass traits of 16,970 Hanwoo. As a result, 222, 98, 142, and 119 significant SNPs were found in association with carcass weight (CW), backfat thickness (BFT), eye muscle area (EMA), and marbling score (MS), respectively. The most significant marker was the same in the results of CW and EMA, which was located at 19,524,263 on chromosome 14 involving the *PRKDC* gene. Four candidate genes were newly discovered for MS. The SNPs identified in this study will be useful markers for carcass traits in Hanwoo commercial farms.

Introduction

Hanwoo is the native cattle of Korea and has been inhabiting the Korean Peninsula for over 2,000 years. The improvement of Korean beef is cultivated by intensive selection for economic characteristics including meat quantity and meat quality (Edea *et al.* 2018). This was further accelerated by genome-wide panels of single nucleotide polymorphisms (SNPs), enabling the implementation of genome predictions in livestock. In addition, many genome-wide association studies (GWAS) are being conducted to explore quantitative trait loci (QTL) due to SNP information (Kanai *et al.* 2018; Bedhane *et al.* 2019) In previous studies, genetic variation for complex traits has a small effect, which accounts for only a small fraction of heritability (missing heritability problem), which may reduce the power of GWAS studies in complex traits (Weng *et al.* 2011; Dao *et al.* 2022). Koufariotis *et al.* (2014) showed that missense variants in the protein-coding region had the highest proportion in genetic variance explained by SNP according to functional annotation class in dairy and beef cattle. This kind of variants makes changes in both the genetic sequence and the amino acid sequence of the protein, being potential to be harmful or beneficial. This study aims to search for genetic mutations that change the amino acid sequence of proteins in carcass traits by conducting an association analysis using the functional aspects of SNPs in Hanwoo.

Materials & methods

Animals, genotyping, and quality control. Commercial Hanwoo (n=16,970) with an average age of 30.2±1.9 month (born in 2006-2016) were used in this study (Table 1). The animals were slaughtered in 2013-2019 at a commercial abattoir and produced phenotypes (carcass traits). The genomic DNA was extracted from longissimus thoracis muscle samples using a DNeasy Blood and Tissue Kit (Qiagen, Valencia, CA, USA). DNA concentration and purity were determined using a NanoDrop 1000 (Thermo Fisher Scientific, Wilmington, DE, USA). Samples were genotyped using the Bovine SNP50 BeadChip (Illumina Inc., San Diego, CA, USA). Imputation was carried out in 2-step using minimac3 (Das *et al.* 2016). Using the Illumina Bovine HD BeadChip (777K) of 1,295 Hanwoo, the 50K genotype of 16,970 Hanwoo was imputed to a high-density level. After that, it was imputed to the sequence level using whole genome sequence data of 203 Hanwoo. SNPs with R^2<0.6 were excluded in the imputation steps and SNPs on the sex chromosomes were excluded from the analysis. The imputed sequence level data was annotated

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_382, © D.J. Lee *et al.* 2022

using the SnpEff program (Cingolani *et al.* 2012). Genomic coordinates were converted from *UMD 3.1.1* to *ARS-UCD1.2.* Quality control was performed using PLINK1.9 software (Purcell *et al.* 2007) based on the following SNP selection criteria: call rate (CR) higher than 0.90; minor allele frequency (MAF) higher than 0.01; *P*-value for Hardy-Weinberg equilibrium test (HWE) higher than 10^{-4}.

Statistical model. A region-specific genomic data was used by extracting only missense variants from the sequence level data. The phenotypic data on carcass traits were pre-adjusted for fixed effects including growing sites, birth year, season of birth, and slaughter age using a mixed linear model implemented in R software 3.3.1 (R Foundation for Statistical Computing, Vienna, Austria). The adjusted phenotypes and SNP information were subsequently applied for GWAS using a linear mixed model. The linear mixed model can be written as:

$$y = a + bX + g + e \quad (1)$$

Y: phenotype; *a*: the mean term; *b*: additive effect (fixed effect) of the candidate SNP; *X*: SNP genotype indicator variable coded as 0, 1 or 2; *g*: the accumulated effect of all SNPs captured by the genetic relationship matrix (GRM) calculated using all SNPs; *e*: residual.

Results

Genomic region-specific association test. In this study, 28,114 missense variants that met the quality control criteria were used for the genomic region-specific association test. The results are shown in Figure 1. The Bonferroni correction method was used for the significance test (0.05/number of SNPs) in the genomic region-specific association test. In association with carcass weight (CW), back fat thickness (BFT), eye muscle area (EMA), and marbling score (MS), 222, 98, 142, and 119 significant SNPs were found, respectively. Among them, information of the three most significant SNPs based on *P*-value and information of novel marbling genes discovered are shown in Table 2. For the trait of carcass weight (CW), the marker at position 19,524,263 in chromosome 14, located in the *PRKDC* gene, was found to be the most significant. Interestingly, a significant association was detected between the eye muscle area (EMA) and the same SNP. For backfat thickness (BFT), the most significant marker was found at position 107,160,304 in chromosome 2, involving the *ZFAND2B* gene. The most significant marker for marbling score (MS) located in the *ACVR2B* gene at position 11,849,704 in chromosome 22 was also detected. In the case of marbling, one significant SNP was found in each of chromosomes 9, 14, 18, and 19, and they were novel genes that have not yet been identified.

Table 1. Descriptive statistics of phenotypic data in commercial Hanwoo.[1]

Traits	N	Mean	Median	Min.	Max.	SD	CV
CW	16,970	441.21	441.00	159	692	50.89	11.53
BFT	16,970	14.26	14.00	1	48	4.92	34.50
EMA	16,970	95.81	95.00	30	156	12.21	12.74
MS	16,970	5.94	6.00	1	9	1.85	31.14

[1] BFT = backfat thickness (mm); CV = coefficient of variation (%); CW = carcass weight (kg); EMA = eye muscle area (cm^2); MS = Marbling score (1-9); SD = standard deviation.

Table 2. Information on the three most significant SNPs and novel genes for carcass traits.

Trait	Chr:Pos	Allele	RefSNPID	Gene	SNPeffect	*P*-value
CW	4:9502576	C/T	rs445255852	*LRRD1*	9.460	1.734E-38
	6:37403795	T/C	rs109696064	*LCORL*	15.561	7.073E-50
	14:19524263	T/C	rs449968016	*PRKDC*	17.855	2.707E-92
BFT	2:107160304	G/T	rs109446852	*ZFAND2B*	0.811	1.639E-20
	13:52131127	C/T	rs210393794	*MRPS26*	-0.612	2.129E-17
	22:51866351	T/C	rs110365838	*MAP4*	0.549	1.629E-10
EMA	3:97976928	C/T	rs526496677	*SLC5A9*	1.649	1.349E-09
	6:37403795	T/C	rs109696064	*LCORL*	1.900	1.615E-13
	14:19524263	T/C	rs449968016	*PRKDC*	1.807	3.044E-17
MS	3:19028381	G/A	rs210416891	*MRPL9*	0.226	7.210E-11
	9:60753528	C/T	rs382343369	*Novel gene*	-0.233	1.441E-07
	14:16328530	A/G	rs471616366	*Novel gene*	-0.201	3.281E-08
	18:52030102	T/C	rs378517676	*Novel gene*	-0.183	2.835E-07
	19:7310638	G/C	rs799291287	*Novel gene*	-0.236	1.133E-08
	22:11849704	G/C	rs799031002	*ACVR2B*	0.320	9.777E-16
	25:36843468	A/G	-	*ZNF655*	0.296	1.327E-11

Discussion

Candidate genes for carcass traits. Several markers in the *NCAPG-LCORL* locus have been reported to be significantly associated with feed intake for weight gain and feed intake in beef cattle (Lindholm-Perry *et al.* 2011). In particular, the SNP c.1326T>G in the *NCAPG* (non-SMC condensin I complex, subunit G) gene identified in this study has also been reported to be a significant SNP associated with height, length, and width of puberty (components of body skeleton size) in the Japanese Black steer and F2 bulls of Charolais × German Holstein cross (Setoguchi *et al.* 2011). The PRKDC gene is important for embryonic development, interferon tau expression, and trophoblast development rate, and it has been reported that PRKDC is required during early bovine embryonic development (Ealy *et al.* 2012). Activin-type II receptor B (*ACVR2B*) is a myostatin-related gene involved in the regulation and signaling of myostatin. Genetic mutations in myostatin in cattle show marked effects on skeletal muscle, resulting in a dual muscle phenotype in certain breeds (Walsh *et al.* 2007).

Acknowledgements

This work was supported by Korea Institute of Planning and Evaluation for Technology in Food, Agriculture, Forestry and Fisheries (IPET) funded by Ministry of Agriculture, Food and Rural Affairs (MAFRA) (321082-3)

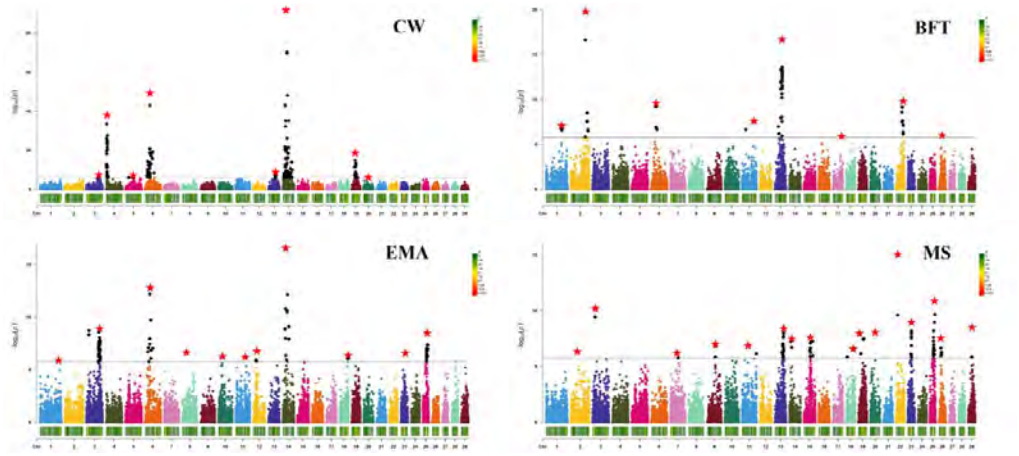


Figure 1. Manhattan plots of genomic region-specific association test using missense variants for carcass traits.

References

Bedhane M., Van Der Werf J., Gondro C., Duijvesteijn N., Lim D. *et al.* (2019) Frontiers in genetics 10:1235. https://doi.org/10.3389/fgene.2019.01235

Cingolani P., Platts A., Wang L.L., Coon M., Nguyen T. *et al.* (2012) Fly 6(2):80-92. https://doi.org/10.4161/fly.19695

Dao C., Jiang J., Paul D. and Zhao H. (2022) Journal of Statistical Planning and Inference 220:15-23. https://doi.org/10.1016/j.jspi.2022.01.003

Das S., Forer L., Schönherr S., Sidore C., Locke A.E. *et al.* (2016) Nature Genetics 48(10):1284-1287. https://doi.org/10.1038/ng.3656

Ealy A.D., Yang Q.E., Ozawa M. and Johnson S.E. (2012) Biology of Reproduction 87:211-211. https://doi.org/10.1093/biolreprod/87.s1.211

Edea Z., Jeoung Y.H., Shin S.-S., Ku J., Seo S. *et al.* (2018) Asian-Australas J Anim Sci 31(3):327-334. https://doi.org/10.5713/ajas.17.0276

Kanai M., Akiyama M., Takahashi A., Matoba N., Momozawa Y. *et al.* (2018) Nature Genetics 50(3):390-400. https://doi.org/10.1038/s41588-018-0047-6

Koufariotis L., Chen Y.-P.P., Bolormaa S. and Hayes B.J. (2014) BMC Genomics 15(1):436. https://doi.org/10.1186/1471-2164-15-436

Lindholm-Perry A.K., Sexten A.K., Kuehn L.A., Smith T.P.L., King D.A. *et al.* (2011) BMC Genetics 12(1):103. https://doi.org/10.1186/1471-2156-12-103

Purcell S., Neale B., Todd-Brown K., Thomas L., Ferreira M.a.R. *et al.* (2007) The American Journal of Human Genetics 81(3):559-575. https://doi.org/10.1086/519795

Setoguchi K., Watanabe T., Weikard R., Albrecht E., Kühn C. *et al.* (2011) Animal Genetics 42(6):650-655. https://doi.org/10.1111/j.1365-2052.2011.02196.x

Walsh S., Metter E.J., Ferrucci L. and Roth S.M. (2007) Journal of Applied Physiology 102(6):2142-2148. https://doi.org/10.1152/japplphysiol.01322.2006

Weng L., Macciardi F., Subramanian A., Guffanti G., Potkin S.G. *et al.* (2011) BMC Bioinformatics 12(1):99. https://doi.org/10.1186/1471-2105-12-99

383. Genetic mechanisms underlying feet and legs malformation in Nellore cattle: prioritization of GWAS results

T.L. Silva[1*], C. Gondro[2], P.A.S. Fonseca[3], D.A. da Silva[1], G. Vargas[1], H.H.R. Neves[4], I. Carvalho Filho[1], C.S. Teixeira[1], L.G. Albuquerque[1,5] and R. Carvalheiro[1,5]

[1]Department of Animal Science, Sao Paulo State University, School of Agriculture and Veterinarian Sciences, Jaboticabal, Brazil; [2]College of Agriculture and Natural Resources, Department of Animal Science, Michigan State University, East Lansing, MI, 48824, USA; [3]Department of Animal Biosciences, Centre for Genetic Improvement of Livestock, University of Guelph, N1G 2W1, Guelph, Canada; [4]Gensys Associated Consultants, 90680-000 Porto Alegre, Brazil; [5]Researcher at National Council for Scientific and Technological Development (CNPq), Brasília, Brazil; tl.silva@unesp.br

Abstract

A genome-wide association analysis was performed to identify genomic windows and the underlying positional candidate genes associated with feet and legs malformation (FL) in Nellore cattle. Single Nucleotide Polymorphism (SNP) effects were estimated from 295,031 pedigree records from animals with phenotypes, 12,537 genotypes for FL and 461,057 SNPs from the Alliance Nellore database using the weighted single-step genomic BLUP method. Twelve non-overlapping windows of 20 adjacent SNPs that explained more than 1% of the additive genetic variance were selected for candidate gene annotation. Functional and gene prioritization analysis of the candidate genes identified four genes (*ATG7*, *EXT1*, *ITGA1*, *PPARD*), previously reported in dairy and beef cattle populations, with a known role in skeletal muscle development, aberrant bone growth, lipid metabolism and intramuscular fat deposition. These candidate genes are believed to be associated with the incidence of FL in Nellore cattle.

Introduction

The feet and legs malformation (FL) is a functional binary trait already reported in the literature with a heritability of 0.18 (SE=0.04) and an incidence of approximately 5% in Nellore cattle (Vargas *et al.*, 2017). Animals affected by feet and legs malformations are unable to satisfactorily perform their productive and reproductive functions, which leads to considerable economic losses. The objectives of the present study were: (1) to find SNP windows significantly associated with FL in Nellore cattle using a weighted single-step approach; and (2) to identify positional candidate genes and their potential role in conformational structure of feet and legs.

Materials & methods

Pedigree data of 295,031 Nellore cattle from the Alliance Nellore (www.gensys.com.br), all with phenotypic information, collected between 2001 and 2017 were used in this study. Contemporary groups were defined considering the effects of herd, year and season of birth, sex, management group at weaning and yearling, date of measurement at yearling. Only data from contemporary groups with more than 10 records, variability in the trait and at least 10 genetic links to the main dataset were kept for the analyses.

A total of 12,537 genotyped animals from the genomic reference population of Alliance Nellore, all with their own phenotypes and progeny evaluated for FL, were used for GWAS. Animals genotyped with 50K SNP chip were imputed to 700K using FImpute V3 Software (Sargolzaei *et al.*, 2014) and 6,624 HD genotyped samples as reference. Using QCF90 (Masuda *et al.*, 2019) SNPs from non-autosomal regions, with minor allele frequency <0.02 and a call rate <0.9 were excluded from the analysis. After the quality control, there were 461,057 SNPs left.

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_383, © T.L. Silva *et al.* 2022

GWAS. Programs from the BLUPF90 suite (Misztal, 2020) were used to estimate SNP effects by the weighted single-step genomic BLUP method (WssGBLUP). The observed phenotypes of FL were used as dependent variables in a single-trait threshold animal model:

$$y = Xb + Z_a a + e \quad (1)$$

where y is a vector of underlying liabilities of FL; b is a vector of systematic effects; a is a vector of random additive direct genetic effects; e is a vector of random residual effects, and X and Z_a are incidence matrices that relate the phenotypes in y to the effects in b and a, respectively.

The solutions of SNP effect estimates (û) were obtained as a function of the genomic estimated breeding values (GEBVs) using the formula:

$$\hat{u} = DZ'[ZDZ']^{-1}\hat{a}_g \quad (2)$$

where D is a diagonal matrix with weights, assigned based on SNPs solutions, Z is an incidence matrix of genotypes for each locus, and $\hat{a}_g$ is the vector of GEBVs of the genotyped animals. The û vector and the matrix D were iteratively recomputed using the algorithm proposed by Wang *et al.* (2012). Thus, the weighted genomic relationship matrix and $\hat{a}_g$ were recalculated for the prediction of SNP effects at each iteration. The algorithm was run over three iterations in total.

Gene annotation. The sum of the proportion of variance explained by SNP effects within each non-overlapping window of 20 adjacent SNPs was adopted as the criterion to identify potentially important genomic regions. Windows that explained more than 1% of the additive genetic variance were selected for candidate gene annotation using the .gtf annotation file corresponding to the bovine gene annotation from the ARS-UCD1.2 assembly and the GALLO package (Fonseca *et al.*, 2020). The presence of genes in adjacent 100 Kb windows to the left and right of the main windows was also evaluated, as their effect can be captured by neighboring SNPs due to linkage disequilibrium (O'Brien *et al.*, 2014).

Gene prioritization analysis. A 'trained list' of the top 100 genes associated with relevant trait keywords[1] was created using GUILDify (Guney *et al.*, 2014) and a species-specific (*Homo sapiens*) interaction network. This list of 'trained' genes was used in the ToppGene Suite (Chen *et al.*, 2009) simultaneously with the list of 'test' genes, containing genes located within windows that explained more than 1% of the genetic variance. Gene Ontology terms (molecular function, biological process, and cellular component), human and mouse phenotypes, metabolic pathways, PubMed publications, coexpression pattern and diseases were used to retrieve the functional information for the genes in both lists. The prioritized significant genes were selected based on a multiple correction false discovery rate of 5% ($P \leq 10^{-3}$). This means that these genes have a functional profile that is significantly similar to the functional profile of the list of 'trained' genes.

Results

Twelve windows, located on chromosomes BTA2, BTA3, BTA5, BTA6, BTA14, BTA18, BTA20, BTA22, BTA23, BTA27 and BAT29, explained >1% of the genetic variance of FL. Together, these windows explained 16.37% of the total genetic variance of FL (Figure 1). A total of 35 positional candidate genes for FL were identified, including the genes found 100 Kb upstream and downstream from the windows. The window that harbored the largest number of genes (eight genes) for FL was located on BTA3 and the window that explained the largest proportion of variance (1.7%) was located on BTA27. The candidate gene prioritization analysis identified 4 prioritized candidate genes (*ATG7*, *EXT1*, *ITGA1*, *PPARD*) (Table 1).

[1] Keywords: 'mobility aplomb', 'feet', 'foot angle', 'leg', 'rear legs', 'leg conformation', 'leg structure', 'conformational structure', 'feet quality', 'leg quality', 'articular cartilage', 'osteogenic differentiation', 'lameness resistance' and 'hoof health'.

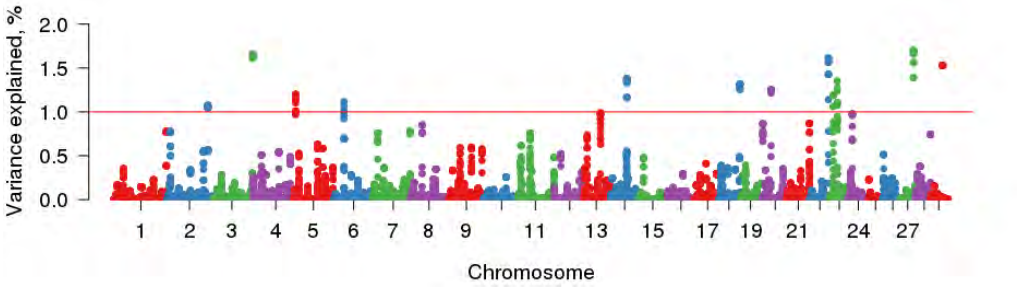


Figure 1. Manhattan plot for percentage of variance explained by top non-overlapping windows of 20 adjacent SNPs for FL obtained by WssGBLUP method. The red line indicates the 1% variance threshold explained by the windows.

Table 1. Significant windows explaining more than 1% of the variance for FL in Nellore cattle and the genes located within the windows (plus 100Kb up/downstream).

Window region			Var (%)	Gene names
BTA[1]	Start	End		
27	39,428,405	39,525,725	1.70	*LRRC3B*
3	119,525,250	119,674,756	1.66	*CSF2RA, P2RY8, AKAP17A, OR9S36, OR9S40, OR9S15, COPS9, OTOS*
22	55,269,065	55,321,257	1.61	*ATG7*[2]
29	35,078,320	35,112,359	1.53	*NTM*
14	46,405,542	46,506,899	1.38	*EXT1*[2]
23	20,946,123	21,032,829	1.35	*PTCHD4*
18	62,415,280	62,498,670	1.32	*EPS8L1, RDH13, NCR1, FCAR, KIR2DL5A, KIR3DL1, KIR3DS1*
20	26,181,870	26,332,917	1.25	*ITGA1*[2]*, PELO*
5	6,318,396	6,420,053	1.20	*ZDHHC17, CSRP2, U6, E2F7*
23	9,228,154	9,312,557	1.19	*SCUBE3, ZNF76, DEF6, PPARD*[2]
6	30,757,118	30,803,603	1.11	*ATOH1, GRID2*
2	121,473,182	121,525,407	1.07	*CLDN12, SHOX, bta-mir-2887-1*

[1] BTA – *Bos taurus* autosome.

[2] Prioritized candidate genes identified by GUILDify and ToppGene analysis.

Discussion

The *ATG7* gene (BTA22) encodes an essential activating enzyme for autophagy and cytoplasmic transport to the vacuole. The autophagy flow is important to preserve muscle mass and maintain the integrity of the myofiber. In beef cattle, increased *ATG7* gene expression was correlated with skeletal muscle growth and body weight during the fattening period, regardless of the muscle evaluated (Nakanishi *et al.*, 2019).

The *EXT1* gene (BTA14) encodes a glycosyltransferase responsible for heparan sulfate (HS) polymerization, related to fibroblast growth factor (Nadanaka *et al.*, 2008). The interaction between growth factors and their receptors is regulated by the amount of HS. A defect in HS biosynthesis due to somatic mutations in *EXT1* can cause a localized break in the negative feedback loop that regulates chondrocyte proliferation and maturation, allowing premature differentiation and therefore aberrant bone growth (Duncan *et al.*, 2001).

Another prioritized candidate gene, *ITGA1* (BTA20), encodes a subunit of the cell-surface receptor integrin α1β1, which acts in the regulation of hepatic glucose and lipid metabolism under conditions of overnutrition *in vivo*. On a high fat diet, mice with the *ITGA1* gene inactivated show severe hepatic insulin resistance and decreased hepatic fat accumulation (Williams *et al.*, 2015). In Korean Hanwoo cattle, the *ITGA1* gene was significantly associated with intramuscular fat deposition (de las Heras-Saldana *et al.*, 2020).

A key role for the *PPARD* gene (BTA23) in controlling the biological processes that drive ruminal epithelial cell development in newborn calves was suggested by Naeem *et al.* (2012). The increase in its expression, in response to an improved nutrition plan, would indicate an important role promoting the use of long-chain fatty acids as a substrate for oxidation in cell membranes during differentiation in ruminal development.

The present GWAS allowed the identification of chromosomal regions associated with feet and legs malformation in Nellore cattle. The role of the prioritized candidate genes in skeletal muscle development (*ATG7*), aberrant bone growth (*EXT1*), lipid metabolism (*ITGA1*, *PPARD*) and intramuscular fat deposition (*ITGA1*) adds to the evidence of the importance of these genes for feet and legs malformation in the studied population. Future investigations on these candidate genes and their causal mutations may contribute to the understanding of the genetic and physiological mechanisms that regulate this trait.

Acknowledgements

To 'Coordenação de Aperfeiçoamento de Pessoal de Nível Superior' (CAPES, finance code 001) for financial support and to the Alliance Nellore for supplying the data.

References

Chen J., Bardes E.E., Aronow B.J., and Jegga, A.G. (2009) Nucleic Acids Research 37(2):305–311. https://doi.org/10.1093/nar/gkp427

de las Heras-Saldana S., Chung K.Y., Kim H., Lim D., Gondro C., *et al.* (2020) Genes 11(11):1–17. https://doi.org/10.3390/genes11111381

Duncan G., McCormick C., and Tufaro F. (2001) J. Clin. Invest. 108(4):511–516. https://doi.org/10.1172/JCI13737

Fonseca P.A., Suárez-Vega A., Marras G., and Cánovas Á. (2020) GigaScience 9(12):giaa149. https://doi.org/10.1093/gigascience/giaa149

Guney E., Garcia-Garcia J., and Oliva B. (2014) Bioinformatics 30(12):1789–1790. https://doi.org/10.1093/bioinformatics/btu092

Masuda Y., Legarra A., Aguilar I., and Misztal I. (2019) J. Anim. Sci. 97(3):50–51. https://doi.org/10.1093/jas/skz258.101

Misztal I. (2020), BLUPF90. Available at: http://nce.ads.uga.edu/wiki/doku.php?id=start

Nadanaka S., Ishida M., Ikegami M., and Kitagawa H. (2008) J. Biol. Chem. 283(40):27333–27343. https://doi.org/10.1074/jbc.M802997200

Naeem A., Drackley J.K., Stamey J., and Loor J.J. (2012) J. Dairy Sci. 95(4):1807–1820. https://doi.org/10.3168/jds.2011-4709

Nakanishi T., Tokunaga T., Ishida T., Kobayashi I., Katahama Y. *et al.* (2019) Asian-Australasian J. Anim. Sci. 32(4):592–598. https://doi.org/10.5713/ajas.18.0370

O'Brien A.M.P., Mészáros G., Utsunomiya Y.T., Sonstegard T.S., Garcia J.F., *et al.* (2014) Livest. Sci. 166:121–132. https://doi.org/10.1016/j.livsci.2014.05.007

Sargolzaei M., Chesnais J.P., and Schenkel F.S. (2014) BMC Genomics 15(478). https://doi.org/10.1186/1471-2164-15-478

Vargas G., Neves H.H.R., Cardoso V., Munari D.P., and Carvalheiro R. (2017) J. Anim. Sci. 95(6):2379–2384. https://doi.org/10.2527/jas2016.1327

Wang H., Misztal I., Aguilar I., Legarra A., and Muir W.M. (2012) Genet. Res. 94(2):73–83. https://doi.org/10.1017/S0016672312000274

Williams A.S., Kang L., Zheng J., Grueter C., Bracy D.P., *et al.* (2015) J. Biol. Chem. 290(10):6546–6557. https://doi.org/10.1074/jbc.M114.615716

384. Genome-wide association studies for milk yield in *Bos taurus* × *Bos indicus* crossbred cattle in India

M. Al Kalaldeh[1*], M. Swaminathan[2], V. Podtar[2], S. Jhadav[2], V. Dhanikachalam[2], A. Joshi[2] and J.P. Gibson[1]

[1]Centre for Genetic Analysis and Applications, School of Environmental and Rural Science, University of New England, Armidale, NSW 2350, Australia; [2]BAIF Development Research Foundation and Central Research Station, Uruli Kanchan, Pune, Maharashtra 412202, India; malkala2@une.edu.au

Abstract

In India, crossbred dairy cows are used to exploit milk production potential of exotic *Bos taurus* dairy breeds with the adaptation to difficult environments of indigenous *Bos indicus* cattle. A genome-wide association study (GWAS) for milk yield was undertaken to detect effects of indigenous versus exotic origin at a locus, as well as the effects of quantitative trait loci (QTL) in linkage disequilibrium (LD) with single nucleotide polymorphism (SNP) within ancestral breeds. Analyses found a large number of QTL that define ancestral differences between *B. taurus* and *B. indicus*. The most significant between-breed region was on chromosome 5 located between 44 to 55 Mbp, with the positive allele coming from *B. indicus*. The results indicate that evolutionary differences in milk production (in crossbreds) between *B. indicus* and dairy *B. taurus* cattle are due to many causal mutations, some of which appear to have large effects.

Introduction

India is the world`s largest milk producer, with an annual production of 198 million tons, and increasing by approximately 6% each year (DAHD 2021). Half of the country`s milk production comes from cattle, with the largest proportion from crossbreds between *B. taurus* dairy breeds and local *B. indicus* cattle. The only *B. Taurus* breeds known to have been used in crossbreeding are Holstein-Friesian and Jersey.

Gebrehiwot *et al.* (2020) found literature estimates of age of divergence of *B. taurus* and *B indicus* cattle to range from 200,000 to 2 million years, with most estimates in the range of 200,000 to 300,000 years. Given the time since the divergence between *B. taurus* and *B. indicus*, they are likely fixed for different alleles at many loci affecting milk production, and that different loci are involved in current genetic variation within *B. taurus* and *B. indicus*. If any causative loci are segregating in the two sub-species, it is not expected that they are in the same LD with SNP. In crossbred cattle, SNP alleles can be assigned to their ancestral origin based on the inferred origin of the phased haplotypes in which they sit. This provides an opportunity to identify QTL that segregate in crossbred cows that have been fixed (or nearly fixed) for opposite alleles in *B. taurus* versus *B. indicus* breeds, and also identify QTL that segregate within the ancestral populations. The objective here is to undertake a simultaneous GWAS for milk yield in crossbred dairy in India that separate SNP associations into: (1) the effect of indigenous versus exotic origin (i.e. QTL that are fixed or nearly fixed for opposite alleles in the two ancestral populations); (2) the effect of QTL in LD with SNP within indigenous ancestral breed(s); (3) the effect of QTL in LD with SNP within exotic ancestral breed(s).

Materials & methods

Phenotypes and genotypes. The data analysed originate from the Enhanced Genetic Gains program (EGP) of BAIF Development Research Foundation (https://baif.org.in/). Phenotypes are 123,042 test-day (TD) records from 4,968 crossbred (*B. taurus* × *B. indicus*) cows in smallholder herds that have an average of 1.7 milking cows, collected from 6 states between February 2016 and September 2020. Cows were genotyped with the GGP 50k assay and genotypes imputed for 737,073 SNP. TD records were corrected for non-genetic effects using a mixed model as described in (Al Kalaldeh *et al.* 2021).

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_384, © M. Al Kalaldeh *et al.* 2022

Assigning breed origin to alleles in crossbred cows. This was undertaken using three steps. (1) Phasing the genotypes of both the reference and crossbred cows using Eagle software (Loh *et al.* 2016). Phasing was based on 676,577 common SNP that had minor allele frequencies (MAF) higher than 1%. The reference dataset includes 101 purebred Indigenous animals from the EGP project, 21 pure Holstein and 21 pure Jersey animals from the HapMap project, and 21 pure Friesian animals from the Scottish Rural University College (SRUC), all genotyped with the high density (HD) SNP array. (2) Assigning ancestral origin (*B. taurus* vs *B. indicus*) to the phased haplotypes of crossbred cows using PCAdmix software (*Brisbin et al.* 2012). (3) Assigning breed origin to SNP alleles in crossbred cows. SNP alleles were assigned to their breed origin based on the inferred origin of the phased haplotypes in which they sit. There are 4 possible combinations of ordered breed origin and 4 possible combinations of SNP genotypes, resulting in 16 different ancestry genotype combinations at each locus.

GWAS. Wombat software (Meyer 2007) was used to perform a simultaneous GWAS regression on the allele count for: (1) breed origin; (2) SNP alleles inherited from the indigenous ancestor; (3) SNP alleles inherited from the exotic ancestor. The allele count for breed origin was based on the number of copies inherited from the indigenous ancestor. The genotypes were coded to breed origin as 0, 1, and 2 for no copies, one copy, or two copies coming from the indigenous ancestor, respectively. To avoid confounding of estimates, genotypes were classified into three groups. Group 1 includes 510,243 genotypes that had minor allele frequency (MAF)>0.01 for alleles of exotic and indigenous origin. Group 2 includes 104,292 genotypes that had MAF<0.01 for alleles of indigenous origin. Group 3 includes 62,362 genotypes that had MAF<0.01 for alleles of exotic origin.

For SNPs in Group 1, GWAS model 1 was:

$$\text{trait}=\text{BC}\times\text{ENV}+\text{SNP}_1+\text{SNP}_2+\text{SNP}_3+\text{G}+\text{herd}+\text{e} \quad (1)$$

For SNPs in Group 2, GWAS model 2 was:

$$\text{trait}=\text{BC}\times\text{ENV}+\text{SNP}_1+\text{SNP}_3+\text{G}+\text{herd}+\text{e} \quad (2)$$

For SNPs in Group 1, GWAS model 3 was:

$$\text{trait}=\text{BC}\times\text{ENV}+\text{SNP}_1+\text{SNP}_2+\text{G}+\text{herd}+\text{e} \quad (3)$$

where trait is the adjusted TD milk records averaged per cow, BC×ENV is the fixed effect of breed composition by environment as defined by Al Kalaldeh *et al.* (2021), SNP_1 is the allele substitution effect for breed origin, SNP_2 is the allele substitution effect for SNP alleles inherited from the indigenous ancestor, SNP_3 is the allele substitution effect for SNP alleles inherited from the exotic ancestor, G is the cow's random additive genetic effect estimated using the genomic relationship matrix, herd is the random herd effect, and e is the random error effect. Due to the degree of correlation between alleles of breed origin, alleles of exotic ancestor, and alleles of indigenous ancestor, which can affect the coefficients and *P*-values, the *P*-values were adjusted for genomic control by dividing the chi-squared test statistics by the genomic inflation factor (lambda). Lambda were 0.91, 1.07, and 0.97 for SNP effects of breed origin, alleles inherited from exotic ancestor, alleles inherited from indigenous ancestor, respectively. After performing GWAS, the false discovery rate (FDR) was calculated to control for multiple SNP testing.

Results

The frequency of indigenous breed origin alleles ranged from 0.30 to 0.40, with an average of 0.36, which is the estimated proportion of *B. indicus* ancestry in the crossbreds. The most prominent region detected for breed origin effect was between 44.8 and 55.1 Mbp on chromosome 5, which included 35 SNP that had FDR<5% (Group 1 results shown in Figure 1). This region has not previously been identified as having a large effect on milk yield. The average difference between *B. indicus* and *B. taurus* SNP genotypes is 0.60 kg/day (8% of average TD yield), with the positive allele coming from *B. indicus*. Considering a conservative FDR of 10%, 14 regions pass the threshold, and at the less conservative FDR 30%, a high proportion of the genome passes the threshold. These results indicate that causal differences between *B. indicus* and *B. taurus* occur at very many loci distributed across the genome. Figure 2 and 3 show Manhattan plots for LD within *B. indicus* and LD within *B. taurus* respectively, with no effects detected at FDR<30%.

Discussion

Many regions differentiating *B. indicus* from *B. taurus* ancestors are detected. However, no regions in LD within *B. indicus* and *B. taurus* are detected. Since only 36% of alleles are assigned to *B. indicus* and 64% to *B. taurus* origin, GWAS for within ancestor LD have substantially lower power than the between breed origin GWAS. For example, the power of GWAS based on LD within *B. indicus* is equivalent to performing a purebred GWAS on approximately 1,800 cows rather than the 4,968 cows analysed here. Also, both Holstein-Friesian and Jersey are known to have been used for crossbreeding in the states sampled in our data, and many different and usually undefined indigenous breeds are known to have been used. Thus, LD may well differ between ancestral breeds within the *B. taurus* and *B. indicus* ancestral origins traced here, further eroding the power of detection of most QTL. Analyses are underway to separate the effects due to LD within Holstein-Friesian versus Jersey ancestors, which is relatively accurately determined based on imputed genotypes. Differentiating alternative *B. indicus* ancestors is not possible because of the large number of ancestral breeds and the low SNP genetic variation among *B. indicus* breeds (Strucken *et al.* 2021).

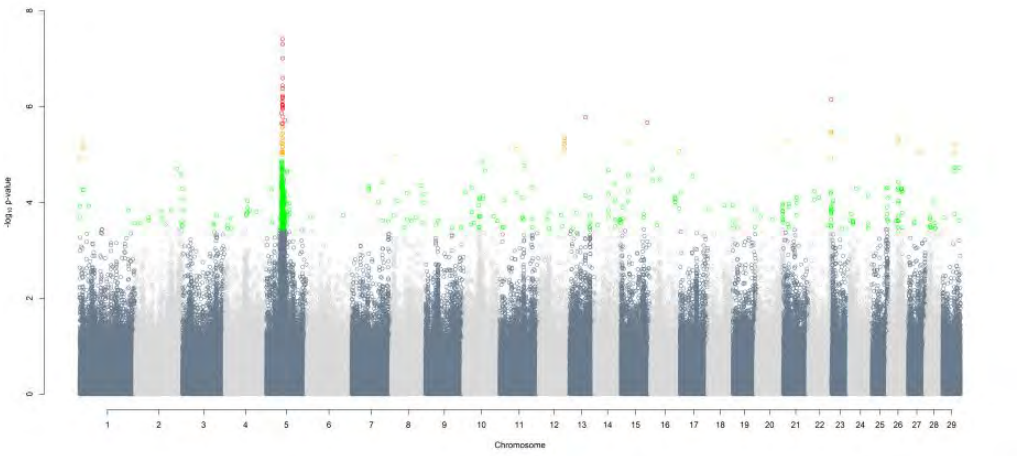


Figure 1. Manhattan plot of SNP effects for breed origin. SNPs passed FDR 5, 10 and 30% are highlighted in red, orange, and green, respectively.

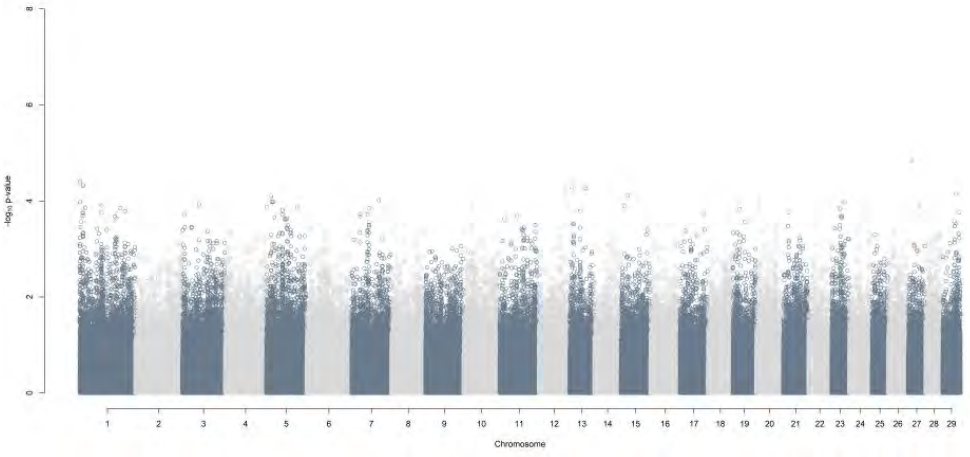


Figure 2. Manhattan plot of SNP effects for SNP inherited from *B. indicus* ancestral origin.

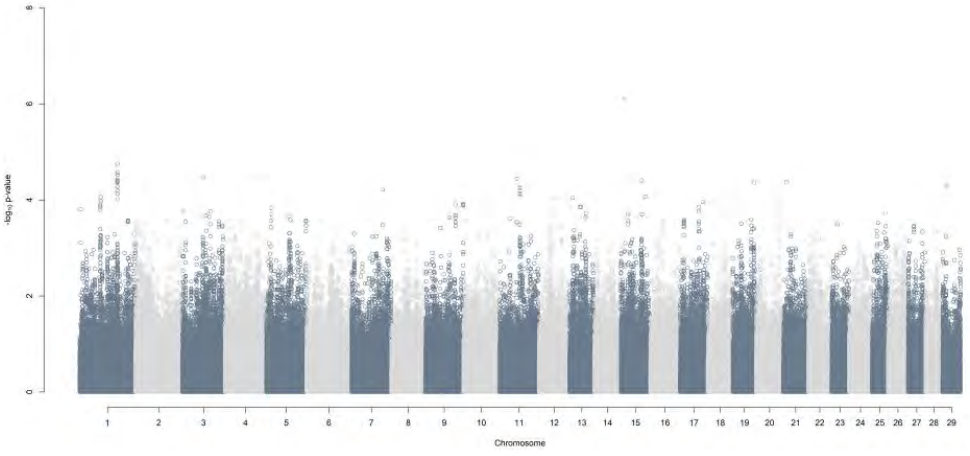


Figure 3. Manhattan plot of SNP effects for SNP inherited within *B. taurus* ancestral origin.

References

Al Kalaldeh M., Swaminathan M., Gaundare Y., Joshi S., Aliloo H. *et al.* (2021) Genetics Selection Evolution 53(1)**:** 73. https://doi.org/10.1186/s12711-021-00667-6

Brisbin A., Bryc K., Byrnes J., Zakharia F., Omberg L. *et al.* (2012) Human Biology 84(4)**:** 343-364, 322. https://doi.org/doi.org/10.3378/027.084.0401

Dahd (2021) Annual Report 2020-2021, pp. Ministry of Fisheries, Animal Husbandry and Dairying, Government of India.

Gebrehiwot N.Z., Strucken E.M., Aliloo H., Marshall K. and Gibson J.P. (2020) BMC Genomics 21(1)**:** 869. https://doi.org/10.1186/s12864-020-07270-x

Loh P.-R., Palamara P.F. and Price A.L. (2016) Nature Genetics 48(7)**:** 811-816. https://doi.org/10.1038/ng.3571

Meyer K. (2007) Journal of Zhejiang University SCIENCE B 8(11)**:** 815-821. https://doi.org/10.1631/jzus.2007.B0815

Strucken E.M., Gebrehiwot N.Z., Swaminathan M., Joshi S., Al Kalaldeh M. *et al.* (2021) Genetics Selection Evolution 53(1)**:** 47. https://doi.org/10.1186/s12711-021-00640-3

385. Grouping of highly correlated variables improves performance of genomic evaluations

D. Wittenburg* and J. Klosa

Research Institute for Farm Animal Biology (FBN), Institute of Genetics and Biometry, Wilhelm-Stahl-Allee 2, 18196 Dummerstorf, Germany; wittenburg@fbn-dummerstorf.de

Abstract

Extensive genetic information enables the identification of genome regions associated with trait expression but this may lead to high model dimensions and the occurrence of dependencies among predictor variables. A grouped penalised regression approach provides a framework to account for these dependencies adequately. Previous studies showed that the overall performance of genomic evaluations improved if genomic markers were grouped according to their strength of interdependence. Here, we investigate different options for grouping taking linkage and linkage disequilibrium into account and additionally considering adjacency between markers. As the extent of dependence is influenced by the family structure in a livestock population, we investigated an empirical and a theoretical measure thereof. Options have been validated with simulated data. Generally, the more groups were determined, the better the performance of the grouped penalisation approach.

Introduction

To investigate the association between genetic and phenotypic variation, genomic markers (such as single nucleotide polymorphisms; SNPs) are used as predictor variables in a statistical model, e.g. $y = X\beta + e$. The $n \times p$ design matrix X contains the genotype codes at p SNPs of n individuals with corresponding regression coefficients $\beta = (\beta_1,\ldots\beta_p)'$. Any fixed or nuisance effects are discarded here for convenience. Due to the (partly) tight linkage of markers, multicollinearity occurs in X. Grouping is a universal strategy to cope with this situation. No matter what the actual grouping technique is (e.g. hierarchical clustering, k-means clustering), it shall employ a measure of association suited to the family structure in a population, which is characterised by non-random mating in livestock. Then, pre-specified groups of highly associated markers can be utilised, for instance, in a grouped penalised regression approach to differentiate genome regions with impact on trait expression (y) from neutral regions. Sparse-group lasso (SGL) has been developed to allow for sparsity of effect estimates between and within groups of predictors (Simon *et al.* 2013; Klosa *et al.* 2021). In an earlier evaluation of penalised regression approaches (Klosa *et al.* unpublished), SGL turned out to be a robust and high-performing method in various scenarios of genetic architecture (differing in heritability, proportion of non-zero effects, etc.). In extension to this, our objective is to investigate suitable options for grouping of markers regarding the family structure and to quantify the impact on the ability to detect trait-associated loci and on the accuracy of genomic prediction with SGL.

Materials & methods

Grouped penalised regression approach. Estimates of genetic effects β captured by SNP markers were obtained with SGL, as implemented in the R package seagull v. 1.1.0, i.e.

$$\min_\beta \frac{1}{2n}\left\|y - X\beta\right\|_2^2 + \alpha\lambda\left\|\beta\right\|_1 + (1-\alpha)\lambda\sum_{l=1}^{L}\sqrt{p_l}\left\|\beta^{(l)}\right\|_2$$

with shrinkage parameter λ>0 and mixing parameter α=0.5. The regression coefficients were grouped into L bins with group size p_l, l=1,…,L. As a baseline approach, a single group of markers was built for each chromosome. Alternatively, involving linkage and linkage disequilibrium (LD) enabled measuring the strength of association between markers more locally. The squared correlation between two columns of X

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_385, © D. Wittenburg and J. Klosa 2022

served as an approximation to population LD (in terms of R-squared) between a pair of SNPs; the matrix including all pairwise *observed* correlations was denoted as R_{obs}. The correlation among SNP pairs was also derived theoretically conditional on the genetic information of parents and a linkage map (Wittenburg *et al.* 2021). Hence, this approach was suited to the underlying family structure in the data. The resulting *expected* correlation matrix was termed R_{exp}. Both R^2_{obs} and R^2_{exp} served as a similarity matrix for a clustering procedure in the following step. As a first option, grouping followed a greedy approach with no restrictions ('noc') on the proximity of markers (Wittenburg *et al.* 2021) as implemented in the R package hscovar v. 0.4.2. This way, markers with squared correlation >0.8, no matter if they were located close to each other or not, were grouped into the same bin. As a second option, we applied hierarchical clustering under adjacency constraints (Ambroise *et al.* 2019), provided in the R package adjclust v. 0.6.3. This approach necessitated a banded structure of the underlying similarity matrix, for which we regarded a bandwidth of 10%. Hierarchical clustering required a decision rule for selecting the optimal number of groups. We considered two possibilities: (1) we selected the same number of clusters as obtained with the greedy approach ('adj1') or (2) we applied a slope heuristics *capushe* ('adj2'; Baudry *et al.* 2012), which did not rely on prior information.

Simulation study. A simulation study was conducted resembling a typical family structure in cattle. In total, 100 samples were generated containing 10 half-sib families with 900, 100 and 100 progeny each in the training data, validation data and independent test data, respectively. Two chromosomes, each containing 1,660 SNPs on one Morgan length, were simulated with realistic extent of LD. Scenarios with three or 9 groups of trait-associated and highly correlated loci per chromosome were set up with varying sparsity within such groups (33 or 66% of loci had non-zero effect). Genetic variation explained 30% of the phenotypic variation. The data and more details on the simulation study are available at https://dx.doi.org/10.22000/432.

Using the validation data, we inspected sensitivity and specificity of SGL's selection property resulting from the different groupings of markers. The number of true positive detections was recorded allowing a window of 0.1 centiMorgan (i.e. one to two SNPs) to both sides of the causal variant. The mean squared error (MSE) of regression coefficients provided another measure of performance. Using the independent test data, precision of prediction was determined as correlation between simulated and predicted phenotype.

Results

Grouping based on the expected (observed) correlation matrix with no constraints yielded on average L=1,480 (1,481) groups of markers over all chromosomes, whereas considering adjacency constraints followed by cluster selection with *capushe* resulted in L=26 groups of markers based on both correlation matrices.

Though sensitivity was at a high level with the baseline approach (2 groups of markers) throughout scenarios, the remaining performance criteria were inferior to other options of grouping. With increasing number of groups of markers, we observed the typical trade-off: sensitivity reduced and specificity increased. Additionally, MSE improved. Precision of prediction was almost unaffected by the grouping method chosen. Against our expectation, grouping with adjacency constraints ('adj1') did not improve the performance of SGL in general, only in case of sparse signals (three groups of trait-associated loci and sparsity 33%). All results are summarised in Table 1; bold entries highlight the best outcome in each setting. Since grouping was very similar with R_{exp} and R_{obs}, performance of SGL did not differ in essence and only results with option 'observed-noc' were presented exemplarily in Table 1. Furthermore, as a representative example, Figure 1 displays estimated marker effects in a single repetition. It shows that few regions with

strong effects can be localised best with option 'noc'. Though the number of groups is equal with 'noc' and 'adj1', spurious non-zero effects appeared more often with option 'adj1'.

Table 1. Accuracy-related measures for SGL with different options of grouping SNPs.

Scenario[1]	Grouping option[2]	Sensitivity	Specificity	MSE[3]	Precision
3 groups and sparsity 33%	per-chrom	**0.96**	0.53	1.17	0.49
	expected-noc	0.39	**0.99**	0.57	0.52
	expected-adj1	0.54	**0.99**	**0.52**	**0.53**
	expected-adj2	0.97	0.59	0.86	0.52
	observed-noc	0.39	**0.99**	0.57	0.52
3 groups and sparsity 66%	per-chrom	**1.00**	0.44	3.71	0.51
	expected-noc	0.86	**0.91**	2.97	**0.53**
	expected-adj1	0.82	**0.91**	**2.93**	**0.53**
	expected-adj2	**1.00**	0.48	3.34	0.52
	observed-noc	0.86	**0.91**	2.96	**0.53**
9 groups and sparsity 33%	per-chrom	0.98	0.43	3.78	0.51
	expected-noc	0.82	**0.87**	2.81	**0.52**
	expected-adj1	0.82	0.84	3.10	**0.52**
	expected-adj2	**0.99**	0.43	3.51	**0.52**
	observed-noc	0.81	**0.87**	**2.80**	**0.52**
9 groups and sparsity 66%	per-chrom	**0.98**	0.32	6.48	0.49
	expected-noc	0.64	0.89	5.67	0.50
	expected-adj1	0.68	0.87	**5.44**	**0.51**
	expected-adj2	0.90	0.47	5.90	0.50
	observed-noc	0.64	**0.90**	5.67	0.50

[1] Number of trait-associated groups per chromosome and percentage of loci with non-zero effect within group.

[2] Per-chrom: one group of markers per chromosome; expected/observed: grouping based on expected/observed correlation matrix; noc/adj: grouping approach without/with adjacency constraints; adj1: number of clusters corresponds to noc; adj2: number of clusters was derived from slope heuristics.

[3] MSE$\times 10^{-4}$ of regression coefficients.

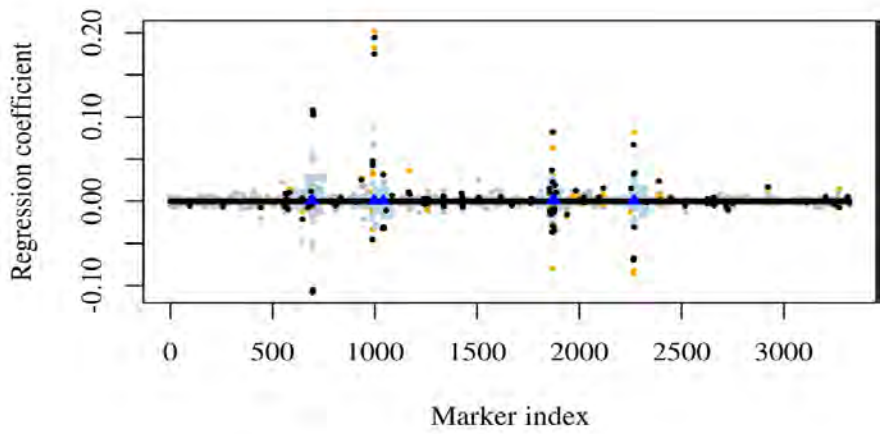


Figure 1. Estimated regression coefficients with sparse-group lasso for a single repetition. Grouping was obtained with greedy approach ('noc', black), hierarchical clustering under adjacency constraints ('adj1' in orange and 'adj2' in lightblue) or chromosome-wise ('per-chrom', grey). Simulated locations of causal variants (three groups with sparsity 66%) are coloured in blue.

Discussion

The aim was to exploit the grouped nature of predictor variables according to the underlying family structure. We worked with the observed and expected correlation between SNP genotypes. By construction, any deviation between observed and expected correlation vanishes with increasing number of families. However, only the expected correlation is able to account for imbalanced family sizes if they exist. Setting up the elements of R_{exp} considers the haplotype distribution within family and its contribution to the population-wide correlation depending on family sizes. We expect that if the number of families is smaller than investigated here or the family design becomes imbalanced, the larger the benefit from R_{exp} compared to R_{obs} will turn out. This study, however, concentrated on the grouping approach. Influences arising from sample and family sizes will be verified in future research.

The SGL is not only able to cope with multicollinearity in X, it can also incorporate different sources of data, such as expression or metabolomics data. In those applications, similar to Boulesteix *et al.* (2017) and Klau *et al.* (2018), predictor variables may be grouped according to their origin and prior knowledge (e.g. gene pathways).

Conclusions

This study showed that considering adjacency for grouping of markers did not improve the performance of SGL *per se*. The number of groups was decisive – in particular, the more groups, the better the specificity of detecting trait-associated loci. Grouping with or without adjacency constraints is applicable to any study design (family structure, SNP panel, etc.) and requires in the simplest case only the observed pairwise correlations between SNP genotypes. Hence, GWAS can rely on all marker information available without prior pruning.

References

Ambroise C., Dehman A., Neuvial P., Rigaill G. and Vialaneix N. (2019) Algorithms Mol Biol 14, 22. https://doi.org/10.1186/s13015-019-0157-4

Baudry J. P., Maugis C. and Michel B. (2012) Stat Comput 22:455-470. https://doi.org/10.1007/s11222-011-9236-1

Boulesteix A.-L., De Bin R., Jiang X. and Fuchs, M. (2017) Comput Math Methods Med 2017:7691937. https://doi.org/10.1155/2017/7691937

Klau S., Jurinovic V., Hornung R., Herold T. and Boulesteix A.-L. (2018) BMC Bioinf 19:322. https://doi.org/10.1186/s12859-018-2344-6

Klosa J., Simon N., Westermark P. O., Liebscher V. and Wittenburg D. (2021) BMC Bioinf 21:407. https://doi.org/10.1186/s12859-020-03725-w

Simon N., Friedman J., Hastie T. and Tibshirani R. (2013) J Comp Graph Stat 22:231-245. https://doi.org/10.1080/10618600.2012.681250

Wittenburg D., Doschoris M. and Klosa J. (2021) BMC Bioinf 22:79. https://doi.org/10.1186/s12859-021-04010-0

386. A Monte-Carlo approximation of the null hypothesis for genome wide association sing the windows variance approach

D. López-Carbonell[*#], M. Ramírez[#], J. Altarriba, H. Srihi and L. Varona

[†]Instituto Agroalimentario de Aragón (IA2), Facultad de Veterinaria, Universidad de Zaragoza, Calle Miguel Servet, 177, 50013, Zaragoza, Spain; 767339@unizar.es; [#]these authors contributed equally

Abstract

The interest in Genome Wide Association Studies (GWAS) has widely increased over the last few years, and Frequentist and Bayesian methods have been developed. The use of genomic windows has been proposed to gather SNP information. Nevertheless, spurious associations have been detected, which may be attributed to different causes. The aim of this study is to detect these spurious associations and weigh SNP importance to avoid them. This can be reached starting from a real data set, that was analysed using single step GWAS methodology. Thereafter, new simulated data were generated, using statistical simulation, as well as following data and pedigree structure. The same methodology was performed to analyse simulated data. Lastly, real and simulated results were compared in order to distinguish those spurious associations that appear in both simulated and real data, and those who were associated with genetic variance in the trait, which sometimes can be concealed.

Introduction

Genome Wide Association Studies (GWAS) have become a common technique to locate Quantitative Trait Loci (QTL) (Visscher *et al.*, 2017). Some authors have proposed using genomic windows (Sahana *et al.*, 2010) instead of testing significance for each single nucleotide polymorphism (SNP) to avoid the low amount of information provided for each single SNP. This kind of approach is possible due to the equivalence of GBLUP and SNP-BLUP methods (Wang *et al.*, 2012) and some Frequentist (Chen *et al.*, 2017) and Bayesian (Fernando *et al.* 2017; Legarra *et al.* 2018) test have been proposed. However, Li *et al.* (2021) have noticed that the results of the GWAS analysis using genomic windows are strongly influenced by the SNP density. As a result, they have offered two solutions: (1) using windows size of equal numbers of SNP; and (2) using a specific test for each genomic window.

Here we noticed that the results can also be strongly affected by some other causes such as the LD (linkage disequilibrium), the allelic frequencies (AF) or the degree of genetic similarity across individuals within each genomic region (DS). We propose a simple Monte-Carlo approximation in order to take into account all the different possible circumstances.

Materials & methods

Data. The data used in this study consist of 16,320 records of *Pelvis Width* measured by trained classifiers in the Pirenaica beef cattle population and a pedigree of 35,161 individual-sire-dam entries. Additionally, there were 755 genotypes at our disposal generated with Axiom Bovine platform (*Axiom_BovMDv3*) from ThermoFisher Scientific (Waltham, MA USA). The filtering consisted of excluding genotypes with a minor allele frequency lower than 0.05 and a SNP call rate less than 0.90 in the overall population. After this, 31,509 SNP markers were used.

Model of analysis. The data were analysed using the *airemlf90* software (Misztal *et al.*, 2002) with the following model:

$$y = Xb + Wp + Zu + e \quad (1)$$

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_386, © D. López-Carbonell *et al.* 2022

where b is the vector of systematic effects that include sex -2 levels-, classifier -25 levels- and age in years -9 levels-, p is the vector or random herd-year-season, u is the vector of breeding values and e is the residuals. Later, we used the *postGSf90* software to calculate the percentage of additive genomic variance explained by genomic windows of 12 SNP (approx.. 1 Mb). This process was repeated iteratively 3 times for SNP weighing as suggested by Wang *et al.* (2012).

Simulation. We generated 1000 populations using statistical simulation over the available pedigree. The statistical simulation calculates the simulated breeding values (u_s) and the herd-year-season (p_s) effects with the following multivariate Gaussian distribution:

$$\mathbf{u}_S \sim N(\mathbf{0}, \mathbf{A}\hat{\sigma}_A^2) \quad (2)$$

$$\mathbf{p}_S \sim N(\mathbf{0}, \mathbf{I}\hat{\sigma}_P^2) \quad (3)$$

where $\hat{\sigma}_A^2$ and $\hat{\sigma}_P^2$ are the estimated additive genetic and herd-year-season variances. Furthermore, the simulated data were generated in the phenotyped individuals using the same model as (1), using the estimated systematic effects and residuals sampled from a Gaussian distribution with a mean of 0 and the estimated residual variance. The analysis of each simulated dataset was performed with the same procedure as that of the real one.

Results

The results of the GWAS analysis of the real data set using *postGSf90* software are presented in Figure 1. The results of the analysis identified two genomic regions at BTA6 (bp 64,514,509) and BTA16 (bp 25,796,893), which explains more than 2% of the additive genetic variance. However, when the simulated datasets were analysed (Figure 2), the results showed that the average estimates of the percentage of variance explained by the genomic segments is heterogeneously distributed along the genome, albeit the data sets were simulated under a pure polygenic model.

These discrepancies can be explained by several reasons, such as the patterns of LD, AF, or DS. If we focus on the two highlighted segments of the first analysis, it can be noticed (Figure 3) that the result obtained in the genomic region of BTA6 (2.24%) is clearly higher than all the results generated by the simulated datasets in the same genomic region. In contrast, simulated data sets explained more percentage of variance in BTA16, and 35 of 1000 explained a higher percentage of variance than real data set (2.75%).

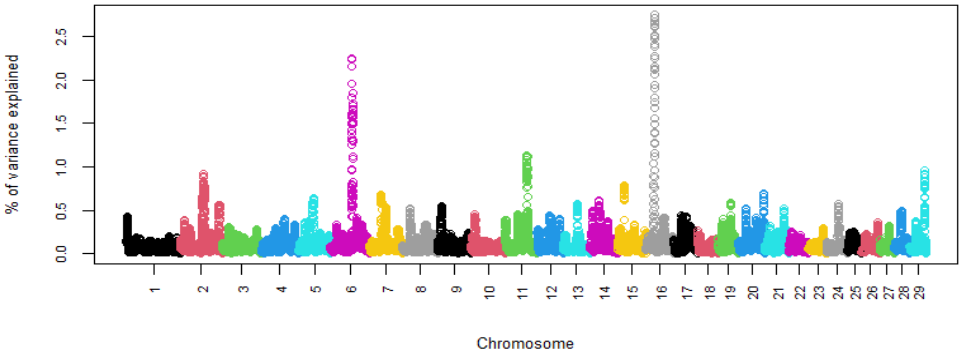


Figure 1. Manhattan plot for the percentage of additive genetic variance explained by windows of 12 SNP in the real dataset.

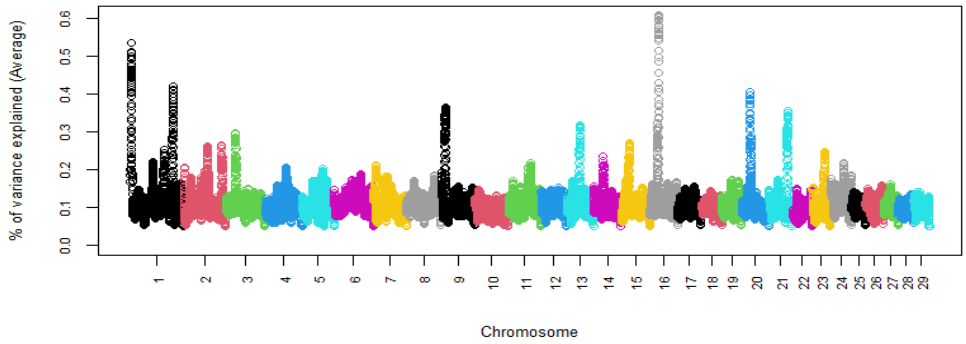


Figure 2. Manhattan plot for the average percentage of additive genetic variance explained by windows of 12 SNP in the simulated data sets.

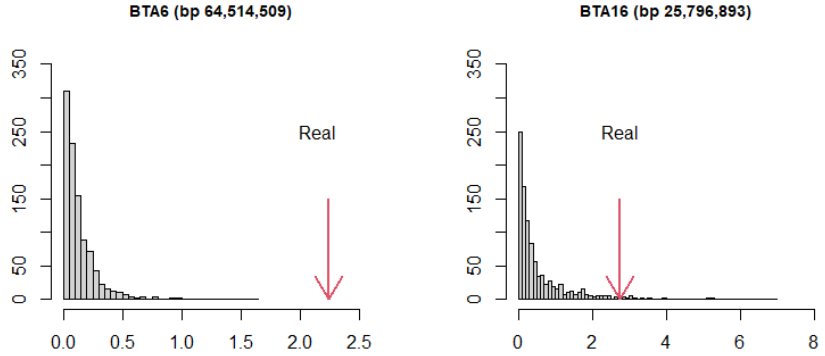


Figure 3. Empirical distribution of the percentage of additive variance explained by the genomic regions of 12 SNP centered at BTA6 and BTA16 in simulated data sets and the percentage of additive variance explained by the same regions in the real data set (red arrow).

Finally, we estimate an empirical *P*-value by using the empirical distribution generated with the simulated datasets at each genomic region. The results are presented in Figure 4.

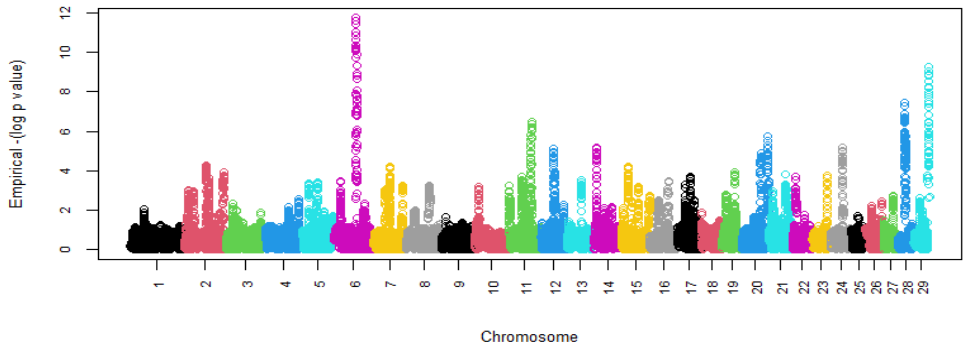


Figure 4. Manhattan plot of empirical log *P*-values using genomic regions of 12 SNP along the autosomal chromosomes.

Having corrected the real data using simulated ones, the results strongly confirmed the QTL detected in BTA6 but showed that the QTL previously identified at BTA16 was spurious and additionally point some other interesting genomic regions at BTA11, BTA12, BTA14, BTA20, BTA24, BTA28 or BTA29 as potentially candidates of carrying genes that affect the additive genetic variation of the trait.

Discussion

We have shown that the results of GWAS study based on the variance explained by genomic segments were not only affected by the genomic density but also by some other circumstances associated with the chip design or the population genomic features. To generate a null hypothesis, we have suggested using statistical simulation, which is based on the fact that breeding values follow a multivariate Gaussian distribution. Therefore, it reflects a pure polygenic model that considers the resemblance among relatives caused by the additive genetic variation. The case study illustrates that the inferences can be strongly modified even when the number of SNP is set to a fixed number.

References

Chen C., Steibel J. P., and Tempelman R. J. (2017) Genetics 206(4):1791–1806. https://doi.org/10.1534/genetics.117.202259

Fernando R, Toosi A, Wolc A, Garrick D, Dekkers J. (2017) J Agric Biol Environ Stat 22(2): 172–193, https://doi.org/10.1007/s13253-017-0277-6.

Legarra A., Ricard A., and Varona L. (2018). G3: Genes, Genom, Genet 8(7):2301–2308. https://doi.org/10.1534/g3.118.200336

Li J., Wang Z., Fernando R., and Cheng H. (2021). Genet Sel Evol, 53(1):45. https://doi.org/10.1186/s12711-021-00638-x

Misztal I., Tsuruta S., Strabel T., Auvray B., Druet T., *et al.* (2002) Proc. of 7th World Congr. Genet. Appl. Livest. Prod. - Montpellier, France.

Sahana G., Guldbrandtsen B., Bendixen C., and Lund M. S. (2010) Anim Genet 41(6):579–588. https://doi.org/10.1111/j.1365-2052.2010.02064.x

Visscher P. M., Wray N. R., Zhang Q., Sklar P., McCarthy M. *et al.* (2017) Am J Hum Genet 101 (1):5-22. https://doi.org/10.1016/j.ajhg.2017.06.005

Wang H., Misztal I., Aguilar I., Legarra A., and Muir W. M. (2012) Genet Res 94(2):73–83. https://doi.org/10.1017/S0016672312000274

387. Sequence-based association analyses on X chromosome in six dairy cattle breeds

M.P. Sanchez[1*], C. Escouflaire[2], A. Baur[2], C. Hozé[2], A. Capitan[2], S. Fritz[2] and D. Boichard[1]

[1]Université Paris Saclay, INRAE, AgroParisTech, GABI, 78350 Jouy-en-Josas, France; [2]Allice, 75595 Paris, France; marie-pierre.sanchez@inrae.fr

Abstract

Genomic evaluations in cattle are commonly based on effects of autosomal SNPs, the X chromosome (BTAX) being generally excluded because of its hemizygous state in males. We estimate here the effects of imputed sequence variants of BTAX in six dairy breeds and for 11 complex traits related to milk production, milk composition, mastitis resistance, fertility and stature. We detected QTL in almost all breeds and for almost all traits. They were distributed throughout BTAX and explained between 0.1 and 5.2% of the total phenotypic variance of the trait. In each of these QTL, we identified genes (*GPC3*, *SEPTIN6*, *ZC3H12B*, *COL4A6*, *IDS*, *SLC16A2*, *MIR363*, *HS6T2*, *MAMLD1*, *COL4A6*, *DGKK*, *DDX3X*, *GPC3*, *ERCC6L*, *FRMPD4*, *ENOX2*, *KLHL13*, *HMGB3*, and *CDKL5*) for which the functional link with the studied traits remains to be established. We therefore show here the importance of BTAX in the genetic determinism of complex traits in dairy cattle.

Introduction

Due to its specificity (hemizygous state in males, dosage compensation in females, different genomic relationship matrix), BTAX is often excluded from genetic analyses in cattle. Yet, BTAX is the second largest chromosome of the bovine genome and it is particularly rich in genes. Excluding BTAX from genomic evaluations can thus lead to the loss of a significant part of the genetic variability of traits (Su *et al.*, 2014). To assess the effects of BTAX on dairy traits, we report here results of association analyses from imputed sequences of cows of six French breeds for eleven traits related to milk production, mastitis resistance, fertility and stature.

Materials & methods

Cows and traits. We analysed 236,496 cows from three national (MON, NOR, HOL) and three regional (ABO, TAR, VOS) breeds with 50k genotypes and phenotypes (Table 1). Traits included were milk (MY), protein (PY) and fat (FY) yields; protein (PC) and fat (FC) contents; somatic cell score (SCS) and clinical mastitis (MAST; except for VOS); calving-first insemination interval (ICFI), heifers' (HCR) and cows' (CCR) conception rate; stature (STAT). Phenotypes used were yield deviations (YD) i.e. phenotypes adjusted for non-genetic effects.

Table 1. Features of populations and sequence variants analysed.

Breed	Abbrev.	# 50k genotypes	# HD genotypes	# WGS animals[2]	# variants after filtering[1]	Mean[1] imputation R^2	Mean[1] MAF
Abondance	ABO	7,449	199	9	154,966	0.69	0.18
Tarentaise	TAR	3,969	179	12	181,473	0.79	0.19
Vosgienne	VOS	2,910	181	4	170,560	0.77	0.20
Montbéliarde	MON	61,881	522	63	186,368	0.81	0.18
Normande	NOR	78,472	526	45	190,280	0.82	0.17
Holstein	HOL	81,815	804	1,053	201,554	0.81	0.17

[1] Variants with a MAF≥0.005 and with an imputation R^2≥0.20.

[2] 2,712 multi-breed sequences used for imputation.

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_387, © M.P. Sanchez *et al.* 2022

Imputation and association analyses. We considered only the X-specific non-pseudo autosomal region (non-PAR) that covers the main part of BTAX (0-133.3 Mbp on the ARS-UCD1.2 reference genome; Johnson *et al.*, 2019). For imputation, males were considered homozygous for all SNPs and pedigree information was not included. HD genotypes of 32,268 SNPs were first imputed from genotypes of 1,147 EuroGMD SNPs with FImpute (Sargolzaei *et al.*, 2014) using animals with HD genotypes as a reference (Table 1). Then, 778,576 sequence variants were imputed using a multi-breed population of 2,712 animals from the RUN8 reference panel of the 1000 Bull Genomes consortium (Bouwman *et al.*, 2018) with Minimac (Howie *et al.*, 2012). Allele dosages were tested in within breed association analyses using GCTA software (Yang *et al.*, 2011). All phenotypes were measured on females, we therefore applied the following model: $y = 1\mu + xb + u + e$, where y is the vector of YD; μ is the overall mean; b is the additive fixed effect of the variant tested; x is the vector of imputed allele dosages; $u \sim N(0,G\sigma_u^2)$ is the vector of random polygenic effects, with G the genomic relationship matrix based on autosomal 50k SNPs, and σ_u^2 is the polygenic variance; $e \sim N(0,I\sigma_e^2)$ is the vector of random residual effects. We analysed variants with a MAF≥0.005 and with a Minimac imputation $R^2 \geq 0.20$ i.e. from 154,966 to 236,011 variants depending on the breed (Table 1). A unique threshold was used corresponding to 5% significance after Bonferroni correction for ~160,000 variants ($-\log_{10}(P)=6.5$). In each breed and trait, variants with significant effects located less than 10 Mbp apart were grouped together to define QTL. Confidence intervals (CIs) of QTL were then determined based on the positions of variants in the upper third of the peak (van den Berg *et al.*, 2016). The percentage of phenotypic variance (σ_p^2) explained by each QTL was calculated for the variant with the most significant effect as $\%\sigma_p^2 = 100.2p(1-p)\alpha^2/\sigma_p^2$, with p the frequency and α the estimated allelic substitution effect.

Results

Association analyses detected QTL in all breeds, except TAR and VOS, and for all traits, except HCR (Table 2). We identified 191 QTL (0 to 10 QTL per breed × trait analysis). For milk traits, we detected QTL in four breeds, the most significant ones being mainly located at the beginning of BTAX for PC and FC and throughout BTAX (~30, 60, 85, 102, 110, 130 Mbp) for MY, PY, and FY (Figure 1). We found a higher number of QTL in national breeds (63, 57, and 63 in MON, NOR, and HOL, respectively) than in regional breeds (8, 0, and 0 in ABO, TAR, and VOS, respectively). The cumulative effects of the QTL per breed explained from 0.1% (ICFI in HOL and CCR in MON) to 5.2% (FY in ABO) of σ_p^2. CIs of the QTL contained 1 to 2,193 variants with significant effects. In total 39,429 unique variants were identified.

Most variants included in CIs were located in intergenic regions (72%) while the other (28%) were in genes or in the vicinity of genes, mainly in introns (20.3%) (Table 3). The best-ranked variants in the peaks i.e. the top 10 and top 1 variants, were more frequently located in genes (37.3 and 47.8%, respectively). Non-intergenic variants ranked first in the peaks were located in 58 different genes. The top 1 variants with the most significant effects ($-\log_{10}(P)\geq 20$) were located in *GPC3* (FC), *SEPTIN6* (PC), *ZC3H12B* (PC), and *COL4A6* (STAT) in MON; *IDS* (STAT) in NOR; and *SLC16A2* (FC) and *MIR363* (PC) in HOL. For fertility traits (ICFI and CCR), the genic top 1 variants of the QTL, detected in national breeds (MON, NOR

Table 2. Number of QTL [total % of phenotypic variance explained by the QTL].

Breed[1]	MY	PY	FY	PC	FC	SCS	MAST	ICFI	HCR	CCR	STAT
ABO	1 [0.9]	1 [0.9]	2 [5.2]	2 [4.1]	1 [1.5]	0	1 [0.9]	0	0	0	0
MON	10 [1.8]	7 [1.4]	7 [1.2]	9 [2.9]	9 [1.5]	3 [0.4]	0	8 [1.3]	0	1 [0.1]	9 [1.5]
NOR	8 [0.8]	10 [1.0]	7 [0.6]	7 [0.8]	5 [0.7]	4 [0.4]	0	8 [1.2]	0	0	8 [3.1]
HOL	8 [0.5]	7 [0.6]	9 [1.0]	10 [1.2]	10 [1.8]	9 [1.1]	0	1 [0.1]	0	1 [0.2]	8 [0.9]

[1] No QTL were found in regional TAR and VOS breeds.

Table 3. Functional annotation of all, top 10 or top 1 variants located in CIs of the QTL.

	All variants		Top 10 variants		Top 1 variants	
	Number	%	Number	%	Number	%
Intergenic region	28,168	72.0	1,084	62.7	106	58.2
Intronic region	7,955	20.3	467	27.0	58	31.9
Upstream region	1,471	3.8	87	5.0	8	4.4
Downstream region	1,231	3.1	75	4.3	7	3.8
Synonymous variants	87	0.2	8	0.5	2	1.1
3'UTR	80	0.2	3	0.2	0	0
Missense variants	79	0.2	3	0.2	0	0
5'UTR	33	0.1	0	0.0	0	0
Other exonic variants	28	0.08	3	0.2	1	0.5
Total	39,132	100	1,730	100	182	100

or HOL), were located in *HS6T2*, *MAMLD1*, *COL4A6*, *DGKK*, *DDX3X*, *ENSBTAG00000048527*, *GPC3*, *ERCC6L*, and *FRMPD4* genes while for udder health traits (MAST and SCS in ABO, MON, NOR, and HOL), they were located in *ENOX2*, *KLHL13*, *HMGB3*, and *CDKL5*.

Discussion

In mammals, the non-PAR region of BTAX differs from autosomes because of its hemizygous state in males (XY). Mechanisms of dosage compensation occur in females (XX) to inactivate one BTAX copy and therefore ensuring an equal expression of genes located on BTAX in both sexes. Little is known about the complex mechanisms of BTAX inactivation in cows but both copies appear to be equally expressed in the mammary gland at the population level suggesting a random inactivation of either copy (Couldrey *et al.*, 2017). Because of its specificities, BTAX is usually excluded from genetic analyses and few GWAS were conducted in cattle (e.g. Fortes *et al.*, 2020). In the present study, we focused on this particular chromosome to assess the effects of its variants imputed at the sequence level on various complex dairy traits (milk production, udder health, fertility, and stature) in cows from six dairy cattle breeds. After adapting the imputation process (males assumed homozygous and no pedigree) and considering paternal and maternal X chromosomes equally expressed at the population level, we applied the model used for autosomes in association analyses. Except for breeds with the lowest number of cows (TAR and VOS) and for one fertility trait (HCR), we identified QTL for all breeds and traits analysed. As expected, due to the number of cows analysed in each breed, the number of QTL was higher in the largest breeds and these QTL had larger effects. A higher number of QTL with more significant effects was also found for the most heritable traits (milk production, milk composition, and stature). Depending on the breed and on the trait, these QTL explained up to 5.2% of σ_p^2. We also highlighted QTL for less heritable traits that may explain up to 1.3% (fertility) and 1.1% (udder health) of σ_p^2 in some breeds. In particular, for SCS in MON, NOR, and HOL, we found that 0.4, 0.4, and 1.1% of σ_p^2 was explained by 3, 4, and 9 QTL located on BTAX, respectively. These values can be compared to the values we estimated in a previous study for the same trait in MON, NOR, and HOL bulls i.e. 6.3, 0, and 2.6% for 3, 0, and 2 QTL detected on all the 29 bovine autosomes, respectively (Tribout *et al.*, 2020). As illustrated in Figure 1, QTL detected on BTAX presented large CIs. This phenomenon is probably due to the linkage disequilibrium, more extended in the non-PAR region where no recombination occurs in males (Zhang *et al.*, 2020). We show here that dozens of candidate genes are located in X-linked QTL. However, functional links between these genes and traits analysed remain to be established. This study shows the importance of BTAX in the genetic determinism of complex traits in dairy cattle and supports the results of Su *et al.* (2014) who found that markers on BTAX contribute to accuracy of genomic predictions in Holstein breed.

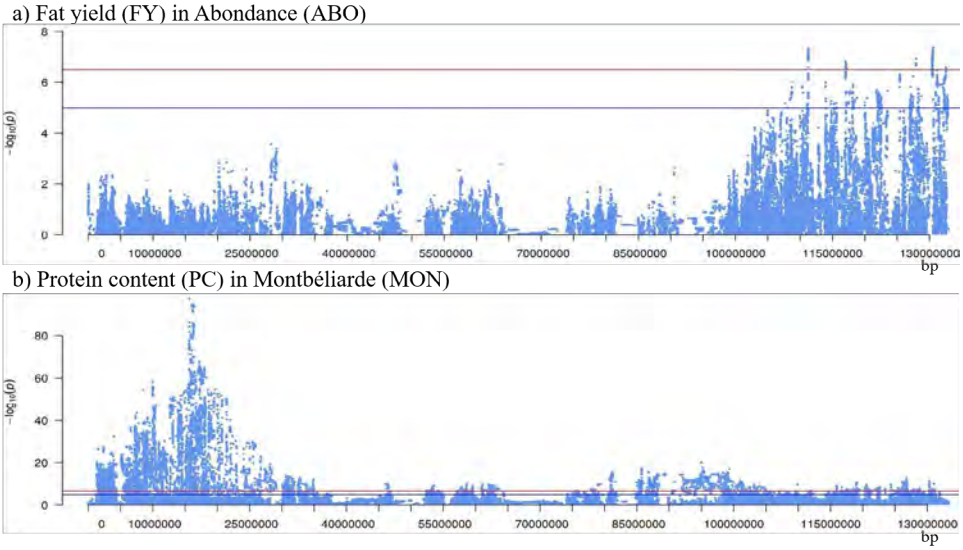


Figure 1. $-\log_{10}(P)$ values plotted against the position of variants on X chromosome.

References

Bouwman A. C., Daetwyler H. D., Chamberlain A. J., Ponce C. H., Sargolzaei M. *et al.* (2018) Nat Genet 50(3):362-367. https://doi.org/10.1038/s41588-018-0056-5

Couldrey C., Johnson T., Lopdell T., Zhang I. L., Littlejohn M. D. *et al.* (2017) J Dairy Sci 100:5491-5500. https://doi.org/10.3168/jds.2016-12490

Fortes M., Porto-Neto L. R., Satake N., Nguyen L. T., Freitas A. C. *et al.* (2020) Genet Sel Evol 52:46. https://doi.org/10.1186/s12711-020-00563-5

Johnson T., Keehan M., Harland C., Lopdell T., Spelman R. J. *et al.* (2019) J Dairy Sci 102:1-5. https://doi.org/10.3168/jds.2018-15638

Howie B., Fuchsberger C., Stephens M., Marchini J., and Abecasis G. R. (2012) Nat Genet 44(8):955-959. https://doi.org/10.1038/ng.2354

Sargolzaei M., Chesnais J., and Schenkel F. (2014) BMC Genomics 15(1):478. http://doi.org/10.1186/1471-2164-15-478

Su G., Guldbrandtsen B., Aamand G. P., Stranden I. and Lund M. S. (2014) Genet Sel Evol 46:47. http://doi.org/10.1186/1297-9686-46-47

Tribout T., Croiseau P., Lefebvre R., Barbat A., Boussaha M. *et al.* (2020) Genet Sel Evol 52:55. https://doi.org/10.1186/s12711-020-00575-1

van den Berg I., Boichard D., Lund M. S. (2016). J Dairy Sci 99(11):8932-8945. https://doi.org/10.3168/jds.2016-11073

Yang J., Lee S., Goddard M., and Visscher P. (2011) Am J Hum Genet 88(1):76-82. https://doi.org/10.1016/j.ajhg.2010.11.011

Zhang J., Kadri N. K., Mullaart E., Spelman R., Fritz S. *et al.* (2020) Heredity 125:304-316. https://doi.org/10.1038/s41437-020-0341-9

388. Pleiotropic effects on carcass and meat quality traits in crossbred beef cattle

F.M. Rezende[1*], E. Rodriguez[1], J.D. Leal-Gutiérrez[2], M.E. Elzo[1], D.D. Johnson[1], C. Carr[1] and R.G. Mateescu[1]

[1]University of Florida, P.O. Box 110910, 2250 Shealy Dr, Gainesville, FL, USA; [2]University of California, 9500 Gilman Dr, San Diego, CA 92093, USA; frezende@ufl.edu

Abstract

Our objective was to identify genomic regions with pleiotropic effects on carcass and meat quality traits in a multibreed Angus-Brahman population. Data included phenotypes from 2,384 steers, of which 1,038 were genotyped. Single-trait genome-wide association studies were first used to investigate the relevance of additive genetic effects. A second analysis was performed to capture genomic regions with pure direct effects on each trait, removing regions with indirect effects. Four pleiotropic (BTA5, BTA7 and BTA18) and two suggestive pleiotropic regions (BTA10 and BTA19) were identified. Candidate genes encode key players in cell growth, muscle development, lipid metabolism and fat deposition. Functional analysis revealed GO terms directly related to carcass quality, meat quality, and tenderness in beef cattle, including calcium-related processes, and modulation of cell-cell adhesion. These results contribute with novel information about the complex genetic architecture of carcass and meat quality traits in crossbred beef cattle.

Introduction

To attenuate the impacts of tropical and subtropical conditions on growth rate and reproductive performance of Taurine (*Bos taurus taurus*) beef breeds, producers use crossbreeding between European Taurine and Zebu (*Bos taurus indicus*) breeds as a strategy to enhance beef production. In subtropical areas of the United States, Angus-Brahman crosses are preferred for beef production over other Taurine-Zebu combinations.

While carcass quality is the primary factor determining the value of a carcass in the beef industry supply chain, consumers evaluate beef products at purchase time based on visual quality and at consumption time based on sensory quality. These traits are of even greater importance in crossbred cattle used in subtropical and tropical regions for their superior adaptability because they tend to underperform compared to their purebred counterparts.

Many of these traits are challenging and expensive to measure and unavailable until late in life or after the animal is harvested, hence unrealistic to improve through traditional phenotypic selection, but perfect candidates for genomic selection. Before genomic selection can be implemented in crossbred populations, it is important to explore if pleiotropic effects exist between carcass, visual meat quality and sensory meat quality traits. Therefore, our objective was to identify genomic regions with pleiotropic effects on carcass and meat quality traits in a multibreed Angus-Brahman population.

Materials & methods

The cattle population for this study consisted of 2,384 steers from the University of Florida multibreed Angus-Brahman herd born between 1989 and 2018 and raised under very similar management system throughout the years. The breed composition of animals in this unique multibreed population ranged from 100% Angus to 100% Brahman, including all crosses in between them. Phenotypes for marbling (MARB), ribeye area (REA), tenderness (TEND), juiciness (JUIC), color, texture (TEXT) and firmness (FIRM) were analyzed. Genotyping was carried out on 1,038 of the 2,384 animals using the Bovine GGP F250 array,

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_388, © F.M. Rezende *et al.* 2022

which contains 221,115 SNP. A total of 125,042 autosomal SNP markers with MAF>0.01 and call rate higher than 90% were included in the genomic analysis.

Variance components, heritabilities, additive genetic correlations, and phenotypic correlations were estimated using single-trait and two-trait single-step genomic best linear unbiased prediction (ssGBLUP). Animal mixed models included the direct additive genetic and residual as random effects, year of birth as a class effect, and age at slaughter as a covariate, except for TEND and FLAV where age at slaughter was not significant. Computations were performed with the *airemlf90* package from the BLUPF90 (Misztal *et al.*, 2009).

Genome-wide scan for pleiotropic effects. Single-trait genome-wide association studies were carried out using the weighted ssGBLUP (WssGBLUP; Wang *et al.*, 2012) procedure to scan for genomic regions with pleiotropic effects on carcass and meat quality traits. First, the relevance of direct additive genetic effects was investigated fitting all fixed and random effects from the variance component models. Second, conditional genome scan fitting all other traits as covariates in the aforementioned model allowed correcting for indirect effects and capturing genomic regions with pure direct effects on the trait under analysis. Genomic windows explaining more than 1% of direct additive genetic variance were considered associated with the analysed trait. Common genomic regions involving overlapping windows associated with two or more phenotypes in both analyses were considered as pleiotropic regions. Additionally, common regions from overlapping windows explaining more than 1% of the direct additive genetic variance for one trait and between 0.9 and 1% of the direct additive genetic variance for another trait were considered as suggestive pleiotropic regions.

Functional analysis. Top 20 SNP markers with the largest absolute estimated effect across two or more traits within each pleiotropic region were used to identify genes with pleiotropic effects. A SNP marker was assigned to a particular gene if it was located within the gene. Genes with a known function directly or indirectly associated with carcass and meat quality traits were defined as candidate pleiotropic genes. Following, GO and pathway enrichment and clustering analyses of all annotated genes within pleiotropic regions were performed.

Results

Estimates of heritabilities, phenotypic and genetic correlations between analyzed traits are presented in Table 1. Heritability estimates for MARB, REA and TEND were moderate, ranging from 0.44 to 0.53, and low estimates were observed for the other sensory panel and visual meat quality traits (0.12-0.19). Ribeye area had consistently the lowest phenotypic correlations with all other traits (-0.01 to 0.10). Positive moderate phenotypic correlations existed between MARB and TEND (0.32), MARB and JUIC (0.32), TEND and JUIC (0.51), and JUIC and COLOR (0.36). Negative moderate phenotypic correlations were estimated between MARB and FIRM (-0.37) and JUIC and FIRM (-0.33). High and favorable additive genetic correlations was observed between MARB and JUIC (0.66) and MARB and FIRM (-0.38), TEND and JUIC (0.64) and TEND and TEXT (-0.53).

The WssGBLUP analyses correcting for indirect effects identified a total of 3,460 non-overlapping 1-Mb genomic windows for MARB, 3,083 for REA, 3,218 for TEND, 3,557 for JUIC, 3,502 for COLOR, 3,315 for TEXT, and 3,341 for FIRM. Out of these, 4, 8, 5, 7, 6, 2, and 5 windows explained more than 1% of the direct additive genetic variance for MARB, REA, TEND, JUIC, COLOR, TEXT, and FIRM, respectively. Four pleiotropic genomic regions and two suggestive pleiotropic regions were identified. The first pleiotropic region on BTA5 (26.7-27.5 Mb) was significantly associated with REA (2.68%) and MARB (1.27%), while the second region (56.2-56.9 Mb) explained a high proportion of the direct additive genetic variance in

Table 1. Heritabilities (h^2, diagonal), phenotypic (above diagonal) and direct additive genetic (cor, below diagonal) correlations between carcass and meat quality traits in a multibreed Angus-Brahman population.[1]

Trait	MARB	REA	TEND	JUIC	COLOR	TEXT	FIRM
MARB	**0.49**	0.19	0.32	0.32	0.03	-0.22	-0.37
REA	-0.03	**0.53**	0.10	-0.03	0	-0.01	0.04
TEND	0.21	0	**0.44**	0.51	0.16	-0.08	-0.17
JUIC	0.66	-0.15	0.64	**0.15**	0.36	-0.01	-0.33
COLOR	-0.19	0.02	0	-0.54	**0.15**	0.23	-0.19
TEXT	-0.30	0.24	-0.53	-0.99	0.02	**0.12**	0.16
FIRM	-0.38	0.24	-0.16	-0.32	-0.22	-0.24	**0.19**

[1] Standard error (SE): $0.04<SE_{h^2}<0.07$ and $0.06<SE_{cor}<0.5$.

REA (4.78%), TEND (2.45%), MARB (2.09%), TEXT (1.55%) and FIRM (1.34%). One genomic region on BTA7 (51.5-52.5 Mb) had effect on MARB (2.24%), TEND (1.63%), and TEXT (1.36%). The genomic region between 61.9 and 62.5 Mb on BTA18 accounted for 2.17 and 1.10% of the genetic variance for COLOR and TEND, respectively. A suggestive pleiotropic region on BTA10 (76.2-77.2 Mb) explained 1.04 and 0.99% of the genetic variance for JUIC and MARB. The suggestive pleiotropic region on BTA19 (27.0-28.0 Mb) explained 1.12 and 0.96% of the genetic variance for TEND and TEXT.

These pleiotropic genomic regions contained about 260 genes with effect across two or more traits. Two candidate genes were identified in the first pleiotropic region on BTA5, seven in the second region on BTA5, one gene in the pleiotropic window on BTA7, and two genes in the pleiotropic region on BTA18. In addition, two genes were mapped in the suggestive pleiotropic region of BTA10, and the region on BTA19 contain four candidate genes. Overrepresented terms for GO Biological Processes within pleiotropic regions included 'Regulation of Apoptotic Process', 'Regulation of Cell Proliferation', 'Linoleic Acid Metabolic Process', and 'Cell Adhesion via Plasma Membrane Adhesion Molecules'. Moreover, overrepresented terms for GO Molecular Functions included 'Iron Ion Binding', 'Calcium Ion Binding', 'Steroid Hormone Receptor Activity', 'DNA Binding', 'Translation Initiation Factor Activity', and 'Transcription Factor Activity'.

Discussion

Heritability estimates for carcass and meat quality traits in the multibreed Angus-Brahman population were generally consistent with values reported in the literature (Mateescu *et al.*, 2014). Examination of direct additive genetic correlations between traits in this study is important to understand the challenges and limitations that could result from the inclusion of any of these traits in selection schemes. The moderate favorable direct additive genetic correlation observed between MARB and TEND was lower than other estimates of 0.40 (Reverter *et al.*, 2003) and 0.61 (Wheeler *et al.*, 2010), but comparable to estimates for purebred Brahman (Riley *et al.*, 2003), reinforcing the long-held belief of a unique fat-tenderness relationship in *B. t. indicus* versus *B. t. taurus* cattle.

The two candidate genes (*CSAD* and *TNS2*) in the first pleiotropic region on BTA5 are directly involved in muscle physiology and lipid metabolism (Wen *et al.*, 2019; Ishii *et al.*, 2013). The second BTA5 region is of particular importance because of its pleiotropic effects on most of the traits under investigation. This region harbors three candidate genes (*LRP1*, *MYO1A*, *NACA*) involved in muscle development, myoblast proliferation, and lipid metabolism (Dato and Chiabrando, 2018; Lv *et al.*, 2019, Berger *et al.*, 2012). Besides the obvious effect on MARB and REA, these genes could also affect TEND, TEXT and FIRM given the impact of muscle fiber diameter and density on these traits (Pearson, 1990, Lv *et al.*, 2019). The region on

BTA7 contains *Protocadherin Beta 1* (*PCDHB1*), which may directly impact marbling, tenderness, and texture. Cadherins are structural proteins that play important roles in cell adhesion and differentiation in several bovine tissues (Martignani *et al.*, 2020), thus could be involved in processes that lead to less tender and visually coarser meat. The suggestive pleiotropic region on BTA10 harbors *SYNE2*, and obvious candidate gene due to its possible role in proteolysis and cell compartmentalization (Zhang *et al.*, 2007).

Many of the biological pathways enriched within these regions were previously reported to be important for carcass quality, meat quality, and tenderness in beef cattle (Leal-Gutiérrez *et al.*, 2019). Numerous genes identified in pleiotropic regions are involved in calcium-related processes such as calcium ion binding, calcium channel, and calcium channel regulator. It was anticipated that calcium and potassium play a major role in meat tenderness because of their contribution to the proteolytic system responsible for muscle contraction and postmortem tenderization. Taken to together, our results contribute with novel information on the complex architecture of direct additive genetic correlation between carcass and meat quality traits in crossbred beef cattle.

References

Berger, F., Berkholz, J., Breustedt, T., Ploen, D., and Munz, B. (2012) Eur. J. Cell Biol. 91:150–155. https://doi.org/10.1016/j.ejcb.2011.10.004

Dato, V.A., and Chiabrando, G.A. (2018) Int. J. Mol. Sci. 19:1780. https://doi.org/10.3390/ijms19061780

Ishii, A. *et al.* (2013). Anim. Sci. J. 84, 675–682. https://doi.org/10.1111/asj.12063

Leal-Gutiérrez, J.D., Elzo, M.A., Johnson, D.D., Hamblen, H., and Mateescu, R.G. (2019) BMC Genomics 20:151. https://doi.org/10.1186/s12864-019-5518-3

Lv, C., Niu, S., Yan, S., Bai, C., Yu, X., Hou, J., *et al.* (2019) Poult. Sci. (98):3418–3425. https://doi.org/10.3382/ps/pez168

Martignani, E., Ala, U., Sheehy, P.A., Thomson, P. C., and Baratta, M. (2020). Sci. Rep. 10:14183. https://doi.org/10.1038/s41598-020-71179-4

Mateescu, R. G. (2014) Genetics of meat quality. In: D. J. Garrick and A. Ruvinsky (eds) The Genetics of Cattle. CABI, Wallingford, UK, pp. 544–570.

Misztal, I., Legarra, A., and Aguilar, I. (2009) J. Dairy Sci. 92, 4648–4655. https://doi.org/10.3168/jds.2009-2064

Pearson, A.M. (1990) Crit. Rev. Food Sci. Nutr. 29:167–196. https://doi.org/10.1080/10408399009527522ps/pez168

Reverter, A. *et al.* (2003) Aust. J. Agric. Res. 54:119–134. https://doi.org/10.1071/AR02086

Riley, D.G. *et al.* (2005). Meat Sci. 70:347–356. https://doi.org/10.1016/j.meatsci.2005.01.022

Wang, H., Misztal, I., Aguilar, I., Legarra, A., and Muir, W.M. (2012) Genet. Res. 94:73–83. https://doi.org/10.1017/S0016672312000274

Wen, C. *et al.* (2019) Mol. Nutr. Food Res. 63:1800536. https://doi.org/10.1002/mnfr.201800536

Wheeler, T.L., Cundiff, L.V., Shackelford, S.D., and Koohmaraie, M. (2010) J. Anim. Sci. 88:3070–3083. https://doi.org/10.2527/jas.2009-2497

Zhang, X. *et al.* (2007). Development 134, 901–908. https://doi.org/10.1242/dev.02783

389. Sequence-based multi-trait genome-wide association study for linear classification traits in Belgian Blue beef cattle

J.L. Gualdron Duarte[1,2*], A.S. Gori[2], G. Costa Monteiro Moreira[1], M. Georges[1], C. Charlier[1] and T. Druet[1]

[1]Unit of Animal Genomics, GIGA-R, University of Liège, 4000 Liège, Belgium; [2]Walloon Breeders Association (awe groupe), 5590 Ciney, Belgium; jlgualdron@awegroupe.be

Abstract

We performed a genome-wide association study in Belgian Blue beef cattle for 11 traits related to muscular development and height. Phenotypes were available for 14,762 genotyped cows and imputation was performed to the sequence level (11,537,240 variants). A total of 37 genome-wide significant QTL were identified with a single-trait analysis, corresponding to 11 unique pleiotropic QTL. We then used and compared two multiple-trait approaches to combine information from different traits to fine-map the QTL and increase the power. Although these approaches allowed identifying additional QTL, other were lost. Among the identified QTL, four corresponded to previously identified variants having negative effect when homozygous. In addition, three new missense variants in genes related to muscular development and height were identified.

Introduction

Genome-wide association studies (GWAS) are performed to identify variants associated with traits of interest Generally, they are performed with single-trait (ST) approaches although it is not uncommon to identify variants associated with several traits. Multiple-trait (MT) approaches could be advantageous in such cases, by improving the fine-mapping resolution or the mapping power. These MT approaches have a high computational cost and they can thus be applied on a limited number of traits. Bolormaa *et al.* (2014) proposed an approximate MT approach that requires only to perform ST approaches for each trait.

With the implementation of genomic selection in Belgian Blue beef (BBB) cattle, the number of genotyped individuals is continuously increasing, offering the possibility to perform GWAS at increasing power. A series of linear classification traits, related to conformation and muscular development, are routinely recorded on adult females. Such traits are typically correlated and affected by shared variants. Our objective was to conduct a GWAS in BBB cattle for 11 linear classification traits and to assess the benefit of MT approaches (Stephens, 2013), including the indirect approach proposed by Bolormaa *et al.* (2014).

Materials & methods

Genotype imputation. 14,762 cows with phenotypes were genotyped with custom low-density (LD) arrays with 8,576 to 32,318 SNPs or with the Bovine50K genotyping array (MD). To impute genotypes to the sequence level, we relied on a set of 890 bulls genotyped with the BovineHD array and on a sample of 242 whole-genome sequenced bulls. The imputation was achieved in three steps with ShapeIT4.2 (Delaneau *et al.*, 2019) and Minimac4 (Howie *et al.*, 2012), from LD (7,525 SNPs) to MD (28,893 SNPs), from MD to HD (611,322 SNPs) and from HD to whole-genome sequence level (11,537,240 selected SNPs and indels).

Single-trait GWAS. Association was tested with a Linear Mixed Model (LMM) approach implemented in GEMMA (Zhou and Stephens, 2012). The genomic relationship matrix obtained with the MD genotypes was used to capture the polygenic effects and the population structure. Each cow was phenotyped for 11 traits including body length, pelvis length, height, chest width, pelvis width, rib shape, rump, top muscling, shoulder muscling and buttock muscling (side and rear view). Phenotypes were corrected for fixed effects

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_389, © J.L. Gualdron Duarte *et al.* 2022

from the genetic evaluation. The number of independent SNPs was estimated to be equal to 500,000 whereas we determined that 7 independent traits were analysed. Accordingly, we set the significance threshold at 1.43e-8 after application of a Bonferroni correction for 3,500,000 tests.

Multiple-trait GWAS. We first applied the MT approach proposed by Stephens (2013) and implemented in GEMMA. In addition, we used the indirect approach proposed by Bolormaa *et al.* (2014) relying on a score computed from the signed t-values obtained for the SNPs and for each traits (in ST analyses). For one SNP, the score is computed as tVt' where t is the array of t-values estimated for the tested SNP for each of the N traits and V is the N × N correlation matrix between t-values, estimated from all the tested positions. This score is approximately distributed as a chi-squared with N degrees of freedom. For MT analysis, the significant thresholds were modified to 1e-7 and 5e-8 when one or two genome scans were performed (e.g. all traits simultaneously or two groups of six traits).

Results

The ST GWAS analysis identified a total of 37 QTL exceeding the significance threshold. However, several of these QTL co-localized in the same interval, suggesting pleiotropic QTL. All the identified QTL are summarized in Figure 1, where the 37 QTL could be condensed in a set of 11 unique QTL. As illustrated for a QTL on chromosome 19 (Figure 2), high correlations between SNP significance levels for several traits further support the presence of a pleiotropic QTL.

The first MT approach (Stephens, 2013) could not be applied on all traits simultaneously. We used the results from the QTL sharing with the ST approach to define two groups of 6 traits to perform MT analysis. By combining information from several traits, the MT approach helps to identify the causative variants. For instance, three previously identified causative variants are clearly the lead SNP with the MT approach whereas this was not the case for all ST analyses (Figure 2 and 3). The association signals obtained with the second MT mapping approach were highly correlated with those from the first approach (correlations ranging from 0.90 and 0.99) and pointed towards similar sets of causative variants, indicating that the

Figure 1. Heat-map showing association between QTL and traits. The circle sizes indicate strength of association, red borders around circles are used for significant associations, and the colour intensity denotes effect magnitude.

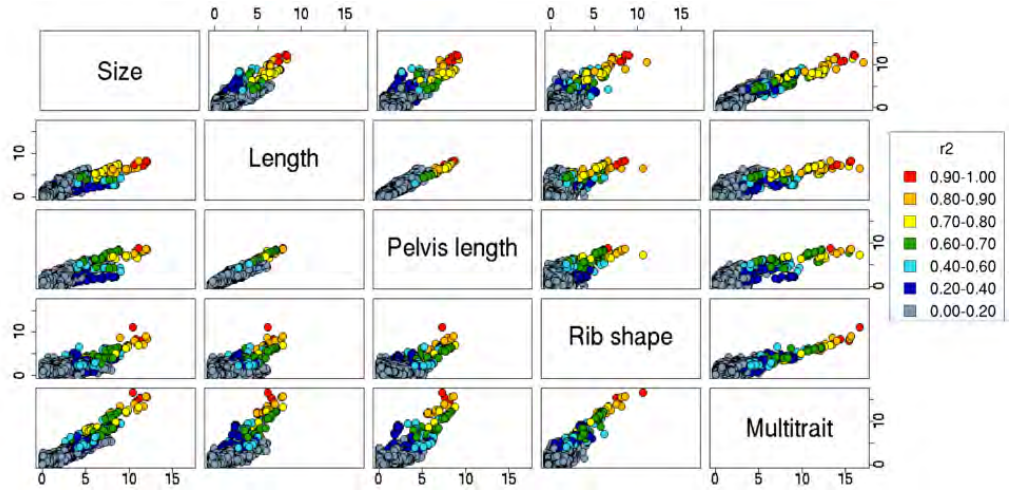


Figure 2. Correlation plot between ST and MT *P*-values (on -log10 scale) for a QTL on chromosome 19. Variants ±2.5 Mb from the QTL are included. The dot colours indicate the linkage-disequilibrium (r^2) between each variant and the lead SNP (on each row, the lead SNP corresponds to the analysis indicated on the diagonal).

indirect approach was a good approximation (although with slightly lower *P*-values). Although both approaches identified respectively two and three new QTL, two and one QTL were no longer significant (on BTA16 and 23). Inclusion of traits not associated with a QTL in MT approaches might thus reduce the power of association. Therefore, the optimal MT approach should not necessarily combine all traits simultaneously. These observations were confirmed by a MT analysis including 11 traits: although the mapping resolution was improved for the most pleotropic effects, the power decreased for some QTL associated with fewer traits (QTLs on BTA16 and 23 were lost). Nevertheless, five new QTL were identified thanks to increased power for pleiotropic QTL (and correction for fewer tests).

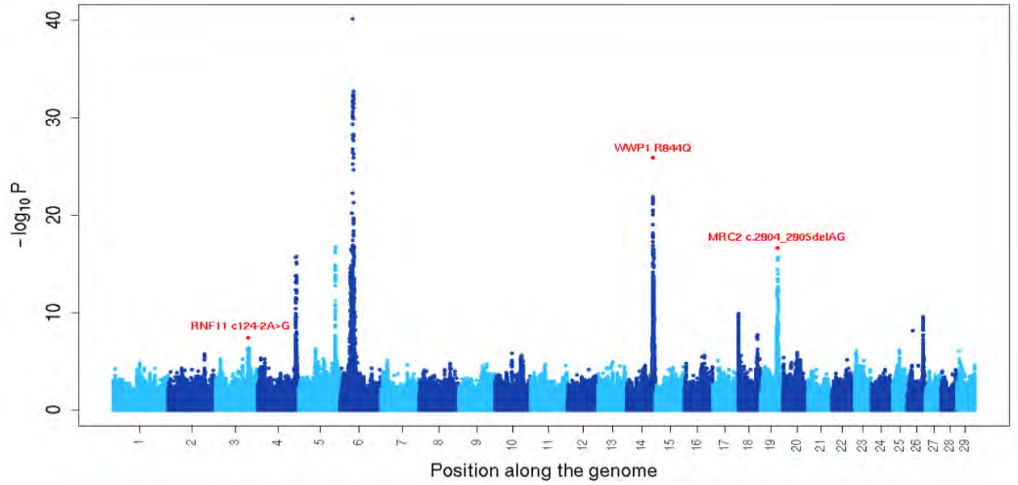


Figure 3. Example of Multiple-trait GWAS. Six traits were included: body length, pelvis length, height, chest width, pelvis width and rib shape. The red points are previously identified causative variants.

Four of the QTL are associated to variants altering the protein sequence previously identified, whereas three additional missense variants in genes related to height or muscular development were also identified.

Discussion

Several QTL detected in this study were previously identified in other studies. For instance, we found QTL on BTA5 (CCND2), BTA6 (LCORL and NCAPG), BTA23 and BTA26 reported for stature in Bouwman *et al.* (2018). In addition to these QTL, four QTL associated with deleterious variants presenting heterozygous advantage (e.g. Gualdron *et al.,* 2020), and other QTL might be more specific to the BBB cattle breed. The association with these recessive deleterious variants are still highly significant although their frequency has been decreasing in the population as a result of selection against carriers.

MT approaches allow obtaining a more precise association signal, in particular when most of the combined traits are associated with the variant. Although fewer tests were performed with such approaches, few new QTL were identified and some ST QTL were even lost, possibly because the MT approach included too many un-associated traits. Finally, the indirect MT approach from Bolormaa *et al.* (2014) was an excellent approximation of the standard MT approach with a lower computational cost. However, the *P*-values obtained with this second approach were less conservative and it remains to be determined whether the higher number of significant results we obtained results in increased power or a higher rate of false positives.

References

Bolormaa S., Pryce J.E., Reverter A., Zhang Y., Barendse W. *et al.* (2014) A Multi-Trait, Meta-analysis for Detecting Pleiotropic Polymorphisms for Stature, Fatness and Reproduction in Beef Cattle. PLoS Genet 10(3): e1004198. https://doi.org/10.1371/journal.pgen.1004198

Bouwman A.C., Daetwyler, H.D., Chamberlain A.J., Hurtado Ponce C., Sargolzaei M. *et al.* (2018) Meta-analysis of genome-wide association studies for cattle stature identifies common genes that regulate body size in mammals. Nat Genet 50, 362–367. https://doi.org/10.1038/s41588-018-0056-5

Delaneau O., Zagury JF., Robinson M.R., Marchini J.L., and Dermitzakis E.T. (2019) Accurate, scalable and integrative haplotype estimation. Nat Commun.10(1):5436. https://doi.org/10.1038/s41467-019-13225-y

Gualdrón Duarte J.L., Gori AS., Hubin X., Lourenco D., Charlier C. *et al.* (2020) Performances of Adaptive MultiBLUP, Bayesian regressions, and weighted-GBLUP approaches for genomic predictions in Belgian Blue beef cattle. BMC Genomics 21, 545. https://doi.org/10.1186/s12864-020-06921-3

Howie B., Fuchsberger C., Stephens M., Marchini J., Abecasis GR. (2012) Fast and accurate genotype imputation in genome-wide association studies through pre-phasing. Nat Genet 44(8): 955–959. https://doi.org/10.1038/ng.2354.

Stephens M. (2013) A Unified Framework for Association Analysis with Multiple Related Phenotypes. PLoS ONE 8(7): e65245. https://doi.org/10.1371/journal.pone.0065245

Zhou X., and Stephens M. (2012) Genome-wide efficient mixed-model analysis for association studies. Nat Genet 44(7): 821–824. https://doi.org/10.1038/ng.2310

390. Integrating imputed structural variants and tandem repeats in a GWAS for growth and carcass traits in F_2 pig crosses

I. Blaj[1*], J. Tetens[2,3], J. Bennewitz[4], G. Thaller[1] and C. Falker-Gieske[2]

[1]Institute of Animal Breeding and Husbandry, Kiel University, Hermann-Rodewald-Straße 6, 24118 Kiel, Germany; [2]Department of Animal Sciences, Georg-August-University, Burckhardtweg 2, 37077 Göttingen, Germany; [3]Center for Integrated Breeding Research, Georg-August-University, Wilhelmsplatz 1, 37073 Göttingen; [4]Institute of Animal Science, University of Hohenheim, Garbenstraße 17, 70599 Stuttgart, Germany; iblaj@tierzucht.uni-kiel.de

Abstract

Structural variants and tandem repeats are relevant sources of genomic variation but are not commonly used in genome-wide association studies due to challenging identification and genotyping. Here, we profiled these variants via state-of-the-art strategies in the founder animals of four F_2 crosses based on whole-genome sequence data (20×). After imputation, an association study in the F_2 generation was carried out for average daily gain, backfat thickness, meat to fat ratio, and carcass length. 56.56% of the significant variants were not in high LD with previously significantly associated SNPs and small indels identified for the same traits in an earlier study and thus presumably not tagged in a standard association study. Besides proposing specific causative mutations related to *COL21A1, BMP5, BMP2, ISL1*, or *SHANK2*, the findings of this study demonstrate that structural variants and tandem repeats play an important role in biological regulation often via mechanisms driven by long non-coding RNAs.

Introduction

A genome-wide association study (GWAS) is an effective approach to discover phenotype-genotype associations with growing potential when using whole-genome sequence (WGS) data and advanced analytical methods. The common understanding of the term genotype refers to SNPs and possibly small indels (<50 bp), nevertheless additional polymorphisms classes exist such as structural variants (SVs) and tandem repeats (TRs). SVs are described as large-scale (>50 bp) genomic alternations (e.g. deletions, insertions, duplications, inversions, and translocations), while TRs are characterised by their repetitive motifs and highly polymorphic nature. The contribution of SVs and TRs to quantitative trait variation is currently not fully understood as they are not routinely screened for. Therefore, the aim of this study was to conduct a GWAS based on imputed SVs and TRs and to assess their implications for growth and carcass traits in four pooled F_2 pig crosses.

Materials & methods

Data sources. The four F_2 pig populations included in this analysis were entirely genotyped with the Illumina PorcineSNP60 BeadChip and, additionally, WGS data was available for the F_0 (20×) and F_1 (1×) individuals. The following traits were used as phenotypes: average daily gain (ADG), backfat thickness (BFT), meat to fat ratio (MFR), and carcass length (CRCL). A full description of the crosses, as well as, the genomic and phenotypic data was published by Falker-Gieske *et al.*, 2019.

Structural variants and tandem repeats profiling. Using WGS data from the founders, we profiled SVs using three independent callers: smoove v0.2.6 (Brent, 2018), DELLY v0.7.7 (Raush *et al.*, 2012), and manta v1.6.0 (Chen *et al.*, 2016). Further, a high confidence set was obtained with SURVIVOR v1.0.7 (Jeffares *et al.*, 2021) followed by discarding SVs with a genotyping rate <0.8 and QUAL<6,000. Concerning the TRs, we screened all tandem repeats with a motif length of 2-20 bp with GangSTR v2.4.6 (Mousavi *et al.*, 2019)

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_390, © I. Blaj *et al.* 2022

which we then subjected to a call-level quality control filtering (min DP 10 and quality score >0.8). The final TRs dataset only contained variants with a call rate >0.8.

Imputation step and genome-wide association study. The joint set of curated SVs and TRs were then imputed to the F_2 individuals (n=2,657) based on the low coverage F_1 and F_2 SNP array data using the same imputation strategy we applied in Falker-Gieske *et al.*, 2019 for SNPs and small indels. Beagle 5.2 was employed for phasing, while Beagle 4.0 (Browning *et al.*, 2021) was utilized for imputation as it considered pedigree information. For single trait association analysis, we used GCTA v1.93.2 (Yang *et al.*, 2011) by incorporating the SVs and TRs as genotypes to be tested for association and ADG, BFT, MFR, and CRCL as phenotypes. A genomic relationship matrix was created from the F_2 SNP array data for the mixed linear model 'leave one chromosome out' analysis which also included a cross covariate (4 classes representing each of the 4 crosses). Following this model, imputed SVs and TRs with MAF>0.01 were tested. The significance level for the marker effects was set by the corresponding Bonferroni correction to a *P*-value of 0.05/number of tests.

Results

In the first part of this study, we profiled 13,201 high confidence SVs and 103,730 polymorphic TRs in the founders. These variants were considered for imputation. After MAF filtering on the imputed variants in the F_2 individuals, a final SVs and TRs dataset comprising 54,076 variants was tested for association. From the GWAS results, the following cumulated number of SVs and TRs exceeded the genome-wide significance threshold: 17, 156, 105, and 339 for ADG, BFT, MFR, and CRCL, respectively. Overall, among these, 88 were SVs whereas 529 were TRs. Compared to our previous SNPs and small indels based GWAS study (Falker-Gieske *et al.*, 2019), here we detected additional significant signals on *Sus Scrofa* chromosome 1 (SSC1) and SSC16 for ADG, on SSC16 for BFT, on SSC14 and SSC15 for MFR, and on SSC4, SSC8, SSC9, SSC14 and SSC16 for CRCL. Nevertheless, common GWAS peaks in both studies were localized over all traits on SSC1, SSC2, SSC4, SSC5, SSC7, SSC17, and SSC18. On these chromosomes, we proceeded by prioritizing SVs and TRs based on LD. Therefore, we discarded for each trait all significant SVs and TRs that were in high LD (r^2>0.8) with the previously identified significant SNPs and small indels. The number of non-tagged SVs and TRs were 12 (out of 17) for ADG, 112 (out of 156) for BFT, 89 (out of 105) for MFR, and 237 (out of 339) for CRCL. 388 genes incorporated or lied in the proximity of these prioritized SVs and TRs from which 75.51% were protein coding and 19.33% lncRNA (long non-coding RNA), while the remaining 5.16% were represented by miRNA (microRNA), processed pseudogene, pseudogene, snoRNA (small nucleolar RNA), and snRNA (small nuclear RNA). Table 1 presents the chromosome-wise top three genes (when present) per trait.

Table 1. Chromosome-wise top three associated genes (if present) for average daily gain (ADG), backfat thickness (BFT), meat to fat ratio (MFR) and carcass length (CRCL).[1]

Trait	Top three associated genes (Chromosome: genes)
ADG	1: *ZFAND5, TMC1*; 2: *AP2A2, ENSSSCG00000012835*; 4:*UBE2V2*; 7: *U6**, *ENSSSCG00000044971***, *HMGCLL1*; 16*: ENSSSCG00000050375, CDH10*.
BFT	1: *ENSSSCG00000048964***, *ENSSSCG00000043016***, *ENSSSCG00000051116***; 2: *SHANK2, STX3, TNNT3*; 4: *PREX2, CHD7*; 7: *COL21A1, DST, ENSSSCG00000001612*; 16: *ENSSSCG00000050819***, *ISL1*.
MFR	1: *PLPP7, PRRC2B, RALGPS1*; 2: *STX3, CBLIF, MRPL16*; 4: *ENSSSCG00000005127***, *LRP12, CHD7*; 5: *VDR, RPAP3*; 7: *COL21A1, ENSSSCG00000002296, VTI1B*; 14: *BTRC, XKR6, ENSSSCG00000045361***; 15: *IGFBP5, TNP1*; 18: *HERPUD2, ENSSSCG00000042613***, *AOAH*.
CRCL	4:*UBE2V2*; 7: *COL21A1, HMGCLL1, DAAM2*; 8: *NOCT, ENSSSCG00000049816***; 9: *PON2, ASB4, ENSSSCG00000049806***; 14: *ANKRD13A, SPPL3, WSCD2*; 16: *ENSSSCG00000050819***, *ISL1*; 17: *ENSSSCG00000044805***, *PLCB4, ENSSSCG00000043546*.

[1] * = snRNA; ** = lncRNA.

Many of the highly associated regions overlapped between the traits leading to common candidate genes. *COL21A1* was singled out due to an intronic TR ($(TTTG)_3/(TTTG)_5$, SSC7:29,488,854) causing the top GWAS peak on SSC7 for BFT, MFR, and CRCL. Further, for BFT, MFR, and CRCL, *DST* was associated, whilst for ADG, BFT, and MFR, the genes *AP2A2, ENSSSCG00000012835, SHANK2* were shared. Finally, *U6, ENSSSCG00000044971, HMGCLL1, DAAM2, PRIM2, ENSSSCG00000001612* influenced ADG, BFT and CRCL.

Discussion

Here, we investigated the role of structural variants and tandem repeats in a series of association studies using pooled F_2 pig crosses characterized for growth and carcass traits. The influence of SVs and TRs on quantitative traits and to which extent these variants were tagged by SNPs and small indels from a prior GWAS was addressed. Regarding the latter and assessed across all traits, 56.56% of the highly significant variants were not in high LD (r^2<0.8) with a previously associated variant, demonstrating that part of the molecular mechanisms behind complex traits could be driven by SVs and TRs. Nevertheless, in both GWAS studies many signals coincided leading to overlapping associated genes. An intronic TR variant was identified for *COL21A1* and was proposed as a straightforward candidate for functional validation. In addition, for ADG, BFT, and CRCL, *BMP5* was in the vicinity of the top genes *U6* (snRNA) and *ENSSSCG00000044971* (lncRNA). Similar lncRNA-related mechanisms were found for CRCL with *BMP2* and nearby lncRNAs (upstream: *ENSSSCG00000043546, ENSSSCG00000047884*; downstream: *ENSSSCG00000044805*). Li *et al.*, 2021 indicated that a SNP (rs320706814) was the main cause of the effect on carcass length, however, this variant was not significantly associated nor in LD with the significant SNPs, small indels, SVs or TRs we have identified. The associated genes found for BFT and CRCL on SSC16 (Table 1) indicated the existence of pleiotropic loci which we also detected in the previous GWAS but on SSC7. The peak on SSC16 was led by a lncRNA (*ENSSSCG00000050819*) together with *ISL1* and has not been reported for both traits in the previous GWAS as well as no reports in the region were found for CRCL in the AnimalQTLdb (Hu *et al.*, 2022). *ISL1* is known to regulate pancreatic development and insulin secretion (Yan *et al.*, 2016). The findings related to lncRNAs add to the growing body of evidence indicating that these genes could be involved in different mechanisms of gene regulation as they can control the expression of nearby genes by influencing their transcription (Statello *et al.*, 2021). For SSC2, *SHANK2* was part of the BFT and MFR GWAS peaks due to several highly significant TRs and to one deletion (SSC2:2,753,494, length=-556 bp). *SHANK2* has been associated with the autism spectrum disorder (Lu *et al.*, 2018) that was linked with feeding and eating disorders in humans (Carpita *et al.* 2020). Therefore, the differences in BFT and MFR phenotypes could be a consequence of pig behavioral changes related to feed intake. This study provides an efficient strategy for utilizing existing populations with whole-genome sequence, SNP array, and phenotypic data available to better understand mechanisms behind quantitative traits. Overall, this paper can be regarded as a valuable resource for future studies examining how structural variants and tandem repeats regulate gene expression and ultimately contribute to complex trait variation.

Acknowledgements

The study was funded by the German Research Foundation (Deutsche Forschungsgemeinschaft, DFG).

References

Brent P. Smoove (2018) Available at: https://brentp.github.io/post/smoove/

Browning B. L., Tian X., Zhou Y., and Browning S. R. (2021) Am. J. Hum. Genet. 108(10):1880-1890 https://doi.org/10.1016/j.ajhg.2021.08.005

Chen X., Schulz-Trieglaff O., Shaw R., Barnes B., Schlesinger F., *et al.* (2016) Bioinformatics 32(8):1220-1222 https://doi.org/10.1093/bioinformatics/btv710

Carpita B., Muti D., Cremone I. M., Fagiolini A., and Dell'Osso L. (2020) CNS Spectr. 9:1-9 https://doi.org/10.1017/S1092852920002011

Falker-Gieske C., Blaj I., Preuβ S., Bennewitz J., Thaller G., *et al.* (2019) G3: Genes, Genomes, Genet. 9(9):2823-2834 https://doi.org/10.1534/g3.119.400452

Hu Z.-L., Park C. A., and Reecy J. M. (2022) Nucleic Acids Res. 50(D1): D956-D961 https://doi.org/10.1093/nar/gkab1116

Jeffares D. C., Jolly C., Hoti M., Speed D., Shaw L., *et al.* (2017) Nat. Commun. 8(1):1-11 https://doi.org/10.1038/ncomms14061

Li J., Peng S., Zhong L., Zhou L., Yan G., *et al.* (2021) Genet. Sel. Evol. 53(1):1-13 https://doi.org/10.1186/s12711-021-00689-0

Lu Z. A., Mu W., Osborne, L. M., and Cordner Z. A. (2018) BMJ Case Rep. bcr:2018 http://dx.doi.org/10.1136/bcr-2018-225119

Mousavi N., Shleizer-Burko S., Yanicky R., and Gymrek M. (2019) Nucleic Acids Res. 47(15):e90-e90 https://doi.org/10.1093/nar/gkz501

Rausch T., Zichner T., Schlattl A., Stütz A. M., Benes V., *et al.* (2012) Bioinformatics 28(18):i333-i339 https://doi.org/10.1093/bioinformatics/bts378

Statello L., Guo C. J., Chen L. L., and Huarte M. (2021) Nat. Rev. Mol. Cell Biol. 22(2): 96-118 https://doi.org/10.1038/s41580-020-00315-9

Yan C., Yu C., Zhang D., Cui Y., Zhou J. *et al.* (2016) Sci. Rep. 6(1):1-11 https://doi.org/10.1038/srep39308

Yang J., Lee S. H., Goddard M. E., and Visscher P. M. (2011) Am. J. Hum. Genet. 88(1): 76-82 https://doi.org/10.1016/j.ajhg.2010.11.011

391. Detecting QTL for two lowly correlated traits using multi-trait and meta-analysis approaches in Swiss Large White pigs

A. Nosková[1*], A. Mehrotra[1], N.K. Kadri[1], S. Neuenschwander[2], A. Hofer[3] and H. Pausch[1]

[1]ETH Zürich, Universitätstrasse 2, 8092 Zürich, Switzerland; [2]ETH Zürich, Tannenstrasse 1, 8092 Zürich, Switzerland; [3]SUISAG, Allmend 8, 6204 Sempach, Switzerland; anoskova@ethz.ch

Abstract

We performed single-trait and multi-trait genome-wide association studies (GWAS) to detect quantitative trait loci (QTL) for body length and feed intake in 3,638 Swiss Large White pigs that had genotypes at 43,994 SNPs. Single-trait GWAS (ST-GWAS) revealed two non-overlapping QTL for each trait. Multivariate GWAS implemented with GEMMA software yielded 75 significant SNPs ($-\log_{10}(p)>6$) at the same four QTL. A meta-analysis of summary statistics from the two ST-GWAS resulted in 83 significant SNPs at the same QTL. After including all samples that had at least one phenotype recorded (in total 5,109), the *P*-values of top variants decreased suggesting higher statistical power due to a larger sample size. Fine mapping using imputed sequence variant genotypes confirmed the presence of three previously reported candidate causal variants in *MC4R*, *VRTN* and *BMP2* genes among the most significantly associated variants.

Introduction

Most of the traits considered in pig breeding programs have a complex genetic architecture. Genome-wide association testing between dense markers and phenotypes in reasonably large cohorts is a powerful approach to detect loci underpinning complex traits. Recent investigations suggested that markers associated with complex traits often exhibit pleiotropic effects (Bolormaa *et al.*, 2014). Despite an increasing number of GWAS being carried out in pigs, only a handful of causal variants have been identified.

In this study, we explored and compared single-trait and multi-trait GWAS methods to detect QTL for two lowly correlated traits in the Swiss Large White (SLW) pig breed. The detected QTL were characterized using sequence variant genotypes imputed from a cohort of 421 sequenced animals.

Materials & methods

Input data. Genotypes of 17,006 pigs from different breeds were obtained using 5 Illumina porcine SNP Bead chips with various densities. The physical positions of the SNPs were determined based on the *Sscrofa*11.1 assembly of the pig genome. Using plink (v. 1.9; Purcell *et al.*, 2007), we kept 43,994 unique autosomal SNPs that had minor allele frequency above 1%, less than 10% missing genotypes per variant, and did not deviate from Hardy-Weinberg proportions. We retained only samples that had genotypes for at least 80% of the markers. Sporadically missing genotypes were subsequently imputed and phased for 14,292 Large White pigs using Beagle (v. 5.0; Browning *et al.*, 2018). Using our in-house reference panel of 421 pigs with whole-genome sequence data (average depth of coverage 7.1×, ranging from 2.4× to 37.5×), we imputed the array-derived genotypes up to sequence level. In total 17,600,482 bi-allelic SNPs passing the aforementioned filtering criteria, and 3,638 pigs with phenotype records for the studied traits were used to fine-map the identified QTL.

Phenotypes were available in the form of de-regressed estimated breeding values (drEBV; Garrick *et al.*, 2009). The breeding values of 5,753 SLW pigs were routinely estimated using a BLUP multi-trait multi-breed animal model. Only drEBV with $r^2>0.3$ and in the range of mean ±5 standard deviations were considered, resulting in 3,638 records for body length (BL) and 5,109 for average daily feed intake (ADFI).

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_391, © A. Nosková *et al.* 2022

Genome-wide association studies (GWAS). We applied linear mixed model-based approaches implemented in the Genome-wide Efficient Mixed Model Association software (GEMMA; v. 0.98.5; Zhou and Stephens, 2014) to detect association between genotypes and phenotypes while considering either one (ST-GWAS) or two traits at a time in a multivariate GWAS (mtGWAS). A centered genomic relatedness matrix was fitted to control for the relationship among the individuals. A score test was applied to obtain the significance level of the association tests.

Results from mtGWAS were compared with those obtained from a meta-analysis (metaGWAS) of summary statistics from ST-GWAS following the approach proposed by Bolormaa *et al.* (2014). The t-values were calculated based on regression coefficients and corresponding standard errors from the ST-GWAS summary statistics. Multi-trait χ^2 statistic was subsequently calculated based on the SNP-based t-values and corresponding variance-covariance matrix. The metaGWAS was calculated on results obtained from individuals that had phenotypes for both traits (n=3,638) as well as for at least one of the traits (n=5,109). The GWAS results were visualized using the qqman package (Turner, 2018), with *suggestiveline* = -log10(1e-06).

Results

Single-trait GWAS. Linear mixed model-based ST-GWAS were performed between two traits (body length and average daily feed intake; BL and ADFI) and microarray-derived genotypes at 43,994 SNPs in 3,638 SLW pigs. The ST-GWAS revealed four major QTL (Figure 1A). The most significant SNPs (Table 1) for BL were MARC0038565 on SSC7 and INRA0052780 on SSC17. The top SNPs at two QTL for ADFI were both located on SSC1 (ASGA0004992 and ASGA0008077).

Low correlation of the de-regressed breeding values (r=-0.06 from 3,638 pigs) and distinct QTL revealed by ST-GWAS indicated that the traits were largely independent. Next, we investigated if multi-trait association methods might increase the power to detect QTL also for seemingly unrelated traits.

Multi-trait GWAS and meta-analysis GWAS. We applied multivariate association testing (mtGWAS) and meta-analysis of ST-GWAS (metaGWAS) to consider both traits (BL and ADFI) simultaneously. Both approaches revealed similar results (Table 2). Both methods discovered the four QTL that were previously shown by ST-GWAS, including the same top SNPs. The metaGWAS yielded 8 more markers above the suggestive significance level ($-\log_{10}(p)>6$) than mtGWAS.

Following our confirmation that both methods provide almost identical results, we took advantage of the fact that the metaGWAS approach does not rely on the availability of phenotypes for all individuals. We conducted the analysis for BL and ADFI in 5,103 pigs including 1,471 animals that had missing phenotypes for BL. The results (Figure 1B) mirrored previous findings and further increased the significance of top marker on SSC1 by on average 17% when compared to first metaGWAS. The significance level of the markers on SSC7 and SSC17 slightly decreased (-0.85%).

Table 1. The most significant markers at QTL detected for two traits.

SNP	rsID	Chr	Position	MAF	Candidate gene
ASGA0004992	rs80843328	1	160,210,902	0.300	*MC4R*
ASGA0008077	rs80964940	1	270,968,825	0.144	*FIBCD1*
MARC0038565	rs80894106	7	97,652,632	0.287	*VRTN*
INRA0052780	rs333311249	17	15,659,761	0.448	*BMP2*

Table 2. Top markers and their corresponding *P*-values in each method.

SNP	BL[1*]	ADFI[1*]	mtGWAS[1]	metaGWAS[1]	metaGWAS[2]
ASGA0004992	2.20E-02	2.06E-13	3.41E-13	2.76E-14	1.74E-15
ASGA0008077	4.00E-03	9.63E-19	5.56E-19	2.86E-20	1.08E-27
MARC0038565	5.05E-25	2.20E-01	4.60E-25	4.47E-27	6.76E-27
INRA0052780	8.90E-20	1.94E-02	8.37E-21	1.46E-22	2.41E-22

* Single-trait GWAS for body length (BL) and average daily feed intake (ADFI).

[1] Number of individuals = 3.638.

[2] Number of individuals = 5.109.

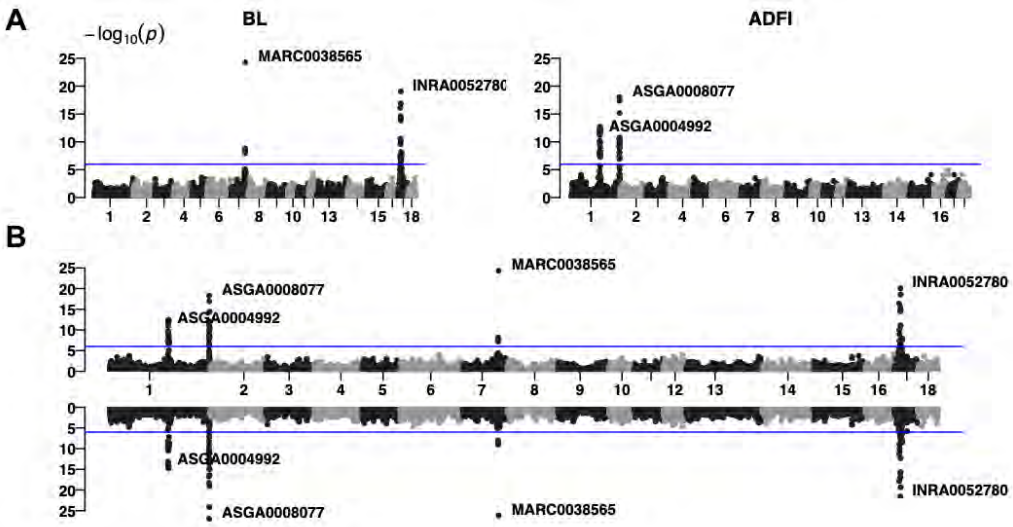


Figure 1. Manhattan plots of (A) ST-GWAS for body length (BL) and average daily feed intake (ADFI) and of (B) mtGWAS (top; n=3,638) and metaGWAS (bottom; n=5,109).

Fine mapping. We considered sequence variants imputed from 421 pigs into our mapping cohort to fine-map the QTL on SSC1 (160 Mbp), SSC7 (98 Mbp) and SSC17 (16 Mbp). All three top sequence-derived SNPs were in high LD with the top array-derived markers as well as with previously described candidate causal variants (rs81219178, rs709317845 and rs320706814 in or nearby *MC4R*, *VRTN* and *BMP2*, respectively). The QTL on SSC17 was strongly associated with BL (*P*-value 5.17E-61) and weakly with ADFI (*P*-value 2.7E-07), unveiling a hitherto undetected potential pleiotropic effect.

Discussion

We compared mtGWAS and metaGWAS of two almost uncorrelated traits and fine-mapped the QTL using imputed sequence variants in 3,638 SLW pigs. Similar to Fang and Pausch (2019), the *P*-values of the top SNPs were lower in multi-trait methods than in ST-GWAS (Table 2). The four QTL described here have been detected in several pig populations and *BMP2* (Li *et al.*, 2021), *VRTN* (Duan *et al.*, 2018), and *MC4R* (Bruun *et al.*, 2006) have been proposed as candidate genes. The QTL on SSC1 (271 Mbp) was found to be associated with backfat thickness (Gozalo-Marcilla *et al.*, 2021). It is close to *FIBCD1*, which is associated with body mass index in human (Pulit *et al.*, 2019). Our sequence-based fine-mapping approach

revealed three previously detected candidate causal variants near *VRTN*, *MC4R* and *BMP2* among the most significantly associated variants. However, the previously reported candidate causal variants were not the top variants. This may indicate sampling bias (Barendse, 2011) or superior association power due to strong LD with adjacent causal variants (Bickel *et al.*, 2011), or that they are not causal but in LD with the true causal variant. Another likely reason could be low accuracy of imputation due to the low marker density in the array panels. Low accuracy of imputation from a marker panel with similar size was reported previously (van den Berg *et al.*, 2019).

Multivariate association testing as implemented with GEMMA requires individual-level genotype and phenotype data and considers only animals with non-missing phenotypes across the traits. To consider all animals, missing phenotypes need to be imputed upfront. In contrast, the meta-analysis of summary statistics that are measured on a various number of individuals in contributing studies is somewhat agnostic to an unbalanced number of animals per trait. We observed an increase in statistical significance when additional individuals (n=1,471) were considered that had phenotypes only for one trait. However, this increase in statistical power is likely higher for correlated traits and pleiotropic variants.

References

Barendse W. (2011) BMC Genom. 12(1):1–12. https://doi.org/10.1186/1471-2164-12-232

Van den Berg S., Vandenplas J., van Eeuwijk F.A., Bouwman A.C., Lopes M.S., *et al.* (2019) Genet. Sel. Evol. 51(2):1-13. https://doi.org/10.1186/s12711-019-0445-y

Bickel R. D., Kopp A., and Nuzhdin S. (2011) PLoS Genetics 7(1):e1001275. https://doi.org/10.1371/JOURNAL.PGEN.1001275

Bolormaa S., Pryce J. E., Reverter A., Zhang Y., Barendse W., *et al.* (2014) PLoS Genetics 10(3):e1004198. https://doi.org/10.1371/journal.pgen.1004198

Browning B. L., Zhou Y., and Browning S. R. (2018) Am. J. Hum. Genet. 103(3):338–348. https://doi.org/10.1016/j.ajhg.2018.07.015

Bruun C. S., Jørgensen C. B., Nielsen V. H., Andersson L., and Fredholm M. (2006) Anim. Genet. 37(4):359–362. https://doi.org/10.1111/J.1365-2052.2006.01488.X

Duan Y., Zhang H., Zhang Z., Gao J., Yang J., *et al.* (2018) Int. J. Biol. Sci. 14(6):667–681. https://doi.org/10.7150/ijbs.23815

Fang Z. H., and Pausch H. (2019) BMC Genom. 20(1):1–15. https://doi.org/10.1186/s12864-019-6066-6

Garrick D. J., Taylor J. F., and Fernando R. L. (2009) Genet. Sel. Evol. 41(1):55. https://doi.org/10.1186/1297-9686-41-55

Gozalo-Marcilla M., Buntjer J., Johnsson M., Batista L., Diez F., *et al.* (2021) Genet. Sel. Evol. 53(1):76. https://doi.org/10.1186/s12711-021-00671-w

Li J., Peng S., Zhong L., Zhou L, Yan G., *et al.* (2021) Genet. Sel. Evol. 53(94):1-13. https://doi.org/10.1186/s12711-021-00689-0

Pulit S.L., Stoneman Chr., Morris A.P., Wood A.R., Glastonbury C.A., *et al.* (2019) Hum. Mol. Genet. 28(1):166-174. https://dx.doi.org/10.1093%2Fhmg%2Fddy327

Purcell S., Neale B., Todd-Brown K., Thomas L., Ferreira M.A.R., *et al.* (2007) Am. J. Hum. Genet. 81(3):559–575. https://doi.org/10.1086/519795

Turner S. (2018) J. Open Source Softw. 3(25):731. https://doi.org/10.1101/005165

Zhou X., and Stephens M. (2014) Nat. Methods 11(4):407–409. https://doi.org/10.1038/nmeth.2848

392. A Forward algorithm to identify credible sets in Bayesian variable selection models

G. de los Campos[1,2,3*], A. Samaddar[2], B. Valente[4], C.-Y. Chen[4], W. Herring[4] and A. Grueneberg[1]

[1]Department of Epidemiology & Biostatistics, Michigan State University, USA; [2]Department of Statistics & Probability, Michigan State University, USA; [3]Institute for Quantitative Health Science and Engineering, Michigan State University, USA; [4]The Pig Improvement Company, Genus plc, Hendersonville, TN, USA; gustavoc@msu.edu

Abstract

In regression models, high collinearity can severely reduce the power to detect associations between individual SNPs and a phenotype. We present a simple algorithm that uses the samples from the posterior distribution of a Bayesian Variable Selection model to identify sets of SNPs (i.e. Credible Sets, CS) that (jointly) are confidently associated with a trait. The proposed method leads to multi-resolution inferences identifying, depending on the degree of collinearity and the LD between the SNPs and causal variants, CSs of variable size. Using simulations, we show that the proposed approach increases power with adequate FDR-control. The method described in this study was implemented in the R language and is available with the BGLR R-package.

Introduction

Modern data sets often include a large number of predictors (e.g. SNPs, gene expression, methylation, milk spectrometry) that cluster in complex ways. For instance, in genomic data, linkage disequilibrium (LD) coupled with high marker density can lead to very high collinearity between SNPs. In genome-wide association studies (GWAS) mapping is done using either marginal association test-statistics or multiple regression models. In multiple linear regression models, high collinearity can reduce the power to detect individual variants associated with QTL. For instance, when Bayesian Variable Selection methods are applied to SNPs in high LD often no individual SNP reaches the posterior probability thresholds that are required to achieve a low false discovery rate (FDR, e.g. $\pi_j = \text{p}(\beta_j \neq 0 \mid data) > 0.95$ which corresponds to a local FDR of 0.05); thus, reducing power or requiring to tolerate a high error rate to achieve high power. At the same time, the joint probability of association of the SNPs in segments harboring causal loci (i.e. for a SNP set) $\text{p}(\beta_1 \neq 0 \text{ or } \beta_2 \neq 0 \text{ or } \beta_s \neq 0 \mid data)$ can be very high, even close to 1. In such situations, we may have strong evidence of an association for a set of SNPs in a chromosome segment and high uncertainty about individual SNP effects.

To address the problem above-described, studies often use window methods whereas statistics (variances, joint probability of association) are produced for chromosome segments instead of individual variants. However, there is no consensus as to how wide these windows should be, whether a fixed or variable window size is needed, and whether non-overlapping or sliding windows are best (Beissinger *et al.*, 2015). In this study, we propose an alternative approach that uses the samples from the joint posterior distribution to identify sets of SNPs with high joint posterior probability of inclusion. When the number of SNPs in the model is large, exploring all possible credible sets (CS) is not feasible. Thus, we designed a forward search algorithm to identify sets of SNPs jointly associated with the outcome. Here, we described the proposed algorithm and evaluate its performance using extensive simulations applied to two large data sets with distinct characteristics.

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_392, © G. de los Campos *et al.* 2022

Materials & methods

The joint probability of association of two SNPs with a phenotype is:

$$p(\beta_j \neq 0 \text{ or } \beta_{j'} \neq 0) = p(\beta_j \neq 0 + p\beta_{j'} \neq 0 - p(\beta_j \neq 0 \text{ and } \beta_{j'} \neq 0$$

Above, β_j and $\beta_{j'}$ are SNP effects in a multiple linear regression model. The above expression can be extended to more than two SNPs. However, evaluating the joint probability of inclusion of more than two SNPs requires evaluating multiple joint probabilities of inclusion of 2, 3, or more SNPs. To avoid performing an exhaustive search for CSs, we propose to use a forward algorithm that builds CSs by adding to an existing CS the SNP that maximizes the joint probability of association of the CS with the phenotype. To simplify computations, at each step of the forward search we maintain a boolean vector $Z = \{Z_s\}$ where $Z_s = \{1$ if at least one element of the CS was active; 0 otherwise$\}$. Using this, at each step of the forward search we decide which SNP to add to the CS by maximizing

$$p(Z_s = 1 \text{ or } \beta_j \neq 0) = p(Z_s = 1) + p(\beta_j \neq 0) - p(Z_s = 1 \text{ and } \beta_{j'} \neq 0) \quad (1)$$

We use the above expression to decide, at each step in the forward search which SNP to add to the CS. Schematically, the algorithm proceeds as follows:

- *Step 1*: Select the SNP with the highest posterior probability of inclusion.
- *Forward search*: Using Equation 1, select the SNP that, when added the current CS, yields the maximum joint probability of inclusion ([1]) of the CS. Add the selected SNP to the CS. Repeat this process until one of the following stopping rules are met.
- *Stopping rules*: At each step stop the forward algorithm: (1) if the size of the CS is larger than a pre-specified maximum size; or (2) if the SNP to be added produces an increase in the probability of association of the CS by less than a pre-specified threshold (e.g. 0.03); or (3) if the joint probability of association of the CS is greater than a pre-specified value (e.g. 0.98).

We implemented the algorithm described above in the R-function findCS(); the function is included in the development version of the BGLR R-package (https://github.com/gdlc/BGLR-R).

Data sets. To evaluate the performance of the proposed algorithm we conducted a simulation study with two data sets (Table 1). The human data set uses ultra-high-density (imputed) SNP genotypes from distantly related individuals from the UK Biobank. The second data set originates from a pig breeding population from the PIC® and includes high-density SNP genotypes from a breeding line. The two data sets have contrasting features and are representative of the challenges that are faced in GWAS applied to data from distantly related individuals and from breeding populations.

Simulation. In both cases, in each MC replicate we simulated one 30 Kbp chromosome segment with three causal variants separated by 10 Kbp. In the human data set, each causal variant explained either 1/12th- or 1/3rd-of-1% of the phenotypic variance. In the pig data set the sample size available does not offer high power to detect very-small effect variants, thus, each causal variant explained either 1% or 1/3rd-of-1% of the phenotypic variance.

Table 1. Data sets used in the simulation study.[1]

Data set	Sample size	# of SNPs	Sample type	LD
Human	100,000*	13.1 million (~4,366 SNPs/Mbp)	Distantly related**	Short span
Breeding (swine)	25,000	43,837 (~15 SNPs/Mbp)	Family data	Long span

[1] Mbp: mega-base-pair; *: we run analyses with N up to 300,000; due to space-restriction we present results up to n=100,0000 only; **: genomic relationship <0.05.

Analysis. We fitted a Bayesian model with a point of mass at zero and a Gaussian slab prior (aka BayesC in animal breeding, Habier *et al.*, 2011) to a chromosome segment consisting of all SNPs within ±500 Kbp (in the human data set, ~4,400 SNPs) or ±5 Mbp (in the pig data set, ~160 SNPs) of the 30 Kbp segment holding the 3 causal variants. We used much shorter segments in the human data set because LD decays very fast in this data set. We fitted models using the BGLR R-package (Perez and de los Campos, 2014); for each simulation scenario, we conducted 300 MC replicates. For each MC replicate we run five chains; then, using the combined samples, we obtained discovery sets by identifying either individual SNPs or CS with a posterior probability of inclusion greater than 0.95, 0.90, 0.80, and 0.50. To make a fair comparison between individual-SNP and CS inference, we considered a SNP in a discovery set to be a true (false) discovery if its distance to the closest causal variant was smaller or equal (larger) than xKbp (x=10, 30, 50, 100, and 500 in the human data set and x=500, 1000, 2,000, and 3,000 in the pig data set). Using this, we computed the false discovery proportion for each MC replicates and mapping resolution. Likewise, we considered that a causal locus was mapped if the distance of it to the closest SNP in the discovery set was smaller than xKbp – we used this to estimate power.

Results

Figure 1 shows the power/FDR curves obtained by data set, method (either single-SNP or CS), sample size, and heritability. Due to space restriction, we only present results for a mapping resolution of 50 Kbp in the human data set and 2 Mbp in the pig breeding data set. In all cases, the CS method achieved higher power than inferences based on individual SNPs. However, a stringent probability cutoff (e.g. >0.95) was needed to achieve effective FDR control. In the human data set, for h^2=0.33% and a mapping resolution of 50 Kbp, the CS method achieved high power with an FDR near 0.1 for N≥30,000. For lower heritability (0.08%) a sample size of ≥100,000 was needed to achieve high power. On the other hand, in the pig breeding data set the mapping resolution that was needed to obtain a good power/FDR performance was much coarser (2 Mbp); this is a direct consequence of lower marker density and the longer span of LD present in this breeding population. In this data set, even with a 2 Mbp mapping resolution, the FDR was high for the low heritability scenario (h^2=0.33%) and sample size <10,000.

Discussion

High LD between SNPs can reduce the power to detect the association of individual variants with a phenotype because collinearity can lead to uncertainty about individual SNP effects. Therefore, when multiple variants are in LD with the same causal variant often multiple variants show elevated inclusion probability with no individual variant reaching the inclusion probability cutoffs that are used in GWAS to achieve adequate FDR control. In these cases, we may have strong evidence supporting the presence of signal variants in a region, without enough evidence to perform individual variant mapping. The credible set algorithm addresses the problem just described. The simulation study showed that the use of CS can lead to a remarkable increase in power relative to individual SNP inferences. In most scenarios, the CS method showed a flat power-FDR curve (Figure 1) which suggests that one should use this method with a stringent probability inclusion cutoff (e.g. 0.95 or 0.98) – lowering the probability cutoff can increase the FDR without much improvement in power. This way, the approach leads to high power with a reasonably low error rate. The size of the CS produced by the proposed algorithm varies, when individual variants reach a high posterior probability of inclusion some CS may contain only one variant; in cases where multiple variants show elevated posterior probability of inclusion with no individual variant reaching the probability cutoff, the resulting CS include multiple variants.

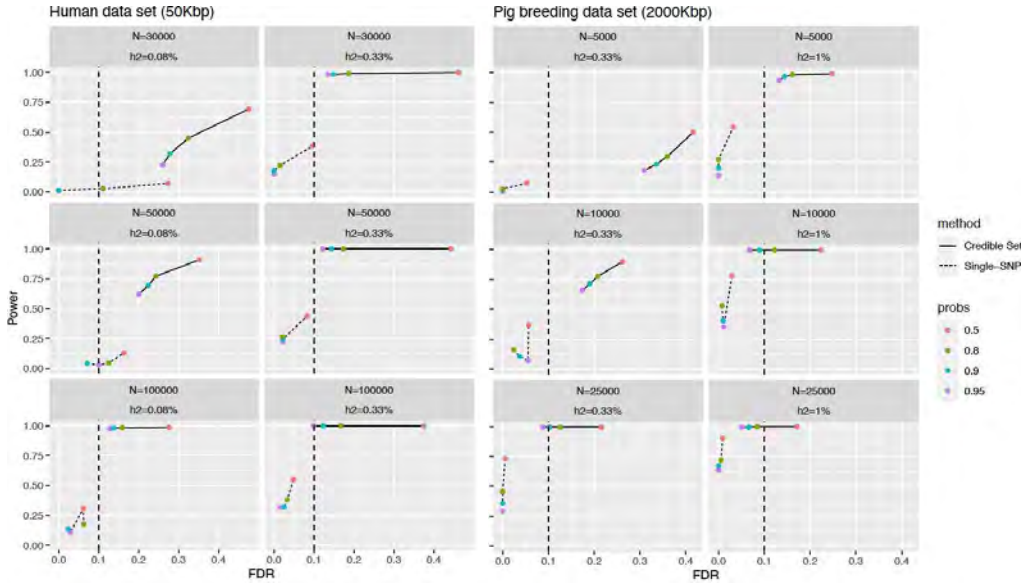


Figure 1. Power-FDR performance by data set, sample size, heritability, and posterior probability threshold used to define credible sets.

References

Beissinger T.M., Rosa G.J.M, Kaeppler, S.M., Gianola D., and de Leon N. (2015) Genetics Selection Evolution 47 (30). https://doi.org/10.1186/s12711-015-0105-9

Habier D., Fernando R.L., Kizilkaya K., and Garrick D.J. (2011) BMC Bioinformatics 12 (186). https://doi.org/10.1186/1471-2105-12-186

Pérez-Rodriguez P. and de los Campos G. (2014) Genetics 198(2):483-95. https://doi.org/10.1534/genetics.114.164442

Part 21

Statistical genetics: machine learning, deep learning and artificial intelligence

393. Multivariate application of artificial neural networks for genomic prediction

S.O. Peters[1,5*], K. Kızılkaya[2], M. Sinecen[3] and M.G. Thomas[4]

[1]Department of Animal Science, Berry College, Mount Berry, GA 30149 USA; [2]Department of Animal Science, Faculty of Agriculture, Aydin Adnan Menderes University, Aydin, 09100, Turkey; [3]Department of Computer Engineering, Faculty of Engineering, Aydin Adnan Menderes University, Aydin, 09100, Turkey; [4]Department of Animal Science, Colorado State University, Fort Collins, CO, USA; [5]Department of Animal and Dairy Science, University of Georgia, Athens, GA, USA; speters@berry.edu

Abstract

Phenotypic data for growth traits including birth, weaning and yearling weights and genomic data of 50K SNP markers were collected from 738 Brangus. Multivariate artificial neural networks (ANN) with 1 to 10 neurons models including the inputs from genomic relationship matrix were created and applied with the learning algorithms (Levenberg-Marquardt, Bayesian Regularization and Scaled Conjugate Gradient) and activation functions (tangent sigmoid and linear) for the analysis of growth data. Pearson correlation coefficients were used to evaluate multivariate ANN model performances about genomic prediction in datasets. In multivariate ANN models, the prediction performance of the combination of learning algorithms and activation functions changed for growth traits and there were no superior multivariate ANN models with learning algorithms and activation functions. The application of different activation functions did not make any significant difference on the genomic prediction performance of ANN models with different learning algorithms.

Introduction

The development and use of artificial neural networks (ANN) for artificial intelligence was inspired from the operations of nerve cells (known neurons) in the human brain. The ANN is a statistical modeling of the human brain functions and represents a new generation of information processing systems (Pereira *et al.*, 2009). The learning ability of ANN without using any prior assumption and sophisticated statistical models for linear and nonlinear relationships in the information processing systems is a very attractive feature. Therefore, ANN statistical models have been used to predict genomic values of individuals for complex traits for last decade because complex quantitative traits such as growth and production traits in animal and plant breeding are controlled by a network of numerous genes (Howard *et al.* 2014). This study aimed to compare multivariate ANN models with the learning algorithms (Levenberg-Marquardt (LM), Bayesian Regularization (BR) and Scaled Conjugate Gradient (SCG)) and activation functions (tangent sigmoid and linear) using 1 to 10 neurons models based on the input from G genomic relationship matrix.

Materials & methods

Phenotypes. Phenotypes including birth weight (BW), weaning weight (WW) and yearling weight (YW) as growth traits were from 738 Brangus heifers (3/8 Brahman × 5/8 Angus) from Camp Cooley Ranch in east central Texas, and the Chihuahuan Desert Rangeland Research Center and Campus Farm of New Mexico State University described by Luna-Nevarez *et al.* (2010), and Peters *et al.* (2012). Contemporary groups based on year of birth (2005 to 2007) and calving season (spring or autumn) and dam age were also extracted from database of International Brangus Breeders Association.

SNP marker genotypes and G matrix. SNP marker (53,692 SNP) genotypes were obtained from the BovineSNP50 Infinium BeadChips (Illumina, San Diego, CA). The quality control of SNP markers (Granato *et al.*, 2018) removed the animals having higher than 50% missing data, SNP markers having higher than

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_393, © S.O. Peters *et al.* 2022

5% missing data or lower than 95% call rate and lower than 10% minor allele frequency and resulted in 738 animals with 35,351 SNP markers. The genomic relationship (G) matrix was calculated based on Habier *et al.* (2007) and vanRaden (2008).

Artificial neural networks. The Multi-Layer Perceptron Artificial Neural Network (MLPANN) includes many interconnected neurons which are grouped into an input layer, an intermediate or hidden layer and an output layer. Genomic relationships (g_{ij}) between animals are independent variables and the input neurons in the input layer of the MLPANN. Also, the growth traits are dependent variables and the output neurons in the output layer of the MLPANN. Training process in the MLPANN focus on increasing the estimation accuracy between predicted values ($\hat{y}$) and the phenotypic values (y) using the number of neurons between 1 and 10, learning algorithms (LevenbergMarquardt (LM), Bayesian Regularization (BR) and Scaled Conjugate Gradient (SCG)) and tangent sigmoid and linear activation functions in the hidden layer and producing the multivariate outputs in the output layer.

Cross-Validation and application of MLPANN models. The genotyped animals were partitioned into 10 cross-validation sets by using k-means clustering. The accuracy of phenotypic values of animals was quantified in training and validation sets using the Pearson's correlation coefficient ($r_{y_k,\hat{y}_k}$) between the observed (y_k) and predicted phenotypic ($\hat{y}_k$) values.

MATLAB Neural Network Toolbox (2000) was used to fit the MLPANN models. The learning algorithms were applied by utilizing the functions of trainlm(.), trainbr(.) and trainscg(.) in MATLAB Neural Network Toolbox. The tangent sigmoid and linear activation functions for each learning algorithm were applied by using in MATLAB Neural Network Toolbox, respectively.

Results

Genomic prediction for three growth traits were studied by using multivariate ANN models with 1 to 10 neurons, BR, LM and SCG learning algorithms and Tangent Sigmoid and Linear activation functions. The average Pearson correlation coefficients (accuracies) obtained from 10-fold cross validation datasets were given in Table 1 for training datasets and in Table 2 for validation datasets for growth traits. Table 1 showed that the predictive accuracy results changed from -0.06 to 0.85. The best accuracies for BW, WW and YW were obtained from BR learning algorithm (0.00 to 0.85) and Linear activation function (0.22 to 0.84). However, SCG learning algorithm produced the worst accuracies results (-0.06 to 0.05) regardless of the trait and activation function and there were no big differences among accuracy results from Tangent Sigmoid and Linear activation functions. Table 2 indicated that accuracy results from cross-validation dropped remarkably and changed from -0.09 to 0.16. The best accuracy results (0.06 to 0.16) were obtained from the combination of BR learning algorithm and linear activation function for BW trait. However, learning algorithms and activation functions produced very low and similar accuracy results for BW, WW and YW traits.

Discussion

As seen in Table 1 and 2, the average accuracy results showed that the predictive performance of multivariate ANN models depended tightly on the number of neurons used (architecture of network) in the hidden layer of the network, the type of learning algorithms and activation functions and genetic architecture of growth traits.

In multivariate ANN models, as the number of neurons and the number of output increases, the number of parameters to be estimated also increases. Therefore, genomic relationship among animals could not provide enough information through input neurons to obtain high accuracy from multivariate ANN

models and the predictive performance results from cross-validation datasets were not consistent with increasing number of neurons within each growth trait for different learning algorithm and activation functions.

Table 1. Predictive abilities of multivariate application of MLPANN models for growth traits in training procedure.

	BW			WW			YW		
Neuron	**BR**	**LM**	**SCG**	**BR**	**LM**	**SCG**	**BR**	**LM**	**SCG**
Tangent sigmoid									
1	0.22	-0.02	-0.02	0.85	-0.03	-0.02	0.63	-0.03	0.04
2	0.20	0.03	0.05	0.70	0.12	0.03	0.56	0.00	0.02
3	0.35	0.09	0.04	0.36	0.22	0.03	0.20	0.13	-0.03
4	0.21	0.12	-0.03	0.41	0.19	-0.06	0.31	0.09	-0.06
5	0.16	0.20	0.01	0.45	0.09	0.04	0.30	0.07	0.05
6	0.05	0.24	0.02	0.12	0.20	0.05	0.07	0.19	0.02
7	0.04	0.30	0.00	0.09	0.10	-0.05	0.00	0.06	-0.06
8	0.10	0.28	-0.03	0.26	-0.03	-0.01	0.18	-0.02	-0.01
9	0.06	0.30	0.02	0.16	0.11	0.01	0.10	0.12	-0.02
10	0.06	0.27	0.01	0.12	0.14	-0.01	0.11	0.10	-0.02
Mean	0.15	0.18	0.01	0.35	0.11	0.00	0.25	0.07	-0.01
Std	0.10	0.12	0.03	0.26	0.09	0.04	0.21	0.07	0.04
Linear									
1	0.23	0.04	-0.04	0.84	0.11	-0.01	0.63	0.08	0.01
2	0.22	0.02	-0.02	0.82	0.10	0.00	0.69	0.05	0.00
3	0.43	0.17	0.01	0.36	0.14	0.00	0.24	0.20	0.00
4	0.43	0.24	0.00	0.40	0.15	0.00	0.29	0.08	0.00
5	0.47	0.17	0.01	0.37	0.16	0.00	0.27	0.10	-0.02
6	0.45	0.21	0.02	0.36	0.22	0.00	0.25	0.17	0.01
7	0.48	0.23	-0.01	0.36	0.12	-0.02	0.25	0.10	-0.02
8	0.43	0.21	-0.02	0.37	0.13	0.03	0.26	0.11	0.01
9	0.44	0.24	-0.04	0.38	0.13	-0.02	0.26	0.12	0.01
10	0.44	0.33	0.03	0.40	0.17	0.00	0.28	0.13	-0.01
Mean	0.40	0.19	-0.01	0.47	0.14	0.00	0.34	0.11	0.00
Std	0.10	0.09	0.02	0.19	0.03	0.01	0.17	0.04	0.01

Table 2. Predictive abilities of multivariate application of MLPANN models for growth traits in cross-validation procedure.

	BW			WW			YW		
Neuron	**BR**	**LM**	**SCG**	**BR**	**LM**	**SCG**	**BR**	**LM**	**SCG**
Tangent sigmoid									
1	0.06	0.06	0.04	0.04	-0.08	-0.01	0.05	0.01	-0.03
2	0.05	0.11	0.03	-0.01	-0.03	0.03	0.02	-0.02	0.10
3	0.15	-0.01	0.01	0.03	0.01	0.04	0.05	0.03	0.06
4	0.03	0.06	0.06	-0.01	0.00	-0.09	0.02	-0.03	-0.06
5	0.11	0.02	-0.04	0.04	0.01	-0.09	0.03	0.00	-0.07
6	-0.05	0.06	0.02	0.03	0.00	-0.03	-0.04	0.01	-0.02
7	-0.04	0.06	-0.04	-0.04	0.06	-0.06	0.03	0.04	-0.06
8	0.01	0.07	-0.08	-0.03	0.01	-0.08	0.02	0.05	-0.05
9	0.02	0.02	0.00	0.05	-0.01	-0.03	0.08	0.00	-0.01
10	0.01	0.08	-0.03	0.01	0.01	0.02	0.00	0.01	0.02
Mean	0.03	0.05	0.00	0.01	0.00	-0.03	0.03	0.01	-0.01
Std	0.06	0.04	0.04	0.03	0.04	0.05	0.03	0.03	0.06
Linear									
1	0.06	-0.02	0.01	0.01	0.08	0.00	0.04	0.04	0.01
2	0.06	-0.01	0.00	0.00	-0.06	0.05	0.04	-0.05	0.04
3	0.16	0.09	-0.03	0.06	0.01	0.02	0.04	0.07	0.02
4	0.10	-0.01	-0.04	0.02	0.01	-0.06	0.00	0.04	-0.01
5	0.13	0.05	-0.03	0.05	0.03	-0.02	-0.01	0.04	-0.05
6	0.13	0.04	0.05	0.06	0.01	-0.03	0.03	0.03	0.00
7	0.14	0.07	-0.01	0.03	0.04	0.03	0.02	0.08	-0.03
8	0.13	0.05	-0.02	0.02	0.02	0.06	0.02	0.06	0.03
9	0.13	-0.01	0.01	0.03	0.03	0.01	0.01	0.03	0.01
10	0.15	0.05	0.03	0.01	0.02	-0.02	-0.02	0.13	-0.02
Mean	0.12	0.03	0.00	0.03	0.02	0.00	0.02	0.05	0.00
Std	0.03	0.04	0.03	0.02	0.04	0.04	0.02	0.05	0.03

References

Demuth, H. and Beale M. (2000) Neural Network Toolbox User's Guide, The Mathworks, Natick.

Habier, D., Fernando, R.L. and Dekkers, J.C.M. (2007) Genetics 177(4): 2389–2397.

Howard, R., Carriquiry, A.L. and Beavis, W.D. (2014) G3 Genes|Genomes|Genetics 4(6): 1027–1046. https://doi.org/10.1534/g3.114.010298

Luna-Nevarez, P.D., Bailey, W., Bailey, C.C., VanLeeuwen, D.M., Enns, R.M., *et al* (2010) J. Anim. Sci. 88:1891–1904.

Pereira B. de B., Rao C.R., and Rao M. (2009) Data Mining using Neural Networks: A Guide for Statisticians, 1st ed. Chapman and Hall/CRC.

Peters, S.O., Kizilkaya, K., Garrick, D.J. Fernando, R.L. Reecy, J.M. *et al.* (2012) J. Anim. Sci. 90: 3398–3409.

vanRaden, P.M. (2008) J. Dairy Sci. 91(11): 4414–4423.

394. Comparison of GBLUP and machine learning breeding values for binary disease traits in working dogs

J.A. Thorsrud[1], K.M. Evans[2], D.M. Holle[2], H.J. Huson[1*]

[1]Department of Animal Sciences, Cornell University College of Agriculture and Life Sciences, 201 Morrison Hall, 507 Tower Road, Ithaca, NY 14853, USA; [2]The Seeing Eye Inc, 1 Seeing Eye Wy, Morristown, NJ, 07960, USA; hjh3@cornell.edu

Abstract

Guide dogs are a class of highly trained working dogs used by individuals with vision impairment to increase independence and quality of life. Through structured breeding programs, genetic improvements can advance their health and performance. Breeding values are a valuable tool that allow for heritable traits of interest to be selected for or against in animal colonies. Three binary health traits from a cohort of 1,097 guide dogs were utilized to construct breeding values to predict phenotypes using genome best linear unbiased predictor (GBLUP) and four machine learning (ML) techniques; support vector machine (SVM), random forest (RF), gradient boosting (GB), and neural network (NN). Across the traits studied, there was not a significant difference between GBLUP and ML performances. These results provide an experimental comparison indicating that ML approaches can construct breeding values at a similar level to the more traditional GBLUP.

Introduction

A variety of breeds are suitable for work as guide dogs which are defined by their ability to complete rigorous training to be selected for working (Fishman 2003). Health conditions can make it more difficult for a dog to pass their training and even preclude them from usage or breeding. The present study investigated data for class II mandibular distocclusion, distichiasis, and acral lick dermatitis traits within a cohort of working dogs. Mandibular distocclusion class II is a case of the mandible being too short and results in a misalignment between the lower and upper teeth. Dogs with the condition can experience difficulty eating, and pain. Treatments can include orthodontics and extractions (Hoyer and Rawlinson 2019). Distichiasis occurs when an extra layer of cilia is present on the margin of the eyelid. Eye irritation and infections can result, with surgical interventions available for severe conditions (Palella Gómez *et al.* 2020). Dogs with acral lick dermatitis often exhibit excessive licking of the extremities due to neuronal, behavioral, or external triggers. Persistent bacterial infections can be common and require medical and behavioral interventions (Paterson *et al.*, 2007).

All of the traits identified present challenges for completing any intensive training regimen, and breeders for guide dogs have incentives to remove genetic causes of these traits from their populations. A method to identify the genetic merit of an individual is breeding values which can be constructed with a variety of techniques. GBLUP was built on the improvement of genetic data over pedigree data used in the BLUP method. A relationship matrix is constructed, which employs the observed genetic similarity to determine the merit of individuals and can be used to predict phenotypes based on the genotypes of animals outside of the training dataset (Clark and van der Werf 2013).

Many different ML techniques have been utilized to construct breeding values in various species of both animals and plants. SVM segregates data into groups using kernels to enlarge the feature space (Suthaharan 2016). Studies have successfully used SVM for breeding values in species from swine to maize (Zhao *et al.* 2020). RF works by constructing individual decision trees, which operate as an ensemble to provide a single output prediction expected to outperform any individual tree (Qi 2012). It is a highly versatile ML approach that has been used extensively for genetic purposes (Naderi *et al.* 2016). GB is similar but continually constructs

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_394, © J.A. Thorsrud *et al.* 2022

trees to outperform and replace the previous model in the iteration, which makes it useful for bioinformatics (Srivastava *et al.*, 2021). Neural networks aim to artificially replicate how biological neurons work for pattern recognition in high-dimensional data. This approach is flexible but can be computationally costly (Nicholls *et al.*, 2020). In genetics, NN have been used for phenotypic prediction, genome wide association studies, and breeding value construction with success (Liu *et al.*, 2019). The current project aims to identify the effectiveness of ML methods compared to GBLUP in binary classification breeding value creation.

Materials & methods

Genotypes and phenotypes. A total of 1,097 dogs were genotyped from a colony of guide dogs. This included 322 German Shepherds, 578 Labrador Retrievers, 127 Golden Retrievers, and 70 Golden Retriever Labrador Retriever crosses. The animals were genotyped on the Illumina Canine HD beadchip with 220k single nucleotide polymorphisms (SNPs). After the SNP call rate of >0.9, number of alleles <2, and minor allele frequency of >0.03 thresholds were set, 158,895 SNPs remained. A sample call rate of >0.9 cutoff removed 76 dogs. Genotypes were coded as 0, 1, 2 for all models based on the minor alleles. The percentage of dogs with class II mandibular distocclusion was 5.8%, for distichiasis it was 4.9%, and acral lick dermatitis had a percentage of 4.8%.

Statistical methods. Five-fold cross-validation was used to evaluate the models. The dataset was broken into five segments and each segment had a run as the test dataset while the other four were used for training. The models were scored based on the area under the receiving operating characteristic curve (AUROC) to adjust for the high probability of prediction accuracy alone being greater than 90% for all models due to the small percentage of cases. An analysis of variance (ANOVA) test compared the performance of the models across their cross-validations to determine the significance of any differences in performance.

Model creation. For the GBLUP, the efficient mixed-model association algorithm was used to identify the restricted maximum likelihood. The genomic-heritability was calculated for the three traits as well. The GBLUP model creation and testing were done through the SNP and Variation Suite v8.9.0 (Golden Helix, Inc., Bozeman, MT, www.goldenhelix.com).

The scikit-learn python package was used for all ML methods to create and test the models. Multiple versions of the five-fold cross-validation were run with different parameters to determine the optimal settings for each approach. The highest scoring parameters were used in the model included in the analyses. When testing the SVM classifier, the radial basis function kernel proved most accurate for the current dataset. For the RF classifier, the number of trees built before the average was taken, the maximum number of features before splitting a node, the maximum number of levels in a tree, the minimum number of data points before a node split, and bootstrapping were optimized. The maximum depth of regression estimators, maximum leaf nodes, number of boosting stages, and learning rate were the parameters that were set for the individual models while they all performed best with Friedman mean square error for the criterion. A multi-layer perceptron classifier was used for the neural network. The parameters varied for individual models included the learning rate, hidden layers, and hidden units.

Results

Initially, the genomic heritabilities were calculated with mandibular distocclusion, distichiasis, and acral lick dermatitis, propducing values of 0.532, 0.240, and 0.071, respectively. The top-performing model for each ML approach's results are listed in Table 1, along with the results from the GBLUP for the average AUROC and standard deviation. An ANOVA test was used to determine significant differences between the model's AUROC with the *P*-values being 0.182, 0.569, and 0.148 for mandibular distocclusion, distichiasis, and acral lick dermatitis respectively.

Table 1. AUROC average and standard deviation for the models.

	GBLUP	SVM	RF	GB	NN
Mandibular distocclusion	0.811±0.032	0.744±0.084	0.677±0.107	0.699±0.096	0.723±0.097
Distichiasis	0.939±0.040	0.894±0.125	0.906±0.085	0.919±0.078	0.832±0.155
Acral lick dermatitis	0.660±0.067	0.686±0.053	0.631±0.114	0.591±0.053	0.695±0.024

Discussion

For all three traits, there were similar levels of prevalence within the population of five percent. This low rate can make classification problems difficult as the class imbalance can lead to a bias towards the negative. With that, there was not a significant difference in the performance of any of the models in the present study with a trend for GBLUP to outperform the other models. This matches previous work with breeding values using regression methods (Srivastava *et al.*, 2021). The genomic heritabilities for the three traits studied varied, but the impact on model performance matched previous work using classifiers scored by AUROC comparing RF and GBLUP with simulated data (Naderi *et al.* 2016). The genomic heritability for distichiasis also aligned with previous estimates based on pedigrees (Peterson *et al.*, 2015). SVM trended as the highest performing ML model, which may indicate that it is better suited for binary traits compared to regression problems where it was not found to trend higher than GB or RF (Srivastava *et al.*, 2021).

In simulated studies, the number of quality trait loci plays an important factor in the effectiveness of RF in particular (Naderi *et al.* 2016). The usage of NN for association studies and quality trait loci identification could help provide a pipeline to help identify which model would be most appropriate for breeding value construction based on the results (Liu *et al.*, 2019). Other approaches have used RF and GB to subset SNPs which are then used to construct genomic relationship matrices for breeding value construction (Li *et al.*, 2018). This may indicate that a blending of approaches may be needed to maximize the ability to predict phenotypes based on genetic data.

ML models have become an increasingly popular tool across bioinformatics. Research into when different approaches can be more or less effective than other methods can allow researchers to better assist breeders in creating healthier and higher performing animals.

References

Clark, S. A., and van der Werf, J. (2013). Methods in Molecular Biology, 321–330. https://doi.org/10.1007/978-1-62703-447-0_13

Fishman, G. A. (2003). Survey of Ophthalmology, 48(4), 452–458. https://doi.org/10.1016/s0039-6257(03)00052-3

Hoyer, N. K., and Rawlinson, J. E. (2019). Journal of Veterinary Dentistry, 36(4), 251–256. https://doi.org/10.1177/0898756420905136

Li, B., Zhang, N., Wang, Y.-G., George, A. W., Reverter, A., *et al.* (2018). Frontiers in Genetics, 9. https://doi.org/10.3389/fgene.2018.00237

Liu, Y., Wang, D., He, F., Wang, J., Joshi, T., *et al.* (2019). Frontiers in Genetics, 10. https://doi.org/10.3389/fgene.2019.01091

Naderi, S., Yin, T., and König, S. (2016). Journal of Dairy Science, 99(9), 7261–7273. https://doi.org/10.3168/jds.2016-10887

Nicholls, H. L., John, C. R., Watson, D. S., Munroe, P. B., Barnes, M. R., *et al.* (2020). Frontiers in Genetics, 11. https://doi.org/10.3389/fgene.2020.00350

Palella Gómez, A., Mazzucchelli, S., Scurrell, E., Smith, K., and Pinheiro de Lacerda, R. (2020). Veterinary Ophthalmology, 23(3), 506–514. https://doi.org/10.1111/vop.12748

Paterson, S., Midgley, D., and Barclay, I. (2007). In Practice, 29(6), 328–332. https://doi.org/10.1136/inpract.29.6.328

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., *et al.* Scikit-learn: Machine Learning in Python. Available at https://scikit-learn.org/stable/index.html

Petersen, T., Proschowsky, H. F., Hardon, T., Rasch, S. N., and Fredholm, M. (2015). Canine Genetics and Epidemiology, 2(1). https://doi.org/10.1186/s40575-015-0024-7

Qi Y. (2012) Random Forest for Bioinformatics. In: Zhang C., Ma Y. (eds) Ensemble Machine Learning. Springer, Boston, MA. https://doi.org/10.1007/978-1-4419-9326-7_11

SNP and Variation Suite v8.9.0 (Golden Helix, Inc., Bozeman, MT, www.goldenhelix.com)

Srivastava, S., Lopez, B. I., Kumar, H., Jang, M., Chai, H.-H., *et al.* (2021). Animals, 11(7), 2066. https://doi.org/10.3390/ani11072066

Suthaharan S. (2016) Support Vector Machine. In: Machine Learning Models and Algorithms for Big Data Classification. Integrated Series in Information Systems, vol 36. Springer, Boston, MA. https://doi.org/10.1007/978-1-4899-7641-3_9

Zhao, W., Lai, X., Liu, D., Zhang, Z., Ma, P., *et al.* (2020). Frontiers in Genetics, 11. https://doi.org/10.3389/fgene.2020.598318

395. Predicting the polled genotype of ungenotyped animals using machine learning methods

A.B. Gjuvsland

Geno, Storhamargata 44, 2317 Hamar, Norway; arne.gjuvsland@geno.no

Abstract

Selection for polled in Norwegian Red cattle started 30 years ago and intensified the last 5 years. The SNP chip includes the (celtic) polled mutation, making selection very accurate for genotyped animals. However, ~85% of calves are ungenotyped and phenotypic horn status has been used to predict the polled genotype. This prediction has low accuracy (r^2=0.69) which limits genetic progress. To improve predictions, a broad set of variables related to horns were extracted from the herd registration system. Data for ~134,000 genotyped animals was used to train and compare different machine learning methods for predicting polled genotype. Gradient boosting machines performed best, followed by random forests, deep neural networks, and generalized linear models. The best gradient boosting machine gave a large improvement in accuracy (r^2=0.87). We have implemented this in the routine evaluations and expect increased accuracy of selection for ungenotyped animals and faster increase in polled allele frequency.

Introduction

The main genetic determinant of horns in cattle is the Polled locus, with the polled allele (*P)* dominant over horned (*p*). However, the genetics of horns is more complex (Grobler *et al.* 2021). Heterozygous polled (*Pp*) animals sometimes develop loose horn-like structures called scurs, and the inheritance and phenotypic expression of scurs is dependent on sex and breed. A number of scurs loci have been mapped on different chromosomes in different breeds (Gehrke *et al.* 2002) and in Norwegian Red cattle (NR), a scurs locus has been identified on BTA5 (Gjuvsland *et al.* 2019)

Historically, most NR were horned, but the celtic *P* allele (Medugorac *et al.* 2012) segregated in the population. In 1990 when selection for polled started the *P*-frequency was 2-3%. With genomics came new possibilities, a haplotype test was developed and later the causal mutation was added to the chip. When large-scale cow genotyping started in 2017, polled was included in the index (current weight 5%). For new-born calfs the *P*-frequency is now ~20%. For genotyped animals the number of P alleles is weighted into the index. However, ~85% of NR calfs are not genotyped and for these the predicted number of *P* alleles is used. Farmers register the phenotypic horn status for new-born calfs, and in each category the mean number of *P*-alleles among genotyped animals has been used for the ungenotyped animals. However, the accuracy of prediction is low, limiting the genetic progress for polled.

The machine learning ecosystem has developed rapidly in the last decade, with methodological breakthroughs and striking applications in many fields. The speed and breadth of development makes it challenging to identify the best algorithm and software for the problem at hand. Automated machine learning (AutoML) aims at automatizing the process of training models, tuning hyperparameters and comparing a wide range of algorithms as much as possible. Several powerful and easy-to-use AutoML packages are now available (Feurer *et al.* 2021, Ledell and Poirier 2020).

The goal of this work was to make a better predictor of polled genotype for ungenotyped animals. The approach taken here to achieve this is to use a wider set of relevant registrations and test a range of machine learning algorithms.

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_395, © A.B. Gjuvsland 2022

Materials & methods

Dataset. Records from 1,154,254 animals with published breeding values were extracted from the database used for routine evaluations, 134,155 of the animals were genotyped. Each animal in the data set was characterized by birthyear, gender and the variables in Table 1.

Cow features: For cows additional variables were computed: number of calfs (*Noff*), total number of *P* alleles for sires of the calves (*Npnum_sires*), number of calves registered as phenotypically 'Polled' (*Npolloff*) and number of dehorned calves (*Ndhoff*). Two additional variables were derived as: *off_score1=Npolloff-Npnum_sires* and *off_score2=Ndhoff/Noff*.

Herd features: Registration quality varies between herds and some herd variables were defined in an attempt to account for this: within-herd correlation between being registered as horned and being dehorned (*corr_HDH*), within-herd correlation between not being registered as polled and being dehorned (*corr_NPDH*), proportion of herd registered as horned (*rate_H*), proportion of herd not registered as polled (*rate_NP*) and the number of animals (*N_herd*).

Prediction method 1 Registered horn status, category-means. The first method for predicting genetic horn status (*Pnum*) for ungenotyped animals was based on the registered phenotypic hornstatus (*Hornbud*). Within each phenotypic category (Horned, Polled, Unknown) the mean number of *P-alleles* for genotyped animals was used as the predicted number *of P*-alleles for all the ungenotyped animals. This method was used in the routine runs for NR until 2021.

Prediction method 2 Genotypes of parents or male ancestors. Here we computed *Pnum_ped,* the expected number of *P* alleles inherited from the parents (Equation 1) or male ancestors (eq.2, if dam is not genotyped) of an animal. Equation 1 was used for animals with genotyped dams, and Equation 2 for animals with ungenotyped dams.

$$Pnum_ped = 0.5 \times Pnum_sire + 0.5 \times Pnum_dam \quad (1)$$

$$Pnum_ped = 0.5 \times Pnum_sire + 0.25 \times Pnum_mgs + 0.125 \times Pnum_smgd \quad (2)$$

Table 1. Records extracted for all animals in the dataset.

Variable	Description	Values
Pnum	Number of *P* alleles for the animal. NA if not genotyped.	0, 1, 2, NA
Hornbud	Horn bud registered by farmer at birth	horned, polled, unknown
Dehorned	Dehorning registered by veterinarian	0, 1
Dehorned_age	Age (days) at dehorning	integer
Pnum_sire, Pnum_dam, Pnum_mgs, Pnum_smgd	Number of *P* alleles for sire, dam, maternal grandsire and sire of maternal granddam of the animal. NA if relative is unknown or ungenotyped.	0, 1, 2, NA
Hornbud_dam	*Hornbud* for dam of animal	horned, polled, unknown
Dehorned_dam	*Dehorned* for dam of animal	0,1
Snum_sire	Number of scurs alleles for sire of animal	0,1,2,NA

Prediction method 3 Automatic machine learning with all variabless. Here we used automatic machine learning to predict genetic horn status (*Pnum*) from all the other variables. We used the AutoML functionality (LeDell and Poirier, 2020) in the H2O (http://h2o.ai) R-package v3.32.1.1. The following machine learning algorithms were included: generalized linear models with elastic net regularisation, distributed random forests, gradient boosting machines, fully-connected multi-layer artificial neural network and ensemble models. The automl function was run on 20 CPUs for 4 hours and 10-fold cross-validation was used for tuning of hyperparameters and model comparison.

Results & discussion

Distribution of horn bud and dehorning registrations. Phenotypic registrations of horn status (*Hornbud)* and dehorning correspond fairly well with polled genotype. The majority of *pp* animals are registered as horned, while most *Pp* and *PP* are registered as polled (Figure 1). For both *pp* and *Pp* animals there are small, but clear sex-differences in registered horn status, with male calves being registered as horned more often than females. When it comes to dehorning there is a substantial sex-difference for *Pp* animals, with 21.8% of the males and 3.5% of the females being dehorned.

Automatic machine learning. Gradient boosting machines (GBM) outperformed the other methods, while random forests, deep learning networks and generalized linear models had quite similar performance (Table 2). For many applications with complex data types like images, deep learning networks are superior, but for tabular data gradient boosting is often better (Shwarz-Ziv and Armon 2022). In many machine learning use cases an ensemble of all models outperform the best performing single model, but here there is practically no difference between the ensemble and the best GBM. GBMs are complex, non-linear models, but 'variable importance' allows for some interpretability. The best GBM ranks *Hornbud* as the most important variable, followed by *Pnum_sire, Dehorned* and *Hornbud_dam.*

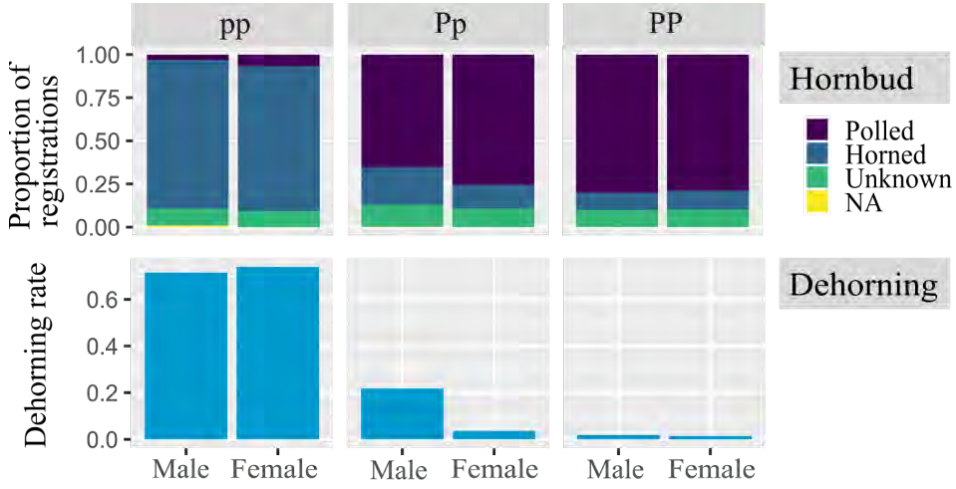


Figure 1. Horn phenotypes conditional on sex and polled genotype for 127K genotyped NR calves (92K females / 35K males) born 2005 or later.

Table 2. AutoML leaderboard.

Best model per algorithm	RMSE
Ensemble of all models	0.2959
Gradient boosting machine	0.2960
Distributed random forest	0.3208
Deep learning[1]	0.3336
Generalized linear model[2]	0.3385

[1] Fully connected multilayer neural networks.
[2] With elastic net regularization.

Comparison of prediction methods. Cross-validation (10-fold) on genotyped animals was used to evaluate and compare the prediction methods. *Pnum* and the predictions are numeric so we used correlation, rather than classification-metrics, to measure accuracy. Prediction of number of *P* alleles by registered horn status (Figure 2 red line) separates (on average) *pp* animals from *Pp* and *PP*, but it is unable to separate *Pp* from *PP* due to dominance. The accuracy, measured as the correlation to the true genotype (black line), is fairly low (cross-validated r^2=0.69). Using the expected number of *P* alleles from the genotype of parents or male ancestors (green line) separates *pp* from *Pp* and *Pp* from *PP*, but the accuracy goes down (r^2=0.61). Prediction with gradient boosting (blue line) separates well between the three polled genotypes and the accuracy increases substantially (r^2=0.87).

Conclusions and implementation

Machine learning with a broad set of variables gives a substantial increase in a prediction accuracy compared to prediction based on registered horn status. The prediction models are trained and validated on genotyped animals, however the practical use in the breeding programme is for predicting polled genotypes of ungenotyped animal. The value of the predictions therefore relies on the genotyped animals being representative for the whole population. We expect the genotyped animals to be representative in many aspects, but when it comes to things like unknown parents and pedigree errors there are systematic

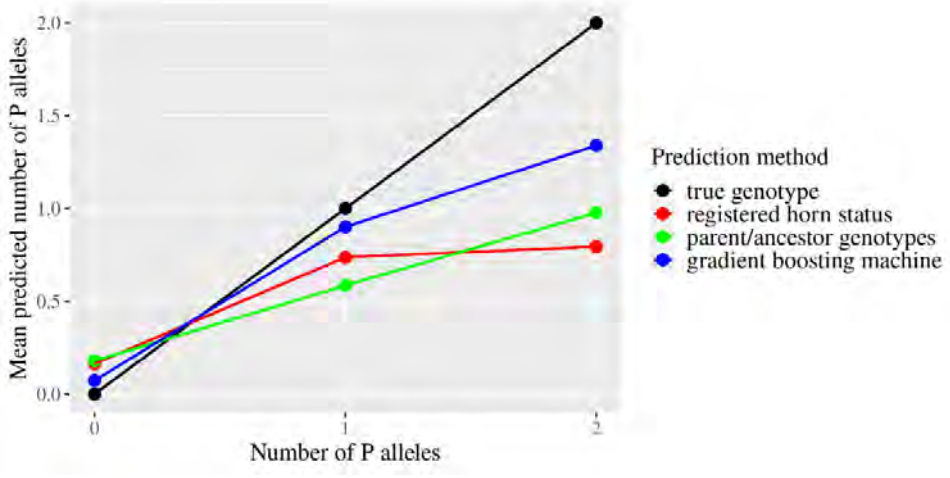


Figure 2. Mean predictions of polled genotype for genotyped animals. True number of *P*-alleles (x-axis) versus the mean predicted number (y-axis) when the genotype is masked in cross-validation.

differences and the prediction accuracy for ungenotyped animals will be lower than observed here. However, considering the low numbers of pedigree errors and unknown parents in the NR population we expect the predictions to be quite accurate for most animals and have implemented the gradient boosting method in our routine evaluations. This will give more accurate selection for polled in particular on the female side where many breeding animals in the herds and dams of potential selection candidates are ungenotyped.

References

Feurer, M., Eggensperger, K., Falkner, S., Lindauer, M., & Hutter, F. (2021). *arXiv preprint arXiv:2007.04074*. https://arxiv.org/abs/2007.04074v2

Gehrke L.J., Capitan A., Scheper C. *et al.* (2020) *Gen Sel Evol* 52:6. https://doi.org/10.1186/s12711-020-0525-z

Gjuvsland, A.B., Storlien H., and Larsgard A.G. (2019) Proc. of 70th EAAP, Ghent, Belgium.

Grobler R., van Marle-Köster E. and Visser C. (2021) *Livestock Sci*ence 247:104479. https://doi.org/10.1016/j.livsci.2021.104479

LeDell E.A. and Poirier S. (2020) 7th ICML Workshop on Automated Machine Learning. https://www.automl.org/wp-content/uploads/2020/07/AutoML_2020_paper_61.pdf.

Medugorac I., Seichter D., Graf A., *et al.* (2012) PLoS ONE 7(6):e39477. https://doi.org/10.1371/journal.pone.0039477

Shwartz-Ziv, R. and Armon, A. (2022) Information Fusion, 81:84-90. https://doi.org/10.1016/j.inffus.2021.11.011

396. A machine learning approach to classify high and low feed efficient or grower mink

A. Shirzadifar, G. Manafiazar, P. Davoudi, D.N. Do, G. Hu and Y. Miar*

Department of Animal Science and Aquaculture, Dalhousie University, Truro, Nova Scotia, B2N 5E3, Canada; miar@dal.ca

Abstract

This study aims at evaluating the feasibility of Machine Learning (ML) algorithms in classification of American mink (*Neovison vison*) into high/low performing groups for three different traits. Average Daily Gain (ADG), Feed Conversion Ratio (FCR), and Residual Feed Intake (RFI) were calculated for 1,088 mink over a period of 105 days. The 10% high and low feed efficient and grower animals were identified. To classify animals into the high/low feed efficient and grower groups based on the recorded features (initial weight and length at the bigining of the test, sex, age, and color-type), five common ML algorithms including Random Forest, Neural network, K-Nearest Neighbors, Logistic Regression, and Decision Tree were employed. The results showed that Random Forest was the most accurate algorithm in classification of 10% high and low ADG, FCR, and RFI groups with the accuracy of 98, 95, and 100%, respectively.

Introduction

Feed is the largest expense to the production of American mink (*Neovison vison*), which is the most widely raised fur-bearing animal in the world. Growth and feed efficiency traits such as Average Daily Gain (ADG), Feed Conversion Ratio (FCR), and Residual Feed Intake (RFI) are economically important traits that could be used to select the animals to increase the profitability of mink farmers (Yi *et al.*, 2018). Selection for growth and feed efficiency traits require substantial labor and resources to continually monitor Daily Feed Intake (DFI) and performance measures over a defined period. Alternatively, Machine Learning (ML) algorithms may potentially to be used to classify animals based on their feed efficiency (Nayeri, *et al.*, 2019). For instance, ML algorithms predicted the RFI class based on the gene expression levels in the liver and duodenum with an accuracy of 87% (Piles *et al.*, 2019). Therefore, the objective of this study was to investigate the feasibility of ML algorithms in classification of mink into high/low performing groups for ADG, FCR and RFI using the available features such as initial body weight, and length at the test bigining, sex, age, and color type.

Materials & methods

Data collection. Daily feed intake and body weight were recorded on 1,088 mink from the Canadian Centre for Fur Animal Research (CCFAR) at Dalhousie Faculty of Agriculture (Truro, NS, Canada) for 105 days from August 1st to November 14th in both 2018 and 2019. Additional information including sex, color type, body length, and age were recorded at the beginning of the experiment. Growth and feed efficiency measures (ADG, FCR, and RFI) were calculated using the following formulas (Jiao *et al.*, 2014):

$$\text{Average Daily Gain (ADG)} = \frac{\text{Final weight-Initial weight}}{\text{Test Period (Days)}} \quad (1)$$

$$\text{Feed Conversion Ratio (FCR)} = \frac{\text{Feed intake}}{\text{Average Daily Gain}} \quad (2)$$

$$\text{Residual Feed Intake} = \text{DFI}_{\text{actual}} - \text{DFI}_{\text{expected}} \quad (3)$$

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_396, © A. Shirzadifar *et al.* 2022

$$DFI_{expected}=\beta_0+\beta_1 ADG+\beta_2 BW^{0.75}+\varepsilon \quad (4)$$

$$BW^{0.75}=\left(\frac{\text{Initial BW+Final BW}}{2}\right)^{0.75} \quad (5)$$

Where β_0 is the intercept; β_1 is the partial regression coefficient of DFI on ADG; And β_2 is the partial regression coefficient of DFI on $BW^{0.75}$; and ε is residual error

Data analyses. The 10, 15, 20, and 25% high and low animals based on their ADG, FCR, and RFI records were identified as high and low groups. The middle group between high and low percentage groups was considered as conventional group and animals among this group were not classified. Five ML algorithms including Random Forest, Neural network, K-Nearest Neighbors, Logistic Regression, and Decision Tree were employed to classify the animals into high/low performing groups for each trait (ADG, FCR, and RFI) using initial body weight and length at the beginning of the test (the first day of test period), sex, age, and color-type. Twenty scenarios were considered for each ADG, FCR, and RFI traits including the combinations of five ML algorithms and four data sets (10, 15, 20, and 25% high/low) to test and find the most reliable scenario for the classification of animals into high and low feed efficiency and grower groups.

Classifier performance evaluators. The performance of ML was evaluated through a 10-fold cross-validation method. For this purpose, data were divided randomly into two parts for training (70% of data) and validation (30% of data), and the process was repeated 10 times. For each fold, the ML algorithms were developed with training data set and evaluated with validation data set. Results were compared with the actual data to evaluate the accuracy, sensitivity, specificity, precision, and F-measure. The ML algorithm performance was then evaluated using two classes including high and low groups based on the confusion matrix. Sensitivity and specificity represent the proportion of real high or low animals that are correctly classified. In precision calculation, the number of actual predicted high animals is divided by the total number of predicted high animals that were classified as true or false (Shirzadifar *et al.*, 2020). F-measure defines the harmonic mean of sensitivity and precision.

Results

The performance measures based on the 10, 15, 20 and 25% high and low ADG, FCR and RFI values were calculated for five different ML algorithms. The results indicated the high accuracy of over 90% in identifying the high and the low feed efficient and grower animals using ML algorithms and five features of initial body weight and length at the beginning of the test period, sex, age, and color-type. The most accurate ML algorithm was Random Forest with the accuracy of 98, 95, and 100% for 10% high/low ADG, FCR, and RFI groups, respectively. The confusion matrices of the Random Forest algorithm for high and low 10% groups are shown in Table 1. To assess the classification accuracy of ML algorithms, the total number of 64 (32+32) animals were used for the 10% high and low groups of ADG, FCR and RFI. Table 1 showed that 32 animals from high ADG group were correctly classified, while one animal from low ADG group was misclassified. There were two animals from low FCR group and one from high FCR group that were classified incorrectly. However, all animals in the high and low groups were correctly identified for the RFI data set.

To further evaluate the algorithms performance, sensitivity, specificity, precision, and F-measure were calculated in addition to accuracy. High accuracy (over 90%) was achieved for the 10, 15, and 20% high and low groups in all ML algorithms. Figure 1. illustrated the statistical measurements achieved by different ML algorithms for the 10% high and low groups.

Table 1. Confusion matrix of Random Forest classification based on the 10% high/low ADG, FCR, and RFI data sets.

		Predicted animal group								
		ADG			FCR			RFI		
		High	Low	Total	High	Low	Total	High	Low	Total
Actual animal group	High	32	0	32	31	1	32	32	0	32
	Low	1	31	32	2	30	32	0	32	32
Total		33	31	64	33	31	64	32	32	64

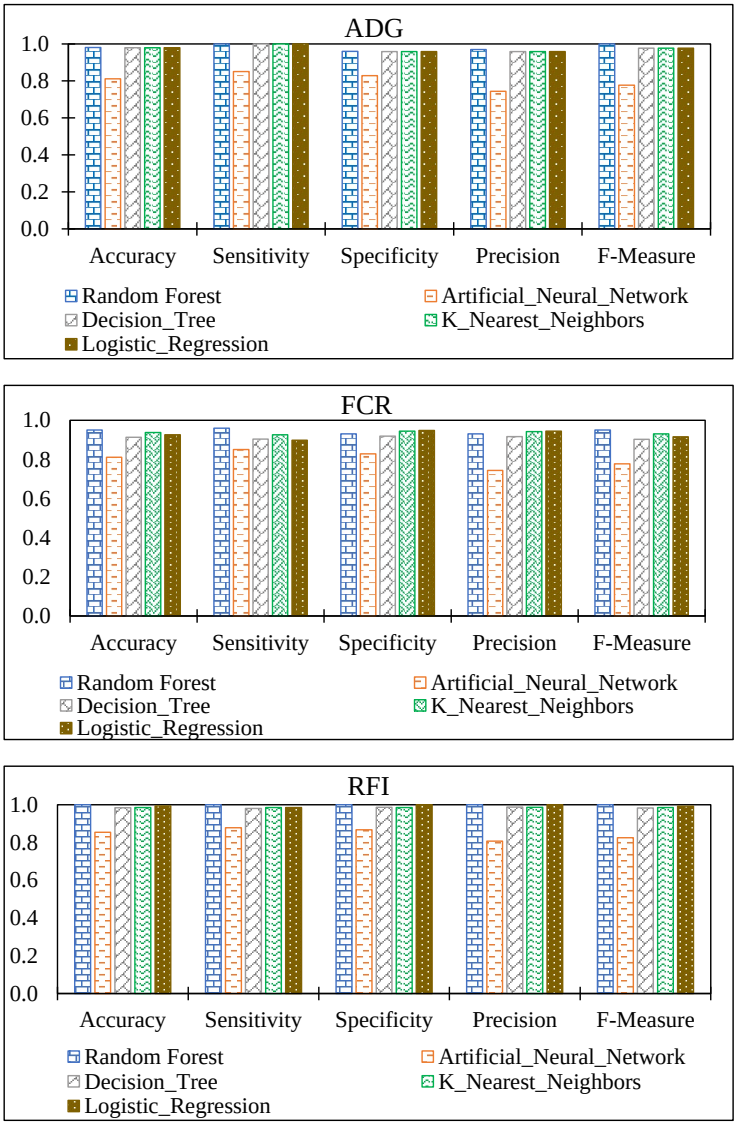


Figure 1. Performance measures achieved by different ML algorithms for the 10% high/low ADG, FCR, and RFI groups.

Discussion

This study demonstrated that ML algorithms were capable of identifying high/low performing mink groups for ADG, FCR and RFI traits based on the features including initial body weight and body length, at the beginning of the test, sex, age, and color-type without relying on expensive and labour intensive measurements. A high level of accuracy in prediction can be attributed to the high correlation between RFI, ADG, and FCR with the features used for prediction Although RF is the best method, all ML algorithms performed quite similarly; therefore, additional data is required to confirm if RF is the most appropriate approach.

References

Jiao, S., Maltecca, C., Gray, K. A., & Cassady, J. P. (2014). J. Anim. Sci., 92(6), 2377-2386. https://doi.org/10.2527/jas.2013-7338

Nayeri, S., Sargolzaei, M., and Tulpan, D. (2019). Anim. Health Res. Rev., 20(1), 31-46. https://doi.org/10.1017/S1466252319000148

Piles, M., Fernandez-Lozano, C., Velasco-Galilea, M., González-Rodríguez, O., Sánchez, J, *et al.* (2019). Genet. Sel. Evol., 51(1), 1-15. https://doi.org/10.1186/s12711-019-0453-y

Shirzadifar, A.M., Bajwa, S., Nowatzki, J., Shojaeiarani, J. (2020). Comput. Electron. Agric., 170, 105276. https://doi.org/10.1016/j.compag.2020.105276

Yi, Zh., Li, X., Luo, W., Xu, Zh., Ji, C., *et al.* (2018). J. Anim. Sci. Biotechnol., 9(1), 1-9. https://doi.org/10.1186/s40104-018-0261-1

397. A machine learning approach for predicting the most and the least feed-efficient groups in beef cattle

A. Shirzadifar[1], G. Plastow[2], J. Basarab[2], Y. Miar[1], C. Li[2,3], C. Fitzsimmons[2,3], M. Riazi[4] and G. Manafiazar[1*]

[1]Department of Animal Science and Aquaculture, Faculty of Agriculture, Dalhousie University, Truro, Nova Scotia, B2N 5E3, Canada; [2]Livestock Gentec, Department of Agricultural, Food and Nutritional Science, University of Alberta, Edmonton, AB, Canada; [3]Lacombe Research and Development Centre, Agriculture and Agri-Food Canada, 6000 C&E Trail, Lacombe, AB, T4L 1W1, Canada; [4]Computing Science Department, University of Alberta, Edmonton, AB, Canada; manafiazar@dal.ca

Abstract

This study evaluated six machine learning (ML) algorithms in classification of beef cattle into the most and least feed efficient groups using genomic information. Residual Feed Intake (RFI; a measure of feed efficiency) and 50k SNP genotype data for 4,050 animals were included from research and commercial herds in this analysis. A genome-wide association study was performed to determine SNP effects on the phenotypic variation of RFI. Next, SNPs were sorted based on their contribution to phenotypic variation, and six new datasets were created containing 500, 1K, 5K, 10K, 15K, and 20K SNPs. To predict the most and least feed efficient groups, six ML algorithms (Random Forest, Neural Network, K-Nearest Neighbors, Logistic Regression Decision Tree, and Support Vector Machine) were tested. K-Nearest Neighbor, with an accuracy of 61%, was found to be the most accurate algorithm to predict the most and least feed efficient groups.

Introduction

Approximately 55-75% of the total cost beef production is attributed to feed, and feed efficiency has a significant role to play in enhancing profitability and competitiveness for the Canadian beef industry. Residual Feed Intake (RFI), as a measure of feed efficiency, is considered a more appropriate measure since it is independent of body weight, growth and body composition and is a reflection of maintenance requirements. The production of beef cattle with low RFI is desirable as they consume less feed than expected for their body weight and growth. RFI is calculated as the difference between an animal's actual feed intake and its expected feed intake based on the regression of DMI on metabolic body weight and post-weaning ADG of its contemporary over a defined feeding period (e.g. 74-84 days). RFI is, however, limited by a lack of facilities to record daily feed intake on individual animals and costs approximately $1.5-2.5 per animal per day, or more. Machine learning (ML) uses big data to learn from and create new opportunities to reveal any possible pattern associated with data. ML is now being applied increasingly in the medical and engineering fields; however, to date it has had limited uptake in livestock mainly due to the limited number of observations and number of recorded traits. Among the limited number of available reports, Piles *et al.* (2019) employed the Elastic Net algorithm to predict RFI class in swine based on gene expression levels found in the liver and duodenum tissues of 32 high and 33 low RFI pigs with 78% accuracy. Compiling RFI data over 12 years from different commercial and research herds in this study availed us to generate a large dataset and to assess the capabilities of ML algorithms to predict the most and least feed efficient beef cattle.

Materials & methods

Data acquisition. This study included RFI performance data and genomic data (genotyped utilizing the Illumina 50K SNP BeadChip) from 4,050 beef cattle collected over a 12-year (2003-2015) period from research and commercial herds. The RFI value for each animal was calculated across contemporary groups, and animals were sorted based on their RFI values. Animals in the 15% top and bottom of the range based on their RFI values were selected and considered as the most (negative RFI; n=607) and least (positive RFI values; n=607) feed efficient groups making a total of 1,214 animals.

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_397, © A. Shirzadifar *et al.* 2022

Genomic analysis. The objective of the genomic analysis was to demonstrate to what degree phenotypic variation in RFI is accounted for by the SNPs of a 50k panel, and to choose a set of highly significant SNPs in order to reduce the dimensionality of the dataset for the development of ML algorithms. Using a quality-control procedure, SNPs with frequencies less than 0.05 and call rates less than 0.95 were excluded from the genomic dataset. A total of 42,689 SNPs was considered for further analysis with a mean frequency of 0.54 and a variation of 0.07.

The method of single-step regression was used to detect SNPs, and SNPs associated with RFI using BLUPF90-family software package program. The variance components of RFI were estimated using the following model:

$$y = Xb + Zu + e \quad (1)$$

where y represents RFI, and X is an incidence matrix relating RFI to fixed effects. In this model b is the vector of fixed effects solutions, Z is the incidence matrix that relates animals to each other and to phenotypes, u is the vector of breeding values, and e is the vector of random residuals. Following the estimation of SNP marker effects using model 1 with the data set of 4,050 beef cattle, SNPs were sorted in descending order based on the percentage of the genetic variance accounted for by each SNP. To reduce dimensionality, 6 different sets of SNPs (500,1K, 5K, 10K, 15K, and 20K SNPs) were selected based on their contribution to phenotypic variation in RFI.

Machine learning analysis. Six common machine learning algorithms were used to predict the most and least feed -efficient groups. The algorithms were Random Forest, Neural Network, K-Nearest Neighbors (KNN), Logistic Regression, Decision Tree, and Linear Support Vector Machine. Thirty-six scenarios were tested to predict the highest and lowest RFI groups in a binary formats with the combinations of six ML algorithms and six genomic datasets (500, 1k, 5k, 10k, and 15k and 20K SNPs) to identify the most reliable scenario for classifying the most and the least efficient animal groups.

Classifier performance evaluators. The performance of the ML algorithms was evaluated using a 10-fold cross-validation procedure. Specifically, the data were randomly divided between training (70%; n=848) and validation (30%; n=366) sets, and the process was repeated ten times. For each fold, the ML algorithms were developed using training datasets and evaluated with a validation dataset. Results were compared with observed RFI groups to evaluate the accuracy, sensitivity, specificity, precision, and F-measure. Following that, we evaluated the performance of the ML algorithms using two classes of the most and least efficient animal groups based on the confusion matrix. In the confusion matrix, each row represents an instance in an actual class, while each column represents an instance in a predicted class, or vice versa (Piryonesi *et al.*, 2020). Sensitivity and specificity are the proportion of actual high- or low-efficiency animals correctly classified in each scenario. In precision calculation, the number of actual predicted most efficient animals is divided by the total number of predicted most efficient animals that were classified as true or false (Shirzadifar *et al.,* 2020). F-measure defines the harmonic mean of sensitivity and precision.

Results

Genomic analysis results. Results from the GWAS analysis and individual SNP effect are illustrated in the form of Manhattan plots (Figure 1). For the Manhattan plot, the Y axis represents the negative logarithm of the *P*-value for each single nucleotide polymorphism (SNP), and the X axis is the chromosomal location of the SNP. Each dot on the Manhattan plot signifies a SNP. The top10 highly significant SNPs are located on chromosomes 14, 2, 16, 18, 1, 5, 3, 22, 8 for RFI.

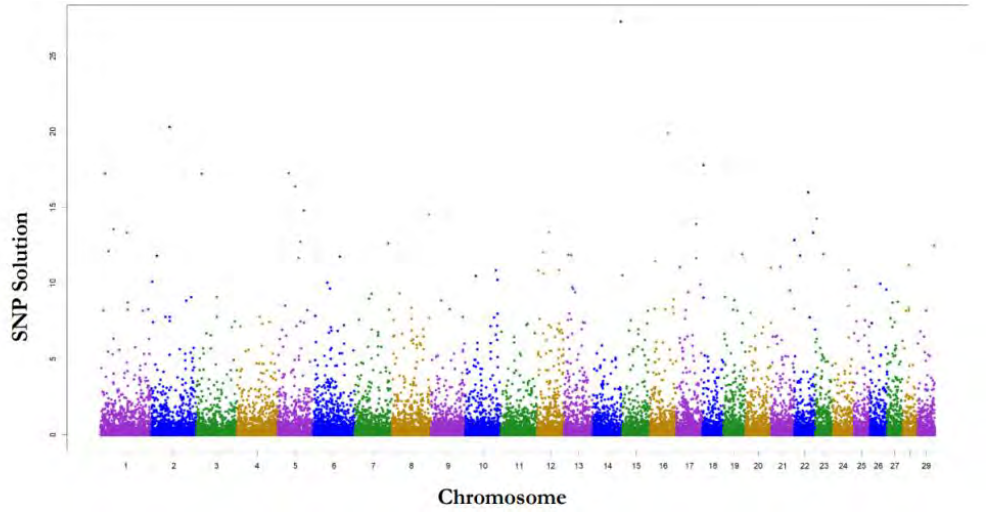


Figure 1. Manhattan plot of single nucleotide polymorphism (SNP) variance explained by each SNP for RFI.

ML algorithm classifier performance results. The performance measures based on genomic datasets including 500, 1k, 5k, 10k, 15k, and 20K SNPs were calculated for six different ML algorithms. The results showed an accuracy of over 53% in identifying the most efficient and the least efficient groups in all selected genomic datasets, excluding logistic regression. On a 15 K genomic dataset, KNN algorithem was the most accurate ML algorithm with an accuracy of 61% (Figure 2.). Nevertheless, the accuracy of the Logistic Regression algorithm decreased dramatically when the numbers of SNPs were increased. To further evaluate the algorithms' performance, sensitivity, specificity, precision, and F-measure were calculated in addition to accuracy. The performance of the KNN algorithm using 15K genomic dataset was the highest among the combination of the other ML algorithms with genomic datasets.

Figure 2 illustrates the statistical measurements of the performance of 15% sorted value for different ML algorithms in 15K dataset. The accuracy, sensitivity, specificity, precision, and F- measure were obtained 61, 60, 64, 61, and 60%, respectively in 15K genomic dataset.

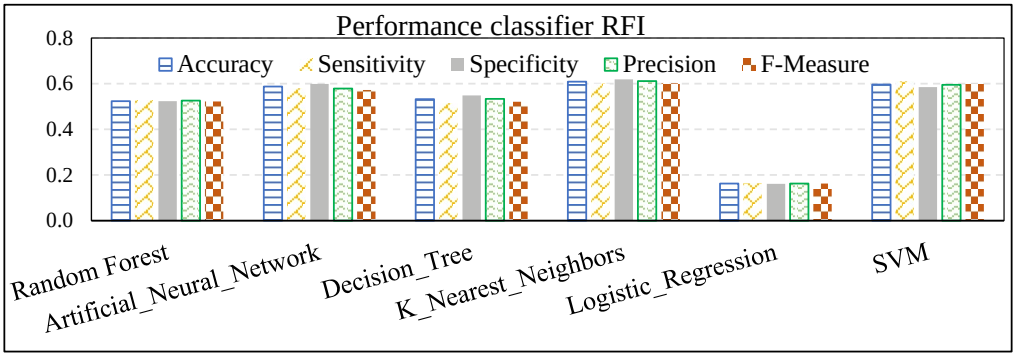


Figure 2. Performance measures for 15% sorted value and different ML algorithms based on RFI.

Table 1 presents the test set validation results (confusion matrix) of KNN analysis for all genomic datasets. It was noted that in classifier performance evaluators, 70% of the dataset were used to develop the model and 30% was selected at random to test the accuracy of the prediction. A total number of 366 animals (183 animals within each group) as a test was used to assess the classification accuracy of ML algorithms using different sets of SNPs, and the results are demonstrated in Table 1. The results based on confusion matrix demonstrated that 15K SNPs had the best predictability such that 101 out of 183 animals were correctly classified in the most efficient group, and 118 out of 183 animals from the least efficient animal group were correctly classified.

Table 1. Confusion matrix of KNN algorithm classifier based on 15% RFI and selected genome datasets.

		Predicted animals											
		500		1K		5K		10K		15K		20K	
		ME[1]	LE[2]	ME	LE	ME	LE	ME	LE	ME	LE	ME	LE
Actual animals	ME	98	85	89	94	90	93	97	86	101	82	95	88
	LE	89	94	66	117	87	96	88	95	75	118	72	111
Total		187	179	155	211	177	189	185	181	176	200	167	199

[1] ME = most efficient.

[2] LE = least efficient.

Discussion

This study demonstrated that KNN algorithms has potential to identify the most and the least efficient animal groups based on genomic information of beef cattle without relying on expensive and hard-to-measure traits such as dry matter intake. The output of this project presents evidence of how new ML-based algorithms have the potential to be used in the livestock breeding industry.

Acknowledgements

We thank financial support from Alberta Innovates and in-kind contributions from Beefbooster.

References

Jiao, S., Maltecca, C., Gray, K. A., & Cassady, J. P. (2014). J. Anim. Sci., 92(6), 2377-2386. https://doi.org/10.2527/jas.2013-7338

Nayeri, S., Sargolzaei, M., and Tulpan, D. (2019). Anim. Health Res. Rev., 20(1), 31-46. https://doi.org/10.1017/S1466252319000148

Piles, M., Fernandez-Lozano, C., Velasco-Galilea, M., González-Rodríguez, O., Sánchez, J, *et al.* (2019). Genet. Sel. Evol., 51(1), 1-15. https://doi.org/10.1186/s12711-019-0453-y

Piryonesi S. Madeh; El-Diraby Tamer E. (2020). J. Infrastruct. Syst. s. 26(1): 04019036. https://doi.org/10.1061/(ASCE)IS.1943-555X.0000512

Seabury, C.M., Oldeschulte, D.L., Saatchi, M. *et al.* (2017). BMC Genom. 18, 386. https://doi.org/10.1186/s12864-017-3754-y

Shirzadifar, A.M., Bajwa, S., Nowatzki, J., Shojaeiarani, J. (2020). Comput. Electron. Agric., 170, 105276. https://doi.org/10.1016/j.compag.2020.105276

Yi, Zh., Li, X., Luo, W., Xu, Zh., Ji, C., *et al.* (2018). J. Anim. Sci. Biotechnol., 9(1), 1-9. https://doi.org/10.1186/s40104-018-0261-1

398. Combination of SNP markers for discrimination of a target chicken population using machine learning algorithms

E. Cho, S. Cho, M. Kim, T.K. Ediriweera, D. Seo and J.H. Lee*

Chungnam National University, 99, Daehak-ro, Yuseong-gu, 34134, Daejeon, Republic of Korea; junheon@cnu.ac.kr

Abstract

Identifying target populations using genetic information is a necessary aspect for the conservation of pure breed and discrimination of chickens in the market scenario. In this study, the optimal combination of single nucleotide polymorphism (SNP) markers for discriminating the target chicken breed (the Yeonsan Ogye breed) was presented using high-density SNP chip data. SNP markers specific to the target population were discovered through case-control genome-wide association study (GWAS) and filtered out based on the linkage disequilibrium blocks, and optimal SNP markers were selected by applying a machine learning algorithm. Through the machine learning approach, the identification power of the 38 optimal SNP marker combination for the specific chicken population was confirmed, and the marker combination demonstrated complete accuracies. Hence, the GWAS and machine learning models used in this study can be efficiently utilized to find out the optimal combination of markers that can discriminate target populations using multiple SNP markers.

Introduction

Yeonsan Ogye, characterized by black feathers, skin, and bones, is one of the Korean native chicken (KNC) breeds and is highly regarded as a natural heritage. Generally, the methods for identifying chicken breeds are based on morphological characteristics, but distinguishing them from other breeds with similar morphology is difficult. From this point of view, genetic information can be used effectively for precise identification. Various genetic markers have been developed and used to obtain genetic information, but the most effective tool of genetic markers for discrimination of target breed is single nucleotide polymorphism (SNP) markers. The SNP genotype data can provide rapid and accurate information with selected marker combinations for breed identification. The objective of this study was to select the optimum SNP marker combination to discriminate a target chicken population (Yeonsan Ogye) from other populations using machine learning algorithms.

Materials & methods

Samples and genotypes. A total of three data sets were used in this study: Set 1 and Set 2 for selecting SNP markers, and Set 3 for validation test. Set 1 and Set 2 consisted of 198 populations (3,904 individuals), genotyped with Affymetrix 600K SNP array (Kranis *et al.*, 2013). Set 3 constituted 127 individuals, genotyped with custom 60K SNP array and Illumina 60K SNP array (Groenen *et al.*, 2011). The imputation step was performed before analysis using 60K genotype data. More information on each set has shown in Table 1.

Pre-processing and SNP pruning. Based on the genotype call rate (missing rate ≥10%) and the minor allele frequency (MAF≥0.05), 542,717 SNPs were derived from the merged common SNPs. These SNPs were analysed for case-control genome-wide association study (GWAS) using PLINK 1.9 software. In this analysis, the case group was the Yeonsan Ogye population, whereas the control group was composed of the other populations. Following the case-control test, the linkage disequilibrium (LD) was calculated, and the LD-block based SNP pruning analysis was conducted to select one SNP per each of the 50 LD blocks.

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_398, © E. Cho *et al.* 2022

Table 1. Summary of the samples used in this study.

Object	Class	Population		Number of animals	Total
SNP[1] selection	Set 1	KNC[2]	Indigenous	208	3,904
			Adapted	218	
		Yeonsan Ogye		189	
	Set 2	Commercial	Broiler	34	
			Layer	20	
		SYNBREED[3]		3,235	
Validation test	Set 3	Yeonsan Ogye		67	127
		KNC		30	
		Crossbreed of Yeonsan Ogye and White Leghorn		30	

[1] SNP, single nucleotide polymorphism.

[2] KNC, Korean native chicken.

[3] Various other global chicken breeds from the SYNBREED project in Germany (Malomane *et al.*, 2019).

Feature selection. The feature selection by machine learning methods was applied for pruned SNP markers. The feature importance value was calculated through the Random Forest machine learning algorithm. Finally, the optimum SNP marker combination was selected with values higher than the average feature importance value.

Evaluation of the accuracy. To confirm the accuracy for discrimination of the target chicken population, eight different machine learning models were used: Decision Tree, AdaBoost, Support Vector Machine, Quadratic Discriminant Analysis, Random Forest, Linear Discriminant Analysis, K-Nearest Neighbours, and Naïve Bayes. To resolve data imbalance before analysis, only one individual from each group was randomly sampled in the control group consisting of 197 populations. Accordingly, the data set was split into two sets based on the five times repeated K-fold cross-validation (K=10): training set (70%) and test set (30%). We used the 'caret' package in R software to build classification models.

Results

Principal component analysis (PCA) was performed to confirm the genetic distribution trend of the entire population using 600K SNP genotype data. As a result, the Yeonsan Ogye population was remarkably clustered from the SYNBREED populations, as well as the KNC populations (Figure 1). A total of 38 SNPs

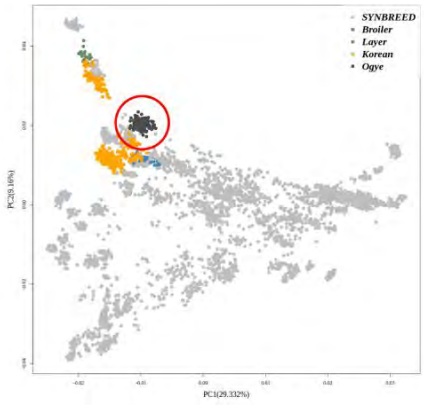


Figure 1. PCA plot based on 600K SNP genotype data. Note that Yeonsan Ogye population (inside red circle) is a distinct population by comparing with the data from Korean breeds and the global SYNBREED populations.

was selected through the SNP pruning and feature selection processes. In the machine learning approach for evaluation of the classification using 38 optimal SNPs, all of the algorithms discriminated the Yeonsan Ogye population precisely (Figure 2). All confusion matrix values (Accuracy, Specificity, Sensitivity, Precision, and F1-score) were also confirmed to be 1.00. In the case of the validation test, the 30 SNPs overlapped with the previously selected markers. Even with these markers, it was possible to sufficiently differentiate the targeted population.

Discussion

Strategies for optimal breed identification are utterly important in livestock conservation and research. The capability to discriminate chicken breeds at the genetic level could increase consumer confidence and protect unique genetic resources. Recently, a large amount of produced genotype data is available as biomarkers in various ways. However, finding optimal SNP markers for identifying targeted populations using high-density SNP chips is still quite complex.

According to previous studies, SNP markers were used for genetic classification based on PCA, *F*-statistics, and genotype frequencies (Suekawa *et al.*, 2010; Wilkinson *et al.*, 2011). In this study, significant markers were obtained based on the GWAS and LD pruning results using 600K high-density SNP chip data. Johnson *et al.* (2010) explained that it is often difficult to determine which genetic markers identified through GWAS are actual causative genes in response to LD. Also, according to Bakshi *et al.* (2016), more effective results were obtained when removing SNPs with strong LD relationships and using them for analysis. In this analysis, it was also demonstrated that the target group was effectively discriminated when the SNP markers were selected in consideration of the LD relationship.

We applied machine learning algorithms to determine efficient marker combinations using preselected SNPs. In previous studies, it was confirmed that machine learning was applied to discover and select SNP markers (Seo *et al.*, 2021; Schiavo *et al.*, 2020). In addition, selecting markers by applying the feature selection technique to the GWAS results can reduce dimensions and overfitting errors, resulting in more accurate predictions (Bermingham *et al.*, 2015).

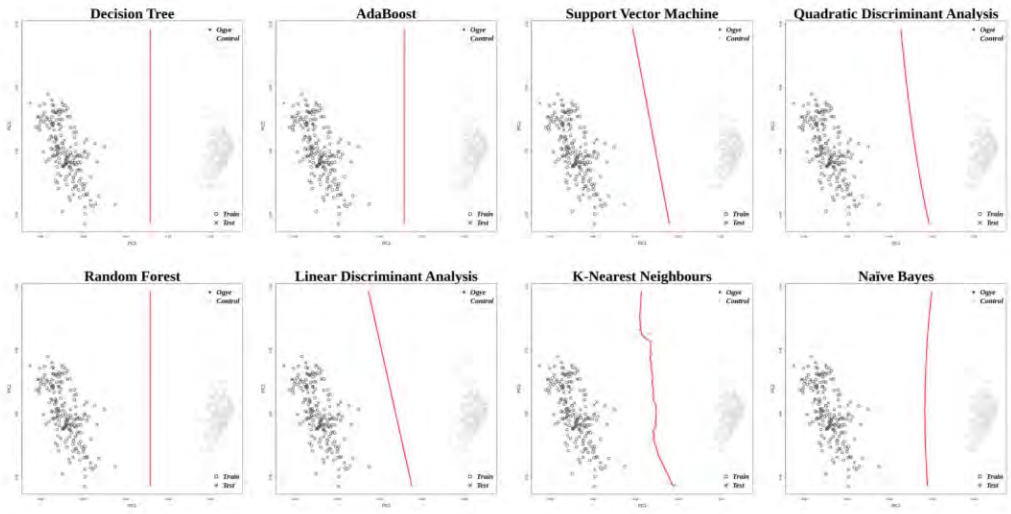


Figure 2. Group clustering trend plot for eight machine learning models using 38 SNPs.

The decisive reason for obtaining such precise results might be that the Yeonsan Ogye was a genetically unique breed. It was verified that the Yeonsan Ogye had an independent gene pool from entirely black chickens in the SYNBREED group such as Cemani and Sumatran from Indonesia, and Silkies from China.

The combinations selected for identification targeting the Yeonsan Ogye pure line showed impressive results in the validation test. The 30 SNPs correctly differentiated KNC and Yeonsan Ogye as in the SNP selection process. In addition, the crossbreed groups between the Yeonsan Ogye and White Leghorn were also clearly distinguished. These marker combinations showed the ability to discriminate the individuals with the same phenotype as those of the purebred Yeonsan Ogye individuals.

Since the breed-specific markers are identified through the pure lines, there is a limitation as an application to breeds that have not been verified under the marker selection process. Of course, although verification analysis was performed on crossbreed in this study, it would be complicated to flexibly react to hybridizations with other breeds other than White Leghorn. Therefore, the accuracy of the discrimination power possessed by the optimal SNP marker combination will be properly recalculated when information composed of more diverse populations is additionally provided.

Identification of populations by combinations of SNP markers can efficiently discriminate different types of populations and confirm their genetic diversity. Ultimately, this can be used as a way to protect the right to use a specific breed population in a society where breeds can secure national competitiveness as a genetic resource.

References

Bakshi A., Zhu Z., Vinkhuyzen A.A.E., Hill W.D., *et al.* (2016) Sci Rep 6(32894). https://doi.org/10.1038/srep32894

Bermingham M.L., Pong-Wong R., Spiliopoulou A., Hayward C., *et al.* (2015) Sci Rep 5(10312). https://doi.org/10.1038/srep10312

Groenen M.A., Megens HJ., Zare Y., Warren W.C., *et al.* (2011) BMC Genomics 12(274). https://doi.org/10.1186/1471-2164-12-274

Johnson R.C., Nelson G.W., Troyer J.L., Lautenberger J.A., *et al.* (2010) BMC Genomics 11(724). https://doi.org/10.1186/1471-2164-11-724

Kranis A., Gheyas A.A., Boschiero C., Turner F., *et al.* (2013) BMC Genomics 14(59). https://doi.org/10.1186/1471-2164-14-59

Malomane D.K., Simianer H., Weigend A., Reimer C., *et al.* (2019) BMC Genomics 20(345). https://doi.org/10.1186/s12864-019-5727-9

Schiavo G., Bertolini F., Galimberti G., Bovo S., *et al.* (2020) Animal 14(2):223-232. https://doi.org/10.1017/S1751731119002167

Seo D., Cho S., Manjula P., Choi N., *et al.* (2021) Animals 11(1):241. https://doi.org/10.3390/ani11010241

Suekawa Y., Aihara H., Araki M., Hosokawa D., *et al.* (2010) Meat Sci. 85(2):285-288. https://doi.org/10.1016/j.meatsci.2010.01.015

Wilkinson S., Wiener P., Archibald A.L., Law A., *et al.* (2011) BMC Genet. 12(45). https://doi.org/10.1186/1471-2156-12-45

399. LASSO and SVM: an alternative approach to identify associated genome regions for simple and complex traits in cattle

S. Widmer[1*], F.R. Seefried[2], C. Drögemüller[1] and P. von Rohr[2]

[1]*Institute of Genetics, University of Bern, Bremgartenstrasse 109a, 3012 Bern, Switzerland;*
[2]*Qualitas AG, Chamerstrasse 56, 6300 Zug, Switzerland;* sarah.widmer@vetsuisse.unibe.ch

Abstract

Variable selection via the LASSO regression analysis followed by a classification of phenotypes into two groups of cases and controls using SVM was performed to identify genomic regions associated with phenotypic traits of interest. This approach was successfully validated by identifying previously known genomic associations for two Mendelian traits and a complex polygenic trait in Swiss cattle populations. Furthermore, new associated regions for the trait multiple birth were found in Holstein. Therefore, this approach proved to be a valuable alternative to identify associated genomic regions for inherited traits in livestock.

Introduction

Support Vector Machine (SVM) is a general-purpose classification method (Cortes and Vapnik 1995) and is used in many fields (James *et al.* 2021). For example, it was able to find SNPs associated with the risk for type 2 diabetes in humans and to perform genotype-based predictions of the affection status (Ban *et al.* 2010). SVM classifies data consisting of different groups by separating hyperplanes. The separating hyperplanes are defined by explanatory variables such as SNP genotype effects. This definition makes SVM suitable to analyse data that can be divided into two groups, e.g.: cases and controls. Although SVM can be used for high-dimensional data, it is important to first identify the subset of relevant explanatory variables. It is necessary to select relevant SNPs to avoid overfitting which might impair classification negatively. Least Absolute Shrinkage and Selection Operator (LASSO) performs variable selection in a linear model using a constraint on the norm of the absolute values of the coefficients (Tibshirani 1996). LASSO regularizes the coefficient estimates, shrinks them towards zero and consequently reduces their variance significantly (James *et al.* 2021). Combining LASSO and SVM together with using raw phenotypes instead of predicted breeding values promises a highly computational efficient method.

In this study, we first validated the proposed method as an alternative approach to detect genomic associations for two Mendelian traits: *CNGB3*-related achromatopsia, also reported as Original Braunvieh Haplotype 1 (OH1) (Häfliger *et al.* 2021) and *TWIST2*-related depigmentation known as belt pattern (Awasthi Mishra *et al.* 2017). Subsequently we applied the workflow on stature in Holstein, a complex but well studied genetic trait (Bouwman *et al.* 2018). Finally, the proposed method was used to find new genomic regions associated with the trait multiple birth in Holstein cattle, representing a still poorly understood polygenic trait.

Materials & methods

Phenotypes. The proposed approach was validated using three different datasets of three different Swiss cattle populations (Table 1). The phenotypic and genotypic data were provided by the Swiss cattle breeding organisations. The definition of cases and controls for both monogenic traits achromatopsia and belt was based on phenotypic observations. Regarding the trait stature, cows in first lactation born between 2016 and 2019 with an age at first calving between 730 and 820 days were selected. From the routine conformation scoring, 500 top / bottom cows for measured sacrum height were used to define the groups of cases and controls, respectively. For multiple birth, dams with at least one multiple birth event (mostly twins) were defined as cases and dams with at least 3 singleton calvings as controls.

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_399, © S. Widmer *et al.* 2022

Table 1. Final datasets per trait and the sizes of the respective case and control groups.

Trait	Population	No. cases	No. controls	No. SNP
Achromatopsia	Original Braunvieh	8	231	670,140
Belt	Brown Swiss	92	1,644	675,828
Stature	Holstein	500	500	680,502
Multiple birth	Holstein	238	919	683,277

Genotypes. Animals were genotyped under the umbrella of genomic selection using different arrays that include between 9k and 850k SNPs. A standard imputation approach was applied, which led to the final datasets (Table 1).

LASSO. LASSO uses the residual sum of squares plus a penalty term to estimate coefficients and to do variable selection in a linear model (Tibshirani 1996). LASSO was described in detail by James *et al.* 2021. The LASSO analyses were carried out using the R-package glmnet (Friedman *et al.* 2010). The parameter λ was estimated using a ten-fold leave-one-out cross-validation. The cross-validation resulted in two different estimates for λ: λ_{min} is the estimate with lowest cross-validation error. Adding one standard error to λ_{min} leads to λ_{1se}. Both estimates of λ were used in the LASSO analyses for all four datasets. LASSO resulted in a smaller set of SNPs compared to the complete marker-set.

SVM. SVM separates two groups of data (such as cases and controls) with the hyperplane that maximizes the distance to the training data of the two groups (Cortes and Vapnik 1995). The separating hyperplane is determined by the training data (James *et al.* 2021) that we defined as a random sample of 80% of the complete dataset. The remaining 20% of the data is used as test data to validate the estimated classification model. A linear kernel was used for classification. For SVM analyses R-package e1071 was applied on the SNPs selected by LASSO for the validation of the variable selection (Meyer *et al.* 2021). The prediction of animals outside the high and low groups is not considered yet.

Results

The results of the LASSO analyses of the four datasets are shown in Table 2. Using λ_{1se} instead of λ_{min} led to a lower number of SNPs selected by LASSO for all datasets.

From the LASSO analysis, the absolute coefficient values were combined using a sliding window approach whereof each window consisted 50 SNPs. Window coefficients are plotted as a function of their chromosomal position (Figure 1). Only windows with a higher impact on the trait and the colocalization of possible candidate genes were considered (Table 3). Therefore, the window coefficients must be higher as the 5-fold standard deviation of all absolute coefficients. For each of the two Mendelian traits, as expected

Table 2. Results of LASSO and SVM analysis for all four traits.

	LASSO No. SNP		SVM rate of correct prediction (%)		SVM F_1 score	
Trait	λ_{min}	λ_{1se}	λ_{min}	λ_{1se}	λ_{min}	λ_{1se}
Achromatopsia	8	5	100	100	1	1
Belt	1,243	420	100	100	1	1
Stature	674	393	100	100	1	1
Multiple birth	232	1	89.7	76.3	0.765	0.035

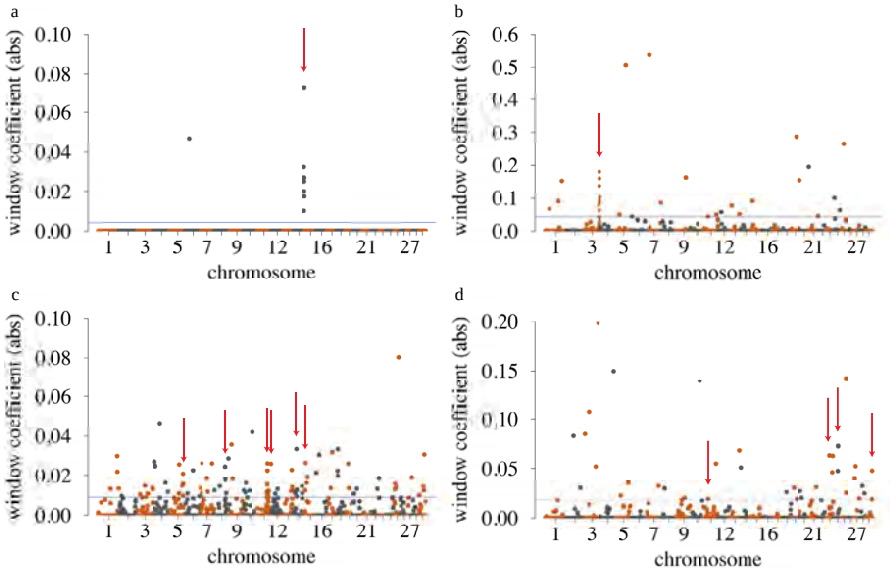


Figure 1. Genome-wide Manhattan plots of the sum of the absolute coefficient values for each window. For (a) achromatopsia and (d) multiple birth λ_{min} was used, while for (b) belt and (c) stature λ_{1se} was used. Red arrows are highlighting the QTL from Table 3 and the blue lines indicate the values of the 5-fold standard deviation of all absolute (abs) window coefficients per trait.

a clear signal was identified on BTA 14 for achromatopsia and BTA 3 for belt, respectively (Figure 1a and 1b). For the trait stature, multiple QTL regions were observed (Figure 1c). Several associated regions were identified for multiple births in Holstein (Figure 1d). The rate of correct predictions from SVM analysis (Table 2) using the test data was 100% (F_1 score=1) for all three validation datasets and for both variants (λ_{1se} and λ_{min}). For the trait of multiple birth, the rate of correct predictions in the test data reached 90% using λ_{min} (F_1 score=0.765).

Discussion

For all four analysed traits, LASSO resulted in a set of SNPs which is considerably smaller compared to the complete marker-set. This forms the basis for the identification of associated genomic regions using raw phenotypes. Using λ_{1se} instead of λ_{min} resulted in a lower number of SNPs selected, as expected from the properties of LASSO. The resulting set of SNPs was used in an SVM classification to group the data into cases and controls and to validate previous results. The rate of correct predictions of SVM for both λ-estimates were comparable and hence λ_{1se} (except for achromatopsia and multiple birth) was favoured because it yields sparser models. Regarding the monogenic recessive disorder achromatopsia, LASSO selected correctly the region of *CNGB3* on BTA 14 (Häfliger *et al.* 2021). For the monogenic dominant inherited belt phenotype, a signal in a window on BTA 3 with *TWIST2* was observed (Awasthi Mishra *et al.* 2017). Concerning stature, two regions on BTA 11 and the signals on BTA 5, 8 and 14 were previously reported as QTL for bovine height in a meta-analysis using the 1000 Bull Genomes project data (Bouwman *et al.* 2018). We propose *MMP13* as a new candidate gene mapping to the region of a QTL found on BTA 15 for stature in cattle, as it influences bone mineralization and growth plate cartilage (Ståhle-Bäckdahl *et al.* 1997). Finally, regarding multiple birth in Holstein, we found multiple associated regions with genes related to the reproduction cycle and oocyte maturation (Table 3). The QTL in the region of *FSHR* and *LHCGR* genes on BTA 11 agrees well with recent findings of GWAS and classical fine mapping approaches

Table 3. Identified associated genome regions with selected candidate genes for the four analysed traits.[1]

Trait	BTA	Start position[2]	End position[2]	Coefficient[3]	Candidate gene[2]
Achromatopsia	14	75,583,503	78,489,957	0.2055	*CNGB3*
Belt	3	117,018,007	118,619,571	0.7183	*TWIST2*
Stature	5	104,712,189	105,942,817	0.0467	*CCND2*
	8	80,544,099	83,316,183	0.0551	
	11	79,311,390	79,594,907	0.0347	
	11	105,098,263	105,220,577	0.0258	
	14	19,996,921	23,379,474	0.0688	*PLAG1*
	15	4,836,530	5,916,553	0.0389	*MMP13*
Multiple birth	11	31,139,464	31,340,099	0.0236	*FSHR, LHCGR*
	23	23,927,635	25,017,848	0.0777	*GSTA1, GSTA2, PAQR8, TMEM14A*
	24	33,994,132	36,319,363	0.1255	*GATA6, ENOSF1, ADCYAP1*
	29	48,563,749	50,109,725	0.0671	*IGF2, INS, CTSD*

[1] For achromatopsia and multiple birth using λ_{min} and for belt and stature using λ_{1se}.

[2] Using the ARS-UCD1.2 reference assembly.

[3] Sum of all absolute values of the coefficient in the interval (typically more than one window).

using predicted breeding values (Widmer *et al.* 2021). Dichotomizing the quantitative trait by selecting extreme phenotypes as cases and controls helped here to unravel the potential genetic factors important for the expression of the phenotypes of interest. The rates of correct predictions from SVM were 100% for the three validation datasets and 90% for multiple birth (using λ_{min}), respectively. Hence, the validation of the herein applied alternative methods was successful.

In conclusion, using the combination of LASSO variable selection and SVM classification, we were able to detect associated genome regions for simple and complex inherited traits in cattle using raw phenotypes. Therefore, we propose this approach as a valuable and efficient alternative to GWAS based on predicted breeding values.

References

Awasthi Mishra N., Drögemüller C., Jagannathan V., Keller I. *et al.* (2017) PLoS One 12(6):e0180170. https://doi.org/10.1371/journal.pone.0180170.

Ban H.-J., Heo J. Y., Oh K.-S. and Park K.-J. (2010) BMC Genet 11:26. https://doi.org/10.1186/1471-2156-11-26.

Bouwman A. C., Daetwyler H. D., Chamberlain A. J., Ponce C. H. *et al.* (2018) Nat Genet 50(3):362–367. https://doi.org/10.1038/s41588-018-0056-5.

Cortes C., and Vapnik V. (1995) Mach Learn 20:273–297. https://doi.org/10.1007/BF00994018

Friedman J.H., Hastie T., and Tibshirani R. (2010) J Stat Softw 33(1):1-22. https://doi.org/10.18637/jss.v033.i01.

Häfliger I. M., Marchionatti E., Stengård M., Wolf-Hofstetter S. *et al.* (2021) Int J Mol Sci 22(22):12440. https://doi.org/10.3390/ijms222212440.

James G., Witten D., Hastie T., and Tibshirani R. (2021) An Introduction to Statistical Learning. Springer New York, New York, USA.

Meyer D., Dimitriadou E., Hornik K., Weingessel A. and Leisch F. (2021) e1071: Misc Functions of the Department of Statistics, Probability Theory Group (Formerly: E1071), TU Wien. Available at: https://CRAN.R-project.org/package=e1071.

Ståhle-Bäckdahl M., Sandstedt B., Bruce K., Lindahl A. *et al.* (1997) Lab Invest 76(5):717–728.

Tibshirani R. (1996) J R Stat Soc Series B Stat Methodol 58(1):267–288.

Widmer S., Seefried F. R., von Rohr P., Häfliger I. M. *et al.* (2021) Genet Sel Evol 53(1):57. https://doi.org/10.1186/s12711-021-00650-1.

400. Quantifying the functional conservation between human and pig using artificial neural networks

J. Li[1], T. Zhao[1], Z. Pan[1], H. Zhou[1], L. Fang[2] and H. Cheng[1*]

[1]*Department of Animal Science, University of California, Davis, One Shield Avenue, Davis, CA 95616, USA;* [2]*MRC Human Genetics Unit, Institute of Genetics and Cancer, University of Edinburgh, Edinburgh EH4 2XU, United Kingdom;* *qtlcheng@ucdavis.edu*

Abstract

Identifying functional conservation between human and pig genomes is an important challenge in pig model studies. To evaluate it, we predicted conservation scores through fitting a neural network model based on the integration of multi-tissue epigenetic and gene expression profiles. We then integrated the conservation score with chromatin state and human expression quantitative trait loci (eQTL). The average conservation score for promoters was higher than enhancers as expected. The conservation scores for human regions with eQTL were higher than the average level of the whole genome. In addition, eQTLs with higher conservative scores had a smaller effect size than those with lower scores. In conclusion, the model reflected the genome-wide functional conservation between human and pig, which can be easily extended to other species, and could be further utilized to reveal the conserved patterns linked to complex traits.

Introduction

The investigation of genome conservation not only reveals evolutionary evidence (Alföldi and Lindblad-Toh, 2013), but also helps with the integration of genetic findings across different species (Raymond *et al.*, 2020). The genomic conservation between human and domestic pig (*Sus scrofa*), a relevant biomedical model for certain human diseases, is of importance and interest. Previous studies investigated the functional conservation between human and pig based solely on the gene expression profile (Sjöstedt *et al.*, 2020) or the epigenomic profile (Pan *et al.*, 2021), but not all available functional data were integrated simultaneously to quantify the conservation. Kwon and Ernst (2021) developed a neural network model to study human-mouse functional conservation based on multi-omics information and showed that their method can be used to explain the phenotype-associated variation. In this study, we aim to implement a neural network model to evaluate genome-wide functional conservation between pig and human based on the integration of large-scale epigenetic data and gene expression data across multiple tissues.

Materials & methods

The overview of the model development is shown in Figure 1. To define pairs of human and pig regions for training and prediction, we first obtained the pairwise sequence alignment between the human genome (hg38) and the pig genome (Sscrofa11) from the UCSC Genome Browser (Navarro Gonzalez *et al.*, 2021). We then divided the alignment into non-overlapping 50-bp windows, which resulted in 38,961,932 aligned pairs. We randomly paired up human and pig regions included in the aligned pairs to get the same number of unaligned pairs.

For each aligned and unaligned pair, we obtained the corresponding human and pig functional features, including gene expression measured by RNA-seq, chromatin accessibility measured by Assay for Transposase-Accessible Chromatin (ATAC-seq), histone modifications measured by Chromatin Immunoprecipitation sequencing (ChIP-seq), and chromatin state annotations (ChromHMM). Human features were collected from ENCODE (The ENCODE Project Consortium, 2012) and Roadmap Epigenomics Project (Kundaje *et al.*, 2015), and pig features were from https://doi.org/10.6084/m9.figshare.13480425 (Pan *et al.*, 2021). In total, we collected 313 and 312 functional features for human and pig, respectively.

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_400 J. Li *et al.* 2022

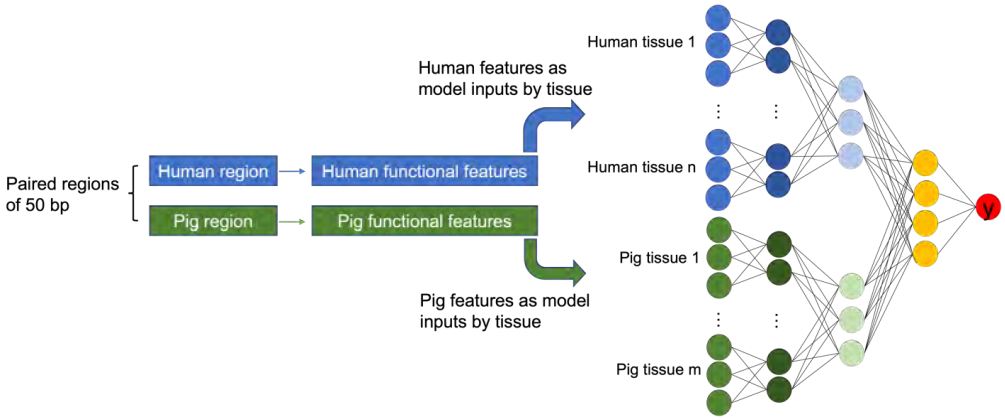


Figure 1. The schematic of the artificial neural network model. The functional features of human and pig in the paired (aligned or unaligned) regions are used as the model inputs by species and tissue, to predict the conservation score (y). When training the model, the response variables (labels) for aligned pairs are coded 1 while those for unaligned pairs are coded 0.

Aligned regions are conservative at the sequence level but not necessarily at the functional level. However, unaligned regions are much less likely to be conservative given different sequence information. Based on this, we presumed the aligned regions to be conservative (coded as 1) and unaligned ones to be unconservative (coded as 0). Thereby a pseudo-Siamese neural network model was trained based on the alignment (1 or 0) as the response variable and corresponding functional features as the predictor variables (Figure 1). To reflect the tissue-specific information in the model structure, the input and first two hidden layers were not fully connected. Instead, the input layer and the first hidden layer were connected by species and tissue, and the first and second hidden layers were connected by species. Although the response variable is binary when training the model, the output was transformed through a sigmoid activation function. Therefore, a continuous conservation score between 0 (not conserved) and 1 (highly conserved), could be predicted for any paired regions based on the functional features.

To predict the conservation score of human regions from even (or odd) chromosomes, and the corresponding paired pig regions, a neural network model was trained based on human odd (or even) chromosomes and paired pig regions. The area under the receiver operating characteristic (ROC) curve was assessed by comparing the predicted results to whether the pairs were aligned or not.

To demonstrate the potential application of the conservative score, we integrated the score with the expression quantitative trait loci (eQTL) in human. The significant loci associated with at least one gene expression from 49 tissues were obtained from the GTEx portal (https://gtexportal.org/home/datasets). We also compared the conservation score and the effect size of eQTLs, which was quantified as log allelic fold change (Mohammadi *et al.*, 2017).

Results & discussion

The predicted conservation scores showed that most of the aligned pairs (~95%) had a small conservation score (<0.2), while only ~0.4% of the aligned pairs had a conservation score greater than 0.8. This indicates that only a small part of the aligned genome between human and pig is functionally conserved. The area under the ROC curve was 0.77, suggesting a decent model prediction.

We next investigated average conservation score for chromatin states in each tissue (Figure 2). Similar results were found for human and pig. The highest average conservation score was found in the strongly active promoters, which was higher than all the scores derived from the enhancers. This result was consistent with the previous finding that promoters are more conservative than enhancers in general (Villar *et al.*, 2015).

The results of the application of the conservative score on human eQTLs are shown in Figure 3. The scores obtained by regions with eQTLs were higher than the average score of the whole genome in all tissues (Figure 3A), which suggests that eQTLs regions were more conservative than genomic regions outside eQTLs. We regressed the eQTL effect size on the conservation score and found that the slope was significantly smaller than 0 for all tissues except uterus and blood (Figure 3B). The negative association between effect size and

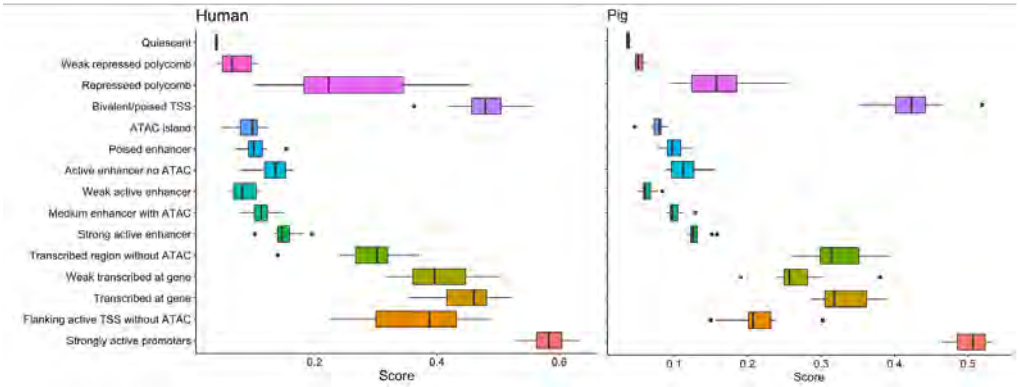


Figure 2. Average conservation scores for different chromatin states across tissues of human (12 tissues) and pig (14 tissues).

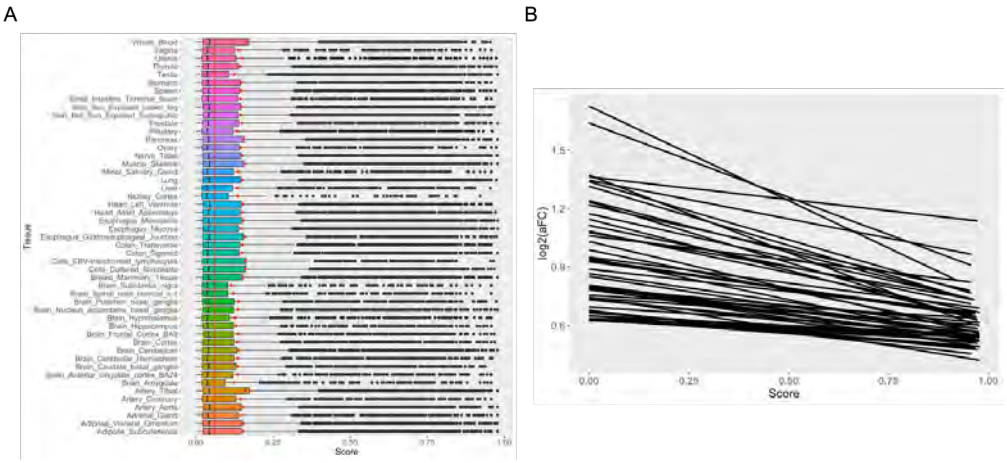


Figure 3. Relationship of conservation scores to human eQTLs. A: Conservation scores on eQTLs across 49 tissues. The red line represents the average score over the whole genome, and the red dot represents the average score of the eQTLs in the tissue; B: Regression lines of the conservation score and eQTL effect size (log_2(aFC) = log allelic fold change) for 49 tissues.

conservation of eQTLs agrees with the previous finding that genomic regions having large effect sizes are less likely to be conservative (Mohammadi *et al.*, 2017).

In conclusion, functional conservation between human and pig predicted by the neural network model had a decent accuracy, and the score reflected the conservation of different chromatin states and eQTLs. The method could be further used to reveal the conserved patterns linked to respective complex traits and diseases between human and pigs, facilitating the recycling information among species. The method can also be easily extended to other species.

References

Alföldi, J., and Lindblad-Toh, K. (2013) Genome Res. 23(7):1063-1068. https://doi.org/10.1101/gr.157503.113

Kundaje, A., Meuleman, W., Ernst, J., Bilenky, M., Yen, A. *et al.* (2015) Nature 518(7539):317-330. https://doi.org/10.1038/nature14248

Kwon, S.B., and Ernst, J. (2021) Nat. Commun. 12:2495. https://doi.org/10.1038/s41467-021-22653-8

Mohammadi, P., Castel, S.E., Brown, A.A., and Lappalainen, T. (2017) Genome Res. 27(11):1872-1884. https://doi.org/10.1101/gr.216747.116

Navarro Gonzalez, J., Zweig, A.S., Speir, M.L., Schmelter, D., Rosenbloom, K.R. *et al.* (2021) Nucleic Acids Res. 49(D1):D1046-D1057. https://doi.org/10.1093/nar/gkaa1070

Pan, Z., Yao, Y., Yin, H., Cai, Z., Wang, Y. *et al.* (2021) Nat. Commun. 12:5848. https://doi.org/10.1038/s41467-021-26153-7

Raymond, B., Yengo, L., Costilla, R., Schrooten, C., Bouwman, A.C. *et al.* (2020) PLoS Genet. 16(9):e1008780. https://doi.org/10.1371/journal.pgen.1008780

Sjöstedt, E., Zhong, W., Fagerberg, L., Karlsson, M., Mitsios, N. *et al.* (2020) Science 367(6482):eaay5947. https://doi.org/10.1126/science.aay5947

The ENCODE Project Consortium. (2012) Nature 489(7414):57-74. https://doi.org/10.1038/nature11247

Villar, D., Berthelot, C., Aldridge, S., Rayner, T.F., Lukk, M. *et al.* (2015) Cell 160(3):554-566. https://doi.org/10.1016/j.cell.2015.01.006

401. Using a U-net architecture to develop an automated ultrasound images analysis method in a rabbit breeding program

C. Herbert[1,2*], C. Ribas[2], K. Kashefifard[2,3], L. Flatrès-Grall[2,3,4] and G. Lenoir[2,3]

[1]HYCOLE, 59159 Marcoing, France; [2]GIS Nord Sélection, France; [3]AXIOM, 37310 Azay-Sur-Indre, France; [4]Chrysalide Loïc Flatrès-Grall Kerbastard, 29150 Dinéault, France; caroline.herbert@hycole.com

Abstract

Since 2014, HYCOLE has been using *in vivo* saddle by ultrasound measurements to estimate the animals' carcass yield. The saddle's area measurement was done manually by specialist operators and was time consuming. HYCOLE therefore searched a method to automatically determine the saddle area and its measurement. Deep learning, using a U-net architecture model, was the appropriate answer to this issue. The model was trained in two phases with 4,124 images over 4 years with an associated mask and then, a correlation validation phase was made with a set of 1,410 which were not used in the training phase and also measured manually by a specialist operator. Correlations between the automatically measured criterion (S23auto) and the manually measured criterion (S23ref) were calculated: the rank correlations with the selection criterion are between 0.80 and 0.94 and between 0.86 and 0.96 for the rank correlations between the areas. A genetic evaluation was carried out to validate the trait and automatic measure gave more accurate estimation with higher heritability. S23auto is validated to use in the genetic program and will be add as a data in the « operator » variable from the carcass yield genetic evaluation.

Introduction

Carcass yield is an important economic criterion in meat rabbit production. At HYCOLE, a rabbit breeding company, the improvement of carcass yield by selection is carried out by *in vivo* measurements on the candidates for selection. The surface area of the *Longissimus dorsi* muscle between the 2nd and 3rd lumbar vertebrae is measured on the HYCOLE X male line by ultrasound measurements at 65 and 66 days of age. This estimation of the saddle surface is done manually by an operator. This process is very time consuming and is a limiting factor for the development of the use of ultrasound images for other carcass quality criteria. HYCOLE therefore wanted to automate the estimation of these measurements and validate the method at the genetic level.

To estimate these measures, different approaches were investigated first by algorithmic means, using predefined models with markers coupled to pixel intensity gradients, and then by deep learning means, using neural network architectures such as the U-net (Ronneberger *et al.,* 2015). It is this last method that will be developed in this article.

Materials & methods

Animals. We used ultrasound images of males and females of the HYCOLE male X line for this study. These are animals in closed nucleus, born and raised on the same farm. Three datasets were created to validate the method: a first one composed of 4,124 images with a manual mask dating from 2018 to 2020 for model learning. A dataset of 1,410 images with a manual mask dating from 2018 to 2020 and not used in the training of the model to validate the model with correlations calculation from an expert operator, and a third set of 1,633 images over 2021 to estimate S23auto heritability, phenotypic and genetic correlations. The pedigree file was made with 88,968 animals born between 1998 and 2021 with no missing parents.

Ultrasound measurements. Rabbits were weighed at 65 days of age and then scanned at 65-66 days of age using an Exapad Mini (IMV Imaging; Angoulême, France) equipped with a linear probe of 7.5 MHz.

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_401, © C. Herbert *et al.* 2022

Measurements were taken at the level of the *Longissimus dorsi* muscle, between the 2nd and 3rd lumbar vertebrae on the right and left. Two images were recorded for each side of the muscle. Operators who recorded measurements are technicians, trained together and evaluated each year to validate the correlations between their measurements, by analyzing several same images by several operators, to not create biases over the years.

S23ref. The reference areas were estimated using a method developed by HYCOLE (Lenoir and Morien, 2015). All recorded images and images with estimated areas are stored as Joint Photographic Experts Group (JPEG) in grayscale.

S23auto. The automatic image analysis process was developed by HYCOLE using all historical data. First, image pre-processing was performed to ensure that the input images were square to obtain 464×464 pixels images and stored into Portable Network Graphics (PNG) format so that the max-pooling operation worked well. A deep learning model was then created based on the U-net architecture (Ronneberger *et al.,* 2015) and was run under python version 3.8.8 using Tensorflow (Raschka, 2015; Tensorflow.org). The U-net architecture is presented in Figure 1: each blue box corresponds to a multi-channel feature map. The number of channels is denoted on top of the box. The x-y-size is provided at the lower left edge of the box. White boxes represent copied feature maps. The arrows denote the different operations (Ronnerberger *et al.,* 2015).

For model learning phase, out of the 4,124 images (26 batches) available, 80% were randomly selected for the model training phase (3,300 images) which lasted 5 rounds. The remaining 20% (824 images) were used for the model verification phase. A morphological filter is finally applied before the surface calculation to refine the results and eliminate areas that would not belong to the saddle. A second validation of the model was carried out using a batch of 1,410 images (6 batches) for which we already had the measurements, and which had not been used during the learning phase. The model was validated using Spearman correlations by rank correlation between the obtained areas and with the selection criterion on R for each batch (R Core Team, 2018).

Genetic parameters estimation. Genetic parameters were estimated using ASREML 4.2 software (Gilmour *et al.*, 2009; Gilmour and Thompson, 2021) with the restricted maximum likelihood methodology

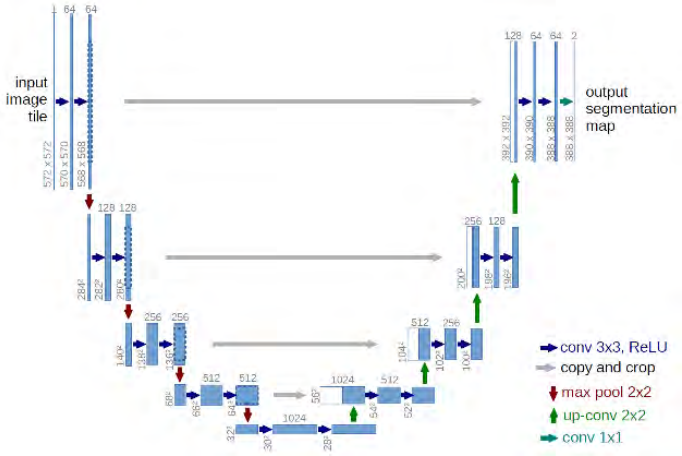


Figure 1. U-net architecture, example for 32×32 pixels in the lowest resolution (Ronnerberger *et al.*, 2015).

applied to a multitrait animal model. Fixed effects were tested at an α-risk of 5% using the Wald test on ASReml (Gilmour *et al.*, 2009; Gilmour and Thompson, 2021). Sex with male and female, batch of control, the ultrasound's operator and the age at weighing were significant for both S23 and were used as fixed effect. The additive animal genetic effect was also added as a random effect.

Results

S23auto development. The learning phase, which is carried out only once, lasted 18 hours. The time spent on a group then, is 10 min (170 animals with 4 photos/animal) from the pre-processing of the image to the opening of the file containing the estimated *Longissimus dorsi* area for each animal. Below, a raw image in Figure 2 and the desired mask during the learning phase and in Figure 3, a raw image and the mask actually produced by the model.

Spearman correlations ranged from 0.80 to 0.94 with the selection criterion and from 0.86 to 0.96 with the measured or estimated areas.

Genetic parameters. The genetic parameters are presented in Table 1. S23auto heritability is high but the standard error is also high and cannot be considered as a difference. Phenotypic and genetic correlations are significatively different from 1 even if genetic correlation is near 1 with correlation respectively at 0.87 and 0.90.

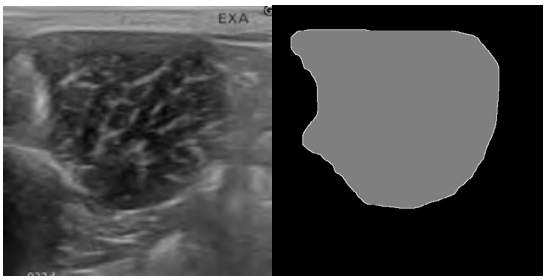


Figure 2. Raw image and desired mask during the model training phase.

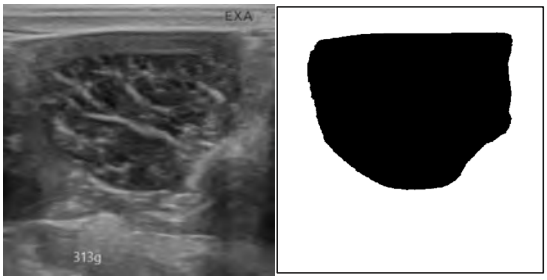


Figure 3. Raw image and mask made during the model validation phase.

Table 1. S23ref and S23auto genetic parameters (heritability on diagonal, phenotypic correlation down to the diagonal and genetic correlation up to the diagonal),

	Animals number	S23ref	S23auto
S23ref	52,472	0.34±0.02	0.90±0.08
S23auto	1,633	0.87±0.06	0.54±0.11

Discussion

S23Auto. Deep learning approach seems to be the best approach for HYCOLE. As Kvam *et al.* (2017) show, this is usable with a large number of images and sufficient image quality. Work on the images' quality will nevertheless have to be carried out in greater depth, as approximately 10% of the images do not pass the quality test. Some common points stand out on the images that do not pass the quality control. The idea is therefore to draw the images manually as before and to make the model relearn by adding these images in order to increase the variability of the model's images quality and to reduce the percentage that do not pass the test. Regarding the mask that came out of the model, we should have three colours on the obtained images: the pixels outside the saddle in black, the pixels inside the saddle in grey and the pixels around the saddle in white. This is not the case after the training phase of the model: the grey colour does not appear, the pixels inside the saddle and its contour are black and those outside the saddle are white. This is the case for all the images that emerge from the model. We do not tried to train the model with only two colours because what was important in our study was to be able to recover the surface of the saddle, which is the case here. Thus, we kept the model as it was.

Genetic validation of S23auto. S23auto and S23ref had very similar values. The carcass yield criteria automatically obtained is well correlated to the one manually developed by Lenoir and Morien (2015). The Spearman correlations show that the rank of the values obtained by manual measurements and by automatic measurements are well correlated and permits to select animals on S23Auto. We should gain in precision when adding new data. Phenotypic and genetic correlations between S23auto and S23ref are close to 1, both approaches are evaluated the same trait. S23Auto heritability is higher than S23Ref even if its standard deviation is also higher.

Conclusions

HYCOLE has developed a deep learning process to estimate the *Longissimus dorsi* area. The Spearman, phenotypic and genetic correlations show that S23Auto and S23ref are the same trait. Automatization permits to get the same *Longissimus dorsi* area or even more accuratly, while reducing a lot the time of measurements. This model has achieved the required objectives. The next work is to improve images quality to reduce the number of images rejected.

References

Gilmour A.R., Gogel B.J., Cullis B.R., Thompson R., 2009. ASReml User Guide Release 3.0. *VSN International Ltd. Hemel Hempstead. HP1 1ES. UK.*

Gilmour A.R., Thompson R., 2021. ASReml Update What's new in release 4.2 *VSN International Ltd. Hemel Hempstead. HP1 1ES. UK.*

Kvam J., Kongsro J., In vivo prediction of intramuscular fat using ultrasound and deep learning, Computers and Electronics in Agriculture, Volume 142, Part B, 2017, Pages 521-523, ISSN 0168-1699, https://doi.org/10.1016/j.compag.2017.11.020.

Lenoir G., Morien F., 2015. Intérêt de l'évaluation par échographie des caractéristiques de carcasse pour la sélection du lapin de chair, *16èmes Journées de la Recherche Cunicole, Le Mans, 24-25 Nov. 2015*, Pages 177-180.

Raschka S., 2015. Python Machine Learning Equation Reference. *Packt Publishing Ltd, Birmingham.*

R Core Team, 2018. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. Available from: https://www.R-project.org

Ronnerberger O., Fischer P., Brox T., 2015. U-net: Convolutional networks for biomedical image segmentation. In MICCAI, pages 234–241.

402. Machine learning and genetic improvement of animals and plants: where are we?

D. Gianola[1*], J. Crossa[2], O. Gonzalez-Recio[3] and G.J.M. Rosa[1]

[1]University of Wisconsin-Madison, WI 53706, USA; [2]CIMMYT, 56237 Texcoco, Mexico; [3]Instituto Nacional de Investigaciones Agrarias, INIA-CSIC, 28040 Madrid, Spain; gianola@ansci.wisc.edu

Abstract

We give a brief overview of knowledge collected in animal and plant breeding since the term 'machine learning' (ML) surfaced in these fields. ML is used to denote members of a broad 'hyper-class' of methods that have been employed, or are undergoing scrutiny, in genome-assisted animal improvement and phenomics. Our focus is on kernel methods and neural networks and on how machine learning can be employed to construct features or target phenotypes of relevance. Mainly, we summarize our experience with these procedures and give scant references for reasons of space.

Introduction

The borders between predictive inference, statistical learning and machine learning are soft. For instance, best linear unbiased prediction (BLUP) can be viewed as an inferential procedure (estimation of breeding value); as a parameter in a Bayesian setting); a prediction method of yet-to-be observed quantities; as a kernel regression method (kernel= genome or pedigree-derived similarity matrix), or as a single-layer neural network, with an identity activation function and a Gaussian penalty for regularization. The Bayesian alphabet represents linear regression methods where the distinct priors induce different regularization and, thus, prediction machines. If the problem is one of prediction or classification, performance must be evaluated via a replicated cross-validation such that extant uncertainty can be assessed.

It is unclear when 'novel' ML methods started being applied in these fields. A predictive GWAS (genome-wide association analysis) was proposed in 2007 by Gianola's group, using a naïve Bayes classifier for SNP set selection. Reproducing kernel Hilbert spaces (RKHS) regression for genomic prediction made its debut in 2006 and the first application to animal breeding was in 2008. The CIMMYT group led by Crossa extensively studied kernel methods for single and multi-environments, with genotype × environment interaction as an important target. Bayesian Regularized Gaussian neural networks (BNN) were suggested in 2011, and a grouped-neuron architecture (additive and dominance groups) with Monte Carlo sampling of the posterior distribution of connection strengths published in 2012. In 2018, Deep learning (DL) was applied to genomic prediction of human traits and by the CYMMIT group to many different setting and types of plants. ML is being actively investigated in Wisconsin by Rosa and Dorea in connection with leveraging data from sensors, images, spectroscopy and other digital tools to identify and monitor animals, spatially and temporally.

Our experience

Our opinion is that there is no universally best genome-based prediction machine. The performance of the various methods depends on many factors, including genomic structure, sample training and testing sizes, and algorithm (optimization vs sampling). Reinoso-Peláez *et al.* (2022) made 205 comparisons in 34 studies done in the last decade with plants and animals. They used a rank-based Thurstonian method to compare prediction models, as when comparing horses in races. Figure 1 shows that the methods that consistently ranked best in terms of a higher Pearson's correlation between predicted and observed phenotypes were RKHS, BNN, naive Bayes, Bayes A, and Bayes B. In general, RKHS and Bayesian methods seemed to outperform other procedures.

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_402, © D. Gianola *et al.* 2022

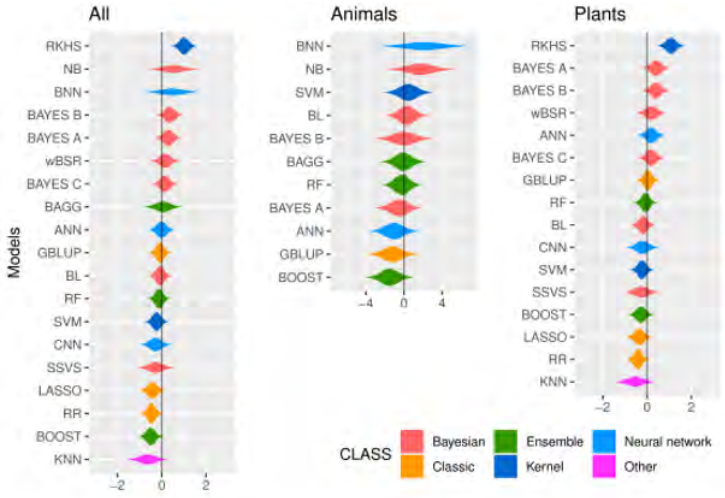


Figure 1. Rank of predictive performance (predictive correlation) of different methods in in 205 comparisons from 34 studies. (with permission of Reinoso-Peláez *et al.* 2022).

CIMMYT extended and applied RKHS methodology in many directions, including multi-kernel fitting, multiple-trait kernels, genotype × environment interaction kernels and deep kernels that, in some sense, are counterparts of deep learners. In multiple-kernel fitting, the basic idea is to create, say, genomic, epi-genomic, meta-genomic or environmental similarity matrices. For instance, such a matrix could represent atmospheric pressure similarity among sites, and another matrix could use some humidity or temperature (or both) index. Interaction kernels are created by Hadamard products. In a single-trait setting, training requires fitting a variance component model. In the multi-trait approaches, one must estimate several covariance matrices. Due to non-orthogonality of kernels, it is difficult to interpret and disentangle the relative importance of the sources of (co)variation.

Neural networks 'resurrected' in the late 80s and early 90s. The appeal of a neural network is provided by one out of many Kolmogorov's theorems. Any continuous function can be approximated by a nested sum of transformed linear combinations of nonlinear functions of inputs and connection strengths. All features of the architecture (number of nested sums, number of terms in the sum, rate coefficients, activation function and statistical distribution of 'residuals') must be learned. Methods of learning have been non-linear least squares for n>p problems, and for n<p situations, empirical Bayes methods and fully Bayesian approaches with Markov chain Monte Carlo sampling have been explored. The use of sampling, instead of optimization, led to improvement of performance (Perez-Rodriguez *et al.* 2012). Zhao *et al.* (2021) run multi-layered neural networks with priors as in the Bayesian alphabet and found a consistent improvement in predictive correlations with both simulated and real data. Figure 2 shows differences between a neural network with 10 layers and an implementation of BayesCπ. An inversion of Fisher's z-transform, implies that a difference in correlations of 0.67 and 0.65 (i.e. 0.02) translates into a difference between 0.81 and 0.73 in a linear scale. The predictive correlation, typically in the 0-1 space, is like a speedometer with an insensitive needle. These results encourage more research on neural networks, or even deep learning, together with Monte Carlo sampling in genome-enabled prediction.

Deep learning (DL), a multi-layered neural network with a new class of algorithms, entered into the picture recently. The CIMMYT group evaluated DL in many different settings (Montesinos-Lopez *et al.*, 2018). DL implementation is challenging. The DL methods do not seem competitive and stable when n<<p, and

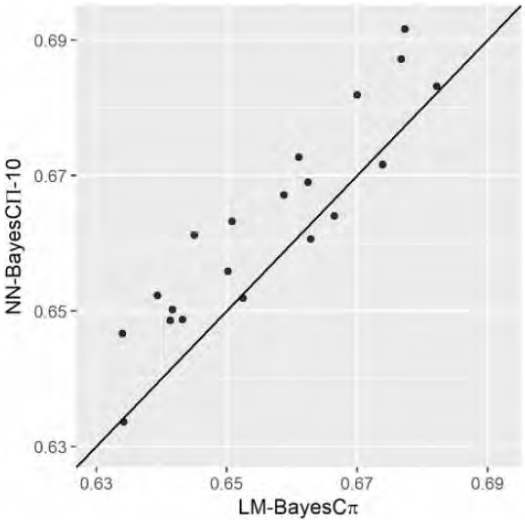


Figure 2. The prediction accuracy of NN-Bayes composed of 10 hidden nodes with BayesCπ priors (NN-BayesCπ-10) versus the linear model with BayesCπ-10 prior (LM-BayesCπ) for the soy dataset. 20 replicates for validation were applied. The diagonal line is used for reference such that a dot above the line represents a validation with higher accuracy for NN-Bayes (Zhao *et al.*, 2021).

there is not much evidence of utility for extracting biological insights from data. The convolutional neural network (CNN) DL topology seems outstanding when inputs are raw images. For this reason, CNNs are being applied in plant science to, e.g. root and shoot feature identification, leaf counting, detecting wheat spikes, and estimating plant morphology and developmental stages. The situation is similar in animal production.

The explosion of ML led to the emergence of 'livestock' prediction analytics. For example, monitoring feed efficiency at the animal level requires measuring individual feed intake, which is extremely difficult. Milk spectrum from mid-infrared spectroscopy has been used to predict milk traits, and it has also been explored as a predictor for other traits such as dry matter intake (DMI). Overall, the use of neural nets improved predictive performance. Image analysis and computer vision systems (CVS) are promising and the power of information has been harvested through DL algorithms. For example, CVS has been used for automated prediction of body weight and body measurements in fish. In a study in Nile tilapia, DL predicted body weight with a predictive R^2 of 0.96.

Conclusions

It seems that SOME machine learning procedures are SOMETIMES better but OFTEN worse than either BLUP or members of the Bayesian alphabet, as there is no universally best method. In genome-enabled prediction DL has not delivered clearly and consistently better results. However, ML and DL offer promise for targets derived from images, sounds, sensors and motion.

The gold standard is single-step genomic best linear unbiased prediction (ssGBLUP). A competing ML procedure should accommodate cross-sectional, longitudinal and multiple-trait settings, and should be as flexible and computable as the gold standard. Kernel methods may meet this challenge. Wisconsin has shown how a more general kernel can be used in lieu of a genomic relationship matrix in ssGBLUP.

CIMMYT proposed a Bayesian multi-trait kernel method: the Gaussian kernel outperformed GBLUP by 10.4, 17.5 and 16.4% in three datasets. There is work going on in extending the multi-trait kernel method to a ssGBLUP setting. The more flexible kernel methods may capture non-additive gene action and accommodate genotype-interactions and may (or may not) outperform the current gold standard.

Defining breeding objectives is crucial. Until buzz terms such as, e.g. eco-system, robustness, resilience, sustainability, etc., are defined precisely and unambiguous metrics are developed, researchers will navigate in a sea of vagueness. AI, ML and DNN may lead to the discovery of more relevant phenotypes. A salutary impact of ML and AI has been to help us take the naïve theory of quantitative genetics less seriously. In ML, heritability and genetic correlations cease to have an existential meaning and become transitional tools that help us go from past and current phenotypes to expected phenotypes.

Current emphasis on 'big data', 'massive computing' and 'visualization' perhaps will diminish basic science education. Students must understand the meaning of a probability distribution, and attain knowledge of experimental design, randomization or causality, before running algorithms (downloading is easy), that somebody else (perhaps a robot) has written. Strong theory and basic biology must continue being taught if animal breeding wishes to maintain a status as a science.

References

Montesinos-López, A., Montesinos-López, O.A., Gianola, D., Crossa, J., Hernández

Suárez, C.M. (2018)G3: Genes|Genomes|Genetics 8(12): 3813-3828.

Perez-Rodriguez, P., Gianola, D., Weigel, K.A., Rosa, G. J. M., Crossa, J. (2013) Journal of Animal Science 91: 3522-3531

Reinoso-Peláez, E. L., Gianola, D., y González-Recio, O. (2022). Genome-enabled prediction methods based on machine-learning, in Genomic prediction of complex traits. Springer.

Zhao, T., Fernando, R., Cheng, H. (2021) *G3 Genes|Genomes|Genetics* 11 (10), October 2021, jkab228.

403. Predicting pig digestibility coefficients with microbial and genomic data using machine learning prediction algorithms

C. Carillier-Jacquin[1*], V. Deru[1,2], L. Tusell[3], A. Bouquet[2,4], L. Jacquin[5] and H. Gilbert[1]

[1]GenPhySE, Université de Toulouse, INRAE, ENVT, Castanet Tolosan, France; [2]France Génétique Porc, 35651 Le Rheu, France; [3]Animal Breeding and Genetics Program, Institute of Agriculture and Food Research and Technology (IRTA), Barcelona, Spain; [4]IFIP-Institut du Porc, 35650 Le Rheu, France; [5]CIRAD, BIOS, UMR AGAP, Montpellier, France; celine.carillier-jacquin@inrae.fr

Abstract

Classical methods as genomic BLUP performs well for genomic prediction of polygenic trait, but does not consider interaction between genes or between genes and other information such as host genetic or microbial data. This study aims at comparing several methods including parametric and machine learning methods to predict digestive coefficient using genomic, microbial and both genomic and microbial information. Considering only microbial data led to the best prediction accuracies for digestive coefficients, whereas considering only genomic data performed worst. BLUP, RKHS and GSVM gave the best prediction accuracies except when combined genomic and microbial data was used. Combining microbial and genomic data did not improve prediction accuracies for all traits and methods considered in this study. Thus, considering microbial information is crucial to predict digestive efficiency and interactions between host genetic and faecal microbial information seem to be limited.

Introduction

Genetic evaluation as a sum of additive trait effects using genomic Best Linear Unbiased Predictor (BLUP) method, as defined by Meuwissen *et al.* (2001), is known to be more accurate than other methods for polygenic traits in animal genetics (Zhu *et al.*, 2021; Wang *et al.*, 2019). Nevertheless, some non-parametric machine learning methods allow considering interacting genes, or interaction between microbiota and host genetic and major effect of some gene or OTUs (Operational Taxonomic Unit). The objective of this study was to compare parametric approaches, and machine learning methods that can capture such interactions, in terms of prediction accuracy of digestive efficiency considering microbial data, genomic data, or both information. Indeed, in pigs, earlier studies showed that both host genetic and faecal microbial information contribute to digestive coefficients variability (Déru *et al.*, 2021a).

Materials & methods

Statistical analyses. Corrected phenotypes were predicted using six different methods: BLUP, Least Absolute Shrinkage and Selection Operator (LASSO), Elastic Net (EN), Reproducing Kernel Hilbert Space (RKHS), Support vector Machine using Linear (LSVM) and Gaussian (GSVM) kernel. BLUP (aka Ridge regression) and LSVM are linear methods, which can capture additive polygenic effects of SNP. The major difference between BLUP and LSVM lies in the loss function used for measuring the corrected phenotype prediction error i.e. squared euclidean loss versus epsilon-insensitive loss respectively. EN and LASSO are variable selection methods, which allow considering non-polygenic traits or more important effects of some OTUs. RKHS and GSVM methods are approaches using kernels, which allow considering additive effects and interactions between genes or OTUs. BLUP, LASSO and EN were performed using glmnet R package (Friedman *et al.* 2010), whereas RKHS and SVM were performed using KRMM (Jacquin *et al.*, 2016) and kernlab R packages (Karatzoglou *et al.*, 2004), respectively. Hyperparameters for RKHS and SVM methods were optimised by cross-validated grid-search as in Jacquin *et al.* (2016). For LASSO, the regularization parameter was optimised using the cv.glmnet function with its default values (Friedman *et al.*, 2010). For each trait, the six prediction methods were used with genomic data only, microbial

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_403, © C. Carillier-Jacquin *et al.* 2022

data only, and combined genomic and microbial information. Microbial data was included in the model as the centered and standardized log abundancies of each OTU (i.e. matrix of 2,399 OTUs for 1,082 animals). Genomic data was the matrix of 48,919 SNP for 1,082 animals and did not include any pedigree information. Combining genomic and microbial data was considered using the concatenation of the matrix of SNP and the matrix of OTUs (i.e. matrix of 51,318 columns (48,919 SNP + 2,399 OTUs) and 1,082 rows (animals)). Considering the largest amount of SNP compare to the amount of OTUs, combining genomic and microbial information could not be ideal for methods that did not reduce the number of variables. For this reason, we also tested combination of microbial data with genomic data reduced to 10,000 SNP selecting equidistant SNP.

Prediction accuracies. Cross validation was used to estimate prediction accuracy, by splitting randomly our population of 1,082 animals into two sets: a training set (722 animals, i.e. 2/3 of the total population) and a validation set (360 animals). Prediction accuracy was estimated for each trait and predicting method for the validation set, as the Pearson correlation between records corrected for diet and batch and predicted traits. Confidence intervals of the Pearson correlation were computed using cor.test function of psych R package (Revelle, 2021).

Data. Data used for 1,082 Large White male pigs feed with a conventional European diet or a high-fiber diet were available (Déru *et al.*, 2021b). Digestibility coefficients of energy (DC_E), and nitrogen (DC_N) were predicted from faecal sampling at 16 weeks of age using near infrared spectrometry (Déru *et al.*, 2021a). Microbial information was obtained from sequencing of the V3-V4 regions of the 16S rRNA contained in the same faecal samples as used for DC predictions. After data curation, the 2,399 operational taxonomic units (OTUs) were kept filtering out OTUs present in less than five samples and with an average abundance lower than 0.001%. Genomic data from the 70K SNP GeneSeek GGP Porcine chip was filtered according to minor allele frequency, call rate, call freq and Hardy-Weinberg equilibrium leading to 48,919 markers available for all pigs.

Results & discussion

Prediction accuracies and confidence intervals estimated in the validation set for DC of nitrogen and energy using the six prediction methods are presented in Table 1. Rankings of the six methods were similar for all DC. Among methods, BLUP, RKHS and GSVM gave better accuracies compared to other methods when considering genomic or microbial information separately. Selection variable methods used in this study (i.e. LASSO and EN) gave the worst prediction accuracies with separate information, and were best when combined microbial and all the genomic information was used. If some genes or OTUs had higher effects on trait than other ones, we expected that these methods outperformed BLUP. We can suppose that all SNP and OTUs seemed to contribute equally to digestive coefficients. LSVM did not performed as well as BLUP, probably due to loss function not adapted to our data. GSVM and RKHS did not outperform BLUP, which could be explained by only few interactions between host genes, between OTUs and between host genetic and microbial information. The better prediction accuracy obtained for BLUP and RKHS in this study is similar to results obtained in other species for a wide range of traits (Wang *et al.*, 2021; Zhu *et al.*, 2019). However, Maltecca *et al.* (2019) have reported similar results with LASSO and RKHS when using microbial data to predict pig growth and carcass traits.

Table 1. Prediction accuracies and confidence intervals for digestive coefficients of energy (E) and of nitrogen (N) with six prediction methods.[1]

		BLUP	LASSO	EN	RKHS	LSVM	GSVM
	N	0.38	0.29	0.29	0.38	0.33	0.36
Genomic		[0.3;0.46]	[0.2;0.38]	[0.2;0.38]	[0.29;0.46]	[0.24;0.42]	[0.27;0.45]
	E	0.43	0.42	0.44	0.47	0.40	0.45
		[0.35;0.51]	[0.34;0.5]	[0.36;0.52]	[0.39;0.54]	[0.31;0.48]	[0.36;0.52]
	N	0.67	0.60	0.60	0.67	0.58	0.65
Microbial		[0.61;0.72]	[0.53;0.66]	[0.54;0.66]	[0.61;0.72]	[0.51;0.64]	[0.59;0.70]
	E	0.69	0.63	0.64	0.69	0.62	0.67
		[0.63;0.74]	[0.57;0.69]	[0.57;0.69]	[0.63;0.73]	[0.56;0.68]	[0.61;0.72]
Genomic and	N	0.49	0.61	0.61	0.48	0.47	0.48
Microbial		[0.41; 0.56]	[0.54;0.67]	[0.54; 0.67]	[0.40;0.55]	[0.38; 0.54]	[0.40; 0.55]
	E	0.51	0.62	0.60	0.52	0.49	0.52
		[0.43;0.58]	[0.55; 0.67]	[0.53;0.66]	[0.45;0.59]	[0.41;0.56]	[0.45; 0.59]
Genomic (10k)	N	0.57	0.58	0.59	0.57	0.55	0.56
and Microbial		[0.51;0.64]	[0.52;0.65]	[0.53; 0.65]	[0.49; 0.63]	[0.49;0.62]	[0.52;0.65]
	E	0.58	0.60	0.59	0.59	0.56	0.59
		[0.43,0.58]	[0.55,0.67]	[0.53,0.66]	[0.52;0.65]	[0.41;0.56]	[0.45;0.59]

[1] BLUP = best linear unbiased predictor; EN = elastic net; GSVM = support vector machine using Gaussian kernel; LASSO = least absolute shrinkage and selection operator; LSVM = support vector machine using linear kernel; RKHS = reproducing kernel Hilbert space.

For all traits, considering only microbial data led to the best prediction accuracies, whereas considering only genomic data performed worst. The decrease in prediction accuracies when using only genomic information compared to using microbial data only can be explained by the large part of variance of DC traits explained by microbial information in these traits (Déru *et al.* 2021c). Compared to the use of microbial data only, Pearson correlations decreased when both microbial and genomic data were analysed jointly except for variable selection methods certainly due to much larger amount of genomic data (48,919 SNP) compared to microbial data (2,399 OTUs). In practice, LASSO and EN were able to select more OTUs than SNP in the model, thus prediction accuracies were close to the ones obtained with microbial data only. When reducing genomic data to 10,000 equidistant SNP, other methods (i.e. BLUP, RKHS and SVM) performed similarly to the selection variable methods. This confirms our hypothesis that the highest amount of SNP compared to OTUs did not allow methods to perform as well as using only microbial data, at least in the case of DC traits where microbial information seem to be preeminent on genomic information to explain the traits.

Comparing DC traits, Pearson correlations obtained for DC of nitrogen using only genomic data were lower to the one obtained for DC of energy, which is not the case when considering only microbial or combined data. This could be explained by a lower contribution of host genetic to DC_N variance, as reported also by Déru *et al.* (2021c) with different approaches.

Conclusions

We conclude that microbial information is crucial to predict digestive coefficients. Machine learning methods did not outperform BLUP, which suggest few interactions between SNP, between OTUs, and between microbial and host genetic information. For further study, it could be interesting to estimate prediction accuracies without filtering out rare OTUs, to evaluate a potential prediction improvement

when combing both data. The interest of random forest variable importance methods could also be used to explore genomic and microbial data combination.

Funding

This study was supported by the Feed-a-Gene European H2020 project. The authors acknowledge the breeding companies Axiom and Nucleus for providing animals and the UEPR phenotyping station staff in Le Rheu for animal raising and data recording.

References

Déru V., Bouquet A., Labussière E., Ganier P., Blanchet B. *et al.* (2021a) J Animal Breed Genet 138 :246-58. Déru V., Bouquet A., Zemb O., Blanchet B., De Almeida M.L. *et al.* (2021b) Preprint bioRxiv.

Déru V., Tiezzi F., Carillier-Jacquin C., Blanchet B., Cauquil L. *et al.* (2021c) Proc. of the 72nd EAAP, Davos, Switzerland.

Friedman J., Hastie T., and Tibshirani R. (2010) J. Stat. Softw. 33:1.

Jacquin L., Cao T.V., Ahmadi N. (2016) Front Genet, 7,145.

Karatzoglou A., Smola A., Hornik K., Zeileis A. (2004) J. Stat. Softw. 11, 1–20.

Maltecca C., Lu D., Schillebeeckx C., McNulty N.P., Schwab C., *et al.* (2019). Sci Rep 9, 6574.

Meuwissen T., Hayes B.J., Goddard M. (2001) Genetics 157: 1819–1829

Revelle W. (2021). psych: Procedures for Psychological, Psychometric, and Personality Research. Available at: https://CRAN.R-project.org/package=psych.

Wang X., Miao J., Chang T., Xia J., An B., *et al.* (2019) PLoS ONE 14(2).

Zhu S., Guo T., Yuan C., Liu J., Li J., *et al.* (2021) G3, 11(11).

404. (Quasi) Multi-task support vector regression for whole-genome prediction of carcass traits in commercial broilers

A.A.C. Alves[1*], A.F.A. Fernandes[2], V. Breen[2], R. Hawken[2] and G.J.M. Rosa[1]

[1]University of Wisconsin, Department of Animal and Dairy Sciences, 1675 Observatory Dr, Madison, WI 53705 USA; [2]Cobb-Vantress, Inc., 4703 US Highway 412 East, Siloam Springs, AR 72761, USA; carvalhoalve@wisc.edu

Abstract

This study investigates non-linear kernels for multi-trait genomic prediction using support vector regression (SVR) models. Phenotypic traits from broiler chickens purebred lines were used. The predictive ability for two carcass traits was assessed according to different single-trait (ST) and multi-trait (MT) models, as well as training strategies. The MT models used performance traits measured *in vivo* (Growth Trait – *Growth* and Feed Efficiency Trait – *FE*) as secondary traits. The proposed (Quasi) multi-task SVR (*QMTSVR*) outperformed the traditional *MTGBLUP* model, improving the predictive ability for the *post-mortem* traits when phenotypes for secondary traits were available. Additionally, the weighted *QMTSVR* accounting for the relative importance of different genomic regions provided a bias reduction on the predicted values.

Introduction

In animal breeding programs, many quantitative traits of economic importance are sex-specific, laborious, time-consuming, or expensive to measure. Hence, strategies that allow rapid, accurate, and resource-efficient predictions are of particular interest to fully explore the potential of genome-enabled selection. In this regard, the interest in multi-trait genomic prediction (MT-GP) for aggregating the complementary information of correlated traits has been largely increased in the last few years (Karaman *et al.* 2018, Gianola and Fernando 2020, Manzanilla-Pech *et al.* 2020).

MT-GP is particularly interesting for carcass traits as they generally require slaughtering the animals for assessing their phenotypic performance; therefore, exploring the information of secondary traits measured on live animals might benefit the genomic evaluation of *post-mortem* traits. In practical applications, the implementation of MT-GP models can be challenging as such methods are computationally intensive and require the knowledge of covariance parameters that are hard to estimate when many traits are analysed jointly. Additionally, most MT-GP models assume only additive effects for the genetic markers, while potential non-linear effects are ignored. In this study, we propose non-linear kernels for aggregating multi-trait genomic information in support vector regression models to overcome these problems.

Materials & methods

(Quasi) Multi-task support vector regression (QMTSVR). In standard support vector regression (SVR) models the ϵ-insensitive loss function is generally optimized by using quadratic programming. Considering the information from multiple correlated traits, the following data augmentation scheme can be proposed for this loss function:

$$arg\ min_{\alpha,\beta}\ \frac{1}{2}\begin{matrix}\boldsymbol{\alpha}^{(1)}-\boldsymbol{\beta}^{(1)\,\mathbf{T}}\\ \boldsymbol{\alpha}^{(2)}-\boldsymbol{\beta}^{(2)}\\ \vdots\\ \lfloor\boldsymbol{\alpha}^{(t)}-\boldsymbol{\beta}^{(t)}\rfloor\end{matrix}\mathbf{Q}\begin{matrix}\boldsymbol{\alpha}^{(1)}-\boldsymbol{\beta}^{(1)}\\ \boldsymbol{\alpha}^{(2)}-\boldsymbol{\beta}^{(2)}\\ \vdots\\ \lfloor\boldsymbol{\alpha}^{(t)}-\boldsymbol{\beta}^{(t)}\rfloor\end{matrix}+\varepsilon\begin{matrix}\boldsymbol{\alpha}^{(1)}-\boldsymbol{\beta}^{(1)\,\mathbf{T}}\\ \boldsymbol{\alpha}^{(2)}-\boldsymbol{\beta}^{(2)}\\ \vdots\\ \lfloor\boldsymbol{\alpha}^{(t)}-\boldsymbol{\beta}^{(t)}\rfloor\end{matrix}\mathbf{1_n}\otimes\mathbf{1_t}-\begin{matrix}\boldsymbol{\alpha}^{(1)}-\boldsymbol{\beta}^{(1)\,\mathbf{T}}\\ \boldsymbol{\alpha}^{(2)}-\boldsymbol{\beta}^{(2)}\\ \vdots\\ \lfloor\boldsymbol{\alpha}^{(t)}-\boldsymbol{\beta}^{(t)}\rfloor\end{matrix}\begin{matrix}\mathbf{y}^{*(1)}\\ \mathbf{y}^{*(2)}\\ \vdots\\ \lfloor\mathbf{y}^{*(t)}\rfloor\end{matrix}$$
$$subject\ to\ \begin{cases}0\le\ \alpha_i^{(t)},\beta_i^{(t)}\ \le C, i\ =\ 1,2,\dots,N\\ \quad[\boldsymbol{\alpha^{(t)}}-\boldsymbol{\beta^{(t)}}]^T\mathbf{1_n}=0\end{cases} \quad (1)$$

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_404, © A.A.C. Alves *et al.* 2022

in which t refers to the number of traits, $\alpha^{(t)}$ and $\beta^{(t)}$ are the vectors of positive Lagrange multipliers related to trait t, 1_t is a t-dimensional vector of ones, $\otimes$ is the Kronecker product, $\left[\mathbf{y}^{*(1)}, \mathbf{y}^{*(2)}, \ldots, \mathbf{y}^{*(t)}\right]^T$ are vectors of observed values for t target variables standardized to the same scale, and $\mathbf{Q}$ is a block symmetric matrix, partitioned into t^2 blocks of dimension $N \times N$. Each block matrix in $\mathbf{Q}$ is computed as $\mathbf{Q}_{mn} = \mathbf{Q}_{nm} = \rho_{mn} exp\left(-\theta_{mn}\mathbf{D}^2 p^{-1}\right)$, where ρ_{mn} is a constant between -1 and 1 for $m \neq n$ (m, n = 1, 2, …, t) or 1 otherwise, θ_{mn} is a block-specific bandwidth hyperparameter for traits m and n, $\mathbf{D}$ is an $N \times N$ Euclidean distance matrix (EDM), and p is an scalar representing the number of markers *loci*. Please notice that the regularization parameter (C) and epsilon value (ε) are assumed to be constant across traits, therefore this model can be considered as a (Quasi) Multi-task Support Vector Regression (*QMTSVR*). The block matrices in $\mathbf{Q}$ can be further expanded to account for the loci-specific variance in each trait combination by building block-specific weighted EDMs ($\mathbf{WD}_{mn}$); this method will be hereinafter termed as *QMTWSVR*. The weighting factors were obtained with a marker-wise genome-wide association study (*GWAS*), performed only on the training population for each trait and normalized by dividing the -log10(p-values) by their maximum value. For $m \neq n$ the weighting factors were found by multiplying the -log10(p-values) computed for the same marker on each trait-specific *GWAS* and rescaled to lie between 0 and 1. Further, in the *QMTWSVR* method, the constant p in $\mathbf{Q}_{mn}$ is replaced with $\sum_{l=1}^{p} w_l^{(mn)}$, in which $w_l^{(mn)}$ is the weighting factor for *locus l* considering the $N \times N$ block for traits m and n.

Phenotypic and genotypic data. The dataset used in this study was provided by Cobb-Vantress Inc. (Siloam Springs, AR) and comprised observations related to broiler chickens from purebred lines, bred in 20 overlapping mating groups. The phenotypic data included observations of two carcass traits (Carcass Trait 1 – *CT1* and Carcass Trait 2 – *CT2*) and two *in vivo* traits (Growth Trait – *Growth* and Feed Efficiency Trait – *FE*). All traits were adjusted for the fixed effects of contemporary group (GC) with at least 5 observations and sex. The number of genotyped animals with observations for *CT1* and *CT2* (the primary traits) was 8,749, being all males. All animals were measured also for *Growth*, whereas a subset of 3,703 animals had phenotypes available for *FE*. The genotypic data were obtained using a moderate-density Illumina (San Diego, CA) chip, containing 60k single nucleotide polymorphism (SNP) markers.

Training and testing populations. The predictive ability of the *SVR* and *QMTSVR* models was assessed by splitting the data into training and testing datasets according to the chronological information of the mating group. The animals from mating groups 1 to 19 (n=8,337) were used for fine-tuning and training the models, whereas the data related to mating group 20 (n=412) was held out as the testing set. Three different training strategies were used: *TS1*, *TS2*, and *TS3*. The *TS1* consisted of using solely the information available for the primary trait on the training set. For the *TS2*, the primary traits (*CT1* and *CT2*) were analysed jointly with a secondary trait in a bi-trait model (*CT1* × *Growth* and *CT2* × *FE*), considering that the observations for both the primary and secondary traits were missing in the testing group. The *TS3* was similar to *TS2*, with the exception that all available observations for the secondary trait were considered for training the models (including those from the mating group 20). The *SVR* models can be summarized according to the training strategies as *SVR-TS1*, *QMTSVR-TS2*, *QMTSVR-TS3*, *QMTWSVR-TS2*, and *QMTWSVR-TS3*. These models were benchmarked with ST and MT genomic best linear unbiased prediction models (*GBLUP* and *MTGBLUP*, respectively), following the same training strategies. The criteria adopted for assessing the methods' predictive ability were the prediction accuracy (*ACC*), root-mean-squared error (*RMSE*), mean absolute prediction error (*MSE*), and inflation factor (b; values closer to one are preferable). For finding an optimal combination within the hyperparameter multidimensional space we used a genetic algorithm (GA) to perform a heuristic search by minimizing the metric 1 – ACC^2 measured in the 19th mating group. All GA-optimized support vector regressions models were fitted using an under development package that wraps the *kernlab* library (Karatzoglou *et al.,* 2004) as backend. The *GBLUP* and *MTGBLUP* models were fitted using the BLUPF90 suite programs (Misztal *et al.*, 2018).

Results

The observed values for the predictive ability metrics varied remarkably according to the trait, model, and training strategy, ranging from 0.41 to 0.63 for *ACC*, 1.52 to 1.78 for *RMSE*, 1.23 to 1.41 for *MAE*, and 0.91 to 1.31 for *b*, considering the *CT1*, and from 0.51 to 0.55 for ACC, 0.26 to 0.27 for *RMSE*, 0.20 to 0.21 for MAE, and from 0.78 to 0.85 for *b*, considering the *CT2* trait (Figure 1). According to those metrics, the *GBLUP* was slightly better than the *SVR* and *WSVR* models for *CT1* (higher *ACC*, lower *RMSE* and *MAE*, and *b* closer to one), whereas for *CT2* the opposite was observed. In general, there was an evident improvement for all predictive ability metrics when using the MT models, especially comparing the *TS1* with *TS3* (Figure 1).

The *QMTSVR-TS3* provided the highest *ACC* for both traits, 0.63 and 0.55 for *CT1* and *CT2*, respectively, which correspond to a relative gain of about 22 and 3.6% compared to the *MTGBLUP-TS3*. Incorporating the prior information for the *loci*-specific variance on the *SVR* models did not improve the overall *ACC* even though this approach was notably successful in reducing the prediction errors in both traits, as evidenced by the lower *RMSE* and *MAE*. Further, weighting the kernel Q according to individual and pleiotropic effects was also important for attenuating the prediction bias, more notably for *CT1*, with the *b* coefficient reducing from 1.31 to 0.96, comparing *QMTSVR-TS3* with *QMTWSVR-TS3* (Figure 1). The necessary time for the GA to find the best hyperparameter configuration was on average 69.13±15.54 (minutes) and 298.91±44.24 (minutes) for the *SVR* and *QMTSVR* models, respectively. This was the time required for the GA to compare 750 specific hyperparameters combinations for each trait and training strategy. The (co)variance components on the *GBLUP* and *MTGBLUP* models were estimated via AIREML, with the algorithm taking on average 871.42±196.66 minutes to converge for the bi-trait analyses.

Discussion

Since early developments by Hernderson and Quaas (1976), multiple-trait analysis has gradually become the state-of-art methodology for the joint selection of many traits. This approach has a solid theoretical base

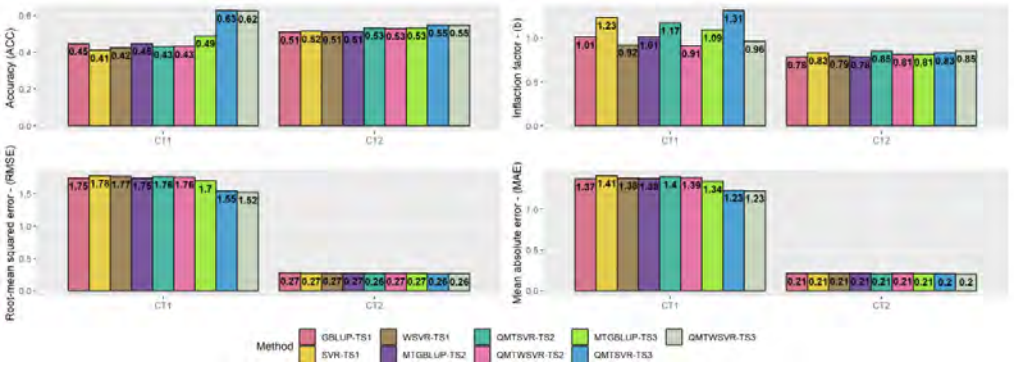


Figure 1. Predictive ability metrics for carcass traits measured in broiler chickens according to different single-trait and multi-trait genomic prediction methods. GBLUP-TS1: Genomic Best Linear Unbiased Predictor using TS1; MTGBLUP-TS2: multi-trait GBLUP considering that the information for the secondary trait (Growth or FE) was available only for the training population. MTGBLUP-TS3: MTGBLUP considering that the information for the secondary trait was available for both the training population and testing population. SVR-TS1: Support Vector Regression; MTQSVR-TS2 and MTQSVR-TS3 are multi-task SVR models with similar training and testing set schemes as in the MTGBLUP model. WSVR and QMTWSVR are similar models but with weighted kernels to reflect genetic architecture.

for pedigree-based selection and is becoming popular for GS over recent years, with different methods and strategies being proposed (Cheng *et al.,* 2018, Karaman *et al.* 2018, Gianola and Fernando 2020, Manzanilla-Pech *et al.* 2020). Here, we presented a simplified approach for expanding the SVR to the MT case under a genomic prediction context. This model shares some computational advantages of the GBLUP (e.g. both are based on $N \times N$ kernels) while being capable of exploring non-linear relationships within and between traits, without any strong assumption for the markers effects. Results showed that the *QMTSVR* improved the predictive ability for the primary traits (*CT1* and *CT2*) when compared to the traditional *MTGBLUP*, especially when the phenotypes for secondary traits were available on the testing population (Table 1). This is a reasonable scenario since broiler chickens are selected before the slaughtering process but after some routinely measured traits are already available, especially growth-related traits.

The burden of the fine-tuning process for the MT-QSVR method was alleviated by assuming some global hyperparameters for all traits analysed jointly, leading to $2 + t[t+1]/2$ hyperparameters to be tuned. A GA was then used for avoiding an extensive grid search. Results suggested that finding an appropriate hyperparameter configuration with the GA for the *QMTSVR* would take less computing time than estimating the *MTGBLUP* (co)variance components with an AIREML algorithm. Similar to what has been proposed for *GBLUP* (Karaman *et al.*, 2018) we further expanded the *QMTSVR* method to account for the relative importance of different genomic regions, by computing trait-specific and pleiotropic weighted Kernels. While this strategy did not imply higher prediction accuracy than the non-weighted counterpart, it was noticed a bias reduction with the *QMTWSVR* method (Figure 1). One must highlight that the weighting factors used in this study were based on a simple marker-wise *GWAS*; thus, further studies using weighting factors derived with more sophisticated methods may provide a better picture of the usefulness of the *QMTWSVR* model. Lastly, future research performing a comprehensive comparison of the proposed approach with other methods and different traits is encouraged.

References

Cheng H., Kizilkaya K., Zeng J., and D. Garrick (2018) Genetics 209: 89–103. https://doi.org/10.1534/genetics.118.300650

Gianola D., and Fernando R. (2020) Genetics 214:305–33. https://doi.org/10.1534/genetics.119.302934

Henderson C.R., and Quaas R.L., (1976) J. Anim. Sci. 43:1188–1197. https://doi.org/10.2527/jas1976.4361188x

Karatzoglou A., Smola A., Hornik K., and Zeileis A. (2004) J. Stat. Soft. 11:1–20. https://doi.org/10.18637/jss.v011.i09

Karaman E., Lund M.S., Anche M.T., Janss L., and Su G. (2018) G3 8(11):3549-3558. https://doi.org/10.1534/g3.118.200673

Manzanilla-Pech C.I.V., Gordo D., Difford G.F., Løvendahl P., and Lassen J. (2020) J. Dairy Sci. 103: 9195-9206. https://doi.org/10.3168/jds.2019-17857

Misztal I., Tsuruta S., Lourenco D.A.L., Masuda Y., Aguilar I., *et al.* (2018) Manual for BLUPF90 family programs. University of Georgia. Available at: http://nce.ads.uga.edu/wiki/doku.php?id=documentation

405. Genomic prediction of longitudinal body weights in pigs using a neural network

Z. Ni[1*], R.L Fernando[1], H. Yu[1], E.F. Knol[2] and J.C.M. Dekkers[1]

[1]Iowa State University, Ames, Iowa 50011, USA; [2]Topigs Norsvin, 6640 AA Beuningen, the Netherlands; zqni@iastate.edu

Abstract

Neural networks have drawn considerable interest for their use in genomic prediction in recent years. As a flexible universal approximator, nonlinearity can be easily accommodated in a neural network, allowing non-additive genetic effects and genotype by environment interactions to be captured. However, SNP effects must be formulated as additive effects for use in genetic improvement. We developed a neural network model for genomic prediction for body weight of pigs. The model consists of two parts: (1) additive effects of SNPs using genomic prediction; and (2) one layer of a feedforward neural network to predict longitudinal body weights based on the SNP effects. By decoupling the prediction of additive SNP effects from the prediction of phenotype, the model is expected to better predict body weight across populations and environments.

Introuction

Genomic prediction has been shown to improve genetic progress by allowing more accurate estimated breeding values (EBV) at an early age (Meuwissen *et al.*, 2001). However, the fundamental assumption of additive models for observed phenotypes, as applied in most genomic prediction models, remains debatable. Over the last decade, there have been multiple attempts to incorporate non-additive genetic effects and genotype by environment interactions (GxE) into genomic prediction. As a universal approximator, neural networks have been shown to improve the accuracy of phenotype prediction in several settings (Waldmann 2018; Gianola *et al.* 2011). Whether neural networks can be used for genomic prediction based on longitudinal trait data, such as body weight, which is a complex quantitative trait with moderate heritability, deserves investigation. Several studies have attempted to integrate non-linear mechanistic models into genetic evaluation of body weight in pigs (Doeschl-Wilson *et al.*, 2007). These mechanistic models usually require specific assumptions of the relationship between latent variables and phenotypes, which may not be valid across populations and environments. Neural networks don't require such assumptions. Against this background, the objective of this study was to develop an assumption-free neural network model for genomic prediction of longitudinal body weights in pigs.

Materials & methods

Neural network model for genomic prediction. Figure 1 illustrates the neural network model with 3 neurons that each consist of a pair of latent variables, b_{ni} and a_{ni}, that model the intercept and slope, respectively, of body weight against centered age. The latent variables are linked to the observed longitudinal body weight of pig_n using an one-layer feedforward neural network, with nonlinearity introduced through the activation function Relu (Stursa and Dolezel 2019). Similar to a random regression model (Jamrozik *et al.*, 1997), each b_{ni}/a_{ni} pair is linked to SNP genotypes using an additive whole-genome prediction model, as linear combinations of an intercept (μ_{ani}/μ_{bni}), a fixed contemporary group effect (CG_{bni}/CG_{ani}), random SNP effects for the slope and intercept ($\delta_{jbni}/\delta_{jani}$), and a residual ($\varepsilon_{bni}/\varepsilon_{ani}$). The number of b_{ni}/a_{ni} pairs is a tunable parameter. Increasing the number of pairs increases the expressive power of the model, but predictive ability may decrease if the size of the dataset is limited. The final predictions of body weight with age are obtained as a weighted sum of the predictions from each b_{ni}/a_{ni} pair, equivalent to an ensemble learning process (Hashem 1997).

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_405, © Z. Ni *et al.* 2022

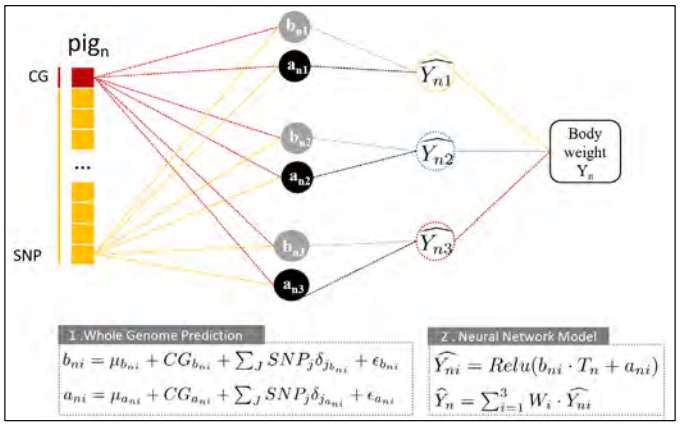


Figure 1. Structure of the neural network model for genomic prediction of body weights based on 3 neurons, each consisting of a pair of latent variables (g and t).

The neural network was built using the Flux.jl package (Innes *et al.* 2018) and trained for 5,000 iterations using the minibatch gradient descent method (Goodfellow *et al.*, 2016) and the ADAM optimizer (Kingma and Ba, 2014). A mean square error loss function that incorporates the L2 regularization (Friedman *et al.*, 2001) was used, which avoids overfitting, since the number of SNPs is typically much larger than the number of records. The model has 4 hyperparameters: the number of neurons, the L2 regularization coefficient λ, the size of the minibatch, and the learning rate of the ADAM optimizer. The first step in applying the model is tuning the hyperparameters, which was based on a grid search by training the model for a given set of hyperparameters and evaluating its prediction accuracy in validation data to identify the set with the highest accuracy.

Application. The model was applied to a dataset of daily body weight records from 79 to 197 days of age from 3,934 purebred boars in 97 contemporary groups. Evaluation of predictions was by forward cross-validation by using the first-born 2,000 boars as the training set and the remaining boars as the test set. To avoid information leakage, tuning of the hyperparameters was conducted by forward cross-validation within the 2,000 training animals, by fitting the model to the first 1000 of the 2,000 training set animals, and computing the accuracy of predictions using data from the second set of 1000 animals. For this purpose, the accuracy of predictions for a given set of hyperparameters was computed for each day of age based on the correlation of predictions with observed body weights adjusted for contemporary group effects for that day of age. Choice of the best set of hyperparameters was based on the average accuracy across days 103 to 153, which is when most pigs had records. The optimal set of hyperparameters was then used to train the model on the full training set of 2,000 and applied to forward-predict the test set. The accuracy of predictions in the test set was computed for each day of age as for tuning. The bias of predictions was computed for each day of age based on the regression coefficient of adjusted body weights on predictions and would be 1 if there's no bias.

Results

The grid search identified the optimal hyperparameters to be 5 neurons, a batch size of 1000, λ=240,000, and a learning rate of 1.0 e^{-5}. Example predictions in the test set at example days are in Figure 2, compared to predictions based on a hierarchical Bayesian model that fits the Gompertz function, similar to Cai *et al.* (2012). The cross-validation accuracy and bias of predictions of body weights for each day are shown in Figure 3. Results from the Bayesian hierarchical model are provided for comparison. The average accuracy

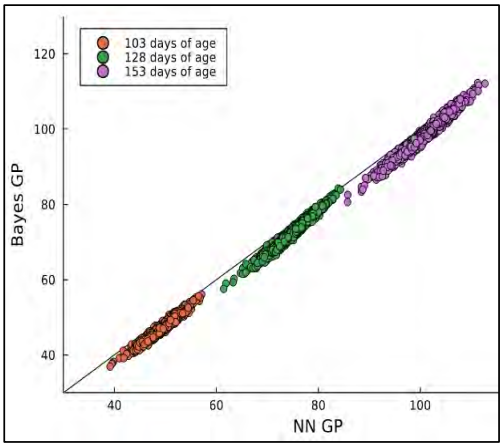


Figure 2. Predictions of body weight (kg) using a Neural Network Model (NN GP) versus the Bayesian method (Bayes GP) at 103, 128, and 153 days of age.

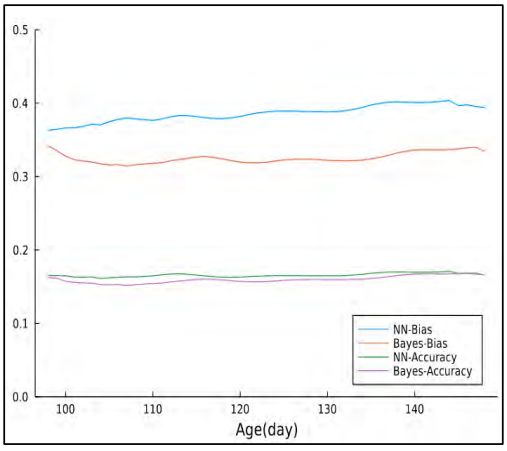


Figure 3. Comparison of bias and accuracy of the Neural Network Model (NN) and the Bayesian method (Bayes).

across days was slightly higher for the neural network model (0.17) than for the Bayesian hierarchical model (0.16). The neural network model also had a lower bias of 0.39 compared to 0.33 for the Bayesian hierarchical model.

Discussion

We found that the neural network model obtained slightly higher accuracy and less bias than the Bayesian hierarchical model that was based on the Gompertz function as the underlying biological model. This may be because the Gompertz function models the S-shaped body weight curve over a longer lifespan, which contains much nonlinearity, compared to the fairly linear part of the curve that our data were obtained in. Nevertheless, the results suggest that the neural network can predict longitudinal data well when trained using properly tuned hyperparameters in the framework of random regression. And indeed,

Sonoda and Murata (2017) previously proved that a neural network using Relu functions could act as a universal approximator. However, because a neural network model is intrinsically a data-driven method, extrapolating predictions, e.g. to other ages, remains problematic. This is less of an issue for mechanistic models. Interpretability of a neural network model is also lacking under most circumstances. However, an appealing feature of the neural network model is that its computing time is much shorter than that of the Bayesian hierarchical model: less than 5 minutes on a Geforce Nvidia 3090 GPU given a specific set of hyperparameters, compared to at least 3.5 hours on an Intel Core i7-9800X CPU for the Bayesian model. But the current grid search requires numerous trials over different combinations of hyperparameters, which is time consuming. More sophisticated hyperparameter tuning methods such as heuristic algorithms need to be investigated to reduce tuning time.

In conclusion, the expressive power of a neural network can be incorporated into genomic prediction using a decoupling 2-step design, enabling the model to capture non-additive genetic effects and G×E, hence increasing the accuracy of predictions and reducing their bias. The predictive performance of the neural network model was comparable to a Bayesian hierarchical model that relies on a mechanistic biological model and associated assumptions. The neural network avoids the possible negative effects of improper assumptions of mechanistic models on predictions, especially under specific scenarios, which supports the broader application of predictions across different scenarios.

Acknowledgements

This research was funded by USDA-NIFA grant # 2020-67015-31031.

References

Cai, W., Kaiser, M. S., and Dekkers, J. C. M. (2012) J. Animal Sci. 90(1):127-141. https://doi.org/10.2527/jas.2011-4293

Doeschl-Wilson, A. B., P. W. Knap, B. P. Kinghorn, and H. A. M. Van der Steen. (2007). Animal 1(4): 489–99. https://doi.org/10.1017/S1751731107691848

Friedman, J., Hastie, T., Tibshirani, R., (2001). The elements of statistical learning. Springer Press, New York, USA.

Gianola, D., Okut, H., Weigel, K.A., Rosa, G.J., (2011). BMC Genet 12(87). https://doi.org/10.1186/1471-2156-12-87

Goodfellow, I., Bengio, Y., Courville, A., (2016). Deep Learning. MIT Press, Cambridge, USA.

Hashem S., (1997). Neural Networks.10(4):599-614. https://doi.org/10.1016/S0893-6080(96)00098-6

Innes, M., Saba, E., Fischer, K., Gandhi, D., Rudilosso, M.C., *et al.*, (2018). Available at: http://arxiv.org/abs/1811.01457

Jamrozik, J., L. R. Schaeffer, and J. C. Dekkers., (1997). J. Dairy Sci. 80(6):1217–26. https://doi.org/10.3168/jds.S0022-0302(97)76050-8

Kingma, D.P., Ba, J., (2014). Available at: http://arxiv.org/abs/1412.6980

Meuwissen, T.H., Hayes, B.J., Goddard, M.E., (2001). Genetics (157):1819–1829. https://doi.org/10.1093/genetics/157.4.1819

Sonoda, S., Murata, N., (2017). Appl. Comput. Harmon. Anal. (43):233–268. https://doi.org/10.1016/j.acha.2015.12.005

Stursa, D., Dolezel, P., 2019. (2019). Proc. of 22nd International Conference on Process Control, Štrbské Pleso, Slovakia.

Waldmann, P., (2018). Genet. Sel. Evol.(50):70. https://doi.org/10.1186/s12711-018-0439-1

Yu, H., van Milgen, J., Knol, E., Fernando, R., and Dekkers, J. (2022) WCGALP 2022

406. Compressing microbiota information using an autoencoder to predict growth traits in swine

C. Maltecca[1*], J. Jiang[1], J. Fix[3], C. Schwab[3,4] C. Shull[4] and F. Tiezzi[1,2]

[1]Department of Animal Science, North Carolina State University, Raleigh, NC 27695, USA; [2]Department of Agriculture, Food, Environment and Forestry, (DAGRI) University of Florence, 50144 Firenze, Italy; [3]Acuity Ag Solutions, Carlyle, IL 62231, USA; [4]The Maschhoffs LLC, Carlyle, IL 62231, USA; cmaltec@ncsu.edu

Abstract

Microbial composition represents a promising tool in precision farming. In the current paper, we evaluated the power of fecal microbial composition to predict growth performance across swine farming systems. We used different dimensionality reductions to select microbial features to be included in the predictive models, ranging from random selection to selection based on association to data compression using a sparse autoencoder. We compared these methods with a model including all information available. We found that microbial information can predict performances for growth and fat deposition. We found that in most cases, the use of all microbial information resulted in the highest predicting performance regardless of the trait or the populations used for training and prediction. Our results suggest that including all available microbial information might be the best option when using it to predict performance.

Introduction

Previous research used microbiota information to predict livestock performance (Maltecca *et al.*, 2019). In principle, microbiota abundance counts are akin to genomic information in format. Yet, they present peculiar analysis challenges related to their compositional nature and the fact that they are temporally and spatially dependent (Maltecca *et al.*, 2020). This is particularly relevant when transferring microbiota predictions from different settings or experimental designs. Autoencoders are a class neural network used, among other things, to learn an efficient reduced representation of the input data (Pasolli *et al.*, 2016). They are often proposed as a way to denoise data and extract salient descriptors. This paper compares autoencoders vs raw counts as a preliminary step in predicting swine growth parameters using microbiota information in nucleus and terminal systems.

Materials & methods

Experimental design and data collection. Phenotypic records were obtained from a commercial and a nucleus farm operated by The Maschhoffs LLC (now Acuity, Carlyle, IL, USA). The data spanned two connected systems: a Duroc nucleus purebred population (NU) and a terminal commercial crossbred population (TE), both sired by 28 Duroc founding boars. The NU population consisted of 769 Duroc individuals. Individuals were raised in a fixed-time system. Individuals entered performance test at an average of 88 days of age and were taken off-test at an average of 178 days of age. The TE population consisted of 1,141 individuals generated by crossing the Duroc sires with two commercial sow lines. Crossbred commercial individuals were raised in a fixed-weight testing system and harvested at an average weight of 98.8 kg and 97.90 kg for the males and females. In both systems, a contemporary group was defined as the group of animals entering a given facility at the same time. Microbiota composition from all pigs was obtained at off-test. 16S rRNA gene sequencing and quality control of the data were conducted as described by Lu and colleagues (2018). 3,001 OTUs were available for further analyses. Additional details can be found in Maltecca and colleagues (2021).

Phenotypic information. Growth traits were collected in the two populations according to the standards of the respective production systems. Phenotypes for NU individuals were obtained at off-test while at

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_406, © C. Maltecca *et al.* 2022

market weight for TE individuals. Backfat (BF), average daily gain (ADG), and loin depth (LD) were used for the subsequent analyses.

Training and testing sets. Solutions (either OTUs or compressed representations) were obtained for each trait in one system serving as the training set. These solutions were then used to predict performance in the opposite system serving as the prediction set (i.e. NU to predict TE or TE to predict NU).

Models. Solutions were obtained fitting microbial information with a Bayesian Ridge Regression model (BRR). A preliminary model was fitted to correct for any systematic effects in the data using the BGLR package in R (Pérez and Campos, 2014). The models for NU and TE were similar. In TE, models for each trait included the effects of sire, contemporary group, maternal genetic line, and pen. In the NU population, the models included the effects of sires, contemporary group, and biological litter. Adjusted phenotypes were subsequently used for the microbiota analyses. All microbial models were fitted as implemented by the Python Scikit Learn (Pedregosa et al., 2011). The predictive performance of each model was evaluated through Pearson's correlation coefficient, between microbial predicted values (MPV) and observed phenotypes, as well as root means squared error (RMSE).

Autoencoder. Autoencoders are similar to PCA (equivalent under linear transformation) and have two portions: an encoder and a decoder. The first condense the data to a reduced representation state (bottleneck, or latent space). The second takes the output of the bottleneck to recreate the input. For the predictive part of the model, only the bottleneck is retained and used as input in the subsequent analysis. Based on a preliminary grid search analysis (data not shown), we defined an encoder with three hidden layers, the first with 1,500 inputs, the second with the same number of inputs as OTUs (3,001), and the third (the bottleneck layer) with 400 inputs. A rectified linear unit (ReLU) was used as activator function. The autoencoder was fitted using the Python package Keras (Chollet *et al.*, 2015). The output of the models at the bottleneck was used as a feature vector in the BRR autoencoder model (AUm). These models were compared to models with all features included (ALLm), and models where 400 features were selected based on association of OTUs (MWASm) or randomly (RANm). MPVs in the prediction set were obtained as $X_{ip}m_{it}$ where X_i is the feature matrix of the correspondent model *i*, m_i is the vector of solutions for features of the correspondent model *i,* and t and p correspond to the training or prediction sets. Features were only selected from association (MWASm) in one popopulation, and used to predict the opposite.

Results

Prediction accuracies for all the models traits combinations are reported in Table 1. Predictions were highest for fat deposition and lowest for loin depth for all scenarios. This is in agreement with previous studies (Bergamaschi *et al.*, 2020). Predictions from nucleus training to terminal testing set were higher in most cases than the opposite, except for loin depth that was predicted better from terminal to nucleus (0.08 vs 0.24 in the full model), for the ALLm model, accuracies ranged from 0.08 (NU_TE LD) to 0.36 (NU_TE BF). Models including features selected through association were the second-best performing ones with Pearson's correlations ranging from 0.10 (LD NU_TE) to 0.34 (BF NU_TE).

In most cases, MWASm had less accuracy than ALLm except for LD predicted from NU to TE (0.12 vs 0.08). The autoencoder model was, on average, the third performing one, and it only outperformed the MWASm for only one trait (LD), for which it was better than ALLm for NU_TE (0.10 vs 0.08) and performed the same as MWASm. The scenario with 400 randomly chosen featured was the one that performed worst in all cases, as expected, with correlations ranging from (0.01 to 0.10). Results for RMSE are reported in Table 2. These, for the most part, were in agreement with accuracies. Interestingly the AUm models had, on average, a lower RMSE than the MWASm models. In all cases, ALLm models were the ones with the lowest RMSE.

Table 1. Correlation coefficients, (*r),* between microbial predicted values (MPVs) and phenotypes, for backfat (BF), average daily gain (ADG), and loin depth (LD), for the OTU model (ALLm), autoencoder model (AUm), association OUT model (MWASm) and random OTU (RANm), when predicting from nucleus to terminal (NU_TE) and from terminal to nucleus (TE_NU).

	ALLm *r(MPVp,Yp)*		**Aum** *r(MPVp,Yp)*		**MWASm** *r(MPVp,Yp)*		**RANm** *r(MPVp,Yp)*	
	NU_TE	TE_NU	NU_TE	TE_NU	NU_TE	TE_NU	NU_TE	TE_NU
ADG	0.32	0.27	0.29	0.26	0.31	0.25	0.10	0.05
BF	0.36	0.29	0.31	0.30	0.34	0.31	0.08	0.03
LD	0.08	0.24	0.10	0.23	0.10	0.22	0.02	0.01

Table 2. Mean squared error (RMSE) in prediction, for backfat (BF), average daily gain (ADG), and loin depth (LD), for the OTU model (ALLm), autoencoder model (AUm), association OUT model (MWASm), and random OTU (RANm), when predicting from nucleus to terminal (NU_TE) and from terminal to nucleus (TE_NU).

	ALLm R*MSE*		**AUm** R*MSE*		**MWASm** R*MSE*		**RANm** R*MSE*	
	NU_TE	TE_NU	NU_TE	TE_NU	NU_TE	TE_NU	NU_TE	TE_NU
ADG	3.15	3.35	3.97	4.19	3.85	4.23	6.91	6.87
BF	4.32	5.12	6.11	6.25	5.03	6.14	9.12	9.85
LD	1.82	1.41	2.83	2.18	2.15	2.14	4.13	5.09

Discussion

In the current paper, we evaluated the power of fecal microbial composition to foretell growth performance across farming systems. We took advantage of an experimental design where the microbial composition of purebred or terminal crosses progeny of 28 Duroc sires was collected during the performance test. We then used this information to investigate how microbial information could be used to predict phenotypes in the nucleus or terminal populations using different methods. In our research, prediction accuracy from TE to NU was higher than the opposite for most traits. Possible reasons for this could be related to a more considerable environmental and genetic variability in commercial settings leading to a broader microbial representation in TE, as well as the difference in sample size in the training populations. Prediction accuracies were larger than zero for all the traits and sizable for most of them, particularly for fat deposition and growth. The use of an autoencoder did not improve the stability of predictions across systems, and the model with all features included was the best performing in all cases. This, in addition to the added computational costs of the model, suggest that including all available microbial information might be the best option when using it to predict performance.

References

Bergamaschi, M., Tiezzi, F., Howard, J., Huang, Y.J., Gray, et al (2020). Gut microbiome composition differences among breeds impact feed efficiency in swine. Microbiome 8, 110. https://doi.org/10.1186/s40168-020-00888-9

Chollet, F., *et al.*, 2015). Keras. GitHub. Retrieved from https://github.com/fchollet/keras

Maltecca, C., Bergamaschi, M., Tiezzi, F., (2020). The interaction between microbiome and pig efficiency: A review. J. Anim. Breed. Genet.137, 4–13. https://doi.org/10.1111/jbg.12443

Maltecca, C., Dunn, R., He, Y., McNulty, N.P., Schillebeeckx, *et al.*, (2021). Microbial composition differs between production systems and is associated with growth performance and carcass quality in pigs. Anim. Microbiome 3, 57.

Maltecca, C., Lu, D., Schillebeeckx, C., McNulty, N.P., Schwab, *et al.*, (2019). Predicting growth and carcass traits in swine using microbiome data and machine learning algorithms. Sci. Rep. 9, 6574.

Pasolli, E., Truong, D.T., Malik, F., Waldron, L., Segata, N., (2016). Machine Learning Meta-analysis of Large Metagenomic Datasets: Tools and Biological Insights. PLoS Comput. Biol. 12, e1004977. https://doi.org/10.1371/journal.pcbi.1004977

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., *et al.*, (2011). Scikit-learn: Machine Learning in Python. J. Mach. Learn. Res. 12, 2825–2830.

Pérez, P., Campos, G. de los, (2014). Genome-Wide Regression and Prediction with the BGLR Statistical Package. Genetics 198, 483–495. https://doi.org/10.1534/genetics.114.164442

Part 22

Society – genetics in the tropics

407. Providing genetics for the dairy industry in the tropics – the Brazilian Dairy Gir breeding program

J.C.C. Panetto[1,7*], M.V.G.B. Silva[1], M.A. Machado[1], R.S. Verneque[1], A.R. Fernandes[2,3], C.H.C. Machado[2], M.F. Martins[1], D.R.L.R. Faza[1], C.A.V. Borges[1], H.T. Ventura[4], M.A. Pereira[4], J.C. Oliveira[2], L.A. Glatzl Junior[5], A.O. Garcia[6] and F.D. Leandro[7]

[1]Brazilian Agricultural Research Corporation (EMBRAPA), Rua Eugenio do Nascimento, 610, 36038-330 Juiz de Fora, MG, Brazil; [2]Brazilian Association of Dairy Gir Breeders, Praça Vicentino Rodrigues da Cunha, 110, 38022-330 Uberaba, MG, Brazil; [3]Associated Colleges of Uberaba (FAZU), Av. do Tutuna, 720, 38061-500 Uberaba, MG, Brazil; [4]Brazilian Association of Zebu Breeders (ABCZ), Praça Vicentino Rodrigues da Cunha, 110, 38022-330 Uberaba, MG, Brazil; [5]Federal University of Juiz de Fora (UFJF), Rua Jose Lourenço Kelmer, s/n, 36036-900 Juiz de Fora, MG, Brazil; [6]Federal University of Viçosa (UFV), Av. Peter Henry Rolfs, s/n, 36570-900 Viçosa, MG, Brazil; [7]Jose do Rosario Vellano University (UNIFENAS), Rodovia MG 179, Km 0, 37132-440 Alfenas, MG, Brazil; joao.panetto@embrapa.br

Abstract

Milk production is usually challenging under tropical environments. Heat stress and/or recurrent parasite infestations frequently cause cattle's reduced fitness and productivity. As a way to overcome this challenge, the Gir indicine cattle breed has increasingly been used in crossbred schemes to constitute milk production herds in Brazil and other countries in tropical regions. The Brazilian Dairy Gir Breeding Program was established in 1985, based on a partnership between the Brazilian Agricultural Research Corporation and the Brazilian Association of Dairy Gir Breeders. It was the first breeding program outlined with a progeny test for the improvement of an indicine breed for milk production in the world. Several technical innovations were employed along the time. Now, producers can access genomic predictions for thousands of Gir males and females. Those predictions are used mainly by Brazilian breeders, but also increasingly in other tropical countries from South America, Central America and Asia.

Introduction

India is the world's largest milk producer, responsible for more than 20% of the global production, followed by the United States of America, China, Pakistan and Brazil. A significant part of the global milk production occurs under tropical conditions. European cattle breeds are very successful under temperate or subtropical climates, but struggle to keep good production and reproduction performances in tropical regions, especially because of heat stress and/or recurrent parasite infestations. Thus, better adapted cattle genetics is needed to constitute tropical dairy herds, especially in some parts of Asia, Africa, Central and South Americas.

The Gir indicine breed has been increasingly used in successful crossbreeding schemes to compose dairy herds in the tropics. The initial idea of such crossbreeding schemes was to combine the productivity from one European breed, mainly Holstein, to the rusticity from one indicine breed, mainly Gir. Some breeders in Brazil, though, started selecting purebred Gir for milk production during the last century, with the idea of improving the performance of crossbred production herds. Those breeders later created the Brazilian Association of Dairy Gir Breeders (ABCGIL, from the initials in Portuguese), and stablished a partnership with the Brazilian Agricultural Research Corporation (EMBRAPA), to initiate the Brazilian Dairy Gir Breeding Program (PNMGL, from the initials in Portuguese), back in the 1980s. It was the first breeding program in the world outlined with a progeny test for the improvement of an indicine breed aiming milk production. The objective of this paper was to describe a very successful public-private partnership in Brazil, in which EMBRAPA provides the technology and detachment needed for reliable genetic evaluations,

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_407, © J.C.C. Panetto *et al.* 2022

serving the breeders with information to improve their herds, while breeders provide the data allowing EMBRAPA and partners to continue developing relevant research projects.

Materials & methods

The PNMGL started in 1985 by selecting a group of purebred Gir sires, based on the data available at that time. Semen from those sires were distributed to small holder dairy producers, being used in purebred Gir, Holstein/Gir crossbreds of various genetic compositions or even purebred Holstein cows. EMBRAPA's and ABCGIL's personnel started making systematic visits to those herds in order to keep reproductive and productive records of the resulting female progeny. This progeny test scheme continues until the present day, with a new group of sires being included every year. Together with the purebred Gir genealogical and productive data, recorded by the Brazilian Association of Zebu Breeders (ABCZ, from its Portuguese initials), and the composite Girolando breed genealogical and productive data, recorded by the Brazilian Girolando Association, the progeny test composes the basis for the Brazilian Dairy Gir genetic evaluations. The Dairy Gir Sire Summary published in 2021 included a list of more than 500 bulls, evaluated for up to 26 different traits.

Traits recorded. In the beginning of the program, only milk yield, milk fat and age at first calving were recorded. In 1994, the program started recording 18 linear conformation and handling traits (*body traits*: rump height, heart girth, body length, rump length, pin width, hip width and rump angle; *legs and feet traits*: foot angle, rear legs side view and rear legs rear view; *mammary system traits*: fore udder attachment, rear udder height, rear udder width, udder depth, teat length and teat diameter; *handling traits*: milking ease and temperament). In 1999, milk protein, lactose, total solids and somatic cell count started being recorded. From 2001 biological samples from sires and production cows started being collected for DNA extraction and storage. Some markers of interest, related to milk properties started being analysed and published from 2006 and analyses for genetic diseases were included in 2013. The beta-casein genotype started being published in 2016, allowing some breeders to select their herds for the production of the A2A2 milk. Even though a number of traits are already included and many others still need to be properly recorded, the main selection objective of Dairy Gir breeders in Brazil still is 305 days milk yield.

Methods applied. First sire groups were formed by breeder's choices, based on the information they had from the genetic evaluations of parents, or the milk yield of the dam of the young bull. Along the time, the technical coordination followed the results of each group of sires, while implementing various innovations and rules to the program and the choice of sires to enter the progeny test. The animal model was implemented in 1995, replacing the formerly used sire model. One important rule was included to the group of sires selected in 2008, in which only young bulls from the top 10% dams of each herd, in terms of milk yield Expected Breeding Values ($EBV_{milk\ yield}$), should participate on the progeny test. The last breaking point of the program was the implementation of the genomic selection to the young bulls, that began in the group of sires selected in 2016, after structuring of the Gir genotyped reference population (Boison *et al.*, 2017). From this point, breeders had more accurate information to choose the best young animals within their herds and the program was able to select the most promising animals, amongst all breeders, to compose the groups of sires.

Results

Average milk yield of participating herds has considerably grown since the beginning of the breeding program. The phenotypic trend for this trait in the population can be observed through illustration in Figure 1.

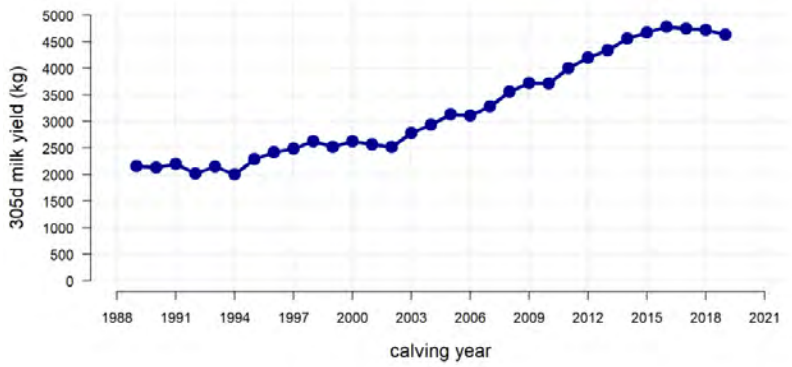


Figure 1. Phenotypic trend for 305 days milk yield since the beginning of the Brazilian Dairy Gir breeding program.

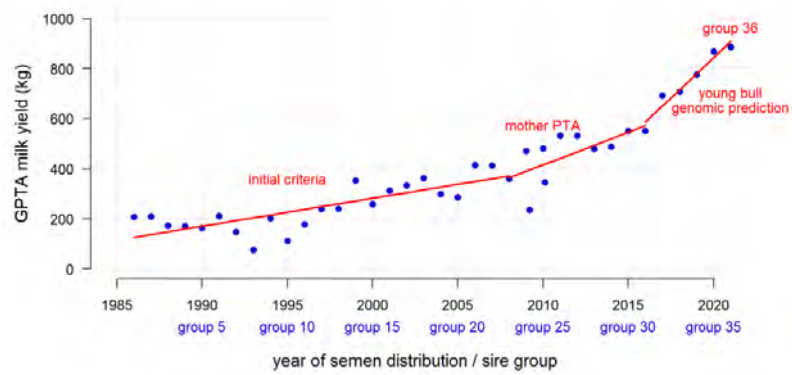


Figure 2. Genetic trend of the groups of sires since the beginning of the PNMGL: effect of change of criteria for the inclusion of young sires.

The impact of two actions taken by que program coordination on the genetic evolution for milk yield can be observed by the change in the slope of the genetic trend indicated by the average genomic values estimated for the groups of sires according to the year they have been included in progeny test (Figure 2).

In Figure 2, the blue dots indicate average genomic predicted transmitting abilities for milk yield of each group of sires. From the year 1985 to 2008 a small rate of genetic evolution can be observed (initial criteria). For this period, the milk yield of the dam was the main criteria used by the breeders to choose young sires to be included in the progeny test. For the second period (2008 to 2016) the mother of the young bull should be among the top 10% dams within the herd. This new rule, as expected, resulted in an increased genetic evolution in the population. For the most recent period observed (2016 to 2021) the increase in the rate of genetic evolution was even bigger. The use of genomic values for the selection of young bulls to be included in the progeny test allowed breeders to include more animals as candidates, by genotyping all born males at a young age. Thus, with better accuracy of genetic predictions and higher selection intensity, the rate of genetic evolution resulted in the program having its best Dairy Gir sires included in recent groups.

Discussion

Despite higher input costs and unfavourable weather conditions, Brazil's milk output is likely to rise by 1.0% year-on-year due to increasing dairy cattle numbers, increased herd sizes, improved genetics, and high yields in large-scale farms (FAO, 2021). Such conclusion from FAO's technical report is in accordance to the results observed in the Brazilian Dairy Gir breeding program, partially presented here. Genetic composition of dairy cattle herds in Brazil and other tropical countries include Gir and Holstein as the main breeds. Girolando is a composite breed formed by these two breeds, largely used for milk production and continually including improved sires and dams from both, Gir and Holstein breeds.

Genetic improvement of Holstein, Girolando and Gir populations have great impact on the milk production and dairy industry in the tropics. Considering the advances recently achieved by the Brazilian Dairy Gir Breeding Program, we can assume that newly obtained sires, and also genomic selected dams, have the potential to cause great improvements for the dairy herds in Brazil and other tropical countries, if the best animals are properly used for reproduction. Actually, exportation of Dairy Gir genetics from Brazil to tropical regions is clearly growing in recent years (unpublished data). Main countries importing Dairy Gir from Brazil are Colombia, Ecuador and Costa Rica. Even India, which is the country of origin of the Gir breed, can greatly benefit from improved Dairy Gir from Brazil.

Research groups should work as closely as possible the production environment, benefiting from the possibilities of obtaining field data and stimulating improved data annotation and collection. As an outcome of this approximation, innovation and technological solutions can be provided, resulting in quality and productivity advances to the sector.

Acknowledgements

The authors received financial support from EMBRAPA (Brazilian Agricultural Research Corporation); CNPq: INCT-CA (National Council for Scientific and Technological Development: Animal Science National Institute); FAPEMIG (Minas Gerais Research Foundation); and CAPES (Coordination for the Improvement of Higher Education Personnel). The authors would like to thank the ABCGIL (Brazilian Association of the Dairy Gir Breeders) for coordinating all the progeny test and pre-selection processes in the field and financial support to the program and ABCZ (Brazilian Association of Zebu Breeders) and Girolando breeders association for providing important additional data.

References

Boison S.A., Utsunomiya A.T.H., Santos D.J.A., Neves, H.H.R., Carvalheiro, R. *et al.* (2017) J. Dairy Sci. 100(7):5479-5490. Available at: http://doi.org/10.3168/jds.2016-11811

Food and Agriculture Organization of the United Nations (FAO), Dairy Market Review 2021. Available at: www.fao.org/3/cb5635en/cb5635en.pdf

Panetto J.C.C., Silva M.V.G.B., Verneque R.S., Machado, M.A., Fernandes A.R. *et al.* (2021). Programa Nacional de Melhoramento do Gir Leiteiro: sumário brasileiro de touros. Embrapa, Juiz de Fora - Brazil. Av. at: https://www.embrapa.br/gado-de-leite/publicacoes

408. Modelling phenotypic plasticity relative to a temperature humidity index as an indicator of adaptability in beef cattle

W. Shaffer[1], J.A.H. Moreno[2], N.M. Bello[3], R. Weaber[1], J. Bormann[1] and M. Rolf[1*]

[1]Kansas State University, 232 Weber Hall, Manhattan, KS 66506, USA; [2]University of Georgia, 425 River Road, Athens, GA 30602, USA; [3]The Ohio State University, 2029 Fyffe Court, Columbus, OH 43210, USA; megrolf@ksu.edu

Abstract

Genetic selection for phenotypic plasticity may be a solution to develop sires suitable for a variety of climates. Phenotypic plasticity can be modelled with random regression, where the additive genetic intercept represents production in a baseline environment and the slope of a first order model represents the environmental sensitivity and target of selection for phenotypic plasticity. Dry matter intake (DMI) was modelled as a function of a temperature humidity index (THI). The mean phenotypic trajectory point estimate was negative, indicating DMI decreases as THI increases. However, there was enough additive genetic slope variation for selection to move the population phenotypic slope towards zero and phenotypic plasticity. Selection for plasticity would likely be advantageous for DMI, because additive genetic correlations for DMI between divergent levels of THI (70 to 85) fell as low as 0.50. This indicates selection of sires using data from different environments may be inaccurate.

Introduction

Climate change is a serious issue contributing to increasing average global temperatures and greater variability in climate conditions (Nardone *et al.*, 2010; IPCC, 2007). Heat stress is known to negatively affect cattle (Fuquay, 1981; Nardone *et al.*, 2010) and can be measured by temperature humidity indices (THI). Thus, it is likely prudent to select for cattle adaptable to heat stress conditions.

Phenotypic plasticity represents the degree of sensitivity to an environmental variable and is characterized by stable performance across the environmental continuum (West-Eberhard, 2008). The sensitivity, or degree of change, of additive genetic potential as a function of the environment implies genetic-by-environment interaction (G×E) effects. The degree of environmental sensitivity, or the magnitude of G×E effects, can be characterized by the slope of a linear reaction norm. Individual animal reaction norms capturing the additive genetic component of phenotypic plasticity can be modelled via random regression. Thus, the objectives of this study were to characterize the genetic parameters for dry matter intake (DMI) with respect to THI using random regression models and the extent of G×E in a feedlot population of admixed steers.

Materials & methods

Population and phenotype collection. The study design, ration, and pre-processing details are described in Ahlberg et al. (2019). In total, seven cohorts of animals, comprised of 830 admixed steers, were on trial during all seasons from May 2014 to May 2018. All procedures involving the use of animals were approved by the Institutional Animal Care and Use Committee at Oklahoma State University (protocol AG13-18).

Phenotypes were collected for 70 days after an initial acclimation period. An Insentec system (Hokofarm Group, Marknesse, the Netherlands) was utilized to collect individual feed intake events, which were then summed to obtain daily DMI values. The THI was calculated using a livestock-specific equation (NOAA,

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_408, © W. Shaffer *et al.* 2022

1976; Bohmanova *et al.*, 2007), where T_{db} is the dry-bulb temperature and RH is the percent relative humidity.

$$THI = (1.8 \times T_{db} + 32) - (0.55 - 0.55 \times RH / 100) \times (1.8 \times T_{db} - 26) \quad (1)$$

Temperature and humidity data were obtained from the Oklahoma Mesonet Stillwater tower in five-minute intervals. The THI was then calculated according to Equation 1 for each five-minute interval and daily DMI observations were paired with the maximum THI observed for the day to capture the effects of heat stress on feed intake. The THI values were subsequently centred around thermoneutral conditions by subtracting 70 from the THI value. Observations with a centred THI value that was negative were removed from the analysis as the analysis focused primarily on heat stressed conditions.

Statistical model and other computations. Random regression animal models with permanent environment effects were utilized to evaluate the influence of the THI on predictions of estimated breeding values (EBV). The following random regression model was utilized for DMI

$$y_{ijk} = \eta + f_i + \beta_1 x_{ijk} + \beta_2 v_{ijk} + a_j + b_j x_{ijk} + g_j + h_j x_{ijk} + e_{ijk} \quad (2)$$

where y_{ijk} is a DMI observation for an individual, η is the intercept, f_i is the effect of the i^{th} group, β_1 is the effect of a one unit increase of THI on DMI, x_{ijk} is the centred THI value associated with a DMI observation, β_2 is the effect of a one kg increase in initial body weight on DMI, v_{ijk} is the body weight measurement, in kg, associated with a DMI observation, a_j is the additive genetic intercept of animal j, b_j is the additive genetic slope for animal j, g_j is the PE intercept for animal j, h_j is the PE slope for animal j, and e_{ijk} is the residual for a given DMI observation. Notably, the model included heterogeneity of residual variances by group, a covariance between the additive genetic intercept and slope, a covariance between the PE intercept and slope, and a between-group variance. Location (fixed) parameters included the population intercept and slope for the THI covariate and the covariate for initial body weight. Dispersion parameters (random) included the group effect, additive genetic intercept and slope for the THI covariate, PE intercept and slope for the THI covariate, and the residual effect. Analyses were conducted in a Bayesian framework with improper, uninformative priors and posteriors for all parameters were sampled.

Additive genetic variances for various levels of THI and additive genetic covariances and correlations between each pair of THI values were computed as follows:

$$\sigma_{u|x,u|x'} = \text{cov}(a_j + b_j x, a_j + b_j x') = \sigma_a^2 + xx'\sigma_b^2 + (x + x')\,\sigma_{ab} \quad (3)$$

where $\sigma_{u|x,u|x'}$ represents the additive genetic covariance for DMI between THI values x and x'.

Results

Parameter point estimates from the random regression models are in Table 1. The DMI intercept was positive at 5.79 kg DMI, but the population slope was negative, with a decrease of 0.046 kg DMI/ 1 unit THI. The DMI additive genetic intercept point estimate was 2.01 kg DMI^2, the additive genetic slope variance was 0.0078 (kg DMI / 1 unit $THI)^2$, and the additive genetic covariance between the intercept and slope was -0.10 kg DMI^2 / 1 unit THI. Permanent environment variance components were at least one magnitude smaller than their corresponding additive genetic counterparts. Residual variances by group ranged from 2.07 to 3.22 kg DMI^2. The genetic correlation between the additive genetic intercept and slope was negative at -0.66.

Table 1. Posterior median point estimates for model parameters, including lower and upper bounds of 95% highest posterior density credible intervals for daily dry matter intake (DMI) measured as kg. Slope parameters represent the DMI response to the temperature humidity index covariate.

Parameter	Estimate	Lower	Upper
η	5.79	4.88	6.66
β_1	-0.046	-0.053	-0.039
σ_f^2	0.85	0.10	3.92
σ_a^2	2.01	1.54	2.49
σ_{ab}	-0.10	-0.13	-0.07
σ_b^2	0.0078	0.0061	0.0096
σ_g^2	0.27	0.00	0.59
σ_{gh}	-0.0066	-0.03	0.00
σ_h^2	0.0007	0.0000	0.0020
$\sigma_{e_2}^2$	3.221	1.86	4.71
$\sigma_{e_6}^2$	2.231	2.13	2.33
r_{ab}	-0.78	-0.86	-0.69

[1] The largest and lowest residual variance point estimates were for groups 2 and 6, represented by the subscript. Residual variance point estimates for groups 1, 3-5, and 7 were between 2.23 and 3.22.

Point estimates for the DMI additive genetic correlations between different levels of THI are in Figure 1. The genetic correlations between different levels of THI ranged from near one at close intervals of THI to 0.5 as pairwise values of THI diverged.

Discussion

The population mean phenotypic trajectory for daily DMI decreased on average, equating to a mean loss of nearly 0.5 kg DMI per day for every increase of 10 units of THI. Compounded over many days and many head of cattle, this may equate to a substantial decrease in performance in both DMI and average daily gain (Santana *et al.*, 2014). Fortunately, there is sufficient genetic variation with respect to the environmental sensitivity (i.e. the additive genetic slope) for selection to move the mean phenotypic slope towards zero and thus increase phenotypic plasticity. Using empirical rule, 95% of EBVs would be expected to be within the interval of -0.177 to 0.177 kg DMI / 1 unit THI. Given the population phenotypic mean trajectory is

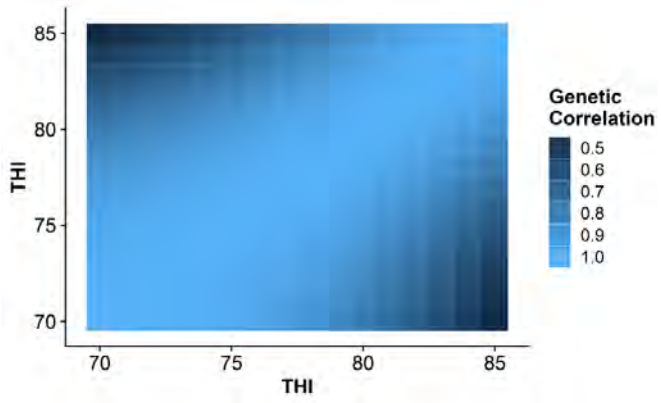


Figure 1. Genetic correlations for dry matter intake between different values of the temperature humidity index (THI).

-0.046 kg DMI / 1 unit THI, this indicates selection can move the population to be more phenotypically plastic on average.

The negative additive genetic correlation between the additive genetic intercept and additive genetic slope indicates selection to increase DMI in non-heat stressed conditions will decrease the slope on average, which is not favourable for phenotypic plasticity given the negative population trajectory. These results agree with the majority of G×E studies using the same THI function, such as Aguilar *et al.* (2009), where similar unfavourable relationships between plasticity and production exist. However, Sungkhapreecha *et al.* (2021) have reported a favourable additive genetic correlation between selection for milk production in non-heat stressed environments and environmental sensitivity to the THI in a Holstein × Thai *Bos indicus* cross and Bradford *et al.* (2016) demonstrated a favourable correlation between the intercept and slope for maternal weaning and yearling weight respective to THI in Angus animals. This may indicate inclusion of *Bos indicus* alleles could shift production under heat stressed scenarios to be favourable.

Given the low genetic correlations for DMI between diverse values of THI, sire EBVs would likely reranked based on the environment they are utilized in. Failure to account for these G×E effects would lead to decreased accuracy in cross-environment selection, thus decreasing the rate of genetic progress. Given the correlations for DMI between different levels of THI are low, selection to increase phenotypic plasticity is advisable.

References

Aguilar I., Misztal, I., and Tsuruta S. (2009) J. Dairy Sci. 92(11)5702-5711. https://doi.org/10.3168/jds.2008-1928

Ahlberg C.M., Allwardt A., Broocks A., Bruno K., Taylor A., *et al.* (2019) J. Anim. Sci. 97(12):4770-4782. https://doi.org/10.1093/jas/skz354

Bohmanova J., Misztal I. and Cole J.B. (2007) J. Diary Sci. 90(4)1947-1956. https://doi.org/10.3168/jds.2006-513

Bradford H.L., Fragomeni B.O., Bertrand J.K., Lourenco D.A.L. and Misztal I. (2016). J. Anim. Sci. 94(10):4143-4150. https://doi.org/10.2527/jas.2016-0707

Fuquay J.W. (1981) J. Anim. Sci. 52(1):164-174. https://doi.org/10.2527/jas1981.521164x

Intergovernmental Panel on Climate Change (2007). The intergovernmental panel on climate change 4th assessment report. Jackson Inst., Univ. College, London, UK.

Nardone A., Ronchi B., Lacetera N., Ranieri M.S., and Bernabucci U. (2010) Livest. Sci. 130(1-3):57-69. https://doi.org/10.1016/j.livsci.2010.02.011

National Oceanic and Atmospheric Administration. 1976. Livestock hot weather stress. US Dept. Commerce, Natl. Weather Serv. Central Reg., Reg. Operations Manual Lett. C-31-76. US Govt. Printing Office, Washington, DC, USA.

Santana M.H.A., Oliveira Jr. G.A., Gomes R.C., Silva S.L., Leme P.R., *et al.* (2014) Livest. Sci. 167:80-85. https://doi.org/10.1016/j.livsci.2014.06.002

Sungkhapreecha P., Misztal I., Hidalgo J., Steyn Y., Buaban S., *et al.* (2021) J. Dairy Sci. 104(12):12703-12712. https://doi.org/10.3168/jds.2021-20151

West-Eberhard M.J. (2008). Phenotypic Plasticity. In: Jørgensen, S.E. and Fath B. D. (eds) Encyclopaedia of Ecology. Elsevier, Amsterdam, the Netherlands, pp. 2701-2707.

409. Differentiated genomic regions in the Afrikaner and Brahman cattle of South Africa reveal parallel adaptive mechanisms

S. Mdyogolo[1,2], M.D. MacNeil[1,3,4], F.W.C. Neser[3], M.M. Scholtz[1,3] and M.L. Makgahlela[1,3*]

[1]ARC-Animal Production Institute, Private Bag X2, Irene, 0062, South Africa; [2]Department of Life and Consumer Sciences, University of South Africa, Florida, South Africa; [3]Department of Animal, Wildlife and Grassland Sciences, University of the Free State, P.O. Box 339, Bloemfontein, 9300, South Africa; [4]Delta G, 145 Ice Cave Rd., Miles City, MT 59301, USA; mmakgahlela@arc.agric.za

Abstract

Driven by the diversity of the Afrikaner and Brahman cattle, regarding their diverged origins, different migration routes to South Africa and performance under harsh environmental conditions; this study was designed to investigate genomic differentiated regions that may be evidence of parallel adaptive mechanisms. Across breed, edited genotype data included 83,306 SNPs common for 309 Afrikaner and 523 Brahman animals. Estimation of differentiation and signatures of selection explored two statistical complementary tests i.e. pairwise F_{ST} and FLK. Average F_{ST} was 0.318, showing moderate population structure. Regions of differentiation were located on Chromosomes 1, 6 and 21 with 105 putative candidate genes (65 annotated and 40 unannotated) using F_{ST}, and Chromosomes 4, 6, 11, 12 and 21 with a combined 4,697 candidate genes using FLK method. Shared regions were on Chromosomes 6 and 21, with underlying genes (e.g. SEM4B and SEC11A) responsible for compensatory growth and mammary involution during dry period and environmental heat stress.

Introduction

Different historical origins and divergent routes into South Africa (SA) characterize the Afrikaner and Brahman cattle breeds. The Afrikaner cattle are Sanga type (*Bos taurus africanus*), typically small to medium frame size and have a unique evolutionary adaptation to harsh climatic conditions and the Brahman cattle are Zebu type (*Bos indicus*) medium framed cattle, developed from the Nellore, Gir, Indo-Brazil, Guzerat and Krishna Valley cattle, which themselves originated from Pakistan, India and Brazil (Hartati *et al.*, 2015). The Afrikaner cattle have low susceptibility to endemic diseases, tolerance to trypanosomiasis, and great adaptability to extensive grazing in the arid areas of SA and survive on low quality forage (Scholtz *et al.*, 2016). The Brahman breed is heat tolerant, has resistance against internal and external parasites and can survive on low quality forage. This breed has desirable traits that separate it from other indigenous breeds, such as maintenance of a balance between low birth weight and higher weaning weight and has good mothering abilities. However, the Brahman cattle lacks tolerance against trypanosomiasis (Mattioli *et al.*, 2000).

The Afrikaner and Brahman cattle differ in historical origins and development, as also shown by their phenotypic disparities. However, harsh SA climatic conditions, characterized by dry and extreme weather conditions may have driven selection in both breeds to favour development of parallel adaptive mechanisms at the genome level. There is limited information about the shared genes associated with adaptation between the Afrikaner and Brahman, and the parallel evolution of adaptive mechanisms between these breeds. This information will be useful in breeding programmes to improve their genetic response to environmental stressors. In this study, we explore 2 statistical complementary tests i.e. fixation index (F_{ST}) and marker differentiation (FLK) (Bonhomme *et al.*, 2010) to assess genetic differentiated regions that may be evidence of parallel evolution of adaptive mechanisms due to similar environmental conditions in South African Afrikaner and Brahman cattle breeds.

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_409, © S. Mdyogolo *et al.* 2022

Materials & methods

Resource population and genotyping. Influential animals explaining the genetic diversity of the breeds were defined by their marginal genetic contributions to the entire population. The ARC-Animal Production ethical committee (APIEC18/03) approved sample collection (hair and semen samples) and genotyping procedures for this study. A total of 373 Afrikaner and 523 Brahman cattle were genotyped for 141,746 SNP markers using GeneSeek Genomic Profiler. Minor allele frequencies (MAF) and call rates (CR) for the individual SNP and animals were edited within and across breeds using Plink v1.9, where markers with MAF<0.02, CR<0.95, and individuals with call rate <0.90 were excluded. Within breeds, the edited data contained 92,291 and 99,086 autosomal SNP, respectively, for 310 Afrikaner and 523 Brahman animals. These data were used to estimate linkage disequilibrium (LD), genetic diversity and haplotype block structure. Across breeds, the edited data contained 83,306 SNP common to both breeds, which were used to detect differentiated regions containing putative selective sweeps.

Statistical analysis. Population admixture was investigated in the Afrikaner and Brahman after selecting 30,000 SNP markers pruned for LD<0.03 using a sliding window of 50 SNP and forward shift of 3 SNP in order to reduce the number of markers in LD. The genetic admixture with cross-validation (CV) for K values ranging from 1 to 3 was performed using Admixture v1.22 default parameters, were SA Limousine breed (Bos taurus) was used as an out-group. Clustering among breeds was visualized with a multi-dimensional (MDS) plot in Plink v1.9. Pairwise F_{ST} statistic was calculated as 1000 kb sliding windows with a step size of 50 kb. Genomic regions with average $F_{ST} \geq 0.2$ were examined further. The FLK statistic, which is a single SNP test that accounts for population size heterogeneity and structure to compute a global F_{ST} for each SNP was used to detect loci differing between breeds. The calculation of FLK exploits a kinship matrix where each chromosome was analysed independently, and the probability of the difference between loci was based on the $\log_{10}$ of probability values (*P*-values).

Ensembl variant effect predictor (VEP) was used to annotate the genome regions harbouring potential signatures of selection. The annotated variants from Ensemble were used for general 'gene to biological pathway' classification in PANTHER 15.0 and the NCBI gene database manual search was conducted to identify the candidate genes underlying signatures of selection.

Results

The Afrikaner cattle showed more ancestral admixture while the Brahman and Limousine had negligible levels of ancestral admixture (Figure 1 left). The separation of Brahman from Afrikaner and Limousine on the 1st MDS coordinate likely reflects more indicine heritage compared to the taurine ancestry of Afrikaner and Limousine breeds, while similarity of Brahman and Afrikaner at the 2nd MDS coordinate could be tracing back the indicine ancestor (Figure 1 right).

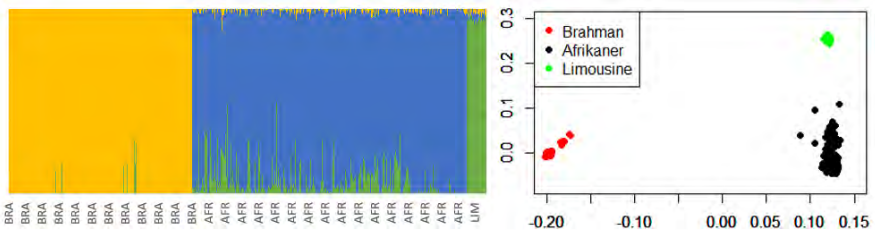


Figure 1. Admixture of the Afrikaner (AFR) and Brahman (BRA) ancestry at K=3 with the Limousine (LIM) as outgroup (left) and the Multi-dimensional plot of the distinct separation of the breeds (right).

Table 1 presents putative signatures of selection that are indicative of differentiation between Afrikaner and Brahman cattle and the number of underlying candidate genes for these breeds.

The average estimated F_{ST} was 0.318, and subsequent regions of differentiation between the breeds were identified in chromosomes 1, 6 and 21, with chromosome 1 having the longer window than 6 and 21. A total of 105 putative candidate genes were revealed with 65 annotated and 40 unannotated genes.

The FLK statistics located regions of differentiation in chromosomes 4, 6, 11, 12 and 21 with a combined 4,697 candidate genes underlying these regions, Chromosome 4 had the longest SNP window that consisted of 1,175 genes, while Chromosome 21 had a shortest window consisting of 814 candidate genes. Shared regions were observed between F_{ST} and FLK analysis on Chromosomes 6 and 21. Candidate genes identified in the shared regions included *FABP2, SEM4B and SEC11A*.

Discussion

Genetic structure and clustering showed some degree of common ancestry between the Afrikaner and Brahman. The levels of ancestral admixture of the Afrikaner versus Brahman and Limousine were previously observed by Makina *et al.* (2016), where ancestral genetic contributions making up the Afrikaner was quantified as 70% African taurine and 28% indicine. Population admixture and clustering for the Afrikaner was consistent with the suggested migration route and point of introgression between its ancestral breeds (African taurine from Egypt and Sudan; and *Bos indicus* breeds from Asia).

Differentiated genomic regions identified a higher proportion of candidate genes associated with adaptation than production in both breeds. For example, *GTPBP8* gene in Chromosome 1, was previously associated with post-partum anoestrus, resumption of ovulation, hormone activity and energy balance in tropical Brahman cattle (Fortes *et al.*, 2014). In terms of parallel adaptive mechanisms, shared regions between breeds and methods located on chromosomes 6 and 21 were underlined by candidate genes such

Table 1. Identified putative selection signatures indicative of differentiation between Afrikaner and Brahman cattle, and the underlying candidate genes using F_{ST} and FLK statistics.

Chromosome	Fixation index (F_{ST})				Marker differentiation (FLK)		
	Position (Bp)	Window length (Mb)	No. of SNP	Candidate genes	Position (Bp)	Window length (Mb)	Number of candidate genes
1	57500000~58500000	21.256	33	*CD200R1L, GTPBP8, NEPRO, BOC, CFAP44, SPICE1, SIDT1, USF3, NAA50, ATP6V1A*			
4					1440342~118804272	117.364	1 175
6	5500000~6500000	0.935	1	*TRNAG-CCC, PDE5A, FABP2, C6H4orf3, USP53, TRNAC-GCA, MYOZ2, SYNPO2, TRNAE-UUC, SEC24D*	4702047~113636044	108.934	873
11					6152931~105985714	99.834	1 264
12					16365586~90449343	74.084	571
21	21500000~22500000	10.524	17	*SEMA4B, CIB1, GDPGP1, NGRN, VPS33B, PRC1, TRNAG-CCC, RCCD1, UNC45A, HDDC3, MAN2A2, FES, FURIN, BLM, CRTC3, TRNAW-CCA, IQGAP1, ZNF774, ZSCAN2, WDR73, NMB, SEC11A, ZNF592, MIR2364, ALPK3*	13510976~68948621	55.434	814
21	29000000~30000000	18.671	28	*PCSK6, SNRPA1, CHRNA7, OTUD7A*			
21	68500000~69500000	17.966	28	*TDRD9, ASPG, KIF26A, C21H14orf180, TMEM179, INF2, ADSS1, SIVA1, AKT1, ZBTB42, CEP170B, PLD4, AHNAK2, CLBA1, CDCA4, GPR132*			

the *FABP2, SEM4B and SEC11A*, which have been previously associated with compensatory growth and mammary involution during dry period and environmental heat stress. These results showed inherited genomic characteristics by the Afrikaner and Brahman from their ancestral breeds, with small common ancestry.

Acknowledgements

Authors acknowledges the Red Meat Research and Development of South Africa, The Technology and Innovation Agency an implementing agency of the Department of Science and Innovation (DSI) and The National Research Foundation of South Africa for their financial support and the Centre for High Performance Computing for computational resources.

References

Bonhomme, M., Chevalet, C., Servin, B., Boitard, S., Abdallah, J. *et al.* (2010) Genet. 186, 241–262. https://doi.org/10.1534/genetics.110.117275

Chen, J., Tellez, G., Richards, J.D. and Escobar, J. (2015) Front. Vet. Sci. 2, 14-14. https://doi.org/10.3389/fvets.2015.00014

Fortes, M.R.S., Suhaimi, A.H.M.S., Porto-Neto, L.R., McWilliam, S.M., Flatscher-Bader, T. *et al.* (2014) Livest. Sci. 166, 158–166. https://doi.org/10.1016/j.livsci.2014.06.017

Hartati, H., Utsunomiya, Y.T., Sonstegard, T.S., Garcia, J.F., Jakaria, J. *et al.*, (2015BMC Genet. 16, 75. https://doi.org/10.1186/s12863-015-0229-5

Makina, S.O., Muchadeyi, F.C., van Marle-Köster, E., MacNeil, M.D. and Maiwashe, A. (2014) Front. Genet. 5 (333). https://doi.org/10.3389/fgene.2014.00333

Mattioli, R.C., Pandey, V.S., Murray, M. and Fitzpatrick, J.L. (2000) Acta Trop. 75, 263-277. https://doi.org/10.1016/s0001-706x(00)00063-2

Scholtz, M.M., Neser, F.W.C. and Tissier, L. (2016) The Afrikaner breed of cattle, South Africa's heritage of food security. 1st edition. Agricultural Research Council, Pretoria, South Africa.

410. Genomic analysis of milk yield and heat tolerance in small holder dairy system of sub-Saharan Africa

C.C. Ekine-Dzivenu[1*], R. Mrode[1,2], R.D. Oloo[1,4], D. Komwihangilo[3], E. Lyatuu[3], G. Msuta[3], J.M. Ojango[1] and A.M. Okeyo[1]

[1]International Livestock Research Institute, Animal Biosciences, P.O. Box 30709 Nairobi, 00100, Kenya; [2]Scotland Rural College, Easter Bush, Roslin Institute Building, EH25 9RG Midlothian, United Kingdom; [3]Department of Animal Breeding and Husbandry in the Tropics and Subtropics, University of Hohenheim, Garbenstr 17, 70595 Stuttgart, Germany; [4]Tanzania Livestock Research Institute, P.O. Box 834, Dodoma, Tanzania; cekine@cgiar.org

Abstract

This study was undertaken to estimate the additive genetic variances of milk yield and heat tolerance and the genetic correlation between these effects at different thermal loads as a first step in exploring the possibility of including heat tolerance in selection indices for dairy cattle selection in the sub-Saharan African climate of Tanzania. Data obtained from aWhere, an agricultural weather data platform (http://www.awhere.com) was analysed with 15,703 test day milk records of 9,922 dairy cows collected between 2016 and 2020. Temperature humidity index (THI) was calculated from daily maximal temperatures and daily minimum humidity. Variance components were estimated using a fixed regression model with random regressions on a function of THI for additive genetic and permanent environmental effects modelled with an intercept and a slope. Results suggested that milk yield is antagonistic with heat tolerance, and selection for only increased milk production might reduce heat tolerance in this population.

Introduction

Climate change is expected to have major impacts on livestock, including increased heat stress in animals in many parts of the tropics and in low and middle-income countries (LMICs), where the cost of heat stress abatement measures may be prohibitive (Thorton *et al.* 2021). One alternative approach to mitigating the impact of climate change in dairy cattle, especially in smallholder dairy systems of LMICs which are more vulnerable to the impacts of a changing global climate than other dairy systems would be selection for improved heat tolerance in production traits, to which several authors have demonstrated substantial between animal genetic variation and low to moderate heritability (Ravagnolo and Misztal, 2000; Bernabucci *et al.*, 2014). Using temperature-humidity index (THI), a single value that combines the effects of air temperature and humidity as an indicator of heat load, Ekine-Dzivenu *et al.* (2020) evaluated the effect of heat stress on milk production and described the pattern of response of milk yield to increasing heat load in small holder dairy cattle populations in the sub-Saharan African climate of Tanzania. Results from their study indicated that cows experienced heat stress between THI values of 67 and 76 after which milk loss plateaued. Identifying sires and dams to improve thermotolerance in smallholder farming systems in sub-Saharan Africa in the face of rising temperatures is of vital importance for small holder dairy profitability and will require availability of genetic parameters which is presently lacking. Within the described background, the goal of this study, was to evaluate changes in genetic parameters for milk yield and heat tolerance in small holder dairy cattle population of Tanzania under different THI thresholds as a first step toward including this trait in selection indices for dairy cattle in the region.

Materials & methods

Data. This study was carried out as part of a wider project called African Dairy Genetic Gains (https://www.ilri.org/research/projects/african-dairy-genetic-gains) which is developing and testing a multi-country genetic gains platform using smallholder dairy cattle performance and genomic information. Performance

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_410, © C.C. Ekine-Dzivenu *et al.* 2022

data consisting of 134,987 test day records of 14,741 cows from 8,735 farms, with pedigree information comprising 67,910 animals were extracted from the International Livestock Research Institute (ILRI) database (https://www.adgg.ilri.org/uat/auth/auth/login) for cows registered on the ADGG project from November 2016 until May 2020 in Tanzania. The Data was subjected to several edits as previously described by Mrode *et al.* (2021). Climate data was obtained from aWhere (http://www.awhere.com), an agricultural weather data platform using GPS coordinates of the farms at 10 km grid cell resolution. This included daily maximum (Tmax, °C) and minimum relative humidity (RHmin, %). Daily THI was calculated by applying the National Research Council (1971) formula as follows:

$$\text{THI} = (1.8 \times \text{Tmax} + 32) - [(0.55 - 0.0055 \times \text{RHmin}) \times (1.8 \times \text{Tmax} - 26.8)] \quad (1)$$

Each THI was calculated using a 4-day average of Tmax and RHmin from measures on the test day and 3 days prior. After production and weather data were merged and duplicate THI values per animal removed, 15,703 records for 9,922 animals remained for analysis.

Genotypes. Genotypic data for this study were from 1,916 cows and bulls genotyped with the GeneSeek Genomic Profiler Bovine 50K chip. These were imputed to the Illumina HD chip using a reference population consisting of crossbred cattle from East Africa and several European purebred animals (Aliloo *et al.*, 2018). After quality control, 681,511 SNPs remained for inclusion in the analysis.

Statistical model. Variance components were estimated using a fixed regression model with random regressions on a function of THI for additive genetic and permanent environmental effects using the Wombat program (Meyer, 2006). The model fitted was as follows

$$y_{ijklmno} = \text{wym}_i + \text{breed}_j + \text{par}_k + \text{age}_l(\text{par}_k) + \text{dim}_m(\text{par}_k) + a_n + \text{aht}_n\,[f(THI)i] + \text{pe}_n + \text{peht}_n\,[f(THI)_i] + e_{ijklmno} \quad (2)$$

where $y_{ijklmno}$ is the o^{th} test-day milk yield of cow n; wym_i is the random effect of the i^{th} combination of ward-year-month of milking; breed_i is the fixed effect of the i^{th} breed group; par_k is the fixed effect of the k^{th} parity; age_l is the fixed effect of the l^{th} age group nested within parity; dim_m is the fixed effect of the m^{th} dim group nested within parity; a_n is the random additive genetic effect of cow n without consideration of heat tolerance (intercept); aht_n is the random additive genetic effect of heat tolerance for cow n (slope); pe_n is the random permanent environmental effects of cow n without consideration of heat stress (intercept); $peht_n$ is the random permanent environmental effect of cow n for heat tolerance (slope); $f(THI)_i$ is a heat stress function for ward-year-month of milking and $e_{ijklmno}$ is the random residual effect. The THI function was defined as follows:

$$f(THI) = \{0, THI \leq THIthreshold\ (no\ heat\ stress)$$

$$THI - THIthreshold, THI > THIthreshold\ (heat\ stress) \quad (3)$$

using various heat stress threshold points (69, 71, 78), previously reported by Ekine-Dzivenu *et al.* (2020) for the same population. Genomic analyses were implemented with single-step genomic best linear unbiased prediction (ssGBLUP) where inverse of the pedigree-genomic relationship matrix H was used for the estimation of genetic parameters. Heritability for production of milk at heat stress levels, genetic correlations between the intercept and slope of the additive genetic effects and between the intercept and slope of the permanent environmental effects was calculated based on Ravagnolo and Misztal (2000).

Results

A summary of the mean, standard deviation, and range of test-day milk yield including a summary of weather data characterizing THI during the period of this study is presented in Table 1. Estimates of genetic parameters for milk yields along the THI gradient are presented in Table 2 and shows a decrease of variance for additive genetic effect of the intercept (9.83±1.21 to 7.52±1.39) but an increase for additive genetic variance for the slope (0.10±0.61 to 2.15±0.95). Covariance between the intercept and the slope for additive genetic effects were negative and ranged from -0.24±0.72 to -2.10±0.93 and correlation ranged from -0.06 to -0.99 but variances and covariance between the intercept and the slope for permanent environmental effects were negligible across THI levels (not shown). Heritability estimates for milk yield showed an increase across THI gradient and ranged between 0.38 to 0.50.

Discussion

Our study detected antagonistic relationships between milk yield (intercept) and thermal tolerance for heat stress (slope) which suggests that selection for increased milk yield disregarding heat tolerance will result in deteriorating heat tolerance in this population. Antagonism between milk production and heat tolerance has also been found in several other studies in cattle (Ravagnolo and Misztal, 2000; Bernabucci *et al.*, 2014; Carabaño *et al.*, 2014). Estimates for permanent environmental variances were very small which might indicate that sensitivity for heat stress is not cow-specific but genetic. It is important to note that not observing animals across the THI gradient due to low retention of cows in the small holder farms may contribute to the low permanent environmental variances observed. The moderate heritability estimates for milk yield over different heat stress levels found in this study show potential to genetically improve

Table 1. Basic description of milk yield and weather data.

	Mean	Std	Range
Milk, liters	7.90	4.21	1-38
Temp, °C	21.05	5.50	11.6-30.64
RH, %	80.48	3.73	68.06-89.61.
THI	76.3	3.78	69-82

Table 2. Estimates for genetic parameters and heritability for milk production at different THI thresholds (heat stress levels).[1]

Variance components	THI thresholds		
	69	71	78
σ^2_a	9.83(1.21)	9.16 (1.32)	7.52(1.39)
σ^2_{aht}	0.10(0.61)	1.96(1.20)	2.15(0.95)
$\sigma_{a(a, aht)}$	-1.00(0.59)	-2.10(0.93)	-0.24(0.72)
$r_{g(a, aht)}$	-0.99(NE)	-0.50(0.15)	-0.06(0.19)
$r_{pe(pe,peht)}$	-0.92(NE)	0.64(NE)	0.90(NE)
σ^2_{wys}	8.85(0.38)	6.65(0.37)	5.47(0.43)
σ^2_e	5.47(0.12)	4.79(0.12)	4.15(0.18)
h^2	0.38	0.44	0.50

[1] σ^2_a= additive genetic variance of animal, σ^2_{aht} = additive genetic variance of heat tolerance effect, $\sigma_{a(a,aht)}$= covariance between additive genetic variance of animal and heat tolerance effect, $r_{g(a,aht)}$= correlation between additive genetic effect and additive genetic heat tolerance effect, $r_{pe(pe,peht)}$= correlation between permanent environmental variance of animal and heat tolerance effects, σ^2_{wym}= variance due to ward-year-month of calving, σ^2_e = residual variance, h^2= heritability. NE= not estimable.

milk yield for climate conditions with high THI values howbeit with care to avoid unfavourable correlated response of thermal stress on milk production.

References

Aliloo, H., Mrode, R., Okeyo, A. M., Ni, G., Goddard, M. E., et. Al. (2018). The feasibility of using low-density marker panels for genotype imputation and genomic prediction of crossbred dairy cattle of East Africa. Journal of dairy science, 101(10), 9108-9127. https://doi.org/10.3168/jds.2018-14621

Bernabucci, U., S. Biffani, L. Buggiotti, A. Vitali, N. Lacetera, and A. Nardone. (2014). The effects of heat stress in Italian Holstein dairy cattle. J. Dairy Sci. 97:471–486. https://doi.org/10.3168/jds.2013-6611

Carabano, M.J., Bachagha, K., Ramon, M., Díaz, C., 2014. Modeling heat stress effect on

Holstein cows under hot and dry conditions: selection tools. J. Dairy Sci. 97 (12), 7889–7904. https://doi.org/10.3168/jds.2014-8023

Ekine-Dzivenu, C. C., Mrode, R., Oyieng, E., Komwihangilo, D., Lyatuu, E., *et al.* (2020). Evaluating the impact of heat stress as measured by temperature-humidity index (THI) on test-day milk yield of small holder dairy cattle in a sub-Sahara African climate. Livestock science, 242, 104314. https://doi.org/10.1016/j.livsci.2020.104314

Meyer, K. (2007). WOMBAT – A tool for mixed model analyses in quantitative genetics by restricted maximum likelihood (REML). Journal of Zhejiang University Science B, 8(11), 815-821. https://doi.org/10.1631/jzus.2007.B0815

Mrode, R., Ojango, J., Ekine-Dzivenu, C., Aliloo, H., Gibson, J., and Okeyo, M. A. (2021). Genomic prediction of crossbred dairy cattle in Tanzania: A route to productivity gains in smallholder dairy systems. Journal of Dairy Science,104(11),11779-11789. https://doi.org/10.3168/jds.2020-20052

National Research Council (US), 1971. A Guide to Environmental Research on Animals.

National Academy of Sciences, Washington, US.

Ravagnolo, O., Misztal, I., & Hoogenboom, G. (2000). Genetic component of heat stress in dairy cattle, development of heat index function. Journal of dairy science, 83(9), 2120-2125. https://doi.org/10.3168/jds.S0022-0302(00)75095-8

Thornton, P., Nelson, G., Mayberry, D., and Herrero, M. (2021). Increases in extreme heat stress in domesticated livestock species during the twenty-first century. Global Change Biology, 27(22), 5762-5772. https://doi.org/10.1111/gcb.15825

411. Genetic analysis of traits needed to optimize improvement in beef cow-calf efficiency in Afrikaner cattle

M.M. Scholtz[1,2*], M.D. MacNeil[1,2,3], M.C. Chadyiwa[1,2] and F.W.C. Neser[2]

[1]Agricultural Research Council, Animal Production, Private Bag X2 Irene 0062, South Africa; [2]University of the Free State, P.O. Box 33009, Bloemfontein, 9300, South Africa; [3]Delta G, 145 Ice Cave Rd., Miles City, MT 39301, USA; gscholtz@arc.agric.za

Abstract

Improving beef cattle productivity will have positive sustainability implications, as it will reduce resource use while improving the economic viability. Achieving optimal cow-calf efficiency is therefore important but measuring it in extensive cow-calf production systems is not easy; however, it is imperative that it should be estimated. In this study data from the indigenous Afrikaner beef breed was used to estimate genetic parameters for the three component traits of cow-efficiency, viz. weaning weight, cow weight and fertility as measured by inter-calving period. The estimates of (co)variance components indicated low to moderate heritabilities for the component traits and the presence of weak genetic antagonisms among the components of cow efficiency. The use of the ratio of calf weight to cow weight should be replaced by an index that includes reproductive performance. Annual cow-calf efficiency can be calculated for every individual, including bulls and those with incomplete records, using their EBV's.

Introduction

The world population is expected to increase dramatically in the coming decades with a commensuration increase in demand for high quality protein products (e.g. beef). This population increase is expected to be mostly in the developing countries of Africa (Bello-Schünemann, 2017). In addition, the beef cattle industry globally faces the challenge to maintain and improve sustainability with the increase in global population. Furthermore, environmental stewardship of beef producers is under scrutiny as many consumers perceive that beef production has an unacceptable environmental cost (Scholtz *et al.,* 2020). Improving production efficiency will have positive implications for sustainability because resource use and greenhouse gas emissions will be reduced whilst improving the economic viability of beef producers (Scholtz, *et al.,* 2013a).

Livestock agriculture is the largest user of land resources in South Africa, where 71% of the surface area is only suitable for extensive livestock farming (Scholtz, *et al.*, 2013b). Consequently, primary beef cattle farming (cow-calf production cycle) is largely extensive. The cow-calf portion of the production cycle (the extensive part in South Africa) accounts for 72% of the energy that is utilized from conception to harvest. (Ferrell and Jenkins, 1982). Improvement of the productivity of beef per unit of feed used over the whole production system would be of significance to beef producers.

Achieving optimal efficiency is therefore a significant goal in beef production. It should be noted that on the farm, an efficient animal would be the one that has minimum maintenance requirements and has the ability to convert available energy (forage) into the greatest possible kilogram of weaned calves, without increasing the mature weight (or the maintenance requirements) of the cow. In spite of the fact that the ratio of calf weight at weaning to cow weight (CW) favours smaller cows, it has long been used as a proxy for efficiency (Dickerson and Grimes, 1947; Frahm and Marshall, 1985; Lemes *et al.*, 2017) and is still used frequently (Thompson *et al.*, 2020; Farrell *et al.*, 2021). Therefore, following from Dickerson (1970), an index (I) of the phenotypes for the three components traits weaning weight (WW), CW and inter-calving period (ICP) is proposed as a metric to evaluate annual cow efficiency:

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_411, © M.M. Scholtz *et al.* 2022

$$I = (365 \times WW) / (ICP \times CW)$$

The objective of this study was to estimate genetic parameters for the component traits in I for the Afrikaner breed, which is indigenous to South Africa.

Materials & methods

The data were extracted from the Integrated Registration and Genetic Information System of South Africa. After being edited, the data comprised of weaning weight records for 29,975 calves, with 19,191 records of cow weights from 8,556 unique cows, and 6,181 records of inter-calving period. Only those cows with at least one recorded measurement of body weight at the time of weaning and their progeny were retained in the dataset for analysis. The pedigree that was used to quantify relationships among animals contained 38,766 records.

Multiple-trait analyses using restricted maximum likelihood methods were implemented using the software of Boldman *et al.* (1995). The linear model for weaning weight contained a linear covariate for age of calf at weaning, a categorical fixed effect for age of dam and sex of calf, random direct and maternal genetic effects, and uncorrelated random effects of the herd-year-season (HYS) in which the calf was weaned and the permanent maternal environment provided by the dam. The linear model for cow weight contained a categorical fixed effect for the age of the cow, a random direct genetic effect, and uncorrelated random effects to account for the herd-year-season in which the cow was weighed and to account for repeated observations. The linear model for inter-calving period contained a categorical fixed effect for the age of the cow when she first calved, a random direct genetic effect, and uncorrelated random effects to account for the herd-year-season in which she first calved. Note that with annual calving and a fixed breeding season, the HYS of subsequent calving events is confounded with the herd-year-season of first calving. Convergence to a global maximum was assumed when the variance of values in the simplex was less than 10^{-9} with confirmation that the parameter estimates were unchanged in independent analyses. Adjusted phenotypes for WW, CW, and ICP were calculated by combining the corresponding EBV with the overall means for the respective traits.

Results

Estimates of the phenotypic and genetic parameters for weaning weight, cow weight and inter-calving period for Afrikaner cattle are given in Table 1.

Table 1. Estimates of genetic parameters for weaning weight, cow weight and inter-calving period of Afrikaner cattle.

Effects	Weaning weight		Cow weight		Inter-calving period	
	VC[1]	%[2]	VC[1]	%[2]	VC[1]	%[2]
Direct genetic	131.1	18±1	923.3	33±2	3,184.7	34±3
Maternal genetic	72.6	10±1				
Permanent environmental	58.5	8±1	383.9	14±1		
Herd-year-season	265.6	33±1	691.2	25±1	2,912.1	31±2
Residual	233.7	32±1	724.2	27±1	3,275.6	35±3
Phenotypic variance	730.0		2,722.6		9,372.4	

[1] Variance component.

[2] Percentage of phenotypic variance.

Estimates of the heritabilities and genetic correlations are given in Table 2. The heritability estimates may differ somewhat from those found in other literature because they include the variance of HYS effects in the phenotype variance. This practice allows for the recovery of genetic variance that is present between the HYS classes which make the prediction of breeding values more accurate. The heritability estimates range from low in the case of weaning weight direct (WWd) and weaning weight maternal (WWm) to moderate / high in the case of cow weight (CW) and inter-calving period (ICP). The high heritability of ICP is somewhat surprising, especially considering that the variance among contemporary groups is included in its denominator.

With respect to genetic correlations the following should be noted: (1) there is a moderate negative genetic correlation between WWd and WWm, (2) WWd is moderately positively correlated with CW, indicating selection for it will tend to result in heavier cows, (3) the correlation between WWm and cow weight is low negative, indicating that cows with more milk will tend to be smaller, (4) the low negative correlation between WWd and ICP may indicate that calves that are heavier at weaning may have shorter ICP's and (5) the correlation between CW and ICP is low negative indicating that lighter cows will have longer ICP's.

The estimated breeding values (EBV) resulting from this multiple trait analysis had average accuracies for WWd, WWm, CW, and ICP of 0.59, 0.46, 0.61, and 0.51, respectively. The rank correlations were also estimated and are presented in Table 3. The negative rank correlation of ICP with WWd (-0.46) and CW (-0.46), indicate that more fertile cows will tend to have greater growth potential.

Due to its commonplace use (e.g. Thompson *et al.*, 2020), the ratio of weaning weight to cow weight is used here as a benchmark. The annual efficiency index (I) that includes ICP ranks animals in a manner that is moderately correlated with this ratio (r =0.5). Thus, using the weaning weight to cow weight ratio will explain approximately 25% of the variation in annual efficiency when differences in the probability of producing a calf are taken into account. The rank correlations for the proposed annual efficiency index with its component traits were 0.47 for weaning weight direct, 0.31 for weaning weight maternal, 0.15 for cow weight, and -0.80 for inter-calving period.

Table 2. Estimates of heritability (on the diagonal) and genetic correlation (above the diagonal).

Trait	Weaning weight direct	Weaning weight maternal	Cow weight	Inter-calving period
Weaning weight direct	**0.18**	-0.32	0.55	-0.17
Weaning weight maternal		**0.10**	-0.22	0.14
Cow weight			**0.33**	-0.19
Inter-calving period				**0.34**

Table 3. Estimates of the rank correlations between the EBV's of the different traits.

Trait	Weaning weight maternal	Cow weight	Inter-calving period
Weaning weight direct	-0.02	0.76	-0.46
Weaning weight maternal	-	-0.13	0.06
Cow weight		-	-0.46

Discussion

The estimates of (co)variance components clearly show the presence of modest genetic correlations among the components of cow efficiency. The direct-maternal correlation for weaning weight was clearly negative in these data and should be included in the genetic evaluation for this trait (Lee, 2002). However, the ranking of animals by either component of weaning weight was essentially unchanged. Conversely, the genetic correlations of ICP with WWd and CW were both larger and thus indicated that more fertile cows (low ICP) would be genetically heavier. Favourable genetic relationships between growth and reproduction have been observed (Kluska *et al.*, 2018; Santana *et al.*, 2012); but significant antagonisms between these traits have also been observed (MacNeil *et al.*, 1984). These genetic relationships indicate the need for multiple trait genetic evaluation and selection indexes to improve biological and economic efficiency, and reduce the environmental impact of beef production.

It is recommended that the use of the ratio calf weight to cow weight needs to be replaced by an index that includes reproductive performance. Because the components of cow efficiency are correlated, additional accuracy can be gained in their genetic evaluation using multiple trait models. The recommended index can be derived from the genetic evaluation for the component traits. One advantage of using the EBV's to calculate annual cow-calf efficiency is that it can be calculated for every individual, including bulls and animals with incomplete records.

References

Bello-Schuneman, J. (2017). Africa's population boom: burden or opportunity? Institute for Security Studies. Pretoria, South Africa.

Boldman K., Kriese K.A., Van Vleck L.D., Van Tassell C.P., and Kachman S.D. (1995) U.S. Department of Agriculture, Agricultural Research Service, Washington, DC.

Dickerson G. (1970) J. Anim. Sci. 30(6):849–859. https://doi.org/10.2527/jas1970.306849x

Dickerson G.E. and Grimes J.C. (1947). J. Anim. Sci. 6:265-287. https://doi.org/10.2527/jas1947.63265x

Farrell L.J. Morris S.T. Kenyon P.R. and Tozer P.R. (2021) Agriculture 11:35. https://doi.org/10.3390/agriculture11010035

Ferrell C.L. and Jenkins,T.G. (1982) Efficiency of cows of different size and milk production potential. RLHUSMARC Germ Plasm Evaluation Report 10, 12.

Frahm R.R. and Marshall D.M. (1985) J. Anim. Sci. 61:844-855. https://doi.org/10.2527/jas1985.614747x

Kluska S., Olivieri B.F., Bonamy M., Chiaia H.L.J. Feitosa F.L.B. *et al.* (2018) Livest. Sci. 216:203-209. https://doi.org/10.1016/j.livsci.2018.08.015.

Lee C. (2002) Asian-Australas. J. Anim. Sci. 15(8):1222-1226. https://doi.org/10.5713/ajas.2002.1222

MacNeil M.D., Cundiff L.V., Dinkel C.A. and Koch R.M. (1984) J. Anim. Sci. 58(5):1171-1180. https://doi.org/10.2527/jas1984.5851171x

Santana M.L., Eler J.P., Ferraz J.B.S. and Mattos, E.C. (2012) Animal 6(4):565-570. https://doi.org/10.1017/S1751731111001856

Scholtz M.M. Maiwashe A. Neser F.W.C. Theunissen A. Olivier W.J. *et al.* (2013a) S. Afr. J. Anim. Sci. 43:269-281. https://doi.org/10.4314/sajas.v43i3.4

Scholtz M.M. Neser F.W.C. and Makgahlela M.L. (2020) S. Afr. J. Sci. 116(9/10), Art. #8192. https://doi.org/10.17159/sajs.2020/8192

Scholtz M.M. van Ryssen J.B.J., Meissner H.H. and Laker M.C. (2013b) S. Afr. J. Anim. Sci. 43:247-254. https://doi.org/10.4314/sajas.v43i3.2

Thompson L.R. Beck M.R., Buskirk D.D. Rowntree, J.E and McKendree M.G.S. (2020) Trans. Anim. Sci. 4(3):txaa166. https://doi.org/10.1093/tas/txaa166

412. Genetic diversity of the Cameroon Blackbelly sheep, an indigenous sheep from West Africa

P. Wiener[1*], M. Salavati[1,2], A. Djikeng[2], C.P. Van Tassell[3], B.D. Rosen[3], G.L. Spangler[3,4], G. Simo[5], V. Tanya[6], F. Meutchieye[7] and E.L. Clark[1,2]

[1]Division of Genetics and Genomics, The Roslin Institute and Royal (Dick) School of Veterinary Studies, University of Edinburgh, Midlothian, United Kingdom; [2]Centre for Tropical Livestock Genetics and Health, University of Edinburgh, Midlothian, United Kingdom; [3]Animal Genomics and Improvement Laboratory, USDA-ARS, Beltsville, MD, USA; [4]Epic Sciences, San Diego, CA, USA; [5]Department of Biochemistry, Faculty of Science, University of Dschang, Dschang, Cameroon; [6]Cameroon Academy of Sciences, Yaounde, Cameroon; [7]Department of Animal Science, Faculty of Agronomy and Agricultural Science, University of Dschang, Dschang, Cameroon; pam.wiener@roslin.ed.ac.uk

Abstract

While genomic tools have been applied to address various questions in livestock of European origin, primarily from temperate and boreal ecological zones, less is known about African breeds, many of which are from tropical regions. Our study focused on indigenous Blackbelly sheep from Cameroon, most of which has a tropical climate. We found that Cameroon Blackbelly sheep are genetically distinct from other African sheep populations and are associated with a single ancestry. Genes found in genomic regions of low diversity suggest biological functions that may be under strong selection, including immune function, pigmentation and fertility. Genomic diversity information from indigenous tropical livestock is a valuable source of information to understand environmental adaptation of animals to local conditions and will also inform the development of new and existing genotyping platforms.

Introduction

An important application of genomic technologies over the last several years has involved the use of genetic diversity patterns to help identify genes and functional gene groups associated with environmental adaptation. This approach has been applied in various livestock species and has led to the identification of candidate genes related to adaptation to high altitude (Friedrich and Wiener, 2020) as well as other climatic and disease factors (e.g. hot or cold temperature, aridity (Wiener *et al.*, 2021; Yurchenko *et al.*, 2018); nematode resistance (Ahbara *et al.*, 2021)).

The Republic of Cameroon is a country in west-central Africa. Most of Cameroon has a tropical climate, with wet and dry seasons. It has a diverse geography, including highlands, low-lying coastal plains and tropical forests, suggesting that autochthonous livestock may be adapted to the local environmental conditions. It has also been estimated that 2/3 of the country is at risk from Animal African Trypanosomiasis, due to tsetse fly infestation (Meyer *et al.*, 2016). Compared to cattle, less is known about this endemic African disease among small ruminants, however, there is evidence that some West African sheep and goat breeds are trypanotolerant (Geerts *et al.*, 2009). Anecdotally, the Cameroon Blackbelly sheep, which is indigenous to Cameroon, has been reported to be tolerant to trypanosomiasis, suggesting that they have been under selection for this trait and that this may have left a signature on their genome. Based on similarities in physical appearance and historical information (Meka zibi II. *et al.*, 2019) as well as genetic findings for other West African sheep (Spangler *et al.*, 2017), it is likely that the genetics of the Cameroon Blackbelly sheep has contributed to sheep breeds of the Caribbean (i.e. Barbados Blackbelly) and elsewhere in the world. Capturing the genetic diversity of indigenous breeds such as Cameroon Blackbelly will facilitate characterisation of the genetic basis of environmental adaptation and disease resistance and also help to ensure that such breeds are adequately represented in new and existing genotyping tools.

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_412, © P. Wiener *et al.* 2022

Materials & methods

Samples. Ear notch samples from 111 Cameroon Blackbelly sheep were collected from regions across Cameroon. DNA was extracted using a DNAEasy Blood and Tissue Kit. Samples were genotyped using the Illumina Ovine 50K SNP chip. Additional 50K data for 97 other sheep breeds from across the globe was extracted from public databases. For two additional breeds from West Africa, publicly available whole-genome sequence data was down-sampled to provide 11-22K markers on the 50K array. Highly related individuals (first-degree or greater) were removed from the dataset (including 30 Cameroon Blackbelly sheep). Marker positions were based on the OAR3.1 reference genome assembly coordinates.

Population structure. In order to analyse data across breeds, a subset of 12,047 markers was extracted that were genotyped in most of the down-sampled individuals and which met additional criteria (genotype coverage >0.90, minor allele frequency >0.01 and between-marker r^2<0.5). In order to balance the dataset across breeds, a subset of individuals were selected for analysis. For the set of European + Pacific (Australia/ New Zealand) breeds, 250 representative samples were selected using the Corehunter v3.2.1 method with default settings (De Beukelaer *et al.*, 2018). For other breeds with >10 samples (excluding the Cameroon sheep), 10 individuals per breed were selected, again using Corehunter. All individuals were retained from breeds with ≤10 samples and from the Cameroon sheep. The complete dataset comprised 779 individuals, including 81 Cameroon sheep. Subsequent analyses were carried out for only the African breeds (262 individuals, 11,937 markers) and only the Cameroon sheep (81 individuals, 11,976 markers).

Plink 1.9 (Chang *et al.*, 2015) was used to carry out principal components analysis (PCA) on the full dataset. Admixture software (Alexander *et al.*, 2009) was used to estimate ancestry using the complete dataset, the subset of African breeds and only the Cameroon sheep. Assessment of the number of underlying clusters (K) was assessed by the default (5-fold) cross-validation procedure.

Genomic regions of reduced diversity. Normalized heterozygosity values in the Cameroon sheep were identified using the ZHp statistic (Rubin *et al.*, 2012; Rubin *et al.*, 2010) for 45,066 markers with >0.90 genotype coverage in the 81 unrelated individuals. Window averages (ZHp-window) were calculated in 9-SNP sliding windows across the genome (with step size of 1 SNP), such that the statistic was attached to the central marker in each window. Markers in the bottom 0.001of the ZHp-window distribution (lowest level of heterozygosity) were extracted and signatures of selection were identified as groups of markers within 200 kb of each other. Genes underlying the genomic signatures were identified based on the annotation of the OAR3.1 reference assembly.

Results

PCA. The first principal component (PC) of the PCA (Figure 1) explained 29% of the genotype variation while PC2 explained 23%. The separation of individuals/breeds by PCA was strongly associated with geography, with a few exceptions: (1) Pacific breeds clustered with European breeds and (2) some North and South African individuals clustered together, close to Middle Eastern breeds. Higher PC1 values were associated with African breeds, with highest values for the Cameroon sheep. Lower PC1 values were associated with European/Pacific breeds.

Population ancestry. The Admixture analysis of the complete dataset supported K=14 underlying clusters, where West African sheep were primarily associated with a single cluster. Out of the 81 individuals from Cameroon, all except four had ancestry proportions of >70% for this cluster. Outside of West African sheep, this cluster was represented at <13% ancestry except in the Barbados Blackbelly breed, from the Caribbean, where it was represented at ~30%. The Admixture analysis of only African populations (Figure 2) supported K=6 underlying clusters, where one main cluster was represented at high ancestry in the Cameroon sheep (most individuals with ancestry >80%) and also in the other West African sheep (Figure 2). This cluster was

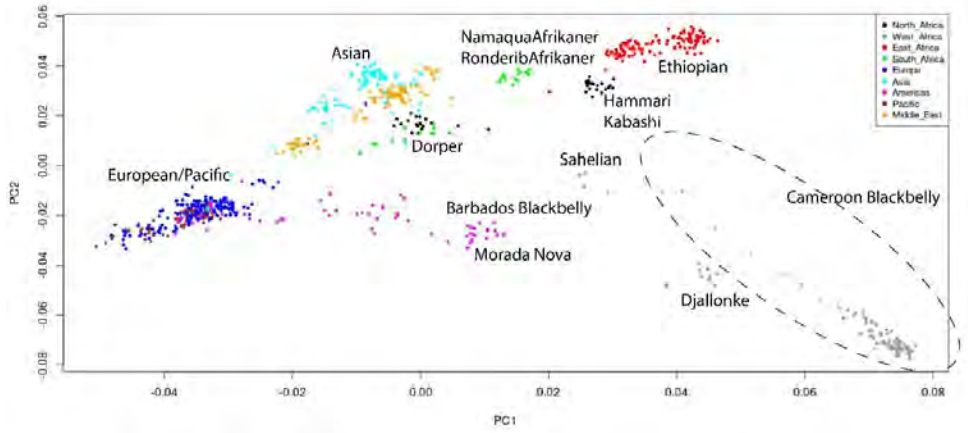


Figure 1. PCA of 779 sheep from across the world, including 81 unrelated Cameroon Blackbelly individuals. Analysis based on 12,047 markers.

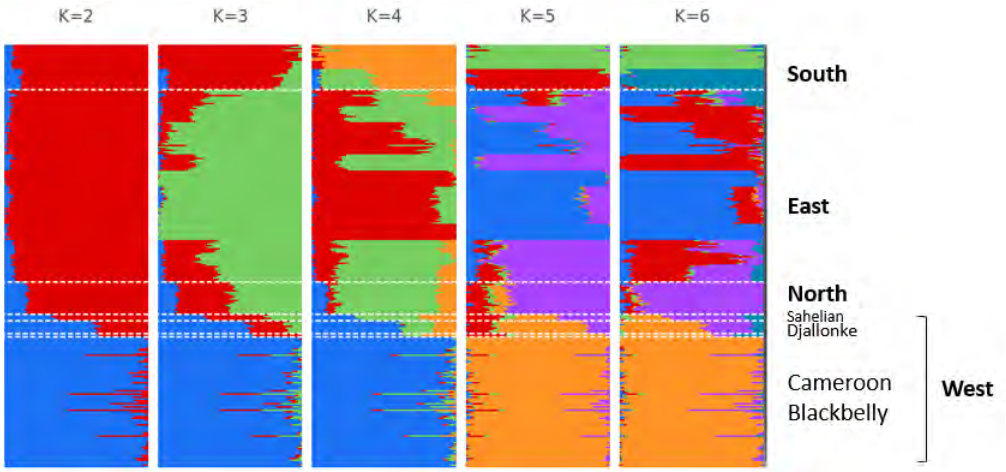


Figure 2. Admixture results for 262 African sheep samples: ancestry proportions for K=2-6 clusters shown on x-axes. North, East, South refer to breeds from those African regions; West African sheep: Cameroon Blackbelly, Sahelian and Djallonke are shown separately.

represented at <6% ancestry in breeds from North, East and South Africa. The Admixture analysis of only Cameroon sheep supported a single underlying cluster (K=1 had the lowest CV error).

Reduced diversity. Twenty-three regions of reduced heterozygosity in Cameroon sheep (lowest values of ZHp) were identified across the genome (OAR1,2,3,4,5,7,9,11,15,17,18 and 23). The most extreme was on OAR3, which overlapped the PUM2 gene (encoding an RNA-binding protein that functions as a translational repressor during embryonic development). The other regions overlapped 10 genes, which include those associated with immune function (IL6R), pigmentation (OCA2) and fertility (SPAG1), suggesting possible functions under selection.

Discussion

Our analysis demonstrates the genetic distinctiveness of West African sheep compared to other African sheep, with separation on both the first and second principal components of the PCA. The ancestry results show that the Cameroon Blackbelly sheep are primarily associated with a single cluster, with limited admixture from other populations. These findings reflect the genetic uniqueness of small ruminants in West Africa and demonstrate the importance of conserving them as a genetic resource and ensuring that indigenous breeds are adequately represented on new and existing genotyping tools. The finding of West African ancestry in Barbados Blackbelly sheep is consistent with previous studies (Spangler *et al.* 2017) and demonstrates the role of human history and migration (e.g. slave trade) in the worldwide distribution of livestock. Additional analysis, including analysis of whole genome sequencing data, will further explore introgression from West African sheep into Caribbean breeds.

Acknowledgements

This work was funded by a BBSRC Global Challenges Research Fund Data and Resources Grant BBS/OS/GC/200012A 'Developing genotyping tools for Cameroon sheep as a model for understanding the genetic basis of trypanotolerance in small ruminants in West Africa'.

References

Ahbara A.M., Rouatbi M., Gharbi M., Rekik M., *et al.* (2021) Sci. Rep. 11. https://doi.org/10.1038/s41598-021-88501-3.

Alexander D.H., Novembre J., and Lange K. (2009) Genome Res. 19:1655-1664. https://doi.org/10.1101/gr.094052.109.

Chang C.C., Chow C.C., Tellier L.C.A.M., Vattikuti S., *et al.* (2015) GigaScience 4:7.

De Beukelaer H., Davenport G.F., and Fack V. (2018) BMC Bioinformatics 19. https://doi.org/10.1186/s12859-018-2209-z.

Friedrich J., and Wiener P. (2020) Anim. Genet. 51:157-165. https://doi.org/10.1111/age.12900.

Geerts S., Osaer S., Goossens B., and Faye D. (2009) Trends Parasitol. 25:132-138. https://doi.org/10.1016/j.pt.2008.12.004.

Meka zibi II. M.A., Meutchieye F., Ntsoli J., Tadakeng Y., *et al.* (2019) PKFokam Journal of Applied Science and Technology 1:28-37.

Meyer A., Holt H.R., Selby R., and Guitian J. (2016) PLoS Negl Trop Dis 10. https://doi.org/10.1371/journal.pntd.0005247.

Rubin C.-J., Megens H.-J., Barrio A.M., Maqbool K., *et al.* (2012) Proc. Natl. Acad. Sci. USA 109:19529-19536. https://doi.org/10.1073/pnas.1217149109.

Rubin C.-J., Zody M.C., Eriksson J., Meadows J.R.S., *et al.* (2010) Nature 464:587-U145. https://doi.org/10.1038/nature08832.

Spangler G.L., Rosen B.D., Ilori M.B., Hanotte O., *et al.* (2017) Plos One 12. https://doi.org/10.1371/journal.pone.0179021.

Wiener P., Robert C., Ahbara A., Salavati M., *et al.* (2021) GBE 13. https://doi.org/10.1093/gbe/evab014.

Yurchenko A.A., Daetwyler H.D., Yudin N., Schnabel R.D., *et al.* (2018) Sci. Rep. 8. https://doi.org/10.1038/s41598-018-31304-w.

413. Phenotypic characterisation of African chickens raised in semi-scavenging conditions

K.M. Morris[1*], K. Sutton[1], M. Girma[2], W. Esatu[2], B. Solomon[2], T. Dessie[2], L. Vervelde[1], A. Psifidi[1,3], O. Hanotte[1,2,4] and G. Banos[1,5]

[1]Centre for Tropical Livestock Genetics and Health, The Roslin Institute, University of Edinburgh, Easter Bush Campus, Midlothian, EH25 9RG, United Kingdom; [2]LiveGene, International Livestock Research Institute (ILRI), P.O. Box 5689, Addis Ababa, Ethiopia; [3]The Royal Veterinary College, Hawkshead Lane, Hatfield, Hertfordshire, AL9 7TA, United Kingdom; [4]Cells, Organisms and Molecular Genetics, School of Life Sciences, University of Nottingham, Nottingham, NG7 2RD, United Kingdom; [5]Scotland's Rural College (SRUC), Animal and Veterinary Sciences, Easter Bush, EH25 9RG, Midlothian, United Kingdom; katrina.morris@roslin.ed.ac.uk

Abstract

In sub-Saharan Africa, most poultry production is traditional with birds being raised by smallholders in free-range semi-scavenging conditions. The aim of our project is to extensively characterise phenotypes of chickens raised in typical African farming conditions, by measuring production, immunity and survival characteristics. In total, 2,573 chickens were raised in five batches in the poultry facility at ILRI in Ethiopia. These chickens were phenotypically characterised and sampled across an eight-week period. Traits measured included weekly body weight, growth rate, breast muscle weight in carcass, mortality/survival, and immunological titres. The population of chickens had extensive variance at these phenotypes. For body weight, 65% of the total phenotypic variance was attributed to the individual birds providing an excellent source of variation for identifying potential selection markers. This data will subsequently be used along with whole genome sequencing data of these birds to identify selection targets to underpin future breeding programs.

Introduction

In recent years much progress has been made in identifying selection signatures in the genomes of poultry to improve poultry production. This has been primarily focussed on commercial species used in large farming enterprises in developed nations (Wolc *et al.* 2016), with huge gains made in growth rate and feed efficiency in the past few decades (Zuidhof *et al.* 2014). In Africa, 80% of poultry production is in the form of smallholder farms, where chickens are typically raised in free-ranging semi-scavenging conditions (Sonaiya 2008), but there is limited knowledge of the phenotypic characteristics and genetics of these chickens. Phenotypic characterisation and identification of genomic selection targets is necessary for the genetic improvement of chickens adapted to extensive semi-scavenging conditions.

In this study we aimed to comprehensively characterise the phenotypes of typical African dual-purpose village chickens, raised in simulated semi-scavenging conditions in Ethiopia, focusing on meat production, immunity and survival.

Materials & methods

Bird trials, phenotyping and sampling. A total of 2,573 T4451 Sasso birds, a dual-purpose chicken, were raised in five batches of approximately 500 birds across the span of a year at the poultry facility of the International Livestock Research Institute in Ethiopia. The birds were raised in outdoor, semi-scavenging conditions (feed mainly from scavenging with some supplementation from day 56 of age, for approximately 8 weeks until they reached an average market weight of approximately 1,500 g (except batch 1 which was raised until 1000 g was achieved). Body weight was recorded weekly. At the beginning of the experiment, blood (from wing vein) and cloacal swabs were collected for genotyping and immune phenotyping, and the

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_413, © K.M. Morris *et al.* 2022

sex of the birds was recorded. Across the trial period the health of the birds was monitored, disease episodes (including coccidiosis) were recorded, and the day and cause of death were recorded, when applicable. At the last day of the experiment (day of slaughter), blood, cloacal and buccal samples were collected for additional immune phenotyping. Breast muscle was excised and weighed, and liver, heart, spleen and ileum samples were collected and stored in RNA-later and frozen at -80 °C for future transcriptomic studies.

Immune phenotyping. Blood from 2,573 birds from day 56 and 2,097 birds from the day prior to slaughter, was allowed to coagulate overnight at room temperature prior to removal of serum. Serum samples were stored at -20 °C. The cloacal samples from the same birds were retrieved using floxed swabs that were subsequently placed in 500 µl of PBS and stored at -20 °C. Serum anti-NDV titres were analysed using commercial IDEXX NDV ELISA kits (serum dilution 1:100). Cloacal samples were analysed for total IgA levels using direct in house developed sandwich ELISA.

Statistical analysis. A mixed model was used to assess the impact of the batch, age and sex on body weight and estimate the proportion of variance attributed to individual birds. A fixed effect model was used to examine the impact of the same effects on the other traits. Model analysis was conducted in ASReml-W 4.2.

Results

Data are summarised in Table 1. The data demonstrated that the chickens were phenotypically diverse with extensive variance in body weight, growth rate and breast muscle weight.

The key quantitative traits measured all significantly differed by batch and age, likely due to variation in season across the batches (Table 2). In addition, sex differed significantly for weight-related traits, as

Table 1. Data summary of key traits from chicken trials.

Batch[1]	F/M[2]	1	2	3	4	5	Total
No. birds		511	520	520	507	515	2,573
No. mortalities		152	21	75	85	175	511
Sex ratio (%F)		79.7	58.5	60.0	61.7	59.4	62.2
Body weight – start (g)	F	495(85)	516(53)	544(95)	525(113)	566(92)	527(93)
	M	548(89)	566(71)	638(88)	671(142)	662(87)	624(109)
Body weight – end (g)	F	1,092(176)	1,434(190)	1,474(259)	1,280(289)	1,269(233)	1,312(271)
	M	1,220(195)	1,572(257)	1,786(314)	1,528(322)	1,543(265)	1,572(319)
Growth rate (g/d)	F	21.1(4.4)	16.4(3.1)	13.8(3.3)	13.3(4.0)	12.5(3.6)	15.6(4.8)
	M	23.8(4.9)	18.0(4.1)	17.2(4.2)	15.3(4.5)	15.7(4.4)	17.3(5.0)
Breast muscle weight	F	173(38)	266(45)	236(54)	201(58)	217(53)	219(59)
	M	190(36)	289(56)	270(59)	226(61)	246(47)	253(63)
Start NDV titres		0.35(0.24)	0.29(0.23)	0.16(0.15)	0.61(0.44)	0.31(0.42)	0.34(0.30)
End NDV titres		0.12(0.13)	0.15(0.22)	0.13(0.12)	0.17(0.14)	0.97(0.28)	0.35(0.36)
Start IgA titres		0.27(0.42)	0.57(0.61)	1.09(0.81)	0.45(0.62)	0.86(0.76)	0.65(0.72)
End IgA titres		0.97(0.58)	0.86(0.93)	0.46(0.60)	1.21(0.81)	1.41(0.86)	0.87(0.83)

[1] Mean(sd).

[2] Female/male.

Table 2. Significance (F-stat) of fixed effects on key quantitative traits.[1]

	Sex	Age	Batch
Body weight	322.65**	242.28**	136.38**
Growth rate	160.58**	355.22**	234.74**
Breast muscle weight	109.88**	152.8**	176.61**
Start NDV titre	5.36*	93.25**	109.34**
End NDV titre	1.84	113.97**	1,351.61**
Start IgA titre	5.14*	152.61**	93.51**
End IgA titre	0.04	111.09**	83.67**

[1] * = P<0.05, ** = P<0.001.

expected, though had less impact on the NDV antibody and IgA titres. The proportion of phenotypic variance in body weight attributed to the individual bird was 65.3% (±0.8%).

Across the five batches, 511 mortalities were recorded for a range of reasons including predation, huddling and disease (primarily coccidiosis infection) (Figure 1), which are common causes of mortality in smallholder African farms. The cause and rate of mortality varied by batch, particularly in terms of predation and diseases.

NDV titres showed extensive variance across the samples, indicating a range in immune response to NDV vaccination (Figure 2). This varied both within and between batch, with titres falling across the test period in batches 1-4 but increasing in batch 5. IgA titres likewise showed strong variance across the dataset. Batch 5 also experienced a large outbreak of the parasitic disease coccidiosis, and NDV titres at the beginning of the experiment were found to be significantly associated with coccidiosis mortality ($P<0.01$).

Discussion

In this study we phenotypically characterised over 2,500 closely monitored chickens raised in simulated semi-scavenging smallholder village conditions in Ethiopia. This includes key production and health traits

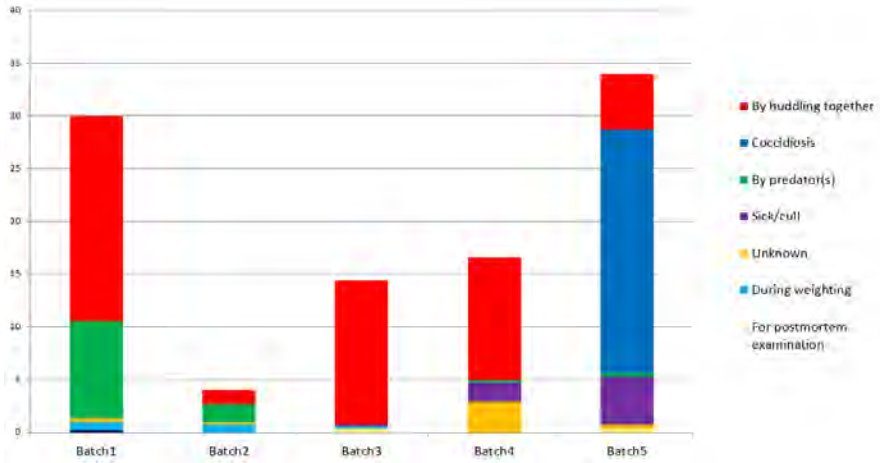


Figure 1. Percent mortalities by batch and cause.

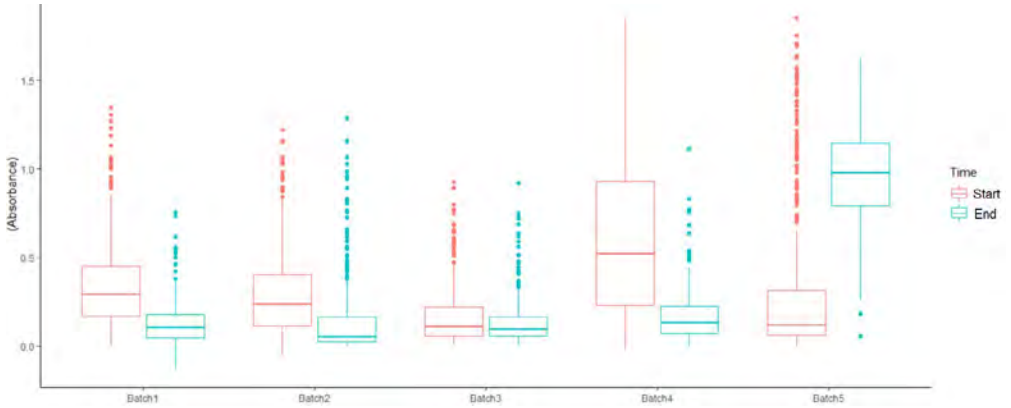


Figure 2. NDV titres of birds at the start and end of the experiment by batch.

including body weight, growth rate, carcass breast weight, disease, mortality, and antibody titres. The analysis of these traits, and selection of chickens adapted to smallholder farm conditions in developing and tropical nations has not achieved the same progress as in the developed world.

We observed extensive variation in the studied traits; we anticipate a proportion of this will reflect genetic variability which can be used for the identification of selection sites for use in breeding programs. Additionally, the substantial number of mortalities observed in this study due to predation and infection will be key for identifying genetic signatures important for survival. These will be critical for African farms as high mortality has been identified as the greatest constraint to poultry production (Sonaiya 2008).

We now aim to genotypically sequence and characterise all of the chickens used in this study. Using the combined phenotypic and genomic data we will identify genes and genomic regions associated with these traits that can be subsequently used as selection targets in breeding programs. In addition, we have identified 48 birds which have the highest and lowest growth and immune trait records, which will be studied using RNA-Seq to further understand the molecular mechanisms underlying these traits. This study will be key for the optimisation of breeding programs and the improvement of poultry in typical smallholder farms in Africa.

Acknowledgements

Funding was from the Bill and Melinda Gates Foundation through CTLGH. Personnel at ILRI farm contributed to bird monitoring and data and sample collection.

References

Sonaiya F. (2008) Smallholder family poultry as a tool to initiate rural development. International Conference Poultry in the Twenty-first Century: avian influenza and beyond: 5–7 November 2007. Bangkok, Thailand: FAO.

Wolc A., Kranis A., Arango J., Settar P., Fulton J.E., *et al.* (2016) Anim. Front. 6 (1):23–31. https://doi.org/10.2527/af.2016-0004

Zuidhof M., Schneider B., Carney V., Korver D., Robinson F. (2014) Poult. Sci. 93:2970–2982. http://doi.org/10.3382/ps.2014-04291

414. Spatial modelling in genetic evaluation of South African Holstein cattle population

I. Houaga[1,2*], T.P. Oliveira[2], I. Pocrnic[2], E. Lavrenčič[3], C.B. Banga[4] and G. Gorjanc[2]

[1]Centre for Tropical Livestock Genetics and Health, Roslin Institute, University of Edinburgh, EH25 9RG, United Kingdom;[2] Roslin Institute, University of Edinburgh, EH25 9RG, United Kingdom; [3]Biotechnical Faculty, University of Ljubljana, Kongresni trg 12, 1000 Ljubljana, Slovenia; [4]Agricultural Research Council (ARC), Old Olifantsfontein Road, Pretoria, 0002, South Africa; ihouaga@ed.ac.uk

Abstract

African dairy production systems are characterized by small herd size and low genetic connectedness between herds. This situation makes it difficult to accurately estimate environmental and genetic effects. We evaluated how accounting for spatial relationship between neighbouring herds impacts genetic evaluation of 305-days milk yield in South African Holstein cattle population. We analysed 305-days milk yield and pedigree data for 98,632 Holstein cows collected between 1957 and 2014 from 1,145 herds. Herd and herd-year effects were modelled as independent, whereas spatial effects were modelled as correlated between regions using the Besag model. The results show a regional variation of estimated breeding values and separation of spatial effect from herd effect. The spatial modelling of herd as random effect showed similar accuracy compared to modelling herd-year. Further studies integrating the genotype-by-environment interactions will be needed to better explore the benefits from spatial modelling in African smallholder dairy farming systems.

Introduction

The significant increase of milk yield in the dairy cattle production sector in developed countries has been associated with selective breeding and herd management strategies (Cole *et al.*, 2020). However, in many low to middle-income countries genetic improvement have led to low or no genetic progress, especially in smallholder dairy production systems, due to a lack of appropriate performance records (Ojango *et al.*, 2019). One of the reasons for this is that African dairy production systems are characterized by small herd size and low genetic connectedness between herds. Such a situation is challenging accurate estimation of environmental and genetic effects (Selle *et al.*, 2020). Genetic evaluation partitions genetic and environmental effects on phenotype, which is critical because similarity between relatives may also be caused by shared environmental conditions. Breeders use statistical models to separate these effects by modelling the genetic effect as random with pedigree or genomic relationships between animals and modelling a herd or herd-season effects as fixed or random contemporary group (Mrode, 2014). Traditionally, the herd effect is modelled by assuming the effect of different farms as independent (Cuyabano *et al.*, 2021). However, nearby farms are likely to have similar climate, soils, management, and even social aspects of production and addressing these aspects with spatial models can improve genetic evaluation (Sæbø and Frigessi, 2004; Cuyabano *et al.*, 2021; Selle *et al.*, 2020).

The aim of this study was to evaluate the impact of modelling regions as correlated effects on genetic evaluation of 305 days milk yield in Holstein dairy cattle population in South Africa.

Materials & methods

The 1,331,015 305-days milk yield (MY305) records from 377,921 South African Holstein cows in 7,121 herds was provided by South African Agricultural Research Council. After the quality control we retained 248,597 MY305 records and a pedigree for 196,778 animals, including 98,632 cows from 1,145 herds. The average number of animals per herd across the whole period was approximately 91 cows per herd (8-2,230).

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_414, © I. Houaga *et al.* 2022

We used postcodes of the herd's location to assign them to administrative regions. The herds were spread over 46 of the 52 South African regions.

Statistical modelling

The baseline model (G_1) for MY305 was accounting for the overall mean, a polynomial function of age within parity, random effect of herd (h), random additive genetic effect of the animal (a), random permanent environmental (p) and random residual (e): $y = Xb + Z_1h + Z_2a + Z_3p + e;\ h \sim N(0, I\sigma_h^2), a \sim N(0, A\sigma_a^2)$, $p \sim N\left(0, I\sigma_p^2\right)$ and $e \sim N(0, I\sigma_e^2)$ where X is the design matrix associated with the vector of fixed effects b, Z_1, Z_2 and Z_3 are design matrices respectively associated with herd (h), additive genetic effect (a), and permanent environmental effects (p), and A is the pedigree relationship matrix. We also tested a second model (G_2), where the herd effect was replaced by the herd-year effect (hy), $hy \sim N\left(0,\ I\sigma_{hy}^2\right)$. We extended models G_1 and G_2 using an intrinsic conditional auto-regressive (ICAR) model component (Besag, 1974) to account for spatial variation associated with herd's region: (1) $GR_i = G_i + R$, where R is the fixed effect of herd's region and i is the index of the models (i=1,2); (2) $GRU_i = G_i + RU$, where RU is an independent random effect of herd's region; and (3) $GRC_i = G_i + RC$, where RC is a correlated random effect of herd's region via neighbourhood-based covariance matrix (Besag, 1974). We defined scaled (GRC_1) and unscaled ($GRCNS_1$) inverse of the neighbourhood-based covariance matrix using R-INLA package (Rue *et al.*, 2009). The models were fitted using the BLUPf90 suite of programs (Misztal *et al.*, 2018). The AIC criteria was used to select the most parsimonious model. Moreover, accuracies of estimated breeding values (EBV) were calculated as sqrt(1-PEV/σ^2_g) where PEV is the prediction error variance of EBVs and σ^2_g the additive genetic variance.

Results

Variance components. Variance components and heritability estimates are presented in Table 1. Heritabilities in G_1 models were higher than those from G_2 models (Table 1). The additive genetic, permanent environmental, and residual variances were similar between G_1 models. The same trend was observed between G_2 models. However, the additive genetic and residual variance were higher in G_1 models compared to G_2 models due to large difference between herd and herd-year variance estimates. On the other hand, the permanent environment effect was higher in G_2 models than in G_1 models. The herd effect variance was smaller in GR_1 than in model G_1 since the herd effect in model G_1 captured the regional effects.

Table 1. Variance components and heritability[1].

Model	σ^2g	σ^2h	σ^2hy	σ^2pe	σ^2r	σ^2p	h^2
G_1	836 (23)	1,597 (69)	_	338 (17)	_	4,536	0.18
GR_1	836(23)	1,402 (62)	_	338 (17)	_	4,340	0.19
GRU_1	836 (23)	1,404 (62)	_	338 (17)	205 (64)	4,550	0.18
GRC_1	836 (23)	1,411 (63)	_	338 (17)	217 (80)	4,568	0.18
$GRCNS_1$	836 (23)	1,411 (63)	_	338 (17)	464 (172)	4,815	0.17
G_2	504 (17)	_	2,580 (28)	345(13)	_	4,741	0.11
GR_2	507 (17)	_	2,192 (25)	345 (13)	_	4,365	0.12
GRU_2	507 (17)	_	2,191 (25)	345(13)	646 (147)	5,009	0.10
GRC_2	508 (17)	_	2,190 (25)	345 (13)	1,010 (228)	5,372	0.12
$GRCNS_2$	508 (17)	_	2,191 (25)	345 (13)	2,158 (488)	6,520	0.08

[1] Variances are in 10^3 and standard errors in brackets, σ^2_g: animal additive genetic variance, σ^2_h: herd variance, σ^2_{hy}: Herd-year variance, σ^2_r: regional variance, σ^2_p: phenotypic variance and h^2: heritability.

GRU models had a similar estimate of regional variance as the GRC models and scaling was important to get interpretable estimates (GRCi vs GRCNSi).

Regional effect estimates. Figure 1 presents region estimates using GRC_1 and GRC_2 models. From the GRC_1 model, regions with low estimates (indicated in blue) were observed in the central-western and eastern parts of South Africa (Figure 1, left). Most of the regions with low estimates are neighbouring regions in the semi-arid part of the country with high variability in temperature and precipitations. Regions with medium estimates (indicated in purple) were primarily found in the country's northern, south-central, and south-eastern parts. The few regions with high estimates (indicated in red) were in the country's southwestern and eastern parts. GRC_2 model showed similar observations (Figure 1, right).

Regional variation of estimated breeding values (EBV) and accuracy. Figure 2 presents the regional variation of EBV measured by the standard deviation (SD, left) and prediction error variance (PEV, right). The region in the northern part with a very low SD of EBVs had a very high PEV. Similarly, the region in the western part with a low SD of EBVs had a very high PEV. However, most of the regions in the country with medium to high SD of EBV had low to medium PEV. The same trend was seen for GRC_2 models (results not shown). The few regions with low SD of EBV and high PEV suggest that there was not much information on animals in those regions and the difference in distributions is more due to the amount of information. Considering the EBVs of all animals in the pedigree, similar accuracies were observed across the G_1 and G_2 models (results not shown).

Discussion

The study evaluated the impacts of modelling spatial variation via neighbouring herd's region on genetic evaluation of MY305 compared to treating herds as entirely independent units. The results show that spatial modelling separated spatial effect from herd effect when treated as correlated and did not negatively impact prediction accuracy. Regional distribution of EBVs was observed. A recent study showed that modelling proximity between herds better separated genetic effects from environmental effects and increased

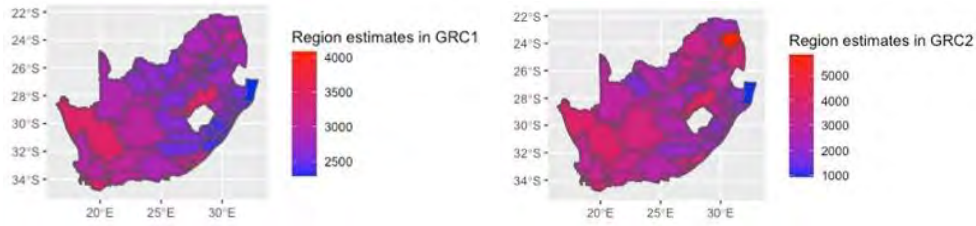


Figure 1. Regional South Africa effect estimates in GRC_1 model (left) and GRC_2 model (right).

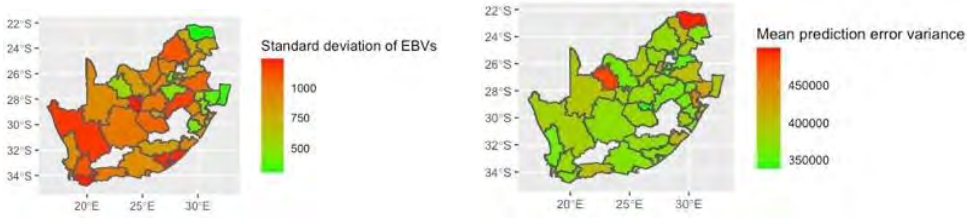


Figure 2. Standard deviation of EBVs (left) and mean of prediction error variance of EBVs (right) in GRC_1 model within south Africa

prediction accuracy in smallholder breeding programmes (Selle *et al.*, 2020). The spatial modelling increases accuracy because it induces environmental connectedness and hence accurately separates a proportion of herd effect that other models may assign to breeding values. Our initial analysis showed that modelling herds using a spatial correlation structure between herd's regions using the Besag model (Besag, 1974) does not increase the prediction accuracy. A recent study on modelling proximity between herds showed increased reliabilities of genomic EBV, but no increase in prediction accuracy of phenotypes (Cuyabano *et al.*, 2021). However, a deeper understanding of data structure, genetic connectedness between herds, the frequent movement of cows between herds, typical to the African livestock production system, is the next important step in this study.

Significant region effects were observed in the present study. These effects are likely due to differences in climatic conditions across regions. To account for large scale environmental effects, Sæbø and Frigessi (2004) modelled veterinary districts as correlated via neighbourhood in genetic evaluation of mastitis resistance in Norwegian Cattle and discovered regional effects in the south-eastern parts of Norway. Despite the benefits from regional modelling in previous studies, coordinate-based models would account more precisely for spatial relationships between farms (Selle *et al.*, 2020; Cuyabano *et al.*, 2021). Furthermore, the current study ignored the genotype-by-environment interaction effect, which are likely significant given the wide range of environmental differences in South Africa.

Conclusions

This study evaluated incorporating region-based spatial effects in genetic evaluation of 305-days milk yield in South African Holstein. Results showed large differences between regions, in some cases more than double, indicating the importance of modelling environmental variation. Further work is required to model environmental variation at a herd level by using a coordinate-based model and possibly expand that model with genotype-by-environment interaction.

References

Besag J. (1974) Spatial Interaction and the Statistical Analysis of Lattice Systems. *JSTOR*, *36*(2), 192–236. http://www.jstor.org/stable/2984812

Cole J. B., Eaglen S.A.E., Maltecca C., Mulder H.A., and Pryce J.E. (2020) Anim. Front. 10(2):37–44. https://doi.org/10.1093/af/vfaa007

Cuyabano B.C.D., Rovere G., Lim D., Kim T.H., Lee H.K., Lee S.H. and Gondro C. (2021) Animals 11(7):2050. https://doi.org/10.3390/ani11072050

Misztal I., Lourenco D., Aguilar I., Legarra A., and Vitezica Z. (2018) Manual for BLUPF90 family of programs. http://nce.ads.uga.edu/wiki/lib/exe/fetch.php?media=blupf90_all7.pdf

Mrode R. (2014) Linear models for the prediction of animal breeding values (3rd ed.). CAB International

Ojango J.M.K., Mrode R., Rege J.E.O., Mujibi D., Strucken E.M., *et al.* (2019) JDS 102(6):5266–5278. https://doi.org/10.3168/jds.2018-15807

Rue H., Martino S., and Chopin N. (2009) J R Stat Soc Series B. 71(2):319–392. https://doi.org/10.1111/j.1467-9868.2008.00700.x

Sæbø S., and Frigessi A. (2004) GSE 36(5):527–542. https://doi.org/10.1186/1297-9686-36-5-527

Selle M. L., Steinsland I., Powell O., Hickey J. M., and Gorjanc G. (2020) GSE 52(1):69. https://doi.org/10.1186/s12711-020-00588-w

415. Testing phenotypes for degree of resilience using fluctuations in milk yield of dairy cows in sub-Saharan Africa

R.D. Oloo[1,2*], C.C. Ekine-Dzivenu[2], J.M.K. Ojango[2], R. Mrode[2], A.O. Mwai[2], and M.G.G. Chagunda[1]

[1]Department of Animal Breeding and Husbandry in the Tropics and Subtropics, University of Hohenheim, Garbenstr 17, 70593 Stuttgart, Germany; [2]International Livestock Research Institute, P.O. Box 30709, 00100 Nairobi, Kenya; richard.oloo@uni-hohenheim.de

Abstract

Despite the relevance of dairy production in the fight against food insecurity and unemployment in sub-Saharan Africa (SSA), negative effects of climate change and general changes in the production environment pose huge challenges to its profitability. Thus, there is a need to improve resilience capacity of dairy animals to adapt to this changing environment. In the current study, we tested two indicators of resilience, log-transformed variance (LnVar) and Skewness (Skew) of deviation, based on fluctuations in animals' milk yield. Further, we assessed the effects of genotype, agroecological zone, and genotype by agroecological zone (G×E) interaction for these phenotypes. Cows with less than 50% of exotic genetics had higher degree of resilience ($P<0.05$). Cows performing in semi-arid zones had higher resilience capacity compared to those in semi-humid environment ($P<0.05$). G×E did not significantly influence both indicators. The results provide valuable information that would inform dairy cattle improvement initiatives in SSA.

Introduction

An animal's degree of resilience to a disturbance is its capacity to be minimally affected by the perturbation or rapidly return to the state pertained before exposure to the disturbance. The performance of the resilient animal need not be the same as without a disturbance, but rather, the negative change in its performance should be relatively lower compared to non-resilient individuals performing in the same conditions. Dairy production in sub-Saharan Africa (SSA) is still low compared to the demand. As a result, pressure for genetic improvement for milk production is still accumulating. However, in the wake of climate change, general environmental changes and their impact, the focus of dairy production must shift from increased production to efficiency and sustainability of milk production. Sustainable dairy production practices that ensure food security to the growing population and overcome the negative impacts of climate change on dairy cattle need to be adopted. SSA is confronted with environmental disturbances, most of which are causes of nature that cannot be modified in the favor of the cattle through good husbandry practices. Therefore, one robust way into this is through breeding for resilience in dairy cattle to enhance their ability to withstand environmental stressors and maintain optimal production levels. Deviations from the normal performance has been utilized to derive indicators that can quantify the general resilience of the animals (Berghof, *et al.*, 2019; Elgersma *et al.*, 2018; Poppe *et al.*, 2020). However, these methods have so far not been applied in SSA. The current study aimed to test two indicators of resilience based on deviation in milk yield and to assess the effect of genotype, agroecological zone and their interaction on these indicators.

Materials & methods

Data used in this study came from dairy cows from three different herds, each representing one of the following agroecological zones of Kenya: semi-arid arable (SAA), semi-arid pasture-based (SAP), and semi-humid (SH). All the herds are kept under extensive dairy production system with occasional supplementation feeding for the lactating cows. Rotational cross breeding is adopted for two herds performing in the semi-arid regions. The herd in semi-humid zone is made up of a stable intermating population of composite cattle originating from crossbred parents.

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_415, © R.D. Oloo *et al.* 2022

Data. The original data set for this analysis contained 2,640 lactations with 62,321 bi-monthly milk yield records of 1,490 multibreed cows from three large-scale dairy farms in Kenya born between January 2000 to December 2017. Data for the first parity were extracted and assessed for quality before the analysis. Cows that had at least two breed types in their genetic make-up based on the information provided by the farmers were used. All cows were required to have 5 or more milk records per lactation and all records were used up to 400 days after calving. To correct for season and year of calving, contemporary grouping of year-season (YS) was done with 17 possible years of calving (2003-2019) and 4 possible seasons. YS groups with less than 5 lactations were excluded from the analysis. After editing the data, 14,278 milk yield records from 745 cows were used for the analysis. These animals were grouped into two genotypic classes based on the proportion of exotic genetics in their breed composition, genotypic class 1 (GC1) (≤50% exotic, n=325) and GC2 (>50% exotic, n=420).

Defining resilience indicators based on fluctuations in milk. The deviations in daily milk yield from the mean production was used to define two resilience indicators: log-transformed variance of deviation (LnVar) and skewness of deviations (Skew). LnVar indicates the impact of the disturbance to the performance of an individual animal. Because resilient animals are less affected by the disturbances in their environments, they have a smaller range of deviation from their average performance. Therefore, they have low variance of deviation. Skew indicates the direction of the deviation and captures the level of severity of the disturbance experienced by an individual animal. Resilient animals have skewness around zero because they have almost equal numbers of negative and positive deviations. The less resilient animals are more influenced by disturbances and thus have more negative than positive deviations which leads to a negative skewness. The LnVar of j^{th} individual was calculated as:

$$ln(variance_j) = ln\left(\frac{\sum_{i=1}^{n_j}(x_{ij}-\bar{x}_j)^2}{n_j-1}\right) \quad (1)$$

where x_{ij} is deviation i of the j^{th} individual, $\bar{x}_j$ is the mean of deviations of the j^{th} individual, and n_j is the number of deviation observations of the j^{th} individual. The skewness of deviation j^{th} individual was calculated as:

$$skew_j = \frac{n_j}{(n_j-1)(n_j-2)}\sum_{i=1}^{n_j}\left(\frac{x_{ij}-\bar{x}_j}{\sqrt{s_j^2}}\right)^3 \quad (2)$$

where n_j is the number of deviation observations of the j^{th} individual, x_{ij} is deviation i of the j^{th} individual, $\overline{\bar{x}}_j$ is the mean of deviations of the j^{th} individual, and s_j is the variance of deviations of the j^{th} individual.

Data analyses. The effects of agroecological zone (environment) and genotype as well as the existence of their interaction in the two resilience indicators was analyzed. A multiple linear regression model with explanatory variables of genotypic class, agroecological zone, genotype by agroecological zone interaction, year-season of calving, day in milk (DIM) class of first day in lactation, DIM class of last day in lactation, age of calving, and squared term of age of calving was used. Analysis of variance was performed prior to fitting linear models to determine the significant factors of variation. The following model was fitted:

$$RI_{ijklmn} = U + G_i + E_j + G_i * E_j + YSC_k + dim.f_l + dim.l_l + Age_m + Age_m^2 + e_{ijklmn} \quad (3)$$

where RI_{ijklmn} is the resilience indicator (LnVar and Skew) measurement for n^{th} animal, U corresponds to the population mean, G_i is the i^{th} genotypic class (i = 1-2), E_j is the j^{th} agroecological zone (j=1-3), $G_i \times E_j$ is the interaction between i^{th} genotypic class and j^{th} agroecological zone, YSC_k is the k^{th} year-season of calving (k=1-53), $dim.f_l$ and $dim.l_l$ are the first and the last DIM classes, respectively of the l^{th} DIM (l=1-10), Age_m and Age_m^2 represent m^{th} age and its squared term, respectively (m=22-60) e_{ijklmn} is the residual error.

Least-square means (LSM) of the genotype classes within the different traits were calculated and contrasted across the 3 different agroecological zones.

Results

The raw average estimate (± standard deviation) for LnVar for the whole dataset was 1.4998±0.8420. The minimum and maximum LnVar values were -1.3587 and 3.8916, respectively. The Skew had a raw mean of -0.05316±0.7186 with a minimum of -2.4986 and a maximum of 3.6525. The significant sources of variation for LnVar were environment, genotypic class, age of the animal, and year season of calving. At population level, GC1 cows had significantly lower LnVar than GC2 cows (P>0.05). This shows that GC1 cows had a higher degree of general resilience than GC2 cows. Agroecological zone influenced LnVar, hence level of general resilience significantly (P>0.001). Animals performing SAA zone had lowest LnVar, followed by those in SAP and SH (P<0.01) (Table 1). Within agro-ecological zones, GC1 had significantly lower LnVar than GC2 cows in SAA (P<0.001) and SH (P<0.05) environments. However, in SAP, there was no significant difference in LnVar for the two genotypes groups (Table 2). The significant sources of variation for Skew were same as those of LnVar except age of the animal and its squared term. However, the skewness

Table 1. Least square mean (± standard error) for log-transformed variance (LnVar) and skewness (skew) of deviation in milk yield at genotypic class and agroecological zone levels.[1]

Variable and level	n	LnVar of deviation	Skew of deviation
% Exoticness			
GC1 (0-50%)	325	0.813±0.257a	-0.0455±0.251a
GC2 (>50%)	420	0.992±0.258b	-0.107±0.252a
Agroecological zone			
Semi-arid arable	352	0.541±0.262a	0.0531±0.256a
Semi-arid pasture-based	169	0.902±0.271b	-0.0721±0.264ab
Semi-humid	224	1.263±0.253c	-0.2099±0.247b

[1] The difference between GC1 and GC2 and among the agroecological zones that are significant at P<0.05 have been shown using different small letters. Similar letter denotes that the difference is not significant. GC1 and GC2 represent genotypic class 1 and genotypic class 2, respectively.

Table 2. Least square means (± standard error) for log-transformed variance (LnVar) and skewness (Skew) of deviation in milk yield of genotypic classes across the three agroecological zones.[1]

Resilience indicator	GC1 (0-50%)	GC2 (>50%)	Pairwise comparison (GC2-GC1)	Significance level
LnVar of deviation				
Semi-arid arable	0.405±0.266	0.678±0.264	-0.2726±0.0808	***
Semi-arid pasture-based	0.92±0.268	0.885±0.299	-0.0345±0.1709	ns
Semi-humid	1.114±0.267	1.413±0.254	0.2989±0.1249	*
Skew of deviation				
Semi-arid arable	0.1205 ±0.26)	-0.0143±0.257	-0.1348±0.0789	ns
Semi-arid pasture-based	-0.0866±0.262	-0.0576±0.292	0.0291±0.1668	ns
Semi-humid	-0.1706±0.26	-0.2492±0.248	-0.0786±0.1219	ns

[1] The difference between GC1 and GC2 that are significant have been shown. '***', and '*' indicate significant differences at P<0.001, and 0.05, respectively. ns denote not significant. GC1 and GC2 represent genotypic class 1 and genotypic class 2, respectively.

of deviation did not detect differences in resilience between the two genotypic classes. The only significant difference in skew was between the animals performing in SAA and those in SH where the animals in former environment had a positive and a closer to zero Skew signifying more degree of resilience than their counterparts ($P<0.01$). Genotype by environment (G×E) interaction was not significant for both resilience indicators.

Discussion

This study used deviations in milk yield to test two indicators of resilience for cows performing in the tropical environment of sub-Saharan Africa. The LnVar was stronger in showing resilience of the animals than the skew based on its ability to discriminate degree of resilience of cows more efficiently. Similar findings were reported in past studies (Berghof *et al.*, 2019; Poppe *et al.*, 2020). Animals with lower exotic genetic proportion had higher degree of resilience possibly due to the fact that they have a high proportion of Zebu genes in their genetic make-up which could have conferred the adaptation to the local production environments. The environment also affected the degree of resilience of the animals. In particular, animals performing in semi-arid zones, which are known to have many disturbances related to high temperatures and long periods of dry seasons, had better resilience capacity than those in semi-humid zone. Constant exposure of the animals to the disturbances in semi-arid zones could have activated their innate regulatory pathways and bettered their chances to survive environmental adversities in the long run (Colditz and Hine 2016). G×E interaction was not significant for both resilience indicators. This implies that a resilient genotype is capable of performing in a wide range of environments without its resilience capacity being altered. This study has shown the possibility of utilizing the deviations in milk yield to quantify general resilience of dairy cows performing in the tropical environment of SSA, where climate change is already affecting dairy production.

References

Berghof T.V.L., Bovenhuis H., and Mulder H.A. (2019) Front. Genet. 10(December):1–15. https://doi.org/10.3389/fgene.2019.01216

Colditz I.G., and Hine B.C. (2016) Anim. Prod. Sci. 56(12):1961-1983. https://doi.org/10.1071/AN15297

Elgersma G.G., de Jong G., van der Linde R., and Mulder H.A. (2018) J. Dairy Sci. 101(2):1240–1250. https://doi.org/10.3168/jds.2017-13270

Poppe M., Veerkamp R.F., van Pelt M.L., and Mulder H.A. (2020) J. Dairy Sci.103(2):1667–1684. https://doi.org/10.3168/jds.2019-17290

416. Egg production and genetic parameters for Tanzanian local chickens

T. Magwaba[1*], E. Mollel[2], J.R Mushi[2], M. Walugembe[1], E.N Amuzu-Aweh[3], G.H. Chiwanga[2], N. Chouicha[4], P.L. Msoffe[2], T. Kelly[5], R.A. Gallardo[5], H. Zhou[4], A. Wolc[1,6], S.J. Lamont[1], A.P. Muhairwa[2] and J.C.M. Dekkers[1]

[1]Iowa State University, Ames, IA 50011, USA; [2]Sokoine University, P.O. Box 3000 Chuo Kikuu, Morogoro, Tanzania; [3]University of Ghana, P.O. Box LG 25 Legon, Accra, Ghana; [4]University of California, Davis, CA 95616, USA; [5]School of Veterinary Medicine, University of California, Davis, CA 95616, USA; [6]Hy-Line International, Dallas Center, IA 50063, USA; tapiwam@iastate.edu

Abstract

Local chickens play an important role in food security, nutrition, and livelihoods in rural Africa but they have limited production performance. We quantified egg production of three Tanzanian chicken ecotypes (n=870) and estimated genetic parameters using genotypes from targeted genotyping by sequencing at 5K SNPs that were imputed to 600K SNPs. Age at first egg (AFE) and the number of eggs laid in consecutive 90-day periods after AFE were analysed. The average AFE was 183.0 days and the mean number of eggs per period ranged from 6.8 to 22.3. Estimates of heritability were 0.04±0.10 for AFE and ranged from 0.04 to 0.46 for egg production per period. Genetic correlations among the egg production traits were >0.91 for most trait pairs but had high standard errors. We conclude that egg-laying performance in local Tanzanian chickens is low but can be improved by selection.

Introduction

Local chickens play an important role in the food security, nutrition, and livelihoods in the rural communities of Africa. In most developing countries, chickens are kept for both meat and egg production (Terfa *et al.*, 2019). Seventy-nine percent of the 4.7 million households in Tanzania are reported to keep local chickens, which comprises 56% of the chickens in the country (Ringo, 2018). Annual egg consumption is 1.5 kg per capita (Ministry of Livestock and Fisheries, 2019), compared to 6.2 in South Africa and 15.2 in the USA. Genetic improvement of local chickens for egg production is one way to increase egg consumption per capita. Tanzania has several local breeds or ecotypes, including Ching'wekwe, Kuchi, Morogoro-medium, Pemba, and Unguja (Lyimo *et al.*, 2013). Whilst these local chicken breeds, require low resource input and are adapted to local conditions, they are characterized by low production performance (Pym *et al.*, 2006; Safalaoh, 1997), which cannot match the growing African population and the change of taste and preference towards animal protein and white meat (Henchion *et al.*, 2017). Knowledge of trait genetic parameters forms the basis for genetic improvement. Several studies have been done on the performance of local chickens, but there have been no reports on the egg-laying performance of the local ecotypes evaluated here, i.e. Morogoro-medium, Chingw'ekwe, and Kuchi, which were chosen based on their diversity and prevalence for a large-scale project on resistance to New Castle disease (Walugembe *et al.*, 2019). The objective of this study was to quantify the egg production of these three ecotypes and estimate the associated genetic parameters.

Materials & methods

Egg production data were collected from 2019 to 2021 (103 weeks) on 870 hens from the local Tanzanian ecotypes Kuchi (n=233), Morogoro-medium (n=352), and Ching'wekwe (n=285). Contrary to commercial layers, which are kept in production for ~1 year, African local chickens have a reproductive life span of three and a half years (Mujyambere *et al.*, 2010). The chickens were offspring of parents described by Walugembe *et al.* (2019). Chicks were tagged, weighed at hatch, and raised as one flock. The birds were

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_416, © T. Magwaba *et al.* 2022

separated by sex at three weeks of age. Females were separated into individual cages at the point of lay. The birds were vaccinated against Newcastle disease at 3 and 21 days of age and then every 3 months. Gumboro vaccinations were done at 11 and 18 days of age and Fowl Pox vaccination at 8 weeks of age. Deworming was done every 3 months. The birds were genotyped using targeted genotyping by sequencing at 5K SNPs, which were imputed to the Affymetrix chicken 600K SNP chip, as described in Walugembe *et al.* (2022).

The age of each hen at the time of the first egg (AFE, days) was recorded. Eggs were then collected daily and summarized as the total number of eggs laid during each 90-day period, starting from the point of lay, which will be referred to as periods 1 to 8. Records for periods with zero eggs and those that deviated more than three standard deviations from the population mean for each period were removed. Table 1 shows the remaining number of records for each trait and descriptive statistics.

The following univariate linear model was fitted in ASReml version 4.1 (Gilmour *et al.*, 2015) to estimate variance components for each trait: $y = Xb + Zu + e$, where y is the vector of phenotypes, X and Z are incidence matrices for fixed and random effects, respectively, b is a vector for fixed effects, u is a vector of random genetic effects with covariance structure based on the genomic relationship matrix (VanRaden, 2008), and e is a vector of residuals. The fixed effect was the date of hatch (n=26), along with age at a first egg (not included for AFE) and ancestral population proportions as fixed covariates. Ancestral subpopulation proportions were used instead of ecotype because the three ecotypes were found to be made up of two ancestral populations, as described by Walugembe *et al.* (2019). Bivariate models were run to estimate phenotypic and genetic correlations among the first five egg production traits, along with AFE. The last three egg production periods were excluded because of very low heritability estimates.

Results

The mean AFE was 183 days and the mean egg number per 90-day period generally declined with age from 22.3±5.9 in period 1 to 6.8±2.9 in period 8 (Table 1). Estimates of variance components and heritabilities are in Table 2. The estimate of heritability for AFE was 0.04±0.01. Heritability estimates were low for period 1 (0.04±0.09) but then increased to 0.46±0.26 at period 4, after which it decreased again. The average heritability across the periods was 0.14. The estimate of genetic variance was also highest for period 4 (Table 2), while period 5 had the highest phenotypic variance, and period 1 had the highest environmental variance.

Estimates of phenotypic, genetic, and environmental correlations are in Table 3. Estimates of phenotypic correlations were moderate and positive between consecutive periods but close to zero between non-

Table 1. Descriptive statistics for egg production traits and age at first egg (AFE) after removal of outliers.

Trait	Number	Minimum	Mean	Maximum	Standard deviation
Period 1	624	1	20.3	56	5.9
Period 2	351	1	14.1	53	5.7
Period 3	321	1	12.6	35	5.3
Period 4	312	1	16.3	46	8.9
Period 5	302	1	17.7	58	9.3
Period 6	281	1	13.9	47	5.4
Period 7	243	1	9.8	34	4.1
Period 8	145	1	6.8	23	2.9
AFE	605	79	183.0	426	44.4

Table 2. Estimates (standard error) of variance components and heritabilities for egg production traits and age at first egg (AFE).

Trait	Genetic	Environmental	Phenotypic	Heritability
Period 1	4.2 (9.2)	98.1 (111.3)	102.4 (6.0)	0.04 (0.09)
Period 2	6.8 (12.2)	65.1 (13.8)	72.0 (5.7)	0.10 (0.17)
Period 3	14.9 (12.8)	48.9 (13.6)	63.8 (5.3)	0.23 (0.20)
Period 4	48.7 (27.8)	57.4 (27.7)	106.1 (8.9)	0.46 (0.26)
Period 5	23.8 (22.4)	86.3 (24.0)	110.0 (9.3)	0.22 (0.20)
Period 6	3.6 (8.3)	59.0 (10.2)	62.6 (5.4)	0.06 (0.13)
Period 7	0.2 (5.3)	44.6 (7.3)	44.8 (4.1)	0.00 (0.13)
Period 8	0.0 (0.0)	24.4 (2.9)	24.4 (2.9)	0.00 (0.00)
AFE (days)	43.7 (118.1)	1,191.4 (143.7)	1,235.2 (72.2)	0.04 (0.10)

Table 3. Estimates (standard error) of genetic (above diagonal) and phenotypic (below diagonal) correlations for egg production traits and age at first egg (AFE).

	Period 1	Period 2	Period 3	Period 4	Period 5	AFE
Period 1		0.97 (1.45)	DNC	0.96 (2.05)	DNC	0.87 (4.50)
Period 2	0.41 (0.05)		0.53 (1.10)	0.92 (1.85)	0.99 (0.94)	-0.01 (1.21)
Period 3	0.21 (0.05)	0.33 (0.06)		DNC	DNC	0.60 (3.59)
Period 4	0.08 (0.06)	0.08 (0.07)	0.36 (0.06)		0.99 (0.81)	0.80 (1.66)
Period 5	0.11 (0.06)	0.08 (0.07)	0.12 (0.11)	0.27 (0.06)		0.35 (1.45)
AFE	-0.24 (0.04)	-0.01 (0.06)	-0.05 (0.06)	-0.14 (0.06)	-0.26 (0.06)	

[1] DNC = did not converge.

consecutive periods. Generally, estimates of genetic correlations were greater than 0.92 among all periods, except between periods 2 and 3 (0.53±1.10). However, all genetic correlation estimates had very high standard errors because of limited numbers of observations, and some did not converge. Phenotypic correlations of AFE were negative with all egg production traits, as expected. Conversely, the number of eggs laid had positive genetic correlation estimates with AFE for each period, except for period 2, but standard errors were again very large.

Discussion

The mean egg production was low, as is expected for local chickens. These results are consistent with the findings of Dana *et al.* (2011) for local ecotypes in Ethiopia. The mean AFE of 183 days is within the range of means for AFE of 166 and 190 days reported by (Mujyambere *et al.*, 2010) and Dana *et al.* (2011), respectively. In the univariate analyses, the effects of hatch date and AFE were significant, but the effect of ancestral proportions was not. The estimate of heritability for AFE (0.04±0.10) was lower than the previous report of 0.37 by Acharya *et al.* (1969) but close to the estimate of 0.06±0.15 reported by Dana *et al.* (2011). Heritability estimates differed between successive egg production periods, which is consistent with Wolc *et al.* (2007). The average heritability across the periods was 0.14, which is lower than that published by Biscarini *et al.* (2010) in their study with commercial layers but close to findings by Tongsiri *et al.* (2019) (0.15±0.10) on Thai native chickens. Estimates of genetic correlations among egg-laying periods were very high, except for periods 2 and 3, which had a genetic correlation estimate of 0.57. The traits with high genetic correlations might be controlled by the same genes. This suggests that egg production across periods can be

improved by the selection of birds with high production at an early age. The estimates, however, had very large standard errors because of small data sizes. More reliable estimates will require larger sample sizes. The phenotypic correlation between egg-laying periods and AFE was negative. The standard errors were low, which indicates that the estimates are reliable. Most of the genetic correlations between periods and AFE were positive and high, except for period 2. However, again, standard errors were very high and more data is needed to draw strong conclusions.

Acknowledgements

This study was made possible by the generous support of the American people through the United States Agency for International Development (USAID) Feed the Future Innovation Lab for Genomics to Improve Poultry (cooperative agreement number AID-OAA-A-13-00080). The contents are the responsibility of the Feed the Future Innovation Lab for Genomics to Improve Poultry and do not necessarily reflect the views of USAID or the United States Government.

References

Gilmour, A.R., Gogel, B.J, Cullis, B.R, Welham, SJ and Thompson, R. (2015). VSN International Ltd. http://vsni.de/downloads/asreml/release3/UserGuide.pdf

Acharya, R. M., Dhillon, J. S., & Tiwana, M. S. (1969). British Poultry Science, 10(2), 175– 181. https://doi.org/10.1080/00071666908415756

Biscarini, F., Bovenhuis, H., Ellen, E. D., Addo, S., & van Arendonk, J. A. M. (2010). Poultry Science, 89(9), 1842–1849. https://doi.org/10.3382/ps.2010-00730

Dana, N., vander Waaij, E. H., & van Arendonk, J. A. M. (2011). Tropical Animal Health and Production, 43(1), 21–28. https://doi.org/10.1007/s11250-010-9649-4

Lyimo, C. M., Weigend, A., Janβen-Tapken, U., Msoffe, P. L., Simianer, H., & Weigend, S. (2013). S.A Journal of Animal Sciences, 43(4), 499–510. https://doi.org/10.4314/sajas.v43i4.7.

Mujyambere, V., Adomako, K., Olympio, S. O., Ntawubizi, M., Nyinawamwiza, L., (2010). Poultry Science 101:101547 https://doi.org/10.1016/j.psj.2021.101547

Ringo, J. E., and Mwenda, V., (2018). Kingdom of the Netherlands. https://www.netherlandsworldwide.nl/

Tongsiri, S., Jeyaruban, G. M., Hermesch, S., van der Werf, J. H. J., Li, L., & Chormai, T. (2019). Asian-Australasian J. Animal Sciences, 32(7), 930–938. https://doi.org/10.5713/ajas.18.0690

VanRaden, P. M. (2008). Journal of Dairy Science, 91(11), 4414–4423. https://doi.org/10.3168/jds.2007-0980.

Walugembe, M., Amuzu-Aweh, E.N., K.-S. Lim, K.S., Wang, Y., N. Chouicha, N., *et al.*, (2022) Proc. World Congress Applied to Livestock Production, the Netherlands

Walugembe, M., Mushi, J. R., Amuzu-Aweh, E. N., Chiwanga, G. H., Msoffe, *et al.*, (2019). Genes, 10(7). https://doi.org/10.3390/genes10070546

Wolc, A., Lisowski, M., and Szwaczkowski, T. (2007). Czech Journal of Animal Science. https://doi.org/10.17221/2282-cjas.

417. Expanding the cattle reference graph genome

A. Talenti[1*], J. Powell[1], D. Wragg[1,2], E. Paxton[1], M. Chepkwony[3,4], A. Miyunga[3,4], R. Njeru[3,4], J.D. Hemmink[1,2,3,4], A. Fisch[5], B.R. Ferreira[5], J.A. Hammond[6], A.L. Archibald[1], P. Toye[3,4], T. Connelley[1], L. Morrison[1,2] and J. Prendergast[1,2]

[1]The Roslin Institute, Royal (Dick) School of Veterinary Studies, University of Edinburgh, Easter Bush Campus, Midlothian, EH25 9RG, United Kingdom; [2]Centre for Tropical Livestock Genetics and Health, Easter Bush, Midlothian, EH25 9RG, United Kingdom; [3]Centre for Tropical Livestock Genetics and Health, ILRI Kenya, Nairobi, 30709-00100, Kenya; [4]The International Livestock Research Institute, PO Box 30709, Nairobi, Kenya; [5]Ribeirão Preto College of Nursing, University of Sao Paulo, Ribeirão Preto, SP, Brazil; [6]The Pirbright Institute, Pirbright, Woking, GU24 0NF, United Kingdom; andrea.talenti@ed.ac.uk

Abstract

Recent studies have highlighted several key advantages of graph genomes over standard linear reference genomes. These advantages include improvements in read mapping rates at divergent loci and the ability to more accurately call structural variants. However, the availability of graph genomes that represent the extensive diversity of livestock species remain limited. To address this limitation, we have incorporated 15 cattle genomes into an expanded cattle graph genome, including 8 completely novel high-quality cattle assemblies from 4 divergent breeds (Holstein-Friesian, N'Dama, Boran and Nelore), each with high contiguity (N50>10 Mb). This graph genome incorporates over 250 Mb (9.5%) of novel sequence across the primary chromosomes, providing a better reference representation of the bovine pangenome and a key resource for the livestock community.

Introduction

Recent studies have highlighted the limitations of single linear reference genomes in genomic studies (Hickey *et al.*, 2020). For example, as larger structural variants (SVs) are absent from the reference genome, they cannot be directly genotyped (Hickey *et al.*, 2020). One solution is to use population-specific genomes, that reduce the impact of allelic bias and allow for the investigation of portions of the genome not present in the other references available (Fakhro *et al.*, 2016).

Even though breed specific genomes are useful for single-population studies, they do not provide a solution for multi-breed studies involving distantly related populations. A viable solution is provided by graph genomes, which incorporate variation on a backbone genome in the form of 'bubbles' in a graph (Hickey *et al.*, 2020). Multiple studies have shown how incorporating short variants present in a population leads to large improvements in the resulting genotype calling (Garrison *et al.*, 2018; Pritt *et al.*, 2018; Crysnanto and Pausch, 2020). However, the largest gain comes from graph genomes incorporating structural variants, which allows for large increases in read mapping, as well as allowing for the genotyping of long or complex structural variants represented in the graph itself (Talenti *et al.*, 2021).

In this study, we generated a total of 8 new, high-quality cattle assemblies of 4 divergent breeds. These highly-contiguous, highly complete assemblies have been incorporated, together with 7 further publicly available assemblies, in the largest cattle graph genome so far generated, that will allow us to investigate SV diversity across the cattle pangenome.

Materials & methods

Genomic DNA for the Boran sample was extracted from monocytes using the nanobind big DNA kit from Circulomics. Genomic DNA for the remaining samples was extracted from whole blood using the Paxgene

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_417, © A. Talenti *et al.* 2022

Blood DNA kit. All samples were sequenced on the Pacific Bioscience Sequel IIe platform using the circular consensus sequencing protocol, which generated high-fidelity (HiFi) sequences. We generated 20-35X genome coverage for each sample with >Q20 quality. Resulting raw reads have been assembled using peregrine (v1.6.3; Porubsky *et al.*, 2020). Genome completeness for each assembly has been calculated using BUSCO (v5.2.2; Manni *et al.*, 2021), Merqury (v1.1; Rhie *et al.*, 2020) and custom scripts used to compute the N50, L50 and %GC.

Generation of the graph genome. All 8 novel assemblies (Table 1) were then aligned together with 7 publicly available genomes: Hereford (ARS-UCD1.2), Brahman (UOA_Brahman_1), Angus (UOA_Angus_1), N'Dama (ROS_NDA1), Ankole (ROS_ANK1), Highland (GCA_009493655.1) and Simmental (GCA_018282465.1). This provided a set of 15 diverse assemblies from third generation sequencing technologies, which will allow for a better comparison of highly complicated regions of the genome.

A graph genome was then generated using the CACTUS pangenome pipeline (v2.0.4; Armstrong *et al.*, 2020), specifying the ARS-UCD1.2 genome as reference, and focusing only on the primary chromosomal sequences (autosomes, X, Y and MT). This generated a reference-based graph in VG format, which can be used for downstream analyses.

Novel portions of sequence. After the generation of the graph genome, we identified the non-reference sequence in the dataset. The total length of the graph was extracted using the vg stats -l commands (Hickey *et al.*, 2020), which returns the size of the graph. We then extracted the novel sequences using the nf-GraphSeq workflow (https://github.com/evotools/CattleGraphGenomePaper/tree/master/detectSequences/nf-GraphSeq) described in Talenti *et al.* 2021. This workflow takes a PackedGraph (pg) graph, and iteratively screens the graph for nodes (i.e. fragments of sequences) not represented in any of the Hereford chromosomes. It then extracts and saves them for each of the assembly considered in the study.

Results

The assembly generation created a set of highly contiguous genomes, with contig N50s ranging from 15 to 56 Mb. Estimated completeness ranged from 95.7 to 95.9% with estimated accuracy ranging from 37.2 to 42.4 (Table 1). These numbers are comparable to other assemblies generated using PacBio HiFi sequencing technologies (Vollger *et al.*, 2020).

Table 1. Metrics of the novel assemblies: genome contiguity (N50 and L50), GC content (%), genome completeness as % of complete universal single-copy orthologs and quality values (QV).

Genome	Contig N50 (Mb)	Contig L50 (N)	GC-content	Completeness (BUSCO)	Accuracy (QV)
Nelore 1	52	20	43.80%	95.7%	38.7
Nelore 2	35	24	43.03%	95.9%	40.7
N'Dama 1	53	18	43.10%	95.9%	41.3
N'Dama 2	14	64	42.89%	95.8%	41.6
Holstein 1	31	27	42.90%	95.9%	42.3
Holstein 2	31	24	42.88%	95.7%	42.2
Holstein 3	56	19	43.03%	95.8%	42.4
Boran 1	49	21	43.77%	95.8%	37.2
ARS-UCD1.2	25	32	/	/	48.7

Following the generation of the 8 new assemblies, we proceeded to generate a graph genome, with the objective of identifying the amount of novel sequence found across the assemblies. The graph included sequences from 15 different assemblies, the largest so far, encompassing multiple *B. taurus* and *B. indicus* genomes, as well as a number of representative genomes of breeds from Africa, South America and Europe. The output graph from the CACTUS pangenome workflow was 2.93 Gb long, over 250 Mb (9.5%) longer than the ARS-UCD1.2 primary chromosomal sequence. As may be expected, the indicine genomes contribute the most to the number of novel fragments of sequence and bases in the graph, with over 50 Mb of novel bases in the non-taurine genomes in over 11 M nodes per genome (Figure 1).

Discussion

Progress in the field of graph genomics has made evident the limitations of single linear genomes. Even though multiple studies have highlighted how incorporating short variants that are specific to a population into a graph genome can lead to an improvement in the resulting genotypes, the biggest gain comes from incorporating large structural variations. Incorporating the large structural variants and novel sequences can help in increasing the number of mapped reads, as well as allowing for the genotyping of otherwise inaccessible variants.

In this context, the generation of new assemblies that capture more of the global cattle genetic diversity becomes most relevant, since it allows us to expand the repertoire of accessible sequence in the cattle pangenome. In this study, we generated a total of 8 new cattle HiFi assemblies from 4 divergent breeds, each with high contiguity and completeness. Over half of these are of non-European taurine ancestry, allowing us to characterise otherwise inaccessible portions of the genomes of these populations. In fact, we successfully identify a total of 250 Mb of novel bases, not represented in the current cattle reference genome.

In conclusion, we successfully generated a large number of new genome assemblies, and combined them successfully in a VG graph, opening up the possibility of investigating the role of large rearrangements in global cattle diversity and their effects on phenotype of interest.

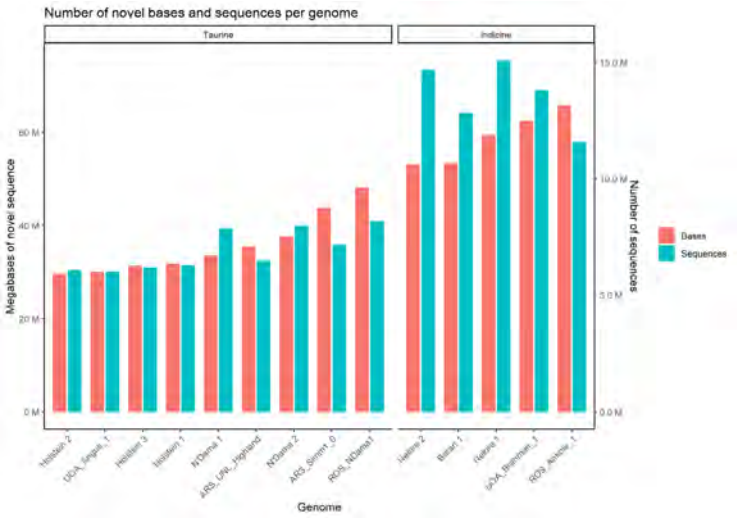


Figure 1. Characterization of the novel sequence identified in the 14 non-Hereford (ARS-UCD1.2) genomes.

References

Armstrong J., Hickey G., Diekhans M. *et al.* – (2020) – Nature 587, 246–251. https://doi.org/10.1038/s41586-020-2871-y

Crysnantp D. and Pausch H. – (2020) – Genome Biol. 21: 184. https://doi.org/10.1186/s13059-020-02105-0

Fakhro K.A., Staudt M.R., Ramstetter M.D. *et al.* – (2016) – Hum Genome Var. 30;3:16016. https://doi.org/10.1038/hgv.2016.16

Garrison, E., Sirén, J., Novak, A. *et al.* – (2018) – Nat Biotechnol 36, 875–879. https://doi.org/10.1038/nbt.4227

Hickey G., Heller, D., Monlong, J. *et al.* - (2020) – Genome Biol 21, 35. https://doi.org/10.1186/s13059-020-1941-7

Manni M., Berkeley M.R., Seppey M. *et al.* – (2021) – Molecular Biology and Evolution, 38-10, pages 4647–4654. https://doi.org/10.1093/molbev/msab199

Porubsky D., Ebert P., Audano P.A. *et al.* – (2020) – Nature Biotech 39, pages 302-308. s41587-020-0719-5

Pritt, J., Chen, NC. & Langmead, B. – (2018) – Genome Biol 19, 220. https://doi.org/10.1186/s13059-018-1595-x

Rhie, A., Walenz, B.P., Koren, S. *et al.* – (2020) – Genome Biol 21, 245. https://doi.org/10.1186/s13059-020-02134-9

Talenti A., Powell J., Hemmink J.D. *et al.* - (2021) – Biorxiv. https://doi.org/10.1101/2021.06.23.449389

Vollger M.R., Logsdon G.A., Audano P.A. *et al.* – (2020) – Annals of Hum Genet 84, 2, 125-140. https://doi.org/10.1111/ahg.12364

418. Genetic structure and diversity of Jersey dairy cattle in Rwanda

O. Opoola[1*], F. Shumbusho[2], R. Mrode[3,4], D. Hambrook[5], M.G.G. Chagunda[6], D. Moran[7] and A. Djikeng[1]

[1]Centre for Tropical Livestock Genetics and Health (CTLGH), The Roslin Institute, University of Edinburgh, Easter Bush Campus, Edinburgh, United Kingdom; [2]Rwanda Agriculture and Animal Resources Development Board (RAB), Kigali, Rwanda; [3]Scotlands' Rural College, Roslin Institute Building, Edinburgh, United Kingdom; [4]International Livestock Research Institute (ILRI), Nairobi, Kenya; [5]Royal Jersey Agricultural & Horticultural Society (RJAHS), Trinity, Jersey; [6]Animal Breeding and Husbandry in the Tropics and Subtropics, University of Hohenheim, Stuttgart, Germany; [7]Global Academy of Agriculture and Food Security, University of Edinburgh, Edinburgh, United Kingdom; oluyinka.opoola@ctlgh.org

Abstract

This study explored the current genetic diversity and structure of Rwanda's dairy cattle population to support the development and use of genomic tools for dairy development and other livestock development initiatives in Rwanda. Dairy cattle in the study (n=1,917) were genotyped at medium density (50K), population structure and diversity analyses were performed. Principal component analysis (PCA), Admixture analysis and measures of expected heterozygosity (He), observed heterozygosity (Ho), minor allelic frequency (MAF) were applied in the study. A dataset of global dairy population of European taurine, African indicus and Zebu breeds (n=250) was used as reference. Results showed Rwanda cattle as a highly admixed population with minor allele frequency of 33% (se=0.001) and 22% (se=0.01) proportion of genes from Jersey Island. The study provides genetic information for future genomic approaches, as well as smallholder farmer-centric decision making tool to enhance productivity and profitability using the Jersey cattle in Rwanda.

Introduction

Genomic selection provides a unique opportunity to optimise genetic gains and breeding potential of animals in the presence of little or no pedigree data records (Thomasen *et al.* 2014) when phenotypes are available. Medium genotyping chips with approximately 50 to 60 thousand single nucleotide polymorphisms (SNPs) have been utilised in global livestock populations for derivation of SNP keys, optimise genetic gains and accuracies of genomic predictions for genomic selection. In Africa, well-structured breeding programs are lacking due to paucity of adequate performance and pedigree data to monitor genetic progress. Therefore, genomic selection in Africa could improve accuracy of selection procedures where data is limited. In addition, productivity gaps of most of the production systems especially, dairy remains significantly high with yield gaps compared to the current levels in advanced economies. The aim of the study was to explore the use of 50-60K SNPs to provide insights to the genetic variation of Rwanda's current dairy cattle population as a smallholder farmer-centric decision making tool to enhance productivity and profitability using the Jersey cattle in Rwanda. We anticipated that the genotyped animals would guide future genomic approaches to support the basis for directional selection for on-farm productivity, genetic progress, healthy and feed-efficient animals that best adapt to the diverse production systems and tropical environment in Africa.

Materials & methods

Jersey cattle of the 'Inka Nziza initiative in Rwanda were selected for genotyping at Geneseek® Neogen, United Kingdom using the Geneseek Genomic Profiler (GGP) 50K SNP array (n=1,917). The Jersey breed of Jersey Island (n=46) were also used as reference population. Other reference breeds were Ankole (Uganda), East African Zebu shorthorn (Kenya), Holstein and non-Island Jersey (West Europe), N'dama (Guinea), Nellore (Brazil), Sheko (Ethiopia), Gir and Sahiwal (India). All SNPs in GGP arrays optimised for use in dairy cattle had the most informative SNPs common to all the platforms for the Illumina Bovine chip®.

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_418, © O. Opoola *et al.* 2022

Data analyses. Quality control of genomic data was performed using the software PLINK (Purcell *et al.*, 2007) considering the following exclusion criteria: non-autosomal SNP, SNP or those with no chromosomal position, minor allele frequency (MAF)≤0.05, call rate (CR)≤0.90, individuals missing more than 90% genotypes (MIND), genotypes missing more than 10% (GENO) of markers, excluding bi-allelic and singletons SNP sites.

Estimation of genetic diversity level and principal components. The SNP filtering based on the Hardy–Weinberg equilibrium (HWE) was performed as we expected HWE deviations in the studied population due to their large and possible diversity and genetic drift. Genetic relationship between and across-breed populations were plotted using the R function 'tidyverse' and its dependencies (Wickham *et al.*, 2019). Average expected heterozygosity (He) was assessed and observed heterozygosity (Ho) were averaged over loci by computing in PLINK under the assumption of HWE (Purcell *et al.*, 2007). Furthermore, the population genetic structure assessment was performed using ADMIXTURE v1.3 (Alexander *et al.*, 2009) to perform cross-validation iterations with a *k*-fold and the increase in log-likelihood between iterations. We performed K=2 to K=12 as ancestral modes in order to identify the optimal number of ancestral populations by detecting the lowest value of cross-validation (CV) error. Admixture output was visualised as barplots in vegan package in R (Jari *et al.*, 2020). After quality control, the total number of genotyped samples and available markers was 1,907 samples and 17,712 SNPs.

Results

Table 1 shows minor allelic frequency, expected, observed heterozygosity within Rwanda dairy population vs reference population.

Population genetic structure, diversity and ancestry. Results of principal component and admixture analyses confirmed the genetic background of Rwanda cattle as an highly admixed (crossbred) population of European taurine; Holstein and Jersey (West Europe); Jersey (Jersey Island) and African indicine and Zebu (Sahiwal, Ankole) with levels of admixture (K=9, 10 & 11) in all the breeding populations.

Figure 1 shows the PCA in Rwanda in comparison to the global dairy population.

Table 1. Sample frequency, minor allele frequency, expected and observed heterozygosity.[1]

Breed	n	H_e	Std	H_o	Std	MAF	Std
ANK	25	0.3	0.001	0.3	0.001	0.2	0.001
EAZ-SH	26	0.3	0.001	0.3	0.001	0.2	0.001
HOL	25	0.4	0.001	0.4	0.001	0.3	0.001
JER_JI	46	0.4	0.001	0.4	0.001	0.3	0.001
JER	25	0.3	0.001	0.4	0.001	0.2	0.001
GIR	25	0.2	0.001	0.2	0.001	0.2	0.001
NEL	25	0.2	0.001	0.2	0.001	0.2	0.001
NDG	24	0.3	0.001	0.3	0.001	0.2	0.001
SHW	13	0.3	0.001	0.3	0.001	0.2	0.001
SHK	16	0.3	0.001	0.3	0.001	0.2	0.001
RWA	2,216	0.4	0.001	0.4	0.001	0.3	0.001
Merged dataset	2,274	0.4	0.001	0.4	0.001	0.3	0.001

[1] H_e (expected heterozygosity); H_o (observed heterozygosity); MAF (minor allelic frequency).

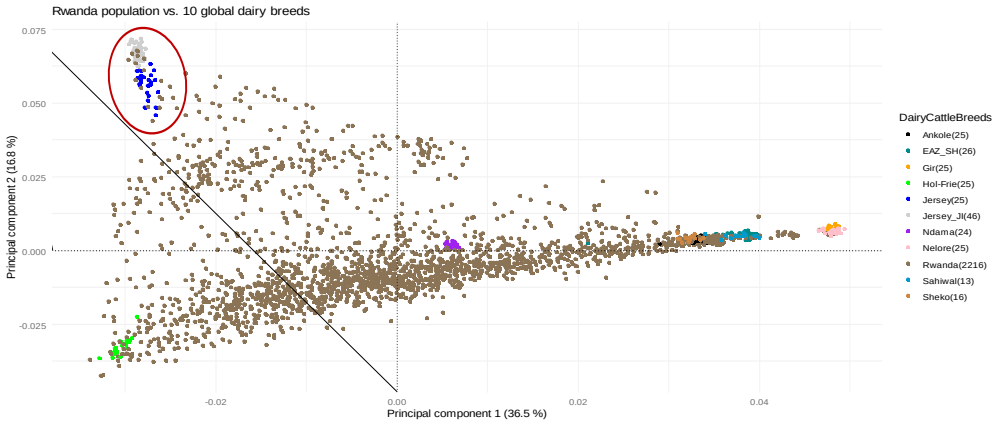


Figure 1. Principal component analysis of Rwanda cattle vs global reference population.

The second principal coordinate vector accounted for 16.8% of the total variation and separated the Island and non-Island Jersey breeds highlighted in the red circle (Figure 1). The first principal coordinate vector accounted for 36.5% of total variation with a significant contribution of European breed (HOL) African *Bos indicus* (NDG), and separated the GIR and NEL breeds. The Rwandan animals dispersed evenly between ANK, SHW, SHK, EAZ-SH and to a lower extent; GIR and NEL breeds. The coefficient of variation (CV) errors for both unsupervised and supervised Admixture learning ranged between 0.60 and 0.65, respectively irrespective of inclusion or exclusion of GIR and NEL breeds. For unsupervised learning, K9 and K11 with NEL and GIR was 0.60. Although not reported here, excluding both breeds had lower but same CV errors at K9 and K11, respectively (0.61; 0.60). For supervised learning with or without GIR and NEL, CV errors were higher than unsupervised learning procedures (K9=0.65; K11=0.64). Figure 2 shows admixture plot at ancestries (clusters) K 9 and 11.

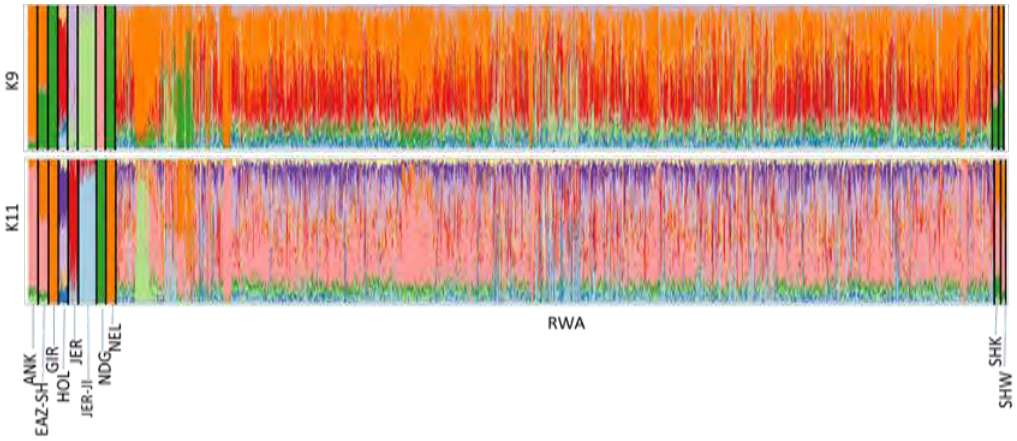


Figure 2. Admixture bar plots showing breed proportions and introgression at selected and assumed ancestry assignment clusters K (9 and 11). Each horizontal bar from left to right, represents Ankole (ANK), East African zebu shorthorn (EAZ-SH), Gir (GIR), Holstein (HOL), non-Island Jersey (JER), Island Jersey (Jersey-JI), N'dama (NDG), Nellore (NEL), Rwanda (RWA), Sheko (SHK) and Sahiwal (SHW).

Average MAF was 34% (±0.0001) and average proportion of exotic Jersey Island exotic genes in the population was ~22% (±0.01) while the greater percentage is of Holstein-Friesian genetics as well as other indigenous African breeds.

Discussion

The Jersey 'Inka Nziza project provided a good platform to assess the genetic diversity of Rwanda's dairy population and generate evidence to support the orientation towards dairy genetics suitable for local production systems. Principal component and admixture analyses confirmed Rwanda cattle as a highly admixed (crossbred) population. The PCA indicated population stratification reflective of sample country of origin, predominantly crossbreds of Holstein breed, followed by Jersey and indigenous breeds of ANK, EAZ-SH, NDG, SHW and SHK. The dispersal of Rwanda animals to the GIR and NEL breed, suggests a contribution of the two breeds albeit to a lower extent. Genetic contributions of N'dama, Gir and European *Bos taurus* breeds have been recently reported in Rwanda (Chagunda *et al.*, 2018). The Holstein-Friesian and Jersey are among the exotic dairy breeds used extensively in pure and crossbreeding in the tropics (Ahunu *et al.*, 1994). Genomic selection has been widely implemented in dairy cattle breeding with the aim to improve performance of purebred animals in developed countries and as well as crossbred cattle (Strucken *et al.*, 2017). By exploring the potential of pure and crossbred animals, there is an untapped potential to select for resilient, productive and ultimately profitable animals by use of genomic tools. Recent studies have demonstrated the opportunities of selection signatures for adaptation traits, disease tolerance parentage assignment and accurate estimation of breed composition (Bahbahani *et al.*, 2017) in cattle in Africa in order to inform future selection of desirable breed traits. The medium-density genomic markers provided insights to the current genetic make-up of Rwanda's dairy cattle population currently practiced in the dairy farming systems. Jersey breed irrespective of the origin showed a diversity of its use within Rwanda. Therefore, we propose that genomic technologies linked to effective dairy breeding systems could close the productivity gaps and more importantly support the development of tools and solutions for use to transform livestock production systems in Africa. This study contributes to a better understanding of the genetic architecture of Rwanda dairy cattle that could best enhance and drive tropical dairy improvement strategies through genomic selection.

Reference

Ahunu B.K., Aboagye G.S., and Arthur P.F. (1994) Proc. of the 5th WCGALP, Ontario, Canada.

Bahbahani H., Tijjani A., Mukasa C., Wragg D., Almathen F., Nash O., et al. (2017). Front. Genet. 8(68):1-20. https://doi.org/10.3389/fgene.2017.00068.

Alexander D., Novembre J., and Lange K. (2009). Genome Res. 19(9):1655-1664. https://doi.org/10.1101/ gr.094052.109.

Chagunda M.G., Mujibi F.D., Dusingizimana T., Kamana O., Cheruiyot E., *et al.* (2018). Front. Genet. 9(438):1-9. https://doi.org/10.3389/fgene.2018.00438.

Jari O.F., Blanchet G., Friendly M., Kindt R., Legendre P., *et al.* (2020). Vegan: community Ecology Package. R package version 2:5-7. Available at: https://CRAN.R-project.org/package=vegan.

Purcell S., Neale B., Todd-Brown K., Thomas L., Ferreira M.A., *et al.* (2007). Am. J. Hum. Genet. 81(3):559-75. https://doi.org/10.1086/519795.

Thomasen J.R., Willam A., Guldbrandtsen B., Lund M.S., and Sørensen, A.C. (2014) J. Dairy Sci. 97(1):458-470, http://doi.org/10.3168/jds.2013-6599.

Strucken E.M., Al-Mamun H.A., Esquivelzeta-Rabell C., Gondro C., Mwai O.A., *et al.* (2017) Genet. Sel. Evol. 49(67):1-18. https://doi.org/10.1186/s12711-017-0342-1.

Wickham H., Averick M., Bryan J., Chang W., McGowan L.D. *et al.* (2019) J. Open Sour. Software, 4(43):1686. https://doi.org/10.21105/joss.01686.

419. Estimation of breed admixture levels in Sri Lanka cattle and its relation to performance traits

M. Spoelstra[1*], J. Sölkner[2], H. Bovenhuis[1], B. Babigumira[2], K. Periasamy[3] and C.M.B. Dematawewa[4]

[1]Wageningen University & Research, P.O. Box 338, 6700 AH Wageningen, the Netherlands; [2]Division of Livestock Sciences, Department of Sustainable Agricultural System, University of Natural Resources and Life Sciences, Vienna, Austria; [3]Animal Production and Health Laboratory, Joint FAO/IAEA Division of Nuclear Techniques in Food and Agriculture, International Atomic Energy Agency, Vienna, Austria; [4]Faculty of Agriculture, University of Peradeniya, Peradeniya 20400, Sri Lanka; mirjam.spoelstra@wur.nl

Abstract

Measuring ancestry levels of local admixed cattle in Sri Lanka provides insight on the most suitable dairy (cross)breeds for the local environment. Performance data of dairy cows (n=197) as well as blood samples for genotyping were collected from two governmental farms in the Kurunegala District in Sri Lanka. Admixture levels were calculated as proportional abundance of the Jersey, Holstein Friesian, and Sahiwal ancestry. The dominance component of heterosis was determined by calculating the abundance of non-homozygous genetic blocks. A mixed linear model was applied to find how performance and admixture levels were associated. For 305-day milk yield, significant predictors were found for Jersey ancestral breed proportion and the dominance component of heterosis for $P<0.05$. Although the feasibility of a crossbreeding program needs to be considered, performance of local Indicine cattle can be improved substantially through crossbreeding with Taurine cattle, mainly through the dominance effects of heterosis.

Introduction

Crossbreeding is a widely used implementation in (sub)tropical countries to increase performance while pertaining favourable characteristics of the indigenous cattle, such as heat tolerance, tick resistance and the ability to perform on low-quality feedstuffs (Galukande *et al.*, 2013). Breeding policies in Sri Lanka have, to large extent, stimulated a continuous grading up of existing cattle using purebred exotic animals (Ministry of Livestock and Rural Community Development, 2010). Continuous upgrading has been reported to decrease performance (Syrstad, 1989) due to a decrease in heterosis as well as the loss of beneficial and necessary key traits, mainly relating to the adaptation of indicine cattle breeds to the local environmental conditions. So far, little research has been performed on breed composition of dairy cattle in Sri Lanka, although identifying admixture levels that correlate with high performance under the local climate(s) of Sri Lanka could substantially improve dairy production.

Materials & methods

Data collection. Phenotypes and blood samples from 197 dairy cows were collected from two governmental farms in Sri Lanka, which were used to estimate the link between performance and admixture level. Phenotypes were collected from 2008 to 2018 and included reproduction data and production data related to milk yield. Ancestry of Sri Lanka cattle were explored in global context and therefore included genotypes of purebred Holstein Friesian, purebred Jersey, local Sri Lankan cattle, purebred Sahiwal as well as several indigenous cattle breeds including Batu Haraka, Thawalam and White cattle. The bulk of genotyping was performed at the laboratories of International Atomic Energy Agency (IAEA) in Seibersdorf (Austria) using an Affymetrix Axiom Bovine chip with a density of 52,518 SNPs. Genotypes of Holstein Friesian and Jersey originated from the International Bovine HapMap Project, genotyped using the Illumina BovineHD chip with a density of 777,962 SNPs (Bovine HapMap Consortium *et al.*, 2009).

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_419, © M. Spoelstra *et al.* 2022

Data editing, selection and admixture analysis. Genotypes were merged and manipulated in PLINK1.9 (Chang *et al.*, 2015), resulting in a total of 35,284 overlapping SNPs. The merged data was explored using multidimensional scaling in the form of a PCA plot and ADMIXTURE analysis (Alexander *et al.*, 2009). To estimate the locus-specific ancestry in the autosomal chromosomes (n=29), data was phased with SHAPEIT v2 (Delaneau *et al.*, 2012) using a cattle recombination map from the Dryad (Ma *et al.*, 2016). Subsequently, locus-specific ancestry was calculated using the Ghap package (Utsunomiya *et al.*, 2020) where the reference populations (k=3) consisted of Holstein-Friesian (n=30), Jersey (n=30) and Sahiwal cattle (n=29). Admixture levels were calculated as proportional abundance of the Jersey, Holstein-Friesian and Sahiwal ancestry for all the generated genetic blocks per animal. Based on the ancestral haplotypes, the dominance component of heterosis was determined by calculating the abundance of genetic blocks of mixed ancestry, following Khayatzadeh *et al.*, 2018).

Mixed linear model. Admixture levels for locus-specific ancestry and the dominance component of heterosis were included in a mixed linear model to estimate how performance and admixture levels were associated. The model was run for milk yield and number of services per conception, using the lme4 package (Bates *et al.*, 2015) in R (R Core Team, 2020) to investigate the breed effect:

$$y_{ijklmnop} = \mu + \beta_1 HF_i + \beta_2 JR_j + \beta_3 Dominance_k + Parity_l + Farm_m + Cow_n + Year_o + e_{ijklmnop} \quad (1)$$

Where: $Y_{ijklmnop}$ is the dependent variable (305-day milk production, number of inseminations), μ is the overall intercept of the model, $\beta_1 HF_i$ is the fixed effect of the proportion Holstein-Friesian admixture level (i=0 to 1), $\beta_2 JR_j$ is the fixed effect proportion Jersey admixture level (j=0 to 1), β_3Dominance$_k$ is the fixed effect of the dominance component of heterosis defined as the percentage homozygous blocks (k=0 to 1). Parity$_l$ is the fixed effect of the k^{th} number of parity (l=1 to 6, where 6 encompasses the parities 6-10*),* Farm$_m$ is the fixed effect of the m^{th} farm (m=Andigama or Maradawila), Cow$_n$ is the random effect of the n^{th} cow (n=1 to 168), Year$_o$ is the random effect of the birth year of the cow (0 = 2005 to 2019), $e_{ijklmnop}$ is the random residual effect with mean 0 and variance σ^2. The model was tested for significant predictors using the lme4 summary statistics and a type III ANOVA test (Wald chi-square test).

Results

Locus-specific ancestry analysis. The supervised locus-specific ancestry analysis used to infer ancestral populations on haplotype level of the admixed cattle on Andigama and Maradawila farm was visualized using Ghap (Utsunomiya *et al.*, 2020, Figure 1). Main breed components consisted of Pakistan Sahiwal (red) and Jersey (green).

Mixed linear model for 305-day milk yield & number of services per conception. The model for 305-day milk yield indicated significant predictors for parity (P=0.0199), Jersey ancestral breed proportion (P=0.0131) and the dominance component of heterosis (P=8.045e-06) for P<0.05 (Table 1). The Holstein-Friesian breed component was showing a high standard error and was indicated as insignificant (P=0.1425) for P<0.05. Both breed components had a negative effect on milk yield, whereas the dominance component of heterosis increased milk yield substantially. The combination of these two components gives us the estimated milk yield response for full heterosis based on dominance effects at haplotype level.

Based on the results of this analysis, a pure Sahiwal would produce 1,076 kg milk at first parity. A first cross Sahiwal × Jersey would result in an estimated milk yield of 1,839 kg for the first parity on Andigama farm (1,076.47 – (789.22×0.5) + 1,156.85) and 2,076 kg for Maradawila farm (1,076.47 – (789.22×0.5) + 1,156.85 + 237.63). A crossbred Sahiwal × Holstein Friesian (not-significant) would result in an estimated milk yield of 1,871 kg for the first parity on Andigama farm (1,076.47 – (724.61×0.5) + 1,156.85) and 2,109 kg

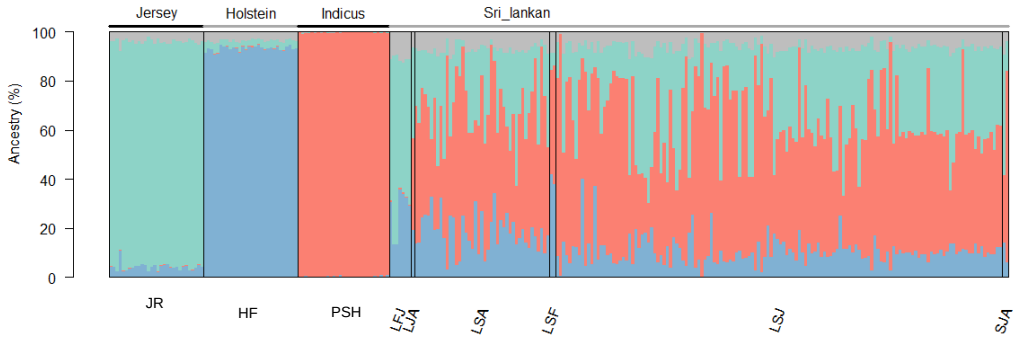


Figure 1. Locus specific ancestry predicted for admixed cattle population based on a reference panel including Jersey (JR: green), Holstein Friesian (HF: blue), and Pakistan Sahiwal (PSH: red). Admixed cattle include crossbred Sahiwal × Jersey × Australian Friesian Sahiwal (SJA); crossbred Sri Lankan Holstein Friesian × Jersey (LFJ); crossbred Sri Lankan Jersey × Australian Friesian Sahiwal (LJA); crossbred Sri Lankan Sahiwal × Australian Friesian Sahiwal (LSA); crossbred Sri Lankan Sahiwal × Holstein Friesian (LSF); crossbred Sri Lankan Sahiwal × Jersey (LSJ).

Table 1. Summary of linear mixed effects model for 305-day milk yield based including the fixed effects: Holstein Friesian proportion; Jersey proportion; parity; farm; and the dominance effect of heterosis.

a. Fixed effects		
	Estimate	**Std. error**
(Intercept)	1,076.47	143.15
Holstein Friesian	-724.61	494.04
Jersey	-789.22	317.93
Parity 2	131.35	62.01
Parity 3	95.04	69.86
Parity 4	10.01	87.27
Parity 5	-106.10	107.54
Parity 6+	-178.19	102.00
Dominance	1,156.85	259.15
Maradawila farm	237.63	91.15

for Maradawila farm (1,076.47 – (724.61×0.5) + 1,156.85 + 237.63). For the model estimating number of services per conception, no significant predictors were found for $P<0.05$.

Discussion

Our findings suggest that an increased response in performance was mainly caused by the dominance effects of heterosis at haplotype level, indicated by the high correlation between the dominance effect of heterosis and 305-day milk yield. Heterosis effects are generally higher in functional traits with low heritability, such as those accounting for reproduction, survival, and fitness, whereas production traits with moderate heritability (e.g. milk yield and growth) are less affected by heterosis. Interestingly, our results showed a higher influence of heterosis on milk yield than on reproduction (number of services per conception). Since the (timing of) insemination of cows is nevertheless under the influence of the farm management, there might be additional influences that we did not account for. Heterosis effects are generally much larger in

crosses between temperate and tropical breeds (Wakchaure *et al.*, 2015. The large genetic distance between the breeds possibly explains the high influence of heterosis on milk yield in our study.

References

Bates, D., Mächler, M., Bolker, B. M., & Walker, S. C. (2015). J. Stat. Softw 67: 1–48. https://doi.org/10.18637/jss.v067.i01

Bovine HapMap Consortium, *et al.* (2009). Science, 324 528–532. https://doi.org/10.1126/science.1167936

Chang, C. C., Chow, C. C., Tellier, L. C. A. M. *et al.* (2015). Gigascience 4:7. https://doi.org/10.1186/s13742-015-0047-8

Delaneau, O., Marchini, J., & Zagury, J. F. (2012). Nat. Methods 9: 179–181. https://doi.org/10.1038/nmeth.1785

Galukande, E., Mulindwa, H., Wurzinger, M. *et al.* (2013). AGR 52:111–125. https://doi.org/10.1017/S2078633612000471

Khayatzadeh, N., Mészáros, G., Utsunomiya, Y. T., *et al.* (2018). J. Anim. Breed. Genet. 135(1), 45–53. https://doi.org/10.1111/jbg.12304

Ma, L., O'Connell, J. R., VanRaden, P. M. *et al.* (2016). PLoS Genet 11:e1005387. https://doi.org/10.1371/journal.pgen.1005387

Ministry of Livestock and Rural Community Development (2010). ISBN: 978-955-0355-01-3. https://www.fao.org/3/bc260e/bc260e.pdf

R Core Team. (2020). https://www.R-project.org/

Syrstad, O. (1989). Livest Prod Sci 23: 97–106. https://doi.org/10.1016/0301-6226(89)90008-0

Utsunomiya, Y. T., Milanesi, M., Barbato, M. *et al.* (2020). Methods Ecol. Evol. 11: 1448–1454. https://doi.org/10.1111/2041-210x.13467

Wakchaure, R., Ganguly, S., Praveen, K. *et al.* (2015). IJAETIS 1: 1-5. http://www.ijaetis.org/asset/images/uploads/144950552993.pdf

420. Economic values for skin grade, days to market and number of hatchlings in the Australian saltwater crocodile industry

S. Hermesch[1*] and S.R. Isberg[1,2]

[1]AGBU, a joint venture of NSW Department of Primary Industries and University of New England, 2351, Armidale, Australia; [2]Centre for Crocodile Research, PO Box 329, Noonamah, Northern Territory 0837, Australia; susanne.hermesch@une.edu.au

Abstract

The saltwater crocodile industry in northern Australia produces skins for high-value leather products. Most eggs are harvested from the wild, providing multiple livelihood benefits to rural and Aboriginal communities in the area. Economic values derived from an independent model for each trait were AU$620.40 for skin grade (a binary trait), AU$-0.95/day for days to market and AU$105.30/hatchling for number of hatchlings per clutch. Economic values were converted to economic weights expressed on the basis of a clutch of the breeding female for a comparison among traits. Based on phenotypic standard deviations of traits, skin grade had the highest emphasis (50%) followed by days to market (38%). Economic values derived in this study should be used to assist management decisions. Setting up structured breeding programs for saltwater crocodiles in northern Australia is less desirable given the wider benefits of harvesting eggs from the wild for rural and Aboriginal communities.

Introduction

The saltwater crocodile industry in northern Australia produces skins for high-value leather products. Saltwater crocodile eggs are collected from the wild, under approved management programs, providing multiple livelihood benefits for rural communities (Fukuda and Webb, 2019, DEPWS, 2020). Harvesting eggs relies heavily on traditional Aboriginal knowledge and royalties are paid for each egg collected, placing economic worth on an apex predator. This sustainable-use model has contributed to full species recovery and conservation of the habitat and co-existing species. Further, the saltwater crocodile industry provides jobs and income to rural communities in remote areas where other income sources are scarce. Fukuda and Webb (2019) also point out 'a key threat, however, to both wild crocodiles and Aboriginal livelihoods, is recent moves by some actors in the fashion industry to stop using wild crocodilian skins.'

The crocodile industry can only provide livelihood benefits to rural communities if it is profitable. Profitability can be increased by improving traits of economic importance. Economic values define economic importance of traits and have been used extensively in animal breeding. However, they may also be used to evaluate alternative management strategies to improve profitability. So far, information about economic values in the crocodilian industry is sparse (Gray *et al.* 2007) and adoption has been hindered by lack of information about input values and specific setup of bio-economic models. The use of independent models (e.g. Amer *et al.* 2014) provides greater flexibility in trait selection for breeding, and management, objectives. The aim of this study was to develop independent economic models to derive economic values for key performance traits in saltwater crocodiles in the Australian context.

Materials & methods

Economic values for skin grade, days to market and number of hatchlings per clutch were derived based on an independent economic model for each trait. The components of these economic models reflect relevant crocodile farming practices which are not widely known and some background information is provided. More detailed information about saltwater crocodile farming and trait definitions was provided by Isberg *et al.* (2005a,b).

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_420, © S. Hermesch and S.R. Isberg 2022

Skin grade. Skins are sold on a 'per cm belly width' basis and returns per crocodile increase linearly for every additional cm in belly width. However, the maximum price received is only obtained for first-grade skins and price is substantially reduced depending on the number and location of any imperfections. Skin grade was defined as a binary trait defining whether skins meet requirements for the first grade (1) or were part of the other grades (0). The economic value for skin grade (EVsg) was derived as: EVsg = (Rsg1 – Rsg0) where Rsg1 was the return for first-grade skins and Rsg0 was the average return for skins in the other three grades. The return per first-grade skin was AU$1,504 based on a payment of AU$40/cm and an average belly width of 37.6 cm (Isberg *et al.* 2005b). The average return for skins from the other grades was AU$883.6 based on prices of AU$30, AU$10 and AU$0 per cm for the second, third and fourth skin grades, respectively, which were assumed to represent the same average dimensions and comprise 15, 2 and 3% of skins.

Days to market. Days to market quantifies the number of days from birth until the age when crocodiles reach market size. Market size is determined by belly width, rather than weight, and saltwater crocodiles are slaughtered when they have reached a belly width of 35 to 45 cm. In northern Australia, saltwater crocodiles hatch from January to July and stay in the hatchling facility until the end of each calendar year, when they are moved to the overall grow-out facilities. Days to market therefore encompasses the hatchling phase and the grower phase. The economic value for days to market (EVdays) was: EVdays = - (Vm – Cm – Ch) / days where Vm was the average market value per crocodile (AU$1,380) based on the assumptions of classification and payments of skin grades outlined above, Cm were the marketing costs per crocodile (AU$ 284) and Ch were costs to produce a hatchling (AU$ 107.3), while days was the number of days to reach market size (1,037.8 days).

Number of hatchlings. Despite the Australian crocodile industry relying heavily on wild egg collection with the primary aims of species conservation and rural livelihood benefits, farms often keep a small nucleus of breeding animals. These breeding crocodiles have been removed from the wild as problem crocodiles, as part of the public relations program of promoting living with an apex predator (DEPWS, 2020). However, no performance selection has occurred and there is an opportunity to optimise future breeder selection and increase farm profitability.

Each female produces one clutch of eggs per year, which are then incubated. Breeding pens vary in their stocking density but mainly consist of one male with one to eight females. Others simulate more natural breeding populations with multiple males and multiple females. Each crocodile farm can produce a fixed number of hatchlings each year due to market constraints and an increase in clutch size would lead to cost savings in the number of females required to produce the target number of hatchlings. The economic value for number of hatchlings (EVnh) was: EVnh = Cf / nh – Clh where Cf was the sum of all annual cost of breeding females (Cf), nh was the number of live hatchlings per clutch (nh: 31.25, Isberg *et al.* 2005a) and Clh was additional labour cost for an additional hatchling (AU$0.96). Costs per female included costs for capital (AU$1,122), labour (AU$1,278) and feed (AU$780) based on one female per male in each breeding pen.

Economic weights. Economic values were derived for two traits of growing crocodiles and a trait of breeding females expressed per clutch, which differ in frequency and timing of expression. These differences were considered through discounted genetic expressions (DGE) to convert economic values of traits of growing crocodiles into economic weights of a breeding objective at the time when number of hatchlings per clutch is expressed. The DGE was derived as: DGE = nh × surv / {(1+r)^(days / 365)} where nh was the number of hatchlings per clutch (31.3), surv was the survival rate of hatchlings until slaughter (81%) and r was discount rate (0.07). The economic weight of each trait was multiplied with the corresponding phenotypic standard deviation to express each trait emphasis relative to the sum of these products. In

animal breeding, this illustration of trait emphasis is usually based on genetic standard deviations which are currently not available for these crocodile traits and phenotypic standard deviations were used instead.

Results

The proportion of first-grade skins was assumed to be 0.8 and the standard deviation of this binary trait was derived (Table 1). First heritability and repeatability estimates for the other two traits were moderate to high. Economic weights expressed relative to the phenotypic standard deviation of traits were highest for skin grade (50%) and days to market (38%, Figure 1).

Discussion

Crocodiles are placed into finishing pens when they are appraised to be close to market size. The number and severity of imperfections on the belly skin determines the amount of time they spend in the finishing pens. Environmental influences such as floor surfaces, hygiene practices and prevalence of mosquito-borne viruses may override expression of underlying genetic variation for skin grade. Given the economic importance of skin grade, this trait should be recorded more closely on farms to identify genetic and non-genetic factors affecting the trait.

Skin grade determines the price per cm belly width obtained for each crocodile skin and a higher belly width leads to higher returns. Further, Gray *et al.* (2007) illustrated a payment system which rewarded larger belly width in the first skin grade while prices for belly width in other skin grades remained constant. This type of payment system can be considered when deriving economic values for both skin grades and belly width through a bivariate distribution of crocodile skins across skin grades and belly-width groups,

Table 1. Data statistics, repeatability or heritability (ratio where known) as well as economic values (EV) and economic weights (EW) for saltwater crocodile traits.

Trait	Mean	SD	ratio	EV	EW [4]
	Traits of growing crocodiles				
First-grade skins (1/0) [1]	0.8	0.40		620.40	12,915.70
Days to market (days) [2]	1,037.8	197.77	0.40	-0.95	-19.83
	Trait of the female (clutch)				
Number of hatchlings [3]	31.3	11.45	0.34	100.80	100.80

[1] Data statistics derived for a binary trait and heritability is unknown.
[2] Data statistics from Isberg *et al.* (2005b), heritability presented.
[3] Data statistics from Isberg *et al.* (2005a), repeatability presented.
[4] The discounted genetic expression to convert EVs of traits of growing crocodiles traits to EW expressed per clutch (AU$/clutch) was 20.82.

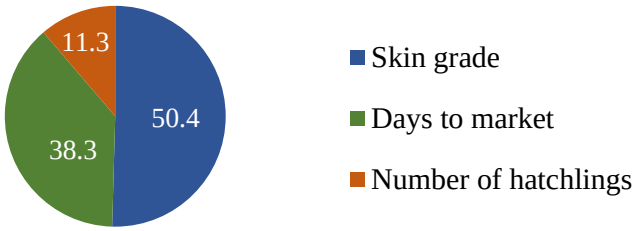


Figure 1. Percentage emphasis of crocodile traits based on phenotypic variation of traits.

as outlined by Hovenier *et al.* (1993) for the univariate case. However, belly width and days to market are associated and trait definitions for belly width and days to market have to ensure that these two traits are independent at both the phenotypic and genetic level. Belly width is similar to market weight in livestock, which is often assumed fixed and not considered separately to days to market in breeding objectives. These considerations need further investigation to model the economic benefits of larger belly width.

Feed costs were not modelled separately and were part of EVdays. Crocodiles require a relatively low amount of feed; e.g. adult breeding crocodiles require only one (females) or two (males) chicken each week. However, costs of feed are not negligible due to high transport and refrigeration costs and should be considered separately as a trait. Of larger consequence is the natural variation of growth rates expressed. Isberg *et al.* (2005b) reported the range in number of days to reach market size to be 634 to 1,859 days. This translates to a difference of 3.4 years in associated production costs, illustrating the potential to improve profitability.

In northern Australia, the majority of eggs to produce hatchlings are sourced from the wild-egg harvest providing income and employment to rural and Aboriginal communities. The cost of wild eggs is similar to the costs of eggs bred in captivity (EVhn) and there is no economic benefit to produce eggs on farms. Given the primary consideration to the income and employment opportunities that wild-egg harvesting provides in northern Australia, economic values derived in this study may best be used to aid management decisions on farms to improve profitability because most farms may not setup a genetic improvement program.

Acknowledgements

Funding from the Centre of Crocodile Research is acknowledged. We thank Kim Bunter, Elena Dehnavi, Pieter Knap and Brad Walmsley for constructive comments.

References

Amer P.R., Ludemann C.I., and Hermesch S. (2014) J. Anim. Sci. 92(12):5345-5357. https://doi.org/10.2527/jas.2014-7943

DEPWS (2020). Wildlife Trade Management Plan- Crocodile Farming in the Northern Territory 2021-2025. https://www.awe.gov.au/sites/default/files/documents/nt-crocodile-farming-wtmp-2021-2025.pdf

Fukuda Y., and Webb G. (2019) Saltwater Crocodile harvest and trade in Australia. Available at: https://cites.org/sites/default/files/eng/prog/Livelihoods/case_studies/CITES_livelihoodsFact_Sheet_2019_Australia_Crocodiles.pdf

Gray E.M., Ahmadi-Eshahani F.Z., and Isberg S.R. (2007) Aust. J. Agric. Resour. Econ. 51:57-72. https://doi.org/10.1111/j.1467-8489.2007.00362.x

Hovenier R., Brascamp E.W., Kanis E., van der Werf J.H.J. and Wassenberg A.P.A.M (1993) J. Anim. Sci. 71:1429-1422. https://doi.org/10.2527/1993.7161429x

Isberg S.R., Thomson P.C., Nicholas F.W., Barker S.G. and Moran C. (2005a) J. Anim. Breed. Genet. 122:361-369. https://doi.org/10.1111/j.1439-0388.2005.00548.x

Isberg S.R., Thomson P.C., Nicholas F.W., Barker S.G. and Moran C. (2005b) J. Anim. Breed. Genet. 122:370-377. https://doi.org/10.1111/j.1439-0388.2005.00549.x

421. Exploiting phenotypic plasticity in animal breeding

F.G. Kebede[1,2*], H. Komen[1], T. Dessie[2], O. Hanotte[2,3], S. Kemp[2], C. Pita Barros[1], R. Crooijmans[1], M. Derks, S.W. Alemu[2,4] and J.W.M. Bastiaansen[1]

[1]Wageningen University & Research, Animal Breeding and Genomics, Droevendaalsesteeg 1, 6708 PB Wageningen, the Netherlands; [2]International Livestock Research Institute, P.O. Box 5689, CMC road, Addis Ababa, Ethiopia; [3]Cells, Organism and Molecular Genetics, School of Life Sciences, University of Nottingham, NG7 2UH, Nottingham, United Kingdom; [4]Massey University, AL Rae Centre for Genetics and Breeding, School of Agriculture and Environment, Palmerston North, New Zealand; fasilgetachew7@gmail.com

Abstract

Livestock populations can have different genetic backgrounds and may vary in their capacity to respond to environmental changes. Our findings suggest that improved chicken breeds differ in growth performance and phenotypic plasticity (yield stability) when they are introduced into new tropical environments. Dual consideration of productivity and phenotypic plasticity gives opportunities to select or recommend genotypes with optimal performance and wider adaptability for smallholder farmers raising livestock in different agroecologies.

Introduction

Phenotypic plasticity refers to the variation in phenotype (e.g. morphological, physiological, behavioural) which results from complex relationships between a genotype and its environment (s). A genotype that shows high phenotypic variation across environments is termed 'phenotypically plastic' while the one that shows low variation is termed 'robust' or 'stable'. Plastic responses for a quantitative trait surfaces as genotype by environment interaction (G×E) during statistical analysis and can be explained through a 'reaction norm'.

Plasticity is thought to be under genetic control. Some genes control both plasticity and mean phenotype while many others are associated only with plasticity (Sieriebriennikov *et al.*, 2018; Lafuente and Beldade, 2019). Phenotypic plasticity can be incorporated as a selectable trait into an animal breeding programme to develop individuals or populations with wider environmental adaptation (De Jong and Bijma, 2002). Plasticity can also be harnessed to screen candidate breeds for introduction at scale to smallholder farmers in the tropics. The objective of the present study is to exploit the concept of plasticity to select and recommend improved chicken breeds with wider environmental adaptation and higher growth performance.

Materials & methods

We distributed five different chicken breeds into three different agroecologies of Ethiopia. The agroecologies were defined following the approaches by Kebede *et al.* (2021) who applied species and phenotypic distribution models. Species Distribution Models (SDMs), implemented in MaxEnt (Phillips *et al.*, 2006) and similar software, predict distribution of a species based on presence-only data, estimate the contribution of environmental predictors, and help identify environmental predictors associated with habitat suitability and phenotypic differentiation in current and future environments. Agroecologies defined through distribution models give a better estimate of G×E compared to conventional agroecologies defined on agronomic attributes (e.g. length of cropping season) and are more appropriate to evaluate phenotypic plasticity in livestock.

A total of 21,562 improved chickens were raised by 2,547 smallholder households and average live body weight of each flock was recorded. The growth performance and plasticity (yield stability) of the five

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_421, © F.G. Kebede *et al.* 2022

chicken breeds until 150-days-of-age (LBW150) across three distribution-model-defined-agroecologies (AEI, AEII, and AEIII) was compared by Additive multiplicative interaction model (AMMI) and mixed-effects (REML)/LMM models.

Results

The phenotypic plasticity for LBW150 among different chicken breeds introduced into different environments based on AMMI model is presented in Figure 1. Performance of the five breeds in the different agroecologies is presented against the first environmental principal component (along the x-axis). The plastic responses of the breeds for LBW150 are characterized by changes in magnitude and rank, signifying the presence of G×E. Breeds with gentler slopes in the reaction norm plots (Sasso, SRIR and Potchefstroom Koekoek) indicate these breeds have wider adaptation (are less plastic). Kuroiler breed, shows a steeper slope (more plastic/less stable) and appears to be a better choice to be introduced specifically into agroecology-I (AEI). Differential growth performance of these breeds in conventionally defined agroecologies was shown by other statistical approaches (Lozano-Jaramillo *et al.*, 2019a,b; De Kinderen *et al.*, 2020; Alemu *et al.*,2021) supporting the presence of G×E.

WAASB (Weighted Average of Absolute Scores) biplots on LMM/REML-based analysis for productivity and plasticity are presented in Figure 2. The breeds or agroecologies included in quadrant I are considered unstable breeds or agroecologies with high discrimination ability, and productivity level below the grand mean. We have Horro in quadrant I. Unstable breeds with productivity above the grand mean are included in quadrant II. The only breed which fell into this category is Kuroiler. Breeds withing quadrant III (SRIR and Koekoek) have close to average productivity but can be considered less plastic (stable) due to the lower value of WAASB. The lower the WAASB stability value, the more stable the breed can be considered. The agroecology included in the third quadrant are considered as unfavourable and having low discrimination ability. The breed within quadrant IV (Sasso) is considered highly productive and broadly adapted (less plastic/more stable).

Discussion

The five chicken breeds showed different levels of growth performance and plasticity for LBW150 in three agroecologies defined by distribution models. That was probably due to differences in their genetic backgrounds (e.g. genes influencing plasticity, selection history, levels of inbreeding). AMMI and LMM

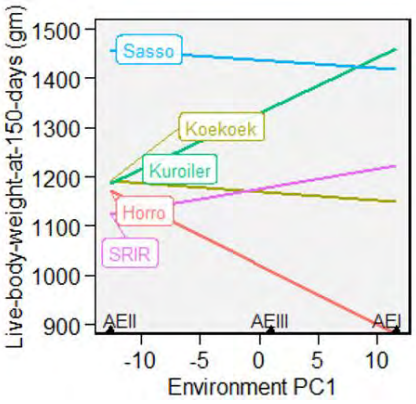


Figure 1. Performance plots showing phenotypic plasticity of five chicken strains for LBW150 against environmental principal component analysed by AMMI model. AEI, AEII, and AEIII, indicate the distance of agroecologies along PC1.

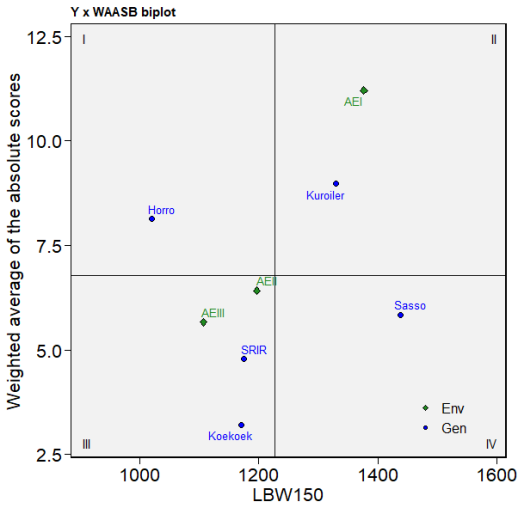


Figure 2. Biplot of LMM analysis for LBW150 showing mean growth performance (x-axis) and plasticity (four-quadrants) of five improved chicken breeds.

detected G×E differently but effectively. But LMM was more sensitive in identifying productive and stable breeds than AMMI. LMM-based models and indexes allowed fitting of random effects and helped generate biplots combining productivity and yield stability. Multi-environment comparison of breeds on productivity and plasticity is essential before new breeds are introduced to suboptimal environments of the tropics. Such comparisons may help identify breeds with high productivity and low plasticity for smallholders.

We recommend that phenotypic plasticity should be harnessed by animal breeders to deal with changes in production environment at present and future scenarios (e.g. climate, health, management levels). In depth studies should be conducted to understand molecular mechanisms controlling plasticity of QTLs. Multi-environment performance models incorporating diverse information (e.g. phenotypic, environmental, and genetic/genomic) should be also developed for comparison of livestock breeds on productivity and plasticity for smallholder systems.

Acknowledgements

The research was funded by the African Chicken Genetic Gains Program (AGGG) of the International Livestock Research Institute (ILRI) with grant received from the Bill and Melinda Gates Foundation (Grant Agreement number: OPP1112198). The authors also acknowledge additional financial support from the Koepon Foundation.

References

Alemu, S. W., Hanotte, O., Kebede, F. G., Esatu, W., Abegaz, S., *et al.* (2021). *Acta Agric. Scand. A Anim. Sci.* 70, 71-77. https://doi.org/10.1080/09064702.2021.1891278

De Jong, G., and Bijma, P. (2002). *Livest. Prod. Sci.* 78, 195-214. https://doi.org/10.1016/S0301-6226(02)00096-9

De Kinderen, M. A., Sölkner, J., Mészáros, G., Alemu, S. W., Esatu, W., *et al.* (2020). *Acta Fytotech. Zootech.* 23. https://doi.org/10.15414/afz.2020.23.mi-fpap.205-213

Kebede *et al.*, K. H., Dessie T, Alemu SW, Hanotte O and Bastiaansen J (2021). *Front. Genet.* 12:723360. https://doi.org/10.3389/fgene.2021.723360

Lafuente, E., and Beldade, P. (2019). *Front. Genet.* 10, 720. https://doi.org/10.3389/fgene.2019.00720

Lozano-Jaramillo, a., Alemu, S., Dessie, T., Komen, H., and Bastiaansen, J. (2019a). *Sci. Rep.* 9, 1-11. https://doi.org/10.1038/s41598-019-51910-6

Lozano-Jaramillo, M., Bastiaansen, J., Dessie, T., and Komen, H. (2019b). A*nim.* 13, 1536-1543. https://doi.org/10.1017/S1751731118003002

Phillips, S. J., Anderson, R. P., and Schapire, R. E. (2006). *Ecol Modell* 190, 231-259. https://doi.org/10.1016/j.ecolmodel.2005.03.026

Sieriebriennikov, B., Prabh, N., Dardiry, M., Witte, H., Röseler, W., *et al.* (2018). *Cell Rep.* 23, 2835-2843. https://doi.org/10.1038/s41598-019-51910-6

422. Whole genome sequencing reveals genetic diversity and heat-stress adaptation in Nigerian indigenous chickens

A.A. Gheyas[1*], M. Rachman[2], O. Bamidele[3], T. Dessie[4], J. Smith[1] and O. Hanotte[2,4]

[1]Centre for Tropical Livestock Genetics and Health (CTLGH), Roslin Institute, University of Edinburgh, EH25 9RG, Edinburgh, United Kingdom; [2]School of Life Sciences, University of Nottingham, NG7 2RD, Nottingham, United Kingdom; [3]African Chicken Genetic Gains (ACGG), Department of Animal Sciences, Obafemi Awolowo University, Ile Ife 220282, Nigeria; [4]LiveGene – CTLGH, International Livestock Research Institute (ILRI), P.O. Box 5689, Addis Ababa, Ethiopia; almas.gheyas@roslin.ed.ac.uk

Abstract

Poultry is a crucial sector for the livelihoods and food security of millions of people in Nigeria. Here we present the first large scale whole-genome sequencing analysis on Nigerian indigenous chickens from different agro-climatic conditions, investigating their genetic diversity and adaptation to tropical hot climates. We observe a large genetic diversity but low levels of population differentiation. Selection signature analyses were performed to identify candidate genes in relation to heat-stress adaptation including those specific to extreme hot-humid or hot-arid conditions. These results have important implications for the conservation of genetic diversity and breeding improvement of chickens for thermo-tolerance.

Introduction

Indigenous livestock populations from different geographic regions constitute important genetic resources for conservation as they represent adaptation to local agro-climatic conditions. Native tropical breeds are particularly crucial, as climate change and global warming is forcing many temperate regions to experience tropic-like conditions and such breeds may hold a genetic solution for climate-resilience.

Nigeria is a tropical lowland country, where poultry farming plays a crucial role in the economy and the livelihood of millions. Nigeria has one of the largest stocks of chickens in Africa, but over half of its birds are still raised in extensive backyard farming systems, which are predominantly represented by unimproved indigenous breeds. Even though these local unimproved breeds have poor productivity, they show superior adaptive ability to harsh tropical climates. The Nigerian agro-climate offers an excellent opportunity to dissect thermo-tolerance in chickens, both under hot-humid and hot-arid conditions. Whilst most Nigerian geographic regions experience very high temperatures (except in the high plateaus), the climate varies from very wet conditions in the coastal south (annual rainfall >3,500 mm, temperature up to 32 °C) to substantially dry conditions in the Sahel region of the north-west and north-east (annual rainfall <600 mm, temperature up to 41 °C). The genetic characterisation of these indigenous chickens (IC) is crucial for conservation of genetic and adaptive diversity and for elucidating the molecular mechanisms of environmental adaptation, particularly to heat-stress.

Most of the genetic studies on Nigerian IC were on mtDNA, with none examining whole-genome sequence (WGS) data. Here we examine the genetic diversity in Nigerian IC with WGS data, and explore genomic signatures for positive selection in response to heat-stress.

Materials & methods

Chicken sampling and genome sequencing. Blood samples were collected from 120 village chickens from 14 populations representing different agro-ecological zones (AEZs) (Figure 1). WGS was performed on the Illumina HiSeqX platform at 30x paired-end coverage.

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_422, © A.A. Gheyas *et al.* 2022

Population	# Samples	Agro-ecology
Sabiyal	6	Sudan Savanna (Arid/Semi-arid)
Jiga	4	Sudan Savanna (Arid/Semi-arid)
WatKaru	10	MidAltitude
DakanKaru	10	MidAltitude
Senchi	4	Northern Guinea Savanna
KWT	6	Northern Guinea Savanna
Ogbondoroko	10	Southern Guinea Savanna
Sanchitagi	10	Southern Guinea Savanna
Gitta Mbasha	10	Derived Savanna
Karu	10	Derived Savanna
Okpofe	10	Humid Savanna/Rain Forest
Odenkume	9	Humid Savanna/Rain Forest
Isiokpo	10	Humid Savanna/ Fresh water Swamp
Degema	10	Humid Savanna/Mangrove

Figure 1. (A) Nigerian agro-ecological map showing sampling locations; (B) details of the studied chicken populations.

Sequence data processing and genomic analysis. Sequence reads were mapped against the chicken reference genome (GRcg6a) using the BWA-mem algorithm. SNP calling was performed following GATK best practice protocol for 'Germline short variant discovery' followed by Variant Calling Score Recalibration (VQSR) for initial filtration. Autosomal SNPs with minor allele frequency >0.05 and genotype quality >15.0 were used for genetic diversity and selection signature analyses (SSA). SSA was performed using Pooled Heterozygosity (*Hp*) (Rubin *et al.*, 2010) and *Fst* (Weir and Cockerham, 1984) approaches in overlapping sliding windows (20 kb size with 10 kb step) with at least 10 SNPs/window.

Results

Nigerian IC show large genetic diversity but low levels of population differentiation. The combined analysis of all samples identified over 17 M SNPs of which 24% (~4.1 M) are novel. This represents an average density of 1 SNP every 62 base pairs. The number of SNPs from different populations varied between 9 M (Jiga) and 12 M (Sanchitagi); variation largely reflecting the number of samples analysed. Nucleotide diversity (pi) – calculated in overlapping windows of 20 kb size and 10 kb steps – was similar across all the populations (0.0031-0.0035). Moreover, a low level of inbreeding (<0.05) was observed for all populations. PCA plots – using 4M LD-pruned SNPs – show overall weak population structure, with most populations clustering together except Degema, DakanKaru and Odenkume (Figures 2A, 2B). Admixture analysis predicted contributions from four ancestral gene pools (Figure 2C) but pairwise *Fst* analyses showed a generally low level of population differentiation (*Fst*<0.05), except in a few cases where moderate differentiation (0.05-0.08) was observed (Figure 2D).

SSA reveals candidate genes for heat-stress adaptations. The climatic conditions of the studied populations showed little variation in annual mean temperature (except populations from the MidAltitude region) (Figure 3A), but large variations in annual rainfall (Figure 3B). SSA was therefore performed in two ways. First, 12 populations (omitting WatKaru and DakanKaru) were combined for *Hp* analysis. The combined analysis allowed reduction of spurious signals from population structure and detection of genomic regions with extreme low heterozygosity (*zHp*<-4) from all the hot-climate populations. These genomic loci and overlapping genes are therefore candidates for heat-stress adaptation, irrespective of humid or arid conditions (Figure 4A). Second, *Fst* analysis was performed among population groups from hot-humid (Degema and Isiokpo) and hot-arid (Jiga and Sabiyal) climates (Figure 3B) to identify candidate regions/genes that show large differentiation (*Fst*>3.5) in relation to heat-stress adaptation under dry and humid conditions (Figures 4B-4D). Selection direction of the *Fst*-sweep regions was decided based on their *Hp* values within groups.

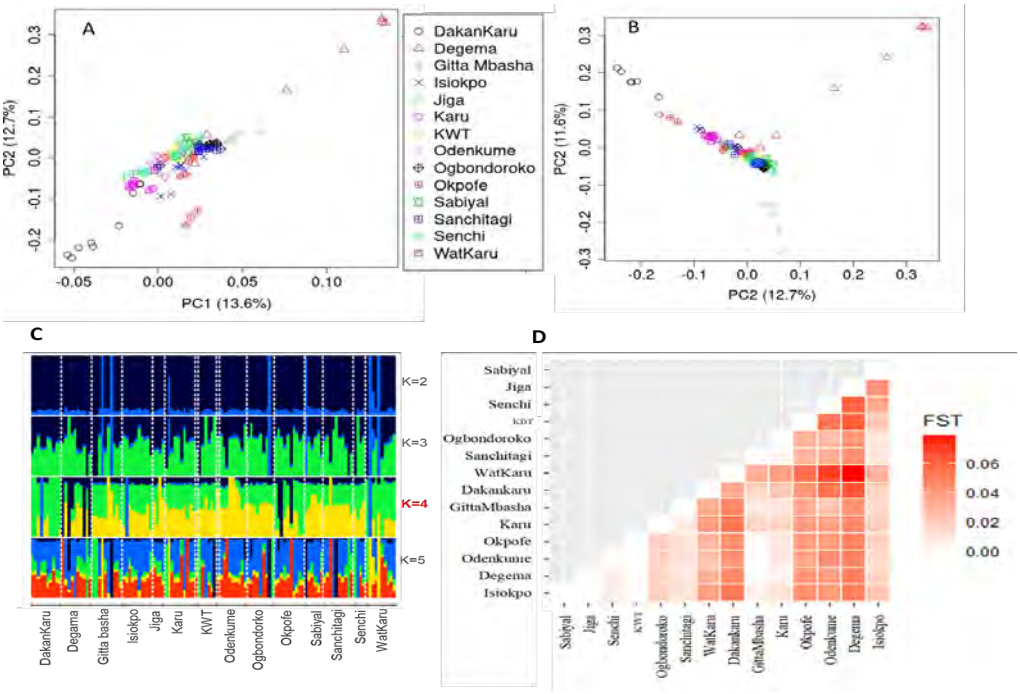


Figure 2. Population genetic diversity. (A & B) PCA plots showing population structure, (C) admixture analyses (K=4 is best prediction); (D) heat map of pairwise *Fst* values.

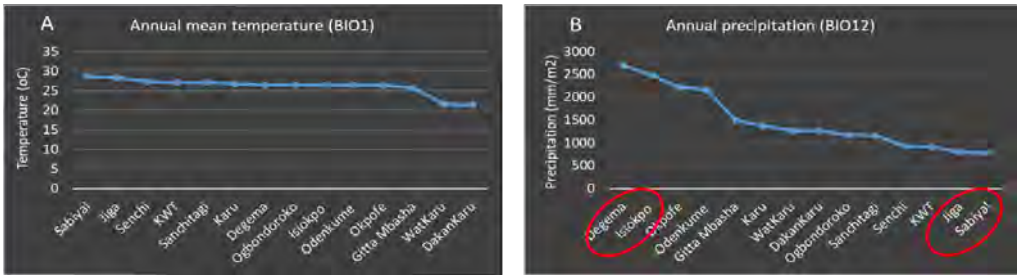


Figure 3. Population gradients based on (A) annual mean temperature and (B) annual precipitation (using the WorldClim database https://www.worldclim.org/data/bioclim.html).

Discussion

This is the first study performing a large scale WGS analysis on Nigerian IC to assess genetic diversity and identify genomic signatures of adaptive selection in relation to hot climates. Our study shows a large genetic diversity in Nigerian chickens that can be harnessed in breeding programmes for improvement of production and performance traits. The study also dissected candidate genes for general heat-stress adaptation (*Hp* analysis) as well as those important in either humid or arid conditions (*Fst* analysis). The *Hp*-based candidates have highly relevant biological functions associated with thermo-tolerance, e.g. *TSHR* – role in thermogenesis (possibly regulated by epigenetic mechanisms, as fixed in most chickens), *RYR1*-like genes (*LOC11233047* and *LOC101748756*) – involved in hyperthermia, *CYP450*-like genes

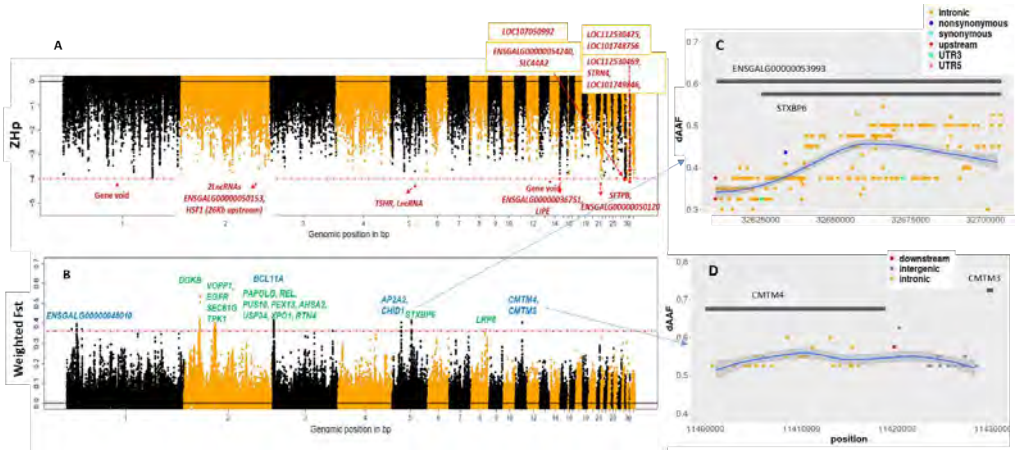


Figure 4. (A) Manhattan plot from *Hp* analysis based on 12 hot climate populations; (B) Manhattan plot for *Fst* analysis between hot-humid and hot-arid groups (candidate genes for humid climate are shown in green and for dry climate in blue); (C) and (D) showing close-up of two selection signature regions from humid and arid regions, respectively (Y axis represents difference in alternative allele frequency between the two groups)

(*LOC11253049* and *LOC101749846*) – roles in oxidative stress response, *SFTPB* – role in respiratory gaseous exchange (affecting heat loss from body), *LIPE* – lipid metabolism, *STRN4* and *SLC44A2* – roles in nervous system processes and immunity, and two lncRNAs with possible cis-regulatory effect on Heat Shock Factor 1 (*HSF1*) gene (Lin *et al.* 2017; Seo *et al.* 2020; Uniprot). The candidate genes associated with hot-arid conditions include: *CMTM3, CMTM4, BCL11A* and *CHID1* – all with roles in immune system development and/or functioning and *AP2A2* and *LRP8* with roles in lipid metabolism (Uniport). Prominent candidates from hot-humid climate include: *DGKB* – involved in many cellular processes including immunity (Zhong *et al.* 2008), *EGFR* – role in converting extracellular cues to appropriate response (Uniprot), *SEC61G* – involved in translocating proteins to Endoplasmic Reticulum, which has an important role in environmental and cellular stress response (Singh *et al.* 2021), *TPK1* – involved in heat and osmo-stress response (Reca *et al.* 2020); *AHSA2* – activator of Heat Shock Protein 90 (*HSP90*) gene (Uniprot), *RTN4* – with many roles including apoptosis, angiogenesis, and immunity (Uniprot). These results will have important implications for breeding improvement to develop heat-tolerant productive breeds for tropical small-holder farming systems.

References

Lin L., Lo L.H.Y., Lyu Q., and Lai K.O. (2017). J. Biol. Chem. 292(23):9451-9464. https://doi.org/10.1074/jbc.M116.772442

Reca S., Galello F., Ojeda, L, Pautasso, C., Cañonero L., *et al.* (2020) Biochim. Biophys. Acta. Gene Regul. Mech. 1863(9):194599. https://doi.org/10.1016/j.bbagrm.2020.194599

Rubin C.J., Zody M.C., Eriksson J., Meadows J.R., Sherwood E., *et al.* (2010). Nature 464:587-591. https://doi.org/10.1038/nature08832

Seo G., Han H., Vargas R.E., Yang B., Li X., *et al.* (2020). *Cell Reports* 32(1):107860. https://doi.org/10.1016/j.celrep.2020.107860

Singh M.B., Lohani N., and Bhalla P.L. (2021). *Front. Plant Sci.* 12: 661062. https://doi.org/10.3389/fpls.2021.661062

Uniport. Available at: https://www.uniprot.org/

Weir B.S. and Cockerham C.C. (1984). *Evolution* 38, 1358-1370. https://www.jstor.org/stable/2408641

Zhong X.P., Guo R., Zhou H., Liu C., and Wan C.K. (2008). *Immunol. Rev.* 224: 249–264. https://doi.org/10.1111/j.1600-065X.2008.00647.x

Part 23

Challenges – delivering genetic progress in systems around the world

423. Impact of industry effort to increase dairy heifer genomic testing in Australia

M. Abdelsayed[1], M.M. Axford[1,2,3*] and G.J. Nieuwhof[1,2]

[1]DataGene Ltd, 5 Ring Road, Bundoora, Victoria 3083, Australia; [2]Agriculture Victoria, AgriBio, Bundoora, Victoria 3083, Australia; [3]School of Applied Systems Biology, La Trobe University, Bundoora, Victoria 3083, Australia; maxford@datagene.com.au

Abstract

The increased use of heifer genomic testing enables further herd improvement on Australian dairy farms. This paper describes pre-competitive, industry efforts to improve the customer experience and increase the rate of heifer genomic testing. Following the implementation of process improvements, the number of females genotyped, annually, more than doubled between years ending 30 June 2020 and 2021. Turn-around time was reduced by two-thirds and 99% of animals now receive results. Process improvements also contribute to the success of activities that increase collaboration and industry extension.

Introduction

Heifer genomic testing delivers value to most Australian dairy farmers as a tool to select replacements, correct pedigree records, control inbreeding and plan mating strategies (Newton *et al.*, 2018). Yet, in 2018/2019, less than 20,000 females (2% of dairy cows) were tested in Australia. An industry-lead initiative began in 2019 to rapidly increase the rate of genomic testing through three work areas:

1. implement a series of process improvements to deliver results more quickly;
2. collaborate with commercial and industry partners to support farmers with clearer and more consistent messages about the value of genomic testing; and
3. develop and deliver communication and extension programs using outcomes from market research activities.

Following the report of collaboration, communication and extension activities presented by Monks *et al.* (2021), the purpose of this paper is to describe the impact of process improvements that delivered an improved customer experience to genomic service providers (customers) and their dairy farmer clients by DataGene.

Materials & methods

In consultation with customers and industry stakeholders, the most important process improvements were prioritised on the basis of likely impact on customer experience. The improvements can be grouped into four areas including faster turn-around time, extending evaluations, systematic correction of data issues and providing secure access to the centralised data repository.

Faster turn-around time. Genetic evaluations were run monthly at the commencement of this work. Accounting for sample transport, processing and lab times, a monthly schedule meant that some animals received results more than two months after samples were collected. This turn-around time did not meet the expectations of customers and so a new target was set for the weekly evaluations of all animals.

Investigators analysed workflows to identify opportunities to improve efficiencies and increase automation during the evaluation schedule. For example, the genetic evaluation team traditionally loaded data from herd recording centres, monthly, with some manual processing required. With automation in place, animal data is now systematically loaded through an application programming interface and quality checked every

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_423, © M. Abdelsayed *et al.* 2022

night from most centres. Provisions are in place for centres that choose to load data, weekly, through an enhanced managed file transfer protocol. In a second example, virtually all genetic evaluation and data quality reports for customers and internal users are now systematically generated as part of the evaluation without manual intervention. Finally, a restructure of the data exchange between genetic evaluation and report delivery systems has reduced the time taken to deliver reports and update all web-based interfaces. A weekly evaluation schedule was then able to be implemented as a result of workflow changes that increased efficiency and reduced manual intervention.

Extend evaluations. Genomic evaluations for animals sired by a Holstein or Jersey sire commenced in Australia in 2011 and 2012 respectively. The project aimed to grow the number of animals that can benefit from genomic testing. Animals sired by the Red breed group (Australian red dairy cattle, Scandinavian red, Ayrshire, Illawarra or Dairy Shorthorn) are the next biggest breed group, comprising 4% of herd-recorded cows in Australia (DataGene (2019/20)).

Building on the work of Van den Berg *et al.* (2020), DataGene scientists implemented a single-step genomic evaluation for four traits for the Red breed group (DataGene, 2021). Traits were selected based on reliability and included milk (L), fat (kg), protein (kg) and somatic cell count (%). When first published in July 2021, average reliability for genotyped females was reported to be 59% for protein (kg) (DataGene, 2021). The National economic index, Balanced Performance Index, is calculated using genomic evaluations for traits where they are available and traditional evaluations for the remainder (DataGene, 2021).

Two other examples where the evaluation has been extended is the calculation of breed-based representation and the reporting of additional haplotypes. Both services deliver additional value to the customer without additional effort as part of the routine evaluation.

Systematic correction of data issues and providing troubleshooting tools. Prior to the commencement of the project, 89% of nominated animals received breeding value results following genomic evaluation. Twenty percent of animals were reported with an inconsistent sire, 10% with an inconsistent dam, 10% without a correct breed, eight percent with an inconsistent birthdate and six percent of samples were rejected for poor quality.

To achieve the target of 95% of animals receiving results, workflows were revised to resolve inconsistent pedigrees where the sire and/or dam could be discovered as part of the evaluation. Importantly, this service is customised to meet the needs of individual customers. The automatic update of pedigree is an 'opt-in' service that service providers can elect to include if it is appropriate for their clients. Secondly, rules were implemented to prevent the changing of a confirmed pedigree, breed or birth date by other parties. This ensures that animals receiving a result don't suddenly fail based on incoming data containing outdated information.

Customers value the ability to quickly detect and resolve an issue related to an animal, genotype or service request with as much capacity for automation as possible. In addition to automatically correcting pedigree, additional static reports and an interactive dashboard with appropriate permissions were built to help customers to troubleshoot issues.

Provide secure access to the centralised data repository to genomic service providers. The number of genomic service providers has doubled since the beginning of the project. By opening access to DataGene's centralised data repository, customers can structure a service according to their business objectives. Previously, customers were limited to supplying data and receiving static reports using file

transfer protocols (ftp) and secure websites. Now, customers can exchange animal data, genotypes and load service requests using a managed file transfer protocol with multiple file loading capacity that can be operated programmatically or by interacting with a user interface. With appropriate security and permissions, customers now have the capacity to load/edit new animals to the centralised data repository, a role that was previously limited to data processing centres.

Results

The turn-around time between data loading and result reporting is a useful metric to monitor as it illustrates the impact of numerous workflow improvements across the system. The mean number of days between loading data and reporting a result has declined from 25 days in June 2020 to 7 days in December 2021 as shown in Figure 1. The steeper decline between quarters ending December 2020 and June 2021 is in line with the transition from monthly to weekly evaluations.

Extending genomic evaluations has increased the number of animals eligible for testing. In the quarter concluding December 2021, four percent of females tested had a sire from the Australian red dairy cattle breed group. This value is identical to the proportion of red dairy cattle animals in the population, indicating a strong, early, uptake of the new evaluations.

In June 2019, June 2020 and June 2021, the proportion of animals receiving results was 89%, 98 and 99% respectively with the most recent two years exceeding the original target of 95%. These results are influenced by the automatic correction of inconsistent pedigree and rules that prevent the over-writing of confirmed animal information. Table 1 shows the proportion of genotyped animals with data warnings or errors and illustrates the impact of process improvements. For example, inconsistent sires were observed in 5% of animals in 2021, compared to 20% in 2020.

Figure 1. Mean days between loading and result delivery.

Table 1. Proportion (%) of genotyped animals with warnings or errors at the time of evaluation.

	Financial year	
Reason	**2020**	**2021**
Inconsistent birth date	8	0
Inconsistent sire	20	5
Inconsistent dam	10	2
Breed unknown or not evaluated	10	1

Discussion

A concerted effort to prioritise and implement substantial system-wide improvements has delivered a faster genomic evaluation service for more animals. The number of females genotyped in the financial year ending 30 June 2021 was more than double that of the previous year. Turn-around time was reduced by two-thirds and 99% of animals now receive results. There is an opportunity to continue to improve service by reducing the turn-around time for animals that arrive with data issues. Additionally, there are opportunities to grow the genomic service by incorporating genomic inbreeding, creating actionable reports using breed-based representation and haplotypes and providing further customer training to optimise their use of the system.

Improving the customer experience is vital to realising results from investment in closer collaboration and stronger communication/extension activities. For example, standardised reporting metrics are regularly shared with commercial and industry partners as they work alongside DataGene to grow awareness and uptake of heifer genomic testing. In a second example, the faster turn-around time and introduction of red breed genomics has attracted farmers to test their heifers for the first time. Farmers that are new to genomic testing have then shared their early experiences in case studies distributed through social and traditional media as part of co-ordinated communication and extension activities organised by DataGene, Dairy Australia and genomic service providers. Success in all three project work areas is vital to realising higher rates of heifer genomic testing.

System-wide process updates that deliver an improved customer experience are expected to lift participation in heifer genomic testing leading to an improvement in heifer selection for the milking herd. More accurate heifer selection means higher rates of genetic gain and faster herd improvement for farmers who choose to utilise heifer genomic testing.

This work was completed as part of the Dairy Australia and DataGene funded project known as 'Accelerating Heifer Genomics'.

References

DataGene, 2019/20. Herd Recording Statistics 2019/20. DataGene, Bundoora, Australia. Available at: https://datagene.com.au/IndustryStatistics

DataGene, 2021. Genomics for Red Breeds. DataGene, Bundoora, Australia. Available at: https://datagene.com.au/sites/default/files/Upload%20Files/Technote%2027%20Red%20Breed%20Genomics_1.pdf

Monks L.C., Scott B.A., Bullen S., Andony J.J. and Axford M.M. (2021). Proc. of the 24th Assoc. Advmt. Anim. Breed. Genet, on-line.

Newton, J.E., Hayes, B.J., Pryce, J.E., 2018. J Dairy Sci 101, 6159-6173. https://doi.org/10.3168/jds.2017-13476

Van den Berg, I., MacLeod, I.M., Reich, C.M., Breen, E.J., Pryce, J.E., 2020. J Dairy Sci 103, 6276-6298. https://doi.org/10.3168/jds.2019-17914

424. Weekly national dairy evaluations – a Vietnam-Australia collaboration

G.J. Nieuwhof[1,2*] and L.T. Ong[3]

[1]DataGene, AgriBio, 5 Ring Road, Bundoora, Victoria 3083, Australia; [2]Agriculture Victoria, AgriBio, 5 Ring Road, Bundoora, Victoria 3083, Australia; [3]TMA Solutions, TMA Tower, Quang Trung Software City, District 12, Ho Chi Minh City, Vietnam; gnieuwhof@datagene.com.au

Abstract

A data warehouse and workflow system was built to store the Australian dairy milk recording data and run genetic evaluations. The involvement of a specialised Vietnamese IT firm ensured a systematic approach that, based on detailed requirements as detailed by stakeholders in Business Requirement Documents, led to a system that is maintainable, agile, adaptable and auditable. It is based on the Data Vault modelling technique, designed at several layers and forms the core of the logic and operations, implemented using mainly SQL and C#. Changes were tested in a testing environment before being brought to the production environment. The system evokes specialised statistical programs for imputation and genetic and genomic solvers. Taking advantage of the latest developments in IT as well as the skills of experts, the new system allows for complete weekly runs with a single user operation, compared to monthly runs depending on a team of four previously.

Introduction

Since its establishment in 1982 the Australian Dairy Herd Improvement Scheme (ADHIS) and its successor DataGene have been expanding and improving the genetic evaluations made available to the Australian dairy industry. The system grew organically; extending the number of traits analysed, followed by the introduction of a national breeding objective and its updates.

The frequency of evaluations increased from a few per year, to about 8 in the years leading to 2020, with a jump to weekly runs in 2021. This paper describes how innovations in the storage of data and use of workflows implemented by a specialised IT company enabled this dramatic rise in run frequency, while improving quality and flexibility in a system that is more user-friendly and delivers a better service to stakeholders.

Conventional evaluations. In 2007, after the national dairy database as maintained by ADHIS had grown for over two and a half decades, both in terms of number of animals and number of traits, it was realised that there was a need for a more flexible, faster and easily maintainable database. An important consideration was the ability to continue the exchange of data. It was realised that a system conceived, written, tested, maintained and run by a small group of specialists (trained in genetics rather than IT), may not take full advantage of the latest IT developments and bears the risk that knowledge residing in a single person could disappear altogether.

Genomics. Then genomics appeared and took priority. The advantage of genomics was that it was new, and therefore a more systematic approach could be taken to implement newly written requirements. The genomic evaluation system called GEN, incorporated many of the requirements for a complete new system, such as a relational database, user-friendly user interfaces and automated workflows. But it was only partial; it started by importing conventional breeding values from the old system and its database only held breeding values and genotypes, no phenotypes. GEN was built by a small local company that was specialised in dairy herd recording. It went live in 2010, delivering Australia's first genomic breeding values.

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_424, © G.J. Nieuwhof and L.T. Ong 2022

2010-2019. The introduction of genomics brought with it an increased interest in cows and an emphasis on a high turn-around. Actually running the evaluations required 4 people for 2 weeks who had to first process incoming data and build databases (c++), extract data and start BLUP runs for each individual trait group, and in the case of conformation, each separate breed, build databases of breeding values, extract them again to blend with MACE proofs, import them again and export basic animal data and breeding values to the GEN system. In the meantime the latest genotypes had to be loaded on GEN, so that imputation could be run, followed by marker effect estimation (SNP BLUP) and in a next step DGV calculation and genomic blending (Nieuwhof *et al.*, 2010). Instruction to run this, for the old conventional system spanned over 100 pages of text that had to be copied, line by line to a unix command line and edited. The GEN system had a more friendly user interface but still required some special manual steps, e.g. to run sql queries to create lists of animals qualifying for imputation, the reference set or genomic blending. Both approaches were prone to human error.

In a separate development the Australian dairy industry realised there would be a great advantage in sharing data, rather than each organisation having their own. This project was called Central Data repository (CDR). It was clear that CDR would become the database from which the genetic evaluation system would obtain its data and the two should be developed together.

The new system

Agile Software Development is now considered state-of-the-art for IT projects (Beck *et al.*, 2010, Wikipedia, 2022). Product owners and the development team collaborate to deliver results faster with fewer overheads. Instead of having everything evident on the table, the agile team deliver work in small, consumable increments. Changes are expected and responded quickly during the project life cycle. Agile environments value interaction, results and customer engagement. By applying the Scrum framework (Schwaber and Sutherland, 2020), the development team and Subject Matter Expert (SME) team can plan, work and validate the result through short iterations (or Sprints), which cover two weeks. Daily Scrum meetings are conducted to react to issues and feedback that may cause the project to deviate from its goal.

Data Vault architecture aligns well with the ideas behind Agile as it avoids having all the data models designed at the beginning when there is still a lot of fuzziness. Business entities are broken down to the smallest grain (hubs) connected through links to depict the relationship. Satellites are attached to hubs or links to describe the entity or relationship attributes. New hubs and links are added later as the team has a better understanding of the requirements.

TMA Solutions in Vietnam was engaged to build the new system. Detailed Business Requirements Documents (BRD) were drafted by DataGene staff and other stakeholders. At a high level, the requirement for the new system combining CDR and the genetic evaluation system (now called GESNP) was to create a relational database to store all Australian milk recording and associated data, including genotypes, combined with automated workflows for running of evaluations and reporting. A User Interface (UI) was to be built to allow both internal and external users to access the central database for loading and interrogating data. All the latest (and some historic) breeding values would be available on a website.

The development team first analysed the requirements and a data warehouse, modelled after the Data Vault technique, was designed at several layers. The data model formed the core of the logic and operations. These logics (or workflows) were implemented using mainly SQL and C#. They closely follow the steps of the old and GEN systems. The data warehouse does most calculations and stores intermediate values between steps. For specialised statistical computations (called Black Boxes) the system invokes programs such as FImpute (Sargolzaei *et al.*, 2014) for imputation and MiX99 (MiX99 Development Team, 2016) for genetic analyses (Figure 1). Minimal human input is required and, therefore, minimal risk of human error.

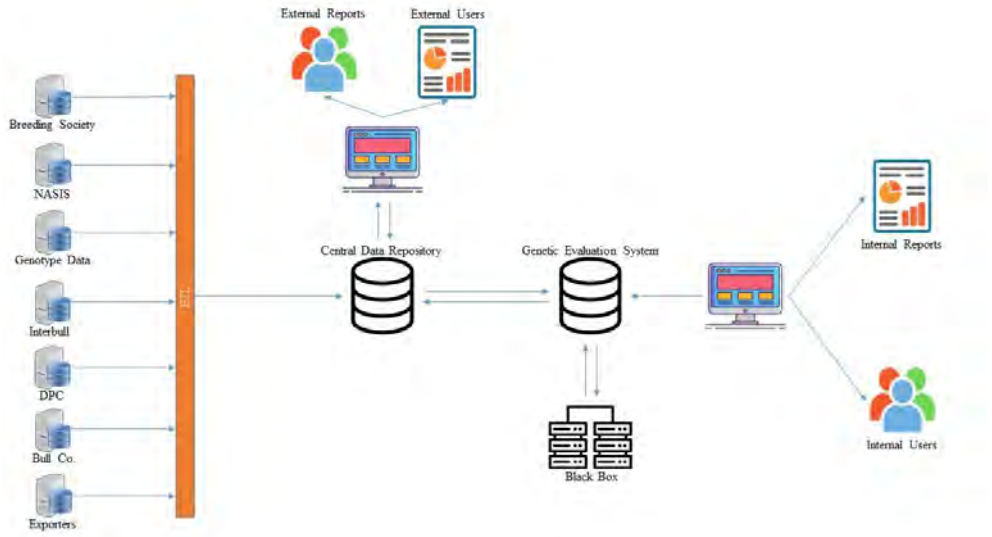


Figure 1. Simplified high-level overview of the CDR/GESNP system.

In its first version, CDR didn't have a lot of features and was mainly for daily data loading from milk recording centres and bull companies. Over time it has completely integrated numerous functions such as bull registration, progeny test reports and genotype services. The system consists of multiple environments for different purposes – pre-production, User Acceptance Testing (UAT) and production. Any changes are first made in the pre-production environment, where they can be tested with dummy data. Secondly, the changes are deployed to UAT, which is very similar to the production environment, to test with actual data. Finally, approved changes go live through periodical production builds. A release note and BRD updates are included to track all the new enhancements in that build.

The main engine of an evaluation run is now called '1-click', although it actually requires six clicks. 1-click runs all steps from initial data extraction to creation of final reports and a report database, in about 3.5 days, without any shortcuts. Prior to the 1-click is a manual exchange of data from CDR to GESNP and the user needs to start imputation with a few clicks on the UI. At the end, the exchange to DataVat for website publication (www.datavat.com.au) is also still a manual process. In incremental steps, and while observing best-practice, these manual steps are being automated.

An important labour reduction is achieved by automatically importing data from the milk recording centres overnight. QA reports can be generated by a few simple clicks on the UI. A typical weekly run consists of:

Wednesday evening:	exchange data from CDR to GESNP
Thursday morning:	run imputation
Thursday midday:	start 1-click
Monday morning:	1-click finishes, QA
Tuesday:	exchange to DataVat, publication of reports

Development is continuous, the next big step is inclusion of the single step genomics for Red Dairy Cattle in the 1-click.

Discussion

The move from a database and workflow system controlled from conception to daily operation by a small group of geneticists to a state-of-the-art IT system is in essence a change in philosophy. While geneticists may at times get frustrated by the new system as it does not allow for quick fixes and requires spelling out of 'obvious' things, in fact the need to think through, document and defend detailed requirements, creates a better system short-term and enables maintenance and upgrades medium and long-term. The approach allows for involvement of stakeholders such as bull companies, breed societies and milk recording centres, as well as non-geneticists within the organisation at both the requirement end and at testing of the final product. IT specialists are very good at checking logic and inconsistencies and they have a better overview of the latest technical developments. Geneticists need to learn to think in terms of requirements, the IT specialists will design and build the solutions.

Conclusions

By engaging an IT firm and going through a systematic approach a new database and workflow system was developed to store Australia's herd recording data and run complete genetic evaluations on a weekly basis. The system is based on solid documentation and independent testing and minimises the risk of human error. Further improvements and expansion will be based on the same philosophy.

References

Beck K., Grenning J., Martin R.C., Beedle M., Highsmith J., et al (2001). Manifesto for Agile Software Development. Available at https://agilemanifesto.org/.

MiX99 Development Team (2016) MiX99: A software package for solving large mixed model equations. Release XI/2016. Natural Resources Institute Finland (Luke). Jokioinen, Finland. Available at http://www.luke.fi/mix99

Nieuwhof G.J., Beard K.T., Konstantinov K.V., Bowman P.J. and Hayes B.J. (2010). Interbull Bulletin 42: 35-39.

Sargolzaei M., Chesnais J. P. and Schenkel F.S (2014) BMC Genomics 15, 478. https://doi.org/10.1186/1471-2164-15-478

Schwaber K., Sutherland J., 2020. The scrum guide. Available at https://scrumguides.org/docs/scrumguide/v2020/2020-Scrum-Guide-US.pdf#zoom=100.

Wikipedia (2022). Agile Software Development. Available at https://en.wikipedia.org/wiki/Agile_software_development.

425. The value proposition for using genomics to confirm parentage in Australian dairy systems

J.E. Newton[1*], M.M. Axford[1,2,3], J.E. Pryce[1,3], S.L. Bullen[4] and M. Haile-Mariam[1]

[1]Agriculture Victoria Research, AgriBio, 5 Ring Road Bundoora, 3083 Australia; [2]DataGene Ltd, 5 Ring Road, Bundoora, 3083, Australia; [3]School of Applied Systems Biology, La Trobe University, Bundoora, 3083, Australia; [4]Dairy Australia, Level 3, HWT Tower, 40 City Road, Southbank, 3006, Australia; jo.newton@agriculture.vic.gov.au

Abstract

To facilitate uptake of female genomic testing, this study sought to estimate and quantify the impact of sire pedigree errors in cow populations. In a sample of 1,863 genotyped females, 26% of females had sire pedigree inconsistencies. The most frequent of these were nominating either unknown or incorrect sires. The additional value correcting pedigree added to GEBV-based heifer selection only represented a small percentage of genomic testing benefits, though this increased as the proportion of females with pedigree errors increased. In a separate scenario, heifers incorrectly identified as either herd replacements, or cull animals due to errors in sire pedigree were expected to cost between $AU 84-253 in lost lifetime profit per mis-identified heifer. While numerically the impact of parentage errors was small, given their importance to farmers, developing value propositions for confirming parentage with genomic testing could help increase female genomic testing and improve genetic progress.

Introduction

Genetic progress in Australian dairy systems is influenced by comparatively low uptake of female genomic testing. In 2020, only 4% of herd recorded cows were genotyped compared to 22% in the USA (Monks *et al.* 2021). The two top reasons given by Australian dairy farmers for engaging in genomic testing were parentage verification and identifying herd replacements (Monks *et al.* 2021). The value of using GEBVs to inform replacement decisions has been extensively modelled (i.e. Newton and Berry (2021). However, while Banos *et al.* (2001) showed that cow pedigree errors bias sire EBVs and reduce genetic progress, the impact of parentage errors on female populations or in isolation from GEBV based selection decisions has not been considered. In pasture-based systems where matching dam to calf is difficult, the value proposition for verifying pedigree is expected to be greater. In Australia, it is common practice to artificially inseminate (AI) cows early in the mating period and use natural matings (a bull) in the later part of the mating period and over heifers. Also, currently only ~50% of commercial Australian dairy farms have access to cow EBVs. To foster increased farmer engagement with genomic testing it is important value proposition development reflects actual on-farm scenarios. The aim of this study was to estimate sire parentage error frequency and type in Australian dairy herds and calculate the potential impact of sire error rates in the milking herd to help facilitate increased uptake of genomic testing.

Materials & methods

Estimating pedigree error rate. Parentage discovery completed during genomic testing was used to calculate frequency of discrepancies in cow pedigree. Discrepancies in sires and dams nominated on-farm and discovered (identified) during genomic testing parentage discovery were summarised from individual parentage discovery reports generated for heifer cohorts in 16 Australian dairy herds (no. females = 1,863) by Holstein Australia. Average across-herd discrepancy rates, standard errors and the types of errors are presented.

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_425, © J.E. Newton *et al.* 2022

Impact of pedigree errors. The additional value of correcting pedigree errors in heifer selection candidates adds to GEBV based heifer selection was calculated using a deterministic excel tool developed by Newton and Berry (2020). The tool estimates the net benefit of selecting replacements on GEBVs over parent-average derived (pedigree) EBVs (PA-EBV) in a discrete generation of heifers. It assumes that the selection index is measured in $/cow.year and that selected replacements express their genetic superiority in additional profit upon entering the milking herd. The tool uses the formula described by Visscher *et al.* (2002), to apply an adjustment to sire reliability based on sire pedigree errors enabling comparison of scenarios with and without pedigree errors. Key model assumptions were: heifer GEBV reliability of 0.66, dam and sire EBV reliability of 0.32 and 0.71, respectively, and standard deviation of the national breeding objective of $AU 90. The proportion of heifers retained as replacements ranged from 50-95% while corresponding replacement rates were 17-32%.

To be reflective of more Australian farms, a simplified heifer selection strategy which did not use EBVs was also modelled. An average fertility herd with 70% AI-sired calves and 30% naturally sired calves was used. It was assumed all AI-sired heifer calves are kept as herd replacements and naturally sired heifer calves are sold. One impact of sire pedigree errors is to accidently sell an AI-sired calf or keep a naturally sired calf. As on average, AI-sired calves have superior genetic merit than naturally sired calves (and the national breeding objective is expressed in $/cow.year), this error can be represented as the expected difference in lifetime profit of the 2 calf types. Genetic merit differences in AI and naturally sired calves ranging from $AU 50 to $AU 150 were considered, reflecting differences observed in industry (M. Axford, pers. comm). However, not all sire errors incur a cost in this strategy, i.e. recording the wrong AI sire for an AI-sired calf is a silent error and has a value of $0 as all AI-sired calves are kept. To get the average cost per heifer with wrong sire pedigree and account for the difference in cost of sire errors, the frequency of each sire error type (Table 1.) was multiplied by the cost of that error. This gave a cost weighted by the frequency of errors.

Results

On average, 26% of females submitted for genotyping had discrepancies in sire nomination and verification (Table 1). The most common error, representing 50% of sire errors was an unknown and ungenotyped bull being nominated as the cow's sire. The next most common errors were a sire being nominated on-farm and, a different AI sire (34%) or no matching sire (11%) being identified during parentage discovery. Two percent of all records had no sire nominated and no sire identified during parentage discovery. In total 26 (1.4%) females had incorrectly assigned dams identified during parentage discovery.

As expected, as the proportion of heifers with incorrect pedigree increased, the value that correcting pedigree added to GEBV based heifer selection increased (Figure 1), however was low overall. At lower replacement rates and higher selection intensity, the added value from correcting pedigree was greater. For example, at a sire pedigree error rate of 26%, correcting pedigree added an additional $AU 3.28, 1.44 and $0.26 per heifer identified as a replacement using GEBVs, for replacement rates of 17, 22.5 and 32%,

Table 1. Sire discrepancies (and s.e.) identified from genomic testing (n=1,863 females).

	Recorded and discovered sires				
Nominated sire	valid AI sire	sire nominated[1]	valid AI sire	unknown sire	none
Sire verified	same AI sire	different AI sire	no sire found	no sire found	no sire found
Average (no. cows/ herd)	85.8 (16.9)	10.4 (4.2)	3.3 (2.3)	15.3 (4.7)	1.8 (0.9)
Totals	1,372	166	52	244	29

[1] Nominated sire could be an unknown natural bull or registered AI sire.

respectively. This represents 3, 6 and 0.4% of net-benefit of genomic testing, respectively. In our simplified replacement strategy, incorrectly identifying a heifer as a herd replacement or cull animal based on sire pedigree, is expected to cost between $AU 84 and $253 per heifer with incorrect sire, depending on the genetic merit difference between AI-sired and naturally sired calves (Figure 2). If there was a $AU 100 or more difference in genetic merit of the two calf types, net cost of pedigree errors exceeds $AU 1000 even at relatively low error rates (6%).

Discussion

In the Australian dataset used here, 26% of cows had inconsistencies in sire pedigree, though large variation existed across herds. This is similar to the 20% sire error rate reported for Australian animals genotyped in 2020 (Axford 2022). Both these values are higher than previously reported in the United Kingdom (Visscher *et al.* 2002), though could be influenced by Australia's larger and pasture-based herds. The 2.6%

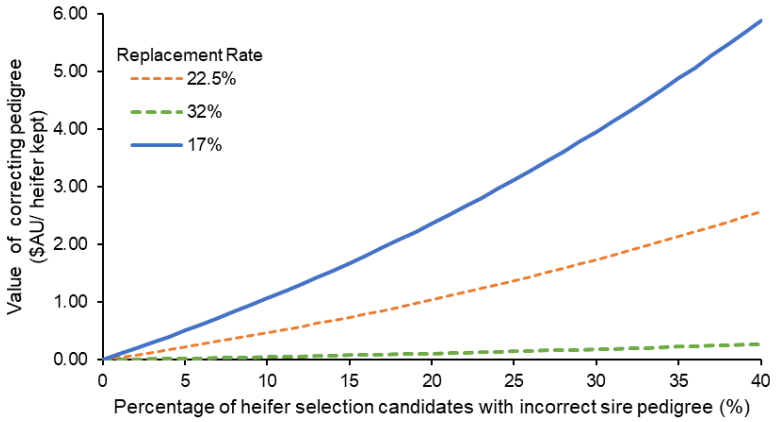


Figure 1. Additional value of correcting pedigree ($AU/heifer retained in the herd) as part of GEBV-based heifer selection decisions for 3 different herd replacement rates.

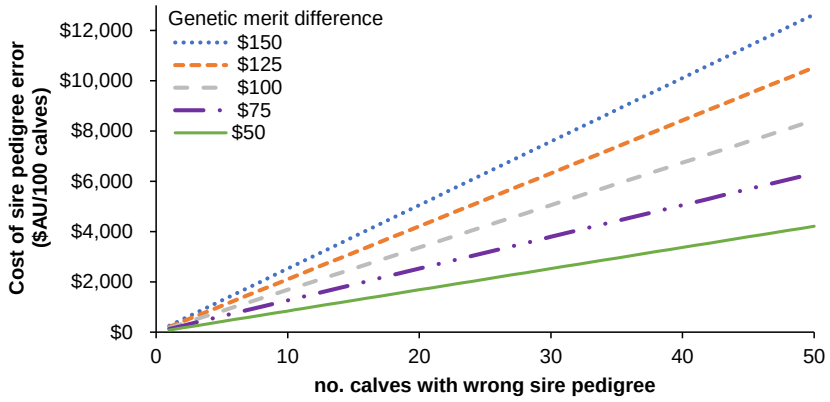


Figure 2. Lifetime cost of sire pedigree error ($AU/100 calves). Different line colours and styles indicate differences in genetic superiority of AI and naturally sired calves.

dam error rate found here is likely lower than the true rate, with few dam genotypes available for parentage verification. This study aimed to value the impact of sire error in cow populations. For an average Australian dairy herd (26% sire error rate, 279 cows, 22.5% replacement rate), the $AU 90 additional value ($AU 1.44 × 22.5% × 279) that correcting pedigree errors added to GEBV based heifer selection was ~6% of the overall value of female genomic testing. This $ value increases as selection intensity or sire pedigree errors increase as pedigree errors reduce accuracy of PA-EBV. Other illustrations of the value of correcting parentage are needed for the ~50% of Australian dairy farmers currently without access to PA-EBVs in their cow herds. Incorrectly identifying a heifer as a herd replacement or cull animal based on sire pedigree could cost an average Australian dairy herd $AU 994 to 2,993 in lost lifetime profit, depending on the genetic merit difference between AI and naturally sired heifers. However, this is less than the expected cost of genotyping at current prices. The full value of having verified pedigree in Australian dairy herds is not captured in these two examples. On-farm, in addition to playing a role in breeding programs, verified pedigree can also lead to 'opportunistic events' like selling surplus heifers into the export market (DataGene, 2021). In the USA, errors in sire pedigree in cow populations have been previously shown to bias sire EBVs, slow genetic trends and decrease genetic correlations in across-country bull comparisons (Banos *et al.* 2001). An opportunity exists to consider a similar study in the Australian population.

Low farmer uptake of female genomic testing is contributing to suboptimal genetic progress. A large industry initiative is currently underway with a target of 300,000 females genomic tested annually by 2025 (DataGene, 2020). Cost of genomic testing still remains a barrier to uptake (Monks *et al.* 2021). The return on investment from genomic testing improves when the information is used multiple times and for multiple purposes. It is therefore important to illustrate multiple uses to farmers and consider alternate value propositions (Newton *et al.* 2021). Aligning research questions to farmer feedback is one strategy to address the challenge of suboptimal genetic progress in female populations. Given the importance farmers place on verifying parentage from genomic testing (Monks *et al.* 2021), future value propositions of genomic testing should consider this aspect, even if numerically small. Extending messages on the value and impact of genomics beyond heifer replacement could help improve farmer uptake of genomic testing resulting in improved genetic progress in Australian dairy systems.

References

Axford M. M. and Nieuwhof G.J. (2022) Proc. of 12th WCGALP, Rotterdam, Netherlands.

Banos G., Wiggans G.R. and Powell R.L. (2001) J. Dairy Sci. 84:2523-2529. https://doi.org/10.3168/jds.S0022-0302(01)74703-0

DataGene (2021) Technote 1: Parentage verficiation. Available at datagene.com.au/technote

DataGene (2020) Accelerating Heifer Genomics. Available at datagene.com.au/node/1476

Monks L.C. *et al.* (2021) Proc. Assoc. Advmt. Anim. Breed. Genet. 24:242-246.

Newton J.E. *et al.* (2021) Anim. Prod. Sci. 61:220-229. https://doi.org/10.1071/AN20168

Newton J.E. and Berry D.P. (2020) Anim. 14:1565-1575. https://doi.org/10.1017/S1751731120000208

Visscher P.M. *et al.* (2002) J. Dairy Sci. 85:2368- 2375.10.3168/jds.S0022-0302(02)74317-8

426. Changes in genetic trends for dairy cattle in the U.S. since the implementation of genomic selection

F.L. Guinan[1*], G.R. Wiggans[2], J.W. Dürr[2], H.D. Norman[2], J.B. Cole[3], C.P. Van Tassell[4], I. Misztal[1] and D. Lourenco[1]

[1]Department of Animal and Dairy Science, University of Georgia, Athens, Georgia, USA; [2]Council on Dairy Cattle Breeding, 4201 Northview Drive, Suite 302, Bowie, MD, USA; [3]URUS Group LP, 2418 Crossroads Drive, Madison, WI, USA; [4]Animal Genomics and Improvement Laboratory, Agricultural Research Service, United States Department of Agriculture (USDA), Beltsville, MD, USA; fguinan@wisc.edu

Abstract

Multistep genomic evaluations for US dairy cattle were first implemented in 2009 for Holstein, Jersey, and Brown Swiss dairy cattle breeds. In contrast, the adoption of genomics for Ayrshire and Guernsey is more recent. We investigated the genetic trends, inbreeding levels, and generation intervals since 1975 for the five dairy cattle breeds receiving genomic evaluations. Overall, production traits have increased steadily over time, as expected. Holstein and Jersey have benefited most from genomics, with up to a 6-fold increase in genetic gain in fat yield for Holstein bulls since 2009, nonetheless, with a rise in inbreeding levels. The progress made by breeds other than Holstein and Jersey is still small because of the limited portfolio of bulls available and the lower number of genotyped and phenotyped animals. Expanding the benefits of genomics for those breeds will require more comprehensive adoption of this technology and market incentives.

Introduction

Official genomic evaluations for dairy cattle in the United States were released in 2009 for Holstein (HO), Jersey (JE), and Brown Swiss (BS), 2013 for Ayrshire (AY), and 2016 for Guernsey (GU) (Wiggans *et al.*, 2017). Those evaluations comprise multibreed BLUP followed by single-breed estimation of SNP effects. Genomic PTAs are obtained as the combination of direct genomic value based on markers and parent average. Since the implementation of genomic selection (GS) in the US, it has been documented that genetic trends have changed significantly for the HO breed (García-Ruiz *et al.*, 2016). Genetic trends measure the success or failure of a breeding program and are needed to justify the usefulness of genetic evaluations and, therefore, the rationale behind using new methodologies (Grosu *et al.*, 2013). There are currently over five million genotyped animals in the US evaluation system (https://queries.uscdcb.com/Genotype/cur_ctry.html), but the uptake of genotyping is not even across breeds. Increased reliability and a reduction in generation interval (GI) are known for driving the accelerated rates of genetic gain under GS. There is little documentation of the improvement and changes in breeds other than Holstein (i.e. AY, BS, GU, and JE) in the US since the introduction of genomic evaluations. Reviewing assumptions of genetic progress since the introduction of genomics is crucial for advising breed associations and non-academic stakeholders on how to utilize this technology moving forward. Therefore, the main purpose of this study was to investigate genetic trends in the five breeds receiving genomic evaluations for production, fertility, longevity, and health traits.

Materials & methods

The data for this study were extracted from the National Cooperator Database maintained by the Council on Dairy Cattle Breeding (CDCB, Bowie, Maryland). The data included (Genomic) Predicted Transmitting Abilities (PTAs) of bulls and cows born since 1975 incorporating pedigree, phenotypic and genomic (when available) data from the August 2021 official evaluation. Each breed is in the within-breed base; therefore, individual plots are not directly comparable. We analyzed milk, fat, and protein yield, somatic cell score (SCS), productive life (PL), daughter pregnancy rate (DPR), and livability (LIV). Data were restricted to a

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_426, © F.L. Guinan *et al.* 2022

minimum of 10 daughters per bull, and a weighted average for bulls was calculated. Cows had to have at least one calf (Masuda *et al.*, 2018), but their PTAs were not weighted by progeny size. Predicted breeding values (PBV) were calculated by multiplying the stored (G) PTA by two.

Previous studies (García-Ruiz *et al.*, 2016), used the four-path selection model was to calculate genetic trends using selection differentials. Research by Dechow *et al.* (2018) has suggested that the four-path selection model was derived for a closed population; however, through the use of artificial insemination commercial herds are not, in fact, closed. Thus, we generated genetic trends based on bull and cow contributions to genetic improvement. Statistical Analysis Software (SAS Institute, 2016) was used to calculate genetic trends, inbreeding levels, and generation intervals. Average PBV, inbreeding, and generation intervals per year were calculated using PROC MEANS. The PROC REG was used to create a linear model to estimate the breakpoints (priors) to be included in PROC NLIN to fit the best line between the breakpoints. The breakpoints were estimated every five years. This aided in plotting genetic trends for the colored breeds when data were limited and based on only a few observations in a given year. Finally, PROC GPLOT and PROC SGPLOT were used to plot the data over time.

Results

Since 1975, there have been 154,602 bulls with at least 10 daughters with milk records in the US evaluation system. Of those bulls, AY, BS, and GU make up less than 2%, whereas HO and JE represent 86.97 and 8.13%, respectively. Interestingly, BS proportionally has the most genotyped bulls. This is a result of the European Brown Swiss Federation Intergenomics project, which established an international database of genotypes for the Brown Swiss populations. In 2005, slightly more than half of the bulls were genotyped, and in 2017 all bulls were genomically tested. Although the number of genotyped cows has increased throughout the years, they still represented only 19.98% of the total cows born in 2017. Proportionally, the JE breed has genotyped the most cows (8.73%); however, the proportion of cows genotyped is still a stark contrast to the bulls (overall 4.62 vs 24.06%).

The most apparent improvement across all traits was observed for production. Since milk historically received so much emphasis in the selection indexes (52% in 1971) and producers were primarily paid for increased production, this is no surprise. Breeding objectives of dairy industries worldwide have been evolving over the years towards functional traits and away from production traits (Miglior *et al.*, 2005). Overall, genetic trends for milk, fat, and protein were similar for both bulls and cows across all breeds. This is partially due to the high genetic correlations between the three yield traits. Tsuruta *et al.* (2004) reported milk-fat genetic correlations ranging from 0.48-0.69. Fat PBV for AY bulls shows a steady increase (Figure 1). Interestingly, the rate of genetic improvement for production traits accelerated in 2009, even though AY did not receive a genomic evaluation until 2013. This is due to producers genotyping the best AY bulls once genomic evaluations were released for HO and JE in 2009, which impacted all the traits. Ayrshire bulls born in 2016 had a mean PBV of 48.12 kg for fat, whereas the ones born in 2012 (i.e. before GS) had a mean PBV of 14.95 kg. A similar trend is evident for the AY cows. In 2006, PBV for fat yield for AY cows became positive (Figure 1). This again highlights the impact of genotyping bulls used when genomic evaluations were initially released.

Genetic trends for BS bulls and cows are comparable, moving upward, especially since 2009 when BS received genomic evaluations. This was also the point where fat PBV became positive (Figure 1). As GU only began receiving genomic evaluations in 2016, and in this dataset, we only have bulls with at least 10 daughters up to 2016, the inference to be made from this trend regarding the use of genomics is limited (Figure 1). However, since 1975, there has been a significant increase in fat production for GU bulls. As our GU cow dataset extends to 2019, we see increased PBV for fat yield since 2009 (Figure 1). This infers

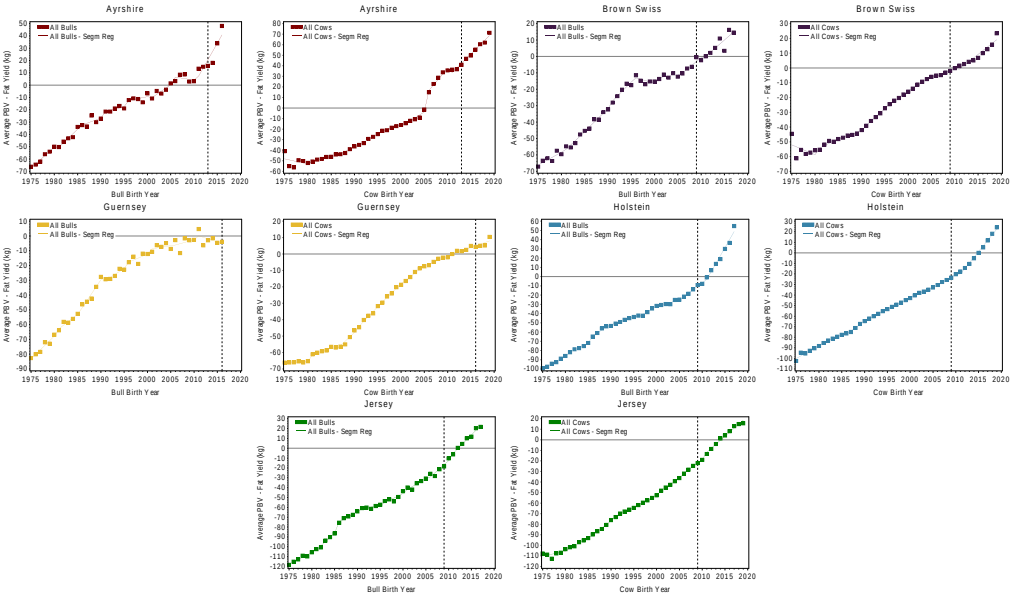


Figure 1. Estimates of genetic change per year from segmented regressions of PBV on birth year for AY, BS, GU, HO and JE bulls and cows for fat yield.

that the best GU were genotyped before receiving a genomic evaluation. Due to a large number of HO bulls and cows, the trends are more defined. HO bulls have potentially made the most progress in the evaluated period. A sharp increase is evident for fat yield since implementing HO genomic evaluations in 2009 (Figure 1). Similarly, fat PBV has increased substantially over time for HO cows (Figure 1). In 2017, JE bulls had an average fat PBV of 21.66 kg. Compared to 2009, this is an increase of 40.04 kg.

Although only fat trends are discussed here, improvements are evident across most breeds for genetic trends. PL increased 4.2-fold for HO bulls since 2009, while JE bulls have increased LIV by 64%. DPR has decreased steadily for JE bulls and cows since 1975; there were 13.98 and 17.97 percentage point reductions, respectively. Because of the increased use of young, genotyped bulls and heifers, a decrease in GI followed the implementation of GS for all breeds, except AY and GU. The GI before and after GS changed from 7.1 to 2.2 years in HO, 7.3 to 3.1 in JE, 8.4 to 5.2 in BS, and 7.6 to 5.3 in AY for sires of bulls. Genomic inbreeding has increased steadily since 2009 for BS cows (6.67-7.82%). The average genomic inbreeding for HO and JE cows in 2019 was 9.1 and 6.4%, respectively. For BS bulls in 2016, genomic inbreeding was extremely high (16.1%); however, it was only one observation and not an average. For HO bulls, the level of genomic inbreeding was 12.73% in 2017, which indicates it would be proper to update GPTA with the genomic inbreeding coefficient. Lastly, JE genomic inbreeding levels averaged 7.22% in 2017.

Discussion

Significant improvements have been made since the implementation of genomic evaluations in the US in 2009. Bulls are making more progress than cows as bulls have more progeny, and a great proportion of bulls (68%) are genotyped versus cows (5%) since 1975. Fat yield increased almost 6-fold for HO bulls between 2009 and 2017. Because of the increased use of young, genotyped bulls and heifers, a decrease in generation interval followed the implementation of GS for all breeds, except AY and GU. As genomic evaluations were only released for GU in 2016, it is too early to see the benefits as clearly as in other breeds. Progress for

smaller breeds will need to be continually monitored to assess the level of genetic progress and to measure the level of genetic gain ten years post-genomic evaluations.

This research provides benchmarks for breed associations to evaluate the updates in genetic evaluations methodology in the genomics era. Genomic information has mainly benefitted Holstein, Jersey and Brown Swiss because the implementation of GS for these breeds happened as early as 2009. Another reason is that HO and JE have larger reference populations and more refined selection indexes. In the short years investigated, there is success in the majority of traits for the breeds other than Holstein. Smaller breeds in the US could significantly benefit from genotyping and phenotyping more of the population and increasing the adoption of genomic methods. This study should be updated in another five years to give these breeds further opportunity to develop.

References

Dechow C.D. and Rogers G.W. (2018) J. Dairy Sci. 101:4312–4316. https://doi.org/10.3168/jds.2017-13571.

García-Ruiz, A., Cole J.B., Vanraden P.M., Wiggans G.R., and Ruiz-López F.J. *et al.* (2016) PNAS. 113(28):E3995-E4004. https://doi.org/10.1073/pnas.1519061113.

Grosu, H. B., Shaeffer L., Oltenacu P.A., Norman D., and Powell R. *et al.* (2013) History of genetic evaluation methods in dairy cattle. The Publishing House of the Romanian Academy, Bucharest, Romania.

Masuda, Y., Vanraden P.M., Misztal I., and Lawlor T.J. (2018) J. Dairy Sci. 101:5194–5206. https://doi.org/10.3168/jds.2017-13310.

Miglior, F., Muir B.L., and Van Doormaal B.J. (2005) J. Dairy Sci. 88:1255–1263. https://doi.org/10.3168/jds.s0022-0302(05)72792-2.

SAS Institute. 2016. Online Doc. Version 9.4. SAS Inst., Inc., Cary, NC.

Tsuruta, S., Misztal I., and Lawlor T.J. (2004) J. Dairy Sci. 87:1457–1468. https://doi.org/10.3168/jds.s0022-0302(04)73297-x.

Wiggans, G.R., Cole J.B., Hubbard S.M., and Sonstegard T.S. (2017) Annu. Rev. Anim. Biosci. 5:309–327. https://doi.org/10.1146/annurev-animal-021815-111422.

427. Community based breeding programme for low input beef cattle production systems in the arid region of Argentina

S. Munilla[1,2*], J.M. Tessi[3] and D.O. Maizon[4]

[1]Universidad de Buenos Aires, Facultad de Agronomía, Departamento de Producción Animal, Av. San Martin 4453 (C1417DSE), Buenos Aires, Argentina; [2]Instituto de Investigaciones en Producción Animal (INPA), CONICET-UBA, Av. Chorroarin 280 (C1427CWO) Buenos Aires, Argentina; [3]INTA, AER Cruz del Eje, Av. Eva Perón 451 (X5280) Córdoba, Argentina; [4]INTA, EEA Anguil, RN 5 km 580 (L6326), La Pampa, Argentina; munilla@agro.uba.ar

Abstract

In this paper, we describe a community based breeding programme designed for beef cattle production systems that range between a smallholder farm and a traditional cow-calf operation. The design involves a network structure with three nodes: a cooperating node, integrated by farmers that organise themselves into a cooperative, a facilitating node that provides technical support, and a centralised nucleus tier. The program was conceptually developed for the case of farmers in the Los Llanos region, La Rioja province, Argentina. After modelling the typical production system of the region, economically relevant traits were identified and a selection index was developed. Finally, the achievable response to selection was estimated. Based on these calculations and projecting the economic benefit over five years, it was estimated that the program would report a 10.6% increase in the producer's gross margins, as a result of an increase in the average weaning weight due to genetic selection.

Introduction

Low input livestock production systems are typically found in regions with less productive potential and far from large markets, conditions that hamper the introduction of their products into formal trade, the purchase of production inputs and the access to financing. In these regions, relatively smallholder farms with limited infrastructure and low adoption of technologies are the norm. These farms, however, generate a large portion of the income necessary to bear the expenses of the family group.

In these contexts, cooperative animal breeding programmes emerge as one of the best known strategies to increase productivity (Olivier *et al.*, 2002). Cooperative programmes involve community based livestock breeding strategies that arise from the needs of the farmers themselves, are proactive, facilitate the organization of the intervening structures and aim not only to attend management of the genetic resources but also to social, cultural and ecological constraints (Mueller *et al.*, 2015, Table 1)

Community based breeding programmes have been developed all around the world for small ruminants (Kosgey *et al.*, 2006, reviewed some of them; a lot of new experiences have been reported since). However, they have attracted little attention to beef cattle farmers. In this paper, we describe a community based breeding programme designed for beef cattle farmers that range between subsistence smallholders and a traditional cow-calf operation. The program was conceptually developed for the case of beef cattle farmers in the Los Llanos region, La Rioja province, Argentina.

Materials & methods

Breeding programme design. The proposed design is based on the general principles discussed by Mueller *et al.* (2015) for community based breeding programmes. The structure involves a network with three nodes: a cooperating node, integrated by farmers that organise themselves into a cooperative, a

R.F. Veerkamp and Y. de Haas (eds)
CC BY

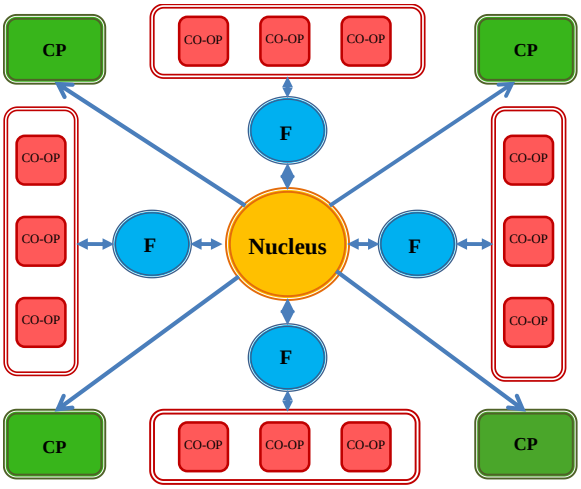
DOI: 10.3920/978-90-8686-940-4_427, © S. Munilla *et al.* 2022

facilitating node, and the nucleus (Figure 1). Co-operators are linked together and with the nucleus through the facilitating node, formed by the professional staffs that provide technical support.

In each production cycle, co-operators transfer to the nucleus the weaned male calves the programme determines. At this point, the selection criterion is based on an intra-herd evaluation coupled with a phenotypic inspection. Once in the nucleus, the young bulls are reared until they are 16-18 months of age, at which time semen is collected from them and subsequently used on cows from the co-operators farms to perform a progeny test. Eventually, this test will provide data on their offspring that will be integrated into the database of a genetic evaluation. With the predicted breeding values, an economic index is calculated and, based on this criterion, outstanding bulls would be ultimately selected to form part of a semen bank. The bank is managed by the nucleus and used to breed back the co-operator's herds via an AI programme.

Case study. Los Llanos, La Rioja province, Argentina, is a typical sample from the arid Gran Chaco biome that comprises about 100,000 km^2 and is framed by mountains ranging 2,000 meters above sea level. It has a dry subtropical climate with annual average temperatures ranging between 17 and 20 °C and rainfall ranging from 500 mm in the east to 300 mm in the west. It is a primary livestock producing region, predominantly dedicated to cattle ranching, characterized by low weaning rates (~50%) and meagre beef yields (~5 kg/ha), mainly associated with the low availability and seasonality of forage and the scarcity of water.

Selection index. To develop an appropriate breeding goal, a cooperative was conceptually created, made up of six cooperating members from the Los Llanos region. Their farms were characterized in terms of both, their productive (management of forage resources, reproductive management of the herd, etc.) and economic (income and expenses, marketing, etc.) factors. Based on this information, a profit equation (Hazel, 1943) comprising three economically relevant traits, birth weight (BW), weaning weight (WW) and mature cow weight (MW), was elaborated. Taking derivatives from the benefit equation with respect to each of these traits, their marginal benefit (or economic weighs) were obtained and finally adjusted by the



F: facilitator; CP: commercial producer; CO-OP: cooperative members.

Figure 1. Community based breeding programme design. F = facilitator; CP = commercial producer; CO-OP = cooperative members.

discounted genetic expressions for terminal sires published by Amer *et al.* (2001, Appendix A) to evaluate the impact that selected bulls would have on this marginal benefit.

Response to selection. Given the selection scheme that derives from the operation of the breeding programme, the response to selection that its potential implementation would generate was calculated. The selection intensity was assessed only for the male's pathway and calculated from the proportion of select bulls in each production cycle. The accuracy, in turn, was that of the progeny testing. Finally, the sire's generation interval was computed by averaging the age of the bulls that make up the semen bank whereas the dam's generation interval corresponded to the average age at which the cows from the co-operators are replaced. For all these calculations, it was assumed that the program was already in place.

Gross margins. The gross margin of each of the six co-operators was calculated and projected over a period of five years. Income derived from the sale of calves (mainly males, as a proportion of the females are retained) and culled cows. In turn, expenses derived from health treatments and veterinary services, labour, and commercialization costs. In addition, feeding costs were calculated relating the grassland area a 400 kg beef cow required in this region and the opportunity cost of this area. It was assumed that over the years the income accrued pulled by increases in the average weaning weight, whereas expenses not directly related to the weight of the calves remained constant.

Results

Table 1 shows the factors that determined the economic values of the selection index proposed for the case study cooperative. The index equation is:

$$\text{Index}(\$) = -0{,}0065 \times \text{EBV}_{BW} + 14{,}95 \times \text{EBV}_{WW} + 0{,}04 \times \text{EBV}_{MW} \quad (1)$$

Notice that the weaning weight is determinant in this index. Then, response to selection was calculated based on this trait.

The response to selection expected per year would be 0.2 genetic standard deviations (σ_A, Table 2). Based on historical data on weaning weight and a presumed heritability of 0.3 for this trait, the genetic standard deviation was estimated as σ_A = 14.4 kg, which represents around 3 kg of annual genetic progress. The

Table 1. Economic values for the traits that compose the selection index.

Trait	Economic weigh ρ	DGE[1]	Economic value $\rho \times$DGE
Birth weight	-0.013	0.500	-0.0065
Weaning weight	32.50	0.460	14.95
Mature weight	2.70	0.014	0.04

[1] DGE = discounted genetic expressions.

Table 2. Response to selection factors.

Pathway	Selection intensity, i	Accuracy, r_{TI}	$i \times r_{TI}$	Generation interval, L	Response to selection, $i \times r_{TI}/L$
Sires	2.097	0.79	23.86	2.50	
Dams	0.000	-	0.00	5.76	
Total			23.86	8.26	0.2 σ_A ~3 kg/yr

projection of the economic benefit over five years indicated a 10.6% increase in the average producer's gross margins.

Discussion

The scale and degree of organization of typical beef cattle farmers in the Los Llanos region, La Rioja province, Argentina, have hindered their involvement in traditional animal breeding programmes. Faced with this scenario, cooperative, community based programmes become an attractive alternative. These programs involve community strategies that arise from the needs of producers and are leveraged on R&D institutions, generally governmental. The link with these institutions is very important, since these programs require great technical, financial and organizational support in their initial stages, as highlighted by Mueller *et al.* (2015).

With that prospect in mind, we designed a community based breeding programme with a centralised nucleus, to which cooperating farmers would contribute by transferring young selection candidate bulls. It was estimated that the implementation of this program would result in a 10.6% profitability increase after 5 years, approximately one generation. Our results are encouraging in terms of the feasibility of a project that would increase the productivity of livestock producers in this region. For this particular case, the National extension service could harbour the nucleus on their local station (INTA, EEA Los Llanos) and provide technical support.

Our analysis did not consider the implementation costs or operating expenses of the cooperative nucleus. However, these expenses could reasonably be covered, once the programme is in place, with the sale of bulls that have been discarded from the selection process and semen straws to other commercial producers in the region. In addition, the programme may also contribute to the more intangible benefit of the conservation and promotion of local animal genetic resources. As it is pointed out in FAO's (2010) 'Breeding Strategies for Sustainable Management of Animal Genetic Resources' guide, these resources contribute to development, food security and poverty alleviation in the regions where they are applied.

References

Amer, P.R., Simm, G., Keane, M.G., Diskin, M.G. and Wickham, B.W. (2001). Livest. Prod. Sci., 67: 223-239. https://doi.org/10.1016/S0301-6226(00)00201-3

FAO. 2010. Breeding Strategies for Sustainable Management of Animal Genetic Resources. Available at: https://www.fao.org/3/i1103e/i1103e00.htm

Hazel, L.N. (1943) Genetics 28: 476-490. Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1209225/pdf/476.pdf

Kosgey, I.S., Baker, R.L., Udo, H.M.J. and Van Arendonk, J.A.M. (2006) Small Rumin. Res. 61: 13-28. https://doi.org/10.1016/j.smallrumres.2005.01.003

Mueller, J.P., Rischkowsky, B., Haile, A., Philipsson, J., Mwai, O. *et al.* (2015) J. Anim. Breed. Genet. 132: 155-168. https://doi.org/10.1111/jbg.12136

Olivier, J.J., Moyo, S., Montaldo, H.H., Thorpe, W., Valle Zarate, A *et al.* (2002) Proc. of the 7th WCGALP, Montpellier, France.

428. A participatory approach to review and update the national breeding objectives for New Zealand dairy cattle

B.F.S. Santos[1*], G.M. Jenkins[2], J. Kok[1], K. Stachowicz[1] and P.R. Amer[1]

[1]AbacusBio Limited, PO Box 5585, Dunedin 9058, New Zealand; [2]Beef + Lamb NZ, P.O. Box 5501, Dunedin 9054, New Zealand; bsantos@abacusbio.co.nz

Abstract

The National Breeding Objective (NBO) for the New Zealand (NZ) dairy industry has been updated in 2021 following a major review. The review included farmer/industry stakeholder consultation, updated methodology to calculate economic values (EVs) for some of the important traits, as well as inclusion of new selection traits. Fertility sub-traits and functionality for non-linear traits were incorporated. The review resulted in new EVs being produced for the updated Breeding Worth index (BW2022) which represents the NBO and is expressed in monetary units (NZ$ per lactation). The changes were modelled to determine the impact to the national dairy herd. The outcomes incorporate farmers' preferences for traits with the intention of increasing adoption levels of the national index as the main reference for selection of dairy cattle in New Zealand. This paper summarizes the NBO review, along with the index testing outcomes and specifications for upgrading BW2022.

Introduction

The NBO for the dairy industry had its previous major review in 2012, and since then it has had body condition score (BCS) added as a new trait and been updated annually for some key performance and price inputs. Following discussions within the dairy sector, New Zealand Animal Evaluation Limited (NZAEL), the DairyNZ wholly owned subsidiary that coordinates national genetic improvement initiatives for dairy cattle, initiated a major revision of the NBO starting in 2020.

The review process consisted initially of stakeholder consultations, most importantly, surveying of stakeholders (industry leaders, scientists, service providers, and breeding companies) and farmers. This consultation was a starting point to inform the wider review process. It was followed by a separate farmer survey which used 1000minds software (Hansen and Ombler, 2009) for trait choice trade-offs, along with parallel survey questions to elicit quantitative differences in trait preferences by New Zealand dairy farmers.

This paper presents the outcomes of the NBO review process, including a high-level summary of the methodology used to meet farmers' trait improvement requirements through EVs representing farm profitability. It also includes an estimate of the impact of changes in the NBO to the national dairy herd.

Materials & methods

The consultation was divided into two specifically designed surveys focused on: (1) stakeholders' views; and (2) dairy farmers' views and trait preferences. The online surveys combined socio-demographic, dairy system, and trait preference questions, similar to the methodology used by Martin-Collado *et al.* (2015). The stakeholder survey reached 280 complete responses and the farmer survey 439 complete responses. Principal component (PCA) and cluster analysis (CA) were used to define dairy farmer typologies and their association with respondents' views.

Following the surveys, the NBO model which calculates EVs for several traits independently of each other (DairyNZ, 2021), was updated to incorporate core inputs. This included changes to milk prices, feeding costs, replacement heifer costs, and other parameters that impact profitability of the dairy operation. A

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_428, © B.F.S. Santos *et al.* 2022

new methodology for fertility EV was developed as the trait was separated out into heifer and cow calving season day (CSD0 and CSD123, respectively), and changed from binary CR42 (% cows calving in first 6 weeks from planned start of calving) to continuous CSD. A new methodology was created to convert residual survival (expressed as the additional average herd age in days) into functional survival (defined as % cows surviving from one parity to the next). Finally, a non-linear model was incorporated for udder overall, a new BW trait where many farmers tend to avoid very bad udders, but do not necessarily want to give up too much progress in other traits to achieve excellent udders.

The changes to BW were modelled to determine the impact of such changes to the national dairy herd through methodology proposed by Zhang and Amer (2021). The analysis was applied to Jersey, NZ Holstein-Friesian, Overseas Holstein and Jersey Holstein-Friesian cross bulls born between 2011 and 2015 in four separate analyses. To be included, bulls had to have at least 50 daughters with production and fertility records, which resulted in a minimum reliability of 85% for milk and 67% for fertility. A minimum reliability threshold of 10% was used for all other trait estimated breeding values (EBVs).

Results

The consultation achieved reasonable coverage of the quite diverse NZ dairy industry, with all farming systems, breeds, and regions participating. There was a lower level of support for the BW index than expected (55% agree with its importance). Fertility and udder traits were stated as the most critical improvement requirements, followed by susceptibility to mastitis and lameness. Environmental traits were also considered important, albeit with substantial variation in stated degree of priority across respondents. On average, farmers are willing to trade-off about 20% of genetic progress in milk solids for progress in Greenhouse gas (GHG) emissions and nitrogen leaching mitigation traits.

Two clusters of farmers were identified according to their preferences for traits (Figure 1). Cluster 1, formed by farmers with strong preference for milk solids, fertility, somatic cell count (SCC), and body condition score; and Cluster 2, formed by those with strong preference for udder, mastitis, and lameness. These preferences were based on intrinsic values and independent of systems, breeds, or region.

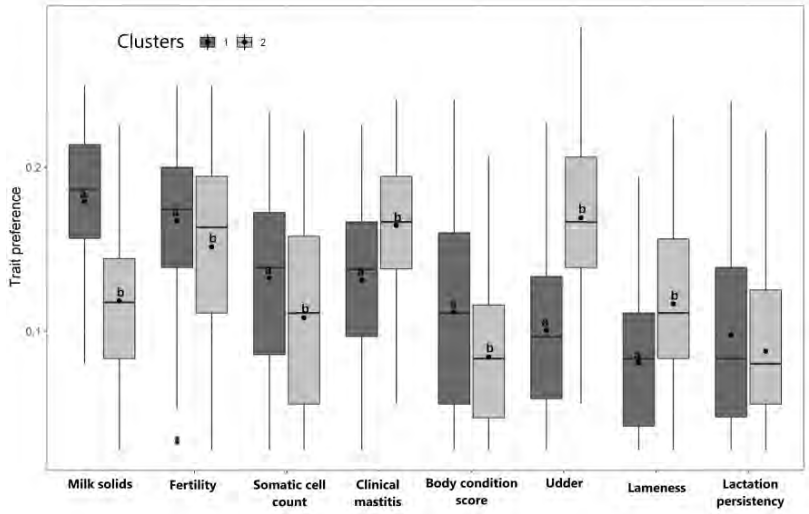


Figure 1. Dairy farmers' clusters based on trait preferences.

The updated trait EVs (BW 2022) are presented in Table 1. The current economic environment, with record high milk prices, resulted in higher emphasis on milk production traits. A small increase in penalty for milk volume (-9.5c/l) resulted from penalties imposed by milk processors and feed costs associated with lactose synthesis. There were no changes implemented for methodology used in the live weight EV calculation, with a further review planned to address the major implications for both within and between breed rankings. The same economic principals were applied, despite the change in the definition of survival, and rising costs of rearing replacement heifers (~NZ$2,000 per heifer), better salvage revenue from cull cows as well as greater loss from culling or death during lactation (impacted by the milk price) were incorporated. The higher SCC penalty is consequence of the increased impact of cell counts on higher value milk and higher replacement costs associated with increased probability of clinical mastitis. There was an increase in the EV of fertility, mostly due to the higher replacement costs, and the separation between heifer and cow fertility traits causing increased effective dispersion in the fertility EBVs. Finally, the increase in BCS EV was largely driven by higher milk prices surpassing the increase in feed costs.

The impact of changes to the NBO were represented both as correlations between trait EBVs and indexes (Table 2 – results averaged from four different breed data sets of bulls), as well as trait responses over a period of 10 years (Table 3).

Discussion

Industry consultation is a powerful tool to improve engagement and to gather direction in the formulation of breeding objectives represented by economic selection indexes. The surveys are not used to establish trait EVs, but rather to identify areas where current EV calculations might be missing true and perceived economic impacts of trait changes. Trait EVs remain strongly influenced by economic modelling, where the impact of each trait is calculated independently. Index formats and gaining a mandate for inclusion of specific traits and economic modelling (e.g. non-linear approaches) and benefited from the industry consultation (Santos *et al.*, 2019).

The consultation indicated that farmers and stakeholder respondents would favour stronger selection effort on functional traits (especially udders), more progress on fertility, and development of national standard auxiliary indexes which address specific industry challenges, e.g. high per cow output systems, once a

Table 1. Economic values (EV) of traits for the updated BW 2022 and current BW 2020.

Trait (units)	Economic Value (NZ$/unit change)		% change
	Updated BW 2022	Current BW 2020	
Milk fat (kg)	$5.18	$4.25	+22%
Milk protein (kg)	$5.21	$4.26	+22%
Milk volume (l)	-$0.0951	-$0.094	+1.1%
Live weight (kg)	-$1.38	-$1.38	0%
Functional survival (% surviving)	$2.65	-	-
Somatic cell score (log of cell counts/ml)	-$42.89	-$37.11	+16%
Fertility (% CR42)[1]	$6.33	$5.92	+7%
Body condition score (units)	$116.93	$101.96	+15%
Udder overall (score)[2]	-$71 to +29	-	-

[1] Fertility was split into CSD in heifers ($-2.40/day) and cows ($-8.08/day) and converted to CR42 at $6.33 per %.

[2] Non-linear function to determine the EV, represented by the quadratic equation $Y=-22.912x^2+54.692x$, where x is the udder overall EBV where Y is the contribution of the Udder overall (score) to an animal's index value.

Table 2. Trait EBV correlations between BW2022 and BW2020 for all breeds.

Trait	BW2020	BW2022
Milk volume	-0.005	0.153
Milk fat	0.734	0.807
Milk protein	0.379	0.520
Somatic cell score	-0.080	-0.083
Fertility (%CR42)	0.258	0.350
Fertility (CSD0)	-0.246	-0.328
Fertility (CSD123)	-0.249	-0.338
Functional survival	0.270	0.283
Body condition score	0.143	0.154
Live weight	-0.230	-0.072
Udder overall (linear)	-0.080	0.069
Udder overall (non-linear)	-0.037	0.119

Table 3. Expected trait unit responses after 10 years of selection on current BW2020 and updated BW2022 by breed.

Traits	Jersey		Crossbred		Holstein Friesian		Overseas Holstein	
	BW2020	BW2022	BW2020	BW2022	BW2020	BW2022	BW2020	BW2022
Milk (kg)	59.36	93.96	-129.7	5.93	-57.07	28.35	83.54	148.22
Fat (kg)	14.26	14.84	12.5	14.15	15.24	16.71	14.71	16.72
Protein (kg)	6.98	7.74	0.91	4.91	3.58	5.94	9.35	10.67
SCC (log cells/ml)	-0.07	-0.07	-0.08	-0.08	-0.04	-0.04	0.03	0.02
Fertility (% CR42)	1.43	2.05	2.16	3.06	1.37	2.04	1.8	2.11
Fertility (CSD0)	-0.337	-0.531	-0.581	-0.876	-0.578	-0.855	-1.100	-1.185
Fertility (CSD123)	-1.017	-1.447	-1.515	-2.139	-0.917	-1.370	-1.136	-1.348
Func. survival (%)	1.04	1.12	0.67	0.74	0.69	0.6	-0.09	-0.04
BCS (score)	0.05	0.06	0	0.02	0.01	0.01	0.03	0.02
Liveweight (kg)	-0.98	1.54	-16.38	-6.02	-5.87	0.16	-7.07	-5.06
Udder over. (score)	-0.56	3.64	-1.11	4.13	-0.98	2.92	-0.13	0.09

day milking, and environmental issues. The non-linear approach to udder overall was well received by stakeholders as it allows poorer bulls to be penalised, but without too much sacrifice in progress in other traits. Correlations between indexes showed that if udder overall had no true value for some farmers, the inclusion of the trait in the BW would cause a loss between 1 and 2% of genetic progress in profitability. The expected improvements in perception and level of support for the national index approach should more than offset these relatively modest impacts.

The framework has been established for incorporation of further changes to BW. The engagement and consultation highlighted under-emphasis on cow functionality as an area of concern relating to the index in use. The BW 2022 has greater weighting on fertility, and udder traits have been included in a way that efficiently penalizes very bad udders. The combination of surveys, rigorous testing, and description of the key input assumptions and methodologies allowed a diverse farmer advisory panel to reach a high level of consensus on the new index.

References

DairyNZ (2021), Economic Values. Available at: https://www.dairynz.co.nz/animal/animal-evaluation/interpreting-the-info/economic-values/

Hansen P., Ombler F. (2009) J. Multi-Crit. Decision Anal. 15:87-107 http://doi.org/10.1002/mcda.428

Martin-Collado D., *et al.* (2015) J Dairy Sci. 98(6). http://doi.org/10.3168/jds.2014-9194

Santos B., *et al.* (2019) Proc. Assoc. Advmt. Anim. Breed. Genet. 23:55-58. http://www.aaabg.org/aaabghome/AAABG23papers/14Santos23055.pdf

Zhang X., Amer P. A. (2021) J Dairy Sci. 104(5). https://doi.org/10.3168/jds.2020-19547

429. Genetic trends for carcass traits within breeds in a multi-breed evaluation

S.A. Mulhall[1,3*], R.D. Sleator[3], R.D. Evans[2] and A.J. Twomey[1]

[1]Teagasc, Moorepark, Fermoy, Co. Cork, Ireland; [2]Munster Technological University, Bishopstown, Co. Cork, Ireland; [3]Irish Cattle Breeding Federation, Highfield House, Shinagh, Bandon, Co. Cork, Ireland; shauna.mulhall@teagasc.ie

Abstract

The focus of this research is to determine the genetic trends in carcass traits within each of the common sire breeds used in both the dairy and beef populations. Estimated breeding values for carcass weight, carcass conformation, and carcass fat were available on 8,375,561 animals from the national carcass genetic evaluation. Larger genetic gain was observed within breeds for beef sires used on beef dams in comparison to beef sires used on dairy dams for all carcass traits. There was a relatively small genetic gain observed across breed for beef sires used on dairy dams although a significant genetic gain was observed within these sire breeds, suggesting breed substitution.

Introduction

Most genetic evaluations and breeding objectives worldwide for beef traits tend to be on a within breed basis (Johnston *et al.,* 2018). Ireland is relatively unique in publishing national multi-breed genetic and genomic evaluation for carcass traits (Purfield *et al.,* 2019). Furthermore, large diversity exists across beef production systems in Ireland, which include suckler-based systems, where cows rear calves for beef (Hocquette and Chatellier, 2011) and dairy systems, where progeny surplus to replacements are reared for slaughter. Genetic selection for beef traits is included in the four national breeding objectives: the economic breeding index for dairy cows, the dairy-beef index for beef sires used on dairy cows, the terminal index to breed animals for beef production, and the replacement index to breed replacement beef cows. Genetic progress has been observed for a range of species and traits including fertility in cattle (Berry *et al.,* 2016) and carcass traits in broilers (Tavárez and Solis de los Santos, 2016). A potential concern for selection based on multi-breed evaluations, and breeding objectives, is that genetic gain observed is due to breed substitution and that there may be little or no genetic gain achieved within individual breeds. The objective of this analysis was to quantify the genetic gain in carcass traits within each breed, for sires used in dairy and beef herds.

Materials & methods

Data. Estimated breeding values (EBV) for carcass weight, carcass conformation, and carcass fat were available from the September 2020 Irish carcass genetic evaluation, from an existing database managed by the Irish Cattle Breeding Federation (ICBF). This multi-trait and multi-breed evaluation is one of the routine evaluations run by ICBF every two months, using Mix99 software (Strandén and Lidauer, 1999) generated EBV's for 20,236,674 animals.

Data edits. Animals with an unknown sire or dam were excluded; after which data on 14,570,772 animals remained. Animals were defined as either beef or dairy based on their dam breed proportion. Dairy dams were defined as dams with a dairy breed (Holstein-Friesian, Jersey, Ayrshire, Norwegian Red, Montbeliarde, Brown Swiss, Normande, and Rotbunt) composition of ≥87.5%. Beef dams were defined as dams with a beef breed (Aberdeen Angus (AA), Belgium Blue (BB), Charolais (CH), Hereford (HE), Limousin (LM), Simmental (SI), Aubrac, Blonde d'Aquitaine, Piedmontese, Parthenaise, and Saler) composition of ≥45%. Animals whose dams were not defined as either beef or dairy were excluded. The sire of each animal was

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_429, © S.A. Mulhall *et al.* 2022

required to be at least 87.5% of its known breed composition. Progeny sired among the most common breeds were retained. For dairy dams only progeny sired by AA, BB, CH, HE, LM were included and for beef dams on only progeny sired by AA, BB, CH, HE. LM, and SI were retained. To ensure an adequate representation of animals per year; animals born prior to 2002 were not considered further for dairy dams and animals born prior to 2005 to beef dams were not considered further. Number of animals sired by each breed and dam type is presented in Table 1.

Data analysis. The mean of the progeny sire EBVs were estimated for each progeny birth year for carcass weight, carcass conformation, and carcass fat. These means were calculated within sire breed (AA, BB, CH, HE, LM) used on dairy dams, and within sire breed (AA, BB, CH, HE, LM, SI) used on beef dams. These means were also calculated on an across breed basis for both beef sires used on dairy and beef dams. A linear regression model was used to estimate the genetic gain within and across each breed from 2008 to 2018, inclusive, for beef sires used on dairy and beef cows separately. The dependent variables were carcass weight, conformation, and fat score. Birth year of progeny was included as a fixed effect. In an additional analysis, sire breed (AA, BB, CH, HE and LM) for dairy dams, and sire breed (AA, BB, CH, HE, LM and SI) for beef dams, was also included as a fixed effect to identify if genetic trend was due to a breed substitution effect.

Results

Dairy dam × beef sire. A genetic gain of 0.15 kg was observed across the most common beef sire breeds (AA, BB, CH, HE and LM) used on dairy dams for carcass weight for the years 2008 to 2018, inclusive. However, when breed was adjusted to a common breed, the trend in carcass weight increased to 0.62 kg per year. The trend for carcass conformation, and carcass fat, reversed when sire breed was adjusted to a common breed. A 0.04 unit decrease was observed per year in conformation EBV's, however the genetic trend reversed when breed was accounted for showing an increase of 0.05 units per year. Similarly, the trend for carcass fat showed an increase of 0.03 units per year for sire EBVs. However, when adjusted for breed, the genetic trend decreased by 0.03 units per year. Carcass weight increased for AA, BB, CH, HE and LM sires used on dairy cows from 2008 to 2018 (Figure 1); the genetic gain was 0.84, 1.27, 1.27, 1 and 0.88 kg per year, from 2008 to 2018 for AA, BB, CH, HE and LM. For carcass conformation sire EBVs increased by 0.01, 0.06, 0.01 and 0.06 units per year for AA, CH, HE and LM, respectively (Figure 2); BB declined by 0.01 units per year, respectively. Finally, a reduction in carcass fat was observed within all five breeds, AA, BB, CH, HE, LM, which was 0.02, 0.04, 0.01, 0.02, 0.01 units respectively, per year.

Beef dam × beef sire. The linear regression for the most recent 10 years for the most common beef sire breeds (AA, BB, CH, HE, LM and SI) used on beef dams for carcass weight, carcass conformation and carcass fat followed the same trend; irrespective of the inclusion or exclusion of sire breed in the model. Carcass weight improved 1.18 kg per year for beef sires used on beef dams, but was 0.83 kg per year when sire breed was accounted for. Similarly, the trend for carcass conformation increased 0.04 units per year irrespective of accounting for sire breed. Carcass fat reduced 0.01 units per year, however, when adjusted to a common sire breed there was 0.04 unit reduction per year for carcass fat. Genetic gain in carcass weight was 1.25, 1.43, 1.54, 1.13, 1.32 and 1.67 kg for AA, BB, CH, HE, LM and SI, respectively, per year from

Table 1. Number of animals sired by each breed.

Dam type	AA	BB	CH	HE	LM	SI	Total
Dairy	823,708	193,804	75,902	496,922	272,519	-	1,862,855
Beef	549,018	322,677	2,570,992	201,001	2,461,768	407,205	6,512,706

2008 to 2018 (Figure 1). For carcass conformation, the largest genetic gain was observed in the continental breeds BB, CH, LM and SI being 0.05, 0.05, 0.07, and 0.05 units per year respectively, whereas a smaller genetic gain was observed in the more traditional breeds AA and HE being 0.01 units and 0.01 units per year respectively (Figure 2). Similar to the previous analysis on dairy cows, a reduction in carcass fat was observed within breed for AA, CH, HE, LM, SI which was 0.02, 0.02, 0.03, 0.02, 0.02 units respectively per year, except BB which remained constant for the years 2008 to 2018.

Discussion

Genetic gain for each of the carcass traits was achieved within each of the sire breeds used on both the dairy and beef dam populations with the exception of conformation in BB sires used on dairy dams. A slight reduction in carcass conformation for BB sires used on dairy cows suggests selection within the breed has put higher emphasis on other traits, such as calving traits.

The genetic gain was larger in beef sires used on beef cows compared to the gain in beef sires used on dairy cows. The poorer genetic gain for carcass traits in beef sires used on dairy cows relative to beef cows is likely due to the higher selection emphasis on calving traits which is known to be negatively correlated (Berry

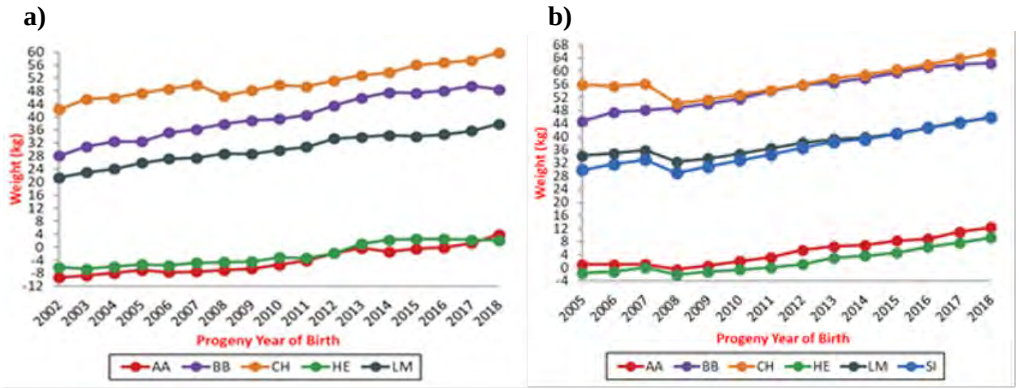


Figure 1. Genetic trend of sire EBVs for carcass weight within breed used on (a) dairy dams and (b) beef dams.

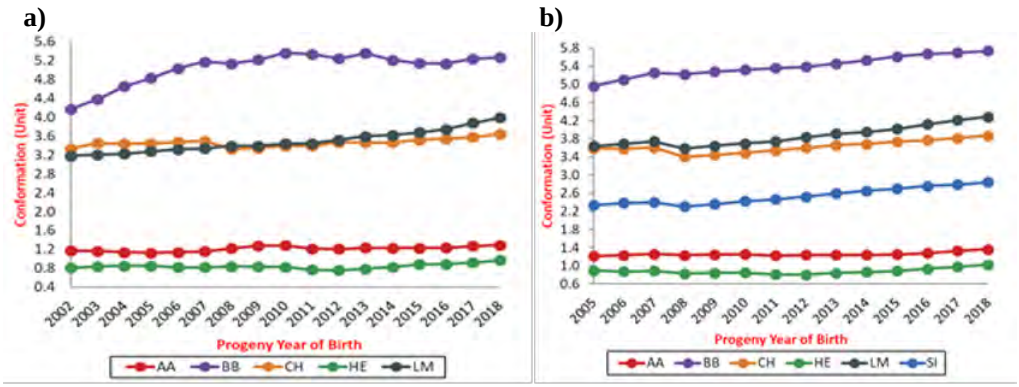


Figure 2. Genetic trend of sire EBVs for carcass conformation within breed used on (a) dairy dams and (b) beef dams.

et al., 2020). Furthermore, the breed substitution observed for genetic trends in beef sires used on dairy cows is linked to a higher selection towards breeds that are renowned for their ease of calving, such as AA and HE. Interestingly, with regards to carcass conformation, the genetic gain estimated within each breed for sires used on both dairy and beef dams was similar except for traditional breeds such as the AA and HE, where a lower genetic gain was observed. The lower genetic gain in these two breeds for conformation is potentially due to the lower genetic variability for carcass conformation within AA and HE compare to LM and CH, (Kause *et al.*, 2015). In additional analysis, the genetic gain in standard deviation units within the traditional breeds, AA and HE, were still lower for carcass conformation (0.02) in comparison to continental breeds like CH and LM 0.08 and 0.09 respectively. Thus the lower genetic gain observed in these two breeds is likely due to a higher intensity of selection for other traits like calving in comparison to the selection of carcass traits in continental breeds.

The genetic gain within sire breeds used on dairy cows suggest that dairy farmers are selecting higher beef merit sires within breed. However, higher gains could potentially be achieved for carcass traits in dairy-beef animals if dairy farmers selected across breed. Furthermore, there was no breeding objective available for beef sire selection for mating on dairy cows in Ireland until 2019 (Berry *et al.,* 2019). Adoption of this breeding objective will likely lead to an improvement of genetic gain across breed for beef sires used on dairy cows. In contrast, there was no breed substitution in the genetic trends for beef sires used on beef dams. Thus, selection of sires in beef dams has been within and across breed.

Acknowledgements

Funding from the Irish Department of Agriculture, Food and Marine Research STIMULUS research grant DAIRY4BEEF is gratefully acknowledged.

References

Berry, D. P., Friggens, N. C., Lucy, M., and Roche, J. R. (2016) Annu. Rev. Anim. Biosci. 4: 269-290. https://doi.org/10.1146/annurev-animal-021815-111406.

Berry, D. P., Amer, P. R., Evans, R. D., Byrne, T., Cromie, A. R., *et al.* (2019) J. Dairy Sci. 102(11): 10056-10072. https://doi.org/10.3168/jds.2019-16912.

Berry, D. P., Ring, S. C., Twomey, A. J., and Evans, R. D. (2020) J. Dairy Sci., 103(2): 1701-1710. https://doi.org/10.3168/jds.2019-17430.

Hocquette, J. F., and Chatellier, V. (2011) *1*(2): 20-28. https://doi.org/10.2527/af.2011-0014.

Johnston, D. J., Ferdosi, M. H., Connors, N. K., Boerner, V., Cook, J. *et al.* (2018) Proc. of the 11th WCGALP, Auckland, New Zealand.

Kause, A., Mikkola, L., Stranden, I., and Sirkko, K. (2015) Animal 9(1): 35-42. https://doi.org/10.1017/S1751731114001992.

Purfield, D. C., Evans, R. D., and Berry, D. P. (2019) BMC Genom. 20(1): 720. https://doi.org/10.1186/s12864-019-6071-9.

Strandén, I., and Lidauer, M. (1999) J. Dairy Sci. 82(12): 2779-2787. https://doi.org/10.3168/jds.s0022-0302(99)75535-9.

Tavárez, M. A., and Solis de los Santos, F. (2016) Anim. Front. 6(4): 37-41. https://doi.org/10.2527/af.2016-0042.

430. Improving growth performance for dual purpose chickens for challenging conditions in Africa

N. Duijvesteijn[1,2*], L. Perrault[2], R. Atikpakpe[3], N. Garcia Martinez[2], J. Renckly[2], T. Arbeau[2] and J.A.M. van Arendonk[1]

[1]Hendrix Genetics, P.O. Box 114, 5830 AC Boxmeer, the Netherlands; [2]SASSO, Route de Solferino, 40630 Sabres, France; [3]Selection Centre Coq de Faso, Boussé, Burkina Faso; naomi.duijvesteijn@hendrix-genetics.com

Abstract

Poultry provides rural families with a low cost, highly efficient and nutritious protein. Providing access to improved chicken breeds suitable for African conditions is one of the key elements to increase the rural poultry productivity at small scale farmer level. A breeding program (BP) including performance testing under African conditions is essential to be able to improve these chicken breeds. We show that we can increase the body weight of a pureline breed housed in France by including crossbred offspring performance from Burkina Faso into the BP of SASSO. The selection differential increased from +15 to +52 considering including and selecting for the performance measured in Burkina Faso. Running an effective BP for developing the best performing breeds under African conditions does not only require collecting performance information locally but also keeping the pureline in a bio-secure environment. Through such a program it is possible to serve customers world-wide consistently.

Introduction

Poultry constitutes an important economic activity for the rural poor in many African countries. Additionally in many poor rural families, poultry provides the family with a low cost, highly efficient and nutritious protein. The local indigenous breeds can be inefficient and unproductive compared to other alternative and improved breed options. Unfortunately, the smallholder farmers in rural areas often do not have access to these improved genetics. Providing access to an improved low-input and dual-purpose chicken could be a contributing element to transforming the rural poultry production chain.

In a number of African countries investments are made for providing access to improved breeds such as the African Poultry Multiplication Initiative (APMI). The APMI model generally operates through capable local private companies to establish a parent stock and hatchery operation for the production of day-old-chicks of improved dual purpose chicken breeds for smallholder farmers. The hatching eggs from the parent stock farm and set and hatched to produce day-old-chicks. The vaccinated day-old chicks are transported to so-called 'Brooder Units' (equivalent to 'Mother Units' or 'Ambassadors') who are specialized in the brooding, feeding, and proper vaccination process for the first 30-40 days of the chicks' life. The brooder units sell these chicks to smallholder farmers for meat and egg production. Through this APMI system, local smallholder farmers get access to improved genetics and the chickens have a high chance to survive due to their good start.

In 2018 Hendrix Genetics received a grant from the Bill & Melinda Gates Foundation to further enhance the use of improved poultry genetics to provide a better living to smallholder farmers in Africa. The program named 'Sustainable Access to Poultry Parent Stock for Africa (SAPPSA)' aims to: (1) secure access to poultry parent stock; (2) grow the APMI-like initiatives within and across countries; and (3) improve the dual-purpose breeds to better meet the needs of African smallholder farmers. In SAPPSA, we built on the experience and the dual purpose breeds developed within SASSO, our specialized colored broiler breeding division.

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_430, © N. Duijvesteijn *et al.* 2022

In this conference paper, we focus on the recurrent test program which we have developed and incorporated in the breeding program for dual purpose birds to better fit the challenging conditions in rural Africa.

Breeding. As part of the SAPPSA project, we developed and implement a recurrent testing scheme to improve performance of dual purpose poultry under African conditions. The recurrent test is directed towards improving performance at the crossbred level under local circumstances, whereas selection is performed at pureline level. To improve performance of the dual-purpose birds, both growth and egg performance on crossbred chickens are measured. In this paper, we present results only on growth performance in which BW_PL reflects the body weight measured at pureline level(BW_PL). Pureline males are used to produce crossbred offspring. The crossbred offspring are measured in our recurrent test facility in Burkina Faso (BW_BF; open housing, local feed supplier and high temperatures) and at recurrent test facility in France (BW_FR; closed housing; label rouge certified feed; mild climate). The recurrent test performances have been included in the genetic evaluation of the pure line.

The data. The data collected is on the slow growing broiler line which is used to produce the commercial cross SASSO Ruby C for smallholder farmers in Sub Saharan Africa. Both pureline data of the line Ruby as crossbred data between a pureline Ruby sire and another pureline line (not always the same) are included in the genetic analyses. Table 1 shows the number of observations in the different tests. Body weight was recorded on in total 7,544 pureline animals and 11,853 crossbred animals of which 7,517 in Burkina Faso.

Important to note is that flock number 1 in France were full sibs of the flock number 2 that was tested 2 weeks later in Burkina Faso. This close genetic link enabled us to estimate the contribution of genotype-by-environment effect from a pureline-crossbred effect. The pureline data consists of 8 discrete generations of animals which were all housed at our bio-secure pedigree complex in Sabres, France.

Fixed effects fitted were sex and date of birth. Linear mixed models were used to estimate the variance components and the data was fitted in the program DMU (Madson and Jensen, 2008). Bivariate or tri-variate analyses were run for BW, where the traits were defined by being measured either in pureline or crossbred animals. In the trivariate analyses, the BW in Burkina Faso, the BW measured in France on crossbreds and purebreds were treated as separate traits. The resulting genetic correlations do reflect a combination of pureline-crossbred performance effect and/or genotype-by-environment effect. Covariances between

Table 1. Overview of the dataset with number of observations, mean and SD of the body weights for pureline and its crosses.

Line	Location	N sires tested[1]	Sex[2]	n	Mean BW	SD BW
Ruby PL	Pureline facility, France		F	4,002	977	224
			M	3,542	1,069	231
Flock 1: Ruby XB	France	81	F	2,130	996	102
			M	2,206	1,078	110
Flock 2: Ruby XB	Burkina Faso	73	F	1,471	630	82
			M	1,362	691	93
Flock 3: Ruby XB	Burkina Faso	31	F	1,680	637	111
			M	1,757	749	123
Flock 4: Ruby XB	Burkina Faso	27	F	642	533	77
			M	605	571	76

[1] Sires with >20 offspring tested.

[2] F = females; M = males.

residual effects in the pureline and crossbred dataset were set to zero, as animals did not have observations in different environments.

To mimic the breeding program, we compared the expected selection gain from selection of sires sires based on EBV calculated on three distinct sources for BW: BW_PL, BW_BF and BW_FR. For each scenario, we selected 30 sires, which reflects the number of sires selected to produce the next generation of purebred animals.

Results & discussion

The h^2 estimated for BW_PL, BW_BF and BW_FR were 0.43, 0.39 and 0.63 respectively (with a standard error of 0.02, 0.03 and 0.09). The genetic correlation based on the bivariate model between pureline and crossbred was 0.45 (se ±0.10) for both data from France and Burkina Faso. The genetic correlations from the trivariate analyses was 0.39 (±0.11) between BW_PL and BW_BF, 0.67 (±0.11) between BW_PL and BW_FR and 0.63 (±0.09) between BW_FR and BW_BF showing clearly a significant pureline – crossbred interaction and genotype-by-environment interaction.

In Figure 1 the realized genetic trend of the three BW traits is shown. The selection has included the RT data from Burkina Faso into the index since generation G005 which resulted in higher genetic progress for the RT_BF and RT_FR traits in G006. Previous generations have solely been selected on BW measured on pureline birds. From generation G005 onwards, 70 sires have been tested in the two recurrent test environments (preselection step to select the best 70 sires based on an index including BW but also other traits). In total 30 sires were selected to renew the generation.

Table 2 shows the differences in selection differential between the three selection scenarios. For each scenario, the breeding goal is to improve BW under challenging conditions in Africa (BW_BF). This tables shows, that by using RT data from Burkina Faso and selecting on EBV for BW_BF, we could more than double the selection differential for the sires (+52 compared to +15) compared to using purebred information. This clearly illustrates the opportunities for making more genetic progress for the environment where the customers are growing the broilers.

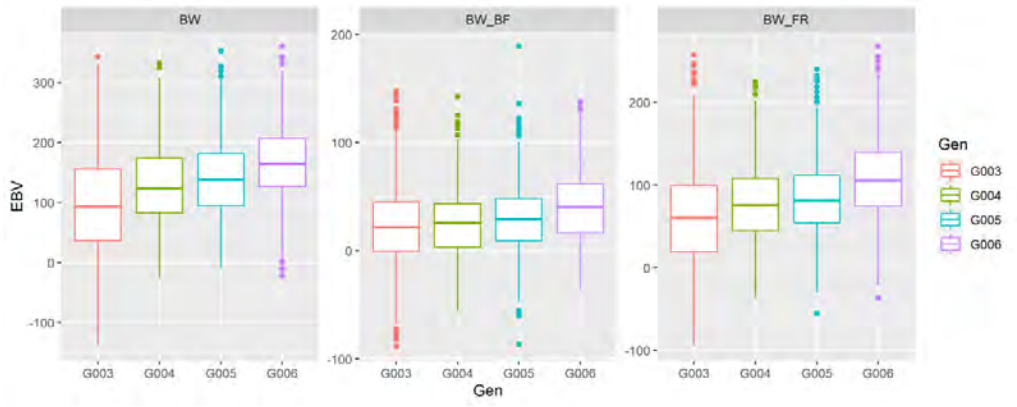


Figure 1. Overview of the genetic trend for the last 4 pureline generations for BW_PL measured on pureline animals, BW_BF measured in Burkina Faso and BW_FR measured in France.

Table 2. The realized selection differential for BW_BF (g/d) for three scenarios which differed in selection criteria.

Selection criteria	Mean all 70 sires	Mean for top 30 sires	Selection differential
EBV BW_PL	28	43	+15
EBV BW_FR	28	52	+24
EBV BW_BF	28	80	+52

Conclusions

Meeting the needs of smallholder farmers in Africa does not only involve developing improved genetics but also ensuring access to the improved genetics. We have learned that the APMI model is a very effective way to provide smallholder farmers access to improved genetics produced by local parent stock farms. By setting-up an international distribution network, we have secured access for our parent stock customers in Africa to high quality material developed in France. To further improve the genetics, we have developed and implemented a recurrent test. The recurrent test enables us to improve the breeding program of dual-purpose breeds by capitalizing on information collected under the challenging conditions in Africa. The results show that both pureline – crossbred correlation and genotype-by-environment interactions exist. This demonstrates the benefit of including performance information collected locally. Running an effective breeding program for developing the best performing breeds under African conditions does not only require collecting performance information locally but also securing the pureline by keeping it in a bio secure environment. Through such a program it is possible to serve customers across the world consistently.

Acknowledgements

This work was supported, in whole or in part, by the Bill & Melinda Gates Foundation with grant number OPP1190193. The authors would like to acknowledge all staff involved at Sasso, France and at the Selection Centre Coq de Faso, Burkina Faso for their contribution to establishing and executing the recurrent test.

References

Madsen P. and Jensen J. (2008): An user's guide to DMU: A package for analysing multivariate mixed models. Aarhus University, Tjele, Denmark. Available at: www.dmu.agrsci.dk.

431. Genomic selection in the US: where it has been and where it is going?

K.J. Retallick[1*], D. Lu[1], A. Garcia[1] and S.P. Miller[1,2]

[1]Angus Genetics Inc., 3201 Frederick Ave., Saint Joseph, MO 64506, USA; [2]AGBU, a joint venture of NSW Department of Primary Industries and University of New England, 2351, Armidale, Australia; kretallick@angus.org

Abstract

Genomic selection (GS) has gained extensive ground in the past decade. The American Angus Association has led the way for incorporation of GS in the US beef industry. Since 2009, GS has been employed for Angus breeders to make selection decisions. Several updates to increase the accuracy of GS has occurred over the years. The most notable being the shift from multistep approaches to single-step genomic BLUP methods to predict genetic differences among individuals. Now after nearly a decade, GS has started to tout its impacts across the industry. Over one million American Angus animals have been genotyped and now enter Angus's genetic evaluation each week. Increased predictive ability and improved rates of genetic progress have supported the growth of this technology across the industry. Selection candidates can now be selected and purchased with more confidence.

Introduction

The American Angus Association (AAA) is the largest beef breed association in the world with over 300,000 annual registrations. The demand for Angus genetics has been one of the strongest in the United States (US). Undoubtedly, some of this success has been driven by the genetic improvement programs that have been offered via AAA. In 1957, AAA launched the Angus Herd Improvement Records program to facilitate the necessary data capture for genetic improvement. The first estimated breeding values on American Angus cattle were published in the Group 1 Report-*Sire Evaluation Report* in 1974. Since that time, significant changes to genetic evaluations have taken place.

Genomic selection (GS) may be one of the most notable changes. Making its debut in late 2009, GS has been making headlines ever since most recently being that of the millionth animal being added to Angus's genomic database in July 2021. While this milestone was certainly celebrated, the success of genomic testing has come about through a series of improvements. This paper's objective is to provide a status overview of GS in American Angus cattle and perspective on how GS has affected producer decisions and the opportunities that lie ahead.

Materials & methods

To illustrate the trend in genetic testing over time, trends were acquired from the AAA. To depict the genetic progress made prior to and presently with the use of GS, genetic trends and variances were acquired. Genetic trend was standardized using each trait's genetic standard deviation. Standard units across traits were determined by regressing average expected progeny differences (EPD) by year from 2001-2009 (prior) and 2010-2021 (present).

Results

Looking back. Companies including Zoetis Animal Health (formerly Pfizer) and Neogen (formerly Igenity owned by Merial Animal Health) took the early initiative to bring GS to the industry. Molecular value predictions described genetic variation captured by markers alleles. These created confusion by placing multiple predictions on the same animals. It was not until AAA made the step towards incorporating GS

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_431, © K.J. Retallick *et al.* 2022

inside genetic evaluation models that widespread adoption took place. The first Angus genomic-enhanced EPD were introduced late 2009, a first for the US beef industry. At the same time, AAA moved to the industry's first weekly genetic evaluation, which was a major change from the previous bi-annual schedule. Enabling breeders to utilize genomic information as quickly as possible was a necessity to increase relevancy of genomic testing, a weekly genetic evaluation supported that initiative.

The first models to augment Angus EPD with genomic information was with a multistep evaluation described by Kachman (2008) and was employed for routine evaluation until 2017. Genomic information was included into genetic evaluations by computing molecular breeding values then fitting these values into genetic models through a correlated trait approach. The challenge to this method was the reliance on training populations, the need for animals in the training population to also be included in prediction pipelines, and the systematic calibration of marker effects as more information was added from the population.

The impact of these effects led AAA to work towards methods that allowed for real time inclusion of genomic information. Several groups had been researching ways to utilize marker information to fit a genomic relationship matrix transitioning genetic evaluations to using SNP genotypes to estimate genomic relationships rather than molecular breeding values (Misztal *et al.*, 2020).

Single-step genomic BLUP (ssGBLUP) had the potential to greatly simplify Angus's genetic evaluation. The method did not rely on deregressed proofs and properly weighted information from genotyped sires and dams. This avoided double counting issues which previously riddled Angus's evaluation as well as properly accounting for preselection bias of genomically tested parents (Legarra *et al.*, 2014). The method aimed to combine pedigree and genomic relationships into a combined genomic relationship matrix (GRM). However, the method relied on the inversion of the GRM and the computational cost of this direct inversion limited its usefulness in large routine evaluations (Aguilar *et al.*, 2011). Then in 2014, Misztal *et al.* presented the algorithm of proven and young (APY). By using approximate inversions of genomic relationships based on the recursions of a fraction of the total population, ssGBLUP was able to be effectively implemented for routine evaluation.

Current evaluation. On July 7, 2017, AAA deployed an updated evaluation. Employing single-step methodology was monumental for Angus and the entire US beef industry. Genetic trends and predictive ability of EPD were analysed using methods employed in 2009 compared to ssGBLUP models. The new ssGBLUP model on average increased the ability to predict future progeny performance for growth traits to 36% compared to previous models at 24%. Genetic trends from ssGBLUP models more closely mimicked genetic trends from traditional models supporting previous multistep trends were overestimated (Lourenco, 2018). At that time, ~300,000 genotypes had been submitted on cattle entering the genetic evaluation. Currently >1.09 million animals have genotypes. Figure 1 outlines the trend in genotype collection since inclusion into the genetic evaluation.

This growing number of genotypes creates alternate challenges. In 2017, a group of high accuracy animals were utilized as core animals for APY (Lourenco, 2018). Overtime this group became less representative of the genotyped population as discussed by Garcia *et al.* (personal communication) in a paper submitted to this congress. Slower convergence and increased run times resulted, making it cumbersome to execute a

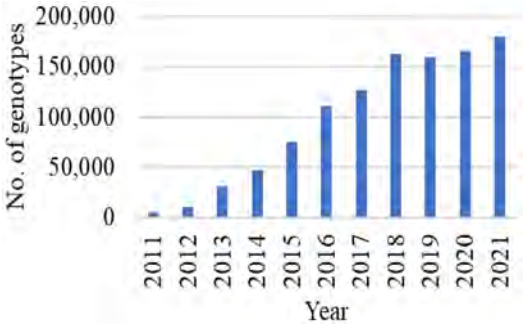


Figure 1. Trend in genotype collection on animals available for Angus's weekly genetic evaluation.

weekly genetic evaluation. The American Angus Association took the step in Summer 2021 to update this core group leading to decreased run times and subtle genetic prediction movements.

Breeders are realizing the promise of genomic testing. The impact of GS can be analysed through standardized genetic progress. Figure 2 illustrates standardized genetic progress prior (2001-2009) to and presently (2010-2021) influenced by genomics for EPD traits with high economic relevance in the US beef industry. Genetic progress for growth and carcass traits have accelerated in favorable directions with the inclusion of genomics with the exception of maternal milk and marbling. The downward trend in backfat is favorable as it would have positive outcomes for carcass yield. The pattern in maternal milk could be explained by breeder selection pressure working to hold a female's maintenance energy requirement to an optimal level. Marbling EPD has not accelerate with genomics. This could be explained by several high genetic bulls for marbling being culled from the Angus population due to lethal genetic defects in 2008.

In addition to accelerated progress, increased EPD accuracy allows users of Angus genetics to purchase bulls with less risk. This risk can be assessed in terms of the number of progeny equivalents added to an individual when genotyped. In Angus, genotyping a non-parent animal is equivalent to an animal producing, on average, 20 progenies, depending on the trait (American Angus Association, 2018). This

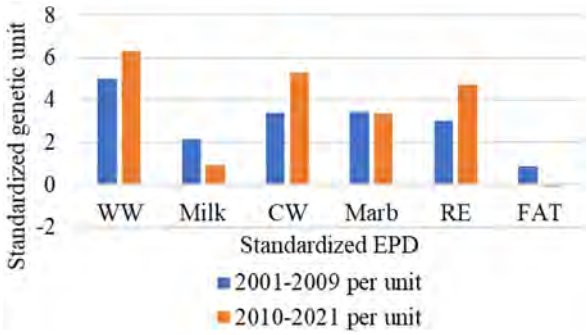


Figure 2. Standardized genetic progress made prior (2001-2009) to and presently (2010-2021) influenced by genomics for EPD traits with high economic relevance in the US beef industry including weaning weight (WW), maternal milk (Milk), carcass weight (CW), marbling score (Marb), ribeye area (RE) and backfat (BF).

accuracy increase is approximately equivalent to the amount of progeny information a sire could manifest with his first calf crop. This increased accuracy has driven the demand for genomic-tested Angus bulls.

Discussion

The future of GS is bright. Over one-half of yearly Angus registrations are now being genotyped. Price reductions have made widespread testing affordable. Sophisticated methodology has improved prediction accuracy and reliability of genetic predictions increasing trust by the user of Angus genetics. Genetic progress has amplified creating favourable marketing conditions for seedstock producers. Based on the number of genotypes submissions, many 'performance orientated' breeders have made genomic testing a standard operating procedure. The demand of genomic-tested bulls has been strong with bull stud companies and bull buyers not only rewarding breeders who test but demanding it.

One area GS has not found a firm foothold is in the US commercial cattle sector. Tests are available to the commercial cattle industry; however, the return on investment has yet to be large enough to pique widespread interest. Unique programs to provide genetic descriptions on groups of calves entering the feeding phase have opened the door to marketing opportunities with genetic predictors. These efforts have allowed for a window of transparency to GS to another sector in the beef supply chain. If the fed cattle industry can utilize GS to better manage their feed yards, increasing production efficiency, commercial cattlemen may be rewarded for providing that information. Continued education on the value of GS along with integrating GS in ways that are accessible, cost effective, and useful to make management decisions will continue to increase awareness and uptake of GS in the future.

References

Aguilar, I., Misztal, I., Legarra, A., and Tsuruta, S. (2011) J. Anim Breed. Genet. 128:422-428. http://doi.org/10.1111/j.1439-0388.2010.00912x

American Angus Association (AAA). American Angus Association Genomic Enhanced EPDs. Available at: https://www.angus.org/university/docs/GenomicEnhancedEPDs.pdf

Kachman, S. D. (2008) 9th Proc. Genet. Pred. Workshop Kansas City, USA

Legarra, A., Chistensen, O. F., Aguilar, I., and Misztal, I. (2014) Livest. Prod. Sci. 166:54-65. https://doi.org/10.1534/genetics.115.177014

Lourenco, D. A. L., Tsuruta, S. Fragomeni, B. O., Masuda, Y., Aguilar, I., *et al.* (2018) 11th Proc. World. Cong. Appl. Livest. Prod., Auckland, New Zealand.

Misztal, I., Legarra, A. and Aguilar, I. (2014) J. Dairy Sci. 97:3942-3952. https://doi.org/10.3168/jds.2013-7752

Misztal, I., Lourenco, D. A. L., and Legarra, A. (2020) J. Anim. Sci. 98:1-14 https://doi.org/10.1093/jas/skaa101

432. Efficient genetic progress for quantitative traits through genomic selection

S. Tsuruta*, D.A.L. Lourenco and I. Misztal

University of Georgia, 425 River Rd, Athens, GA 30602, USA; shogo@uga.edu

Abstract

The objective of this study was to investigate the genetic progress through genomic selection for a quantitative trait over generations, assuming different numbers of QTL and genotyped animals. Simulations were conducted for populations with an effective size of 40 with 20 overlapping generations and phenotypes for a trait with a heritability of 0.5. Genotypes simulated for the last 10 generations comprised 30K SNPs across 30 autosomes including 30 to 3,000 QTL from 200 to 30,000 genotyped animals. Genetic trends were compared when selection was based exclusively on either genomic predictions or traditional genetic predictions. When the population size was small, the genetic gain by genomic selection was smaller than that by genetic selection. The genetic progress through genomic selection was more efficient when the population size was large enough to cover the genomic information in genotypes, especially with more QTL, implying that the trait behaved as polygenic.

Introduction

Genomic selection has been conducted over the past decade to accelerate genetic improvement in various species as genotyping costs dramatically decline. However, field data are often insufficient to quantify the long-term effect of genomic selection for different genetic architectures. Simulations with genotypes can be used to investigate accuracy and bias in genomic predictions (Lourenco, *et al.*, 2017). Additionally, simulations can help identify in which scenarios genomic selection would maximize the genetic progress in breeding population. Genetic trends based on genomic predictions via single-step genomic (ssG)BLUP (Aguilar *et al.*, 2010) and genetic predictions via traditional pedigree-based (A)BLUP start diverging when genomic selection becomes effective (Abdollahi-Arpanahi *et al.*, 2021). This is because ABLUP is biased under genomic selection as this method does not account for genomic information. However, depending on the number of genotyped individuals, how much the selection process relies on genomic information, and the trait architecture, genetic trends based on ABLUP and ssGBLUP may be more or less divergent. The objectives of this study were to understand the nature of genomic selection for quantitative traits and to investigate more efficient genetic progress over generations by assuming various numbers of QTL and genotyped animals.

Materials & methods

For populations with an effective size of 40, 20 overlapping generations with phenotypes for 400, 2,000, 20,000, and 60,000 animals were generated for a trait with a heritability of 0.5 and an additive genetic variance of 50. Genotypes simulated for the last 10 generations comprised 30K SNPs across 30 autosomes including 30, 300, and 3,000 QTL (i.e. 1 to 100 QTL per chromosome) for 200, 1000, 10,000, and 30,000 genotyped animals. In this simulation, it is assumed that QTL explained all the genetic variation for the trait (i.e. no extra polygenic effects). Selection was conducted by mating the top 10% males and 50% females of evaluated animals for 20 generations. We estimated and compared genetic trends when selection was based exclusively on either genomic predictions (ssGBLUP), genetic predictions (ABLUP), or animal predictions without considering relationships among animals (IBLUP). The QMSim software (Sargolzaei and Schenkel, 2009) was used for simulations, running the BLUF90 program (Misztal *et al.*, 2014) inside the QMSim to calculate the predictions and to select animals in each generation over 20 generations. Using all

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_432, © S. Tsuruta *et al.* 2022

phenotypes, genotypes, and pedigrees for all generations, genetic trends were estimated and standardized by the additive genetic standard deviation.

Results

Population size. The genetic gain by genomic selection via ssGBLUP increased as the number of genotyped animals expanded from 200 to 30,000 in the populations with 3,000 QTL (Table 1 and Figure 1a). The realized accuracy calculated by the genetic gain per generation significantly increased as the number of animals increased (Table 1 and Figure 1b), indicating that additional genomic information contributed to the accuracy more than pedigree information in a larger population as shown in the average diagonal of $(G^{-1} - A_{22}^{-1})$ in Table 1.

Number of QTL. Figure 2 shows genetic trend (a) and realized accuracy (b) via ssGBLUP from 30 to 3,000 QTL and via ABLUP and IBLUP in a population including 30,000 genotyped animals. With fewer QTL, the genetic gain by the genomic selection was smaller than that by the traditional genetic selection via ABLUP (Table 1 and Figure 2a). As the number of QTL increased, the genetic gain became larger. However, the gentic trend was lower when the number of QTL was smaller (e.g. 30 QTL = 1 QTL/chromosome) regardless of the population size. The accuracy in ssGBLUP was also lower when the number of QTL was smaller, compared with the accuracy in ABLUP. When the number of genotyped animals was large (i.e. 10,000 and 30,000), the realized accuracies in ssGBLUP were higher than those in ABLUP, but the genetic gains for ssGBLUP were smaller when the number of QTL was 300. As shown in Figure 2a, this contradiction was due to the slow genetic progress. When the number of QTL was smaller (i.e. 30), the genetic progress slowed down more. As predicted, this tendency was more pronounced when the population was smaller than 10,000 (result not shown here).

Table 1. Genetic gains estimated for difference scenarios in the simulation.

No. genotyped animals	No. QTL	Genetic gain / generation	ssGBLUP[1] / ABLUP[2]	Realized accuracy	Diagonal $(G^{-1} - A_{22}^{-1})$
200	ABLUP	0.13	-	0.119	-
	30	0.07	0.57	0.106	0.27
	300	0.20	1.57	0.146	0.31
	3,000	0.18	1.37	0.164	0.33
1000	ABLUP	0.29	-	0.235	-
	30	0.09	0.30	0.181	0.78
	300	0.33	1.15	0.272	0.77
	3,000	0.35	1.21	0.283	0.73
10,000	ABLUP	0.52	-	0.365	-
	30	0.09	0.18	0.200	5.34
	300	0.43	0.82	0.405	5.64
	3,000	0.61	1.16	0.476	5.67
30,000	ABLUP	0.64	-	0.437	-
	30	0.13	0.20	0.249	19.51
	300	0.49	0.76	0.467	21.22
	3,000	0.73	1.14	0.553	21.44

[1] Single-step genomic BLUP.

[2] Pedigree-based BLUP.

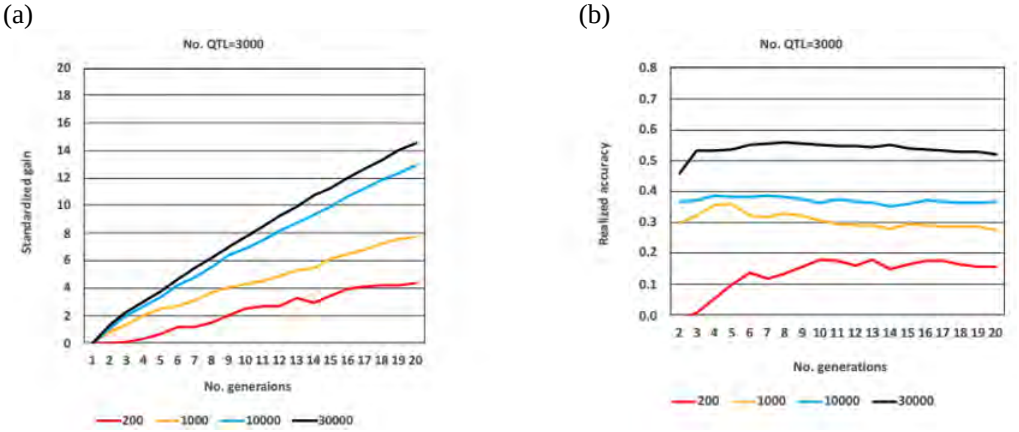


Figure 1. Genetic trend (a) and realized accuracy (b) for numbers of genotyped animals from 200 to 30,000 with 3,000 QTL.

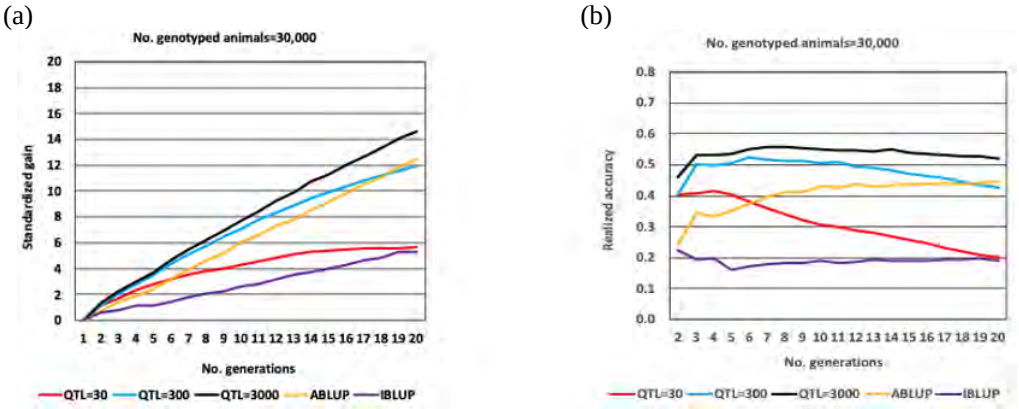


Figure 2. Genetic trend (a) and realized accuracy (b) via ssGBLUP with numbers of QTL from 30 to 3,000 and 30,000 genotyped animals, ABLUP, and IBLUP.

The number of eigenvalues that can explain 95, 98, or 99% in the genomic relationship matrix (G) can be used as a proxy for the number of independent chromosome segments segregating in the population (Pocrnic *et al.*, 2016). Figure 3 shows that for example, in a population of 30,000 genotyped animals with 30K SNPs, 15,000 genotyped animals is required to explain 98% of the variance in G. In a population with the effective population size of 40, when the number of genotyped animals is large enough (>30,000 genotyped animals), the maximum number of independent chromosome segments can be 20,000. Compared with this maximum number, when the population is small (10,000 or fewer genotyped animals), 30K SNPs can explain only less than half of the genomic variation in G. This could be one reason why higher accuracy and larger genetic gain were not achieved in a small population.

Figure 4 indicates that the additive genetic variance (V(g)) estimated with a small number of genotyped animals is largely overestimated, especially when the number of QTL is larger (i.e. the trait is more polygenic). This overestimation of variance components could result in inaccurate genomic predictions in practice.

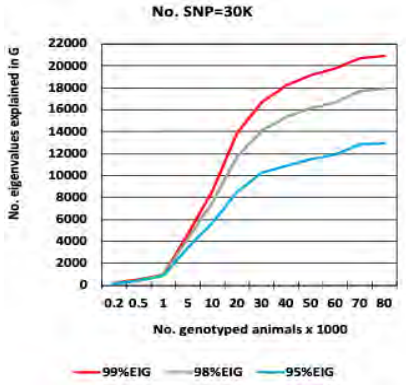


Figure 3. Numbers of eigenvalues explaining 95, 98, and 99% of the variance in the genomic relationship matrix.

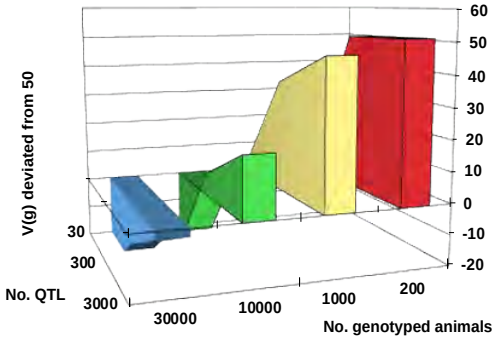


Figure 4. Additive genetic variance (V(g)) estimated with different numbers of QTL and genotyped animals as a deviation from the true V(g) = 50.

Discussion

Genetic trends based on ABLUP and ssGBLUP may diverge depending on the number of genotyped individuals, how much the selection process relies on genomic information, and the trait architecture. Suboptimal gains can be observed if the number of genotyped animals is fewer than that of independent chromosome segments segregating in the population. Therefore, small genotyped populations reap less benefits from genomics. Maximizing the rate of genetic gain under genomic selection would require a large number of genotyped animals that explain sufficient genomic variation, especially for a polygenic trait under directional selection. A trait with fewer QTL or causative SNP (less polygenic) tends toward fixation more rapidly, resulting in low genetic trend. Genetic trend for species with small genomes likely reaches the plateau quickly and then deteriorates due to inbreeding. It is also expected that genomic selection for lowly heritable traits is more effective when the trait is more polygenic and a large genotyped population is available. A long-term genomic selection study on genetic trends with field data should be conducted with caution.

References

Abdollahi-Arpanahi R., Lourenco D., Legarra A., Misztal I. (2021). Genet Sel Evol. 53:89. https://doi.org/10.1186/s12711-021-00683-6

Aguilar, I., Misztal I., Johnson D.L., Legarra A., Tsuruta S. *et al.* (2010) J. Dairy Sci. 93:743:752. https://doi.org/10.3168/jds.2009-2730

Lourenco D.A.L., Fragomeni B.O., Bradford H.L., Menezes I. R., Ferraz J.B., *et al.* (2017) J. Anim. Breed. Genet. 134:463-471. https://doi.org/10.1111/jbg.12288

Misztal I., Tsuruta S., Lourenco D., Aguilar I., Legarra A. *et al.* Manual for BLUPF90 family of programs – Available at: http://nce.ads.uga.edu/wiki/lib/exe/fetch.php?media=blupf90_all7.pdf

Pocrnic, I., Lourenco, D.A.L., Masuda Y., Legarra A., and Misztal I. (2016). Genetics 203:573-581. https://doi.org/10.1534/genetics.116.187013

Sargolzaei M. and Schenkel F.S. (2009) Bioinformatics. 25(5):680-681. https://doi.org/10.1093/bioinformatics/btp045

Part 24

Breeding programs

433. Strategies to support phenotyping for genomic selection R&D and implementation in beef and sheep industries

R.G. Banks

AGBU, a joint venture of NSW Department of Primary Industries and University of New England, Armidale, NSW, 2350 Australia; rbanks@une.edu.au

Absract

Phenotypic recording is central to all genetic improvement. In the pre-genomics era, such recording tended to be focussed on selection candidates, particularly in beef cattle and sheep. In the genomics era, recording and selection can be completely uncoupled. This in turn makes explicit the importance of the key decisions: what to breed for, which leads to what to record, which leads to how to invest in recording (what traits, when recorded, how many records per trait, and which individuals). In single integrated organisations these decisions are relatively straightforward, but in multi-owner or –stakeholder situations, new approaches and levels of investment planning and coordination will be required for optimal sustained genetic progress. Approaches for such planning and coordination have been used (e.g. simulation modelling) are under development (for business aspects of the problem), but organisations and industries will need new models and levels of management for success.

Introduction

A basic algorithm for genetic improvement has been described (Harris and Newman, 1994) and its successful implementation widely documented (e.g. Hill, 2010). The algorithm can be summarised as: (1) definition of the breeding objective; (2) recording appropriate traits; and (3) prediction of genetic merit, leading to iv) selection and mating. When carefully applied, rates of improvement in productivity of 3 to 5% are feasible and simultaneous improvement in productivity and product quality attributes can be achieved (MLA, 2021). Appropriate data collection, or phenotyping is critical.

Since the 1980s, the prediction step in most animal species has been via pedigree-BLUP procedures, enabling use of data that is often variable in terms of what traits are recorded – with recording predominantly for traits that are relatively easily recorded on-farm. This is particularly the case for extensive industries such as beef cattle and sheep, which in most countries are characterised by low levels of organisation or coordination in the sense of not having defined nucleus populations or systematic collection of off-farm data.

More recently, genomic selection (Goddard *et al.*, 2009) has become more widely used, with a key feature being the ability to estimate genetic merit in a population for which a genomic reference population exists. The value of the reference population in terms of accuracy of prediction of genetic merit is broadly understood and underpins an essential feature and opportunity of genomic selection, i.e. the decoupling of phenotype recording from selection. Under pedigree-BLUP, most recording has been conducted by individuals (or enterprises) who then make selection and subsequent marketing decisions on the recorded animals. Under genomic selection, provided a suitable reference population exists, breeding enterprises can operate effectively without conducting any recording themselves. Another way of thinking about this is that it implies scope for explicit investment decisions on what to record, and how much to invest in recording, recognising that accuracy generally increases with level of investment, but with declining marginal return on investment (Goddard *et al.*, 2009). This paper considers some implications of this feature and opportunity, with particular focus on the beef and sheep industries of Australia.

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_433, © R.G. Banks 2022

Implementation of genomic selection for beef and sheep in Australia

The beef and sheep industries in Australia (and elsewhere, to varying degrees) are characterised in genetic terms by wide variation in: enterprise size in the breeding sector (MLA, 2019a); phenotype recording effort (e.g. Johnston, 2007); selection effort (expressed as i/L, selection intensity relative to generation length); and the role and effectiveness of breed associations. Within this context, there have been solid achievements, including development of reference populations through co-investment models; introduction of single-step genomic evaluation; expansion of traits under evaluation; and accelerating growth of genotyping for genomic prediction (MLA, 2021). This early phase implementation provides a basis for confidence in the technology and for continued and potentially increased investment. How best to invest is the challenge.

Challenges to investment in genomic selection

The starting point for considering investment for genetic improvement, at any level of organization, should be to define the breeding objective. Under genomic selection, this need not be constrained either by what is currently recorded or what is currently paid for by product markets or value chains. In this sense, genomic selection shifts genetic improvement towards being a design enterprise and, thus, managers of improvement programs should think broadly about both current and future attributes of their product (e.g. social license traits, see Miller and Retallick, *these proceedings*). Decision-making on breeding objectives needs to take account of all stakeholders in current value chains, including consumers, and to carefully assess what might seems likely to become important in the medium-term. If traits are prioritised that are not currently important, how to fund the data collection and how to report genetic merit for recorded traits in ways that do not distort current enterprise finances, become important challenges.

The practical and financial imperative of this 'design perspective' is that it must be backed by investment in recording the appropriate traits in the reference population. Given that in highly dispersed (in the ownership sense) industries, breeders only capture a small share of total income generated from progeny of bulls or rams (typically 5 to 10% in beef and sheep in Australia), there is a risk of insufficient funds for recording at the required scale. Two broad responses seem possible: firstly, leave this to the market and hope that at least some breeders have sufficient scale to make more speculative investments; or actively invest at a breed or industry level, using whatever funds can be collected, including industry or government R&D funds and, potentially, royalties imposed on genomic testing. Under the more active strategy, two challenges remain: the need for investment in recording is ongoing, in that reference populations need to be 'topped up' with new data so that prediction accuracy is maintained (Dekkers *et al*., 2021); and secondly, some form of coordination of funding becomes essential.

Such coordination is outside the traditional domains of both R&D funding agencies and breed associations, at least in Australia, although both groups have moved in this direction in the last decade (through co-investment in designed recording projects). However, it is not clear to what extent the required permanency of the funding is appreciated. Indeed, the fact that data collection for genomic reference populations is essentially continuing or real-time R&D (with new knowledge being generated continuously) challenges existing definitions of, and funding approaches for, R&D.

The rationale for role of industry or government agencies in funding the R&D aspects of genetic improvement has been considered previously (e.g. Smith, 1981, Elsen *et al*., 1986), but no obvious long-term system or approach has been implemented anywhere, although sustained sequential R&D investment has occurred in several countries (e.g. US (USDA), France, Ireland, and Australia).

Investment questions – what to invest in, how to fund it, and how to make decisions

The challenge in the genomics era for all organisations (public or private) can be distilled to two questions: (1) what to breed for; and (2) how to fund the data collection that will enable high rates of progress towards that goal? The inherent goal is to maximise return on investment. Defining a breeding objective is in principle straightforward and tools are available to assist the process (Barwick *et al*., 2005). For multi-member organisations, or for whole industries, the challenge of obtaining 'buy-in' is real and is likely facilitated by widespread extension and consultation, particularly around traits that are currently not recognised in market prices but which are expected to enhance value in the future. This is somewhat analogous to the challenge of developing and achieving acceptance of multiple goals within a breed, which is technically straightforward but can be very difficult for such organisations (P. Parnell, *pers. comm.*)

Defining a breeding objective directly defines the data collection strategy, which should aim at maximising prediction accuracy for breeding objective traits (Van der Werf *et al.*, 2018). In broad terms, this means recording as many animals as can be afforded (as an approximation, the number to record should be proportional to the size of the commercial or production population, M. Goddard, *pers. comm*.), although an attractive feature of genomic selection is that recording can be concentrated and optimised, rather than depending on 'partial' recording of large numbers of animals across multiple enterprises across the population. Particularly if the objective includes hard-to-measure and/or social license traits, it is likely that these: (1) will not currently be recorded; (2) require expensive equipment; and (3) require consciously planned recording projects. This implies access to funds and sound management of those funds.

To a large extent, funds for such recording have to date been obtained via R&D grants and it may be that this avenue will always be available but this cannot be guaranteed, and mechanisms to collect funds from other avenues will likely be valuable. An efficient and equitable mechanism for sharing recording costs across recorders and users was outlined by Banks (2021). This approach can incorporate funding from diverse sources, such that if industry and (or) government is willing to assist with costs of hard-to-measure and (or) as yet unpriced traits.

At an industry level, investment in recording phenotypes should incorporate breed comparisons where possible, both to facilitate across-breed genomic selection and reduce individual breed recording requirements, but also to inform sire-buying markets, thereby increasing focus on objective merit and likely stimulating faster genetic progress within breeds.

The more rigorously investment thinking is applied to these challenges, the more likely specialised recording operations will become. In this, beef and sheep breeding will become more like the more intensive livestock industries, where ownership and recording is more concentrated. This will bring challenges for the roles of individual enterprises, in particular for 'seedstock breeders', and for how data access is funded. It is plausible but not necessarily optimal that recording and selection would become completely uncoupled business strategies.

Conclusions – key features of possible solutions

Optimally designed and scaled phenotyping is the fundamental issue for all genetic improvement programs into the future and holds special challenges for extensive multi-agent industries such as the beef cattle and sheep industries. Sustained genetic improvement utilising genomic selection seems certain to include the following features:

- definition of breeding objectives that draw on multiple stakeholders' perspectives;
- specialist phenotype recording operations that complement activities of individual breeders and that are optimised at the breed or industry level for scale and traits included;
- mechanisms for harvesting appropriate shares of the value created to maintain funding for optimal recording and that, accordingly, explicitly value and remunerate recording efforts;
- and sound investment management for organisations, including sustained co-investment partnerships with industry and/or government.

Genomic selection increases scope for extensive industries to design and achieve more valuable genetic improvement than hitherto possible, but this will require new models for multi-member organisations and levels of coordination. Ongoing R&D into optimal breeding program design and, in particular, into methods for coordinating investment in phenotypes, will be valuable.

References

Banks, R.G. (2018). Proc. 11th World Congress on Genetics Applied to Livestock Production.2: 758

Banks, R.G. (2021) Proc. 44th ICAR Conference https://www.icar.org/wp-content/uploads/2021/10/ICAR-Technical-Series-25-virtual-meeting-2021-Proceedings.pdf

Barwick, S.A. and Henzell, A.L. (2005) Austr. J. Exp. Agric. 45:923-933

Dekkers, J.C.M, Su, H., and Cheng, J. (2021) Genet. Sel. Evol. 53:55

Elsen, J-M., Bibe, B., Landais, E., and Ricordeau, G. (1986) http://www.wcgalp.org/system/files/proceedings/1986/twenty-remarks-economic-evaluation-selection-goals.pdf

Goddard. M.E. and Hayes, B.J. (2009) Nature Reviews Genetics 10:381-391

Harris, D.L, S. Newman, (1994). J Anim. Sci. 72:2178-2200

Hill, W.G. (2010) Phil. Trans. R. Soc. B 365: 73-85

Johnston, D. J. (2007). Proc. Assoc. Adv. Anim. Breed. Genet. 17:8-15

Meat and Livestock Australia (2019a) L.GEN.1802 https://www.mla.com.au/research-and-development/reports/2019/benefit-and-cost-of-performance-recording-in-the-beef-and-sheep-studs/

Meat and Livestock Australia (2019b) L.GEN.1804 Workshop Report

Meat and Livestock Austalia (2021) L.GEN.1704 Final Report

Smith, C. (1981) Livestock Prod. Sci. 8:193-201

Van der Werf, J.H.J., R.G. Banks and J.C.M Dekkers. 2018. Proc. 11th World Congress on Genetics Applied to Livestock Production.11:889.

434. The infrastructure for cattle data at the Swedish university of agricultural sciences, Gigacow

T. Klingström*, I. Ohlsson[1] and D.J. de Koning

Department of Animal Breeding and Genetics, Swedish University of Agricultural Sciences, Box 7023, 750 07 Uppsala, Sweden; tomas.klingstrom@slu.se

Abstract

Research in animal breeding and precision livestock farming both require vast volumes of highly heterogeneous data, often beyond the capabilities of the universities' own research farms. To facilitate access to detailed phenotypes from a network of farms, we established the infrastructure for cattle data at the Swedish University of Agricultural Sciences (SLU Gigacow). A main achievement of SLU Gigacow has been the development and implementation of automated daily collection of records from each participating farm's herd management system to a centralized server at SLU. SLU Gigacow will monitor around 5,000 cows (including annual replacements) when it is fully operational. Although still under development, SLU Gigacow has already received much attention from various stakeholders, as well as media. In the long-term we expect SLU Gigacow to become a valuable infrastructure for discoveries, as well as a testbed for implementation of new technologies.

Introduction

While a university can do much research in their own dedicated research farms, many studies, especially those related to genetics and genomics, require much larger sample sizes and access to field data. It is very demanding and highly inefficient for individual researchers to recruit farms that are willing to contribute data for each individual research project. To address this challenge, the infrastructure for dairy cattle data at the Swedish University of Agricultural Sciences, Gigacow (SLU Gigacow), was founded in 2018. Its main purpose is to support research projects in dairy cattle genetics, precision agriculture, and other research concerning dairy farms and the environment. In a modern dairy, farm machines and sensors automatically collect data on milk composition, activity, and feeding behaviour of cows to ensure good animal health and high productivity. Genotyping of calves for estimation of genomic breeding values is now also routine procedure at many farms and genotyping with a 45K SNP-chip is available as a commercial service at a low cost. At a farm, the pedigree of each cow is known and the herd structure ensures that between 60 and several hundred animals are exposed to the same environment.

The purpose of SLU Gigacow is to establish a single data collection platform useful to researchers at SLU and elsewhere to accelerate the scientific process. Once an infrastructure has been established, large amounts of data can be collected at a low cost. The cost of establishing a data collection infrastructure with legal agreements and software solutions for automated data collection and data curation is however a major investment. The limitations of funding and a lack of dedicated infrastructure have been identified as the major bottlenecks in phenome research in non-model organisms, as projects are either narrowly focused on a single biological pathway in response to phenotypic change, while ignoring the rest of the dataset, or limited to broad sweeping conclusions about the dataset based on shallow interpretations (Burnett *et al.*, 2020). By providing a rich dataset of phenotypes and associated data, researchers can better plan the scope of projects and engage in exploratory projects with minimal overhead and startup costs. Thus, SLU Gigacow can provide the mechanism necessary to encourage junior and senior investigators to conduct focused, hypothesis-driven research framed in a genome to phenome context and to critically consider experimental design, power, and sample sizes when launching '-omics'-based experiments.

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_434, © T. Klingström *et al.* 2022

Materials & methods

Production data from farms is collected by the software Gigacow Kollector, written in Python 3 (Beaverton, USA), and automates export of data from herd management systems such as DeLaval DelPro (Tumba, Sweden) and Lely T4C (Maassluis, the Netherlands). Data in certain reports on the DeLaval DelPro can be scheduled for automatic exports and other sensor manufacturers also provide similar solutions, enabling data from the herd management system and other equipment to be automatically exported to the temporary file storage area created by Gigacow Kollector. Each day, Gigacow Kollector runs a scheduled task to upload data with Secure File Transfer Protocol (SFTP) to a file storage server located at SLU, which stores the raw files. A verification script run on the server side ensures that all expected reports are uploaded. Data from the SLU Gigacow data collection server is then imported to the Gigacow SQL Server database (DB) by means of multiple processes for extraction, transformation and loading that are maintained with SQL Server Integration Services (SSIS) (Microsoft, Redmond, USA). Researchers can download data directly from the DB via a database data connection or, to facilitate analyses, by connecting the tool of their choice, such as Excel, to a regularly preprocessed and aggregated Online Analytical Processing cube deployed with SQL Server Analysis Services (SSAS).

In addition to data collected on farms, SLU Gigacow also collects data through an agreement with Växa Sverige AB (Uppsala, Sweden), which is a service that aggregates information on subjects such as claw health, veterinary visits, and other key performance indicators related to animal health and productivity. This data is compiled quarterly in fixed length files and sent to the SLU Gigacow server using SFTP. Genome analysis of animals is performed by the service Växa Genomanalys using the 45k EuroG MD beadchip (Amsterdam, the Netherlands). Calves born in herds that participate in SLU Gigacow are genotyped as a part of the tagging process, in which a sample is collected by the farmer and submitted to the genome analysis service. Adult animals are sampled by the farmers or visiting technicians from Växa by collecting blood or hair samples from the tail. Ethics approval from the authorities is required and has been granted for collection of blood samples from adult animals but is not required for hair sample collection.

The data collection system of SLU Gigacow is under continuous development but, in its current form, presents a modular solution to quickly integrate information from many different sources (Figure 1).

Results

Data collection from dairy farms was delayed by the Covid-19 pandemic but was started by remote installations in 2021. In a steady state situation, approximately 5,000 animals on 15 farms will be monitored on an annual basis, with a budgeted annual replacement rate of 35%. So far 1300 animals have been genotyped in collaboration with Växa and the farmers. Using Växa Sverige's commercial service for genome

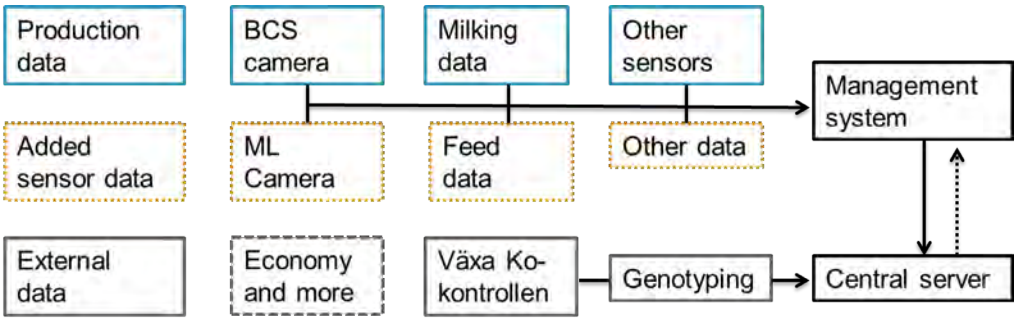


Figure 1. Data collection in Gigacow. Dotted boxes are currently under development.

analysis, the cost per genotyped calf is ~25 Euro at an exchange rate of 0.1 SEK per Euro. Genotyping of adult animals is done by arrangement and cost between 32 and 36 Euro, depending on the size of the farm, who performs the sampling, and labour required to collect animals.

A number of key performance indicators have been tracked during the establishment of SLU Gigacow (Table 1). The SLU Gigacow infrastructure was established not only to support genomics research but also to serve as an interdisciplinary infrastructure that serves all fields related to dairy farming. Education of agronomists specialising in agricultural technology was abolished in Sweden ~20 years ago and SLU is currently working to increase its focus on agricultural technology, bioinformatics, and precision farming. Establishment of SLU Gigacow has created a natural contact point at SLU for stakeholders in the industry and at other universities to bring technology developers and the farming industry closer together. SLU Gigacow has therefore been used as an example in multiple articles and conferences on smart farming and data collection. SLU Gigacow has also participated in numerous grant applications with partners at Linköping University, Uppsala University, Dalarnas University, the University of Oslo, and several nodes in the Elixir European life science infrastructure for biological information (ELIXIR). The total approval rate for grant applications supported by SLU Gigacow has so far been 42%, compared to an average funding rate of 14% for the Swedish Research Council for Sustainable Development (FORMAS) and 20% for grant applications in Natural science and technology by the Swedish Research Council (VR).

Technology development and testing is also an important topic, as the implementation of smart solutions is a key area in the digital transformation of the agricultural sector (OECD, 2019). Gigacow is developed to function as a testbed for innovation and is a member of the Nordic Testbed Network (https://nordictestbednetwork.se/gigacow-2/). Data from farms is continuously collected, facilitating both longitudinal (before and after implementation) and case-control studies, enabling users to 'test before invest'. Activities in this role include projects integrating Edge Artificial Intelligence into farming decision support systems, in collaboration with Uppsala University and a major equipment manufacturer, and assessment of market acceptance for remote sensing applications, together with a major producer of visible, near infrared, and short-wave infrared sensors. To formalise this role, SLU Gigacow also participates in the application for a European Digital Innovation Hub, in collaboration with Linköping University, the Research Institutes of Sweden, and multiple regional actors for technology and innovation.

Discussion

The major challenges during the development of SLU Gigacow have not been how to collect data but to identify what data there is to collect, what agreements are necessary with various stakeholders to collect the data, and finding a robust technical infrastructure that enables the collection of data from heterogeneous

Table 1. Key performance metrics tracked in SLU Gigacow.

Indicator	Number
Animals genotyped	1,294
Articles and presentations of Gigacow	19
Projects supported with data collection	3
Projects supported with expertise	5
Supported applications	20
Supported applications funded	8
Supported applications not funded	11
Supported applications pending	1

data sources and present it in an accessible manner. There is fierce competition in the agricultural sector, both between equipment manufacturers and service providers, which makes protection of international property rights a key concern for these commercial companies. At the same time, partnering with the industry can yield considerable benefits of scale, as well as new research opportunities. The estimated cost of genotyping at the National Genomics Infrastructure would, despite an academic subsidy have amounted to ~52 Euro per animal, even before adding costs for sample collection and DNA extraction, which would have had to be performed in-house at SLU. There is also a strong interest, both among equipment manufacturers and service providers, to further increase the precision and scope of data collection in farming. Close collaboration between researchersm farmers and other stakeholders in primary production can thereby both yield new technology for the study of animals and insights that motivate initiation of new research.

Conclusions

Digital transformation is rapidly changing all sectors of society and many sectors face a lack of computer scientists and bioinformaticians who can process and analyse biological data (Batut *et al.*, 2018). Data collection and analysis is, therefore, an increasingly large part of the workload for many researchers (Klingström, 2017). Automation and collaboration with industry is, therefore, important to reduce the workload of researchers by making sure that data is readily available in a format that is accessible to both domain experts and technology developers who are aiming to bring technological innovations into new domains. Thus, the establishment of livestock data infrastructures can accelerate genetics research by genotyping and phenotyping animals at a scale that is not feasible outside the agricultural sector.

Acknowledgements

SLU Gigacow is financed by an infrastructure grant from SLU and acknowledges in-kind support from VÄXA Sverige. We gratefully acknowledge the Swedish farmers who participate in Gigacow without financial compensation.

References

Batut B., Hiltemann S., Bagnacani A., Baker D., Bhardwaj et. al (2018). Cell Systems 6, 752-758.e1. https://doi.org/10.1016/j.cels.2018.05.012

Burnett K.G., Durica D.S., Mykles D.L., Stillman J.H., Schmidt C., (2020). Integrative and Comparative Biology 60, 397–401. https://doi.org/10.1093/icb/icaa059

Klingström, T., (2017). Data integration and handling, Acta Universitatis Agriculturae Sueciae. 1652-6880; 2017:99 Sveriges lantbruksuniversitet., Uppsala.

OECD, (2019). Digital Opportunities for Better Agricultural Policies. OECD. https://doi.org/10.1787/571a0812-en

435. Optimizing sampling strategies for sib test schemes in a broiler breeder program

C.A. de Hollander[1*], V.P. Breen[1], J. Henshall,[1] F.B. Lopes[1] and M.P.L. Calus[2]

[1]Cobb-Vantress Inc., 4703 US Highway 412 E, Siloam Springs, AR 72761, USA; [2]Wageningen University & Research, P.O. Box 338, 6700 AH Wageningen, the Netherlands; charlie.dehollander@cobbvantress.com

Abstract

In broiler breeding there is a known genotype by environment interaction for bodyweight measured in a bio-secure (BW_B) or in a commercial environment (BW_C), thus a sib test scheme to record and use BW_C records is beneficial. This study estimated a genetic correlation of 0.59 between BW_B and BW_C. To optimize the sib scheme, different genotype proportions (25, 50, 75%) and sampling strategies were studied. No pedigree information was available for birds in the commercial environment unless they were genotyped. Genotyping based on phenotypes, heaviest and lightest, resulted in the highest accuracy compared to random genotyping 0.88 vs 0.80 at 25%, 0.94 vs 0.89 for 50 and 0.97 vs 0.95 at 75% genotyped. No bias was observed with random sampling, however bias was detected with sampling based on phenotypes (0.58 at 25%, 0.76 at 50 and 0.91 at 75%).

Introduction

Broiler poultry breeding companies keep and select their genetic pure lines in a bio-secure environment to reduce the risk of any diseases. However, this environment differs from the environment in which commercial broilers are usually reared. Because the selection environment does not completely reflect the commercial environment, a genotype by environment interaction (G×E) is likely present. The G×E causes the expressions of different phenotypes to be affected (Kolmodin and Bijma 2004) and thereby lead to reranking of families in different environments. A genetic correlation between body weight of the same purebred broiler line in a bio-secure and in a commercial environment was reported between 0.46-0.69 (Chu *et al.* 2019; Kapell *et al.* 2012). Therefore, it would be valuable for broiler breeder companies to set up a sib scheme test design to increase the genetic gain in body weight in a commercial environment. This study explored different genotyping strategies and proportions of birds to be genotyped to optimize a broiler breeding sib scheme set up. The two following strategies were chosen: the first is Random sampling (RND), this strategy is known to result in higher accuracies and lower bias compared to genotyping strategies based on mean phenotypes or best phenotypes (Howard *et al.* 2018; Jiménez-Montero *et al.* 2012). The second strategy is sampling based on Extreme phenotypes (EXT), the heaviest and lightest birds are chosen to be genotyped. This strategy has been found to result in the highest accuracies of breeding values compared to random sampling (Boligon *et al.* 2012; Howard *et al.* 2018; Jiménez-Montero *et al.* 2012). Additionally, three different proportions of genotyped birds were studied: 25, 50 and 75% of the available birds in the commercial environment.

Materials & methods

Phenotypes and genotypes. Data was provided by the poultry breeding company Cobb-Vantress, Inc. (Siloam Springs, Arkansas, USA). The data included performance information collected on pure-line broiler chickens that were either raised in a pedigree bio secure environment (Env B) or in a commercial environment (Env C). Once a bird is placed in Env C, they cannot return to Env B and will therefore be excluded from the breeding program. Of all birds hatched in one selection round, about 30% were placed in Env C, this was found to be sufficient by Chu *et al.* 2018 considering selection intensity, inbreeding and accuracy. These birds were all either full or half sibs of the other 70% birds that were placed in Env B. Data was collected over approximately 3.5 generations. The number of body weight phenotypes and the number

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_435, © C.A. de Hollander *et al.* 2022

of genotypes are represented in Table 1. Body weight was recorded in both environments at the same age. Birds included in this study were offspring from 519 Sires and 2,368 Dams. Genotypes were collected on all birds in Env C but only on a selected amount of birds in Env B (Table 1). All chicks placed in Env B have a known pedigree and the selection candidates are genotyped.

Sampling strategies for Env C. Having all chicks in Env C genotyped, allowed to retrospectively evaluate different sampling strategies. For each strategy, to reflect a scenario in which chicks placed in Env C are hatched in a commercial hatchery, animals that were not sampled in Env C were assumed to have no pedigree, and were therefore not included in the ssGBLUP analysis. The following sampling strategies are studied: Random sampling (RND): based on the proportion, X chicks were randomly sampled. Extreme phenotypes sampling (EXT), based on the proportion, 0.5X birds were sampled based on the highest body weight and 0.5X birds were sampled based on the lowest body weight.

Statistical analysis. A multivariate pedigree based model was used to estimate variance and covariance components. The multivariate model included six traits; BW_B, BW_C, White Meat Percentage (WMPct), Leg Quarter Percentage (LegQPct), Gain and Feed Efficiency (FE). BW_B, WMPct, LegQPct and FE and Gain were only recorded in Env B. BW_C was only recorded in Env C. All traits were corrected for gender and contemporary group. Additionally, for both BW_B and BW_C a random permanent maternal environmental effect was added to the model.

$$y_{BwB + BwC} = Xb + Za + Wc + e \quad (1)$$

$$y_{Gain + RFI + BRM\% + LegQ\%} = Xb + Za + e \quad (2)$$

X, Z and W are incidence matrices; where a, c and e are vectors for direct additive genetic, environmental maternal and residual effects, respectively. These random effects were assumed to be normally distributed $a \sim N(0,A\sigma_a^2)$, $c \sim N(0,I_d\sigma_c^2)$ and $e \sim N(0,I\sigma_e^2)$, where A is the pedigree relationship matrix and I_d the dam's identity matrix and I the identity matrix for all individual birds.

The variance components were estimated with the use of AIREML procedure in the DMUAI program from DMU (Madsen and Jensen 2013). Estimated variance components were used in single-step genomic BLUP (ssGBLUP) to estimate breeding values with the use of BGF90 (Misztal *et al.* 2002). The ssGBLUP models were identical to models described above (Misztal, *et al.* 2009). For computational reasons APY method (Misztal *et al.* 2014) was used to compute the inverse of the G matrix used in ssGBLUP. Core animals included parents and all selection candidates.

Validation. Validation was done using LR method (Legarra and Reverter 2018) to evaluate the differences in GEBV accuracies and bias. Full data contained all genotypes on all birds raised in Env C, the reduced data only contained a proportion of the genotyped birds in Env C based on each strategy. The number of genotypes in Env B remained the same across all strategies. Accuracies of different scenarios were measured by the correlation of GEBVs between the full data and reduced data. Dispersion bias was evaluated by the

Table 1. The number of phenotypes and genotypes used in this study.

	Phenotypes	Genotypes
BW_B	87,381	26,867
BW_C	34,863	34,816

slope of the regression of GEBVs from the full data on the GEBVs of the reduced data for the different strategies.

Results & discussion

Accuracy. Table 2 shows the accuracy of different ratios and different strategies of genotype sampling. The EXT clearly shows an advantage over RND sampling in terms of accuracy, especially at lower sampling rates (0.88 vs 0.80 at 25%, resp.). The magnitude of the advantage disappears as sampling proportion increases (0.94 vs 0.89 at 50% and 0.97 vs 0.95 at 75% of birds genotyped resp.). This increased accuracy of EXT over RND is in accordance with multiple other studies (Boligon *et al.* 2012; Howard *et al.* 2018; Jiménez-Montero *et al.* 2012).

Bias. Table 2 shows that bias is virtually absent for RND, while considerable for EXT, especially at a lower proportion (0.58 at 25% genotyped). When the sampling proportion increases the bias decreases for EXT but is still present even at 75% (0.76 at 50% and 0.91 at 75% genotyped). It is known that with an increase in the number of genotypes there is a decrease in bias (Daetwyler *et al.* 2012) and that genotyping randomly reduces bias compared to sampling based on phenotypes or family index (Boligon *et al.* 2012; Howard *et al.* 2018; Jiménez-Montero *et al.* 2012).

Table 2. Accuracies (r) and regression slopes (β) across strategies and proportions.

Strategy	Ratio	$r_{f\text{-}r}$	SE	$\beta_{f\text{-}r}$	SE
Random	25%	0.80	0.007	0.98	0.008
Extreme phenotypes	25%	0.88	0.005	0.58	0.003
Random	50%	0.89	0.005	0.98	0.005
Extreme phenotypes	50%	0.94	0.002	0.76	0.003
Random	75%	0.95	0.002	1.01	0.004
Extreme phenotypes	75%	0.97	0.001	0.91	0.002

Conclusions

In a sib test scheme where siblings tested in a commercial environment are removed from the breeding program and the correlation between body-weight measured in two different environments is 0.59, the following genotype strategies are advised for different proportions. When the genotyping is limited to less than 75% of all animals, it is advisable to sample animals based on extreme phenotypes as this strategy results in the highest accuracy. However, caution should be taken with interpreting the breeding values as this strategy shows higher levels of bias. When the sampling ratio is 75% or higher, random sampling of animals is recommended as this reduces the bias and shows similar accuracy compared to extreme phenotyping strategy.

References

Boligon, A. A., N. Long, L. G., Albuquerque, K. A., Weigel, D. Gianola *et al.* (2012) JAS 90(13):4716–4722. https://doi.org/10.2527/jas.2012-4857

Chu, T. T., Alemu, S. W., Norberg E., Sørensen, A. C., Henshall, J. Hawken, R. *et al.* (2018) Genet. Sel. Evol. 50(1):52-64. https://doi.org/10.1186/s12711-018-0430-x

Chu, T. T., Bastiaansen, J. W. M., Berg, P., Romé, H., Marois, D. *et al.* (2019) Genet. Sel. Evol. 51(1):50-62. https://doi.org/10.1186/s12711-019-0513-3

Daetwyler, H. D., Swan, A. A., van der Werf, J. H. J., and Hayes, B.J. (2012) Genet. Sel. Evol. 44(1):33-43. https://doi.org/10.1186/1297-9686-44-33

Howard, J. T., Rathje, T. A. Bruns, C. E. Wilson-Wells, D. F. Kachman, S. D. *et al.* (2018) JAS 96(11):4532–4542. https://doi.org/10.1093/jas/sky330

Jiménez-Montero, J A., González-Recio, O. and Alenda, R. (2012) Animal 6(8):1216–1224. https://doi.org/10.1017/S1751731112000341

Kapell, D. N. R. G., Hill, W. G., Neeteson, A. M., McAdam, J. and Koerhuis, A. N. M., *et al.* (2012) Poult. Sci. 91(3):565–574. https://doi.org/10.3382/ps.2011-01934

Kolmodin, R. and Bijma, P. (2004) Genet. Sel. Evol. 36(4):435-454.
https://doi.org/10.1186/1297-9686-36-4-435

Legarra, A. and Reverter, A. (2018) Genet. Sel. Evol. 50(1):53-70.
https://doi.org/10.1186/s12711-018-0426-6

Madsen, P. and Jensen, J. (2013) A User's Guide to DMU.' A Package for Analysing Multivariate Mixed Models Version 6:1–33. Available at: https://www.researchgate.net/publication/291444592_A_user's_guide_to_DMU

Misztal, I., Legarra, A. and Aguilar, I., (2009), JAS 92(9):4648–4655. https://doi.org/10.3168/jds.2009-2064

Misztal, I., Legarra, A. and Aguilar, I., (2014)' Journal of Dairy Science 97(6):3943–3952. https://doi.org/10.3168/jds.2013-7752

Misztal, I. Tsuruta, S. Strabel, T. Auvray, B. Druet, T. *et al.* (2002) Proc of 7th WCGALP, Montpeiller, France.

436. Plasticity of breed direct and individual heterosis effects in beef cattle under extensive conditions

G.M. Pyoos-Daniels[1,2*], M.D. MacNeil[1,2,3], M.M. Scholtz[1,2], M.M. Seshoka[4], H.L. Weepener[5] and F.W.C. Neser[2]

[1]ARC-Animal Production Institute, Private Bag X2, Irene, 0062, South Africa; [2]Department of Animal, Wildlife and Grassland Sciences, UFS, P.O. Box 339, Bloemfontein, 9300, South Africa; [3]Delta G, 145 Ice Cave Rd., Miles City, MT 59301, USA; [4]Northern Cape Department of Agricultural, Land Reform and Rural Development, Private Bag X9, Jan Kempdorp 8550, South Africa; [5]ARC-Soil, Climate and Water, Private Bag X79, Pretoria 0001, South Africa pyoosg@arc.agric.za

Abstract

Over a six-year period, indigenous beef cows (Afrikaner, Bonsmara, Nguni) were crossed with indigenous and exotic (Angus, Simmental) sires in a hot and arid area, to produce 15 breed groups. Inter-year variation among the breed groups in performance was large. This stimulated a study of plasticity of the breed direct effects and heterosis and considerable plasticity of these effects was observed. The interaction of breed direct effects on 205-day weaning weight indicated less plasticity of Afrikaner and Nguni than Angus and Simmental. The individual heterosis effects on 205-day weight did not interact with the year effects. This demonstrates low plasticity in the indigenous breeds, indicating more tolerance to variation in climate. Adequately resourced producers may provide exogenous inputs to offset environmental variation and use genotypes with high plasticity. However, subsistence farmers might choose less plastic genotypes, which can produce irrespective of the environment.

Introduction

A crossbreeding trial is currently running at the Vaalharts research station near Jan Kempdorp in the Northern Cape Province of South Africa, in which Afrikaner, Bonsmara and Nguni cows (Sanga types) are mated with Afrikaner, Bonsmara, Nguni, Angus and Simmental (exotic) bulls in all possible combinations. This creates 15 different genotypes. Large variation has been observed in the weaning weights of the different crosses between years. For example, Sanga sired calves and exotic sired calves had the same 205-day corrected weaning weight (171 kg) in the 2015/16 summer, which was very hot (average maximum summer temperature 32.2 °C) with well below average season rainfall of only 302 mm. In contrast, the 2016/17 summer was cooler (31.1 °C) and wetter (469 mm) and the weaning weights of the exotic sired calves were 27 kg heavier than the Sanga sired calves.

This phenomenon stimulated a study on plasticity of breed direct and individual heterosis effects in beef cattle under the extensive conditions where the crossbreeding is carried out. Plasticity refers to changes in an organism's behaviour or performance in response to a specific, sometimes unique environment (Price *et al.,* 2003). Plasticity is a way for an organism to cope with environmental variation and encompasses all types of environmentally induced changes. These changes may be temporary or permanent (Kelly *et al.,* 2012). In this study, plasticity is higher when the genetic effects show greater variation between years and lower when these interaction effects are smaller.

Santana *et al.* (2013) found that superior genotypes in favourable environments may not necessarily perform the best in unfavourable environments and this was specifically demonstrated for weaning weight. They concluded that there is genetic variation in the sensitivity of animals to a specific environment. This study aims to investigate the plasticity of breed direct and individual heterosis effects of the breeds used in the crossbreeding project.

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_436, © G.M. Pyoos-Daniels *et al.* 2022

Materials & methods

Study site. The Vaalharts research station is located in the centre of South Africa at 27°51' south and 24°50' East at an altitude of 1,175 meters and is in an area with sandy red soil with lime rock underneath (Laker, 2003). The veld type is mixed *Tarchonanthus* veld, Veld type No 16b, 4 (Acocks, 1988). The research station has a recommended carrying capacity of 10 ha/large stock unit. The climate at the Vaalharts research station is characterized by hot summers and cold winters with frost being a common occurrence. The highest monthly average temperature is around ±32 °C and is experienced during December and January. The average precipitation is around 450 millimetres per annum of which 88% is experienced during the summer months from October to April in the form of thunderstorms.

Breeds. The Afrikaner and Nguni breeds are indigenous to Southern Africa and classified as *Bos taurus africanus* (Meyer, 1984), also referred to as Sanga breeds. The Bonsmara is a composite breed developed from the Afrikaner × British Hereford and Shorthorn in a combination of 5/8 Afrikaner: 3/8 British (Scholtz, 2010). The Angus is a British breed and the Simmental a European breed.

Data collection. The data (birth, weaning and cow weights) was collected according to the approved standard operating procedures the National Beef Recording and Improvement Scheme in South Africa, which is accredited with the International Committee for Animal Recording (ICAR). The data resulted from mating Afrikaner (AF), Bonsmara (BN), Nguni (NG), Angus (AN) and Simmental (SM) bulls to AF, BN, and NG cows over a six-year period. In total 212, 324, 303, 179 and 234 calves were sired by AF, NG, BN, AN and SM sires, respectively. There were 182, 633 and 437 calves produced by AF, NG and BN cows, respectively. More details on the data collection and the management can be obtained from Pyoos *et al.* (2020). Breeding occurred on natural veld and all calves were raised by their dams from birth through weaning at approximately 205 days of age. Birth dates were recorded and calves were weighed within 48 hours of birth. Each year, all calves were weaned and weighed on the same day. It should be noted that calves are born in spring and weaned in the autumn of the next year.

The weather data from Vaalharts is unfortunately not complete. Thus, weather data from Hartswater, which is 25 km from Vaalharts, was thus used. The Temperature-Humidity Index (THI) was calculated as follows (Bohmanova *et al.* 2007):

$$THI = 0.81 \times T\ (\text{dry bulb}) + 0.143 \times RH + 0.0099 \times RH \times T\ (\text{dry bulb}) + 46.3$$

where: T (dry bulb) = ambient dry-bulb temperature (°C) and RH = relative humidity (%).

Daily THI values were calculated for the summer months from September to April. The length of the rainfall season was calculated from the first rains (at least 5 mm) to the last rains of the spring to autumn season. The number of heat wave days was calculated as the number of days within the rainfall season with THI values≥72.

Statistical models. The data were analysed using PROC Mixed of SAS. The linear models for birth weight and 205-day weight included effects of year, sex, age of dam, the genetic expectation for breed group as proposed by Dickerson (1969), and the interactions of year with the breed direct and maternal effects and individual heterosis. The direct effects of the Bonsmara in each year were set to zero to make the other effects estimable.

Results

The *P*-values for the interaction terms indicate the degree of plasticity, where a numerically larger *P*-value indicate less genetic variability between environments and hence less plasticity. For birth weight, the *P*-values for the interaction of breed direct effects and year indicate considerable plasticity with respect to the inter-year environment (Figure 1). For weaning weight, the *P*-values for the interaction effects of breed direct effects and year indicate the direct effects of Afrikaner and Nguni were less plastic and showed less variation among years than the direct effects of Angus and Simmental. Overall, birth weight, which is a manifestation of growth *in utero*, appeared more plastic, i.e. more sensitive to the variation among years, than 205-day weight.

Individual heterosis effects on birth weight varied considerably among years with the effects being large in 2014, 2016 and 2018, and essentially nil in the other years (Figure 2). However, the inter-year variation in heterosis effects on 205-day weight was smaller and overall not different from zero.

The rainy season THI during the study period varied between 78.5 and 81.6, whereas Bohmanova *et al.* (2007) indicated that heat stress is induced if THI≥72 in semi-arid regions. This indicates that the environment is very harsh.

The rainfall varied between 240 and 788 mm per season, the length of the rainfall season varied from 156 days to 228 days and the number of heat wave days within the rainfall season between 171 and 232 days.

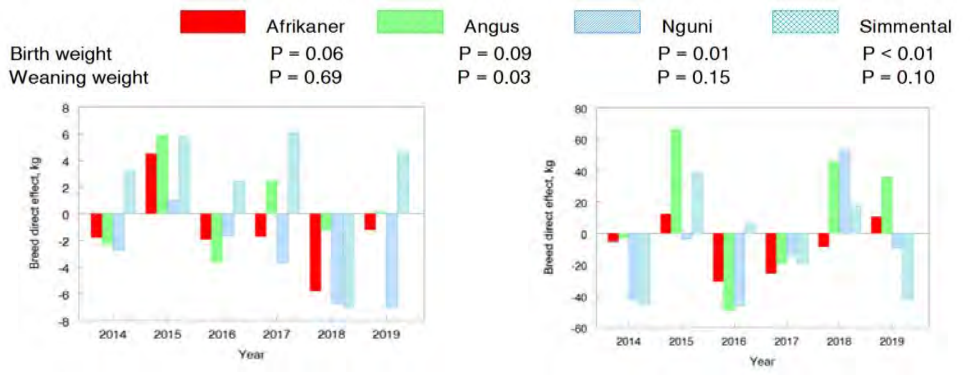


Figure 1. Interactions of the breed direct effects with year for birth (left) and 205-day (right) weights expressed as a deviation from the Bonsmara.

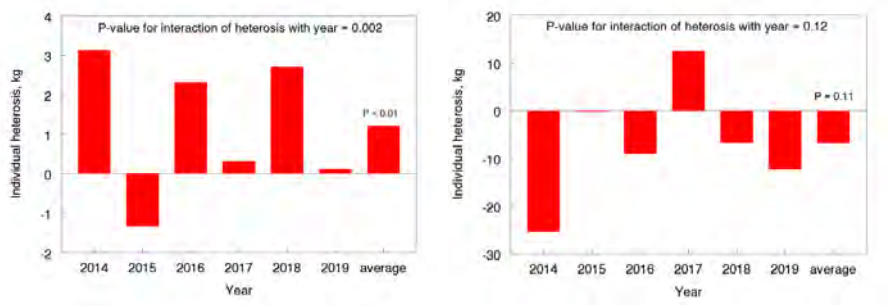


Figure 2. Interactions of the heterosis effect with year for birth (left) and 205-day (right) weights expressed as a deviation from the Bonsmara.

Discussion

It is clear that the breed direct effects for birth weight varied from year to year, and it is possibly influenced by climate. There also appears to be large interactions between breed direct effects and year for 205-day weaning weight. The Afrikaner shows the least interaction, followed by the Nguni. This demonstrates low plasticity in the indigenous breeds, indicating more tolerance to variation in climate. The exotic breeds demonstrated the more plasticity. The next step will be to use weather data such as maximum temperature, seasonal rainfall, and length of the growing season, temperature-humidity index and number of heat waves per season, to see whether it can be linked to the plasticity expressed between different genotypes.

Producing beef cattle can simultaneously optimize the use of additive (breed differences) and non-additive (heterosis) effects (Akanno *et al.,* 2018). However, it is impossible to change the breeding scheme on an annual basis to best account for inter-year environmental variation. Adequately resourced commercial producers may choose to provide exogenous inputs to offset this environmental variation, for example through the purchase of feedstuffs in drought years. They may therefore choose to use breed resources with high average genetic potential for production without regard to their plasticity. However, some commercial producers and probably all subsistence producers might choose a less plastic set of breed resources that can be dependent upon to produce even under less favourable environmental circumstances.

Acknowledgements

The support from the National Research Foundation under grant UID 135438 is acknowledged.

References

Acocks J.P.H. (1975). Memoirs of the Botanical Survey of South Africa. In: Veld types of South Africa. (eds) Botanical Research Institute. Government Printer, Pretoria, RSA.

Akanno E.C., Abo-Ismail M.K., Chen L., Crowley J.J., Wang Z, *et al.* (2018) J. Anim. Sci. 96(3):830-845. https://doi.org/10.1093/jas/sky404.185

Bohmanova J., Misztal I., and Cole J.B. (2007) J. Dairy Sci. 90(4):1947-1956. https://doi.org/10.3168/jds.2006-513

Dickerson G.E. (1969) Anim. Breed. Abstr. 37:191-202.

Kelly S.A., Panhuis T.M., and Stoehr A.M. (2012) Compr. Phys. 2(2):1417–1439. https://doi.org/10.1002/cphy.c110008

Laker M.C. (2003). Soil classification. In: A global desk reference. (eds) CRC Press. Florida, USA.

Meyer E.H.H. (1984) Proc. of the 2nd Wrld. Cong. Sheep Beef Cattle Breed, Pretoria, South Africa.

Price T.D., Qvarnström A., and Irwin D.E. (2003) Biol. Sci. 270(1523):1433–40. https://doi.org/10.1098/rspb.2003.2372

Pyoos G.M., Scholtz M.M., MacNeil M.D., Theunissen A., and Neser F.W.C. (2020) S. Afr. J. Anim. Sci. 50(3):366-377. https://doi.org/10.4314/sajas.v50i3.3

Santana M.L., Eler, J.P., Cardoso, L.G., Albuquerque, L.G., and Ferraz B.S. (2013) Animal 7(2):202-210. http://doi.org/10.1017/S1751731112001711

Scholtz M.M. (2010) Beef Breeding in South Africa. The ARC-Animal Production Institute, Pretoria, South Africa.

437. Polled Accelerator – a unique application of genomic technologies to address a beef breeding challenge

D.J. Johnston[1*], M.H. Ferdosi[1], J. Cook[1] and D.B. Savage[2]

[1]AGBU, a joint venture of NSW Department of Primary Industries and University of New England, 2351, Armidale, Australia; [2]The North Australian Pastoral Company, GPO Box 319, 4000 Brisbane, Australia; djohnsto@une.edu.au

Abstract

Genetic improvement requires selection for all traits in the breeding objective and increasingly this includes consideration of traits associated with animal welfare and social licence. A unique program called the 'Polled Accelerator' was developed for a large beef cattle population to rapidly increase the frequency of polledness. The program was constructed using a unique combination of existing and emerging genomic technologies and methods including DNA tests for polled/horn and Pompe's disease, DNA sire assignment, genomic breed composition, imputation and single-step genomic evaluation. Phenotypically polled young males were harvested from the commercial tier and through the program, high genetic merit polled animals were eligible for promotion to the multiplier and nucleus tiers of the breeding program. The application of multiple genomic technologies will allow the rapid introgression of polledness into this population without compromising the composite breed composition, breeding program structure, future genetic progress or genetic diversity.

Introduction

Surgical removal of horns from beef cattle (called dehorning) is becoming an animal welfare and management issue, and is associated with increased calf losses. Dehorning of calves improves safety of both cattle and people during handing and reduces the risk of bruising to cattle. Pain mitigation strategies are available, however there is a permanent alternative to dehorning through genetic selection, though the speed at which it can be achieved may be slow for breeds with low gene frequency for the polled allele. This is the challenge facing the breeding program of The North Australian Pastoral Company (NAPCO) that developed its own composite breed in an extensive northern Australian production environment. The breeding program is closed and relies on retained heterosis, especially for adaptation and reproduction traits. NAPCO is a vertically integrated operation, retaining ownership and management from breeding to slaughter, including a short phase in the company's feedlot prior to slaughter. The foundation breeds were horned and therefore the opportunity to change from within the nucleus tier would be very slow and there is a risk of increased inbreeding, loss of genetic diversity, change of breed composition and slower genetic progress. Selection occurs based on a defined breeding objective, however there is increasing demand to make the composite genetically polled. The Polled Accelerator (PA) program is addressing this breeding challenge through the use of new and emerging genomic tools to enable the rapid increase in polled allele frequency, without compromising the original composite breed makeup, existing genetic merit, future progress or diversity while maintaining manageable levels of inbreeding.

Materials & methods

The cattle used were a composite beef population developed at Alexandria Station, a 1.6M ha property on the Barkly Tableland region of the Northern Territory in northern Australia. The Alexandria Composite (AC) development commenced in 1982 with the crossing of the foundation breeds of Shorthorn to Brahman, and was followed by crossing those progeny with Charbray and Belmont Red (Kowald 2019). The composite was closed in 1990 and has been *inter se* mated since for 5-6 generations in the nucleus tier, resulting in a theoretical stabilised composition of 3/8 Brahman, ¼ Shorthorn, ¼ Belmont Red, 1/8 Charolais.

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_437, © D.J. Johnston *et al.* 2022

Breeding structure. A classical pyramid breeding design was implemented, with a nucleus (approx. 1,200 cows), a bull multiplier (BM) tier (approx. 2,500 cows) and a commercial tier comprising of approximately 50,000 cows. The initial composite development program and subsequently all 3 tiers of the breeding program have been run in the same location.

Performance recording and genetic evaluation. Intensive recording and genotyping occurs in the nucleus tier with approximately 1000 calves born per year. Recording includes: date of birth, birth weight, weaning weight, 400d and 600d weights, flight time, cow weight, carcase scanning, feedlot performance test, scrotal circumference and days to calving, and recently feed intake and abattoir carcase traits. Data is analysed through AC BREEDPLAN evaluation with genetic parameters and adjustment factors estimated from their own data. The evaluation is full multiple-trait, and recently transitioned to a single-step evaluation (Johnston et al. 2018) using a genomic relationship matrix (GRM) with 10,131 genotyped individuals from the nucleus tier.

Sourcing polled commercial bulls and heifers. During annual branding muster of the commercial herd, structurally sound and phenotypically polled males and females were retrieved, (males remained entire) and DNA samples taken. A total of 1,036 males (termed AC_PA animals) and 1,604 females (termed polled AC_BM breeders) were initially retained in the first two years of the program.

DNA testing. Genotyping AC nucleus animals has been occurring for more than a decade using a range of commercially available SNP chips. AC_PA and recent nucleus cohorts were genotyped and tested for polled/horn (PP, PH, HH) and Pompe's disease. All SNP genotypes were processed through the BREEDPLAN genomic pipeline (Connors et al. 2017) and all genotypes are imputed to a common density using an AC reference with 800K genotypes.

Genomic breed composition. SNP frequencies for AC were used to genomically define breed composition using the procedure described by Borner (2017). Principal component analyses were undertaken to compare the AC to the foundation breeds (viz: Brahman, Shorthorn, Belmont Red and Charolais) and the AC_PA individuals.

Single-step evaluation and selection of AC PA males. AC_PA genotypes were processed through the genomic pipeline and those that could be matched, passed SNP QC, and were >80% related to the AC genomic reference, and which were PP or PH, were included in AC BREEDPLAN single-step evaluation through their inclusion in the GRM (n=421). This generated the full suite of EBVs and AC specific BreedObject selection index ($EBV) for the AC_PA subsets of PP and PH animals, which could be selected based on direct comparisons with all 2018 born nucleus animals (NU18).

Transition of the BM tier from horned to polled. The retained young females (polled AC_BM breeders) were selected as the foundation of a new BM herd. While the contribution of these females to increasing the rate of genetic progress will be limited, their role in introducing polled alleles to the breeding program is valuable. Over a period of 6 years, the existing BM females (which are predominantly horned) will be superseded by polled AC_BM breeders. Only polled sires are selected for mating to the polled AC_BM females and these are sourced from the nucleus tier and AC_PA cohort.

Results

Genomic breed composition of AC is presented in Figure 1a and show AC is a distinct population with the first principal component differentiating on the basis of Brahman versus Charolais and Shorthorn. AC_PA

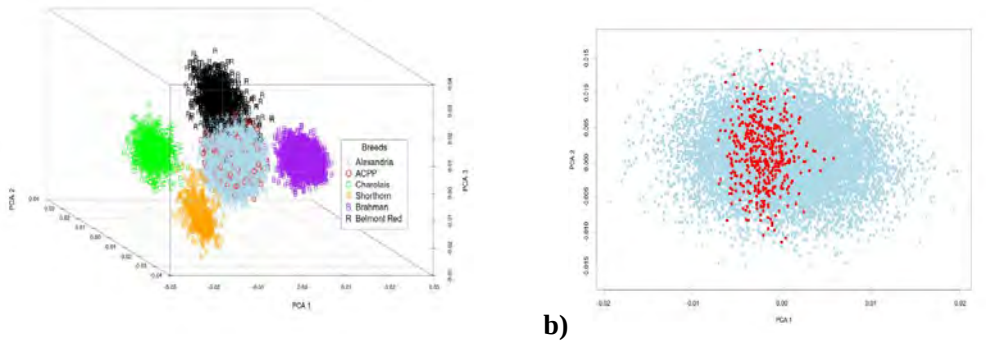


Figure 1. Principal Component analyses (a) AC and foundation breeds and (b) AC (blue) and AC_PA (red)

individuals ranged in genomic breed composition from 40-93% with 69% greater than 80% threshold and are shown in Figure 1b.

Of the 421 AC_PA eligible for inclusion in the GRM, a total of 117 were PP and 304 were PH. The frequency of polled in the NU18 was 11% with only 5 PP bulls in total. Summary statistics for $EBV (Table 1) and key EBVs and accuracies (Table 2) show the EBVs for the two groups of AC_PA (i.e. PP and PH) were similar but lower than NU18, clearly showing the commercial tier animals were genetically inferior. Long term genetic trend of 0.9 kg 600d/yr suggesting the AC_PA animals were genetically about 16 years (or 3 generations) behind the nucleus tier. However, there is considerable spread in AC_PA with a total of 16 and 28 bulls with average or above $EBV and 600d EBV, respectively.

Discussion

The PA program has identified commercial males that could be promoted to the nucleus tier to increase frequency of polled in the AC. The lower average genetic merit and differences in foundation breed percentage highlighted the risk to the program of simply promoting a polled commercial animal to the nucleus. The use of elite AC_PA will still give up some selection differential compared to using top nucleus (horned) bulls but that would likely change if polled was included in $EBV with a high economic value. Also, the AC_PA have lower individual accuracies but this risk could be minimised by using a syndicate of AC_PA selected bulls. AC_PA could be used directly in the nucleus tier (especially PP) but could also be used in the polled BM to increase the number of PP bulls introduced to the commercial tier, resulting in an immediate reduction in horn phenotype (polled being dominant) even though the commercial cow herd would remain predominantly horned in immediate future generations.

Table 1. Summary statistics for self-replacing Index ($EBV) for the 3 groups.

Group[1]	N	Mean	Std	Min.	Max.
PP	117	16.6	8.2	-6.6	40.6
PH	304	17.5	8.7	-4.2	45.6
NU_18	908	34.3	12.7	-2.4	73.5

[1] PP =AC_PA homozygous polled; PH=AC_PA heterozygous polled; NU_18=Nucleus 2018 born animals.

Table 2. Summary statistics for key EBVs from single-step BREEDPLAN evaluation.

		EBV				Accuracy			
trait[1]	Group[2]	Mean	Std	Min.	Max.	Mean	Std	Min.	Max.
BW, kg	PP	0.7	1.0	-1.8	3.3	40.6	6.9	34	57
	PH	0.8	1.1	-2.3	3.9	42.1	7.6	33	69
	NU18	1.6	1.8	-3.9	6.3	70.5	3.4	43	75
600d, kg	PP	17.1	9.2	-6.3	42.9	44.0	7.2	37	59
	PH	17.6	9.7	-4.8	51.5	45.6	7.9	36	71
	NU18	32.3	12.7	-11.9	64.0	73.2	6.1	34	82
DC, d	PP	-0.6	1.3	-4.2	3.5	15.8	5.7	11	29
	PH	-0.7	1.4	-4.7	2.8	17.1	6.4	10	37
	NU18	-2.2	2.6	-10.3	4.5	33.4	4.4	5	42
SC, cm	PP	0.4	0.7	-1.0	2.6	36.1	9.1	28	54
	PH	0.4	0.7	-1.4	3.0	38.2	9.9	27	64
	NU18	1.1	1.1	-2.0	4.7	59.5	6.0	13	68

[1] BW=birth weight; 600d=600 day weight; DC=days to calving; SC=scrotal circumference.

[2] PP=AC_PA homozygous polled; PH=AC_PA heterozygous polled; NU18=Nucleus 2018 born animals.

The PA program highlights the potential risk of simply introducing outside genetics; it could result in a change to the breed composition with no guarantee of improving genetic merit. Depending on the breed of outside genetics chosen, there could also be consequences for progeny adaptation as well as the logistical challenge of sourcing sufficient numbers for such a large herd. There is an opportunity to use mate allocation algorithms to optimise the use of AC_PA animals and to manage potential inbreeding. Gene flow analyses could be used to model the increased polled allele frequencies and to explore where the use of AC_PA animals would have the greatest benefits. EBV accuracies of future AC_PA animals could be increased by expanding the reference population for key traits. Multiple genomic technologies have been exploited in NAPCO's Polled Accelerator program to address a challenge in the breeding program where a non-genomic solution would be slow and introduce additional compromises. Opportunities exist to use more genomic technologies to further increase progress towards a polled population whilst maintaining the benefits and genetic progress for this composite breed.

Acknowledgements

Authors acknowledge the contribution of NAPCO and AGBU staff for their intellectual and operational contributions. Acknowledged is the MLA funding for AGBU's R&D of single-step BREEDPLAN and ABRI for access to NAPCO databases and extracts. Special thanks to Christian Girard for running the evaluations.

References

Borner V. (2017) *Proc. Assoc. Advan. Anim. Breed. Genet.* 22:97-100

Connors N.K, Cook J, Girard C.J. *et al.* (2017) *Proc. Assoc. Advan. Anim. Breed. Genet.* 22:317-320.

Johnston D.J, Ferdosi M.H., Connors N.K. *et al.* (2018) *Proc of 11th WCGALP*, 11:269-277.

Kowald M. (2019) You Still Can't Make It Rain: The North Australian Pastoral Company 1877-2019. North Australian Pastoral Company Pty Limited, Brisbane, Australia

438. An approach for the design of breeding programs using genomics

J.C.M. Dekkers*, H. Su, L. Kramer and H. Yu

Iowa State University, 806 Stange Rd, Ames, IA 50011, USA; jdekkers@iastate.edu

Abstract

Design and optimization of modern breeding programs requires methods and software to predict response to multiple-trait and multiple-stage breeding programs based on estimated breeding values and genomic predictions. Here, we introduce and describe the ShinySelAction software for these purposes. ShinySelAction is based on the widely used software SelAction for breeding programs with discrete generations but with a more exact implementation of multiple-stage selection programs and a more direct implementation of information from genomic predictions. This software can be used by breeders to reliably compare alternative breeding programs and for investment decisions for breeding programs that include genomic information.

Introduction

In the past decades, theory to predict selection response for quantitative traits in pedigree-based multiple-trait livestock breeding programs has been well developed, validated, and implemented, for example in the software SelAction (Rutten *et al.*, 2002). SelAction uses to deterministic recursive equations that were developed based on standard quantitative genetic and selection index theory and pseudo best linear unbiased prediction (BLUP) methodology, to model changes of trait means and variance-covariance structures as a result of selection. It's aim is to predict equilibrium responses to multiple trait selection on estimated breeding values (EBV) that are obtained under the infinitesimal genetic model with consideration of the Bulmer effect but ignoring the effect of inbreeding on genetic (co-)variances. Nowadays, breeding programs are enhanced by genomic EBV (GEBV) based on genomic prediction, which provide more accurate EBV of animals at a younger age and can improve the effectiveness of breeding programs. Schrooten *et al.* (2005) and Dekkers (2007) showed how the GEBV for a trait can be incorporated in standard selection index programs such as SelAction by specifying an additional trait with heritability (close to) one and a genetic correlation with the original phenotype-based trait equal to the accuracy of the GEBV. Genetic and phenotypic correlations of the GEBV with other phenotype-based traits and with the GEBV of other traits were derived by Dekkers (2007). These approaches require estimates of the accuracy of GEBV as inputs, which have been derived by several. Recently, Dekkers *et al.* (2021) derived accuracies of GEBV by formulating the GEBV as an index of the standard pedigree-based EBV and a conceptual EBV based on genomic-deviated-from-pedigree relationships (DEBV). Combined with already available methods to model the accuracy of pedigree-based EBV (PEBV), this allows genomic information to be incorporated following Dekkers (2007) but using the accuracy of DEBV rather than of GEBV. Against this background, the purpose of this paper is to present and describe a computer program called ShinySelAction that extends the SelAction software to incorporate genomic information. In addition, a more exact derivation of predictions for multiple-stage selection programs is implemented.

Materials & methods

Available program code for the discrete generations option of SelAction from Rutten *et al.* (2002) was translated into Python. The user interface was implemented in Shiny from R Studio (https://shiny.rstudio.com/).

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_438, © J.C.M. Dekkers *et al.* 2022

Implementation of the Bulmer effect for multiple-stage selection breeding programs was updated. In SelAction, Bulmer equilibrium parameters were derived by iterating on information available in the final stage of selection, in order to avoid multi-variate normal integrals. Because the final stage has the highest accuracy, this leads to overestimation of the reductions in genetic (co-)variances and accuracies. In ShinySelAction, exact derivations of reductions in (co-)variances of and among the index used at each stage and of true and estimated breeding values at each stage are implemented based on the moment generation function of the n-dimensional truncated multi-normal distribution, following Tallis (1961).

Genomic information is incorporated as in Dekkers (2007) by specifying genomic predictions as correlated traits with heritability (close to) one. However, following Dekkers *et al.* (2021) genomic information was incorporated based on the accuracy of DEBV, rather than that of GEBV. In the following, parameters that are required for DEBV will first be described, followed by formulation of the genetic and phenotypic correlations among DEBV for multiple traits and their correlations with the phenotype-based traits.

DEBV parameters. In ShinySelAction, parameters for genomic information that must be provided include, for each trait, the accuracy of the DEBV in selection candidates (r_{D}) and the proportion of genetic variance captured by the markers (q^2). Both parameters depend on the effective number of chromosome segments, M_e, which is a population parameter that can be derived based on empirical accuracies of GEBV and of PEBV for an available reference data set for the population, as described by Dekkers *et al.* (2021) and implemented below. The required empirical accuracies of GEBV (r_{G_r}) and PEBV (r_{A_r}) in the reference can be estimated based on available phenotypic and genotype data by leave-one-out cross-validation, following Cheng *et al.* (2021). Based on these accuracies, an estimate of M_e for the population can then be derived in the following iterative manner (Dekkers *et al.* 2021, Appendix 1):

1. Compute the accuracy of DEBV in the reference population as: $r_{D_r} = \sqrt{\frac{r_{G_r}^2 - r_{A_r}^2}{1 + r_{A_r}^2(r_{G_r}^2 - 2)}}$.
2. Set the proportion of genetic variance captured by markers for DEBV equal to 1: $q^2 = 1$.
3. Compute θ_{D_r} in the reference population as: $\theta_{D_r} = \frac{r_{D_r}^2(1 - r_{D_r}^2 q^2 h^2)}{q^2 - r_{D_r}^2}$, where h^2 is heritability.
4. Compute M_e in the reference population as: $M_e = Nq^2h^2/\theta_{D_r}$, where N is the size of the reference population.
5. Update q^2 as: $q^2 = M/(M + M_e)$, where M is the number of genotyped markers.
6. Return to step 3 until stable values for q^2 and M_e are obtained.

Accuracies of GEBV and PEBV in the reference population (r_{G_r} and r_{A_r}) for use in step 1 of the above procedure to estimate M_e can be obtained based on data (phenotypes and genotypes) from a population that is under selection and for any trait (or as the average across traits) because M_e is a population parameter that is little affected by selection and heritability of the trait, as shown by Dekkers *et al.* (2021). If a suitable reference population is not available to estimate M_e, predictions of M_e by, e.g. Goddard (2009) or others can be used.

Based on the resulting population estimates of q^2 and M_e, the accuracy of DEBV in an unselected reference population, i.e. ignoring the Bulmer effect, for a trait with base population heritability h^2 and reference size N can then be derived as follows, which is based on the correction of Bijma and Dekkers (2021) (Equation 3b) to Dekkers *et al.* (2021):

$$r_{D_r} = \sqrt{\left[1 + \theta_{D_r} - \sqrt{\left(1 + \theta_{D_r}\right)^2 - 4h^2q^2\theta_{D_r}}\right]/2h^2} \text{ with } \theta_{D_r} = Nq^2h^2/M_e$$

The accuracy of DEBV of selection candidates (=target population), r_D, depends on the accuracy of DEBV in the reference population, r_{D_r}, as derived above, and the loss of accuracy, p_{rt}, from the reference population to the target population due to recombination in each effective chromosome segment, and can be computed following Dekkers *et al.* (2021) as:

$$r_D = r_{D_r} p_{rt} = r_{D_r}(1 - 2L/M_e)^{(l_p + l_m)}$$

where L is the genome size in Morgans and l_p and l_m are the number of generations between the selection candidates and their closest paternal and maternal, respectively, ancestors in the reference population (= 1 if selection candidates are progeny of individuals in the reference population). Accuracy r_D is what is required to be entered into ShinySelAction for each trait with genomic information.

Adding genomic traits. Given the accuracy of the DEBV of selection candidates for each trait i, r_{Di}, and the proportion of genetic variance explained by markers for trait i, q_i^2 (typically assumed the same for each trait), genomic information is incorporated into ShinySelAction by adding the DEBV as genetic traits with heritability equal to 0.999. Following Dekkers (2007), the expanded genetic correlation matrix can then be derived as follows for a 3-trait example:

	BV_1	BV_2	BV_3	$DEBV_1$	$DEBV_2$	$DEBV_3$
BV_1	1	r_{g12}	r_{g13}	r_{D1}	$r_{g12} r_{D2}$	$r_{g13} r_{D3}$
BV_2	r_{g12}	1	r_{g23}	$r_{g12} r_{D1}$	r_{D2}	$r_{g23} r_{D3}$
BV_3	r_{g13}	r_{g23}	1	$r_{g13} r_{D1}$	$r_{g23} r_{D2}$	r_{D3}
$DEBV_1$	Symmetric			1	$r_{g12} r_{D1} r_{D2} / q_1 q_2$	$r_{g13} r_{D1} r_{D3} / q_1 q_3$
$DEBV_2$				$r_{g12} r_{D2} r_{D1} / q_2 q_1$	1	$r_{g23} r_{D2} r_{D3} / q_2 q_3$
$DEBV_3$				$r_{g13} r_{D3} r_{D1} / q_3 q_1$	$r_{g23} r_{D3} r_{D2} / q_3 q_2$	1

which can be computed using the following direct product (∘) of matrices:

$\begin{matrix} 1 & r_{g12} & r_{g13} \\ r_{g12} & 1 & r_{g23} \\ r_{g13} & r_{g23} & 1 \end{matrix}$	$\begin{bmatrix} 1 & r_{g12} & r_{g13} \\ r_{g12} & 1 & r_{g23} \\ r_{g13} & r_{g23} & 1 \end{bmatrix} \circ \begin{bmatrix} r_{D1} & r_{D2} & r_{D3} \\ r_{D1} & r_{D2} & r_{D3} \\ r_{D1} & r_{D2} & r_{D3} \end{bmatrix}$
	$\begin{bmatrix} 1 & r_{g12} & r_{g13} \\ r_{g12} & 1 & r_{g23} \\ r_{g13} & r_{g23} & 1 \end{bmatrix} \circ \begin{bmatrix} 1 & r_{D2}/q_2 & r_{D3}/q_3 \\ r_{D1}/q_1 & 1 & r_{D3}/q_3 \\ r_{D1}/q_1 & r_{D2}/q_2 & 1 \end{bmatrix}' \circ \begin{bmatrix} 1 & r_{D2}/q_2 & r_{D3}/q_3 \\ r_{D1}/q_1 & 1 & r_{D3}/q_3 \\ r_{D1}/q_1 & r_{D2}/q_2 & 1 \end{bmatrix}$

Similarly, the block of phenotypic correlations between BV and DEBV (upper-right block) can be computed as:

$$\begin{bmatrix} 1 & r_{g12} & r_{g13} \\ r_{g12} & 1 & r_{g23} \\ r_{g13} & r_{g23} & 1 \end{bmatrix} \circ \begin{bmatrix} r_{D1} & r_{D2} & r_{D3} \\ r_{D1} & r_{D2} & r_{D3} \\ r_{D1} & r_{D2} & r_{D3} \end{bmatrix} \circ \begin{bmatrix} h_1 & h_1 & h_1 \\ h_2 & h_2 & h_2 \\ h_3 & h_3 & h_3 \end{bmatrix}$$

Note that phenotypic correlations among DEBV are the same as their genetic correlations.

Results & discussion

The approach described enables prediction of response to selection in multi-trait breeding programs with genomic prediction. This is essential for the design and optimization of most modern livestock breeding programs. The multi-variate normal infinitesimal genetic model is assumed but changes in genetic (co-) variances among traits as a result of the Bulmer effect are considered. The effects of inbreeding on genetic

(co-)variances are, however, ignored, such that an equilibrium is reached in terms of genetic parameters and response to selection, following Rutten *et al.* (2002). This equilibrium response is a crucial outcome for comparison and optimization of breeding programs, along with the rate of inbreeding that the breeding program generates. The approach is based on the methodology implemented in the SelAction software of Rutten *et al.* (2002) but with a more exact implementation of the Bulmer effect for multiple-stage selection.

The developed approach implements information from genomic predictions by considering them to be correlated phenotypes for genetic traits that are also assumed to follow the infinitesimal genetic model but with heritability equal to one, following Schrooten *et al.* (2005) and Dekkers (2007). Key input parameters for incorporating genomic predictions are their accuracy and the proportion of genetic variance explained by the markers. Both depend on M_e as a key population parameter, which can be derived as described by Dekkers *et al.* (2021) and summarized in this paper. The approach was validated using simulation in Dekkers *et al.* (2021).

In the proposed approach, the accuracy of DEBV is derived based on a pre-determined reference population size. This does not take into account that, in an ongoing breeding program, the reference population size increases over generations. Assuming a constant size of the reference population may, however, be realistic because there is evidence that data from older generations do not contribute much information to current GEBV and may even reduce their accuracy (Weng *et al.* 2016, Howard *et al.* 2018, Hidalgo *et al.* 2021).

Future developments include allowing for groups of selection candidates with different amounts of information, beyond male versus female selection candidates. This will allow the modelling of breeding programs where not all selection candidates are genotyped or not all selection candidates are phenotyped for all traits.

Acknowledgements

This research was funded by USDA-NIFA grant #2017-67015-26299. The authors thank Piter Bijma for providing the SelAction program code and Julius van der Werf for useful discussions.

References

Bijma, P., and Dekkers, J.C.M. (2022) Genetics Selection Evolution

Cheng J., Dekkers J.C., and Fernando R.L. (2021) J. Animal Breeding and Genetics 138(5):519-527 https://doi.org/10.1111/jbg.12545

Dekkers J.C.M. (2007) J. Animal Breeding and Genetics 124(6):331-341. https://doi.org/10.1111/j.1439-0388.2007.00701.x

Dekkers J.C.M., Su H., and Cheng J. (2021) Genetics Selection Evolution 53(1):1-23. https://doi.org/10.1186/s12711-021-00647-w

Goddard M. (2009) Genetica 136(2):245-57. https://doi.org/10.1007/s10709-008-9308-0

Hidalgo, J., Lourenco, D., Tsuruta, S., Masuda, Y., Breen, V., Hawken, R., Bermann, M. and Misztal, I., 2021. *J. Animal Sci*, *99*(9), p.skab239. https://doi.org/10.1093/jas/skab239

Howard, J.T., Rathje, T.A., Bruns, C.E., Wilson-Wells, D.F., Kachman, S.D. and Spangler, M.L., 2018. *J. Animal Breed.&Genet.*, *135*(4), pp.251-262. https://doi.org/10.1111/jbg.12334

Rutten M.J.M., Bijma P., Woolliams J.A., and Van Arendonk J.A.M. (2002) J. Heredity 93(6):456-458. https://doi.org/10.1093/jhered/93.6.456

Schrooten C., Bovenhuis H., Van Arendonk J.A.M., and Bijma P. (2005) J. Dairy Sci. 88: 1569–1581. https://doi.org/10.3168/jds.S0022-0302(05)72826-5

Tallis G.M. (1961) J. Royal Statistical Society: Series B (Methodological) 23(1): 223-229. https://doi.org/10.1111/j.2517-6161.1961.tb00408.x

Weng, Z., Wolc, A., Shen, X., Fernando, R.L., Dekkers, J.C., Arango, J., Settar, P., Fulton, J.E., O'Sullivan, N.P. and Garrick, D.J., 2016. Genetics Selection Evolution, 48(1) 22. https://doi.org/10.1186/s12711-016-0198-9

439. Should we breed cows like pigs? A new organisation of dairy cattle breeding

M. Kargo[1*], H.M. Nielsen[1], S. Østergård[2,3], R.D. Kring[3], M. Slagboom[1], J.B. Clasen[1] and J.R. Thomasen[4]

[1]Aarhus University, Centre for Quantitative Genetics and Genomics, Blichers Alle 20, 8830 Tjele, Denmark; [2]Aarhus University, Animal Science, Blichers Alle 20, 8830 Tjele, Denmark; [3]SimHerd A/S, Niels Pedersens Allé 2, 8830 Tjele, Denmark; [4]VikingGenetics, Ebeltoftvej 16, 8960 Randers SØ, Denmark; morten.kargo@qgg.au.dk

Abstract

Crossbreeding is the most common breeding system for pigs and poultry, but not for dairy cattle. When selecting purebred lines for crossbred performance, it becomes possible to take advantage of both breed complementarity and heterosis. In this paper, we present a new structure of dairy cattle breeding with the aim to improve crossbred performance in production cows. The structure of the dairy sector can be organised like a pyramid with nucleus herds, multiplier herds, and production herds. Genetic improvement of purebreds takes place in the nucleus. Surplus females from the breeding nucleus are transferred to the multiplier level, where they are used as dams of crossbred animals for production herds. In addition, we hypothesise that joint breeding planning and formulation of breeding goals for each parental breed are necessary in order to optimise crossbred performance.

Introduction

Crossbreeding is the most common breeding system for pigs and poultry. During the last decades, crossbreeding has also increasingly been used in dairy cattle (Clasen, 2021), primarily because of better profitability of crossbred cows due to heterosis for important production and functional traits (Sørensen *et al.*, 2008). Until now, crossbreeding systems in dairy cattle, such as two- or three breed rotational crossbreeding or terminal crossbreeding have only been managed at herd level. To our knowledge, no breeding organisation selects their purebred lines specifically for crossbred performance (CP) and no crossbreeding programs are coordinated at sector level. When purebred lines are specialised to be used in crossbreeding programs, it becomes possible to take advantage of both breed complementarity and heterosis to improve performance of crossbred production animals. Breed complementarity means that the traits of the two breeds complement each other. Breed complementarity may be especially large when traits are unfavourably correlated, such as between milk yield and resistance to mastitis or milk yield and reproduction traits. The use of specialised sire and dam lines as typically used in pig and poultry breeding (Smith, 1964), could previously not be applied for dairy cattle to the same extent because nearly all female offspring were used as replacements due to the low reproduction rate of cattle. Today, the use of sexed semen and multiple ovulation embryo transfer allow for a change of dairy cattle breeding into a multi-layer structure similar to what is found in poultry and pig breeding. Currently, the organisation of dairy cattle breeding is primarily carried out with the purpose of improving the genetic level of purebred populations. Therefore, AI bulls used in crossbred herds can only be selected among AI bulls that are already selected for use in purebred herds. A change in the structure of dairy cattle production at herd and sector level will make it possible to exploit crossbreeding in a better and more efficient way.

In this paper, we present an idea for a new structure of dairy cattle breeding, with the aim to improve CP in production cows. We argue that a multilayer structure as typically seen in pigs and poultry can improve the competitiveness of the dairy sector by utilising the full potential of crossbreeding between dairy breeds.

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_439, © M. Kargo *et al.* 2022

A new dairy breeding structure

The structure of the dairy sector can be organized like the pyramidal structure in pig breeding with nucleus herds, multiplier herds, and production herds (Figure 1).

At the top of the pyramid are the nucleus herds where genetic improvement of the pure breeds takes place. The population sizes of the nucleus herds need to fulfil the requirements for genetic gain given an acceptable level of inbreeding. The breeding scheme in the nucleus herds rely on intensive use of ovum pick-up on young heifers that are selected based on genomic breeding values to increase selection intensity and decrease generation interval. Surplus females from the nucleus herds are transferred to the multiplier herds, where they are used as dams for production of F1 animals. The multiplier herds must be able to deliver enough F1 females to meet the demands from production herds. The demand for F1 females can be reached through intense use of embryo flushing of purebred donors with sexed semen from another dairy breed to produce F1 embryos. Recipients for the embryos will be F1 heifers, which are reared within the multiplier herds. The F1 heifers calve in the multiplier herds and are hereafter sold to the production herds as soon as possible. An alternative is to produce F1 embryos in the nucleus herds, while the recipients remain in the multiplier herds. In that case, management tasks will be more specialized at each level. In either case, the production herds will consist of F1 females and will not be producing their own replacement animals. Instead, all F1 production cows, except if they need to be contracted by the multiplier herds as recipients, are inseminated with beef semen from a terminal beef sire line suited for production of beef on dairy calves (Hjortø, personal communication). Taking this approach, dairy producers in the production herds can fully focus on milk production.

Such a terminal crossbreeding program can be extended by a third breed. The advantage of a three-breed crossbreeding program is to be able to combine the strength of more breeds for specific production systems.

Breeding goals. We hypothesise that joint breeding planning and setting of breeding goals for each pure breed are necessary in order to ensure that genetic levels, heterosis and complementary effects in the crossbred production animals are at the highest possible level. This will lead to higher CP relative to the present situation, where each parental breed is managed independently of the other breeds focusing on purebred performance. Even without assuming non-additive gene action for individual traits, a number of traits combined multiplicatively makes up the total economic merit of an individual. The consequence is that

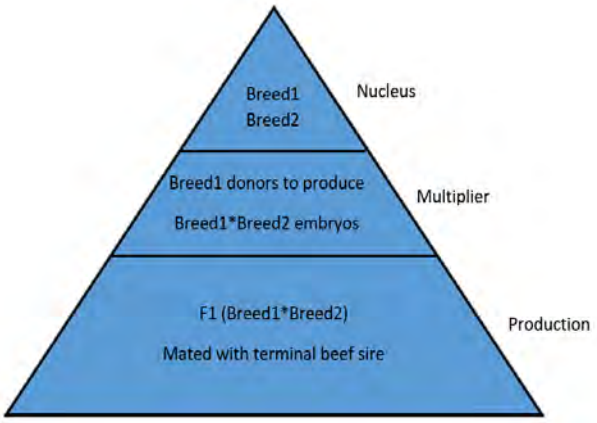


Figure 1. Structure in a terminal F1 dairy crossbreeding program with two breeds.

there is an additional crossbred superiority for total economic merit on top of the heterosis. This additional superiority is larger if the individual traits are highly different in the involved breeds. The economic value of this crossbred superiority should be quantified for the whole sector using decision support tools such as SimHerd Crossbred (Østergaard et al., 2018). In addition, economic weights to be used in breeding goals of the parental breeds need to be derived to obtain the crossbred superiority. Setting diverse breeding goals across the breeds which complement each other will increase the overall genetic diversity and sustainability of the dairy cattle populations globally, on top of creating more profitable production animals.

Breeding value estimation. In the new structure for dairy cattle breeding, the aim is to improve the performance of crossbred cows. This performance depends both on genetic merit of the traits for the parental pure breeds but also on heterosis on a given trait. Therefore, selection for CP involves both improving purebred performance and heterosis. In pig breeding, records from crossbred animals in production herds are often lacking and selection is based on performance of purebreds. On the contrary to pigs, most countries routinely register the performance of the majority of dairy cows, providing the opportunity to select purebreds directly for CP. However, Esfandyari *et al.* (2017) demonstrated that selection for CP is only favourable for a few generations. Therefore, combined selection for CP and purebred performance is expected to result in a better performance of crossbred animals in the long term compared to selection for CP only. Statistical models that utilise phenotypic information in crossbred animals to estimate both genetic merit in the purebreds and crossbreds with high accuracy are essential. In the new structure, it is assumed that the selection of purebreds is based on genomic information and crossbred animals providing phenotypes are also genomically tested.

Discussion

In this paper, we presented a new structure of dairy cattle breeding with the aim to improve CP in production cows. The large advantage with the presented structure is that dairy producers will be able to take the benefits of crossbreeding to an increased level, as parental breeds are selected with the aim of improving crossbred performance. Furthermore, producers will be able to fully focus on milk production since breeding is carried out in the nucleus herds, rearing of heifers in the multiplier herds, and fattening of the beef×dairy offspring by meat producers. The presented structure can however only succeed in a cooperative organisation where the parental breeds are managed together and follow the breeding goals set to optimize crossbred performance. These breeding goals may not be the most optimal in case each of the pure breeds are used for commercial production. That is why the number of animals in the nucleus need to be minimized given the constraint on acceptable genetic gain. Whether or not such a breeding program will be economically feasible will depend on the number of production cows in the crossbreeding program, and on the extra economic gain that can be obtained by using breeding material from pure breeds selected for crossbred performance instead of breeding material from pure breeds selected for purebred performance. The program will be economically feasible if the cost of running such crossbreeding programs will be lower than the extra gain obtained in the production herds due to increased heterosis and improved complementarity from having lines suited for crossbreeding. It needs to be investigated how much extra heterosis is necessary to justify the costs of the breeding scheme, proposed in this paper. In addition we need to estimate how large the crossbred production population needs to be to be able to pay for the extra cost given different levels of increased heterosis and breed complementarity to be achieved. Another advantage of the presented program is increased breed and genetic variation, which may be of value for the future. All these questions are to be answered in the Danish project DairyCross supported by the green development and demonstration programme (GUDP) from the Danish ministry of food, agriculture and fisheries DairyCross project supported by the green development and demonstration programme (GUDP) from the Danish ministry of food, agriculture and fisheries (J. nr. 34009-18-1365). https://qgg.au.dk/projekter/projektdeltagelse/dairycross-krydsning-optimering-paa-tvaers-af-malkeracer/

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_439, © O. Girardie *et al.* 2022

References

Clasen, J.B. (2021) Crossbreeding as a strategy in dairy cattle herds. Swedish University of Agricultural Sciences, Uppsala, Sweden.

Esfandyari H., Henryon M., Berg P., Thomasen J.R., Bijma P. *et al.* (2017) J Hered, 108(3):318-327. https://doi.org/10.1093/jhered/esw123

Ettema J.F., Thomasen J.R., Hjortø L., Kargo M., Østergaard S. *et al.* (2017) J Dairy Sci 100(5):4161-4171. https://doi.org/10.3168/jds.2016-11333

Østergaard S., Ettema J.F., Clasen J.B., and Kargo M. (2018) Proc. of the 11th WCGALP, Auckland, New Zealand

Smith, C. (1964). Ani. Sci. 6, 3: 337-344 https://doi.org/10.1017/S0003356100022133

Sørensen M.K., Norberg E., Pedersen J., and Christensen L.G. (2008) J Dairy Sci 91(11):4116-4128. https://doi.org/10.3168/jds.2008-1273

Part 25

Operational breeding programs

440. Management of undesired alleles at multiple loci for a goat breeding program

J. Raoul[1,2*], J.M. Elsen[1], S.T. Rodríguez-Ramilo[1] and I. Palhière[1]

[1]GenPhySE, Université de Toulouse, INRAE, ENVT, CS 52627, 31 326 Castanet-Tolosan, France; [2]Idele, CS 52637, 31321 Castanet-Tolosan, France; jerome.raoul@inrae.fr

Abstract

As the number of discovered loci carrying undesirable alleles continues to increase, culling carrier animals is not an option. Thus, we used a method that derives a genetic value for each genotype at a given locus. Using stochastic simulations, we assessed various strategies for a goat breeding program that combined two selection criteria used in a preselection step and four mating strategies between sires and dams to produce candidates for artificial insemination. Even without a strategy dedicated to reduce undesirable allele frequency, lethal alleles were purged and their frequencies slowly decreased whereas the average frequency of non-lethal alleles remained constant. Using, in a preselection step, an index including the genetic values for loci likely to carry an undesirable allele, was an efficient alternative to reduce the frequencies of non-lethal alleles and gave higher decrease for lethal alleles frequencies. After selection, the ability to control the frequencies of undesired alleles was similar across mating strategies. However, some mating strategies, based on optimizing an objective function including at least genetic values and increased inbreeding, enable to reduce undesired alleles at no or small cost for both genetic gain and increase in inbreeding.

Introduction

In sheep and goat selected populations, the number of discovered loci carrying undesirable alleles is growing, both including lethal (e.g. Ben-Braiek *et al.*, 2021) and non-lethal (e.g. Martin *et al.*, 2016) alleles. The control of known undesired alleles can be operated at a preselection step by culling carrier animals. This method becomes difficult to adopt due to the increase in the number of undesired alleles to be included. A selection criterion that incorporates both the estimated genetic values for selected traits and a set of genetic values associated to genotypes at these specific loci can be applied (Boichard *et al.*, 2016). The undesirable effects can also be controlled by optimizing mating strategies to avoid producing homozygous carriers as in Cole (2015). In French goat breeding programs, only a few matings are optimized at the individual level, i.e. those to produce male candidates for artificial insemination (AI).

Focusing on a breeding program for goats, we performed stochastic simulations to assess the annual genetic gain, the increase in inbreeding and the control of undesired alleles (known and unknown i.e. genotyped and not genotyped) according to various designs combining a preselection step and strategies to select and optimize matings that produce male candidates for AI. Two cases have been separately considered: lethal and non-lethal alleles.

Materials & methods

Using a stochastic model, we mimicked the Saanen goat breeding program of the French Breeding Society Capgenes, where around 70, 000 females divided in 196 flocks are recorded each year for a maternal trait.

To set up a reference population for genomic selection we first applied a classical design based on the progeny testing of AI sires over 10 years. The next fifteen years, we applied eight different genomic breeding program designs (2 types of selection index × 4 mating systems). The male candidates, elite dams and elite AI sires were selected based on: (1) on GEBV; or (2) an index *ISp* that combines GEBV and genetic values at five specific loci potentially carrying undesired alleles (*R*).

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_440, © J. Raoul *et al.* 2022

For an individual i, $ISp_i = \alpha GEBV_i + \sum_{r=1}^{R} \hat{\beta}_{rg_i}$ where α denotes the relative emphasis put on both the selection trait and allele management, and $\hat{\beta}_{rg_i}$ the genetic value associated to the locus r given the genotype g of the individual i.

As proposed by Boichard *et al.* (2016), $\hat{\beta}_{rg}$ is $2f^2c_r$, $-f(1-2f)c_r$ and $-2f(1-f)c_r$ for $g = ++$, $g = m+$ and $g = mm$, respectively. Where f denotes the frequency of the deleterious allele m, and c_r the cost associated to the genotype mm. These genetic values assume that m is recessive.

Then, we considered 4 mating systems between elite dams and sires that produce male candidates for AI:

1. Elite dams and sires were randomly mated (RAND)
2. Elite dams and sires were mated according GEBV (assortative mating) but matings between animals with a common grandparent were excluded. (AM)
3. Matings were designed based on an objective function OFP_{ij} including the GEBV of a dam i and an elite sire j and and their kinship coefficient F_{ij} (from Pryce *et al.*, 2012): $OFP_{ij} = 0.5(GEBV_i + GEBV_j) - \lambda F_{ij}$, where λ is a penalty coefficient (arbitrarily set to 0.1) and F_{ij} is the pedigree-based inbreeding coefficient of the offspring resulted from mating dam i and elite sire j.
4. Matings were designed based on the previous objective function plus the probability that a mating between a dam i and elite sire j produces an homozygous carrier $g = mm$ for the locus r (from Cole *et al.*, 2014): $OFC_{ij} = OFP_{ij} - \sum_{r=1}^{R} P(g = mm)_{ij,r} \times c_r$.

In a first approach all c_r were arbitrarily set to 1 for all loci r (genetic values and *OFP*).

In (3) and (4), a score based on the objective function was computed for all potential matings between elite AI sires and female candidates for AI (around 80% of adult females). Then 1,050 matings were sequentially selected based on the best score and under constraints (one female is mated with only one male and a maximum of 3,000 females per male are inseminated along his reproductive life).

In addition, to assure the replacement of females, 46% of the breeding females were selected on (G)EBV and mated to an AI sire. Other females and females that did not conceive with AI (fertility after AI was set to 60%) were mated to a natural mating sire. Natural mating sires were all self-replaced within flocks and born from AI sire. The number of progenies per dam depended on the mode of reproduction and parity. Some dams were randomly culled after each reproductive cycle and the maximum parity was five. Around 30% of dams were replaced per year by females that were preferentially chosen among newborn progeny from AI matings. No selection step was considered for newborn females before they were mated.

We simulated individuals and their genome based on a real goat 50 K SNPs chip, as in Raoul *et al.* (2017). In addition, we simulated 5×2=10 loci (5 known and 5 unknown) with a starting frequency respectively equal to 1, 4, 7, 10 and 13%. These loci are independent (no linkage disequilibrium) from the rest of the genome (i.e. 50 K SNPs). Within simulations, undesirable alleles are all lethal (homozygous carriers were stillborn) or all non-lethal (no effect on fitness). In both cases, they did not have any effect on the main selected maternal trait (h^2=0.3, rep=0.5). The model included: (1) the establishment of the founder population; (2) the simulation of 1000 QTL picked from the 50K SNPs and their effect drawn from a Gamma distribution; (3) the simulation of phenotypes and genotypes based on the 50K SNP panels (SNPs selected as QTL were excluded from genotypes); (4) and the estimation of both breeding values and pedigree inbreeding respectively. The last estimations were performed with single step GBLUP using the Blupf90 software developed by Misztal *et al.* (1999) and inbupgf90 software developed by Aguilar *et al.* (2008).

The annual genetic gain ΔG and annual increase in inbreeding ΔF were estimated as the regression slope of the average true breeding value and inbreeding coefficients of first parity females on time for years 10 to 25,

respectively. As well the annual decrease in frequency for known (i.e. genotyped) Δf_k and unknown (i.e. not genotyped) Δf_u loci were estimated from years 10 to 25 and averaged across known and unknown loci separately. The means and standard deviations presented are based on 30 replicates.

Results

Figure 1 gives the annual genetic gain ΔG, increase in inbreeding ΔF and changes in the frequencies of known Δf_k and unknown Δf_u alleles according to 8 breeding strategies. In Figure 1a (lethal alleles) we observed that the selection based on *ISp* gave faster decrease in known undesired alleles frequencies than GEBV only, regardless of the mating design. The frequencies of unknown lethal alleles decreased in all breeding strategies without significant differences on average across breeding strategies. The frequencies tended to converge below 6% at time 10 and below 5% at time 25 for GEBV based strategies (results not shown). For *ISp* based strategies, the trend is the same until time 10 but the frequencies of genotyped lethal alleles tended to converge below 2% at time 25 (results not shown). No significant difference across strategies was observed for the frequencies of ungenotyped alleles. For non-lethal alleles we observed on Figure 1b that ungenotyped alleles are subject to genetic drift and their frequency remains, on average across loci, constant over time. For genotyped alleles, decrease in frequency is significantly higher for *Isp* than GEBV strategies regardless of the mating optimization. The decrease in frequency of the undesired allele at specific loci is poorly influenced by the mating system. We observed a slightly better result for OFP

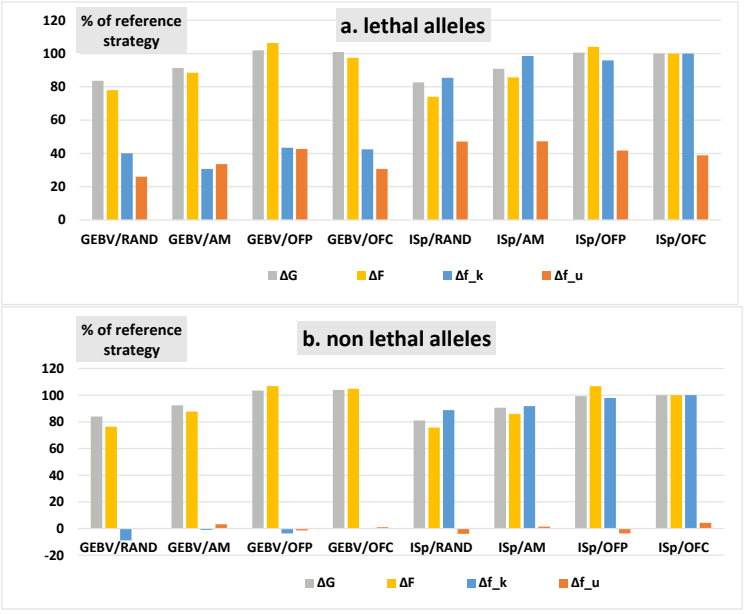


Figure 1. Annual genetic gain (ΔG), increased in inbreeding (ΔF) and changes in the frequencies of known (Δf_k) and unknown (Δf_u) alleles according to the selection criteria × mating design (30 replicates) for lethal (a) and non-lethal (b) alleles. ΔG, ΔF and both Δf_k and Δf_u are respectively expressed as the % of ΔG, ΔF and Δf_k obtained for the reference strategy = ISp/OFC. a. Reference strategy: ΔG=0.180 σa, ΔF=0.33% per generation, Δf_k = -0.15% per year. b. Reference strategy: ΔG=0.179 σa, ΔF=0.32% per generation, Δf_k = -0.38% per year. Selection criteria: GEBV or *ISp* (index combining GEBV and genetic values at five specific loci potentially carrying undesired alleles). Mating designs: RAND (random mating), AM (assortative mating), OFP (objective function including GEBV and penalty to F), OFC (OFP + penalty to probabilities to produce homozygous carriers at five specific loci potentially carrying undesired alleles).

and OFC compared to RAND and AM for non-lethal alleles (significant) and even smaller differences for lethal alleles (non-significant). Avoiding the production of homozygous individuals at specific loci had no or a small effect on the control of the allele frequencies. However, the optimization of matings had significant consequences on the genetic gains and the increase in inbreeding. OFP and OFC gave higher ΔG (but also higher ΔF) than RAND and AM. A better adjustment of λ (penalty coefficient applied to ΔF in OFP and OFC strategies) would be necessary to fairly compare ΔG across strategies.

Discussion

These preliminary results show that including a genetic value at specific loci is an efficient alternative to reduce their frequencies in a goat breeding scheme especially for non-lethal alleles. The genetic architecture simulated in this study, based on an independent segregation of specific loci and other markers as well as the use of a penalization on a pedigree based ΔF, do not enable to assess the apparent contradiction between a general maximization of the SNP heterozygosity and the fixing of favourable alleles at the level of specific loci and QTLs. To go further it is necessary to simulate specific loci which are in linkage disequilibrium with markers and QTLs, use a penalization on a genomic based ΔF and analyse the sensitivity of the result to the value of both λ and c_r.

Acknowledgements

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 772787 (SMARTER) and from the French Minister of Agriculture (CASDAR PRESAGE).

References

Aguilar I., Misztal I. (2008) J Dairy Sci. 2008;91:1669–1672.

Ben Braiek M., Fabre S., Hozé C., Astruc J.M., Moreno-Romieux C. (2021) Genet Sel Evol. 1;53(1):41. https://doi.org/10.1186/s12711-021-00634-1.

Boichard D., Grohs C., Michot P., Danchin-Burge C., Capitan A., *et al.* (2016) Prod. Anim., 29, 353-360. https://doi.org/10.20870/productions-animales.2016.29.5.3003

Cole J.B. (2015) Genet Sel Evol. Nov 30;47:94. https://doi.org/10.1186/s12711-015-0174-9.

Martin P. M., Palhière I., Ricard A., Tosser-Klopp G., Rupp R. (2016) PLoS One. Mar 31;11(3):e0152426. https://doi.org/10.1371/journal.pone.0152426.

Misztal I. (1999) Interbull Bull. 1999;20:33-42.

Raoul J., Swan A.A., Elsen J.M. (2017) Genet Sel Evol. 49(1):76. https://doi.org/10.1186/s12711-017-0351-0.

441. Prediction of gametic variance and its use in bovine breeding programs

C. Hozé[1,2*], A. Baur[1,2], S. Fritz[1,2] and D. Boichard[2]

[1]Allice, 75595 Paris, France; [2]Université Paris Saclay, INRAE, AgroParisTech, GABI, 78350 Jouy-en-Josas, France; chris.hoze@allice.fr

Abstract

With genomic selection, gametic variance has been of increasing interest for breeding top animals. Two gametic variance estimates based on SNP effects estimated in 2018, were obtained deterministically and by simulating progeny. They were compared with each other, with the variance of the genomic breeding values of real progeny in fall 2020, and with the variability of the phenotypes of the female offspring. As expected, both predictors of gametic variances were similar, provided that the number of simulated progeny is large enough (500). The correlation between those predictors and gametic variances observed on real progeny was rather high, although limited by the number of offspring available. The correlation between the gametic variance of the sires and the phenotypic variance ranged from 0.14 to 0.85 (mean=0.33) according to the heritability of the trait. Gametic variance is useful to select parents of future bulls, but its potential to control phenotypic variability is more limited.

Introduction

Since its implementation in France in 2009, genomic selection has had a tremendous development, all bulls are genomically evaluated and more and more females are genotyped annually. Young bulls are selected on the basis of their genomic values among a large number of male progeny of bull sires and bull dams. Under the polygenic model, gametic variances of each parent or, equivalently, the Mendelian sampling (MS) variance of the progeny depend only on inbreeding F_s and F_d of the parents [$Var(MS) = 0.5\ \sigma_g^2\ (1 - 0.5\ F_s - 0.5F_d)$] and is constant without inbreeding. In genomic selection, not only the MS term of the progeny can be predicted but also the gametic variance of each parent, provided that SNP effects and genotype phases of the parents are known. According to the large investments necessary to select the very best young bulls (matings, genotyping, embryo production and transfer), it is essential to maximize the probability to generate and find them. This probability depends not only on the genetic merit of the parents but also on their gametic variance. As already shown in the literature, gametic variance is of major importance when breeding top animals (Bijma *et al.*, 2020; Segelke *et al.*, 2014). However, in spite of its increasing interest, little investigation was done on its practical prediction and use. In this study, we (1) compared two approaches to predict the gametic variance of the genotyped parents; (2) investigated the impact of gametic variance in the selection success of candidates; (3) compared the predicted gametic variances with the variance of genomic breeding values of real genotyped progeny, and (4) with the phenotypic variability of the daughters. In this validation procedure, SNP effects were estimated in French Holstein population with the information available in June 2018 and genomic breeding values and phenotypes of progeny were those available in fall 2021.

Materials & methods

Population studied. We used genotypes and breeding values for 426 widely used bulls in Holstein breeds born between 1991 and 2016. Bulls were selected to have more than 200 daughters with morphology phenotypes, 1000 daughters with lactation phenotypes and 500 genotyped progeny. Genotypes of sires were obtained with different Illumina chip platforms with at least 50K SNPs, and imputed and phased with Fimpute3 software (Sargolzaei *et al.*, 2014).

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_440, © C. Hozé *et al.* 2022

SNP effects were obtained from the French national genomic evaluation (Boichard *et al.*, 2012) computed in June 2018. These effects were used as input for simulation software and gametic variability prediction. Five production and 10 morphological traits were studied.

Simulation procedure and prediction of gametic variance. Two different approaches were used to predict the gametic variance of each bull. The first one was based on simulation (Segelke *et al.*, 2014). For each bull mated to one simulated fully homozygous dam, 500 progeny were simulated. For each progeny, segregations were randomly sampled and recombinations were simulated assuming 1cM corresponds to 1 Mbase. Progeny direct genomic value was obtained by combining the SNP effects with its genotypes. The MS variance was estimated by the variance of the DGV (direct genomic value) of the 500 progeny. Because SNP effects are estimated with an accuracy lower than 1, the MS variance was underestimated.

The second approach is the one described by Santos *et al.* (2019). The gametic variance (V_{theo}) of each sire was predicted from the contrasts within loci with the following formula:

$$V_{theo} = 0.25 \left[\Sigma_i \alpha_i^2 + 2 \Sigma_i \Sigma_{j>i} \alpha_i \alpha_j (1 - 2r_{ij})\right] \quad (1)$$

where: α_i is the signed contrast between effects of the paternal and maternal alleles for marker i, and r_{ij} is the recombination rate between markers i and j. Recombination rate was assumed equal to 0.5 across chromosomes.

DGV and yield deviations (YD) obtained from the French National evaluation run computed in November 2021 were used to validate gametic variance predictions on real data. YD were performances of the female offspring adjusted for fixed effects of the evaluation model and dam genetic merit. For each sire, we computed the variance (V_{obs}) of DGV of progeny genotyped after 2018. The phenotypic variance (V_{yd}) was obtained by computing YD variance. Due to the increase in reference population size and change in French evaluation model (especially change in imputation results due to switch from UMD3 to ARS-UCD1 assembly), markers and haplotypes effects might differ between June 2018 and November 2021, allowing us to test the robustness of our predictors.

Results

Comparison of gametic variance predictions. The predicted and simulation based gametic variances were very similar, with the same mean, very close standard deviations, and correlations in the [0.82-0.94] interval. A larger number of simulated progeny would probably have led to higher correlations. Therefore in practice, the deterministic approach is preferable because it is more accurate and faster to compute. But simulations can be a more flexible tool especially to study covariations between traits or more complex genetic determinisms with dominance or epistasis.

The same indicators computed with 2021 SNP estimates were consistent with those of 2018, showing a good conservation of SNP effect: the correlations between V_{theo} in 2018 and 2021 ranged from 0.87 to 0.99 according to traits. Correlations are presented on Table 1.

Heterogeneity of gametic variance between bulls. Gametic variances observed on real progeny DGVs (V_{obs}) varied widely between sires and ranged from simple to triple for most traits. This is much more than expected under the polygenic model. Figure 1 shows the distribution of V_{obs} for fat content (left) and stature (right). Differences were particularly large for fat content due to the effect of DGAT1 major gene, and lower but still important for stature with most values ranging from 0.3 to 0.9. This large variability of gametic variance observed across animals highlights its interest when breeding top animals. Assuming

Table 1. Correlations between theoretical gametic variance computed in June 2018 (V_{theo}_June18), others indicators of gametic variance and observed progeny variance.

Traits	V_{sim}_June18	V_{sim}_Nov21	V_{theo}_Nov21	V_{obs}_Nov21	V_{yd}_Nov21
Protein yield	0.91	0.76	0.87	0.42	0.44
Fat yield	0.94	0.94	0.98	0.77	0.16
Protein content	0.88	0.89	0.99	0.84	0.46
Fat content	0.95	0.95	1.00	0.93	0.85
Milk	0.90	0.84	0.97	0.73	0.15
Udder support	0.87	0.77	0.88	0.59	0.17
Udder depth	0.84	0.78	0.93	0.37	0.35
Fore udder attachment	0.94	0.88	0.95	0.63	0.23
Rear udder height	0.91	0.88	0.92	0.55	0.14
Fore teat distance	0.90	0.89	0.95	0.64	0.32
Stature	0.80	0.83	0.94	0.64	0.47
Body depth	0.90	0.85	0.94	0.71	0.38
Width at ischium	0.96	0.92	0.96	0.66	0.27
Locomotion	0.87	0.83	0.89	0.51	0.24
Body condition score	0.93	0.83	0.94	0.72	0.35
Average	0.90	0.86	0.94	0.65	0.33

[1] Theoretical gametic variance computed either in November 2021 (Vtheo_Nov21), Simulated gametic variance computed either in November 2021 (Vsim_Nov21) or June 2018 (Vsim_June18), Observed gametic variance computed in November 2021 on DGV (Vobs_Nov21) or on YD (Vyd_Nov21).

these predictions are accurate and reflect true gametic variance (this depends on the accuracy of SNP effects), very good predictions of extreme deciles can be obtained, as well as the probability to obtain an elite progeny above a given threshold. For instance, for milk yield, the prediction of the 90th percentile was: P90% = 0.999 Genomic Pedigree Index + 1.29 σ_{MS} (R^2=0.99), with σ_{MS} the MS standard deviation for a given couple of parents.

Validation of predicted gametic variance. Depending on the traits, the correlations between predicted variance V_{theo} in 2018 and observed variance V_{obs} in 2021 ranged from 0.47 to 0.92 and reached 0.62 on average. Figure 2 presents these results for fat content and stature. These moderate values probably reflect both the limited number of progeny and the variability due to dams.

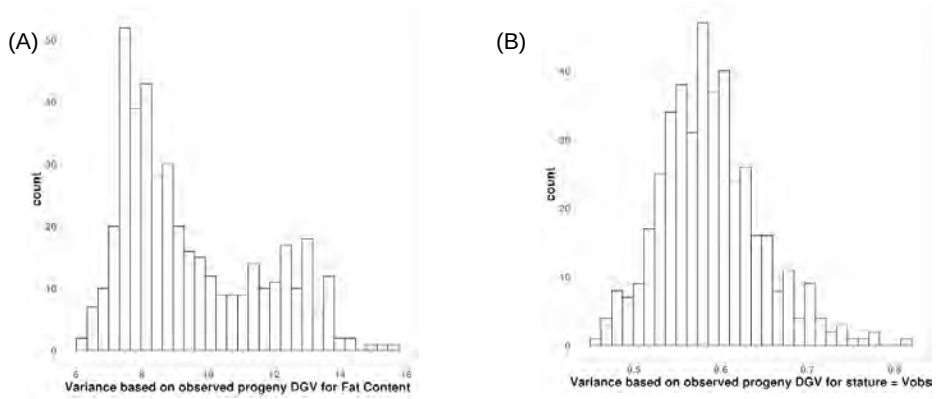


Figure 1. Distribution of progeny DGV variances (V_{obs}) for fat content (A) and stature (B).

The correlation between V_{theo} and the within bull variance of phenotypes was quite low with an average of 0.25 across traits. On the youngest animal (58 bulls born after 2015), correlations were a bit higher with an average of 0.33. Except for a high correlation for fat content (0.85), correlations were moderate for heritable traits (0.44 for protein yield or 0.38 for body depth) and as low as 0.14 for low heritability traits such as rear udder height. All correlations are presented on Table 1.

Discussion

This study in Holstein cattle breed investigates the opportunity to predict progeny variability from genotypes and SNP effects. We observed large variation in progeny variability among bulls supporting the growing interest for an indicator of gametic variability, especially in the choice of bull sires and dams to maximize the probability to conceive top ranking bulls. Validation on real data showed that the proposed theoretical indicator performed quite well to predict the genetic variability of progeny but was not accurate enough to predict the phenotypic performance variability, highly dependent on environmental factors. Computation of the theoretical predictor with Equation 1 is fast, meaning that it can be obtained routinely and used to select animals for the breeding scheme. In this perspective, development of specific selection criteria such as those developed by Bijma *et al.* (2020) is expected on a routine basis. In spite of its cost of computation, the estimation of gametic variance by simulation of progeny may also be interesting especially in complex situations, mainly in case of non-additive determinism.

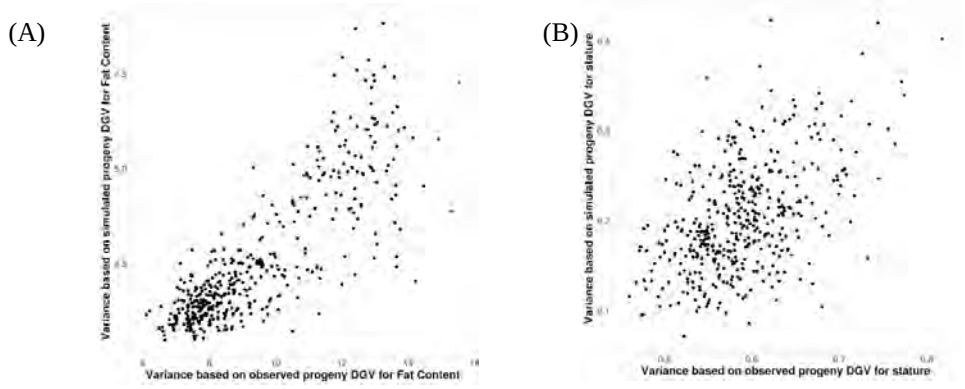


Figure 2. Relationship between DGV within sire variances for simulated and observed progeny, for fat content (A) and stature (B).

Acknowledgements

This research has received funding from European Union's Horizon 2020 research and innovation program – GenTORE – under grant agreement no. 727213.

References

Bijma P., Wientjes Y.C.J., and Calus M.P.L (2020) *Genetics.* 214 (1). 91-107. https://doi.org/10.1534/genetics.119.302643

Boichard D., Guillaume F., Baur A., Croiseau P., Rossignol M. N. *et al.* (2012) *Anim Prod Sci* 52. 115-120. https://doi.org/10.1071/AN11119

Santos D.J.A., Cole J.B., Lawlor T.J.Jr., VanRaden P.M., Tonhati H., *et al.* (2019). J Dairy Sci 102(6) 5279-5294. https://doi.org/10.3168/jds.2018-15971

Sargolzaei M., Chesnais J., and Schenkel F. (2014) *BMC Genomics* 15(1):478. http://doi.org/10.1186/1471-2164-15-478

Segelke D., Reinhardt F., Liu Z. and Thaller G. (2014) *Genet Sel Evol* 46. 42. https://doi.org/10.1186/1297-9686-46-42

442. Implementing genomic management of diversity and selection in French dairy goats

I. Palhière[1*], V. Gousseau[2] and J.J. Colleau[3]

[1]GenPhySE, Université de Toulouse, INRAE, INPT, ENVT, 31326 Castanet Tolosan, France; [2]Capgenes, 2135 Rte de Chauvigny, 86550 Mignaloux-Beauvoir, France; [3]GABI, Université de Paris Saclay, INRAE, AgroParisTech, 78350 Jouy-en-Josas, France; isabelle.palhiere@inrae.fr

Abstract

Optimal contribution selection (OCS) is a selection method that optimises genetic gain and diversity simultaneously. It has been implemented in French dairy goat breeding programmes since 2006. When genomic evaluation was applied in 2018, genomic OCS replaced pedigree OCS, accounting for genomic coancestries and differences of GEBV accuracies between progeny tested sires and untested sires. The genomic OCS efficiency was assessed based on the actual matings scheduled in 2021. Compared to candidates, selected sires had a higher genetic level despite a lower coancestry. Contribution of untested sires was only 30-35%. The future offspring showed a high genetic level and a genomic inbreeding from 3.0 to 3.6%. The survey of the AI bucks born between 2000 and 2022 confirmed the long-term efficiency of OCS: the annual genetic gain was about 0.17 genetic standard deviation on the total merit index whereas the annual rate of inbreeding only ranged between 0.05 and 0.07%.

Introduction

The implementation of genomic selection in many species generates additional genetic gain but raises questions about its impact on genetic diversity. For a large international breed like the Holstein breed, many studies (Doekes *et al.*, 2018; Forutan *et al.*, 2018; Doublet *et al.*, 2019) showed that genetic diversity was dramatically reduced once genomic selection was introduced. OCS is a powerful method (Meuwissen, 1997) maximising genetic gain at restricted rate of inbreeding. Colleau *et al.* (2004) developed an alternative method where the coancestry rate was minimized at fixed genetic gain. In France, two main dairy goat breeds are raised: the Alpine and the Saanen breeds. A total of 600 breeders and 175,000 females is involved in both breeding schemes. Since 2018, genomic selection has been implemented with two main changes: breeding values (GEBV) have been estimated with single step GBLUP (Legarra *et al.*, 2009; Carillier *et al.*, 2014) and 5 month-old AI bucks have been eligible for selection based on their GEBV. Their GEBV accuracy is not high enough to stop progeny testing, unlike dairy cattle. The management of genetic diversity has always been a major concern for the goat breeding schemes. In 2006, pedigree OCS (Colleau *et al.*, 2004) was introduced for producing young AI bucks, born from elite sires and dams. From the start of genomic evaluation in 2018, GOCS has been implemented, accounting not only for genomic coancestries, but also for GEBV accuracies. It was due to the very lower accuracy of not yet progeny-tested sires (NT) compared to progeny tested sires (PT). The goal of this paper was first to present the principles and results of the GOCS actually implemented in the two French dairy goats in 2021. The evolution of the genetic gain and the genetic diversity of AI bucks born between 2000 and 2020 was also analysed, in order to evaluate the long-term impact of OCS.

Materials & methods

Animals involved. GOCS is applied once a year for both breeds, in order to produce the new generation of young AI bucks. The result is a list of matings between elite sires and dams. To perform the optimisation, the breeding organisation provided different datasets and information. The list of candidates dealt with ungenotyped dams (871 Alpine / 822 Saanen) and genotyped elite sires (151 Alpine/ 76 Saanen). Because GOCS should produce minimal coancestry within the pool of males likely to perform AI in the future, the

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_442, © I. Palhière *et al.* 2022

four last cohorts of young bucks (4C) born form OCS the four last years, were also considered (415 Alpine / 302 Saanen). GEBVs for the total merit index (ICC) and 16 other GEBVs were included, in reference with minimum thresholds in order to compute the number of defects for each mating. Genomic data corresponded to 2,901 Alpine and 1,980 Saanen 54K Illumina genotypes for the 46,707 SNP used in the official genomic evaluation. The overall pedigree included 41,557 Alpine and 34,678 Saanen animals.

The survey of AI bucks included 959 AI progeny-tested bucks in Alpine breed and 694 in Saanen breed, born between 2000 and 2020. For cohorts 2021 and 2022, progeny testing was not completed so that their statistics were those of the corresponding mating cohorts (about 700-750 per breed and per year). Trends in genetic gain, inbreeding and coancestry were analysed. Genetic gain was measured for ICC, expressed in σA. GEBV came from the French official genetic evaluation, performed in June 2021.

Genomic coancestry and optimisation methods. Genomic coancestries between pairs of genotyped individuals were computed according to Van Raden's method (2008). Otherwise, they were estimated by the method described by Colleau *et al.* (2017). Finally, SNP genomic coancestries and inbreeding were expressed in a converted scale, established once for all in 2018, based on data not influenced by genomic selection, as proposed by the same authors.

The optimal contributions of candidates and the corresponding mating design were computed based on the simulated annealing Monte-Carlo method (Metropolis *et al.* 1953, Kirkpatrick *et al.,* 1983). Constraints were accounted for by a constrained simulated annealing (CSA), according to the method proposed by Colleau *et al.* (2009), Chapuis *et al.* (2016). For computing optimal contributions, the corresponding CSA considered two major nuisances: the average coancestry coefficient (in a rolling window including 4C and future progeny) and the average inaccuracy (1-CD) of sires. These nuisances were expressed on a 0-1 scale from the minimum to the maximum values. Extreme values were given by dedicated CSA procedures, of course accounting for the constraint imposed for the average ICC of future progeny. Afterwards, scaled nuisances were minimized, while constrained to be equal, because considered of equal importance by the breeding organisation. The mating design basically minimized a linear function involving inbreeding coefficient and number of defects, standardized in reference to the whole variability of possible matings. A special sampling procedure was introduced to avoid extreme percentages of NT sires in the farms holding the selected dams and, thus, to increase acceptability by breeders.

Results

GOCS selected 60 sires (40% of candidates) in the Alpine breed and 37 sires (48%) in Saanen breed. Among them, about 60% were NT sires but their contribution to matings were only 39% in Alpine breed and 34% in Saanen breed. Each PT sire contributed to 19 matings, on average, in the Alpine breed and 33 in the Saanen breed, whereas NT sires contributed to 8 and 10 matings, respectively. The difference of contribution between PT and NT sires was explained by the gap of GEBV accuracy: NT sires had a much lower CD than PT sires (0.56 vs 0.90 on average). In contrast, candidate NT sires showed a genetic superiority over candidate PT sires of 1.12 ICC units (0.36 σA) in the Alpine breed and 0.70 ICC units (0.22) in the Saanen breed. Globally, GOCS selected sires (PT+NT) with a selection differential of 0.29 ICC unit (0.09 σA) in Alpine breed and 0.70 (0.22) in Saanen breed. In comparison with candidates, the average coancestry of selected sires with 4C was decreased by 15% in the Alpine breed and 9% in the Saanen breed. Table 1 shows the results for the 760 matings in the Alpine breed and 700 in the Saanen breeds. The mean ICC was fixed by the breeding organisation. The average genomic coancestry of offspring with 4C was 4.70% in the Alpine breed and 5.35% in the Saanen breed. The average genomic inbreeding reached 2.96% in the Alpine breed and 3.58% in the Saanen breed. Compared to all possible matings between selected sires and dams (before mating optimisation), the average inbreeding was reduced of 40% in the Alpine breed and 35% in

Table 1. Genetic gain, genetic diversity and number of defects of offspring.[1]

Breed	Alpine	Saanen
ICC	7.00 (0.65 / 5.57-9.93)	6.90 (0.58 / 5.58-9.04)
Inbreeding (%)	2.96 (0.57 / 1.16-4.51)	3.58 (0.63 / 1.90-5.57)
Coancestry[2] (%)	4.70	5.35
# defects	0.19 (0.43 / 0-3)	0.09 (0.31 / 0-2)

[1] Mean (standard deviation / min-max.

[2] Average genomic coancestry coefficient between offspring and 4C.

the Saanen breed. The average number of defects was very low, ranging between 0 and 3 depending on the mating. Consequently, offspring had balanced GEBV profile for all the traits.

The annual rate of inbreeding for AI bucks born between 2000 and 2022 (Figure 1) was 0.07% in the Alpine breed and 0.05% in the Saanen breed. It corresponded to an increase of 0.26% per generation in the Alpine breed and 0.21% in the Saanen breed. Before the implementation of OCS (2000-2006) the rate of inbreeding reached 0.30%/year in the Alpine breed and 0.22%/year in the Saanen breed, whereas after the implementation of POCS (2007-2018) it decreased to 0.14%/year and 0.17%/year, respectively. When focusing on the GOCS period (2019-2022), we observed that the rate of inbreeding dramatically decreased: -0.08%/year in Alpine and 0.07%/year in Saanen. The trend in annual coancestry rate was the same, decreasing from 0.01 to -0.19% in the Alpine breed and from 0.01 to -0.06% in the Saanen breed, between POCS and GOCS. The annual genetic gain on ICC was 0.165 σA in the Alpine breed and 0.169 σA in the Saanen breed during the 2000-2022 period. No major difference was observed before or after POCS. But, under the GOCS period, the preliminary results showed that the annual genetic gain substantially increased by 30% (Alpine) to 46% (Saanen).

Discussion

The implementation of genomic evaluation in dairy goat breeding programmes led to modifying OCS. Accounting for genome-based coancestries was necessary, as demonstrated by Sonesson *et al.* (2012). Otherwise, the inbreeding rate would have been underestimated. Including NT sires had contradictory

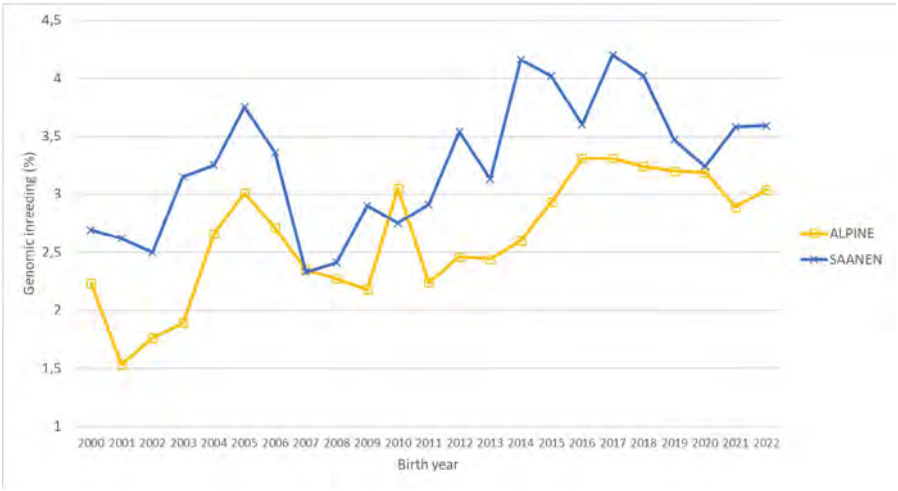


Figure 1. Genomic inbreeding trend for AI bucks between 2000 and 2022.

effects because the value of their genetic superiority for the breeding scheme was tempered by a low at-birth-GEBV accuracy. Currently, goat breeders have not been yet ready to use a high percentage of NT sires. But, the breeding organisation plans in the near future to use 60% of NT sires, still under the current GEBV accuracy, in order to reduce the generation interval. This could generate more genetic gain without decreasing genetic diversity, thanks to GOCS.

Our first results illustrated the efficiency of GOCS that combined moderate inbreeding and balanced GEBV profile of offspring, customized allocation for each NT sire while respecting desired genetic gain. Consequently, the rate of inbreeding for AI bucks remained reasonable, and much lower than in dairy cattle (Doekes *et al.*, 2018; Forutan *et al.*, 2018; Doublet *et al.*, 2019).

References

Carillier C., Larroque H., Robert-Granié C. (2014). Genet Sel Evol 46, 67. https://doi.org/10.1186/s12711-014-0067-3

Chapuis H., Pincent C., Colleau J.J. (2016). J. Anim. Breed. Genet. 133: 1-10. https://doi.org/10.1111/jbg.12178

Colleau J.J., Moureaux S., Briend M., Béchu J. (2004). Genet Sel Evol 36, 373. https://doi.org/10.1186/1297-9686-36-4-373

Colleau J.J, Tual K., de Preaumont H., Regaldo D. (2009). Genet Sel Evol 41, 7. https://doi.org/10.1186/1297-9686-41-7

Colleau J.J., Palhière I., Rodriguez Ramilo S.T., and Legarra A., (2017). Genet Sel Evol 49, 87. https://doi.org/10.1186/s12711-017-0363-9

Doekes H.P., Veerkamp R.F., Bijma P., Hiemstra S.J., Windig J.J. (2018). Genet Sel Evol 50, 15. https://doi.org/10.1186/s12711-018-0385-y

Doublet A.C., Croiseau P., Fritz S., Michenet A., Hozé C. *et al.* (2019). Genet Sel Evol 51, 52. https://doi.org/10.1186/s12711-019-0495-1

Forutan M., Ansari Mahyari S., Baes C., Melzer N., Schenkel F.S., *et al.* (2018). BMC Genomics, 19:98. https://doi.org/10.1186/s12864-018-4453-z

Kirkpatrick S., Gelatt C.D., Vecchi M.P. (1983). Science, 220/671-680. https://doi.org/10.1126/science.220.4598.671

Legarra A., Aguilar I., and Misztal I., (2009). J. Dairy Sci. 92:4656-63. https://doi.org/10.3168/jds.2009-2061

Metropolis N., Rosenbluth A.W., Rosenbluth M.N., Teller A.H., Teller E., (1953). J.Chem.Phys., 21, 1087-1092. https://doi.org/10.1063/1.1699114

Meuwissen T.H.E, (1997). J. Anim. Sci., 75, 934-940. https://doi.org/10.2527/1997.754934x

Sonesson A.K., Woolliams J.A., and Meuwissen T.H.E., (2012). Genet. Sel. Evol.,44:27, https://doi.org/10.1186/1297-9686-44-27

VanRaden P.M. (2008). J. Dairy Sci. 91: 4414-4423. https://doi.org/10.3168/jds.2007-0980

443. Partition in long-term contributions of the genetic trend of Basco-Béarnaise dairy sheep

S. Antonios[1*], A. Legarra[1], J.M. Astruc[2], S.T. Rodríguez-Ramilo[1] and Z.G. Vitezica[1]

[1]GenPhySE, INRAE, INPT, ENVT, 31326 Castanet Tolosan, France; [2]Institut de l'Elevage, 149 rue de Bercy, 75595 Paris, France; simona.antonios@inrae.fr

Abstract

The objective of this work was to study long-term contributions to genetic gain in Basco-Béarnaise dairy sheep breed. From estimated breeding values, contributions from males and females and from five categories were analysed. The AlphaPart R package was used to partition the genetic gain into ancestral Mendelian samplings by sex and by category. The categories were defined as: (1) best males after the progeny testing; (2) males discarded after progeny testing; (3) natural mating males; (4) dams of males; and (5) dams of females. The results showed that the selection scheme works by capturing at each generation the originality (Mendelian samplings) of the animals, and that females contribute largely to the final genetic gain by their originality (or genetic variability).

Introduction

The breeding value of an individual can be expressed as the average of its parental breeding values plus a Mendelian sampling deviation. This last term (the Mendelian sampling) captures the 'originality' of the individual with respect to their parents. In this way, the breeding value of an individual can be decomposed as a sum of Mendelian samplings of all its ancestors.

Analysing long-term contributions of individual Mendelian samplings to the successive generations of the breed allows to understand the different gene flow among, and contributions of, groups and categories of animals (Woolliams *et al.*, 1999; García-Cortés *et al.*, 2008). A retrospective analysis of these long-term contributions in the selection process allows to know the dynamic of the contributions, the ultimate sources of 'originalities' and to predict the future evolution of inbreeding.

The objective of this work was to study long-term contributions to characterize the sources of genetic progress in Basco-Béarnaise (BB) dairy sheep.

Materials & methods

In matrix notation, the vector of breeding values (a) can be written as a linear combination of Mendelian sampling terms and their ancestors as $a = Tw$, where T describes flow of genes through pedigree (Woolliams *et al.*, 1999) and w is the vector of Mendelian samplings. This equation also holds for unbiased predictors as BLUP, $\hat{a} = T\hat{w}$ (García-Cortés *et al.*, 2008), thus Mendelian samplings can be estimated as $\hat{w} = T^{-1}\hat{a}$, where $\hat{a}$ contains the Estimated Breeding Values (EBV).

Our study considered the BB breeding scheme until 2016, time at which genomic selection was introduced. Based on the 2021 official national genetic evaluation (single step GBLUP, Genomic Best Linear Unbiased Prediction), EBVs for milk yield of 138,572 animals born between 1985 and 2016 (136,234 females and 2,338 males) and their pedigree were used in the analyses.

We decomposed the total genetic gain into contributions from males and females, or into contributions from categories defined by the selection pathways. In the BB selection scheme, there are females (recorded for milk yield), Artificial Insemination (AI) males, and Natural Mating (NM) males. Males (AI or MN) are

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_443, © S. Antonios *et al.* 2022

always sons of AI males, while females can be progeny of AI or MN males. AI males produce ~70% of the females, with the remaining ~30% coming from NM males. AI males were selected (until 2016) by progeny testing. Accordingly, we defined 5 categories of animals: (1) AI males (AI_males), the best males kept after progeny testing; (2) discarded AI males (discarded_AI_males), the males discarded after progeny testing; (3) NM males (NM_males); (4) dams of males (DM), elite females chosen to generate AI males; (5) dams of females (DF), who were not chosen as elite females and are considered to be less good than DM.

From EBVs and the pedigree of the breed that shows the flow of genes, the Mendelian samplings terms were computed. The AlphaPart R package (Obsteter *et al.*, 2021) was used to partition the genetic gain into Mendelian sampling trends by sex and by the categories defined above. All results are based on data obtained before the implementation of genomic selection; since then, partitioning probably differs.

Results

Between 1985 and 2016, the total genetic progress was 3.1 l/year, adding up to more that 100 liters. Decomposition of genetic trends into contributions from males versus females allowed the monitoring of their contribution.

Figure 1 shows the decomposition of genetic trend into contributions from Mendelian sampling terms of males and females. The slope indicates that Mendelian sampling terms of dams contribute highly to response, for instance, through their selected sons. The contribution on average was 79 and 21% for females and males, respectively.

Figure 2 presents the average Mendelian samplings gained (selected for) by year. The two categories of dams (DF and DM) and AI males show a progress between 1 and 2 liters per year. The average Mendelian sampling varies between -1 and 1 liter for discarded AI males, and it is around zero for NM males. The discarded AI males are not 'good enough' after progeny testing, and they stop contributing their genes. As for NM males, after initial selection based on Parent Average, NM males are not further selected, and so their Mendelian sampling are unselected for upcoming generations. There is some erratic behaviour of average Mendelian samplings in the last years, particularly for AI males. This happens, first, because accurate estimates of Mendelian Samplings occur several years after the animal has been selected (as sire or dam), when the animal has a large number of progeny or even grand-progeny; and second, because the long-term contributions stabilize only after some generations.

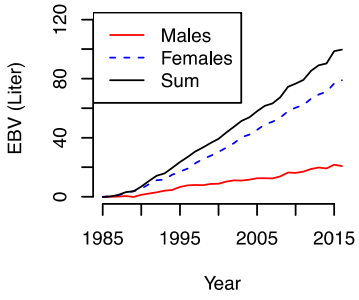


Figure 1. Partitioning the genetic gain into contributions from males and females.

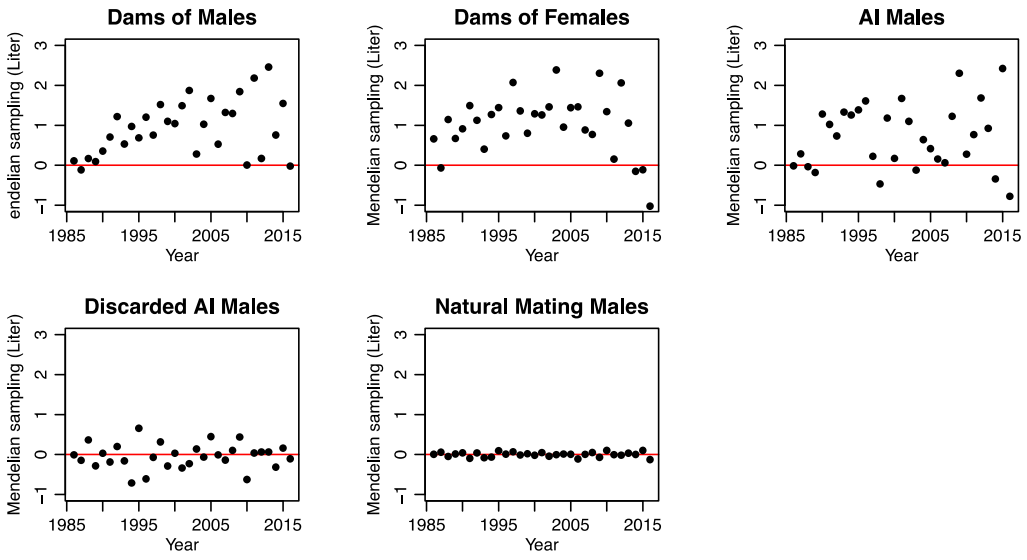


Figure 2. Average Mendelian sampling for each year.

Discussion

To improve a population, we need certain animals to be better than their parents. For this, the originality of the animals with respect to their parents, i.e. the Mendelian sampling, has to be captured. The latter is estimated from different sources of information: (1) own phenotype (milk yield for females, collected in in the selection scheme); (2) phenotypes of the progeny of males (progeny test, i.e. in the AI rams in this scheme); (3) by genomic evaluation (not studied in this work as genomic selection was not yet implemented in 2016). Although the precision in capturing Mendelian sampling is higher in males (by progeny testing), the final genetic progress is largely given by the originality (or genetic variability) of the females, even if their prediction (by own phenotype) is less accurate and their progeny number is smaller. Even if a female has few offspring, a selected female can be the mother or grandmother of an AI male, and consequently have a large number of offspring. Moreover, since the number of eligible females is very high, the selection intensity to be a 'dam of a male' can be very high.

Therefore, fine partitioning in Mendelian sampling trends demonstrates that females are the main reservoir of genetic variability for selection. Still, AI males are crucially important because they allow the spread of best genetics in the population.

Acknowledgements

This study has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No. 772787 (SMARTER), and project ARDI (grant agreement EFA 208/16) from POCTEFA funds.

References

García-Cortés L.A., Martínez-Avila J.C., and Toro M.A. (2008) Animal 2:821-824. https://doi.org/10.1017/S175173110800205X

Obšteter J., Holl J., Hickey J.M., and Gorjanc G. (2021) Genet Sel Evol 53:30. https://doi.org/10.1186/s12711-021-00600-x

Woolliams J.A., Bijma P., and Villanueva B. (1999) Genetics 153(2):1009–1020. https://doi.org/10.1093/genetics/153.2.1009

444. Introgression of prolificacy gene in Tunisian Barbarine sheep through marker-assisted selection scheme

E. Arari[1,2], I. Baazaoui[1,3], N. Letayef[1,2], M. Lahmar[4], M. Hemdène[4], A. Haile[5], M. Rekik[6] and S. Bedhiaf-Romdhani[1*]

[1]Tunisian National Institute of Agronomic Research, Hédi Karray, 1004 Menzah 1, Tunisia; [2]National Agronomic Institute of Tunisia, 43, Avenue Charles Nicolle, 1082 Tunis, Mahrajène, Tunisia; [3]Faculty of Sciences of Bizerte, Jarzouna, 7021 Jarzouna Bizerte, Tunisia; [4]Livestock and Pastures Office, 30, Rue Alain Savary, 1002 Tunis Belvédère, Tunisia; [5]International Center for Agricultural Research in the Dry Areas, ICARDA, Addis Ababa, Ethiopia; [6]International Center for Agricultural Research in the Dry Areas, 8, Complex No 8, Yousef Al Sukkar St, Amman, Jordan; bedhiaf.sonia@gmail.com

Abstract

Sheep litter size is an important economical and reproduction trait. Tunisian Barbarine sheep 'W-INRAT' provide carrier FecX^{Bar} animals, referring to a mutation in the *BMP15* gene. This mutation causes increased ovulation rate in heterozygous ewes. The aim of this work was to introgress the FecX^{Bar} mutation in two conventional multipliers flocks through the development of a marker-assisted selection scheme. The dissemination of the mutation was conducted by artificial insemination of total 350 conventional ewes during two successive years (2019-2020) by four prolific 'W-INRAT' rams (FecX^{Bar}/Y). The F1 offspring were genotyped for FecX^{Bar} mutation by allele-specific PCR. For both years, female lambs (n=92) were heterozygous. They were selected and kept as future prolific breeders. Their mattings with conventional rams will serve to increase prolificacy and produce prolific males. This study shows the importance of genotyping at an early age for the selection and management of mattings for an efficient dissemination of the prolificacy mutation.

Introduction

Reproductive ability has important impacts on the profitability of sheep farming (Abdoli *et al.,* 2016). Reproduction traits such as ovulation rate and litter size are genetically affected by causative mutations in major genes called fecundity (Fec) genes (Davis, 2004). Tunisian Barbarine sheep are a fat-tailed breed found in all production systems and are distinguished by adaptations to environmental conditions (Ben Salem *et al.,* 2011). Barbarine sheep is well known for providing a major source of meat in Tunisia with an acceptable reproductive performance; a high fertility rate (85.2%) but a modest prolificacy mean value of 106.9% with a productivity that does not exceed 0.8 lamb/ewe/year (Atti *et al.,* 2004; Brahmi *et al.,* 2016). The reproduction of Barbarine ewes is, as in most sheep breeds, seasonal and photoperiodically controlled. However, like most breeds from the Mediterranean area, the effect of photoperiod on the seasonality of reproduction is limited, that makes traditional spring (April-May) a mating season (i.e. out of season) (Atti *et al.,* 2004). Furthermore, considering the reproductive ability of the Barbarine, genetic strategies have currently focused on reproduction traits to improve the income of sheep farmers. Thus, a prolificacy-based selection program was implemented since 1979 by the Tunisian National Institute for Agronomic Research (INRAT) in the experimental Oueslatia station Kairouan by screening prolific ewes. The selection process yielded the prolific 'W-INRAT' strain with litter size (LS) of 1.6 and a fertility rate ranging between 82 and 91% (Bedhiaf *et al.,* 2016). Therefore, a novel mutation 'FecX^{Bar}' in the Bone morphogenetic protein 15 (*BMP15*) gene associated with prolificacy was reported for the 'W-INRAT' strain with an additive effect on LS in heterozygous ewes (Lassoued *et al.,* 2017). The FecX^{Bar} mutation is a composite polymorphism associating a single nucleotide substitution (c.301G > T), a 3 bp deletion (c.302_304delCTA) and a C insertion (c.310insC) in the ovine BMP15 cDNA leading to a frame shift at protein position 101. Thus, heterozygous ewes had an increased both ovulation rate (OR) by 0.7 ova and LS by 0.3 lambs (P=0.08)

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_444, © E. Arari *et al.* 2022

but homozygous females carrying $FecX^{Bar}$ exhibited streaky ovaries with a blockade at the primary stage of folliculogenesis. Moreover, inclusion of genetic information of the *BMP15* gene associated with reproductive ability could efficiently enhance the selection response. This is why the 'W' strain selected from INRAT could strengthen the national selection nucleus by providing animals of high genetic potential for multiplier flocks (Bedhiaf *et al.,* 2016). The aim of the present work was to increase prolificacy by creating prolific carrier females to produce carrier males to disseminate the mutation.

Materials & methods

Farm and animal management. The present study was carried out for 2 years (2019 and 2020) in two Black-headed Barbarine sheep farms in the north and north-east of Tunisia (OEP and OTD). The experiment was conducted during the spring (April-May) using a total of 334 and 394 ewes respectively in 2019 and 2020 aged between 2 to 4 years old, with a body condition score of 2.5-3.0. All animals were kept in the same management system. The main source of feed was pasture supplemented with 0.5 to 1 kg hay and 200-300 g concentrate feeds per head per day during the breeding period. In addition, four prolific rams were selected from 'W-INRAT' strain from Ousslatia farm, to use them in FecXBar diffusion.

Selection of animals and marker-assisted scheme. A marker-assisted scheme (MAS) was developed to disseminate the FecXBar mutation from the nucleus (prolific W-INRAT strain) to multipliers flocks in order to increase prolificacy in conventional Barbarine flocks. The FecXBar allele diffusion plan aimed to provide carrier females and to avoid the production of homozygous sterile females. An initial step was established by selecting four carrier rams from 'W-INRAT' strain to inseminate two conventional ewes' flocks from two public farms (OTD and OEP) to obtain heterozygous females among the F1 progeny. The second level consists to inseminate those heterozygous selected females with conventional rams to increase the number of effective carriers at F2 progeny. Diffusion of prolificacy mutation will be realized by male carriers from F2 at commercial farms by adopting this MAS (Figure 1).

Sample collection and allele-specific PCR (AS-PCR) Jugular vein blood samples (5 ml per animal) were obtained from F1 offspring (n=143 in 2019 and n=151 in 2020) during the two years. Genomic DNA was isolated from frozen blood samples using the standard phenol–chloroform–isoamyl alcohol extraction method. The quality, purity and concentration of isolated genomic DNA were evaluated using standard techniques (Sambrook and Russell, 2001). Two independent AS PCRs were used to amplify a fragment of 445 (or 440) bp from *BMP15* exon 1 for each individual, using the same forward primer (F) and one of the two allele specific reverse primers, (R+) for the $FecX^{+}$ wild-type allele and (R-) for the $FecX^{Bar}$ mutated allele. The PCR products were examined on a 1% agarose gel after 35 cycles of amplification (94, 62, 72 °C/30s). Heterozygous females' lams tested positive for both PCRs, but male lambs tested positive for only the wild allele.

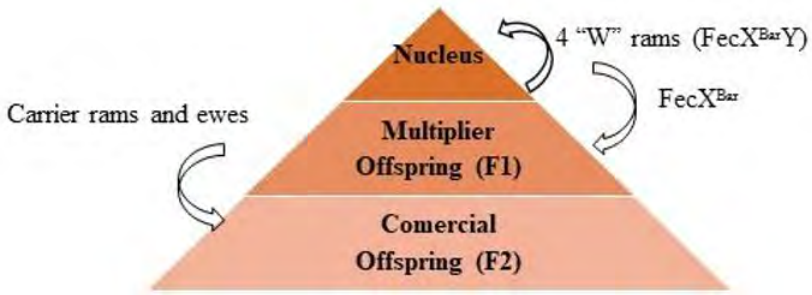


Figure 1. Breeding strategy for Barbarine carrier FecXBar sheep/ Barbarine breeding program.

Table 1. Oligonucleotide primers designed to amplify specific loci of BMP15 gene.

Gene	Region	Oligonucleotide sequence	Product size (bp)	Annealing temperature (°C)
BMP15	Exon 1	F: 5'-TTCCTTGCCCTATCCTTTGTG-3' R+: 5'-GAGGCCTTGCTACACTAGCC-3' R+: 5'-TGAGAGGCCTTGGCTACACA-3'	445	60

Results

As shown in Figure 2, the genotyping results gave two electrophoretic profiles; wild individuals (FR+) and heterozygous individuals (FR+/FR-) (Figure 2). In this context, during the two years, all female lambs produced by artificial insemination (n=92) were heterozygous ($FecX^{Bar+}$). They were selected as future breeders serving to both increase litter size and produce prolific males at F2 progeny. These males (F2) will be used to disseminate the prolificacy mutation at a national level in order to improve commercial flocks' prolificacy.

Discussion

F1 offspring (n=143 in 2019 and n=151 in 2020) during two years resulting from the artificial insemination of 350 ewes by four rams carrying the FecXBar mutation were screened for $FecX^{Bar}$ mutation by AS PCR. Knowledge that these mutations are present in Barbarine sheep breeds is helping in designing and developing breeding strategy that maximize the benefits by increased prolificacy in this breed and their crosses. Artificial insemination programme has been used to successfully introgress the fecundity genes into other breeds in several countries, and these artificial breeding technologies have been particularly useful where only small numbers of progeny tested individuals have been available (Davis, 2004). All offspring female were heterozygous. The findings of the present study are in line with those of Lahoz (2011) who demonstrated that identification $FecX^{R}$ mutation is a reliable tool to identify carrier ewes for an efficient management of the breeding program. In Barbarine ewes, the increase in OR is mainly due to an increase in double ovulations, making the use of the $FecX^{Bar}$ polymorphism suitable for application in commercial farms (Lassoued *et al.*,2017). Selection of heterozygous F1 females as a first step presents an advantage to increase litter size and produce prolific males. These recent results lead to manage the breeding program and follow breeding scheme (MAS) to control of prolificacy rate.

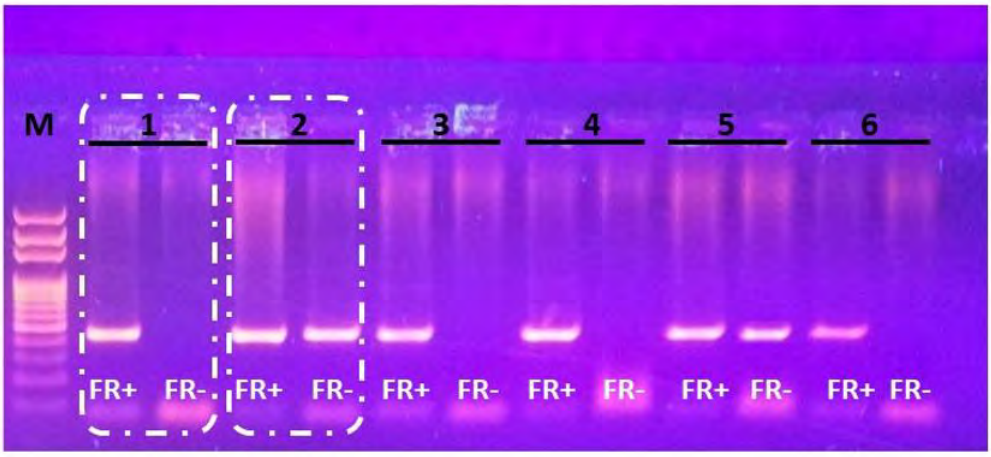


Figure 2. Allele-specific PCR amplification at $FecX^{Bar}$ and $FecX^{+}$ in *BMP15* gene. M: 450 bp ladder DNA marker. FR+ = wild type allele, FR- = mutant allele. Sample 2 and 5: mutant type. Sample 1,3,4 and 6: Wild type genotype.

References

Abdoli R., Zamani P., Mirhoseini SZ, Ghavi Hossein-Zadeh N. *et al.* (2016) Reproduction in Domestic Animals 51: 631–637. https://doi.org/10.1111/rda.12733

Atti N., Bocquier F. and Khaldi G. (2004) Anim. Res. 53:165–176. https://doi.org/10.1051/animres:2004012

Bedhiaf-Romdhani S., Djemali M., Ben Gara A., Rekik B., Ben Hamouda M., *et al.* (2016) Annales de l'INRAT 89:154-157. https://doi.org/10.12816/0028727

Ben Salem H., Lassoued N. and Rekik M. (2011) Tropical Animal Health and Production 43:1357–1370. https://doi.org/10.1007/s11250-011-9863-8

Brahmi A., Bouallègue M.A., Bouzaiène H. and Khaldi G. (2011) Proc. of 100th Options Méditerranéennes Zaragoza, Spain.

Davis G.H. (2004) Animal Reproduction Science 82–83:247–253. https://doi.org/10.1016/j.anireprosci.2004.04.001

Lahoz B., Alabart J. L., Jurado J. J., Calvo J. H., Martínez-Royo A., *et al.* (2011) J. Anim. Sci. 89:3522–3530. https://doi.org/10.2527/jas.2010-3828

Lassoued N., Benkhlil Z., Woloszyn F., Rejeb A., Aouina M., *et al.* (2017) BMC Genetics (2017) 18:43. https://doi.org/10.1186/s12863-017-0510-x

Sambrook J. and Russell DW. (2001) Molecular Cloning: A Laboratory Manual. Cold Spring Harbor, NY, USA.

445. Simulation of animal crossbreeding schemes using ADAM software

H. Liu[1,2*], H.M. Nielsen[2], M. Kargo[2], J.R. Thomasen[3], P. Berg[2,4], M. Slagboom[2], A.C. Sørensen[5] and V. Milkevych[2]

[1]SEGES, Agro Food Park 15, 8200 Aarhus N, Denmark; [2]Center for Quantitative Genetics and Genomics, Aarhus University, Blichers Allé 20, 8830 Tjele, Denmark; [3]VikingGenetics, Ebeltoftvej 16, 8960 Randers SØ, Denmark; [4]Norwegian University of Life Sciences, Universitetstunet 3, 1433 Ås, Norway; [5]The Danish Agriculture and Food Council, Agro Food Park 15, 8200 Aarhus N, Denmark; hliu@seges.dk

Abstract

ADAM software is a stochastic simulation tool intended to model a variety of selective breeding schemes for animals and plants. ADAM can be used to generate genetic data to test statistical tools and the implemented algorithms can evaluate breeding schemes by tracing genetic change under different selective breeding scenarios. In practice, crossbreeding has been used extensively in beef cattle, pig, and poultry production systems, and the use of crossbreeding is increasing in dairy cattle breeding system due to changes in the dairy market. The aim of this paper was to present a new crossbreeding feature of the ADAM software. The software was updated to include multiple purebred populations and crossbred populations in animals with different genetic backgrounds considering both additive and dominance effects. Our simulated results showed that, by using the new features in ADAM, breeding organizations can optimize their design of crossbreeding schemes and, thereby, optimize profitability.

Introduction

The well-known effect of increased performance of crossbreds compared to purebreds has long been used systematically in livestock species such as pig and poultry (Therkildsen *et al.* 2021; Chen *et al.* 2021; Han *et al.* 2020). In dairy cattle, farmers have started to use crossbreeding to utilize the increased performance of crossbreds due to heterosis (Clasen *et al.* 2017). However, tools for the use of crossbreeding are lacking; including tools to predict genetic merit of crossbreds, tools to manage breeding of several breeds simultaneously, and decision tools for farmers to optimally use crossbreeding in a herd.

One of the major decisions to make when managing crossbreeding is which breeding scheme is optimal. Therefore, efficient computational tools that allow stochastic simulation of a variety of breeding schemes are of high importance. Among the existing software (Faux *et al.* 2016; Pook *et al.* 2020) ADAM is a well-established tool intended to model a variety of selective breeding schemes for animals and plants (Pedersen *et al.* 2009; Liu *et al.* 2019). The software simulates populations and traces genetic changes in these populations under different scenarios. ADAM accounts for different population structures and genetic models. It can be used to test different phenotyping/genotyping strategies, estimate breeding values, evaluate different selection and mating strategies and generate genetic data to test statistical tools. It is a compiled software written in FORTRAN.

The purpose of this paper is to present the novel features of the ADAM software dedicated to simulation of multiple populations of (possibly) different genetic backgrounds including dominance genetic effects, demonstrated in dairy cattle crossbreeding scheme. The implemented features naturally extend the existing ADAM functionality (Figure 1) to the case of multiple interacting populations and allow efficient evaluation of crossbreeding effects at genetic and phenotypic levels.

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_445, © H. Liu *et al.* 2022

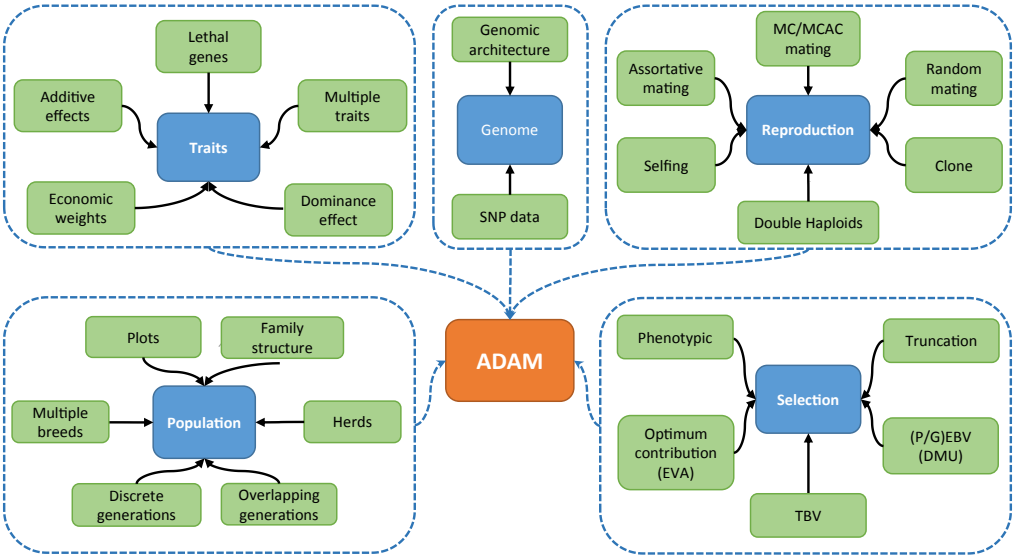


Figure 1. General overview of the ADAM software functionality. MC is minimum co-ancestry; MCAC is mating animals by minimizing the covariance between ancestral contributions; TBV is true breeding values.

Materials & methods

Multiple interacting populations. The model of interacting populations, considered in the context of crossbreed selection, is represented here in terms of evolutionary operators following the approach from Paixão *et al.* (2015). Given a finite population $P(t) = (G^n, \varphi, S_{(\iota,\lambda)})$ of size n at time t characterized by a finite collection $G^n = \{g_i\}$ of genotypes, where $g_i, i = \overline{1,n}$, genotype-phenotype mapping φ, finite collection $S_{(\iota,\lambda)} = \{s_j\}$ of selection candidates $s_j, j = \overline{1,m}$, where i are animals selected for mating and λ are their offspring. Evolution of population P (breeding) in general form:

$$P(t+1) = S_{(\iota,\lambda)}(P(t) \cup (V \circ S_{(\iota,\lambda)})P(t)) \tag{1}$$

where $P(t+1) = (G^l(t+1), \varphi, S_{(\iota,\lambda)})$ is a finite population of size l at time $t+1$; $V: G^n \to G^l$ is a variation operator (mutation, recombination, etc.); $\cup, \circ$ are union and composition operations respectively.

Let a, b, c denote two pure breeds (a,b) and one crossbred (c) populations. Then, the model of interacting populations (crossbreeding):

$$\begin{cases} P_a(t+1) = S_{(\iota\in a,\lambda\in a)}\left(P_a(t) \cup \left(V \circ S_{(\iota\in a,\lambda\in a)}\right)P_a(t)\right), \\ P_b(t+1) = S_{(\iota\in b,\lambda\in b)}\left(P_b(t) \cup \left(V \circ S_{(\iota\in b,\lambda\in b)}\right)P_b(t)\right), \\ P_c(t+1) = S_{(\iota\in c,\lambda\in c)}\left(P_c(t) \cup \left(V \circ S_{(\iota\in c,\lambda\in c)}\right)\{P_a(t), P_b(t)\}\right), \\ P_a(0) = n_{a0}, \\ P_b(0) = n_{b0}, \\ P_c(0) = 0, \end{cases} \tag{2}$$

where n_{a0}, n_{b0}, n_{c0} are sizes of populations a, b and c at the time $t = 0$ respectively.

Dominance. The following sampling was implemented in ADAM to allow a contribution of dominance effects to the total genotypic value (Wellmann and Bennewitz 2011):

$$\delta = \mu + k \cdot \sqrt{\sigma^2} \tag{3}$$

$$d = \delta \cdot |a| \tag{4}$$

where, δ is a dominance degree, $\delta \sim N(\mu, \sigma^2)$; μ is a mean of δ with a default value of 0.193; σ^2 is a variance of δ with a default value of 0.312 (Bennewitz *et al.* 2017); $k \sim N(0,1)$ is a random number; d is a dominance contribution to a genetic value at each locus; a is an additive contribution to a genetic value at each locus.

Data and simulation studies. In order to demonstrate the effect of the novel features implemented in the software, we performed a simple case study simulating a dairy cattle crossbreeding scheme. ADAM can either generate founder haplotypes with user-defined settings for historical populations or read existing haplotypes and genetic architecture into the program. The haplotypes were randomly sampled from the genotype data as used in Thomasen *et al.* (2020) to resemble the Danish dairy cattle populations. A total of 40,000 SNP markers and 1000 QTL were randomly distributed in the genome. A trait with heritability of 0.5, additive genetic variance of 1.00 and dominance variance of 0.25 were considered.

Simulated breeding scheme. A conventional progeny-testing scheme was performed in three purebred populations (population 1-3) in generation 1-10. In each generation and each population, 200 young bulls were progeny-tested and 20 proven bulls (5- to 6-year-old) were truncation-selected for breeding. In addition, 3,000 cows (1- to 5-year-old) were selected out of 10,000 cows for breeding. All progeny-tested bulls were genotyped and included in the sire reference population. In generation 11 to 13, all males and females were genotyped, and a three-way rotational crossbreeding scheme was performed. In each generation, 10 purebred males were mated to 1000 crossbred females, and the crossbred animals are considered to be in a crossbred population (population 4). The breeding scheme simulation was replicated 5 times. The average additive and dominance effects in the population were evaluated for purebred animals in generation 0-10 in population 1 (representing a purebred population) and for crossbred animals in generation 11 to 13 in population 4.

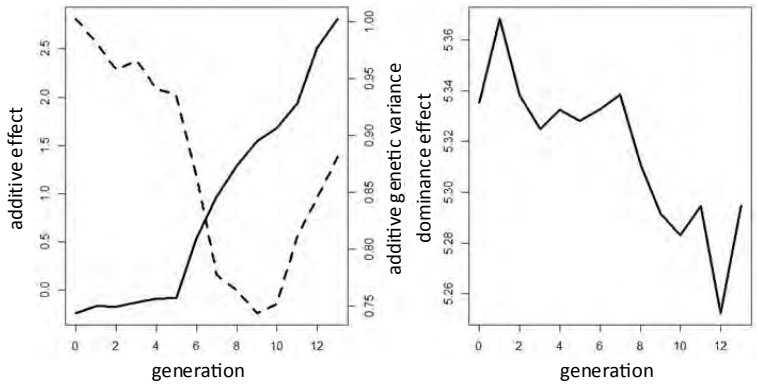


Figure 2. Additive effect (full line), additive genetic variance (broken line) and dominance effect across 13 generations of the simulated breeding schemes.

Results & discussion

The simulation results for the selected breeding scheme are depicted in Figure 2, where the trends of average additive and dominance effects across generations are shown for generation 0-10 in population 1 population and generation 11-13 in population 4. The rate of genetic gain rapidly improved due to crossbreeding in generation 11-13 (0.44 in genetic standard deviation unit per generation) compared to generation 1-10 (0.24 per generation). The dominance effect decreased during generation 1-10 due to selection within purebred populations and increased in generation 11 as the level of heterozygosity increased due to crossbreeding. The dominance effect possibly fluctuated across generations due to a small number of replicates. The introduction of crossbreeding in generation 11-13 increased the additive genetic variance from 0.74 to 0.88 (Figure 2).

This work represents the novel advanced features of ADAM software. By using a simple test case we showed that the software can simulate multiple interacting populations within one breeding scheme while accounting for dominance effects. With the new functionalities, ADAM can simulate a broad range of research-based as well as commercial-based breeding and crossbreeding schemes of variable complexity. By using ADAM, breeding organizations have the opportunity to optimize their breeding strategies and use of resources and, thereby, optimize profitability.

References

Bennewitz, J., C. Edel, R. Fries, T.H.E. Meuwissen, and R. Wellmann (2017) Genet. Sel. Evol. 49 (1):7. https://doi.org/10.1186/s12711-017-0284-7

Chen, C., J. Zhu, H. Ren, Y. Deng, X. Zhang *et al.* (2021) Livest. Sci. 244:104342. https://doi.org/10.1016/j.livsci.2020.104342

Clasen, J.B., E. Norberg, P. Madsen, J. Pedersen, and M. Kargo (2017) J. Dairy Sci. 100 (8):6337-6342. https://doi.org/10.3168/jds.2017-12627

Faux, A.-M., G. Gorjanc, R.C. Gaynor, M. Battagin, S.M. Edwards *et al.* (2016) Plant Genome 9 (3). https://doi.org/10.3835/plantgenome2016.02.0013

Han, Z., Q. Li, S. Liu, and L. Kong (2020) Aquac. 529:735621. https://doi.org/10.1016/j.aquaculture.2020.735621

Liu, H., B.B. Tessema, J. Jensen, F. Cericola, J.R. Andersen *et al.* (2019) Front. Plant Sci. 9 (1926). https://doi.org/10.3389/fpls.2018.01926

Paixão, T., G. Badkobeh, N. Barton, D. Çörüş, D.-C. Dang *et al.* (2015) Toward a unifying framework for evolutionary processes. J. Theor. Biol. 383:28-43. https://doi.org/10.1016/j.jtbi.2015.07.011

Pedersen, L.D., A.C. Sørensen, M. Henryon, S. Ansari-Mahyari, and P. Berg (2009) Livest. Sci. 121 (2):343-344. https://doi.org/10.1016/j.livsci.2008.06.028

Pook, T., M. Schlather, and H. Simianer (2020) G3: Genes Genomes Genet. 10 (6):1915-1918. https://doi.org/10.1534/g3.120.401193

Therkildsen, M., S.K. Jensen, and A. G. Kongsted (2021) Livest. Sci. 250:104584. https://doi.org/10.1016/j.livsci.2021.104584

Thomasen, J.R., H. Liu, and A.C. Sørensen (2020) J. Dairy Sci. 103 (1):597-606. https://doi.org/10.3168/jds.2019-16974

Wellmann, R., and J. Bennewitz (2011) Genet. Res. 93 (2):139-154. https://doi.org/10.1017/S0016672310000649

446. A tool for comprehensive implementation of simple and complex breeding programs

B.P. Kinghorn[1*], A.J. Kinghorn and J.L. Baller[2]

[1]University of New England, Armidale NSW 2351, Australia; [2]ABS Global, Inc., 1525 River Rd., DeForest, WI 53532, USA; bkinghor@une.edu.au

Abstract

This paper gives a brief overview of the tool MateSel, which is used for implementation of breeding programs across many animal and plant species. Progress in algorithms underlying MateSel is briefly described. An example of the Multiple EndUses feature in an industry dataset is used to help illustrate flexibility to control outcomes.

Introduction

There are many issues to handle when running a breeding program. After EBV generation, there has to be consideration of e.g. constraints on numbers of matings per individual, progeny inbreeding, targeted outcomes for specific traits, and migration of stock between groups. Attention should also be paid to Optimal Contribution Selection (OCS) to manage diversity.

Practitioners can spend a lot of time attempting to make balanced decisions across such diverse issues. This paper describes the tool MateSel, which addresses these issues and many more in a single framework, leading to discovery of compromises and opportunities, and production of a mating list to be adopted. Once established in a breeding operation, this tool gives high levels of control and performance. It can also save considerable human time and effort.

MateSel is comprehensively described by Kinghorn and Kinghorn (2022), and also in a series of short videos at https://matesel.com/Website/Support. This paper focuses on a single example that gives a new illustration, using the Multiple EndUses feature. An EndUse is a targeted application for progeny out of a group of matings. Examples are matings to produce bulls for Terminal use versus Maternal use, or for production of progeny targeted at use in the Nucleus versus use in a Multiplier. An EndUse index is the selection index that is used as a criterion to evaluate stock for that EndUse.

Materials & methods

A note on method underlying MateSel. Being used across animal and plant breeding programs, MateSel handles a wide range of biological, logistical and practical constraints. All these issues have to be well accommodated, as the mating list results are used directly in operations. This has led to increasing functionality as new applications arise, but at the time of writing, we know of no outstanding issues that cannot be accommodated. There are parameters for four categories of decision-making that are optimised for the prevailing objective function: (1) selection (matings allocated per candidate); (2) distribution of matings across Male Group × Female Group; (3) allocation of mates to make matings; (4) allocation of matings to different EndUses.

As global optimality is targeted, these parameters are optimised simultaneously in a one-step approach. Optimising selection decisions, then optimising mate allocation given these selection decisions, would be a two-step approach. Some scenarios are very challenging for optimisation, for example large problems with many group mating permission restrictions and multiple thresholds on inbreeding/trait/marker outcomes. This has led to ongoing improvement in method compared to that of Kinghorn (2011), with a mix of

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_446, © B.P. Kinghorn *et al.* 2022

stochastic and deterministic optimisation strategies coordinated in a manner that maintains the one-step approach.

Example using the multiple EndUses feature. We use an example in cattle, with thanks to ABS Global for providing the data and scenario. This ABS job relates to a subset of the breeding program, in which specialised selection for two different indices is required from a single pool of candidates. We will refer to these as EndUse 2 and EndUse 3. Just for illustration, we will later add EndUse 1, to target a fictitious breeding nucleus, with its selection criterion being the scaled average of the EndUse 2 and 3 indices.

One hundred matings are to be generated from 154 female candidates. Seventy matings are required for EndUse 2, and 30 matings for EndUse 3. The Control settings are to give equal emphasis to EndUses 2 and 3 and to practice truncation selection – no OCS required in this subset of the ABS breeding program.

Results

The top row in Figure 1 shows EndUse 2/3 index scatterplots. The left graph shows the females selected for EndUse 2 (orange) and EndUse 3 (blue), and 54 culled females (grey). The corresponding graph for males would be less informative, because males are generally used in multiple matings, and the same male can be used for matings that target different EndUses. However, the combination of male and female selections together with mate allocations is in the right-hand graph, which shows the EndUse 2/3 index scatterplot of progeny mean indices.

The top row results might not please the practitioner: Looking at the bottom-left square box of culled females (grey dots), those near the top-right have been culled, whereas those just outside the top-left and bottom-right corners have been selected, despite the fact that they are very poor for one index, and just marginally across-the-line for the other index. This is rectified in the next row of results by adding some positive emphasis on each of the two indices using MateSel's Trait Management feature (see Kinghorn and Kinghorn, 2022), without altering settings for the Multiple EndUse feature. This can be done on-the-fly, editing the relevant weightings to change the overall objective function *during* the analysis, to find weightings that give satisfactory results. These results may indeed be of higher long-term utility, as Mendelian segregation and environmental deviations can change the ultimate fate of resulting progeny.

In the top row of Figure 1, the equal emphasis given to EndUses 2 and 3 can be seen in the 45-degree line separating the blue and orange dots. In the bottom row, the weighting on Enduse 3 has been doubled, resulting in a lowering of that division between the blue and orange dots, with EndUse 3 benefitting somewhat at the expense of EndUse 2. This contributes to a better result for EndUse 3 in predicted progeny merit.

For Figure 2, 40 matings for EndUse 1 have been added, using black dots. The relative weightings for EndUses 1, 2 and 3 were 1, ½, ½ respectively, ensuring in this case that EndUse 1 (breeding nucleus) had priority over EndUses 2 and 3. The fact that the EndUse 1 index is the average of the EndUse 2 and 3 indices is seen in the truncation slope for the black dots to the top-right in the graph of female selections for truncation selection (Target Degrees (TD)=0, where TD=90 is fully for minimum coancestry selection, Kinghorn and Kinghorn (2022)).

But of course, we should pay attention to genetic diversity in a breeding nucleus. The bottom row in Figure 2 shows that under equal emphasis on parental coancestry (TD = (0+90)/2= 45), the nucleus drops selection of some higher-index females in order to select some lower coancestry females. This impact is usually higher for male selection, due to greater opportunity to vary contributions, seen in the lower response in the nucleus under OCS at TD=45 in the right-hand graph, to maintain more diversity. Dropping some high

index contributions to EndUse 1 gives opportunity for improved response for the other two EndUses, for which genetic diversity is not required.

Discussion

The right-hand graphs in Figure 1 and 2 show greater separation between EndUses than the left-hand graphs. This is due to assortative mate allocation across the multiple EndUses involved, as well as the effects of sire selection. Driving this separation gives better future opportunities, for example for satisfying the diverse needs of different seedstock buyers, or for planned splitting of a breeding line in multiple directions. This can effectively remove a generation of lag, as described here: https://www.youtube.com/watch?v=mAl0YsoFVgM

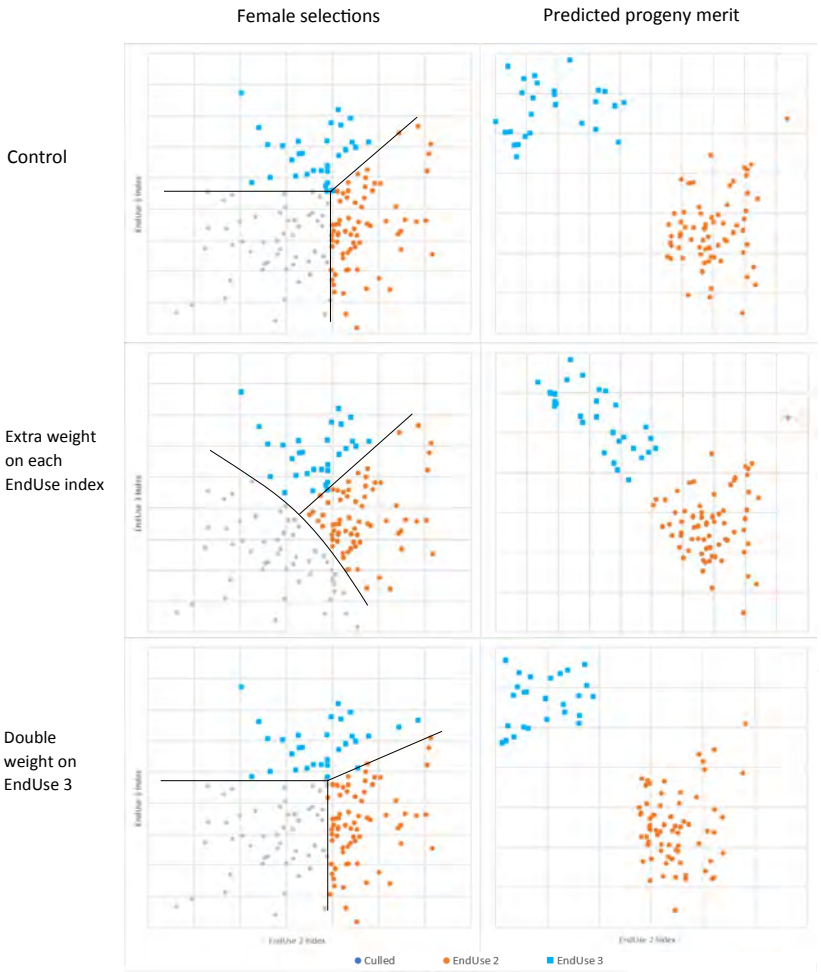


Figure 1. Scatterplots of EndUse 3 index on EndUse 2 index. Color of points shows the fate of each female (left column) and the predicted merit of each mating (right column). Black lines are for visual guidance. See text for more detail.

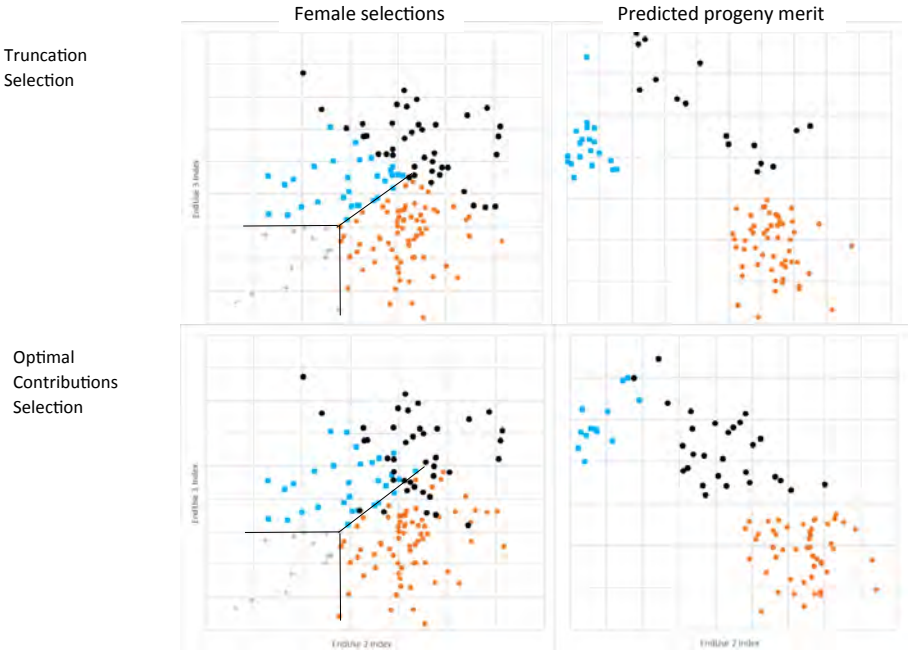


Figure 2. As for Figure 1, but with 40 EndUse 1 (nucleus population) matings added in black, leaving 40 fewer culled females (light grey). See text for more detail.

This session of WCGALP is on Methods and Tools. Most work in this area relates ultimately to genetic evaluation of individuals. MateSel takes off from that point to help fill the gap between EBVs and the implementation of a breeding program. This tool continues to evolve, with updates released every few months over the last several years. The example given illustrates just one specific feature, and is far from a comprehensive overview of the software. For that please see Kinghorn and Kinghorn (2022).

References

Kinghorn, B.P. (2011) An algorithm for efficient constrained mate selection. Genetics Selection Evolution 43:4. http://www.gsejournal.org/content/43/1/4

Kinghorn, B.P. and A.J. Kinghorn (2022) Instructions for MateSel. Available at: https://matesel.com/content/documentation/MateSelInstructions.pdf

447. Beyond scenarios – resource optimization for dairy cattle breeding programs (MoBPSopti)

A. Hassanpour*, H. Simianer and T. Pook

University of Göttingen, Center for Integrated Breeding Research, Animal Breeding and Genetics Group, Thaer-Weg Street 3, 37075, Göttingen, Germany; azadeh.hassanpour@uni-goettingen.de

Abstract

Modern breeding programs are highly complex with various interdependent parameters and determining the optimal breeding scheme is practically impossible. Therefore, it is a common practice to reduce the optimization problem to a set of scenarios that in turn are deeply analyzed in detail. We are proposing a framework to optimize breeding programs beyond just a comparison of scenarios. We first define the design space of all possible breeding programs and approximated their outcome regarding an objective function using stochastic simulation and local smoothing. Next, the space is reduced to the most promising areas of the design space using various optimization techniques. This process is repeated iteratively until an optimal scheme is obtained. We applied our method to a dairy cattle program with an objective function that includes both genetic gain and genetic diversity conservation under a budget constraint. This study provides more detailed insights into potential approaches of designing and optimizing breeding programs by finding the optimum allocation of resources based on each objective's overall impact on mid-and long-term breeding goals.

Introduction

When it comes to addressing global challenges caused by the annual growth rate of the human population and the rapidly changing climate, the breeding sector is required to increase production capacities. Modern breeding systems are highly complex with various decisions that need to be taken, such as parent selection, mating decisions, and managing the rate of inbreeding (Martin-Collado *et al.* 2018; Simianer *et al.* 2021). In addition, the efficient allocation of resources in breeding programs involves competing objectives where the achievement of one objective oftentimes comes at the expense of another. Thus, there is a strong demand of breeders to optimize their breeding schemes and predict the performance of different breeding scenarios before conducting experiments in real-life settings. Optimization is further complicated by breeding actions or changes in parameters of the breeding scheme being highly interdependent of each other. Choosing the number of animals to use in each breeding cycle for mating and selection is one of the first and most important decisions made by a breeder, where breeders seek to maximize the expected gain in target traits by repeated selection rounds while controlling inbreeding and staying within a given budget. To find optimum test resource allocation in breeding programs, simulation techniques can be used to examine all possible resource allocation combinations individually. However, this approach becomes computationally expensive or even prohibitive with the increasing complexity of the breeding program to optimize and resources to allocate. That is why optimization algorithms are vital when the breeding programs are subject to multistage selection with many parameters, where the initial population of breeding units is evaluated in the first stage and only a fraction of the superior candidates are selected in the subsequent stage(s) as parents. Our study aims to determine the best resource utilization or near-optimal resource combination under the existing scenarios based on simulation-based optimization methods and different primary breeding objectives.

Materials & methods

In the following, we will propose a general pipeline for the optimization of breeding programs and will provide an exemplary breeding scenario of a simplified classical dairy cattle breeding scheme (Figure 1).

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_447, © A. Hassanpour *et al.* 2022

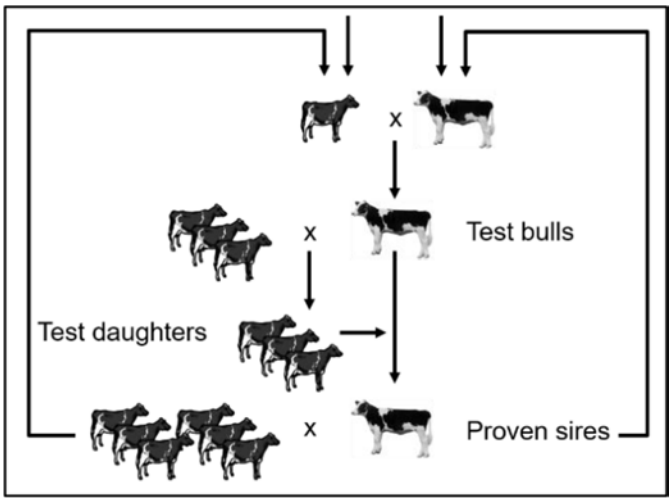


Figure 1. A dairy cattle breeding strategy.

As bulls cannot be phenotyped for performance traits, pre-selected candidates need to be mated to cows to produce test daughters, and test bulls are then selected based on their offspring performance. The selection of bulls is assumed to be based on genomic breeding value estimation, while cows were selected via pedigree breeding value estimation. The last three cow generations were considered for selection with phenotypic information being available for the two older generations. To detect the optimal number of test daughters and bulls for mating and selection in each cycle, the present study attempted to improve their characteristics over 10 generations. 5 burn-in cycles were performed to have a realistic initial population structure. A single trait was simulated with a heritability of 0.3, 1000 purely additive QTLs, a mean genomic value of 100, and a genomic standard deviation of 10. The interested reader is referred to https://github.com/tpook92/MoBPS/ for the exact simulation script of the simulation. Firstly, we have to quantify the search space of all potential breeding schemes and the parameters we are considering for the optimization. In the dairy cattle breeding scheme, the optimization problem can be defined by three main components: the number of test daughters x_1, test bulls x_2, and selected proven sires x_3. When assuming a yearly budget of 10 million euros and housing cost of 3,000 € per bull and 4,000 € per cow and setting reasonable boundaries based on stable capacities, this leads to the following set of constraints:

$$4000x_1 + 3000x_2 - 10000000 \leq 0,\ x_1, x_2, x_3 \geq 0,\ 100 \leq x_2 \leq 700,\ 3 \leq x_3 \leq 30$$

Next, we need to specify our objective function m, which we define as a composite between the genetic gain and inbreeding depending on our set of parameters. This could for example be:

$$m(x_1, x_2, x_3) = f(x_1, x_2, x_3) - 50 * g(x_1, x_2, x_3)$$

with f being the resulting genetic gain and g being the inbreeding rate after 15 years of breeding. As f and g are on a different scale, we additionally performed scaling by a more or less arbitrarily chosen value of 50 for the rate of inbreeding in the composite function. Higher values would put a higher weight on breeding for genetic diversity, while lower values would put more focus on the genetic gain, and depending on the decision-maker, a proposed weighting will of course differ (which in turn also leads to a different optimal breeding program). The next step of our pipeline is the optimization process itself to maximize the objective function within the limits of our constraints. As m cannot be derived directly, we need to approximate it.

For this, we propose to perform stochastic simulation of the breeding scheme using the R package MoBPS (Pook *et al.* 2020). Next, m is approximated using a kernel regression method, a weighted average with values closer to the value of the interest being weighted stronger (Nadaraya 1964; Watson 1964):

$$\hat{m}(x_1, x_2, x_3) = \frac{\sum Y_i * K_h(\frac{x_1 - x_{i,1}}{h_1}, \frac{x_2 - x_{i,2}}{h_2}, \frac{x_3 - x_{i,3}}{h_3})}{\sum K_h(x_1 - x_{i,1}, x_2 - x_{i,2}, x_3 - x_{i,3})}$$

with K being the used Kernel function and h_i being the bandwidth to control which simulations are considered with what weight in the smoothing procedure. Y_i represents an outcome of stochastic simulations with the input parameters ($x_{i,1}$, $x_{i,2}$, $x_{i,3}$). Note that Y_i is a random variable with expected value $m(x_1, x_2, x_3)$ and unknown variance. Subsequently, optimization techniques like gradient descent, simulated annealing, and genetic algorithms are used to obtain potential maxima for $\hat{m}$. Finally, more simulations in the areas of interest can be performed, which in turn allows for a reduction of the bandwidth and a better approximation of m. This step can be iteratively repeated until obtained maxima stabilizes.

Results

In the first step, 60,000 simulations at arbitrary positions in the search space were performed. As our search space in this particular example was 'only' three-dimensional, the initial bandwidth was chosen based on visual inspection as (h_1 = 30, h_2 = 30, h_3 = 1). As inbreeding rates showed a much smaller variance than the obtained genetic gains, bandwidth for inbreeding was divided by 3. The result of simulations in contour maps shows that genetic gain is maximized with the highest function value of 12.56 when applying strong selection and generating a high number of test daughters per test bull (x_1 = 2,418, x_2 = 109, x_3 = 3) (Figure 2a), where inbreeding in such cases is also high (Figure 2b), leading to a not ideal performance in the composite function.

Results are in-line with current breeding theory, where the optimum response to selection can be obtained by maximizing genetic gain with high selection intensity, where a low number of test bulls are beneficial with resulting higher accuracy of the breeding value estimation (due to the higher number of test daughters per bull). On the contrary, a higher number of proven sires and lower selection intensity results in a decreased response to selection and inbreeding rates, where the minima for g was obtained at (2,423, 102, 30) with the best function value of 0.025 (Figure 2b). In the optimization phase for the composite function m, all optimization methods suggested the best function value around 107.86 and the local optimum at (2,371, 172, 14), (2,353, 189, 12), and (2,332, 224, 16) (Figure 2c and 3a). In the second iteration, a further 50,000 simulations were performed with the search space being reduced based on the result of the first optimization to $120 \leq x_2 \leq 260$, $10 \leq x_3 \leq 20$. As this leads to about 10 times as many simulations in the target area, the bandwidth for the second cycle was halved (h_1 = 15, h_2 = 15, h_3 = 0.5) with the best solution 107.89 from optimization suggesting (2,368,

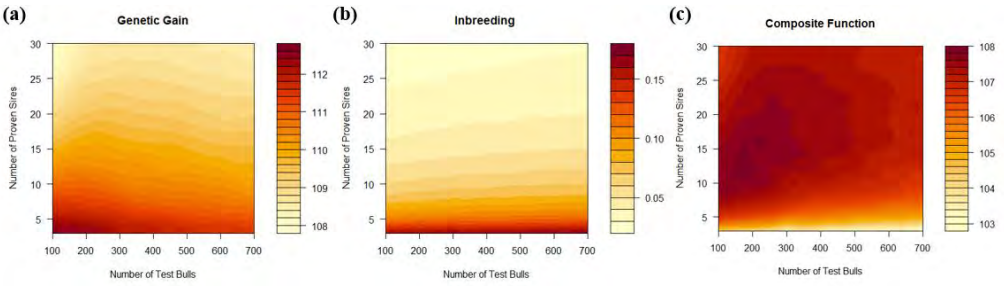


Figure 2. Smoothed surface contour map for full space of (a) f, (b) g and (c) m.

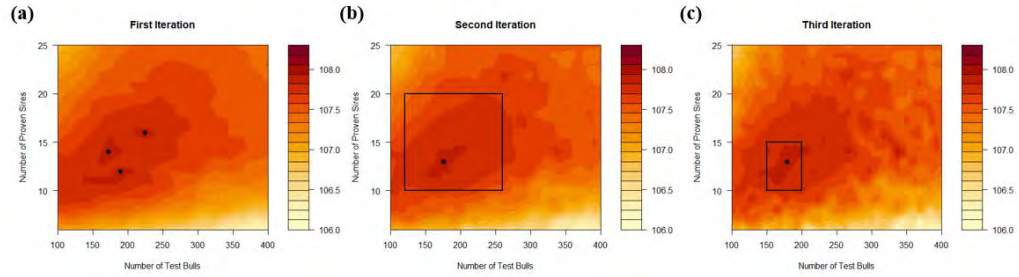


Figure 3. Smoothed surface contour map for the composite function of (a) full space (zoom in): first iteration, (b) second iteration, (c) third iteration. Dots indicate local maxima and squares indicate the search space in the current iteration.

175, 13) as the optimum (Figure 3b). After the third cycle with 12,000 simulations and a further reduced state space ($150 \leq x_2 \leq 200$, $10 \leq x_3 \leq 15$) with bandwidth again halved, the optimization suggests (2,365, 179, 13) indicating stabilization of the optima with the best value of 107.91 (Figure 3c).

Discussion

Optimum allocation of resources has been addressed in breeding mostly focusing on whether the selection decisions should be made upon genotypic or phenotypic information (Henryon *et al.* 2014). Yet, these concessions remain underexploited and little attention had been devoted to the potential benefits of resource optimization in animal breeding. Simulation studies have become increasingly important by enabling breeders to make informed decisions on suitable breeding strategies. Although there is no practical way to exhaustively examine all possible combinations of test parameter values in a complex breeding program, our proposed pipeline is based on a combination of a multivariate optimization approach and systematic zooming in into promising regions of the design space by forming iterative guesses in the proximity of the initial solutions. With gradient descent, multiple optima can be obtained depending on the starting value of the optimizer. Our results provide a promising perspective for test resource optimization by integrating nonparametric approaches that yield the best tradeoff between changes in inbreeding and rates of genetic gain in breeding programs subject to multistage selection. Future research directions include finding the optimal selection of the bandwidth of the kernel using automated techniques and finding a robust optimization framework for problems with a large number of parameters.

References

Henryon M., Berg P., and Sørensen A.C. (2014) Animal-breeding schemes using genomic information need breeding plans designed to maximize long-term genetic gains. In Livestock Science 166, pp. 38–47. https://doi.org/10.1016/j.livsci.2014.06.016

Martin-Collado D., Byrne T.J., Diaz C., and Amer P.R. (2018) Complexity of animal breeding choice making. In Journal of animal breeding and genetics 135 (6), pp. 395–409. https://doi.org/10.1111/jbg.12360

Nadaraya E.A. (1964) On Estimating Regression. In Theory Probab. Appl. 9 (1), pp. 141–142. https://doi.org/10.1137/1109020

Pook T., Schlather M., and Simianer H. (2020) MoBPS - Modular Breeding Program Simulator. In G3: Genes|Genomes|Genetics 10 (6), pp. 1915–1918. https://doi.org/10.1534/g3.120.401193

Simianer H., Büttgen L., Ganesan A., Ha N.T., and Pook T. (2021) A unifying concept of animal breeding programmes. Animal Breeding and Genetics. Volume 138 (2), pp. 137–150. https://doi.org/10.1111/jbg.12534

Watson G.S. (1964) Smooth regression analysis. 359-372: Indian J. Stat (Series A). https://www.jstor.org/stable/25049340

448. Predicting the outcome of a selection scheme in maternal traits

A.V. Mäki-Tanila* and J. Pietarinen

University of Helsinki, P.O. Box 28, 00014 Helsinki, Finland; asko.maki-tanila@helsinki.fi

Abstract

The rate of selection response in maternal traits is predicted with a BLUP approximation allowing an assessment for a wide range of genetic parameters and data recording schemes. The method exploited Willham model separating the direct and maternal effect in the analysis. It is accommodating the changes in genetic (co)variances due to selection. The focus is on uniparous species yielding a simple situation and on the benefits of augmenting the sib information with cousin information under commonly found antagonistic correlation between the effects. There is indication that the cousin information brings qualitative benefits to the approximation.

Introduction

In mammal species, e.g. beef cattle, the dam has an influence (maternal effect) on early growth and other juvenile traits in progeny. According to the Willham (1963) model, the genetic variation in maternal traits is allocated to the components due to the individual's own (direct) effect (a) and the dam's (d) (indirect) maternal effect (m_d) (Bijma 2006). In mammals, these effects are mostly genetically antagonistic. In maternal traits, the genetic evaluation relies on information across generations. For multi-generation data, the BLUP methodology would automatically yield reliable estimates of breeding values (EBV). When we are launching a selection scheme in a new breed or, in particular, in a new species, e.g. reindeer (Muuttoranta *et al.* 2014), we need tools to assess the requirements for organising an optimum recording. Wray and Hill (1979) developed a method, which is approximating the BLUP procedure in predicting the short-term outcome for a selection scheme. We have developed the method for maternal traits with a comparison to simulated results (Pietarinen and Mäki-Tanila 2020). Except the case with substantial genetic antagonism, the method worked in a satisfactory manner. The objective of the present work is stemming from a suggestion by Pieter Bijma (pers. comm.) to improve the approximation by including information on collateral relatives beyond sibs, i.e. cousins. The immediate interest of our research is in planning selection schemes for reindeer, which is a uniparous species like cattle. The methodological development also stays simpler when we consider only half-sibs and their progeny.

Materials & methods

A maternally affected phenotype (p) is expressed as: $p = a + m_d + e$, where e is environmental deviation (variance σ_e^2). We use the notation σ_p^2, σ_a^2 and σ_m^2 for phenotypic variance and direct and maternal genetic variance, respectively, and σ_{am} for the covariance. We assume that the effective population size is so large that selection does not affect the (co)variances σ_{a0}^2, σ_{m0}^2 and σ_{am0} (0 referring to the base generation) due to segregation.

We construct a family index based on the selection candidate's phenotype and its half-sibs' mean $\overline{HS}$ so that its *EBV* for either effect is $b_1(p - \overline{HS}) + b_2\,\overline{HS}$ where b_j is the respective index weight. In a hierarchical mating structure, each sire (s) mates with f dams each having one offspring. Usually, maternal traits are analysed separating technically the direct and maternal effect into two traits. In constructing the index we use the compact approach of Van Vleck (1970). The selection index $I = b'x$ has the coefficient vector $b = P^{-1}\,G\,w$ with the phenotypic (co)variance matrix P and G matrix for the covariance of a and m with the information of sources in the index. The weighting of the effects in the aggregate selection goal are expressed by a vector w. The variances in generation t are according to Willham (1963):

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_448, © A.V. Mäki-Tanila *et al.* 2022

$\sigma^2_{(p-\overline{HS})t} = \frac{f-1}{f}\left(\frac{1}{4}\sigma^2_{a(t-1)d} + \frac{1}{2}\sigma^2_{a0} + \sigma_{am(t-1)d} + \sigma^2_{m(t-1)d} + \sigma^2_e\right)$ and

$\sigma^2_{\overline{HS}(t)} = \frac{1}{4}\sigma^2_{a(t-1)s} + \frac{1}{f}\left(\frac{1}{4}\sigma^2_{a(t-1)d} + \frac{1}{2}\sigma^2_{a0} + \sigma_{am(t-1)d} + \sigma^2_{m(t-1)d} + \sigma^2_e\right)$, and P is diagonal. Using the notation $A_w = \frac{1}{4}\sigma^2_{a(t-1)d} + \frac{1}{2}\sigma^2_{a0} + \sigma_{am(t-1)d}$ and $M_w = \frac{1}{4}\sigma_{am_{(t-1)d}} + \frac{1}{2}\sigma_{am_0} + \frac{1}{2}\sigma^2_{m_d}$), we have:

$$\boldsymbol{G'_t} = \begin{pmatrix} \frac{f-1}{f}A_w & \frac{1}{4}\sigma^2_{a(t-1)s} + A_w/f \\ \frac{f-1}{f}M_w & \frac{1}{4}\sigma_{am(t-1)s} + M_w/f \end{pmatrix}$$

The columns of G are indicated by g_a and g_m. The BLUP approximation is constructed adding the selection candidate's dam's and sire's EBV of a and m and the half-sibs' dams' respective EBV-means amongst the index terms. The $p - \overline{HS}$ is correlated only with $\hat{a}$ and $\hat{m}$, i.e. $\frac{f-1}{f}\left(\frac{1}{2}\sigma^2_{\hat{a}(t-1)} + \sigma_{\hat{a}m(t-1)}\right)$ and $\frac{f-1}{f}\left(\frac{1}{2}\sigma_{\hat{a}m(t-1)} + \sigma^2_{\hat{m}(t-1)}\right)$ while $\overline{HS}$ has the covariances $\frac{1}{2}\sigma^2_{\hat{a}(t-1)} + \sigma_{\hat{a}m(t-1)}$ with both $\hat{a}_d$ and $\overline{\hat{a}_d}$ and $\frac{1}{2}\sigma_{\hat{a}m(t-1)} + \sigma^2_{\hat{m}(t-1)}$ with $\hat{m}_d$ and $\overline{\hat{m}_d}$ and $\frac{1}{2}\sigma^2_{\hat{a}(t-1)}$ and $\frac{1}{2}\sigma_{a\hat{m}(t-1)}$ with $\hat{a}_s$ and $\hat{m}_s$, respectively. In the matrix P the EBV's and their means have (co)variances within individuals and between the effects and their covariances with a and m in G are alike, respectively.

The BLUP approximation may be improved by adding phenotypic information on the sire's female half-sibs' progeny (each sired by a different male) and their parents' EBV. The progeny is a paternal half-cousin (*HC*) with respect to the selection candidate. The essential part in $\overline{HC}$ correlated via the common grandsire (GS) with p and $\overline{HS}$ and sire's EBV's, is $\frac{1}{2}a_{dHC} + m_{dHC}$. We have e.g. $cov(\overline{HS}, \overline{HC}) = \frac{1}{16}\sigma^2_{a(t-2)GS} + \frac{1}{8}\sigma_{am(t-2)GS}$ (cf. Willham 1963; Bondari 1971; Thompson 1976).

A somewhat cumbersome part in the construction is the derivation of the covariances involving *HC*'s (related) dams' EBV. This kind of machinery has been developed in predicting the rate of inbreeding (cf. Woolliams and Bijma 2000). As we have $\hat{a} = \boldsymbol{b'_a}\,\boldsymbol{x}$ where $\boldsymbol{b_a} = \boldsymbol{P^{-1}g_a}$, then for P matrix e.g. $cov(HS, \overline{\hat{a}_{dHC}}) = cov\left(\frac{1}{2}a_s, \boldsymbol{b'_a}\,\boldsymbol{x}\right) = \frac{1}{2}\boldsymbol{b'_a}\,cov(a, \boldsymbol{x})_{HS}$ where $cov(a, \boldsymbol{x})_{HS}$ or $\boldsymbol{g_{aHS}}$ differs from $\boldsymbol{g_a}$ in having the first element A_w/f. Similarly $cov(\hat{m}_s, \overline{pHC}) = cov\left(\boldsymbol{b'_m}\,\boldsymbol{x}, \frac{1}{2}a_{dHC} + m_{dHC}\right) = \frac{1}{2}\boldsymbol{b'_m}\,\boldsymbol{g_{aHS}} + \boldsymbol{b'_m}\,\boldsymbol{g_{mHS}}$ with the first element of $\boldsymbol{g_{mHS}}$ being M_w/f. Obviously there are also covariances between half-sibs' EBV's. For example, $cov\left(\hat{m}_s, \overline{\hat{a}_{dHC}}\right) = cov\left(\boldsymbol{b'_m}\,\boldsymbol{x_i}, \boldsymbol{b'_a}\,\boldsymbol{x_j}\right) = \boldsymbol{b'_m}\,cov\left(\boldsymbol{x_i}, \boldsymbol{x_j}\right)_{HS}\,\boldsymbol{b_a}$. The phenotypic covariance matrix $cov\left(\boldsymbol{x_i}, \boldsymbol{x_j}\right)_{HS}$ has the [1,1] element as A_w/f and the covariance of $p - \overline{HS}$ with $\hat{a}_d$ is now $\left(\frac{1}{2}\sigma^2_{\hat{a}(t-1)} + \sigma_{\hat{a}m(t-1)}\right)/f$ and that with $\hat{m}_d$ is $\left(\frac{1}{2}\sigma_{a\hat{m}(t-1)} + \sigma^2_{\hat{m}(t-1)}\right)/f$ whereas the covariances between half-sibs' (unrelated) dams' EBV's are all 0. The selection is possible only amongst males. The genetic (co)variances among the selected sires are affected by Bulmer (1971) effect. For example, the covariance amongst selected sires is $\sigma_{am(s)} - k\ \sigma_{aI}\,\sigma_{mI}\,/\sigma^2_I$ where $k = i(i - x)$, i is the intensity of selection and x is the truncation point on the standardized scale (cf. Villanueva *et al.* 1993).

Results & discussion

We have developed a BLUP approximation for predicting the short-term outcome from a selection on maternal traits with an allowance for changes in mating ratio. The approximation is based on a selection index of the following information sources: $p - \overline{HS}$, $\overline{HS}$, $\overline{HC}$, $\hat{a}_d$, $\hat{a}_s$, $\overline{\hat{a}_{dHS}}$, $\overline{\hat{a}_{dHC}}$, $\overline{\hat{a}_{sHC}}$, $\hat{m}_d$, $\hat{m}_s$, $\overline{\hat{m}_{dHS}}$, $\overline{\hat{m}_{dHC}}$ and $\overline{\hat{m}_{sHC}}$. The number of index terms generates a huge number of parameters and the research on the selection of indirect effects usually resorts to simulation (Khaw *et al.* 2014; Lillehammer *et al.* 2020). Analytical methods are appealing, because they potentially provide general understanding and predictions for a wide range of design factors.

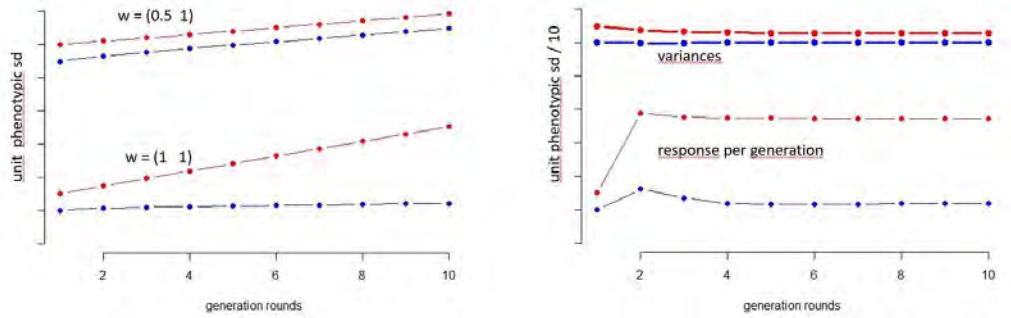


Figure 1. Left: Predicted responses in direct (red) and maternal (blue) effect (in σ_p units of the base generation) when the selection is on an aggregate index of approximated BLUP-EBV's based on selection candidate's, half-sibs' and half-cousins' phenotype and their parents' EBV for two weight combinations (vector w). Right: The changes in the variance and in the rate of progress in the direct and maternal effect across generations for equal economic weights. The heritabilities $\boldsymbol{h_a^2 = h_m^2 = 0.3}$ and r_{am} = -0.5 and the proportion selected in males is 6.7% and f = 30

The rate of progress and the changes in genetic variances behave in an expected manner and like in the previous BLUP approximations, i.e. the asymptotic response per generation and the stable variance levels are reached within few generations in both the effects (Figure 1). Using equal weights and the parameters of Figure 1, with a highly negative genetic correlation (-0.5) between the effects the asymptotic maternal response is negative (-0.017 per generation) vs positive (0.056) without vs with *HC* data while the respective direct gain is 0.254 vs 0.205. The simulation (for details, see Pietarinen and Mäki-Tanila 2020) yields an average response over 7-10th generation as 0.217 for direct and 0.055 for maternal effect. With no genetic correlation between the effects the predicted (simulated) response is 0.288 without *HC* data vs 0.304 with *HC* data (0.304) for direct and 0.152 vs 0.125 (0.163) for maternal effect. There are indicative benefits from including the cousin information in the approximation.

The BLUP approximation is very flexible and has been extended to a multi-trait context (Villanueva *et al.* 1993) and in assessing the major advantages from the use of genomic information in genetic evaluation (Dekkers 2007). These issues are of much interest also for the genetic improvement of maternal traits.

Acknowledgements

We have funding from the Nordic Centre of Excellence ReiGN project to carry out the research.

References

Bijma P. (2006) Journal of Animal Science 84: 800-806.

Bulmer M.G. (1971) American Naturalist 105: 201-211.

Bondari K. (1971) A study of genetic maternal effects in a designed experiment using Tribolium. PhD thesis, Iowa State University, Ames, Iowa.

Dekkers J.C. (2007) Journal of Animal Breeding and Genetics 124: 331-341.

Khaw H.L., Ponzoni R.W., and Bijma P. (2014) Genetics Selection Evolution 46, 39. http://www.gsejournal.org/content/46/1/39

Lillehammer M., Sonesson A.K., Klemetsdal G., Blichfeldt T., and Meuwissen, T.H.E. (2020) Journal of Animal Breeding and Genetics 137: 384-394.

Muuttoranta K., Holand Ø., Røed K.H., Tapio M., Nieminen M. *et al.* (2014) Rangifer 34: 21-36.

Pietarinen J., and Mäki-Tanila, A. (2020) Agricultural and Food Science 29: 395-404.

Thompson R. (1976) Biometrics 32: 903-917.
VanVleck L.D. (1970) Biometrics 26: 477-483.
Villanueva B., Wray N.R., and Thompson R (1993) Animal Production 57: 1-13.
Willham R.L. (1963) Biometrics 19: 18-27.
Woolliams J.A., and Bijma P. (2000) Genetics 154: 1851-1864.Wray N.R., and Hill W.G. (1989) Animal Production 49: 217-227.

449. Inbreeding in circular and panmictic mating systems

B. Guldbrandtsen

Department of Veterinary and Animal Sciences, Grønnegårdsvej 3, 1870 Frederiksberg, Denmark;
berntguldbrandtsen@sund.ku.dk

Abstract
The performance of circular and panmictic breeding schemes are compared based on the eigenvalue effective population size. A simple coalescent approximation is used to demonstrate that for larger populations, the asymptotic behaviours of the models are identical. Consistent with previous studies circular mating schemes are found to lead to early increases in realised inbreeding. A circular mating scheme is associated with a disadvantage. However, the long-term rate of inbreeding is lower for a circular mating scheme, but only for small populations.

Introduction
Among the challenges facing small populations are loss of genetic variation and build-up of homozygosity. Loss of genetic variation leads to loss of adaptive potential, e.g. future selection response. Increasing homozygosity leads to inbreeding depression. Inbreeding depression is the general decline in performance observed in inbred individuals, for a review see, e.g. Leroy G. (2014).

An effective method frequently employed to constrain loss of genetic variation is optimal contribution selection (Meuwissen, 1997). This method has been implemented in tools such as EVA (Berg, 2006). Here, selection of breeders requires striking an optimal balance between genetic gain and the concurrent increase in coancestry in the population. However, the deployment of this kind of scheme requires complete data of high quality and it requires centralized decision-making.

Often one or both of these conditions are not satisfied. In this situation, circular mating schemes (Kimura and Crow, 1963) are used in the management of some small and endangered breeds as a tool to restrain the increase in inbreeding levels. In a circular scheme, the population is divided into a number of subpopulations. In each generation, a number of animals (typically males) are transferred from each subpopulation to the next subpopulation in the circle. The behaviour of such a scheme was previously analysed by Nomura and Yonezawa (1996). By computing the eigenvalue effective population sizes (Ewens, 1982), the performance of a circular mating scheme is compared to the performance of a panmictic mating scheme with an undivided population. In addition to deriving a coalescence approximation, I here focus on aspects relevant to whether to choose a circular mating over a panmictic scheme.

As an alternative to forward simulation, the coalescent process (Kingman, 1982) in many situations produces good approximations to the long-term behaviour of population genetic models. Such an approximation is obtained here. The results are compared to the explicit results for the eigenvalue effective population size.

Materials & methods
Eigenvalue effective population size. For inbreeding schemes, it is straightforward to set up linear systems of equations describing the probability of sampling alleles identical by descent (IBD) from animals of all possible degrees of relationship. For schemes that converge to a state where all alleles are IBD, the equations can be converted to a homogeneous system of linear equations describing the decay of outbreeding, shown in Equation 1.

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_449, © B. Guldbrandtsen 2022

$$\boldsymbol{f}' = \mathbf{M}\boldsymbol{f} \tag{1}$$

f is a vector non-IBD probabilities for pairs of alleles drawn from the same individual (f), two different individuals in the same subgroup (g_0) and two individuals from subgroups i steps apart, respectively (g_i, i=1, …, n–1), in some generation. F' is the vector of probabilities in the next generation.

For a scheme with n=5 subgroups, the equations obtained are given in Equation 2.

$$\mathbf{M}=\begin{Bmatrix} 0 & 1 & 0 & 0 & 0 & 0 \\ \frac{1}{2N^*} & \frac{1}{2}-\frac{1}{N^*} & \frac{1}{2} & 0 & 0 & 0 \\ 0 & \frac{1}{4} & \frac{1}{2} & \frac{1}{4} & 0 & 0 \\ 0 & 0 & \frac{1}{4} & \frac{1}{2} & \frac{1}{4} & 0 \\ 0 & 0 & 0 & \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \\ 0 & \frac{1}{4} & 0 & 0 & \frac{1}{4} & \frac{1}{2} \end{Bmatrix} \tag{2}$$

where $N^* = \frac{4N_mN_f}{N_m+N_f}$ with N_m and N_f being the numbers of male and female breeders in each subpopulation. Note that the first term in the second row differs from the equivalent term in Nomura and Yonezawa (1996), where the term in the present notation is given as $\frac{1}{N^*}$. The equations as stated in Equation 2 can be reduced to lower dimension by exploiting that with n subpopulations, we have that $g_i = g_{n-i}$, where g_i is the probability of sampling IBD alleles from subpopulations i steps apart.

The eigenvalue effective population size (e.g. Ewens 1982) is then derived from the leading eigenvalue of the matrix describing the development of probabilities of non-IBD for various combinations of alleles. The eigenvalue effective populations size λ_1 is obtained by solving $N_e = 1 - \frac{1}{2N_e}$.

Coalescent approximation. The coalescent process (Kingman, 1982) describes genealogy of alleles in a population. It assumes that the population is of sufficient size such that at most one coalescence event takes place in any generation. Thus, the approximations are expected to work best for large populations. The expected rate of coalescence permits the definition of a coalescent effective population size. Asymptotically, the probability that two alleles are located in the same subpopulation is $\frac{1}{n}$. The probability of coalescence for two alleles in the same subpopulation (ignoring that alleles may be in the same individual) is $\frac{1}{2N^*}$. Therefore, the asymptotic probability of coalescence is $\frac{1}{2N^*n}$. In a panmictic, population the probability of coalescence is the same, i.e. when the coalescent model is valid and the population is large enough, the rate of coalescence and, hence, the coalescent effective population size are the same.

Analysis. All numerical analyses were done in R (R Core Team, 2021). Eigenvalues were computed for specific cases of numbers of breeding males and females in each subpopulation (represented by N^*=4 or 100). The equations leading to Equation 1 were iterated to illustrate the time course of the decline in outbreeding. Results were plotted as $\log(1 - g_i)$ against generation number. The value of the leading eigenvalue was compared to the eigenvalue effective population size of a panmictic population of dioecious individuals with nN_m males and nN_f females. For such a population the eigenvalue effective population size is given by

$$N_e^p = \frac{1-\frac{1}{N'}+\sqrt{1+\frac{1}{N'^2}}}{2} \tag{2}$$

where $N' = \frac{4nN_mN_f}{N_m+N_f} = nN^*$. Whenever N' is large, $N_e^p \approx N'$.

Results

The results for iterations where each subpopulation has N^*=4 or N^*=100 are shown in Figure 1. In both situations, there is a rapid initial build-up of relationship in each subpopulation (red and blue). In a panmictic population, this happens more slowly. For the small population in the long term (more than about 40 generations), inbreeding levels in the panmictic scheme exceed those in the circular scheme. The asymptotic rate of inbreeding in the panmictic scheme is faster than in the circular scheme. In contrast, in the large population, inbreeding levels in the circular scheme always remain higher. The asymptotic rates of inbreeding are indistinguishable. After the first generations, the IBD probability in the panmictic population is indistinguishable from the IBD probability for alleles sampled from subpopulations two steps apart. In Figure 1, all probabilities increase at the same (exponential) rate, but the curves for the different statistics are displaced by a number of generations with the within-population statistics leading the between-population statistics.

Discussion

Short-term behaviour. The short-term behaviour of the circular mating scheme produces relatively high levels of realised inbreeding (as measured by f) and coancestry inside subpopulations (as measured by g_0). As previously pointed out by Nomura and Yonezawa (1996), this is caused by the restriction of choice of males in each subpopulation. Sire and maternal grand-sire of any individual will often be closely related. In a panmictic population, sire and maternal grand-sire would be drawn from a large pool of males.

In many conservation schemes the primary interest tends to be on situations with very small populations. Hence, the situation demonstrated in the left panel of Figure 1 is important. This example is extreme. In this example, the circular mating scheme is superior for a period of up to 40 generations.

Long-term behaviour. As expected from the structure of the model, asymptotically all probabilities increase at the same rate, but they are displaced by a number of generations (by 5-10 generations in Figure 1).

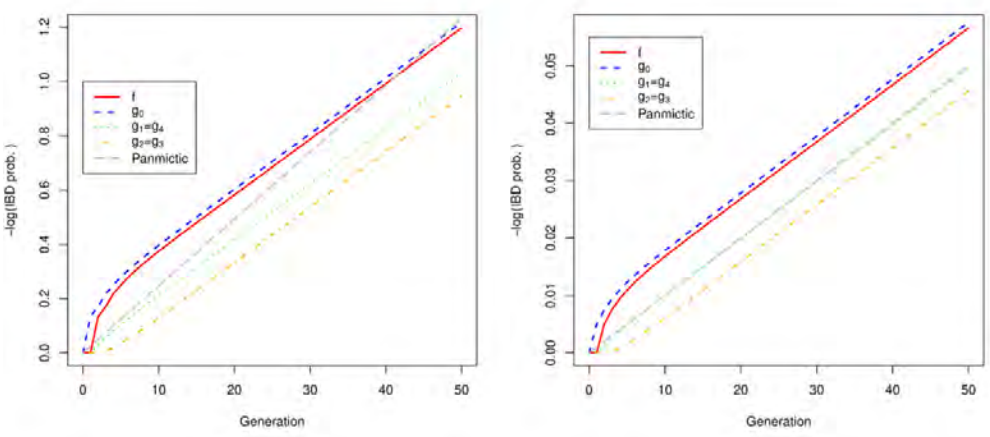


Figure 1. Time course of minus the logarithm of inbreeding levels for the first 50 generations for a circular mating scheme with N^*=4 or N^*=100 (right). f indicates the inbreeding level, whereas g_i are the IBD probabilities for alleles sampled from subpopulations i steps apart. Panmictic (dashed grey) indicates the expected increase in a panmictic population of dioecious individuals with the same number of breeders. In the right panel, panmictic line and the line for g_1 are nearly on top of each other. Note that the y-axes differ in scale.

The coalescent approximation shows that the asymptotic behaviour of the model is the same for the circular scheme and for a panmictic scheme whenever the coalescent model provides a valid description of the population. Primarily, this requires that the population should be large. This agrees with the analysis shown in the right panel of Figure 1. In this case, the behaviour of the IBD probability for alleles drawn from subpopulations two steps apart is indistinguishable from the behaviour of a panmictic population. For smaller populations the situation changes and the asymptotic rate of increase in inbreeding for circular schemes is slower than the rate observed in the panmictic population.

Choice of conservation schemes. Managing a circular mating scheme is more complex than a panmictic scheme. For populations with very small numbers of breeding animals, a circular mating scheme provides slower long-term rates of increase in inbreeding than does a panmictic scheme. However, this occurs at the expense of short-term increases in realised levels of inbreeding. Even for the very small population shown in the left panel of Figure 1, the advantage in lower long-term rate of inbreeding only manifests itself after about 40 generations. Up to that point, the realised inbreeding in a circular scheme is higher than in a panmictic scheme. The difference becomes smaller as populations get larger. In the right panel of Figure 1, it has disappeared altogether.

An additional advantage (not modelled here) of a circular scheme is that it provides some protection against (intentional or unintentional) overuse of particular animals (say, popular sires) in the breeding scheme. Disadvantages to circular schemes include susceptibility to errors. In addition, in numerically small schemes, subgroups may fail to produce males or females, which would have to be addressed by deviating from the scheme.

References

Ewens W.J. (1982) Theoretical Population Biology 21(3):373–378. https://doi.org/10.1016/0040-5809(82)90024-7
Kimura M, and Crow JF. (1963) Genet. Res., Camb. 4:399-415. https://doi.org/10.1017/S0016672300003797
Kingman J.F.C. (1982) J. Appl. Prob. 19A: 27–43. https://doi.org/10.2307/3213548
Leroy G. (2014) Animal Genetics 45(5):618–628. https://doi.org/10.1111/age.12178
Meuwissen T.H.E. (1997) Journal of Animal Science 75(4):934–940. https://doi.org/10.2527/1997.754934x
Nomura T., and Yonezawa K. (1996) Genetics Selection Evolution 28(2): 141–159. https://doi.org/10.1051/gse:19960202
R Core Team R (2021) A Language and Environment for Statistical Computing. Available at: https://www.R-project.org

450. Genomic rotational crossbreeding including advanced optimum contribution selection methods in local dairy cattle breeds

J. Stock[1*], H. Esfandyari[2], D. Hinrichs[3], R. Wellmann[1] and J. Bennewitz[1]

[1]Department of Animal Genetics and Breeding, University of Hohenheim, Garbenstr. 17, 70599 Stuttgart, Germany; [2]TYR, Storhamargata 44, 2317 Hamar, Norway; [3]Department of Animal Breeding, University of Kassel, Nordbahnhofstr. 1a, 37213 Witzenhausen, Germany; joana.stock@uni-hohenheim.de

Abstract

Many local dairy cattle breeds are facing extinction due to a large proportion of foreign genes which have been introgressed in the past. In addition, the performance gap to popular high-yielding breeds is increasing. In this simulation study it was investigated whether a rotational crossbreeding scheme including Optimum Contribution Selection methods can promote Angler, a typical local breed, as a crossing partner for the high-yielding breed German Holstein, while regaining the genetic uniqueness of Angler. Three scenarios that differed in the constraints for Optimum Contribution Selection were evaluated. The results showed that the reduction of migrant contributions e.g. from German Holstein in Angler slowed down their genetic gain remarkably. However, the potential of preservation of the breed increased. Not restricting the amount of migrant contribution or keeping it at a constant threshold resulted in much higher response to selection and thus might be most appealing for breeders that are primarily interested in their herd's performance.

Introduction

European Red Dairy Breeds (ERDB) usually have small population sizes and thus a slower rate of genetic gain than high-yielding breeds. This caused the performance gap between ERDB and high-yielding breeds to increase, especially since genomic selection is widely used in the latter. This led to the fact that many local breeds are facing extinction from introgression of high-yielding breeds, and replacement by high-yielding breeds for economic reasons. The introgression of foreign breeds to increase the performance of local breeds was common in the past, which has caused high foreign genetic material in many local breeds, i.e. Vorderwald cattle (Hartwig *et al.*, 2014), Angler, and Red and White Dual Purpose (Schmidtmann *et al.*, 2021). To recover the original genetic background, advanced Optimum Contribution Selection (aOCS, Wellmann 2019) methods can be used for selecting dams and sires as parents of the next generation. aOCS can decrease the level of migrant contributions and recover the original parts of a breed, while achieving genetic gain and controlling the rate of inbreeding at native alleles (Wang *et al.*, 2017). Nevertheless, the genetic gain achieved by aOCS methods might not be sufficient to compete with high-yielding breeds. The low to medium performance level of local breeds and the increasing performance gap to high-yielding breeds caused many farmers to switch to high-yielding breeds. In a previous study (Stock *et al.*, 2021) we showed that to promote ERDB and to improve their competitiveness the implementation of a genomic rotational crossbreeding scheme can be advantageous. ERDB benefited due to the enlarged reference population including both ERDB and crossbred animals, and the increased performance level of crossbred animals. The aim of the present simulation study was to examine the benefit of rotational crossbreeding schemes at the example of German Angler, as a typical ERDB, and German Holstein, as a typical high-yielding breed. Several and conflicting objectives were studied, i.e. the reduction of foreign genes in Angler, the increase in inbreeding in Angler, the genetic gain of Angler for purebred and for crossbred performance and maximisation of heterosis effects in crossbred offspring.

Materials & methods

A two-way rotational crossbreeding scheme for a ERDB (Angler) and a high-yielding breed (Holstein) was simulated, using the R package xbreed (Esfandyari and Sørensen, 2017). The genotypes (Illumina

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_450, © J. Stock *et al.* 2022

BovineSNP50 BeadChip, 50k) and pedigree data from 268 Angler cattle (Wang *et al.*, 2017) and 1,935 German Holstein (Streit *et al.*, 2013) were used as input to create the founder generations of 3,000 Angler and 6,000 Holstein animals. The first crossing was purebred Angler × purebred Holstein, followed by the rotational crossbreeding scheme, in which Angler and Holstein bulls were mated alternately to crossbred dams. Thus, purebred lines of Angler and Holstein were generated, as well. The number of animals in each population was kept constant, either 3,000 or 6,000, as well as a constant effective population size (N_e) of 100. Discrete generations were considered for both pure and crossbreeding schemes. Allocation of sires and dams was always random. Genetic gain, mean kinship, native kinship and migrant contributions in Angler and their rotational crossbred progeny with Holstein were examined in 3 different scenarios (Table 1). The simulated scenarios differed in the selection criteria of Angler. Selection of individuals to become parents was based on classical or advanced optimum contribution selection (OCS) methods. The classical OCS scenario maximized genetic gain while restricting the increase of mean kinship in the population. The advanced OCS (aOCS) scenarios additionally restricted the increase of native kinship and the increase of migrant contributions (MC) in the population. The overall optimization goal in each scenario was to maximize genetic gain, such as to maximize the breeding values. Genetic gain was measured in true genomic values (TGV), such as:

$$TGV_i = \sum_{j=1}^{1000} (a_{ij} + (2 - x'_{ij})d_{ij})x'_{ij}$$

where a_j and d_j are the true additive and dominance effects of the j^{th} QTL, and x'_{ij} = 0, 1, or 2, depending on the respective genotype.

Results & discussion

As shown in Figure 1 Angler performed best in Scenario 1, in which only the mean kinship was restricted, and worst in Scenario 3, in which the native kinship was constrained and the MC was defined to decrease 1% in each generation. This result shows that genetic gain was in conflict with the genetic uniqueness in this population. In addition, it shows the importance of Holstein contribution to the genetic gain in Angler. Crossbreds slightly outperformed Angler and Holstein in the beginning. They performed best and most stable in Scenario 1 and 2 and then converged to the TGV of Holstein as of generation 5. In Scenario 3 the genetic gain of Angler started to flatten by generation 5 and thus caused the TGV of crossbreds to drop notably. Decreasing the Holstein contribution in Angler slowed the increase of the mean kinship in the Angler population.

Thus, constraining native kinship and reducing MC (Figure 2) resulted in a higher level of genetic diversity in the Angler population, as shown by the least decrease of additive variance (Figure 3). The overall reduced genetic gain of Angler in Scenario 3 led to the largest difference between the performance of the parental breeds Angler and Holstein. The slower rate of genetic gain of Angler arises from the removal of the Holstein contributions, and from the additional constraints that were imposed. In addition, these

Table 1. Overview of simulated scenarios.

Scenario	Selection method of Angler (optimization goal)	Constraints
1	OCS (genetic gain)	Mean kinship (max. ΔF = 0.5%)
2	aOCS (genetic gain)	Native kinship (max. ΔF = 0.5%), Migrant contribution (as in founder generation)
3	aOCS (genetic gain)	Native kinship (max. ΔF = 0.5%), Migrant contribution (-1% each generation)

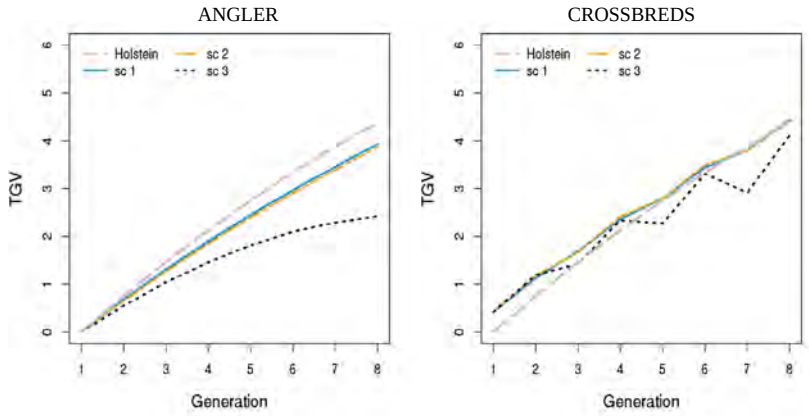


Figure 1. True genomic values (TGV) of Angler and Holstein (left), crossbred animals and Holstein (right) over the simulated time period of 8 generations of rotational crossbreeding. Sc1 = Scenario 1, sc 2 = Scenario, sc 3 = Scenario 3.

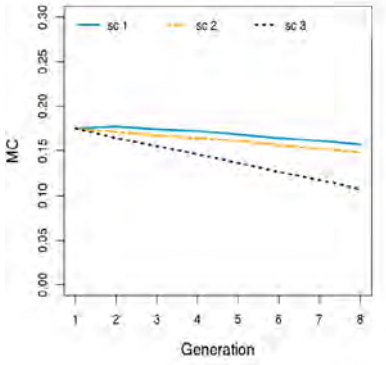


Figure 2. Amount of migrant contribution (MC) from Holstein in the Angler population over the simulated time period of 8 generations of rotational crossbreeding. sc1 = Scenario 1, sc2 = Scenario, sc3 = Scenario 3.

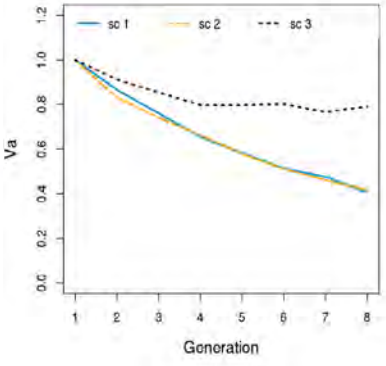


Figure 3. Additive variance (V_a) in Angler over the simulated time period of 8 generations of rotational crossbreeding. sc1 = Scenario 1, sc2 = Scenario 2, sc3 = Scenario 3.

constraints reduced availability of animals to become parents, leading to a low selection intensity and thus loss of opportunity for selection and consequent reduced rate of genetic gain in this scenario.

The difference between the genetic gain of Holstein and Angler was too large to achieve benefits from rotational crossbreeding by the exploitation of heterosis. Hence, the additional heterosis arising from making the breeds genetically more distinct did not compensate for the reduced rate of genetic gain. In addition, the large differences in the performance of crossbred animals from generation to generation can be problematic in practice, i.e. regarding the management of the herd or economic factors.

Conclusions

The results confirmed that the implementation of a genomic rotational crossbreeding scheme could be an interesting option to promote numerically small breeds like German Angler, if a suitable partner breed is available for such a scheme. If this is the case, this might increase their attractiveness, and thus might

reduce the risk of numeric extinction. The results further showed that it is in general possible to combine a genomic rotational crossbreeding scheme with the maximization of genetic gain, while simultaneously restricting the increase of inbreeding and reducing migrant contribution. The latter one turned out to be the most compromising restriction with regard to genetic gain in purebreds, and resulted in remarkable waves of genetic gain observed in the crossbred individuals. The reduction of migrant contribution with the main aim to increase heterosis-effects in crossbreds is, however, not a recommendable breeding strategy. This is because the possible gain in heterosis cannot compensate for the reduced genetic gain imposed by the additional OCS restrictions.

Acknowledgements

This work was financially supported by the German Federal Ministry of Food and Agriculture (BMEL) through the Federal Office for Agriculture and Food (BLE), grant number 2817ERA10D. The project has received funding from the European Union's Horizon 2020 Research and Innovation Program under grant agreement no. 696231 – ReDiverse (Biodiversity within and between European Red Dairy Breeds). Joana Stock was partly supported by the H. Wilhelm Schaumann Foundation, Hamburg, Germany, which is gratefully acknowledged.

References

Esfandyari H. and Sørensen A.C. (2017) Proc. of the 68th EAAP. Tallinn, Estonia.

Hartwig S., Wellmann R., Hamann H. and Bennewitz J. (2014) J. Anim. Breed. Genet. 131, 496–503. https://doi.org/10.1111/jbg.12099.

Schmidtmann C., Schönherz A., Guldbrandtsen B., Marjanovic J., Calus M., *et al.* (2021) Genet. Sel. Evol. 53, 6. https://doi.org/10.1186/s12711-021-00613-6.

Stock J., Esfandyari H., Hinrichs D. and Bennewitz J. (2021) J. Dairy Sci. 104, 6873–6884. https://doi.org/10.3168/jds.2020-19927.

Streit M., Wellmann R., Reinhardt F., Thaller G., Piepho H.-P. *et al.* (2013) G3 (Bethesda) 3, 1085–1093. https://doi.org/10.1534/g3.113.006536.

Wang Y., Bennewitz J. and Wellmann R. (2017) Genet. Sel. Evol. 49, 45. https://doi.org/10.1186/s12711-017-0320-7.

Wellmann R. (2019) BMC 20, 25. https://doi.org/10.1186/s12859-018-2450-5.

451. Allele frequency differences at epistatic QTL explain different genetic trends in number of teats in two pig lines

C.A. Sevillano[1*], B. Harlizius[1], M.F.L. Derks[1,2], M.S. Lopes[1,3], M. van Son[4] and E.F. Knol[1]

[1]Topigs Norsvin Research Center, P.O. Box 43, 6640 AA Beuningen, the Netherlands; [2]Wageningen University & Research, P.O. Box 338, 6700 AH Wageningen, the Netherlands; [3]Topigs Norsvin, Visconde do Rio Branco 1310, 80.420-210 Curitiba, Brazil; [4]Norsvin, Storhamargata 44, 2317 Hamar, Norway; claudia.sevillanodelaguila@topigsnorsvin.com

Abstract

Several QTL regions affecting number of teats have been detected in commercial pig lines. In this study we follow the indirect effects of index selection on different QTL regions for number of teats in two maternal lines in 40K animals with imputed 555K SNP. In total, 3 QTL regions overlap between the two populations. For a QTL on *Sus scrofa* chromosome 7 (*SSC7*), the underlying functional variation affecting number of thoracic vertebrae located in the vertnin gene has also been genotyped showing an allelic substitution effect of nearly 0.4 teats in both lines. However, allele frequencies show an opposite trend at the vertnin gene locus in the two lines. Moreover, epistatic effects between two QTL regions on *SSC7* and *SSC12* are investigated at the molecular and phenotypic level.

Introduction

Some mammals, such as mice and pigs, show variation in the number of mammary glands. In pigs, number of teats (NTE) are monitored in commercial breeding and are part of the overall selection index. The aim of this work is to identify how the selection index indirectly affects QTL regions for NTE. In a previous GWAS (van Son *et al.* 2019), we showed in different lines the segregation of the two variants at the vertnin *(VRTN)* locus which is known to cause the development of additional thoracic vertebrae and ribs. Every rib has the potential to develop an additional mammary gland and therefore also explains the effect of this QTL on NTE (Veltmaat, 2017). In addition, several other QTL regions for NTE have been reported in pigs for which the molecular basis remains unknown (Moscatelli *et al.,* 2020, Zhuang *et al.,* 2020). In this study, we identified two additional QTL in our maternal lines on *SSC10* and *SSC12.* At all three QTL, allele frequencies change over time, however, for some loci allele frequencies change in the opposite direction in the two lines. Moreover, the allelic effects of the second largest QTL located at *SSC12* differs between homozygous genotypes at the *VRTN* locus on *SSC7.* This difference can be due to distinct interactions of the *VRTN* locus with different local genetic background, i.e. epistasis (Mackay, 2014).

Materials & methods

Phenotypes, genotypes and pre-processing. We evaluated NTE in two pig populations L (Landrace) and LW (Large-White based). NTE was recorded at birth on both males and females. The L population animals were born between 2013 and 2020. The LW population animals were born between 2009 and 2020.

Two datasets from each population were used: ALL and GENOTYPED. The dataset ALL consisted of all genotyped animals and their contemporaries that had phenotypes (around 300,000 animals per line). Using ALL, the phenotypes were pre-corrected for all non-genetic effects (sex, heard-year, and litter). The non-genetic effects were estimated with a pedigree-based linear model in ASReml v3.0 (Gilmour *et al.*, 2009). The dataset GENOTYPED was a subset of ALL consisting of all animals that had both phenotypes and genotypes (39,956 L and 39,945 LW). This dataset was used to perform the GWAS. All GENOTYPED animals were genotyped on Illumina GeneSeek custom SNP chips (25K, 50K or 80K) (Lincoln, NE, USA). Genotypes of the animals were imputed within population using Fimpute v3 (Sargolzaei *et al.*, 2014) to

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_451, © C.A. Sevillano *et al.* 2022

a set of 48,075 SNPs. In addition, some boars were genotyped using the Axiom porcine 660K array from Affymetrix (Affymetrix Inc., Santa Clara, CA, United States). A second imputation was performed within population using Fimpute v3 (Sargolzaei *et al.*, 2014) from the 48,075 SNPs set to a set of 554,768 SNPs. Quality control was performed at each level of imputation as described in Van Son *et al.* (2019).

Genome-wide association analysis. A single-SNP GWAS was performed with the GENOTYPED dataset within each population using a linear animal model in GCTA software (Yang *et al.*, 2011). Only SNPs with a MAF≥0.01 were used in this step. Significant SNPs and QTL were detected using a $P<1.0\times10^{-8}$.

Results

The mean value for NTE is around 1 teat higher in the L compared to LW (16.72 vs 15.85 for GENOTYPED animals). The L line has gained 1 extra teat in the last 8 years, while the LW line has only gained 0.6 extra teat in the last 10 years (Figure 1). The heritability (h^2) was 0.39 for L and 0.34 for LW.

The GWAS results for NTE in both lines clearly show the QTL on *SSC7* which has been reported earlier to be associated with NTE (van Son *et al.* 2019). In addition, 5 other QTLs segregate on several other chromosomes in the L line, and another 4 QTLs in the LW line. Only 3 QTL regions overlap between the two populations, including the QTL on *SSC7* and the QTL regions on *SSC10* and *SSC12*. The position of the most significant SNP (top SNP) differs slightly between populations for these 3 QTL regions. On *SSC7*, the *VRTN* promoter SNP (VRTN_RS709317845) is the most significant in LW, while the SNP AX-116329721 is the top SNP in L. However, the *VRTN* promoter SNP and the AX-116329721 SNP are in high LD (r^2) (0.98 in L and 0.91 in LW). On *SSC10*, the top SNPs for each line are 124 kb apart and they show low LD in both populations (0.004 in L and 0.002 in LW). On *SSC12*, the top SNPs for each line are 96 kb apart and they are in LD in L population (0.67) but not in LW population (0.002).

In Table 1, we show the parameters for the top SNPs in the QTL regions of *SSC7* and *SSC12*, as they have the largest effect for NTE. For *SSC7,* we limited to the *VRTN* promoter SNP as it is in high LD with the other top SNP in the same region. The top SNP in *SSC12* for line L (AX-116444814) is not present in the GWAS results in LW line due to MAF lower than 0.01. The frequency of the allele C of the *VRTN* promoter SNP increasing NTE is more than four times higher in the L line compared to LW (0.75 vs 0.17) (Table 1). Figure 2 shows that the frequency has been increased over the years in L line, whereas it is dropping in the LW

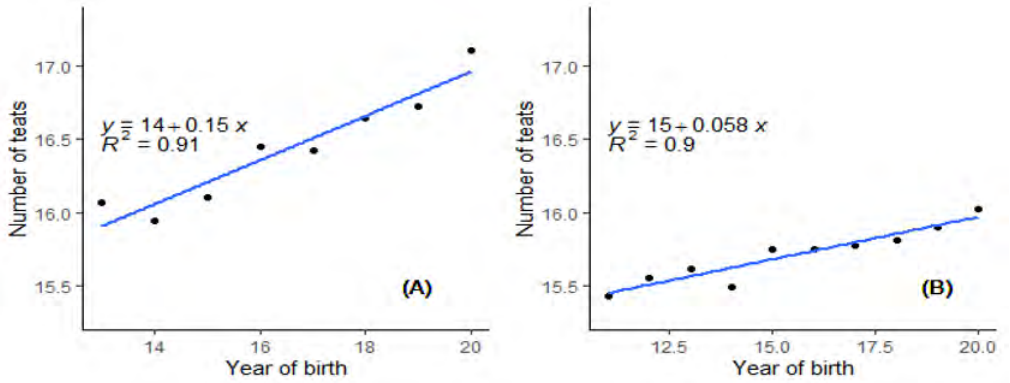


Figure 1. Phenotypic trend for number of teats. (A) Landrace population from 2013 to 2020. (B) Large White population from 2011 to 2020.

Table 1. Parameters for the SNPs in the 2 top QTL associated with NTE in each line.

Line[1]	SSC	SNP	Position (Mb)	*P*-value	Freq.[2]	Effect[2]	SD[2]
L	7	VRTN_RS709317845	98	3.05e-58	0.75	0.37	0.02
	12	AX-116444814	51	4.97e-37	0.17	0.29	0.02
	12	AX-116758961	51	2.68e-10	0.22	0.13	0.02
LW	7	VRTN_RS709317845	98	3.11e-97	0.17	0.38	0.02
	12	AX-116444814	51	NA	0.01	NA	NA
	12	AX-116758961	51	2.73e-41	0.70	0.16	0.01

[1] Line Landrace (L) and Large White (LW).

[2] Relative to the allele increasing NTE.

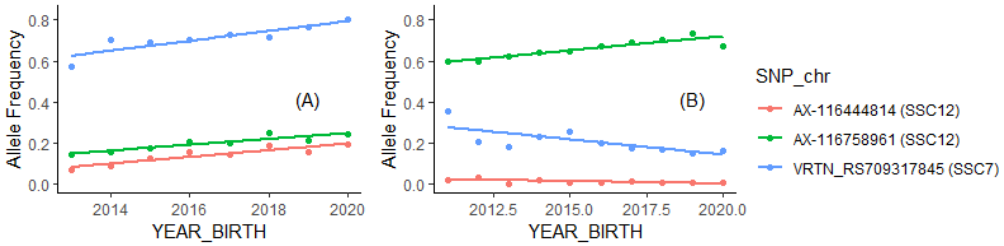


Figure 2. Allele frequency trend for three SNPs associated with number of teats. (A) Landrace population from 2013 to 2020. (B) Large White population from 2011 to 2020.

line. At the same time, the allele increasing NTE on *SSC12* is slightly increasing in both lines. The allele of the top SNP increasing NTE on *SSC12* for LW is much more frequent in LW compared to L (0.22 vs 0.70).

Table 2 shows that allele substitution effects at the *SSC12* QTL are dependent on genotype at the *VRTN* locus on SSC7 and differ between lines. In general, the size of the effect is much higher in *VRTN*-CC animals than in *VRTN*-AA animals. The top SNP in LW (AX-116758961) at the QTL on *SSC12* has a much larger effect in LW than in L but only in *VRTN*-AA animals (0.14 vs 0.06). At the same time, the top SNP in L (AX-116444814) at the QTL on *SSC12* has a lower effect in LW than in L, especially for *VRTN*-AA animals where the effect is even negative (-0.07).

Discussion

This study shows that selection acts on different QTL for NTE in the two sow lines although three QTLs overlaps across lines. However, epistatic effects seems to be playing a role between the two most significant QTL on *SSC7* and *SSC12* that overlap across the two lines. In both lines, CC animals at the *VRTN* locus have a higher genetic potential to develop additional teats than AA animals. Furthermore, the allele frequencies and their trend are the opposite in both lines. At the *VRTN* locus, the allele increasing NTE is going to fixation in L and rather low and declining in LW. At the *SSC12* QTL, the frequency of the allele increasing NTE is very high (0.70) in LW while it is only 0.20 in L. In addition, the top SNP in both lines are different and not in LD in LW. Altogether, this results in different selection pressure at the two QTL. A large proportion of the L animals carry the C allele increasing NTE whereas more than 50% of the LW animals are homozygous for the A allele. Therefore, the unknown causal mutation at the QTL on *SSC12* contributes more to the overall genetic potential and selection space for NTE in LW. Finally, the allele increasing NTE at the top SNP of the *SSC12* QTL in L is hardly present in LW and does not increase NTE in *VRTN*-AA

Table 2. Allelic effects of the largest QTL on SSC12 according to genotypes at the *VRTN* locus on *SSC7*.

Line	SNP	*VRTN*		N of animals[1]	
		CC	AA	CC	AA
L	AX-116444814	0.32	0.12	22,569	2,521
	AX-116758961	0.15	0.06		
LW	AX-116444814	0.27	-0.07	1,245	27,358
	AX-116758961	0.16	0.14		

[1] Number of animals per *VRTN* genotype.

animals. In any case, we clearly show an increased genetic potential to develop additional teats in *VRTN*-CC animals. Altogether, this example shows that rare variants are neglected by genomic selection and could get more emphasis by marker-assisted selection approaches in the breeding program. Additional analyses including data on number of vertebrae and ribs, as well as detailed sequence data, should help to pinpoint the molecular basis of this increased selection space of the *VRTN*-C allele for NTE.

References

Gilmour, A. R., Cullis, B. R., Gogel, B. J., Welham, S. J., and Thompson, R. (2009). ASReml User Guide Release 3.0. VSN International Ltd, England.

Mackay TF. Epistasis and quantitative traits: using model organisms to study gene–gene interactions. Nat Rev Genet. 2014;15:22. https://doi.org/10.1038/nrg3627

Moscatelli G, Dall'Olio S, Bovo S, Schiavo G, Kazemi H, *et al.* Genome-wide association studies for the number of teats and teat asymmetry patterns in Large White pigs. (2020) Animal genetics 51(4):595-600. https://doi.org/10.1111/age.12947

Sargolzaei, M., Chesnais, J. P., and Schenkel, F. S. A new approach for efficient genotype imputation using information from relatives. (2014) BMC Genomics 15:478. https://doi.org/10.1186/1471-2164-15-478

Van Son M, Lopes MS, Martell HJ, Derks MF, Gangsei LE, *et al.* A QTL for number of teats shows breed specific effects on number of vertebrae in pigs: bridging the gap between molecular and quantitative genetics. (2019) Frontiers in genetics 10:272. https://doi.org/10.3389/fgene.2019.00272

Veltmaat, J. M. (2017). Prenatal Mammary Gland Development in the Mouse: Research Models and Techniques for Its Study from Past to Present. In: Martin F., Stein T. and Howlin J. (eds) Mammary Gland Development. Humana Press, New York, NY, pp. 21–76. https://doi.org/10.1007/978-1-4939-6475-8_2

Yang, J., Lee, S. H., Goddard, M. E., and Visscher, P. M. GCTA: a tool for genome-wide complex trait analysis. (2011) The American Journal of Human Genetics 88:76–82. https://doi.org/10.1016/j.ajhg.2010.11.011.

Zhuang, Z., Ding, R., Peng, L., Wu J, Ye Y., *et al.* Genome-wide association analyses identify known and novel loci for teat number in Duroc pigs using single-locus and multi-locus models. BMC Genomics (2020) 21:344. https://doi.org/10.1186/s12864-020-6742-6

Part 26

Developing countries

452. Phylogenetic analysis of local cattle population in northern province of Sri Lanka

K.G.C.B. Wijebandara[1], M.S. Kurukulasuriya[1], P.C.G. Bandaranayake[2] and G.L.L.P. Silva[3*]

[1]Department of Animal Science, Faculty of Animal Science and Export agriculture, Uva Wellassa University of Sri Lanka, Badulla, 90000, Sri Lanka, [2]Agricultural Biotechnology Centre, Faculty of Agriculture, University of Perdeniya, 20400, Sri Lanka, [3]Department of Animal Science, Faculty of Agriculture, University of Perdeniya, 20400, Sri Lanka; pradeepas@agri.pdn.ac.lk

Abstract

The present study focused to identify phylogeny of phenotypically short, local cattle population found in Northern Province using mitochondrial Cytochrome Oxidase Sub-unit I. Two cattle populations; Mannar and Jaffna cattle from Northern Province were included in the study. The results showed that Northern local cattle have low genetic polymorphism. Phylogenetic analysis of Northern local cattle and *Bos indicus* clustered together in a separate group. The shortest evolutionary divergence reported between Mannar and Jaffna cattle identifying them in one population, distributed in two geographical destinations. Morphometric parameters of Northern local cattle showed that they are phenotypically smaller compared to other local cattle populations in Sri Lanka. However, their phylogenetic relationship with the other local cattle in Sri Lanka is yet to be identified.

Introduction

Among the cattle breeds in Sri Lanka, Lankan (*Batu Haraka)*, Thawalam and Thamankaduwa cattle breeds have been identified as the indigenous breeds (Chandrasiri, 2002). Apart from the identified breeds, a phenotypically short cattle population has been reported in Northern Province reared for milk, meat, draught and manure. A small proportion of males in this population are reared mainly for breeding and traditional gaming activities (Dias, 2019).

Due to marginal contribution to milk production and less use of them in agricultural activities, cattle in Northern Province have been subjected for genetic improvement, especially after the civil war (Szent-Ivanyi, 2017). The valuable adaptive characteristics of the native cattle breeds were neglected during the breeding activities due to the production-oriented breeding (Abayawansa, 1992). Several studies have been conducted with local (*Batu Haraka)*, Thawalam cattle and Thamankaduwa cattle populations for morphological and quantitative trait analysis (Silva *et al.*, 2010) and genetic analysis (Senavirathne and Lokugalappatti, 2020). However, the indigenous cattle population in the Northern Province of Sri Lanka are yet to be genetically characterized. Therefore, the present study was conducted to identify the phylogeny of Northern cattle population in Sri Lanka and to identify the population specific phenotypic characters.

Materials & methods

Sample collection. Sampling sites were selected according to the information gathered from the Northern Provincial Department of Animal Production and Health on availability of non-crossbred, non-descriptive, short local cattle. A total of 34 blood samples were collected from external jugular veins of 15 cattle (six males; nine females) from Mannar district (Nanattan, Manthai West Veterinary regions) and 19 cattle (seven males; 12 females) from Jaffna district (Maruthankarny Veterinary region) into EDTA tubes for DNA extraction. The selected animals were at two to three years of age. The cattle breed was confirmed using phenotypic characters and the information gathered from the farmers and the Veterinary Office regarding breeding activities. In order to ensure the genetic non relatedness, maximum of two genetically non-related animals were selected from a single farm. Phenotypic information on height at withers, rump

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_452, © K.G.C.B. Wijebandara *et al.* 2022

height, heart girth, coat colour, hair coat type and presence or absence of horns of the selected animals were recorded.

Total genomic DNA were extracted from EDTA blood using the MasterPure™ DNA Purification Kit. Agarose gel electrophoresis and nanodrop spectrophotometry were used for the visualization and quantification of the extracted DNA.

Amplification of mitochondrial cytochrome oxidase sub-unit I (COI) using polymerase chain reaction (PCR) and sequencing. The complete mitochondrial COI region was amplified using the primers: (F:5'-GGTCAACAAATCATAAAGATATTGG-3';R:5'-TAAACTTCAGGGTGACCAAAAAATCA-3') and expected PCR product size is 680 bp. Polymerase Chain Reaction was carried out with 150-175 ng genomic DNA in a 25 µl reaction volume. The reaction mixture contained 14.9 µl of nuclease free water, 5.0 µl PCR Buffer 5 X, 1.5 µl 25 mM $MgCl_2$, 0.5 µl 10 µM dNTP mix, 0.5 µl each of 10 µM primer, 0.1 µl 5U/ µl Taq DNA polymerase and 2.0 µl of genomic DNA. The PCR conditions were: 1 min at 94 °C for 60 S, 51 °C for 90 S and 72 °C for 90 S and completed with final step at 72 °C for 5 min, then cooled to 10 °C. The size of amplified product was confirmed by 2% agarose gel electrophoresis and sequenced by Sanger sequencing principle using the service provided by Macrogen's sequencing service, South Korea.

Data analysis. The DNA sequences were edited and aligned using ClustalW version 2.0 programme (Larkin *et al.*, 2007) of CodonCode aligner software (Version 9.0.1; CodonCode Corporation, Centerville, MA, USA). Sequence sites representing gaps and unspecified bases were excluded from the analysis and seven COI sequences from the study population were obtained for final analysis. The reference sequences of genus '*Bos*' from several geographical locations were downloaded from NCBI database. The reference sequences include 10 *Bos taurus* breeds (NC_006853.1 *B. taurus*-Reference, DQ124411.1 *B. taurus*-Holstein, AY676857.1 *B. taurus*-Angus, AB074963.1 *B. taurus*-Japanese black, DQ124377.1 *B. taurus*-Korean native, EU177816.1 *B. taurus*-Chianina, EU177835.1 *B. taurus*-Iraqi, EU177849.1 *B. taurus*-Greek, EU177858.1 *B. taurus*-Iranian, FJ971088.1 *B. taurus*-Mongolian), seven *B. indicus* breeds (NC_005971.1 *B. indicus*-Reference, KX575711.1 *Bos indicus*–Sahiwal, AY126697.1 *B. indicus*–Nellore, MF667929.1 *B. indicus*–Vechur, MK335920.1 *B. indicus*–Trinket, MF667930.1 *B. indicus*–Kasargode, MF667931.1 *B. indicus*–Wayanad) and six ancestral and non-prominent breeds (NC_024818.1 *Bos gaurus*, NC_025563.1 *Bos mutus*, NC_013996.1_1 *Bos primigenius*, NC_012706.1 *Bos javanicus*, NC_006380.3_1 *Bos grunniens*, NC_036020.1 *Bos frontalis*).

Nucleotide polymorphism was identified by software DnaSP version 5.1 (Librado and Rozas, 2009). The phylogenetic clusters were identified by maximum likelihood method with 1000 bootstrap replications using the software MEGA X (Kumar *et al.*, 2018).

All reference samples were included into four groups; Yak, *B. taurus*, *B. indicus* and Gaur according to Felius *et al.* (2014) and Mannar and Jaffna cattle populations are considered as two populations for presenting the evolutionary divergence (mean genetic distances). The evolutionary divergence over sequence pairs between groups were estimated based on Tamura-Nei model using MEGA X software. The morphometric data of phenotypic information were subjected to basic descriptive statistical analysis using Minitab 16 software.

Results

Final analysis includes a total of seven sequences (609 bp in length) from the samples (C2, C5, C8 and C13 from Mannar cattle population and C17, C18 and C28 from Jaffna cattle population) and 13 reference sequences. The mean percentages of the four nitrogen bases were 27.7% for adenine (A), 27.2% for thymine

(T), 28.8% for cytosine (C), and 16.3% for guanine (G). A total of 58 polymorphic sites were identified and 53 of them were parsimony-informative sites while having five singleton variable sites. However, Northern cattle population had only one polymorphic site.

Phylogenetic analysis. Local cattle from Northern Sri Lanka were clustered together with *B. indicus* separating from *Bos taurus* cattle creating two monophyletic clades (Figure 1). *B. indicus* were placed in a one clade with Northern local cattle in Sri Lanka, however, they can be clearly identified as a separate group in the phylogenetic tree. Further, the evolutionary divergence over sequence pairs between groups showed, lowest mean genetic distance between Jaffna and Mannar cattle populations (0.0004) while having 0.0037 and 0.0041 mean genetic distances between Jaffna and *B. indicus* populations, and Mannar and *B. indicus* populations, respectively.

Morphological analysis. Local cattle in Northern Province are short and small animals (Table 1). According to the qualitative characters, local cattle from Northern Province were having uniform coat colours. The coat colour varied from among red, white, black, brown and grey colours, however, predominantly they were in white (38.10%) and brown (33.33%) colours. Their hair is short and straight. The shape of the body was short and stocky. Mostly animals had horn (90.43%).

Discussion

According to the evolutionary history, the local cattle found in Northern Province are inheriting from *B. indicus* cattle. Low genetic diversity and separating as a group in the *B. indicus* clade reflects that this group

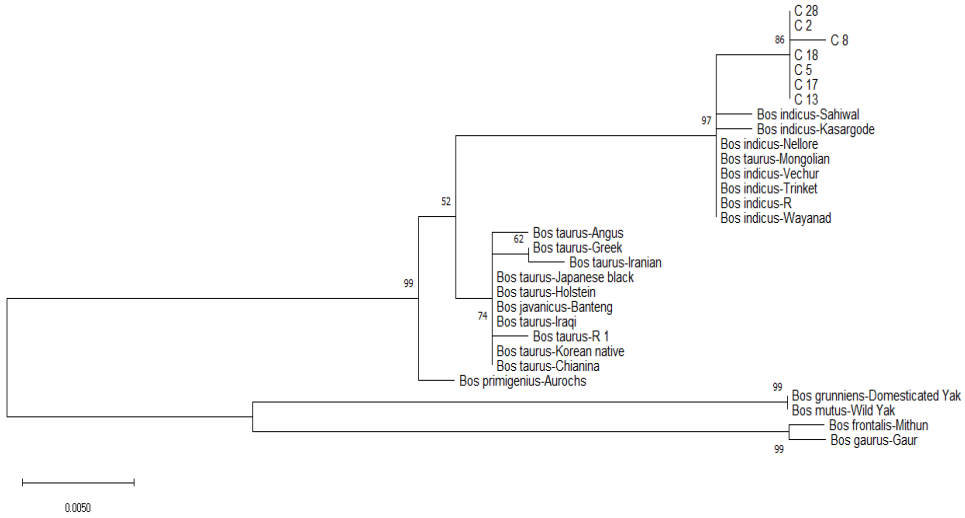


Figure 1. Maximum likelihood tree displaying the variation of cattle breeds based on phylogenetic analysis of the COI gene.

Table 1. Morphometric parameters of Northern cattle population (mean ± SD).

Gender	Height at withers	Rump height	Heart girth
Male	90.05±3.03	97.44±1.54	110.00±2.45
Female	88.14±1.52	94.49±2.17	105.20±3.44

of cattle might have developed phenocopies owing to the environmental conditions prevailing in the area in which they also have subjected to isolated breeding. These two actions might have resulted their small body conformation leading to produce distinct group of cattle unique to the area. The shortest evolutionary divergence was reported between cattle populations in Mannar and Jaffna emerging as one population, distributed in two geographical destinations. Indigenous livestock breeds are known for their hardiness, survival on little inputs and adaptability to the variable environments and disease resistance. Such breeds are therefore believed to be more economically efficient than other breeds when raised in their home tract (Ayalew *et al.,* 2003). Therefore, the Northern local cattle could have evolved over generations under low level of selections followed in traditional animal husbandry system in the area. As a result, Northern local cattle population has evolved with the adaptability to thrive in harsh native environment, showing resistance to tropical diseases and parasites and sustaining on low quality roughages and grasses. Morphometric parameters reflect that these local cattle populations in Northern Province are phenotypically small animals compared to the other local cattle populations in Sri Lanka which is in accordance with the results reported by Silva *et al.* (2010). However, their phylogenetic relationship with the other local cattle populations in Sri Lanka is yet to be identified.

References

Abayawansa W.D., Abeygunawardena H., Jayatilaka M.W.A.P., and Ratnayake D. (1992). Tropical Agricultural Research 4:66-80.

Ayalew W., Rischkowsky B., King J.M., and Bruns E. (2003) Agric. Syst. 76:1137-1156. https://doi.org/10.1016/S0308-521X(02)00033-1

Chandrasiri A.D.N. (2002) Country Report on the State of Animal Genetic Resources (Sri Lanka). The Food and Agriculture organization of the United Nations, Rome.

Dias V.C. (2019) Cattle Farming Systems in Mannar District, Undergraduate Research Dissertation, Uva Wellassa University, Badulla, Sri Lanka.

Felius M., Beerling M.L., Buchanan D.S., Theunissen B., Koolmees P.A *et al.* (2014). Diversity 6(4):705-750. https://doi.org/10.3390/d6040705

Larkin M.A., Blackshields G., Brown N.P., Chenna R., McGettigan P.A *et al.* (2007) Bioinformatics 23(21): 2947-2948. https://doi.org/10.1093/bioinformatics/btm404

Librado P., and Rozas J. (2009) Bioinformatics 25(11):1451-1452. https://doi.org/10.1093/bioinformatics/btp187

Senavirathne K.M.S.M.K., and Lokugalappatti L.G.S. (2020) Proc. of the Undergraduate Research Symposium of the Faculty of Veterinary Medicine and Animal Science, University of Peradeniya, Sri Lanka.

Silva P., Jianlin H., Hanotte O., Chandrasiri A.D.N. and Herath H.M.S.P. (2010). Indigenous cattle in Sri Lanka: production systems and genetic diversity. In: Odongo N.E., Garcia M. and Viljoen G.J. (eds) Sustainable Improvement of Animal Production and Health. Food and Agriculture Organization of the United Nations, Rome, Italy, pp. 45-48.

Szent-Ivanyi T, A Fresh Start with Milk (2017). Available at: https://akzente.giz.de/en/artikel/fresh-start-milk.

453. Accounting for heterogeneity of data in implementing genomic selection models applicable to crossbred cattle in India

A. Guintard[1*], L. Bourguignon[2], G. Restoux[1] and V. Ducrocq[1]

[1]Université Paris Saclay, INRAE, AgroParisTech, GABI, 78350 Jouy-en-Josas, France; [2]Institut Agro (Agrocampus Ouest), 65 rue de Saint Brieuc, 35 042 Rennes, France; ana.guintard@inrae.fr

Abstract

Indian dairy production takes place in very diverse environments. The objective of this study was to identify genotype by environment interactions in the context of establishment of a reference population for genomic selection. Data from 9,265 cows in six states were analysed. The high variability in the distribution of breed types and milk yield across India, as well as the wide range of climatic zones existing in the country are highlighted. This heterogeneity must be taken into account, defining, for example, groups of districts (administrative regions within states), rather than considering India as one homogeneous country, in order to better predict the genomic merit of individuals in their local environment.

Introduction

The dairy sector plays a major role in India, contributing 5.3% of the gross domestic product, and representing an essential source of animal protein in a country where one-third of the population is vegetarian (Belhekar and Dash, 2016). Improving productivity at the animal level is necessary to meet the demand and to ensure self-sufficiency (Chawla, 2009). Genomic selection is a way to increase the milk production in Indian smallholder farming systems (Al Kalaldeh *et al.*, 2021). To implement it, BAIF Development Research Foundation (https://baif.org.in/) has been collecting performances and genotypes for the past four years to set up a reference population. However, the heterogeneity of collected data reflects a huge diversity of production systems over India (across states, climatic zones, herd size, breeds, etc.). Thus, the aim of this study was to analyse phenotypic data from BAIF's database in order to characterise the heterogeneity of performances to account for in the establishment of a reference population for genomic selection.

Materials & methods

Data used for this study originates from the BAIF database and encompass phenotypic data recorded in the states of Andhra Pradesh, Bihar, Gujarat, Jharkhand, Maharashtra, Odisha, Punjab, Rajasthan and Uttar Pradesh between August 2016 and November 2021. Animals were crossbred cattle ranging from pure indigenous *Bos indicus* to pure exotic *Bos taurus* (Holstein or Jersey) cows. The database, contains information on daily milk yield, day in milk, lactation number, last calving date, breed type, the type of recording (by an enumerator or by the farmer himself) the farm, the CDC (Cattle Development Centre, which is a local centre providing services such as artificial insemination for eight to ten villages), the district and the state.

In order to make the dataset suitable for analysis, we filtered the data as follows: (1) Andhra Pradesh and Gujarat states were removed because data were available only on pure zebus (breeding tracks), and were not representative of the diversity of breeds in these states; (2) data from cows with fewer than four records per lactation and with a lactation lasting less than 200 days were removed; (3) data were retained only if they were recorded between eight and 340 days after calving; (4) data from cows with a coefficient of variation in daily milk yield greater than 0.8 on a given lactation or a daily milk yield greater than 25 kg were removed, being more than three times greater than the average daily milk yield (Al Kalaldeh *et al.*, 2021); (5) data from cows with a lactation rank greater than seven were deleted; and (6) for each CDC and breed, data were deleted if they were recorded on fewer than ten cows.

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_453, © A. Guintard *et al.* 2022

Weather data (from worldweatheronline.com), temperature, humidity and precipitation were also considered at different levels of season which were defined as: winter (January to February), pre-monsoon (March to May), monsoon (June to September) and post-monsoon (October to December).

After data filtering and integration of weather data (not available for all the districts), 211,545 records from 9,265 cows in six states (Bihar, Jharkhand, Maharashtra, Odisha, Punjab and Uttar Pradesh), 26 districts, 75 CDC and 4,168 farms were retained for the analysis. The number of cows per herd ranged from one to 47 with an average of 2.2. Overall, 44% of the herds were composed of a single cow while 85% of the herds had three cows or less. The number of cows per CDC varied from ten to 456.

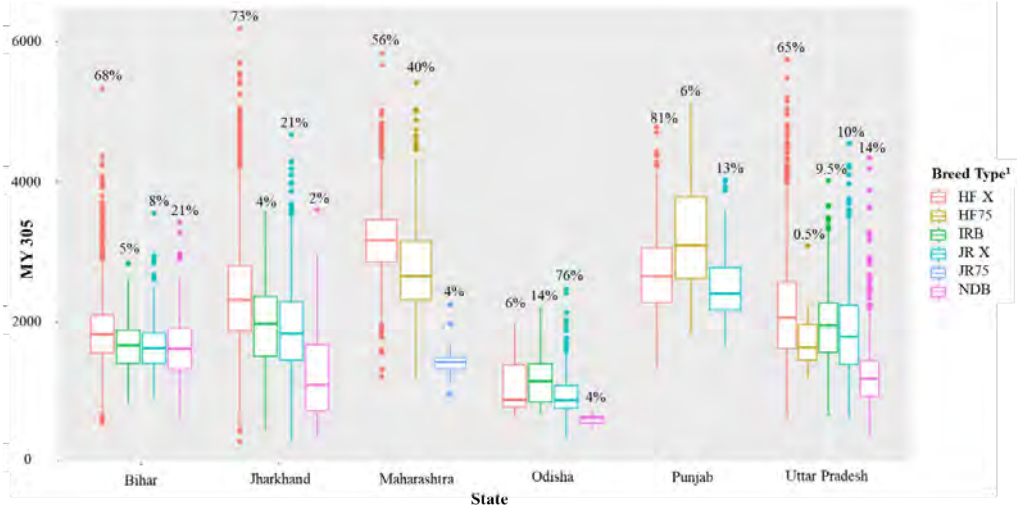
The daily milk yield was transformed into a 305-day production and corrected, using a linear mixed model including the effects of breed type, calving season, lactation rank and type of recording, all treated as fixed.

A recursive partitioning classification using decision trees allowing for clustering of individuals in homogenous sub-populations according to explanatory variables (Breiman *et al.*, 1984), was performed on corrected milk production to identify regions and subsequent characteristics with contrasted production levels.

Statistical analyses were done using RStudio (version 3.5.0) and the ggplot2 package. Recursive partitioning classification was performed using the rpart package.

Results & discussion

Breed distribution. To understand the variability of milk yield, we first looked at the distribution of uncorrected milk production according to breed types between and within states (Figure 1). The percentage of each breed between states varies a lot, ranging from places where indigenous zebus and non-descriptive



[1] HF 75: Crosses with more than 75% of Holstein origin, HF X: Crosses with less than 75% of Holstein origin, JR75: Crosses with more than 75% of Jersey origin, JR X: Crosses with less than 75% of Jersey origin, IRB: Indian Recognised Breed, NDB: Non-Descriptive Breed.

Figure 1. Milk yield production (MY 305) and percentage of each breed type within each state.

cows are an important part of the herd (Bihar, Odisha and Uttar Pradesh), to states that orientated their breeding towards high levels of exotic breeds (Maharashtra, Jharkhand).

Holstein crosses usually show the highest milk production in most of the states. However, we can identify some scaling effect (a significantly higher production in Maharashtra than in other states) and re-ranking effect (in Odisha, Holstein crosses have, on average, a lower milk yield than indigenous breeds).

Figure 1 shows a high variability in the distribution of breed types across the six states. Corrected milk yield was significantly variable among districts (ANOVA, F=170.5, df=25, $P<0.05$).

Milk production under diverse climatic conditions. The recursive portioning classification separated districts into four climatic clusters (Figure 2), with the following predicted average corrected milk productions: 2,999 kg cluster A: Ahmednagar, sd= 649.4 kg, cv=0.2), 2,750 kg (cluster B: Faizabad, Hazaribag, Khunti, Lohardaga, Pune, Ranchi, Ropar and Saharanpur, sd=767.3 kg, cv=0.3), 2,459 kg (cluster C: Ara, Bahraich, Chhapra, Gonda, Hardoi, Hoshiarpur, Kannauj and Meerut, sd=759.4 kg, cv=0.4), and 2,237 kg (cluster D: Allahabad, Bargarh, Bolangir, Buxar, Etawah, Lucknow, Sambalpur, Unnao and Varanasi, sd=701.8 kg, cv=0.4).

Maharashtra shows the highest corrected milk production, especially the district of Ahmednagar (cluster A). This district, situated in a rain shadow region, has a tough climate, with very low rainfall, even during

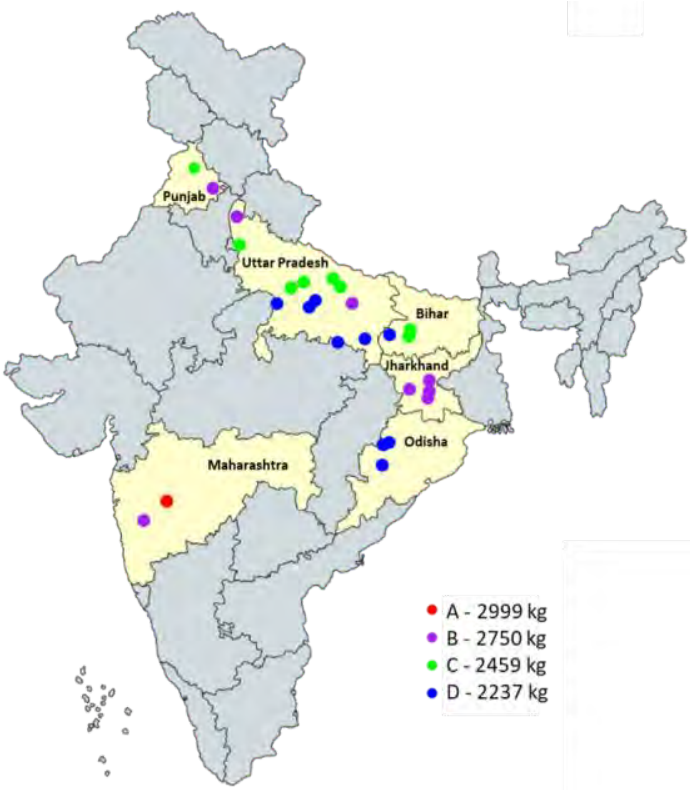


Figure 2. Map of districts partioned by corrected milk production using weather information, each dot corresponding to a homogenous climatic area.

monsoon. The higher average milk production in Maharashtra, is more likely due to a longer technical breeding support of BAIF to local farmers in this state, where BAIF influence started more than 30 years ago. The higher milk production results from years of breeding with exotic breeds, and particularly Holstein cattle. This illustrates that genotype by environment interactions do not only depend on the state, and that beyond climatic conditions, other factors greatly impact production. Particular attention should be paid to the farming system, and further analyses of agro-economic data (major crops, size of the farm, family income, economic status and feeding system of the cows) should help to identify significant determinants of milk yield in India.

Figure 2 also shows that homogeneous groups of weather are not always located within the same state. Indeed, all dots of the same cluster share similar climatic conditions, and we can see that districts of the same cluster can be found in different locations. Mountains and rivers (especially the Ganges river) play an important role in the local weather, resulting in various sub-groups of climates, influencing crops and feedstuff in the area. Considering these regional variations by classifying districts into homogenous climatic zones could help to account for the wide heterogeneity of weathers across India in a genomic evaluation.

Conclusions

India, given its size and its particular milk production, mostly ensured by small and marginal farmers, shows extreme variability in terms of production, breeds and climates, resulting in strong genotype by environment interactions. It, thus, appears crucial to account for them, along with agro-economic characteristics, when establishing a reference population in order to get an appropriate estimation of the genomic breeding value of individuals. In particular, defining a few groups of homogeneous environments across India, rather than considering it as a uniform country, could help to conduct more accurate genomic evaluation in the field.

References

Al Kalaldeh M., Swaminathan M., Gaundare Y., Joshi J., Aliloo1 H., *et al.* (2021) Genet. Sel. Evol. 53:73 https://doi.org/10.1186/s12711-021-00667-6.

Belhekar S., and Dash S. (2016) Paripex - Indian Journal of Research 5(11): 509 – 510 https://doi.org/10.36106/paripex.

Breiman L., Friedman J. H., Olshen R. A., and Stone C. J. (1987) Cytometry, 8(5): 534-535 https://doi.org/10.1002/cyto.990080516.

Chawla A. (2009) Milk and Dairy Products in India – Production, Consumption and Exports. Available at: https://www.hindustanstudies.com/files/dairysept09report.pdf.

454. Connections among local nomenclature, phenotypes and genetic characteristics of native chickens in Cambodia: a synthesis

P. Menghak* and M.G.G. Chagunda

Department of Animal Breeding and Husbandry in the Tropics and Subtropics (490h), Institute of Agricultural Sciences in the Tropics (Hans-Ruthenberg-Institute), Hohenheim University, Garbenstraße 17, 70599 Stuttgart, Germany; menghak.phem@uni-hohenheim.de

Abstract

Currently, there is little existing comprehensive information on Cambodian native chicken (CNC). Further, there is no coordinated phenotypic recording. Predominantly, CNC is known by its local names. However, information linking local names to phenotypic and genetic characteristics of CNC is sparse making phenotypic characterization of these genetic resources difficult. The current study aimed to establish the connection between the local nomenclature of CNC, with their phenotypic and genetic characteristics through a synthesis of some key literature, both published and unpublished. The synthesis involved 7 research articles, one FAO technical report, and local grey literature. Results showed that in terms of nomenclature, CNC are named based on their plumage color and other appearance characteristics such as frizzled and small-sided feathers. There was differing degrees of agreement between local names, phenotypic and genetic characteristics. It is therefore important that the native chickens be phenotypically and genetically characterized incorporating indigenous knowledge in order to initiate a platform for a designed breeding program.

Introduction

In Cambodia, livestock contribute 11.4% to national GDP (MAFF, 2021). The chicken population is approximately 50 million, of which 65% are raised in smallholder farms. The majority of native chicken breeds (98%) are in small flocks of less than fifty birds per household. The remaining 35% of chicken is imported commercial layers and broilers that are reared in intensive systems. Raising native chickens provides a source of food protein, helps farmers to generate other household income, avoids complete dependence on natural fishing, and allows women to earn extra income. The native chicken in Cambodia is also mostly used for ritual and cultural activities. This is despite that the native chicken has low productivity, raised in a poor management system with a high mortality rate due to diseases (Siek *et al.*, 2016). Among local consumers, the native chicken is preferred to the exotic breeds. In most parts of the world, local chickens have different names based on different reasons. Predominantly, the names are derived from appearance, behavior, use, region, and general human perception. However, from a local community acceptance, CNC was named based on their plumage color, feather types and a special characteristic such as frizzled feather. Information linking the local names to phenotypic and genetic characteristics of CNC is never explained. Therefore, the current study aimed to establish the connection between local nomenclature, phenotypes, and possibly genetic characteristics of CNC through a synthesis of some key literature, both published and unpublished to deliver key information for the future breeding plan.

Methods

All research findings and national and international databases, both published and unpublished on native chicken in Cambodia, were reviewed. The review synthesized all relevant information on a production system, external/phenotypic and genetic characteristics from 7 research articles, one technical report from the Food and Agricultural Organization of the United Nations (FAO), and chicken databases from a local institution report. There are external characteristics of chicken adapted from the local knowledge and language.

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_454, © P. Menghak and M.G.G. Chagunda 2022

Results

Classification of chicken breeds. In Cambodia, chicken breeds were categorized into native/local breeds, commercial broilers and layers and crossbreds (Sun, 2018). Very few smallholder farmers used exotic breeds (2-3%) (MAFF, 2021). However, there were a few smallholder producers who had crossed their native line with the commercial lines and fighting cocks from local commercial distributors mainly Thailand, Vietnam. The native breeds were all raised mainly for both meat and egg and in general, the local term is 'Moarn Khmer or Moarn Sre' which means Khmer chicken or indigenous. There is no comprehensive information on native breeds' identities and specific region distribution, the external characteristics are non-descript types varied in plumage color, comb type, shank length and body weight. Sann and Chhum (2004) described eight breeds (ecotypes) with brief external phenotypes summarized in Table 1. In addition, there is another local nomenclature called 'Khmao' which means completely black feather, the breed also varies in body size, shank and skin color. Among all ecotypes, Sampov, Kandong, Skuoy and Kragnas accounted for 55, 10, 7, and 5% respectively, similarly to Sun (2018) mentioned that Sampov accounted for the largest population in the country. Based on the available references, native breeds were most classified by their plumage colors, feather types (frizzle, naked neck), and matured size, except for Red Jungle Fowl and fighting cock. However, the local nomenclature has utilized these external traits to differentiate the native chickens and within the country, people have accepted the names of the native chickens though some characteristics are not distinguished. Future breeding plans may need to include and conserve the local nomenclature and external traits such as plumage, skin, color, shank color.

Phenotypic characteristics of native chickens in Cambodia. In a natural-breeding population without selection, external morphologies of chicken are mixed up and complicated to identify both qualitative and quantitative traits of referenced breeds. Among the native breeds, Khmao and Skuoy were easily identified by the black and barred feathers respectively. However, some ecotypes might not be phenotypically classified as a different breed due to their non-descript characteristics such as qualitative phenotypes; colorful plumage, unidentified skin, shank, earlobe color, comb type, and quantitative phenotypes; body weight, and other growth traits, and egg production traits. Some local nomenclatures referred to feather structures (slow-growing, frizzledness, naked neck), which these phenotypes may occur in any populations

Table 1. The external characteristics of Native Chicken (adapted from Sann and Chhum, 2004).

Local name	Adjusted English name by author[1]	Characteristics adapted by author
Moarn Prey	Red Jungle fowl	Wild chicken (Gallus Gallus) with wild color feather
Sampov	Sailing ship	Large body size and its name refers to the ancient sailing ship, colorful (black, wild, silver, gold, white)
Skuoy	Adlay plant/barred feather	Large body size, sex-link barred feather/color
Kandong	No tail/less feather	Large body size, less body feather, no main tail feather, colorful (black, wild, barred, silver, gold, white)
Kragnas	Frizzle feather	Large/small body size, frizzle feather and tail, colorful (black, wild, barred, silver, gold, white)
Samley	Cotton wool	Small body size/white dwarf bantam, white feather-like cotton
Moarn Cher	Comb, small size	Dwarf local chicken/dwarf bantam, has large single comb, colorful (wild, white)
Moarn Chul	Fighting chicken	The female identity was hidden, the male is noted by the shank color (yellow, and dark green), shank and spur characteristics Colorful (Wild, black, gold, white), fighting behavior

[1] The adjusted names based on the local expert knowledge.

under natural selection for adaptation to the external environment. According to FAO (2009), body weight and shank length distribution of the native chicken were different between male and female, and agro-ecologic zones. The male chickens had heavier and taller or larger body sizes than the female, particularly the chickens in the central plain indicated the highest weight, and the smallest ones found in the coastal region but taller than the chickens in highland. The major shank color was yellow representing 50% of the total, and white and yellow skin was 50 and 35% respectively in all regions, regardless of coastal zone which had the highest yellow skin and shank color up to 60% of the total. The results illustrated differing ranges of agreement between the local names and the phenotypic characteristics of all recent ecotypes identified and accepted as breeds by the local communities. Therefore, there should be more described phenotypic traits to name the chicken breeds also they may differentiate by adapting to different agro-ecologic zones.

Genetic characteristics of native chickens in Cambodia. The Red Jungle Fowl (*Gallus gallus*) occurs over a wide geographical range in Southeast Asia. According to a complete mtDNA D-loop study by Osman and Nishibori (2014), the Red Jungle Fowls (RJFs) from Cambodia showed higher diversity than those in Laos and Myanmar. Okumura *et al.* (2006) illustrated that native chicken populations in South-East Asia are genetically similar including Thailand, Laos, Myanmar and Indonesia due to closed distance and random mating. Nishibori *et al.* (2006) conducted a genetic diversity study from 2,245 native chickens in 17 provinces in Cambodia. The genes considered for morphologic traits; plumage colors, shank colors, comb types were studied; the gene frequency of the loci of CNC except one showed a similar tendency to those of Thai native chickens, particularly those from Ratanakiri province. For dominant white plumage color (I/i locus), the frequency of the I allele was very low (0.007) than that of other Asian countries. In contrast, extend black plumage color (E/e+/e), the frequency of the E allele was higher than Asian countries (0.31), and the frequency of e+ was higher and e was lower than in Asian counties. For shank color, the gene frequency for yellow/white was slightly lower than Asian countries except for China, and females presented higher than males did. On average, this finding was similar to phenotypic frequency (50%) by FAO (2009) but quite low in a few provinces around the capital city where there are high numbers of commercial chicken farms. Cambodian native chickens could be regarded as the first to be domesticated and may share similar genetic resources to those in South-East Asian and Asian countries. Among the native population, they may be genetically different due to natural and artificial selection. Recently, Thailand conserved 4 native chicken breeds and was named them based on plumage color, utilization and main region distribution (Dorji *et al.*, 2011). These studies all suggested a further study on genetic characterization with the relevant breed name, phenotypic characteristics, breed utilization in different agro-ecological zones. Connections among local nomenclature, phenotypes and genetic characteristics of native chickens would be informative in designing breeding plans for improvement, preservation and exploiting the genetic diversity.

Table 2. Comparison of body weights and shank length in different regoins (adapted from FAO, 2009).

Sex	Costal	Central plain	Highland
Body weight (g)			
Males	1,322	1,719	1,469
Females	1,315	1,494	1,320
Shank length (cm)			
Males	9.5	9.8	8.7
Females	8.6	8.4	7.3

Conclusions

Cambodian native chicken breeds are most classified by their plumage colors, feather types (frizzle, naked neck), and matured size by the local nomenclature. Where quantitative information is available, body weight and shank length vary in different sexes and agro-ecological zones. Genetically, CNC is highly diverse. Local nomenclature utilizes external traits to differentiate the native chickens although some characteristics are not clearly distinguished. Future breeding plans may need to include and conserve the local nomenclature although robust phenotypic and genetic characterization will be required.

References

Dorji N., Daungjinda M., Phasuk Y. (2011) Trop. Anim. Health. Prod. 43:779-785. https://doi.org/10.1007/s11250-010-9763-3

FAO (2009) AHBL-Promoting strategies for prevention and control of HPAI, Rome, Italy. Available at https://www.fao.org/3/al677e/al677e.pdf.

Ministry of Agriculture, Forestry and Fishery (MAFF), Annual report of 2020, 2021.

Nishibori M., Nozawa K., Mannen H., Kurosawa Y., Nomura K., *et al.* (2006) Rep. Soc. Res. Native Livestock 23:93-102.

Okumura F., Shimogiri T., Kawabe K., Okamoto S., *et al.* (2006) Anim. Sci. J. 77:188-195. https://doi.org/10.1111/j.1740-0929.2006.00336.x

Osman S.A.M. and Nishibori M. (2014) J. Poult. Sci. 51:248-261. https://doi.org/10.2141/jpsa.0130190

Sann V. and Chhum P.L. (2004) Proc. of Tropentag, 'Rural Poverty Reduction through Research for Development and Transformation', Berlin, Germany.

Siek D., Xu S., Yu W., Ahmed A.G., and Din, C. (2016) Open J. Soc. Sci. 04(05):295-299. https://doi.org/10.4236/jss.2016.45032

Sun H. (2018) Aust. Agribus. Perspect. 21(3):1442-6951. Available at https://cpb-ap-se2.wpmucdn.com/blog.une.edu.au/dist/4/1340/files/2018/07/AAP-2015-Paper-106-Pathiraja-et-al-22p0qfb.pdf.

455. Experience in creating a female *Bos taurus* × *Bos indicus* reference population under smallholder system in India

K.G. Bhave*, S. Joshi, Y. Gaundare, A. Joshi, T. Shirsath, N. Punde, V. Potdar, V. Ratnakar and M. Swaminathan

BAIF Development Research Foundation, Uruli Kanchan, Pune, 412202, India; kaustubh.bhave@baif.org.in

Abstract

BAIF Development Research Foundation is a large Indian NGO currently involved in establishing a female reference population (FRP) of crossbred *Bos taurus* × *Bos indicus* cows in smallholder farms in nine Indian states, with the ultimate goal to develop a genomic selection based dairy cattle genetic improvement program. Here we describe the specific challenges encountered in the data collection and its management, as well as the corrective measures adopted to resolve them.

Introduction

In developing countries like India, the majority of milk production comes from smallholder farmers maintaining mixed types of dairy animals having little or no phenotypic and pedigree information. However, limited efforts have been made to undertake genetic selection based on progeny testing for milk production traits. Furthermore, other traits related to adaptation to the local climatic conditions have not been considered either. Conventional genetic evaluation systems are logistically demanding and expensive, especially under the fragmented smallholder farms where generalized identification is lacking. Given the constraints and challenges posed by classical genetic evaluation systems as implemented in other countries, a simplified digital approach for collection of multiple-trait phenotypes is considered. Recorded animals can be genotyped in order to build up a female reference population to carry out genomic selection. Then, by applying selection with an increased level of accuracy for a sustainable selection objective under one or several specific production environments can be envisioned.

BAIF (BAIF Development Research Foundation, https://baif.org.in/) is a large NGO in India which pioneered the frozen semen technology and delivery of AI (Artificial Insemination) service at the farmers' doorstep. One of the core activities of BAIF is cattle and buffalo breed improvement. As part of the genetic improvement program, the 'Enhanced Genetics Project' (EGP) funded by the Bill and Melinda Gates Foundation (https://www.gatesfoundation.org/) is being implemented with the objective to leverage genomics tools in breed improvement. As part of this project, BAIF is developing a female reference population (FRP) of crossbred animals (and in due course, of indigenous breeds and buffaloes) in order to demonstrate the feasibility of genomic selection under the Indian context (see Ducrocq *et al.*, this congress for details). This required the development of an interactive and flexible system to collect a vast amount of information on different traits from crossbred populations spread across several Indian states.

The present paper describes the recording system implemented to build the FRP, the software and its deployment using appropriate digital devices, as well as the challenges faced and mechanisms adopted to resolve various issues during the implementation.

Materials & methods

Implementation of a data collection system and subsequent hurdles faced. The data collection system was implemented in 9 Indian states to accommodate poor to high production systems (Punjab, Uttar Pradesh, Bihar, Jharkhand, Odisha, Maharashtra, Gujarat, Rajasthan and Andhra Pradesh) with 63 data enumerators visiting the farmers' herds and updating the data in the data logger in offline mode,

synchronising the data base once in a day. The data was collected either by the enumerator on a monthly basis and by the farmer. The farmers were provided with data recording cards on which they filled up information such as milk yield on a weekly basis. The enumerator then visited the farmer and collected the data from card and entered the data into the software. To ensure correct data collection by the farmer, unannounced visits by the enumerator were conducted. The option of recording by the farmer was explored specifically to evaluate a low-cost sustainable model of data recording. This architecture is described in Figure 1.

The selection of areas for recording were made in such a way that performance recording could be done comfortably while maintaining 90% of animals from small herd size (2-4 animals of mixed breed type) and 10% from large herds. Every enumerator collected the data from 2 to 3 cattle development centres (CDC) covering 4 to 5 villages each. For every farmer, a baseline survey was carried out to record the farmer's identity and characteristics (activity, sources of income and assets). A herd management survey (herd composition, equipment, etc.) was also conducted at the time of registration of the farmers into the program.

Production performance. Individual production data of each cow and buffalo included test day milk yield using GPS enabled weighing machine and milk components such as fat%, protein%, SNF%, etc. obtained using a portable Milk Analyser. This was recorded once a month together with dry period and calving information. Information regarding calf suckling was added to the list since this is commonly practiced in indigenous cows and buffaloes.

Breeding performance. The system captures insemination details such as natural vs artificial service, pregnancy diagnosis by rectal palpation between 60-90 days. It also captures pedigree details whenever available. Breeding records are collected as and when available.

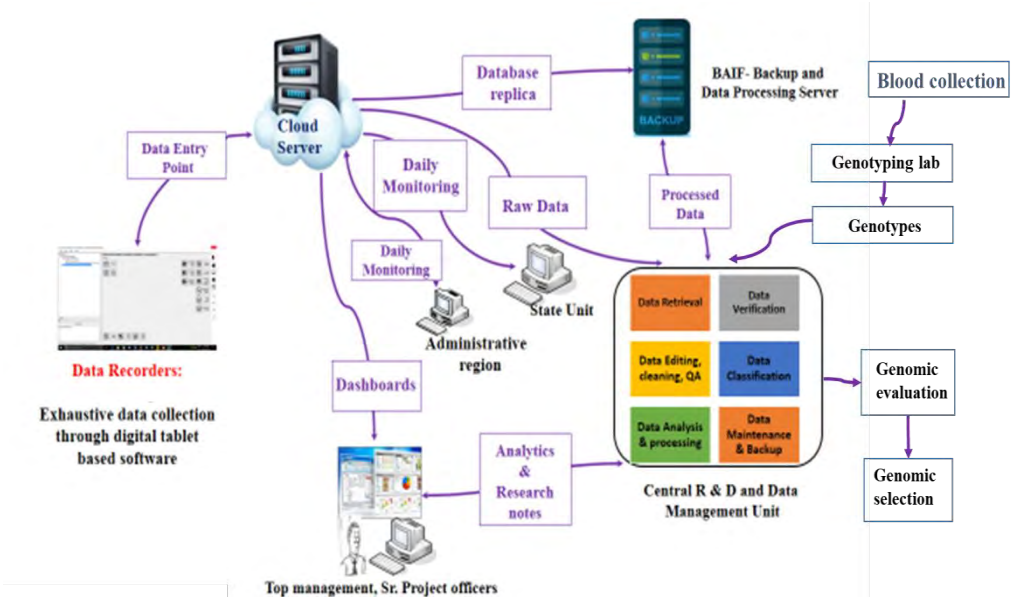


Figure 1. Overview of data collection from farms to the BAIF central research station.

Morphometric and type trait information. As exact weight measurement is not possible at farmers' level; estimated weight is calculated by measuring girth and length with a measuring tape. Morphometric information is collected once in two months. In total, 26 body type traits are also measured or recorded by a trained technician once in a lifetime of every first lactation animal usually in morning hours after milking.

Mastitis incidence. One of the aims of the project was to determine the mastitis incidence as it causes major economic loss. The occurrence of mastitis was recorded when the farmer or the enumerator observed clinical mastitis on the day of recording or during the interval since the last recording. Somatic Cell Count (SCC) information is captured using TANUCHECK SCC kits (Nithya *et al.* 2017) and measured in 5 classes of no. of cells in thousands per ml. This results in a low-cost and simple test at farmer's doorstep to know the udder health status of animals.

Results

Table 1 and 2 describe the distribution of the data collected in the country and its genetic type. Table 3 reports the information already collected at various levels.

Discussion

Data validation and checks. The hierarchical system was developed for data collection and monitoring. 10% of random data recorded by enumerator was cross-validated by a supervisor and 5% by a state coordinator paying an actual visit to the herd. The date of registration, calf birth, performance recording, artificial insemination (AI), calving, etc. were checked for consistency through the software solution. Restrictions were placed on body measurements, on feeding quantities and milk components. Bluetooth facility was used for automatic entry of milk yield into the software. Manual entry of data by enumerators was reduced wherever possible by providing drop down options.

Table 1. Characteristics of data collected between October 2016 and October 2021.

Villages	CDC	Districts[1]	Enumerators	Herds
1,390	155	42	63	13,607

[1] District: administrative region within a state.

Table 2. Breed wise animals under performance recording.

Species	Buffalo		Cow				Total animals
Breeds	Non-descript[1]	Indigenous[2]	Holstein crosses	Indige-nous[2]	Jersey crosses	Non-descript	Recorded
Total	6,836	2,292	14,415	4,134	4,676	1,824	25,049
Genotyped	-	-	6,485	2,250	1,863	517	11,115

[1] Non-Descript: animals without specific characters known to any registered breeds.

[2] Indigenous breeds: Animals with known characters and registered as a breed by NBAGR.

Table 3. Records collected between October 2016 and October 2021.

Milk yield	Milk components	Body measurement	AI records	Body type traits	Mastitis incidence
617,134	237,174	134,288	12,707	5,289	312,153

Difficulties faced and evolution of the data collection and retrieval system. Initial animal identification was difficult, as most animals did not have a tag number. Also, information on villages, cattle development centres (CDC- representing an area of 4-5 villages covered by a same AI technician), etc. was filled manually. To avoid errors, computerised unique IDs were generated for each animal, village, CDC and farmers. Another issue was the absence of uniformity in milk yield data entry. Given the vast region under recording, it was difficult to monitor whether the enumerator actually visited farmers and recorded the animal performance. We introduced GPS-enabled weighing machines with Bluetooth facility. This solved both issues regarding data monitoring as it captured geo-location as well as exact weight of morning and evening milk yield. To avoid erroneous values in milk components, enumerators were equipped with a portable milk analyser which estimated milk components and were advised to calibrate the analyser every 3 months from the local dairies. Considering the very small herd size (sometimes just one animal per herd) observed in earlier recordings, the enumerators were asked to collect data from herds with 2 or more animals in order to separate additive genetic variance from herd variance in statistical analysis.

Future prospects. Multi breed and multi trait selection program for Indian dairy cattle and buffaloes is proposed. Extra non-productive functional traits such as fertility, productive life, farmer preferred (phenotypic) traits such as animal temperament (especially valued by women farmers, panting score (to understand heat stress) will be also recorded. To streamline sample collection process from field to lab, Laboratory Information Management System (LIMS) will be introduced. To refine the genomic prediction system and increase the reference population base; a larger number of animals will be genotyped for multi trait genomic selection program to derive better prediction accuracy. It is also planned that some selected animals will be genotyped on a high density chip or on their whole genome for future studies e.g. causative SNPs for current or possibly novel traits.

References

BAIF Development Research Foundation. Available at: https://baif.org.in/.

Bill and Melinda Gates Foundation. Available at: https://www.gatesfoundation.org/.

DAHD (2021) Department of Animal Husbandry and Dairying Government of India. Available at: https://dahd.nic.in/annual_report/.

Nithya P., Sivakumar K., Aruljothi A., and Tirumurugaan K.G. (2017) International Journal of Livestock Research 7(8), 173-178. http://doi.org/10.5455/ijlr.20170605023657

456. Learnings and experiences from a community-based goat breeding program in India

A. Schurink[1*], S. Kumar[1], R. Chakraborty[2], T. Sawhney[2], C. Nimbkar[3] and P. Amer[1]

[1]AbacusBio Limited, 442 Moray Place, Dunedin 9016, New Zealand; [2]Aga Khan Foundation (India), 6 Bhagwan Dass Road, New Delhi 110001, Delhi, India; [3]Nimbkar Agricultural Research Institute, P.O. Box 44, Phaltan 415523, Maharashtra, India; aschurink@abacusbio.co.uk

Abstract

The impact and progress made in the early stages of Project Mesha's community-based goat breeding program (CBBP) are encouraging. Up-scaling and out-scaling of the CBBP, expanding on its early successes, is considered a sensible priority to reach our ultimate goal of impacting many more landless, marginal and small farm households across a larger area in India. Several factors to consider when expanding a CBBP and ensuring its sustainability stem from challenges we have observed during these early stages that are similar to those faced by other small ruminant CBBPs. Our paper highlights Project Mesha's CBBP impact, elaborates on factors such as the need for community leadership for further up-scaling and out-scaling, and discusses the opportunities that genomics can provide.

Introduction

CBBPs have the potential to genetically improve livestock in developing countries where required infrastructure for setting up national breeding schemes is unavailable (Haile *et al.*, 2020). Additionally, CBBPs engage with and incorporate the views and needs of end users considered crucial for a successful breeding program and therefore progress, while centralised government-controlled schemes seem less successful (Haile *et al.*, 2020). CBBPs generally commit to locally adapted animals, among other things to prevent erosion of these genetic resources (ICARDA, 2018; Haile *et al.*, 2020). In view of the above, the community-led program called Project Mesha that is implemented by the Aga Khan Foundation (AKF) since 2016 in Bihar, one of the poorer states of India with a large goat population (BAHS, 2019), extended its outreach with a CBBP in 2018.

Goats provide livelihood support to over 70% landless, marginal and small farm households in India. Project Mesha (meaning 'buck' in Sanskrit) aims to transform and improve the quality of life of such households in Bihar through improving the productivity and production of local Black Bengal goats owned by them and primarily reared by women. The ultimate aim of the project is to increase the income of these households and empower these women goat rearers. Project Mesha is layering goat productivity improvement interventions on the existing platform of women self-help groups, targeting goat nutrition, preventive health (vaccinations and deworming), low-cost housing and marketing as well as breeding. It also supports adoption of improved goat husbandry and management practices, with all activities being carried out by trained project staff. Preventive health care is delivered through fee-based services by pashu sakhis (meaning 'women friends of livestock'). Currently, the program is serving about 50,000 households in 240 villages in four out of 16 blocks of the Muzaffarpur district in Bihar with a cadre of 250 pashu sakhis.

A high death rate and low production of goat flocks adversely impact the livelihood potential of goat rearing. To generate genetic improvement in these traits, a CBBP has been implemented since 2018 in initially four selected villages, one in each block of the Muzaffarpur district where Project Mesha is operating. The CBBP involves tagging of Black Bengal goats (the predominant breed, being highly prolific and reared for meat production; Dey *et al.*, 2007), data recording on goat reproductive performance, weight, matings and other

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_456, © A. Schurink *et al.* 2022

measurements in the Dtreo platform (https://dtreo.io), and selecting and rotating breeding bucks among participating villages.

Project Mesha is supported by the Bill & Melinda Gates Foundation, with expert consultancy and training in breeding and genetics being provided by AbacusBio Limited and the Nimbkar Agricultural Research Institute. Our paper has a dual objective of presenting the status and impact of the CBBP implemented through Project Mesha, and elaborating on potential pathways and challenges related to up-scaling and out-scaling of a CBBP based on our learnings and experiences and those of other CBBPs reported in the literature.

CBBP status and impact

Project Mesha's CBBP is the first systematic and organized CBBP in India with operational support and daily management by AKF. Its impact has expanded since the implementation in 2018, with goat data recording currently taking place in 16 villages, and selected breeding bucks with superior performance being placed in additional villages.

Awareness among the resource-poor women goat owners on concepts related to genetic improvement and (in)breeding has increased and resulted in their empowerment as demonstrated by a recent impact study carried out by AKF surveying over 350 households participating in the CBBP. Considering inbreeding, it is typical that only free-roaming bucks are available for mating and they tend to mate most of the does in a village likely including their own progeny. By coordinating movement of selected bucks to distant villages inbreeding risks have been mitigated.

The availability and use of the Dtreo platform, AbacusBio's database system for customized animal recording, is crucial for data recording by the field team, tracking of targets and reporting on any level (e.g. individual breeding buck performance, number of doe assessments done in a village, weight trends). At present, records of 2,357 goats have been entered in the database with almost 6,500 records on weight and assessments of health, adaptive ability, previous litter size and kid survival (Table 1). Around 500 does have so far been mated with just over 30 selected breeding bucks. Data analysis to evaluate the superiority of their progeny is planned for the near future.

Table 1. Number of goats and data from Project Mesha's CBBP recorded in the Dtreo platform to date (December 2021).

Animals	**Total number with records**	**Number alive**[1]
Goats	2,357	1,783
Kids (<120 days old)	284	197
Breeding does (≥120 days old)	1,592	1,300
Bucks (≥120 days old)	481	286
Selected breeding bucks	23	23
Records	**Number**	
Kid weights	4,107	
Doe weights	925	
Doe assessments	1,232	

[1] Not representing mortality, but rather the number of animals that are still alive and therefore available for breeding, or selection in case of kids.

Recent analyses of own performance data (Nimbkar *et al.*, 2021) showed that about half of the does gave birth to twins, 26% to triplets, 20% to singles and 3% to quadruplets. Estimated average daily gain up to the age of 120 days was 57.5 g and ranged from 10 to 140 g per day, indicating substantial variation that enables the selection of superior bucks based on an index score that combines data on breeding goal traits (weight and early growth as kid traits, and size, health, twinning ability and kid survival as dam traits; identified as breeding goal in consultation with goat rearers). The production performance of the Black Bengal breed in the project area in the Muzaffarpur district has thus been characterized.

Expansion of CBBPs

Considering Project Mesha's CBBP is still in its early stages, the activities within the program and the impact and progress made so far are encouraging. Up-scaling and out-scaling of the CBBP, expanding on its early successes, is considered a sensible priority to reach our ultimate goal of impacting many more landless, marginal and small farm households in Bihar and other states of India. Factors to consider when expanding a CBBP are elaborated below.

Self-sustainability of a CBBP is key in the long term but can be difficult to realize due to various dependencies, like financial, technical and organisational support (Mueller *et al.*, 2015; based on a review of 8 CBBPs across different countries). Both Mueller *et al.* (2015) and ICARDA (2018) rightfully state that for a CBBP to be sustainable, long-term commitment by a local government or autonomous government-supported institution is essential as a CBBP typically progresses beyond the initial funding duration. Reaching the right people within such institutions and convincing them might prove to be a long-term process that requires both patience and dedication. A local knowledgeable and dedicated stakeholder (like AKF) having a good rapport with these institutions is more likely to enlist the support of such institutions.

Monitoring, evaluation and documentation of a program's impact are key to replicating and expanding a CBBP. Also, rigorous publication of its impact through various forums contributes to attracting additional funds (Mueller *et al.*, 2015). Moreover, a strong program monitoring and evaluation advises on implementation and strategy changes required to achieve the desired results. Our recent impact study identified such amendments (e.g. the need for stronger financial support for women buck keepers). A comprehensive data analysis to determine the socioeconomic impact of the CBBP and obtained genetic gain, such as that presented in Haile *et al.* (2020), is not feasible yet. Nonetheless, available data (Table 1) and findings from our recent impact study can already provide helpful trends and pointers.

Community ownership and leadership are important and inherent to any CBBP (e.g. Zoma-Traoré *et al.*, 2021), and crucial when up-scaling and out-scaling. Our recent impact study highlighted that the pashu sakhis, being local facilitators, have a big influence on the adoption and attitudes of participants towards interventions. The existing platform of women self-help groups provides the needed infrastructure to support a community-led approach to breeding buck selection, distribution and improved mating practices. This follows the work by Zoma-Traoré *et al.* (2021) who highlights the importance of a community's sense of CBBP ownership.

The Black Bengal is a goat breed spread over at least three states of India and there is demand for improved breeding bucks. Currently, some government programs source breeding bucks by purchasing entire males from trading markets with no background knowledge of their performance. There is thus an opportunity to develop breeding buck supply units/clusters for more income-generation. An example of income-generation is described by Gizaw *et al.* (2014) where there is a revolving fund for designated ram keepers in an Ethiopian CBBP. The income from selling old breeding rams is used to buy young candidates and to financially support the ram keepers. A similar scheme could be designed for Project Mesha's CBBP, as

keeping breeding bucks is empowering women goat rearers but often no surplus income is generated due to the reasonably high expenditure (e.g. buck housing and feed).

Training and upskilling are important for both CBBP maintenance and expansion, and should ideally be provided through a train-the-trainer model, with continuous development and updating of training material targeting both staff and goat rearers in their local language and in appropriate ways (for example the educational videos in the local language prepared by Project Mesha, available at https://akflearninghub.org/instructional_videos/video/agriculture-and-food-security/goat-rearing-best-practices/). Scientists and consultants play an important role in training and providing guidance to the CBBP's governing bodies on a regular basis based on updates coming from field progress reports and database statistics. Covering some basic infrastructure costs through an annual contribution from a government or autonomous governmental-supported institution targeting the social and economic empowerment of the rural poor could be a feasible solution, particularly for expanding CBBPs.

Exploiting the potential of genomics

Genomics provides significant opportunities to CBBPs through accurate measures of coancestry and pedigree reconstructions (e.g. ICARDA, 2018). Also, genomics allows for the identification of novel genomic regions and/or variants associated with breeding goal traits like growth and prolificacy (e.g. Haile *et al.*, 2020). Using these novel and known variants through either a marker-based approach or a more advanced genomic selection approach and the earlier mentioned possibility of pedigree reconstructions, allow for faster genetic improvement and better inbreeding management.

Plans are in progress to sample goats for genotyping or sequencing within Project Mesha aimed at exploiting the opportunities that genomics provides. A careful selection of goats should be made, considering both the availability of phenotypes as well as knowledge on potential coancestry. More challenging are the practicalities around sample collection, distribution and storage, as well as data repositories. Nonetheless, we would advice setting up a system to collect DNA samples from the implementation of a CBBP, or even before, so long as individual animal identification has been done.

References

Dey A., Barari S.K., and Yadav B.P.S. (2007) Livest. Res. Rural Dev. 19(9):123.

Basic Animal Husbandry Statistics (BAHS), Government of India, Department of Animal Husbandry and Dairying (2019). Available at: https://dahd.nic.in/circulars/basic-animal-husbandry-statistics-2019.

Gizaw S., van Arendonk J.A.M., Valle-Zarate A., Haile A., Rischkowsky B., *et al.* (2014) J. Anim. Breed. Genet. 131(5):350-357. https://doi.org/10.1111/jbg.12102

Haile A., Getachew T., Mirkena T., Duguma G., Gizaw S., *et al.* (2020) Animal 14(7):1362-1370, https://doi.org/10.1017/S1751731120000269.

International Center for Agricultural Research in the Dry Areas (ICARDA), Guidelines for setting up community-based small ruminants breeding programs (2018). Available at: https://cgspace.cgiar.org/rest/bitstreams/172182/retrieve.

Mueller J.P., Rischkowsky B., Haile A., Philipsson J., Mwai O., *et al.* (2015) J. Anim. Breed. Genet. 132:155-168, https://doi.org/10.1111/jbg.12136.

Nimbkar C., Amer P., Schurink A., Sawhney T., and Azam R. (2021) Proc. of the 24th AAABG conference, Adelaide, Australia. Available at: http://www.aaabg.org/aaabghome/AAABG24papers/68Nimbkar24272.pdf.

Zoma-Traoré B., Probst L., Ouédraogo-Koné S., Soudré A., Ouédraogo D., *et al.* (2021) Sustainability 13:2499, https://doi.org/10.3390/su13052499.

457. Response to index selection for temperate dairy cattle breeds in tropical Sri Lanka

A.M. Samaraweera[1*], J.H.J. van der Werf[2], V. Boerner[1] and S. Hermesch[1]

[1]AGBU, a joint venture of NSW Department of Primary Industries and University of New England, 2351, Armidale, Australia; [2]School of Environmental and Rural Science, University of New England, Armidale 2351, NSW, Australia; amamalsh@gmail.com

Abstract

This study aimed to predict selection responses for breeding objective traits based on large-scale dairy farms rearing temperate dairy breeds in Sri Lanka. Responses were predicted selecting 10% of bulls at 4 and 9 years of age assuming the availability of progeny records for the selection of older bulls. When economic weights were expressed relative to the trait genetic standard deviation, milk yield (MY, 74%) had the highest relative emphasis on selection. Favourable responses were only observed for MY and calving interval reflecting the high selection emphasis on MY to counteract with other traits. Predicted responses relative to the current mean varied from -0.67% (protein yield) to 0.40% (MY). Progeny selection increased the accuracy at the expense of higher generation interval. Predicted increase in profit due to genetic improvement was slightly higher under progeny testing compared to family selection (2,624 vs 2,506 rupees or 14.1 vs 13.5 USD per cow per year).

Introduction

Selection of dairy cattle based on estimated breeding values is currently not practiced in Sri Lanka. However, the need for development of dairy cattle breeding programmes has been highlighted. Therefore, as a precondition for establishment of a breeding programme in Sri Lanka, genetic parameters for the main traits of interest in a breeding objective for temperate dairy cattle in Sri Lanka were recently estimated, and economic values (EVs) were derived (Samaraweera 2020). Prediction of annual genetic response of traits in the breeding objective is a key criterion in evaluating the selection decisions and understanding the direction and magnitude of individual trait responses. Therefore, the objective of this study was to predict response to multi-trait selection for breeding objective traits in large-scale dairy farms in Sri Lanka.

Materials & methods

The breeding objective traits of the large-scale dairy farms rearing temperate dairy breeds were annual milk yield (MY, kg/cow/year), annual fat yield (FY, kg/cow/year), annual protein yield (PY, kg/cow/year), age at first calving (AFC, days/cow), number of services per conception (NSC, counts/cow/conception), calving interval (CI, days/cow), and number of episodes of mastitis (MS, counts/cow/lactation). Please see Samaraweera *et al.* (in press) for trait definitions and EVs.

Economic weights. The EVs were transformed into economic weights (EWs) by multiplying them by discounted genetic expression (DGE) coefficients to account for the different frequency and timing of expressions of the traits (McClintock and Cunningham 1974). Since AFC is expressed once per cow and early in the cow's lifetime, the DGE coefficient for AFC was equal to one, and EWs were expressed per cow at first calving. All other traits were expressed once per lactation. Therefore, DGE coefficients for MY, FY, PY, NSC, CI, and MS were calculated as the sum of all timings and the frequency of expressions over five lactations as follow:

$$\mathrm{DGE} = \sum_{\mathrm{Lc}=1}^{\mathrm{Lc}=5} \frac{\boldsymbol{\rho}}{(1+\mathrm{r})^{\mathrm{Lc}}} \quad (1)$$

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_457, © A.M. Samaraweera *et al.* 2022

Where ρ = a vector with each element being the proportion of cows surviving in each lactation relative to the total number of surviving cows, weighted by lactation number, r = the annual interest rate of 5%, and Lc = lactation number reflecting lactations 1 to 5. Survival rate from one lactation to the next lactation was assumed to be 90%. The EVs and calculated economic weights are given in Table 1. When expressed relative to the trait genetic standard deviation, MY (74%) had the highest relative emphasis followed by FY (11%), CI (7%), and AFC (4%) (Table 1).

Selection index. Response to selection per year was predicted for two scenarios, i.e. (1) where bulls were selected at 4 years of age based on information from their dam and paternal half sibs (scenario 1); and (2) selection at 9 years of age based on progeny test (scenario 2). The breeding programme was based on a government nucleus farm with 2,000 cows, and 10% of male calves were selected as suitable to become breeding bulls in the nucleus. Therefore, selection intensity was 1.75 in both scenarios. Higher number of dam, half-sib, and progeny records were assumed for scenario 2 than scenario 1 (2, 5, and 10 vs 1, 2, and 0). The number of half-sib and progeny information were based on the assumption of limited matings per sire in Sri Lanka. Response to selection was predicted using a multi-trait selection index programme, MTINDEX (Van der Werf 2018) which expresses selection response for each trait per standard deviation of the index. To obtain the response per year, these values were multiplied by the intensity of selection and divided by the appropriate generation interval. The genetic parameters that were used to predict response to multiple trait selection are given in Table 2.

Results

Trait accuracies were higher from progeny testing (scenario 1) compared to selection based on dam and paternal half sib information (scenario 2) (Table 3). After selection, scenario 2 had the highest response for all traits except CI and MS. Favourable (in $ terms) responses were observed only for MY and CI in both scenarios. Even though EWs were negative for FY and PY, response was positive and therefore unfavourable for FY and PY mainly due to the positive correlations between these trats with MY. The predicted increase in profit due to genetic improvement per year was slightly higher in scenario 2 than scenario 1 (2,624 vs 2,506 rupees or 14.1 vs 13.5 USD per cow per year). Predicted genetic response to selection relative to the current mean (% change, Table 3) varied from -0.67% (PY) to 0.40% (MY).

Table 1. Economic values (EV), the discounted genetic expression (DGE) coefficients, economic weights (EW), rupees per genetic standard deviation (SD), and relative emphasis (RE) of each trait.

Parameter	Traits[1]						
	MY	FY	PY	AFC	NSC	CI	MS
EV[2] (rupees[3]/cow/year)	107	-162	-15	-59	-270	-84	-8,303
DGE	2.36	2.36	2.36	1.00	2.36	2.36	2.36
EW (rupees/cow/year)	253	-382	-35	-59	-637	-198	-19,595
rupees per genetic SD	50,496	-7,341	-730	-3,122	-1,126	-4,357	-305
RE (%)	74	11	1	4	2	7	1

[1] MY = annual milk yield (kg); FY = fat yield (g); PY = protein yield (g); AFC = age at first calving (days); NSC = number of services per conception; CI = calving interval (days); ME = number of episodes of clinical mastitis.

[2] EV: from Samaraweera *et al.* (in press).

[3] USD = 186 rupees, May 2020.

Table 2. Phenotypic standard deviation (σ_p), heritability (diagonal), genetic (below the diagonal) and phenotypic (above the diagonal) correlations of the traits used in the multiple-trait selection index.

	σp	MY	FY	PY	AFC	NSC	CI	MS
MY[1,2]	707[5]	**0.08**[5]	0.81[9]	0.81[9]	0.00	0.15[10]	0.23[10]	-0.13[11]
FY	48[14]	0.63[9]	**0.16**[14]	0.88[10]	0.00	0.14[10]	0.20[10]	-0.11[11]
PY	45[9]	0.63[9]	0.80[10]	**0.21**[3]	0.00	0.16[10]	0.22[10]	-0.11[11]
AFC	100[4]	0.31[7]	0.00	0.00	**0.28**[4]	0.45[12]	-0.01[12]	0.00[13]
NSC	10.2[5]	0.25[10]	0.25[10]	0.21[10]	0.87[12]	**0.03**[5]	0.73[10]	0.00
CI	98.3[5]	0.11[8]	0.49[6]	0.44[6]	-0.52[12]	0.71[10]	**0.05**[5]	0.03[13]
MS	0.09[11]	0.32[11]	0.22[11]	0.29[11]	-0.18[13]	0.00	0.061[3]	**0.03**[11]

[1] Traits and units are as given in Table 1.

[2] For reference 1 US dollar is equal to 186 rupees (May 2020).

References for the parameters are indicated by superscripts: [3]Bang *et al.* (2019); [4]Montaldo *et al.* (2010); [5]Samaraweera *et al.* (2020); [6]Adapted from Haile-Mariam *et al.* (2003); [7]Salah Khattab *et al.* (2005); [8]Samaraweera (2020); [9]Adapted from Carabano *et al.* (1989); [10]Kadarmideen *et al.* (2003); [11]Carlén *et al.* (2004); [12]Eghbalsaied (2011); [13]Pryce *et al.* (1997); [14]Wahinya *et al.* (2020). Values without a reference are indicated as zero.

Table 3. The trait accuracy (Acc), annual response to multiple-trait selection in trait units (Tu) and rupee value (Rs), and percentage change of the traits in breeding objective relative to the current mean (% change) when sires are selected at 4 (Scenario 1) and 9 years of age (Scenario 2).

Trait[1]	Scenario 1				Scenario 2			
	Acc	Tu	Rs	% Change	Acc	Tu	Rs	% Change
MY	0.22	18.9	2,722	0.38	0.53	19.9	2,878	0.40
FY	0.30	0.6	-133	-0.22	0.66	0.8	-176	-0.34
PY	0.35	1.17	-23	-0.61	0.72	1.2	-25	-0.67
AFC	0.34	4.6	-156	-0.43	0.72	3.8	-127	-0.34
NSC	0.21	0.05	-20	-0.75	0.52	0.03	-11	-0.31
CI	0.29	-1.08	122	0.24	0.63	-0.81	92	0.18
MS	0.16	0.0005	-5.55	-0.17	0.40	0.0006	-6.84	-0.21
Total[2]			2,506				2,624	

[1] Traits and units are as given in Table 1.

[2] Total: rupee value per cow per year.

Discussion

The two scenarios were compared to understand the response to selection from early selection with lower accuracy using family information versus selection based on more accurate progeny test information but with a longer generation interval. The accuracy of selection and slightly higher responses were obtained for all traits except CI and MS in scenario 2. However, the response in scenario 1 is likely overestimated since selection was based on family information, and genetic variance is reduced due to selection (Bulmer 1971). The genetic variance can be further reduced in family selection due to inbreeding. Despite the overestimated response in scenario 1, the family records can be used to achieve annual responses over a four-year period in a breeding programme until the progeny records become available.

Among the milk yield traits, the selection for FY and PY is not economically viable in both scenarios, even though, increased FY and PY in milk is favoured. Therefore, increased payment for FY and PY or reduced feed costs are recommended (Samaraweera, in press). A high relative emphasis on MY, FY, and PY (92%)

in the selection index is not beneficial for a tropical country like Sri Lanka where fertility, disease, and heat tolerance traits are equality important in the breeding objective for a more resilient dairy population. Therefore, inclusion of these functional traits in the breeding objective is recommended.

Genomic selection is advantageous in reducing the generation interval and increasing the genetic gain compared to conventional progeny testing. In Sri Lanka, genomic selection has not been evaluated so far, and a major problem for functional dairy cattle breeding programmes is the absence of reliable pedigree information. Genomic selection would benefit such systems given parentage can be traced using genetic markers. However, the cost of implementation and monetary benefits should be carefully considered before implementing genomic selection in developing countries.

In conclusion, progeny selection increased the accuracy of selection and response per year at the expense of higher generation interval. Favourable responses after selection were only predicted for MY and CI due to the very high selection emphasis on MY and its unfavourable genetic associations with other traits.

Acknowledgements

The first author acknowledges financial support from University of New England via International Postgraduate Research Award to pursue PhD studies and first author's current employer Angus Australia for funding the conference participation.

References

Bang N., Hayes B., Randhawa I., Lyons R., Gaughan J.et al. (2019) Proc. of the 23rd AAABG, Armidale, Australia.

Bulmer M. (1971) The American Naturalist 105(943):201-211. https://doi.org/10.1086/282718

Carabano M., Van Vleck L.D., Wiggans G., and Alenda R. (1989) J. Dairy Sci. 72(11):3013-3022. https://doi.org/10.3168/jds.S0022-0302(89)79454-6

Carlén E., Strandberg E., and Roth A. (2004) J. Dairy Sci. 87(9):3062-3070. https://doi.org/10.3168/jds.S0022-0302(04)73439-6

Eghbalsaied S. (2011) Trop. Anim. Health Prod. 43(4):811-816. https://doi.org/10.1007/s11250-010-9767-z

Haile-Mariam M., Morton J., and Goddard M. (2003) Proc. of the 15th AAABG,

Kadarmideen H.N., Thompson R., Coffey M.P., and Kossaibati M.A. (2003) Livest. Prod. Sci. 81(2):183-195. https://doi.org/10.1016/S0301-6226(02)00274-9

McClintock A., and Cunningham E. (1974) Anim. Sci. 18(3):237-247. https://doi.org/10.1017/S0003356100022418

Montaldo H., Castillo-Juárez H., Valencia-Posadas M., Cienfuegos-Rivas E., and Ruiz-López F. (2010) J. Dairy Sci. 93(5):2168-2175. https://doi.org/10.3168/jds.2009-2050

Pryce J., Veerkamp R., Thompson R., Hill W., and Simm G. (1997) Anim. Sci. 65(03):353-360. https://doi.org/10.1017/S1357729800008559

Salah Khattab A., Atil H., and Badawy L. (2005) Arch. Anim. Breed. 48(1):24-31. https://doi.org/10.5194/aab-48-24-2005

Samaraweera A.M., van der Werf J.H.J., Boerner V., and Hermesch S. (in press) J. Anim. Breed. Genet. https://doi.org/10.1111/jbg.12667

Samaraweera A.M. (2020) Development of a temperate dairy cattle breeding programme in Sri Lanka using milk, fertility and udder health traits. PhD thesis. University of New England Armidale, Australia.

Samaraweera A.M., Boerner V., Cyril H.W., van der Werf J., and Hermesch S. (2020) Asian Austral. J. Anim. 33(11):1741-1754. https://doi.org/10.5713/ajas.19.0798

van der Werf J.H.J. (2018) Software MTINDEX-multiple trait selection index 20 trait version. Available at https://jvanderw.une.edu.au/software.htm.

Wahinya P.K., Jeyaruban G., Swan A., and Magothe T. (2020) J. Anim. Breed. Genet. 137:495-509. https://doi.org/10.1111/jbg.12473

458. Challenges on construction and use of female reference populations for genomic selection under Indian smallholder systems

V. Ducrocq[1*], M. Swaminathan[2] and J.P. Gibson[3]

[1]Université Paris Saclay, INRAE, AgroParisTech, GABI, 78350 Jouy-en-Josas, France; [2]BAIF Development Research Foundation, India; [3]Centre for Genetic Analysis and Applications, University of New England, Armidale, NSW 2350, Australia; vincent.ducrocq@inrae.fr

Abstract

BAIF Development Research Foundation is a large Indian NGO helping smallholder farmers, in particular through artificial insemination of their crossbred cows at their doorstep. With international support, BAIF has created for the past 5 years a large female reference population of *Bos taurus* × *Bos indicus* cows in order to enhance the genetic merit of this crossbred population through genomic selection. We review the overall context of the project and its objective, the characteristics of the reference population and the choice of the recorded traits. Efficient ways to use the results of genomic evaluations are proposed, underlying the necessary changes required, in particular in the management of the bull stud. Finally, opportunities of upscaling this initiative to other partners in India are discussed.

Introduction

The Indian subcontinent is expected to contribute to more than half of the growth in world milk production over the next ten years (OECD-FAO Agricultural Outlook 2020-2029) and it may provide more than 30% of world milk production in 2030, with 80 to 95% coming from smallholder herds of a few cows or buffaloes. In order to limit detrimental effects on the climate and environment, the increasing demand should not be met by simply increasing the number of cows but by improving the genetic merit of the existing ones. Obviously, the existing cattle populations cannot meet such a challenge: the pure 'exotic' *Bos taurus* breeds are not adapted to the local harsh environments, while indigenous animals – often *Bos indicus* cows or buffaloes – often have a very low productivity. Hopefully, this can change: in developed countries, the development of affordable genomic tools led to a doubling of yearly genetic gains at a lower cost than previously (Boichard *et al.*, 2016). It is now time to adapt these tools to countries in (sub)tropical areas: the constitution of a representative genotyped and phenotyped female population can pave the way to genomic selection of crossbred animals in order to improve production in low-input systems at a more reasonable cost than before (e.g. Ducrocq *et al.*, 2018, Mrode *et al.*, 2019).

BAIF (BAIF Development Research Foundation, https://baif.org.in/) is a large Indian NGO located at Pune, India. BAIF manages 2 large bull studs producing 16 million semen doses of Holstein-Friesian, Jersey and pure *B. indicus* breeds as well as *B. taurus* × *B. indicus* crossbred bulls and various breeds of buffaloes. BAIF has a network of several thousands of artificial insemination (AI) technicians delivering AI at the doorstep of small to marginal farmers. With the financial support of the Bill and Melinda Gates Foundation (BMFG, www.gatesfoundation.org) and scientific support from the Australian Universities of New England (UNE) and of Queensland (UQ) as well as INRAE, France, BAIF is developing a female reference population (FRP) of more than 20,000 crossbred cows in smallholder farms (average herd size: 2 cows). Details on the information collected can be found in Bhave *et al.* (this congress).

This communication reports some lessons learnt during the constitution of this FRP and the anticipated future challenges.

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_458, © V. Ducrocq *et al.* 2022

Materials & methods

Characteristics of the BAIF reference population. BAIF has a long experience of interaction with smallholders in various Indian states through its CDCs (Cattle Development Centers covering 8 to 10 villages, with an AI technician in charge). Recruitment of farmers for performance recording was done in selected CDCs, initially in 6 states (Bihar, Jharkhand, Maharashtra, Odisha, Punjab, and Uttar Pradesh – with the planned addition of 3 others: Andra Pradesh, Gujerat and Rajasthan).

The degree of crossing is most often unknown or very inaccurate because of absence of pedigree recording. Even *B. indicus* composition is seldom known (leading to 'non-descript animals'). However, proportion of Jersey versus Holstein-Friesian vs *B. indicus* breed composition can be determined accurately through genotyping (Figure 1), even with 500 SNP (Strucken *et al.*, 2021a). Interestingly, while close to 50% of SNP molecular genetic variation is between breeds in *B. taurus*, only about 2% of variation is between breeds in *B. indicus*, meaning that *B. indicus* breeds are very difficult to distinguish from each other using SNP data (Strucken *et al.*, 2021b). In practice, all animals in a farm are enrolled, whatever their breed composition.

The area covered by the FRP is characterized by contrasting climate and farming systems, even within most of the states considered in the project. Large variations exist between states in level of % *B. taurus*, in level of production (see Guintard *et al.*, this congress) but also for other traits (e.g. body weight. A key challenge is how to properly account for G×E sources, not only on production but also on other traits.

Choice of recorded traits. The initial focus of the project was on milk production because it is the main and most immediate source of income and it is easy to record. This is also the priority at the Indian federal level. For example, use of semen of purebred *B. taurus* bulls is currently permitted only when the dam of the bull has one recorded lactation above 9,000 kg, a level of production difficult to achieve locally without preferential treatment! Such restrictions favor imports from developed countries of semen which is completely unsuitable for Indian conditions. In fact, in the BAIF context with a continuous range of crossbreeding level, milk production could be regarded as the easiest trait to improve, simply by tuning the breed composition of the new generation in terms of % Holstein or % Jersey. To achieve this, BAIF have developed and is testing in a pilot program a cheap SNP assay with few hundred SNP to routinely determine breed composition of new animals (Stricken *et al.*, 2021a). Such strategy is often overlooked but has the advantage to leave room to stronger selection on other economically important traits. The traits which are currently recorded at BAIF (see Kaustubh *et al.*, this congress) are fat% (a major determinant of milk price in India), protein%, milk quality (somatic cell counts assessed through a simplified recording system with 5 categories), body weight (calculated based on tape measures), udder conformation, feet and legs, age

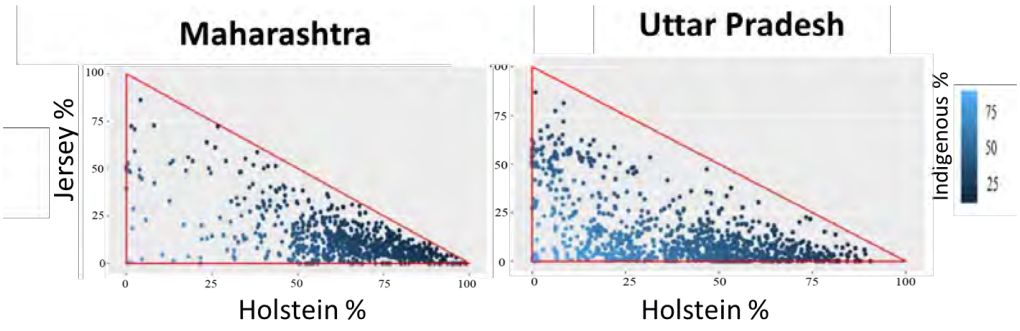


Figure 1. breed composition of the female reference population in two Indian states (here based on imputed high density genotypes).

at first calving, female fertility, etc. The range of some of these traits is much larger than what is found in developed countries (e.g. adult weights from 250 to 600 kg with huge variations between and within breed composition classes and between states). In some cases, important recorded traits are currently not usable due to unsystematic recording (e.g. mastitis incidence) but such issues can progressively be corrected. Moreover, for better acceptance in the field, other traits that are not necessarily relevant in developed countries may need to be also considered. For example, milking ease and temperament when cows are hand-milked by women should not be overlooked.

From genomic evaluation to actual selection at BAIF. Initial genomic analyses of the current FRP – so far essentially on milk production alone – illustrate the feasibility of genomic evaluations in the Indian context (e.g. Al Kalaldeh *et al.*, 2021), as long as specific features relative to smallholders are properly accounted for. As an example, among others, the joint modelling of the local environment and the breed composition in farms with on average two cows cannot be done using a usual fixed herd-year-season effect.

However, how to properly *use* the results of these evaluations to generate a genetic trend in the entire population of inseminated cows should not be neglected. The fact that BAIF has a very large and nationally significant semen production and insemination in the field is a strong point for dissemination of genetic progress. Nevertheless, to be effective requires huge strategic changes at various levels. For instance, the recruitment of a large number of young bull dams to conceive young males should be based on genomic evaluation results without excessive extra constraints (such as the regulations on minimum levels of production mentioned above). Male (but also female) progeny of these cows should be genotyped and get their own genomic evaluation as soon as possible and a drastic selection intensity should be applied to keep only the best ones. Indeed, it would be counterproductive to keep most of the genotyped animals. (There may be a tendency to assume that the animals are good simply because they were genotyped). Then, semen collection (for males) or AI (for females) should start as early as possible. Indeed, in developed countries, a large part of the increase in genetic gain due to genomic selection will be due to a drastic reduction of the generation interval and a fast turnover of bulls proposed for AI. This is a complete revolution compared with the current situation where crossbred bulls are used for a very long period.

Discussion

Upscaling genomic selection to other Indian populations. To increase and broaden the BAIF reference population, a new BMFG 5-year project started in late 2021, including three new partners (State Government breeding organisations). Extension of genomic selection to other populations of pure *B. indicus* breeds as well as buffalo breeds are also considered in a near future. Other Indian large organisations are following the same objectives (e.g. National Dairy Development Board – NDDB) and it is hoped that reference populations will be merged at some point, as happened in Europe with the Eurogenomics consortium (www.eurogenomics.com) or in North America. This will have multiple benefits: (1) an overall reduction of basic costs (under the condition that the same genotyping chip per species is used and that performance recording of the same traits are comparable); (2) a much larger shared reference population leading to more accurate genomic evaluations; (3) the possibility to eventually have a unique genomic evaluation center run by competent staff members, forming a group large enough to avoid collapse in case of departures of key players; (4) a secure connexion with a large international scientific community to help in capacity building and in issuing realistic recommendations to governmental authorities.

Interested readers can view a video overview of the whole BAIF project as presented at the FAO/IAEA international symposium on sustainable animal production and health (on-line) June 28-July 2, 2021 J.P. Gibson, O.A. Mwai, M. Swaminathan. Genetic improvement of smallholder dairy cattle for sustainable dairy systems. Traditional and genomic approaches. streaming.iaea.org/21644 (from 3:42:37 to 4:12:02)

References

Al Kalaldeh M., Swaminathan M, Gaundare Y. Joshi S, Aliloo H, Strucken E.M., Ducrocq V, Gibson J.P. (2021) Genet. Sel. Evol. 53:73. https://doi.org/10.1186/s12711-021-00667-6

Boichard D., Ducrocq V., Croiseau P., Fritz S. (2016). C. R. Biologies 339 (2016) 274–277. http://doi.org/10.1016/j.crvi.2016.04.007

Ducrocq V, Laloe D, Swaminathan M, Rognon X, Tixier-Boichard M, Zerjal T (2018). Front. Genet. 9:251. https://doi.org/10.3389/fgene.2018.00251

Lund M.S, De Roos A.P.W., De Vries A.G., Druet T., Ducrocq V. Fritz S., *et al.* (2011) Gen. Sel. Evol., 43:43 https://doi.org/10.1071/AN11119

Mrode R, Ojango JMK, Okeyo AM, Mwacharo JM (2019) Front. Genet. 9:694. https://doi.org/10.3389/ fgene.2018.00694

OECD-FAO Agricultural Outlook 2020-2029. Available at: www.oecd-ilibrary.org/sites/aa3fa6a0-en/index.html

Strucken E.M., Swaminathan M, Gibson J.P. (2021a) J Anim. Breed. Genet. 138:698. https://doi.org/10.1111/jbg.12544

Strucken, E.M., Gebrehiwot, N.Z., Swaminathan, M., Joshi, S., Al Kalaldeh, M. and Gibson, J.P. (2021b) Genet Sel Evol 53:47 https://doi.org/10.1186/s12711-021-00640-3

459. Bridging the gap in data from smallholder dairy systems; developing the Africa dairy genetic gains (ADGG) data platform

J.M.K. Ojango[1*], A.M. Okeyo[1], R. Mrode[1,2], E. Chinyere[1], G. Gebreyohanes[1], S. Meseret[1], D. Mogaka[1], E. Lyatuu[1], D. Komwihangilo[3], G. Msuta[3], S. Kahumbu[4], J.P. Gibson[5] and S. Kemp[1]

[1]International Livestock research Institute, Box 30709-00100, Nairobi, Kenya; [2]Scotland's Rural College, Easter Bush, EH25 9RG, Edinburgh, United Kingdom; [3]Tanzania Livestock Research Institute, Box 834 Dodoma, Tanzania; [4]Green Dreams Tech, Box 24844-00502 Nairobi, Kenya; [5]University of New England, NSW 2351 Armidale, Australia; j.ojango@cgiar.org

Abstract

Functional and long-term breeding programs with comprehensive recording of phenotypic information from smallholder dairy production systems in Africa are few, limiting herd productivity gains and profitability. To bridge the gap, a pilot information and communication technology-based platform for collation, evaluation and feedback on phenotypic and genomic data from smallholder farms in Tanzania and Ethiopia has been developed through the Africa Dairy Genetic Gains (ADGG) program initiated in 2016. In 2020, pedigree data from 65,624 and 110,959 animals in Tanzania and Ethiopia respectively was evaluated to develop restricted selection indexes and a genomic prediction pipeline for animals in the unique smallholder production environments. Integrated into ADGG, the iCow platform provides digital extension messages to farmers, sharing results from the genomic evaluation and breeding management support for improving productivity of their dairy animals. The genomic evaluation of crossbred dairy animals enables selective breeding for enhanced productivity in smallholder farming systems of Africa

Introduction

In rural areas of Africa, dairying is a mainstay for the majority of resource-poor households and a way out of poverty towards food security. It is estimated that 80% of the milk produced in the continent is from smallholder production systems (FAO, 2016), however comprehensive recording of animal performance is limited. Only two of the 44 countries in Sub-Saharan Africa are reported to have comprehensive national livestock identification, monitoring and recording systems in place in the State of the World Animal Genetic Resources reports (FAO, 2015; AU-IBAR, 2019). Limited information is available on animal productivity levels, costs of production and breeding technology adoption in smallholder farming systems (Rege *et al.*, 2011).

Data collation from dairy farming systems in Africa over time is implemented through small scale piloting of pre-designed tools and database management systems from countries with more advanced dairy sectors (Mrode *et al.*, 2020b). However, though successful with larger scale farming operations, infrastructure limitations and an inability to adequately identify, classify and evaluate the combinations of cross-bred animals reared in the smallholder farming systems restricts buy-in and use of data to inform decisions.

Most smallholder farmers in Africa own low milk-producing indigenous cows, and the few with genetically improved crossbred cows suffer heavy productivity losses due to the practice of continuous upgrading with exotic semen to levels that are beyond what the production systems can optimally support (Galukande *et al.*, 2013; Ojango *et al.*, 2017). The lack of functional local data collection and analytics systems that enable effective and sustained selection and use of uncertified crossbred bulls for breeding results in random backcrossing of crossbred cows to uncharacterized and less-productive bulls. The net result is continuous erosion of some of the gains from crossbreeding with long-term implications on the entire population.

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_459, © J.M.K. Ojango *et al.* 2022

Rapid expansion of information and communication technologies (ICT) alongside mobile based solutions for 'paperless' exchange of information has catalyzed a transformation for information sharing in the smallholder farming systems of Africa (World Bank, 2011; Mrode *et al.*, 2020b). Using ICT, from 2016 the African Dairy Genetic Gains (ADGG) project has sought to address limitations in data from smallholder dairy systems to identify and promote the use of more productive dairy genetics suited to the local production environments. This paper presents information on the multi-country dairy genetic data platform (ADGG-Platform) bridging the gap in information on the system characteristics, productive performance and genomic information on animals reared in African farming systems.

Methodology

The ADGG program initially chose Ethiopia and Tanzania as pilot countries to demonstrate how technological and institutional innovations can be integrated into interventions that enhance dairy productivity across diverse smallholder farming systems. In Sub-Saharan Africa, Ethiopia has the largest cattle population (65 million), while Tanzania has the third largest cattle population (30 Million) (FAOSTAT, 2022), reared by large numbers of smallholder farmers. The ADGG program supported through the Bill and Melinda Gates Foundation established centralized Dairy performance recording centre's (DPRC) in each country through which an initial 12,000 smallholder dairy herds were enrolled by trained national performance recording agents (PRA's). Desktop, Web, mobile tablet and sms based solutions were developed with dynamically defined user roles, using a RESTful Application Programming Interface (API) to safely authenticate and expose data from diverse sources through a single access point with individual country data mirrored at respective DPRC (https://portal.adgg.ilri.org/). The ADGG platform is designed to collate, process, interpret and generate feedback on both phenotypic (animal and environment descriptors, productive and reproductive performance measures, animal management practices and health interventions) and genomic (hair samples for genotyping animals) data from herds (Figure 1). Through partnership with Green Dreams Tech, the ADGG platform is integrated with iCOW (https://icow.co.ke/) to provide digital extension support and feedback from data analyses for farmers using text messages.

Results

Ethiopia and Tanzania have harmonized national cloud-based databases for registering and monitoring dairy cattle productivity that enable genetic evaluation of dairy animals across all farming systems in the countries. The number of farmers registered, and animals monitored over time through the ADGG platform are presented in Table 1. In 2021, up to 200 smallholder farmers enrolled in the program have begun to directly submit data on their animals onto the ADGG platform.

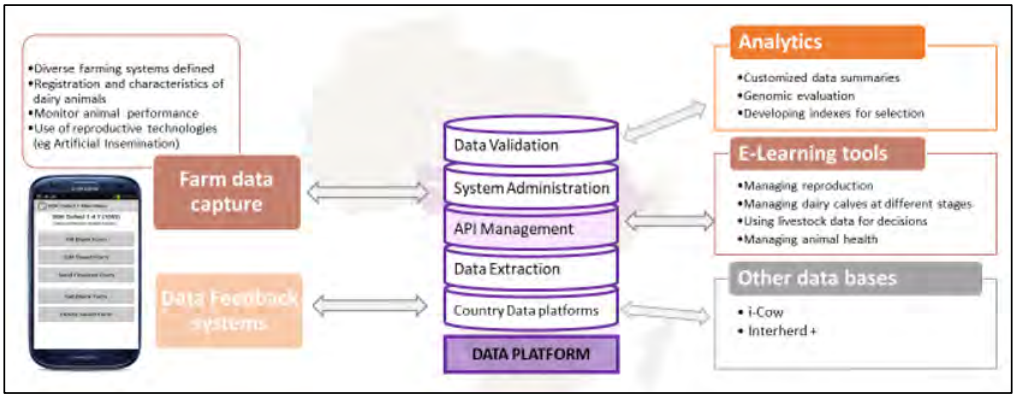


Figure 1. Overview of the ADGG data platform.

Table 1. Numbers of Farmers and animals registered, monitored and genotyped on the ADGG platform 2017-2021.

	Smallholder					Medium and large scale			
		Animals					Animals		
Country	Farms	Registered	With milk data	Test-day records	Animals sampled for genotyping	Farms	Registered	With milk data	Test-day records
Ethiopia	71,215	105,924	32,716	67,929	8,940	772	19,268	7,076	97,291
Tanzania	33,630	66,917	18,076	134,987	6,000	36	2,071	69	750
Total	104,845	172,849	50,792	202,916	14,940	808	21,339	7,145	98,041

Using genomic evaluation methodologies (Mrode *et al.*, 2019), phenotypic and genomic information on animals is evaluated annually on a country basis to determine best crossbred sires and cows for national use. In 2020, using pedigree files with 65,624 animals in Tanzania and 110,959 animals in Ethiopia, milk yield and body weight data for registered animals were evaluated and used to develop a restricted selection index to increase milk production with no change in body weight in order to improve efficiency such that animals produce more milk with less feed for maintenance (Mrode *et al.*, 2020a). Results of genomic prediction analyses for Ethiopia data were used to develop the first animal catalogue and certification system for Ethiopia (ILRI, 2021).

A genomic prediction pipeline developed for animals in the unique production environments (Mrode *et al.*, 2021) was used in preliminary principal component analyses of pooled genomic data for 2,363 Tanzania and 1,642 Ethiopian cows with milk records. Results indicate some degree of relatedness between animals in the two countries that will enable the implementation of regional joint evaluation of animals.

Medium and large-scale farmers who keep larger numbers of animals are known to provide breeding stock and replacement animals for smallholder farmers in Africa. Through combining data from animals reared in these systems with that from the smallholder farming systems more complex genomic prediction models can be explored in analyses to determine best animals for different management conditions.

Through policy dialogues with key stakeholders and government institutions at country level, frameworks and critical infrastructure, including partnerships with telecommunication companies have been developed at country level to support institutionalization of ICT systems for monitoring animal performance on smallholder farms (Gebreyohanes *et al.*, 2021).

Discussion

The ADGG data platform presents the first large scale source of genomic data and information on dairy cattle from smallholder farming systems in Africa, aligned to the Agenda 2063 of the African Union (AU) which seeks to harness Africa's animal genetic resources to address food and nutrition security in Africa. Through the ADGG training, information and feedback mechanisms, livestock keepers continue to adapt improved standards for producing clean milk, improve the quality of data on their animals, and adapt improved breeding and related management practices for their dairy animals. The growing data enables more accurate and reliable evaluation of crossbred dairy animals under smallholder dairy systems.

For the first time results from genomic evaluation of dairy animals reared under smallholder farming systems of Africa are available and being used to inform better use of crossbred dairy animals and provide top ranking locally adapted breeding bulls. Partnership arrangements with international genetic companies are underway to enable wider promotion and use of ADGG certified bulls.

References

AU-IBAR (2019) The State of Farm Animal Genetic Resources in Africa. AU-IBAR, Nairobi, Kenya.

FAO (2015) The Second Report on the State of the World's Animal Genetic Resources for Food and Agriculture. FAO, Rome, Italy.

FAO (2016) Dairy Producion and products. FAO, Rome Italy.

FAOSTAT (2022) FAO Statistical database. FAO,Rome, Italy. https://www.fao.org/faostat/en/#data/QCL accessed 04-01-2022

Galukande, E., Mulindwa, H., Wurzinger, M., Roschinsky, R., Mwai, A. O., *et al.* (2013) Animal Genetic Resources 52; 111–125 https://doi.org/10.1017/S2078633612000471.

Gebreyohanes, G., Meseret, S., Raphael, M., Ojango, J., & Ekine, C. *et al.*, (2021) Proceedings of 29th Ethiopian Society of Animal Production (ESAP) Conference. Addis Ababa, Ethiopia.

ILRI. 2021. Bulls and cows directory in Ethiopia. ILRI, Nairobi, Kenya. https://cgspace.cgiar.org/handle/10568/113558

Mrode, R., Ojango, J.M.k., Chinyere, E. & Mwai, O. (2020a). The Platform for Africa Dairy Genetics Gain (ADGG) New Index for dairy cattle in Tanzania and guidelines on its use. Nairobi, Kenya. https://africadgg.files.wordpress.com/2020/08/adgg-tzn-index-2020.pdf

Mrode, R., Dzivenu, C. E., Marshall, K., Chagunda, M. G. G., Muasa, B. S., *et al* (2020b). Animal Frontiers 10, 6–11. https://doi.org/10.1093/af/vfaa002.

Mrode, R., Ojango, J., Ekine-Dzivenu, C., Aliloo, H., Gibson, J., *et al.* (2021) Journal of Dairy Science 104, 11779–11789. https://doi.org/10.3168/jds.2020-20052.

Mrode, R., Ojango, J. M. K., Okeyo, A. M., & Mwacharo, J. M. (2019). Frontiers in Genetics. 9 https://doi.org/10.3389/fgene.2018.00694.

Ojango, J. M. K., Mrode, R., Okeyo, A. M., Rege, J. E. O., Chagunda, M. G. G., *et al.* (2017). Improving smallholder dairy farming in Africa.Pages 337–362 in Achieving sustainable production of milk Volume 2: Safety, quality and sustainability. van Belzen, N., ed. Burleigh Dodds Science Publishing Limited, Cambridge, UK.

Rege, J. E. O., Marshall, K., Notenbaert, A., Ojango, J. M. K., & Okeyo, A. M. (2011) Livestock Science 136; 15-28. https://doi.org/10.1016/j.livsci.2010.09.003.

World Bank. 2011. ICT in Agriculture e-Sourcebook. https://openknowledge.worldbank.org/handle/10986/12613

460. Attitudes of Peruvian llama and alpaca farmers towards genetic improvement

M. Wurzinger* and G. Gutiérrez

Universidad Nacional Agraria La Molina, Avenida La Molina s/n, Lima, Peru; mwurzinger@lamolina.edu.pe

Abstract

This study aimed to assess llama and alpaca farmers' attitudes towards genetic improvement and compare them with currently implemented breeding practices. After data quality control, 76 interviews were considered for the llama survey and 82 for the alpaca survey. Breeders' attitudes were assessed through a series of attitudinal statements. In addition, information about farmers' profiles and their current breeding practices was collected in the online surveys. The majority of the participants use controlled mating and performance recording, whereas artificial insemination, parentage testing, embryo transfer can be neglected as the technology is not well developed for both species. Purchase of animals from other farmers is the most common sourcing of breeding stock. Overall, farmers had positive attitudes towards the use of breeding and reproduction technologies. They also agreed that comparing and exchanging breeding animals is useful and being a member of an association helps in genetic improvement.

Introduction

Llamas and alpacas play an essential role in the rural economy of the Peruvian High Andes. So far, there are some isolated initiatives to develop and implement breeding programs, but no national breeding programs are in place for both species (Gutiérrez *et al.*, 2018).

The successful implementation of breeding strategies depends on various factors. The readiness of farmers to actively participate in breeding programmes has been identified as an essential factor among others (Kosgey and Okeyo, 2007, Wurzinger *et al.*, 2011, Getachew *et al.*, 2018). However, farmers' willingness is influenced by various factors such as their knowledge, perception and attitude towards the proposed new technology. Farmers' attitudes are affected, in turn, by many personal factors (e.g. gender, age, marital status), socioeconomic factors (e.g. income, assets, education), and personality factors (e.g. self-confidence, readiness for innovation) as well as familiarity with the technology (Meijers *et al.*, 2015). In recent research, farmers' attitudes towards different breeding tools and how they can influence their interest in breeding programs have been investigated (Martin-Collado *et al.*, 2021, Zoma-Traoré *et al.*, 2021).

This study aimed to assess the attitudes of llama and alpaca farmers in the Peruvian Andes towards genetic improvement and contrast them with the current breeding practices.

Materials & methods

The survey for llama farmers was conducted between November and December 2020 and for alpaca farmers from April to May 2021. Both surveys were conducted online due to the COVID-19 pandemic. Invitations with a link to the survey were sent by email and to topic-specific WhatsApp groups. Contacted persons were also encouraged to forward the invitation to other interested farmers. Due to the online format, there is a limitation that farmers without internet access could not participate and therefore total inclusivity could not be ensured.

In total, 80 people answered the llama survey, and 85 people answered the alpaca survey. After data quality control, 76 interviews were considered for the llama survey and 82 for the alpaca survey. Breeders' attitudes were assessed through a series of attitudinal statements, towards which breeders stated their agreement,

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_460, © M. Wurzinger and G. Gutiérrez 2022

using a five-level Likert scale: strongly disagree, disagree, somewhat disagree, somewhat agree, and agree. The option 'I do not know/have no opinion on this' was also included.

The list of attitudinal statements was defined from a study carried out by Martin-Collado *et al.*, 2021. The questions were adapted to local conditions taking into account local practices.

The final list consisted of 11 questions in which they should express their degree of agreement or disagreement. Additional information on the farmer' profile and currently used genetic improvement strategies and tools were also collected.

For data analysis, the questionnaires from departments Huánuco, Junín and Pasco were grouped into the 'Central region' and departments Arequipa, Ayacucho, Cusco, Huancavelica and Puno in the 'South region'. Descriptive statistics were performed using R software.

Results

Table 1 gives an overview of the age and education level of the interviewees from both surveys. The age range was between 21 and 70 years. The education level showed a high proportion (47-66%) of participants with a university education. Full-time but also part-time farmers took part in the survey.

The current use of breeding and reproduction technologies is summarized in Table 2. The results of the different study regions, but also both species, are very similar. The most common breeding practice is controlled mating, followed by keeping production records of individual animals. Farmers hardly use parentage testing, artificial insemination and embryo transfer.

63 and 26% of the llama farmers and 37 and 45% of the alpaca farmers from the Central and South region stated that they are currently a member of a farmer association.

Llama farmers specified that the most frequent source for a breeding male was the direct purchase from another farmer (37 and 39%), followed by selection within own herd (29% in both regions). 21% of the

Table 1. General information about llama and alpaca farmers.

	Llama farmers		Alpaca farmers	
	Central	South	Central	South
	n=38	n=38	n=19	n=63
Age of farmers, year				
mean	39	36	46	40
range (min-max)	21-64	20-65	32-61	20-70
Education (% of respondents):				
primary	5	3	0	2
secondary	15	14	47	15
technical training	21	35	6	17
university	59	48	47	66
Time for livestock keeping (% of respondents)				
≤50%	25	37	16	38
>50%	25	46	47	39
full time	50	17	37	23

Table 2. Current use of breeding and reproduction technologies (% of respondents).

	Llama farmers		Alpaca farmers	
	Central	South	Central	South
	n=38	n=38	n=19	n=63
Pedigree recording	16	24	26	27
Performance recording	55	66	47	77
Controlled mating	82	76	79	86
Parentage testing	3	11	10	5
Artificial insemination	3	0	16	8
Embryo transfer	5	5	5	8

respondents stated that they use a mix of different sources. Purchase at an animal fair, exchange or rent of breeding males were less frequently mentioned.

Alpaca farmers indicated that the most common practice is the direct purchase from another farmer (68 and 41%), followed by selection within own herd (21 and 23%). In the southern region, a combination of different sources was mentioned by 18%, whereas 28% of farmers from the central region used different sources. Exchange or rent of breeding males plays only a minor role as three farmers only mentioned it.

Table 3 shows the attitudes of llama and alpaca farmers towards different breeding tools. Llama and alpaca farmers value performance more than the 'beauty' of their animals (statement 9). There is the K'ara type (meat type) and Chaku type (woolly type) in llamas. As these types are phenotypically very different and also have distinct production purposes, farmers would like to maintain these types and avoid crossbreeding. The same results can be observed in the alpaca survey, where farmers confirm their preference to maintain the Huancaya and the Suri type. These types show very distinct fibre characteristics. Farmers of both surveys stated that performance recording is needed (statement 2, 4) to improve their animals. Also, in both surveys, reproduction technologies were seen as useful and genomic information was rated as an important method in animal breeding (statement 3, 6). Statements 7, 10 and 11 are concerned with farmers' attitudes

Table 3. Attitudes of llama and alpaca farmers towards genetic improvement.[1]

Statements	Llama farmers		Alpaca farmers	
	Central	South	Central	South
	n=38	n=38	n=19	n=63
1. Important to maintain type	3.7±1.84	3.9±1.79	3.7±1.82	4.3±1.47
2. Phenotype sufficient to indicate performance	3.1±1.50	3.3±1.41	3.0±1.41	2.8±1.36
3. Genomic information is fastest tool	3.9±1.67	4.0±1.62	2.8±2.17	4.3±1.03
4. Performance records not needed	2.1±1.40	1.8±1.24	2.2±1.23	1.7±1.05
5. Phenotype of offspring good indicator	3.7±1.74	3.3±1.48	3.6±1.49	3.6±1.39
6. Reproduction technologies are not useful	1.7±1.60	2.8±1.70	3.1±1.49	2.6±1.52
7. Intra-herd comparison is important	4.1±1.44	4.3±1.27	3.7±1.45	4.2±1.30
8. Avoid crossbreeding of types	3.7±1.73	4.0±1.51	2.8±1.98	4.3±1.41
9. Phenotype more important than production	3.0±1.42	2.9±1.33	2.4±1.43	2.9±1.38
10. Exchange of animals is important	4.1±1.58	4.2±1.39	3.8±1.65	4.3±1.24
11. Membership in association is helpful	3.9±1.53	4.0±1.52	3.2±1.69	3.5±1.61

[1] 5=fully agree, 4=somehow agree, 3=somehow disagree, 2= disagree 1=completely disagree, 0=I don't know.

towards cooperation. In all three statements, farmers showed positive attitudes towards cooperation through animal comparison and exchange and joining an association.

Discussion

A very high number of farmers indicated that they have secondary or even university education, and these figures are far above those reported by Gutierrez *et al.* (2012). The sampling technique could explain this bias as farmers with an already greater interest in improving their herd management were willing to participate in the survey. Therefore, all results of this study should be interpreted with care. The current use of breeding and reproduction technologies did not differ between llama and alpaca farmers. In both species, artificial insemination and embryo transfer are not well developed as in other livestock and therefore not implemented routinely on farms. Natural mating is the method of choice, and therefore a high proportion of farmers use controlled mating.

Overall, farmers showed positive attitudes towards different breeding tools. Performance recording was seen as a helpful tool and is already implemented by more than 50% of the participants. The positive attitude towards the exchange of animals is also reflected in the current practice of purchasing breeding animals from other farmers. Being a member of an association was also perceived positively, although not all interviewees were already members of one. The study shows that farmers are open to new technologies and are interested in the genetic improvement of llamas and alpacas.

Acknowledgements

The author would like to thank all llama and alpaca farmers for taking their time to participate in the survey and to share their opinion with the research team.

Funding

This research was funded by project Concytec-World Bank 'IMAGEN–Innovaciones en la mejora genética altoandina: alpacas y llamas' E038-01, through its executing unit ProCiencia [contract number: 029-2019-FONDECYT-BM-INC.INV].

References

Getachew T., Haile A., Rischkowsky B. (2018) Proc. of the 11th WCGALP, Auckland, New Zealand.

Gutierrez G., Gutierrez J.P., Huanca T., Wurzinger M. (2018) Proc. of the 11th WCGALP, Auckland, New Zealand.

Gutierrez G., Mendoza A., Wolfinger B., Quina E., Rodriguez A., *et al.* (2012) VI Congreso Mundial de Camélidos Sudamericanos, Arica, Chile.

Kosgey I.S., Okeyo A.M. (2007) Small. Rumin. Res. 70: 76–88. https://doi.org/10.1016/j.smallrumres.2007.01.007

Meijer S.S., Catacutan D., Ajayi O.C., Sileshi G.W., Nieuwenhuis, M. (2015) Inter. J. Agri. Sustain. 13: 40–54. https://doi.org/10.1080/14735903.2014.912493

Martin-Collado D., Diaz C., Benito-Ruiz G., Ondé D., Rubio A., *et al.* (2021) Animal 15(2): 100062. https://doi.org/10.1016/j.animal.2020.100062

Wurzinger M., Sölkner J., Iñiguez L. (2011) Small. Rumin. Res. 98:170–175. https://doi.org/10.1016/j.smallrumres.2011.03.035

Wurzinger M., Gutiérrez G., Sölkner J., Probst L. (2021) Front. Vet. Sci. 8:613505. https://doi.org/10.3389/fvets.2021.613505

Zoma-Traoré B., Probst L., Ouédraogo-Koné S., Soudré A., Ouédraogo D., *et al.* (2021) Sustainability 13, 2499. https://doi.org/10.3390/su13052499

461. Development of a low-density SNP panel for local Ghanaian and Tanzanian chicken ecotypes

M. Walugembe[1], E.N. Amuzu-Aweh[2*], K.-S. Lim[1], Y. Wang[3], N. Chouicha[3], T. Kelly[4], A.P. Muhairwa[5], B.B. Kayang[2], P.L.M. Msoffe[5], A. Naazie[2], H.R. Otsyina[6], J.S. Mushi[5], R.A. Gallardo[4], H. Zhou[3], S.J. Lamont[1] and J.C.M. Dekkers[1*]

[1]Department of Animal Science, Iowa State University, Ames, IA, USA; [2]Department of Animal Science, University of Ghana, Legon, Accra, Ghana; [3]Animal Science, University of California, Davis, Davis, CA, USA; [4]School of Veterinary Medicine, University of California, Davis, Davis, CA, USA; [5]Department of Veterinary Medicine and Public Health, Sokoine University, P.O. Box 3000 Chuo Kikuu, Morogoro, Tanzania, [6]School of Veterinary Medicine, University of Ghana, Legon, Accra, Ghana; jdekkers@iastate.edu

Abstract

Genetic selection is important in choosing animals for the next generation. Animal breeders now choose selection candidates based on DNA markers, in addition to animal performance. There has been increased use of high-density SNP chips to enable selection of commercial broiler and layer chickens. However, genotyping a large number of selection candidates still carries a high cost. The aim of our study was to develop an economical 5K SNP panel for local Ghanian and Tanzanian chicken ecotypes using targeted genotyping by sequencing (GBS) for imputation to higher density. Low-density panel SNPs were selected from haplotype blocks and other important genomic regions across the chicken genome based on 600K SNP panel genotypes. A total of 188 birds were genotyped by GBS and an in-house shell script pipeline was utilized to obtain SNP calls. Our in-house pipeline was compared to a standard company pipeline and 600K SNP chip genotypes for validation. Selected SNPs were evenly distributed across the genome, with at least one SNP in each megabase region. Comparison of the two pipelines revealed a good genotype match for the 5K SNP panel. The 5K GBS panel and SNP calling pipeline are important tools to aid selective breeding in African chicken ecotypes.

Introduction

In the past decade, there has been an increase in the use of commercial high-density (HD) SNP chips for genotyping in chickens (Kranis *et al.*, 2013), which has accelerated genomic selection in broiler and layer breeding. The availability of genomic and phenotypic information enables genomic prediction of breeding values for selection candidates. Although HD SNP chips are commonly used, the costs to genotype a large number of selection candidates remains high. To reduce costs, the use of low-density (LD) panels followed by imputation to a high density could be a feasible approach (Habier *et al.*, 2009). Array-based genotyping platforms are commonly utilized in genotyping various livestock species. However, they include SNPs that may not be geographically representative and therefore, population diversity and estimates of recombination rates may be biased (Pérez-Enciso and Ferretti, 2010; Zhan *et al.*, 2011). In poultry, such studies have mostly been conducted in commercial layers and broilers, with limited focus on local African chicken ecotypes. New cost-effective next generation sequencing platforms such as genotyping by sequencing (GBS) have recently been developed. GBS was originally developed for plants (He *et al.*, 2014), and follows a simple protocol that provides high SNP coverage (Sonah *et al.*, 2013). The objective of this study was to develop a low-density SNP panel for Ghanaian and Tanzanian local chickens using targeted GBS.

Materials & methods

Experimental birds. Breeding flocks were established at the University of Ghana, Legon, Accra, Ghana, and Sokoine University of Agriculture, Morogoro, Tanzania. Breeder chickens from three Tanzanian (Kuchi, Morogoro Medium, and Ching'wekwe) ecotypes and three Ghanaian (Coastal Savannah, Forest and Interior Savannah) ecotypes were obtained.

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_461, © M. Walugembe *et al.* 2022

SNP panel genotyping and quality control. Blood samples were collected from 1,440 Ghanaian and 1,399 Tanzanian chickens and stored onWhatman FTA cards (Sigma-Aldrich, St. Louis, MO, United States). All birds were genotyped using the Affymetrix Axiom® 600k Array (Thermo Fisher Scientific Inc., Calsbad, CA, USA) at GeneSeek (Lincoln, NE, USA) and Gallus gallus genome version 5 was used for annotation of the genotyping array. Quality control on the combined genotype data was performed using PLINK 1.9 (Chang et al., 2015) and SNPs were screened based on minor allele frequency (MAF)≥0.1 and call rate >95%. A total of 401,083 SNPs were utilized in the downstream analyses.

Selection of SNPs for the 5K GBS panel. SNPs for the 5K GBS panel were obtained across the chicken genome from various sources including haplotype block estimates based on the 2,839 birds, SNPs in the MHC regions provided by Hy-Line International., genome-wide association studies on the commercial Hy-Line Brown laying line (Rowland et al., 2018; Saelao et al., 2019) and in African local chicken ecotypes (Walugembe et al., 2019; Walugembe et al., 2020), and 379 SNPs selected across the genome to ensure that there was at least one SNP in each Mb region. Imputation for the combined genotype data was performed using Fimpute (Sargolzaei et al., 2014) to fill in missing genotypes. Haplotype blocks were estimated in PLINK 1.9 (Chang et al., 2015) using the following options; --blocks, --blocks-min-maf 0.1, and --blocks-strong-lowci 0.6. A total of 69,622 haplotype blocks were generated across the genome and 4,500 SNPs were selected based on these haplotype blocks. The number of SNPs selected for a chromosome was computed as x*4500/401803, where x is the number of 600K SNPs on the chromosome that passed quality control.

Sequencing and SNP calling. Blood samples were also collected from an additional 92 Tanzanian and 96 Ghanaian chickens, which were sequenced for the 5K low-density GBS panel with 100 bp paired-end on 4 lanes using Illumina Hiseq at GeneSeek. Of these, 25 birds were also genotyped on the Affymetrix Axiom® 600k Array. Sequence data were processed using an in-house pipeline by Interval Bio (Mountain View, CA, USA). With input from Interval Bio staff, we developed our own pipeline to process the raw sequence read data and target variant calling using shell scripts and various publicly available software tools (BWA, SAMtools, PICARDS, and BCFtools). The raw reads were aligned to the *Gallus gallus* 5 reference genome, and genotypes were obtained from the vcf file using in house python script. SNP genotype calls from our pipeline were compared to those of the Interval Bio pipeline for all 188 birds and to the 600K SNP genotypes for the 25 birds.

Results

Selection of SNPs for the 5K GBS panel. Initial selection of SNPs utilized LD based pruning methods, where SNPs in high LD were retained after pruning. Although this resulted in selection of SNPs that were distributed across the genome, most had MAF close to 0.5. This approach method was, therefore, replaced by the haplotype blocks method. The 4,500 SNPs selected based on haplotype blocks were evenly distributed across the genome, with MAF ranging from 0.1 to 0.5, as illustrated in Figure 1. Each Mb included at least one SNP and for regions where this was not achieved, a SNP was selected and added to the panel. Haplotype blocks were determined to contain SNPs that are in very strong LD with each other to minimize intra-block variance and maximize inter-block variance.

Comparison of SNP genotype calls. Figure 2 shows the comparison of genotype calls based on our and the Interval Bio pipeline. SNPs were ranked by the percentage of the 188 birds for which the genotype was called the same by both pipelines (Match%). For most of the 188 birds, the same genotype (Match%) was called by both pipelines, with a few cases where neither pipeline or only one of the pipelines made a genotype call. We removed about 80 SNPs where a different (Non-match%) genotype was called by the two pipelines for more than 20% of the birds.

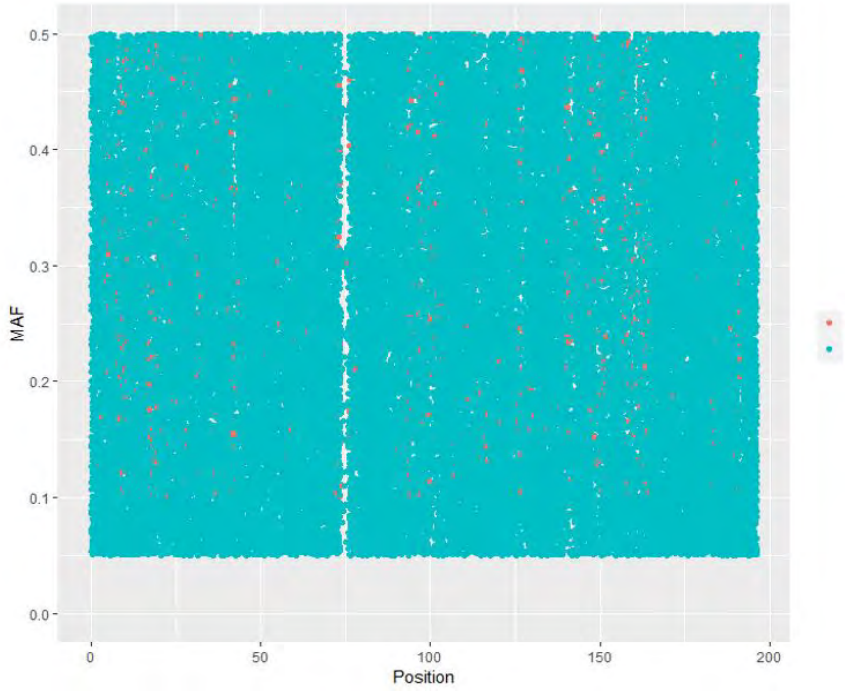


Figure 1. Distribution of the 805 selected SNPs (red) from the 71,725 SNPs that passed quality control on chromosome 1.

Figure 2. SNPs ranked by the % of birds for which the same genotype was called by both pipelines (light blue) along with the % for which different genotypes were called (red), the % for which only our pipeline did not (purple) or did (yellow) make a genotype call, and the % for which no call was made by both pipelines (dark blue).

Both pipelines had matching genotype calls with the 650K genotypes for the 25 birds that were genotyped on both the GBS and the 650K panel. These findings confirmed our low-density SNP panel list that would subsequently be used for genotyping on a GBS platform/protocol.

Discussion

To our knowledge, this is the first low-density SNP panel developed for African local chicken ecotypes, which can enable selective breeding of Africa local chickens using a cost-effective GBS. The panel has been used to genotype animals on a low density panel and imputation to a high density panel will be done using available imputation software programs.

The targeted GBS methodology employed here is one of the current methods of next generation sequencing for genotyping of SNPs and the availability of an excellent reference genome provides the basis for the selection of GBS restriction enzymes (de Donato *et al.*, 2013; Gurgul *et al.*, 2019).

Acknowledgements

This study was made possible by the generous support of the American people through the United States Agency for International Development (USAID) Feed the Future Innovation Lab for Genomics to Improve Poultry (cooperative agreement number AID-OAA-A-13-00080).

References

Chang, C. C., Chow, C. C., Tellier, L. C., Vattikuti, S., Purcell *et al.*(2015). *GigaScience*, *4*(1), 7. https://doi.org/10.1186/s13742-015-0047-8

de Donato, M., Peters, S. O., Mitchell, S. E., Hussain, T., & Imumorin, I. G. (2013). *PLoS ONE*, *8*(5). https://doi.org/10.1371/journal.pone.0062137

Gurgul, A., Miksza-Cybulska, A., Szmatoła, T., Jasielczuk, I., Piestrzyńska-Kajtoch, *et al.* (2019). *Genomics*, *111*(2), 186–195. https://doi.org/10.1016/j.ygeno.2018.02.002

Habier, D., Fernando, R. L., & Dekkers, J. C. M. (2009). Genomic selection using low-density marker panels. Genetics, 182(1), 343–353. https://doi.org/10.1534/genetics.108.100289

He, J., Zhao, X., Laroche, A., Lu, Z. X., Liu, *et al.* (2014). In *Frontiers in Plant Science* (Vol. 5, Issue SEP). Frontiers Media S.A. https://doi.org/10.3389/fpls.2014.00484

Kranis, A., Gheyas, A. A., Boschiero, C., Turner, F., Yu, *et al.* (2013). *BMC Genomics*, *14*(1). https://doi.org/10.1186/1471-2164-14-59

Pérez-Enciso, M., & Ferretti, L. (2010). *Animal Genetics*, *41*(6), 561–569. https://doi.org/10.1111/j.1365-2052.2010.02057.x

Rowland, K., Wolc, A., Gallardo, R. A., Kelly, T., Zhou, *et al.* (2018). *Frontiers in Genetics*, *9*(August), 1–12. https://doi.org/10.3389/fgene.2018.00326

Sargolzaei, M., Chesnais, J. P., & Schenkel, F. S. (2014). *BMC Genomics*, *15*(1). https://doi.org/10.1186/1471-2164-15-478

Saelao, P., Kelly, T., Lamont, S., Gallardo, R., Zhou, *et al.* (2019). Genes, 10(1), 61. https://doi.org/10.3390/genes10010061

Sonah, H., Bastien, M., Iquira, E., Tardivel, A., Légaré, *et al.* (2013). *PLoS ONE*, *8*(1). https://doi.org/10.1371/journal.pone.0054603

Walugembe, M., Amuzu-Aweh, E. N., Botchway, P. K., Naazie, A., Aning, *et al.* (2020). *Frontiers in Genetics*, *11*. https://doi.org/10.3389/fgene.2020.00739

Walugembe, M., Mushi, J. R., Amuzu-Aweh, E. N., Chiwanga, G. H., Msoffe, *et al.* (2019). *Genes*, *10*(7). https://doi.org/10.3390/genes10070546

Zhan, B., Fadista, J., Thomsen, B., Hedegaard, J., Panitz, *et al.* (2011). *BMC Genomics*, *12*. https://doi.org/10.1186/1471-2164-12-557

462. Combining demographic, production and molecular investigations in genetic improvements: case study in indigenous chicken

G.L.L.P. Silva[1*], M.S. Kurukulasuriya[2], P.R.M.K. Fernando[1] and C.M.B. Dematawewa[1]

[1]*Department of Animal Science, Faculty of Agriculture, University of Peradeniya, 20400, Sri Lanka;* [2]*Department of Animal Science, Faculty of Animal Science and Export agriculture, Uva Wellassa University of Sri Lanka, Badulla, Sri Lanka;* *pradeepas@agri.pdn.ac.lk*

Abstract

A study was carried out to assess how knowledge on genetic diversity and molecular analysis could be interpreted in the context of management system and production levels to develop genetic improvement strategies, considering the case of five indigenous chicken populations in Sri Lanka. Combination of molecular investigation with the production and environmental parameters was found to be important to assess the value of contribution of individuals to the diversity and adaptability. Our findings suggested that rearing system of indigenous chicken is the primary component that helps preserving genetic diversity, and the genetic makeup of the individuals play an important role in adaptability than in determining the productivity.

Introduction

Contribution of poultry meat and egg to global animal protein consumption is expected to increase to 38% by 2030 (Wahyono and Utami, 2018). Chicken meat and eggs are the best source of quality protein, particularly eggs carrying 87 NPU (Net Protein Utilization). Chicken products are the primary protein source for most of the rural communities in developing countries where poultry production is mainly based on indigenous ecotypes and scavenging backyard production system (Dematawewa, 2006). As a result of adaptation to varying environmental and management regimes, and due to low level of directional selection pressures, the indigenous poultry show high level of diversity. The pressures of production and economic standards prevailed in the industry have led to create a vacuum in continuous genetic progress of future poultry industry. While the commercial industry is currently shifting its attention more on health, product quality, and animal welfare, which now require access to diverse genotypes (Blackburn, 2006), the non-commercial or indigenous chicken populations are subjected to gradual genetic erosion due to lack of sustainable utilization programs. These two dichotomies would make the industry genetically stagnated in the future.

Despite the overall growth of commercial poultry industry, indigenous chickens remain as a predominant family industry in rural areas, owing to the balance between input and output levels under rural socio-economic conditions (DAPH, 2020). Scavenging or backyard system is the predominant practice found in indigenous chicken rearing in Sri Lanka (Gunaratne *et al.*, 1993; Silva, *et al.*, 2014 and 2016), where both nutritional and genetic improvements need to be addressed. In this context the wealth of molecular information generated so far to understand the extent and level of diversity needs to be interpreted in the context of the wide array of management systems and environmental regimes. This study attempted to relate the knowledge on genetic diversity and molecular analysis with the existing management and production standards to develop a strategy for improvement based on five indigenous chicken populations in Sri Lanka.

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_462, © G.L.L.P. Silva *et al.* 2022

Materials & methods

A two-fold investigation of 132 sampling sites in five different regions in Sri Lanka; namely North Central province (NCP), North Western province (NWP), Western province (WP), Uva province (UP) and Southern province (SP) was carried out. One part of the study consists of molecular characterization of chicken by using Single Nucleotide Polymorphism (SNP) in mitochondrial D-loop of indigenous chicken from different geographical regions of Sri Lanka (results have been described elsewhere), and the other part of the study is to relate the molecular findings with the production and management aspects in development of sustainable utilization strategy.

The sampling locations or villages within the five regions were selected randomly. Sampling was carried out from households scattered within the village, and only one sample was taken from one household to avoid relatives been sampled for molecular analysis, incorporating diverse production and management information as far as possible.

The administration of questionnaire was carried out by a single visit to each household and multiple visits were made to regional veterinary office for data verification on farming conditions of the area.

Results

The molecular investigations related to the present study have been reported elsewhere (Silva *et al.*, 2008). The molecular findings based on the SNP analysis identified a comparatively high diversity with 42 haplotypes (31 unique haplotypes) belonging to six haplogroups, without a particular pattern of distribution with respect to geographical regions studied. Thus, the indigenous chickens in Sri Lanka are a mixture of diverse genotypes distributed all over the country.

The molecular diversity showed no association with the farming duration, purpose of farming, flock size, management practices (Table 1) and production standards (Table 2).

Table 1. Comparison of the management practices in different haplogroups identified among indigenous chicken in Sri Lanka.

Management criteria	Category	Haplogroup					
		A	B	C	D	E[1]	F[1]
Purpose (%)	Egg for hatching	79.5	90.9	81.8	80	100	100
	Egg for sale	18.2	6.8	13.6	20	0	0
	Egg and meat	2.3	2.3	4.6	0	0	0
Duration of farming (number)	≤1yr	6	6	3	1	1	0
	2-10 yrs	25	21	14	12	4	2
	11-20 yrs	9	6	3	0	0	0
	≥20 yrs	4	3	1	1	1	0
Replacement stock (%)	From own farm	95.5	100	95.5	100	100	100
	From outside	4.5	0	4.5	0	0	0
Feed base (%)	Scavenging only	68.2	45.5	68.2	40	60	100
	Scavenging + supplement	31.8	54.5	31.2	60	40	0
Health management (%)	No medicine	65.1	79.1	63.6	73.3	80	100
	With medicine	34.9	20.9	36.4	6.7	20	0

[1] Rear haplogroups with number of observations of ≤5.

Table 2. Comparison of the production levels in different haplogroups identified among indigenous chicken in Sri Lanka.

Production Criteria	Haplogroup					
	A	B	C	D	E[1]	F[1]
Age at lay (w)	27.1±2.3	27.5±2.3	26.8±2.4	27.1±2.3	27.2±.1	26.0
Eggs/week	4.8±1.6	5.2±1.7	5.5±1.7	4.2±1.4	6.0±1.7	5.5
Clutch size	11.2±4.2	11.9±5.1	12.2±5.8	12.9±3.9	11.8±4.9	16.0
Clutch length (w)	3.0±1.8	2.8±0.9	3.1±1.7	4.1±2.9	2.5±0.6	3.0

[1] Rear haplogroups with number of observations of ≤5; only few observations.

Discussion

The management and production characteristics showed a random variation irrespective of the molecular categories (haplogroups), indicating no strong association of genetic categorization with performance of birds in village system. Given the fact that the within region variation of haplogroups is high and no regional specificity was observed in haplogroups (Silva *et al.*, 2008), this observation indicates that the level of performance of indigenous birds is mainly determined by the routine management criteria accepted by farmer communities. Thus, the genetic variation, as identified by the mitochondrial D-loop based haplotypes in this study, is not critical in determining the production in back-yard chicken rearing system. This could be a result of the system that allows no directional selection but stabilized natural selection. However, genotype is playing a key role in the process of adaptability (Barker, 2001), which is the major determinant in sustainability of the system. There was a high sustainability in the poultry production system, where most of the households continue to keep poultry for several decades. On the other hand, high adaptation to the environment by the population of birds reported here was confirmed by the less disease occurrence and absence of routine health care practices by the farming community. Thus, further molecular characterization needs to be done using the variability of the genes responsible for those characteristics.

The remarkably high adaptability of indigenous birds has been highlighted in many instances. This could be attributable to high genetic diversity existing among the indigenous chicken (Barker, 2001, Dessie *et al.*, 2011). O'Brien (1994) explained that genetic diversity can well be linked to the ability to respond to environmental changes. In a more specific study carried out to evaluate the level of diversity explained by the variation of mitochondrial DNA, Bazin *et al.* (2006) related the mitochondrial diversity with adaptive evolution of populations. Looking from a different dimension, Barker (2001) explained that responding to environmental factors or adaptation drives the process of maximizing the fitness, which consequently helps maintaining substantial genetic variance within population. The populations like Sri Lankan indigenous chicken, which evolve with varying environmental conditions, are easily adaptable and instrumental in conserving genetic variation through utilization.

Considering the high diversity in most livestock populations, Barker (2001) emphasized the value of multiple population breeding, where groups in each production system is considered as a separate breeding population. According to Namkoong (1998) the ensemble of breeding populations is a strategy for breeding to preserve production and adaptation in sustainable production system. This is basically because the populations in a system form factorial but not hierarchical breeding populations thus preserving the desired diversity (Namkoong, 1998). According to these observations the back-yard production system is the best arranged to preserve the genetic variation in the population. Therefore, the present study highlights

that the molecular investigations are to be interpreted in the context of the system with a proper evaluation on the trade-off of genes between adaptation and production.

References

Barker J.S.F. (2001). Conservation and management of genetic diversity: a domestic animal perspective. Can. J. For. Res. 31:588-595. https://doi.org/10.1139/cjfr-31-4-588

Bazin E., Glemin S. and Galtier N. (2006). Population size does not influence mitochondrial genetic diversity in animals. Science 312:570-572. https://doi.org/10.1126/science.1122033

Blackburn H. D. (2006). The National Animal Germplasm Program: challenges and opportunities for poultry genetic resources. Poultry Science 85(2):210-215. https://doi.org/10.1093/ps/85.2.210

DAPH (2020). Poultry Sector Analysis and Forecast. Division of Livestock Planning and Economics, Department of Animal Production and Health, P.O. Box 13, Getambe, Peradeniya, Sri Lanka.

Dematawewa C.M.B. (2006). Potential of village chicken in poverty alleviation. Sri Lanka Journal of Animal Production. Vol. 02 (Special Issue):25-52.

Dessie T., Taye T., Dana N., Ayalew W. and Hanotte O. (2011). Current state of knowledge on phenotypic characteristics of indigenous chickens in the tropics. Worlds Poult. Sci. J. 67(3):507–516. https://doi.org/10.1017/S0043933911000559

Gunaratne S.P., Chandrasiri A.D., Hemalatha W.A. and Roberts J.A. (1993). Feed resource base for scavenging village chickens in Sri Lanka. Trop. Anim. Health. Prod. 25(4):249-257. https://doi.org/10.1007/BF02250880

Namkoong G. (1998). The place of adaptation and production in genetic improvement programs for sustainable development. Proc. of WCGALP. 6(28):69-79.

O'Brien S.J. (1994). A role for molecular genetics in biological conservation. Proc. Natl. Acad. Sci. U S A. 91(13):5748-5755. https://doi.org/10.1073/pnas.91.13.5748

Silva P., Guan X., Ho-shing O. and Jones J. *et al.*, (2008). Mitochondrial DNA-based analysis of genetic variation and relatedness among Sri Lankan indigenous chickens and the Ceylon Jungle Fowl (*Gallus lafayetti*). Animal Genetics. 40 (1) 1-9. https://doi.org/10.1111/j.1365-2052.2008.01783.x

Silva P., Abeykone N.D.F., Samaraweera A.M. and Han J. L. *et al.*, (2014). Genetic Diversity and Adaptability Exist among Backyard Poultry Populations in Sri Lanka. Proc. of the 10th WCGALP, Vancouver, Canada.

Silva P., Liyanage R.P., Senadheera S. and Dematawewa, C.M.B. (2016). Monograph on indigenous chicken in Sri Lanka. UNEP-GEF-ILRI FAnGR Asia project, University of Peradeniya, Sri Lanka, pp 81.

Wahyono N.D. and Utami M.M.D (2018). A Review of the Poultry Meat Production Industry for Food Safety in Indonesia. J. Phys. Conf. Ser. 953(1):012125. https://doi.org/10.1088/1742-6596/953/1/012125

Part 27

Breeding goals and selection strategies

463. Relative importance of selection objective traits in participatory breeding programs – MultiAtr software

R.N.B. Lobo

Embrapa Caprinos e Ovinos, Caixa postal 71, Sobral, CEP 62010-970, CE, Brazil; raimundo.lobo@embrapa.br

Abstract

The definition of selection objectives in animal breeding programs in developing countries is not a simple task, due to the lack of economic and productive data. Participatory rural appraisal tools contribute to this process. MultiAtr software was developed through linear programming with the simplex method, and an approximation to PAPRIKA method. Breeders (n=42) in Brazil and Colombia (n=77), participants of participatory breeding programs (PBP), were interviewed to define the traits priorities in their selection objectives for the production of meat from goats and sheep and dairy goats' system, respectively. The algorithm was used to estimate the relative importance of the traits of the selection aims. In Brazilian PBP, the most important traits were number of days for slaughter and animal adult weight. In Colombia, the most important traits were resistance to diseases and parasites, milk quality and milk production.

Introduction

In most developing countries, animal production, especially goats and sheep industries, tends to be disorganized and ultimately informal. Most business relationships are temporary and inconsistent and absent from legal contracts. Therefore, traditional breeding programs have not been effective. The attempts made were discontinued and did not reach the expected success. The reasons for this are due to the low level of production, the variety of local genetic resources (breeds, ecotypes, crossbreed, etc.) and the informality of the markets. A new approach to developing breeding programs in developing countries, especially in poor rural communities excluded from industrial programs, are participatory community-based breeding programs (CBBP), which take into account the needs, visions, decisions, and participation of the farmers, from inception to implementation (Sölkner *et al.* 1998; Wurzinger *et al.* 2011; Mueller *et al.*, 2015). Even in these programs, the decision-making process is complex and often demands time and difficulties for managers, project coordinators and, especially here, specialists in genetic improvement programs. Defining the weights of selection aim for a CBBP is among these challenges.

'Point', 'point-count', 'score' and 'linear' systems are known as additive multi-attribute value models with performance categories. They are solutions to generalized problems of how to combine alternative features on multiple criteria to rank the alternatives. Hansen and Ombler (2009) presented a method for 'scoring' such models to determine their values, based on decision makers' preferences. This method is called by the acronym PAPRIKA (Potentially All Pairwise RanKings of all possible Alternatives), and involves the decision maker's peer ranking. There are several applications of this type of model, to classify specific alternatives known by decision makers, such as prioritizing patients for access to health care, classifying students who apply to higher education schools, selecting immigrants by immigration services, etc. This method could be applicable to define the importance and weights of selection objectives in participatory breeding programs, based on the alternatives chosen and prioritized by the interviewed participants. The aim of this study was to identify the relative importance of economically important traits in two breeding programs (Brazil and Colombia), according to the vision and preference of the main decision makers, that is, the breeders, using a software that applies an approach to the PAPRIKA method.

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_463, © R.N.B. Lobo 2022

Materials & methods

Information from two goats and sheep breeding programs was used, one in the semi-arid region of Brazil and the other in Colombia.

Brazil CBBP. The production system of this CBBP is for the production of meat from goats and sheep, with the animals managed in native pasture typical of the semiarid region of Brazil (Caatinga). Interviews were carried out throughout 2020, in seven districts of Tauá (four communities – Altamira, Lustal, Pitombeira and Veneza) and Parambu (three communities – Juazeiro, Novo Assis and Miranda), in the region of 'Sertões dos Inhamuns', in the state of Ceará, Brazil. All participants were adults and were between 18 and 89 years of age. A total of 42 breeders representing 122 families were interviewed in Tauá (nine in Altamira, five in Lustal, 12 in Pitombeiras and 16 in Veneza) and 40 breeders in Parambu (nine in Juazeiro, 18 in Miranda and 13 in Novo Assis). The traits of the selection objective, previously identified by the breeders were animal adult weight (WEI), average price of the animal at slaughter (AVP), fertility (FER), kids' survival (SUR), number of days for slaughter (DSL), number of deworming treatments (NDT), number of veterinary treatments (NVT), and animal weight loss in dry season (WTL). For each trait, three categories were defined to define the breeders' preferences (WEI = +5 kg on liveweight average, maintain average, -5 kg on liveweight average; AVP = +R$10.00 on the average price paid out by animal, +R$5.00 on the average price, maintain the current average price; FER = +10 kids/flock/year on the current average, +5 kids/flock/year on the current average, maintain the same average; SUR = +10 kids on the average of kids that survive until weaning/year, +5kids on this average, maintain the average; DSL = -2 weeks on average age at slaughter, -1 week on the average slaughter age, maintain the same average; NDT = -2 deworming/animal/year, -1deworming/animal/year, maintain the average; NVT = -2 treatment/animal/year, -1 treatment/animal/year, maintain the average; WTL = loss of until 5 kg on liveweight average, loss between 5 and 10 kg on liveweight average, maintain the current average).

Colombia participatory breeding program. In this country, the goat milk production system was considered. Interviews were carried out between August and November 2021, in three groups of interviewed, with the participation of 77 breeders, supported by Federación Nacional de Caprinocultores y Ovinocultores de Colombia (FEDEANCO), from different regions of the country (47 from Centro Andina Region, 11 from Caribe Region, six from Oriental Region, four from Coffee Axis Region, four from Pacific Region and five from other regions). The traits of the selection objective, previously identified by the breeders were adaptation (ADP), animal adult weight (WEI), fertility (FER), milk production (MLK), milk quality (MQU), mortality (MOR), persistence in lactation (PER), prolificacy (PRO), resistance to diseases and parasites (RES), and udder conformation (UDD). For each trait, three categories were defined to define the breeders' preferences (ADP = +20% on current average, +10% on the average, maintain average; WEI = +15 kg on liveweight average, +10 kg on average, maintain the average; FER = +20% on current average, +10% on the average, maintain average; MLK = +25% on current average, +15% on the average, maintain average; MQU = +25% on current average, +15% on the average, maintain average; MOR = -10% on current average, -5% on the average, maintain average; PER = +15% on current average, +10% on the average, maintain average; PRO = +15% on current average, +10% on the average, maintain average; RES = +30% on current average, +20% on the average, maintain average; UDD = +20% on current average, +10% on the average, maintain average).

MultiAtr software. The MultiAtr software was developed in FORTRAN language for use in command prompt (CMD). From the inequalities (strict preference) and equalities (indifference) corresponding to the explicitly classified pairs, the point values are obtained through linear programming with the simplex method according to an approximation of the PAPRIKA method. Although several solutions are possible for the linear program of inequalities and equalities, all the resulting point values reproduce the same

general classification of alternatives. For each respondent, the algorithm randomly presents the questions to be chosen with the combination between two objectives and two of their categories, in an inverse way in their levels. According to the example below:

****
WHAT IS YOUR PREFERENCE REGARDING THIS CHOICE? FERTILITY MORE 10% AND ADULT WEIGHT EQUAL IS ALREADY OR FERTILITY EQUAL IS ALREADY AND ADULT WEIGHT LESS 10KG TYPE 1 TO OPTION 1, 2 TO OPTION 2, 3 THEY ARE EQUAL, 4 TO SKIP AND 5 TO STOP
****

The maximum number of alternatives is defined by the number of categories to the power of the number of objectives. However, as the alternatives are arranged at different levels between the categories, according to the number of objectives and categories, the number of questions may not reach this maximum. Furthermore, according to the respondent's answers, some alternatives are eliminated as they are contradictory.

The algorithm was used to conduct interviews with participants in the aforementioned countries. Each participant was interviewed separately and their responses entered into the system. At the end of all interviews, the software estimated the priorities for selecting, according to the preference of their participants. Within each country, ANOVA of individual preferences, by region of that country, was performed to assess whether there were statistical differences between regional preferences.

Results

In Brazilian CBBP, the order of priority of the traits was DSL, WEI, FER, AVP, SUR, NVT, NDT and WTL (Table 1). Although there are some differences in this order between municipalities, and between districts, these differences are not statistically significant ($P>0.05$), and the general order should be considered.

In Colombian, for dairy goats' system, the order of preference of the traits, according to the perception of the breeders was RES, MQU, MLK, FER, PER, ADP, UDD, PRO, MOR and WEI (Table 2). Although there are some differences in this order between groups and between regions of the country (according to where the interviewees live), these differences are not statistically important ($P>0.05$), and the general order must be considered.

Table 1. Relative importance (%) of the traits in the objective of selection for meat production with goats and sheep in a Brazilian CBBP, according to the municipalities and districts.

	CBBP	Tauá districts					Parambu districts			
		Tauá	Altamira	Lustal	Pitombeira	Veneza	Parambu	Juazeiro	Miranda	Novo Assis
DSL	19.4 (1°)	20.9 (1°)	20.4 (1°)	20.2 (1°)	20.4 (1°)	21.7 (1°)	17.9 (1°)	16.4 (2°)	17.8 (1°)	18.7 (1°)
WEI	14.1 (2°)	13.2 (3°)	13.8 (2°)	10.2 (5°)	12.7 (5°)	14.3 (2°)	14.9 (2°)	17.2 (1°)	14.1 (3°)	14.9 (2°)
FER	13.4 (3°)	13.3 (2°)	11.7 (5°)	15.3 (2°)	15.2 (2°)	12.1 (4°)	13.6 (3°)	11.7 (5°)	15.9 (2°)	11.9 (4°)
AVP	11.5 (4°)	11.7 (4°)	12.1 (3°)	13.3 (3°)	8.2 (7°)	13.5 (3°)	11.3 (5°)	8.4 (8°)	11.6 (5°)	13.2 (3°)
SUR	11.2 (5°)	11.0 (6°)	9.7 (8°)	12.7 (4°)	13.0 (4°)	9.7 (6°)	11.6 (4°)	11.4 (6°)	12.2 (4°)	10.5 (6°)
NVT	10.9 (6°)	11.4 (5°)	9.8 (7°)	9.6 (7°)	14.2 (3°)	10.6 (5°)	10.2 (7°)	12.8 (3°)	9.6 (7°)	9.8 (8°)
NDT	9.8 (7°)	8.8 (8°)	11.8 (4°)	8.9 (8°)	7.0 (8°)	8.6 (8°)	10.7 (6°)	12.4 (4°)	10.6 (6°)	10.1 (7°)
WTL	9.7 (8°)	9.8 (7°)	10.6 (6°)	9.8 (6°)	9.5 (6°)	9.6 (7°)	9.8 (8°)	9.6 (6°)	8.2 (8°)	10.9 (5°)

Table 2. Relative importance (%) of the traits in the objective of selection for dairy goats in a Colombian participatory breeding program, according to the groups and regions.

	CBBP	Groups			Region					
		1	2	3	A	B	C	D	E	F
RES	12.5 (1°)	14.9 (1°)	10.9 (3°)	12.7 (1°)	11.3 (2°)	12.7 (1°)	9.4 (7°)	13.9 (1°)	15.9 (1°)	10.3 (4°)
MQU	10.9 (2°)	9.6 (7°)	10.9 (4°)	11.9 (2°)	8.5 (9°)	10.9 (3°)	9.6 (5°)	10.9 (4°)	15.0 (2°)	14.4 (1°)
MLK	10.9 (3°)	11.5 (2°)	10.9 (2°)	10.4 (3°)	11.0 (3°)	11.0 (2°)	11.0 (3°)	8.0 (9°)	12.8 (3°)	11.4 (3°)
FER	10.5 (4°)	10.0 (4°)	11.3 (1°)	9.8 (6°)	12.0 (1°)	10.0 (5°)	10.2 (4°)	9.7 (5°)	10.6 (4°)	13.9 (2°)
PER	10.1 (5°)	9.7 (6°)	10.6 (5°)	9.7 (7°)	10.6 (5°)	10.0 (6°)	13.9 (1°)	8.9 (7°)	9.7 (6°)	7.8 (9°)
ADP	10.0 (6°)	9.7 (5°)	9.9 (6°)	10.2 (4°)	9.8 (6°)	10.3 (4°)	9.3 (8°)	9.2 (6°)	9.8 (5°)	8.8 (7°)
UDD	9.7 (7°)	11.2 (3°)	9.8 (7°)	8.5 (9°)	9.7 (7°)	9.9 (7°)	11.5 (2°)	11.9 (2°)	5.7 (10°)	5.5 (10°)
PRO	9.2 (8°)	9.3 (8°)	8.7 (9°)	9.7 (8°)	11.0 (4°)	9.0 (8°)	9.6 (6°)	7.2 (10°)	8.5 (8°)	9.4 (6°)
MOR	8.9 (9°)	7.5 (9°)	9.0 (8°)	9.9 (5°)	9.4 (8°)	9.0 (9°)	7.2 (10°)	8.8 (8°)	9.0 (7°)	8.3 (8°)
WEI	7.5 (10°)	6.6 (10°)	8.4 (10°)	7.3 (10°)	6.8 (10°)	7.0 (10°)	8.5 (9°)	11.5 (3°)	6.0 (9°)	10.0 (5°)

[1] A = Caribe; B = Centro Andina; C = coffee axis; D = Oriental; E = Pacific; F = others.

Discussion

The stated preference tools are widely used for valuation of non-marketed goods and services (Hoyos, 2010). In the livestock sector, a number of researchers applied them to identify trait preferences and estimate economic values for the traits in the breeding objective. This is quite useful and valid in CBBP, especially with regard to involvement and recognition of breeders' perceptions. A robust tool was developed to estimate these preferences, based on the PAPRIKA method, which has proved to be of great value in decision making (Hansen and Ombler, 2009). MultiAtr was able to quickly and consistently estimate the perception of breeders, in two participatory breeding programs, and reveal the importance of the traits of the selection objective of these programs.

The breeders in the Inhamuns region, in Brazil, are more concerned with increasing the animals' growth speed, the final weight of these animals and with the general fertility of the herd, in order to increase their income from the production of goats and sheep. On the other hand, in Colombia, the interviewees showed greater interest in increasing the resistance of their animals against diseases and parasites, not neglecting the aspects of milk production and quality, lactation persistence, adaptation and fertility.

References

Hansen P., Ombler F. (2009) A New Method for Scoring Additive Multi-attribute Value Models Using Pairwise Rankings of Alternatives. J. Multi-Crit. Decis. Anal. 15: 87–107. https://doi.org/10.1002/mcda.428

Hoyos D. (2010). Ecol. Econ. 69, 1595–1603. https://doi.org/10.1016/j.ecolecon.2010.04.011

Mueller J.P., Rischkowsky B., Haile A., Philipsson J., Mwai A.O. *et al.* (2015) J. of Animal Breeding and Genetics 132(2):155–168. https://doi.org/10.1111/jbg.12136

Sölkner J., Nakimbugwe H., and Valle-Zárate, A. (1998) Proc. of the 6th WCGALP, Armidale, Australia.

Wurzinger M., Sölkner J., and Iñiguez L. (2011) Small Ruminant Research 98(1-3):170-175. https://doi.org/10.1016/j.smallrumres.2011.03.035

464. Validation of Angus HeiferSELECT using historical data

P.A. Alexandre[1*], L.R. Porto-Neto[1], B. Hine[2], A. Ingham[1], M. Samaraweera[3], C. Duff[3] and A. Reverter[1]

[1]CSIRO Agriculture and Food, 306 Carmody Rd., St Lucia, QLD 4067, Brisbane, Australia; [2]CSIRO Agriculture and Food, Chiswick, New England Highway, NSW 2350, Armidale, Australia; [3]Angus Australia, 86 Glen Innes Rd., NSW 2350, Armidale, Australia; pamela.alexandre@csiro.au

Abstract

Angus HeiferSELECT is a genomic tool that provides genomic estimated breeding values (GEBV) for thirteen traits, to better inform the selection of replacement heifers. Using historical data, we validated predictions for birth weight (BW), weaning weight (WW), yearling weight (YW) and mature cow weight (MCW), based on the animal's self-performance and the average performance of their progeny. GEBV accuracies were higher for animals in the reference population and when self-performance data was considered. Accuracies were higher for MCW, followed by BW, YW and WW. After ranking animals based on GEBV, average progeny performance in the top versus bottom quartiles differed by 1.7, 5.9, 9.5 and 25.0 kg for BW, WW, YW and MCW, respectively. These numbers were about double when quartile differences were based on self-performance. We conclude that Angus HeiferSELECT GEBV can predict differences observed at the phenotypic level, which can add precision to the selection of superior replacement heifers.

Introduction

Targeted selection of replacement females is crucial to maintaining the best herd structure in commercial beef enterprises. The decision of which heifers to keep in the operation and which to sell potentially affects the profitability of the herd for years to come. One approach is to breed all available females and identify those that breed early. An alternative is to wait until after they wean their first calf and make a more informed decision based on progeny performance. However, for most producers, it is impractical to keep females which are not destined to remain in the breeding herd. Strategies to identify replacement heifers earlier include selection based on their date of birth, bodyweight, phenotypic evaluation including structural assessment, and pedigree information (Meat & Livestock Australia 2022). However, the additional use of genomic information can provide a more reliable method.

Angus HeiferSELECT (Angus Australia 2021) is an advanced genomic tool developed to inform the selection of replacement heifers in commercial beef breeding operations. It includes GEBV for thirteen maternal, growth, feed intake, carcase, and resilience traits. Thus, the tool provides information on traits that are otherwise difficult, time-consuming or expensive to measure using traditional methods, complementing other sources of information for the selection of replacement heifers. Similar tools with a genomic application at commercial level include Angus SteerSELECT (Hine *et al.* 2021), for the drafting of Angus steers, and Neogen's Igenity® Beef (Neogen 2022), which predicts the genetic potential of crossbred beef heifers, steers, and commercial bulls.

The accuracy of GEBV depends on several factors such as the heritability of the trait under evaluation, the size of the reference population, and the relatedness between animals being tested and animals in the reference population (Goddard and Hayes 2009). Hence, the objective of this study was to use historical data to validate accuracies of Angus HeiferSELECT GEBV and their association with performance for a range of growth traits, including birth weight (BW), weaning weight (WW), yearling weight (YW) and mature cow weight (MCW), based on the animal's self-performance as well as that of their progeny.

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_464, © P.A. Alexandre *et al.* 2022

Materials & methods

Dataset. Data used in this study were accessed from the Angus Australia database. Data included: (1) HeiferSELECT GEBV for BW, WW, YW and MCW for all animals with genotype information in Angus Australia's Herd Book Register (HBR), Angus Performance Register (APR) or Angus Commercial Register (ACR) (n=153,978); (2) three generation pedigree information for all animals born from 2000 to 2021 (n=1,888,635); and (3) adjusted phenotypes for BW, WW, YW and MCW for all animals in the pedigree from a contemporary group of size ≥5 (Table 1). Phenotypes were adjusted for age at measurement, age of dam and contemporary group.

Accuracy analysis. For each trait, validation of GEBV (accuracy) was performed based on two Data Types: 1) Self-Performance: Pearson correlation between animals' GEBV and their own (adjusted) phenotype; and 2) Progeny Performance: Pearson correlation between parents' GEBV and the average (adjusted) phenotype of their progeny. In addition, different accuracy values were generated by grouping animals according to: (1) YOB: year of Birth was categorised into one of 5 categories: before 2000, 2001 to 2005, 2005 to 2010, 2011 to 2015, and 2016 or later; (2) Reference: whether or not the animals were included in the Angus HeiferSELECT reference population; (3) Trait: BW, WW, YW or MCW; and (4) Number of progeny: categorised from 1 to 24 and ≥25 for sires and 1 to 5 and ≥6 for dams.

ANOVA. An analysis of variance (ANOVA) was performed using PROC GLM (SAS Inst. Inc.) to identify significant effects influencing accuracy. Factors included in the ANOVA were Data Type, YOB, Reference, and Trait.

Q1Q4 quartile analyses. To evaluate the effect of genomic selection of parents on progeny performance, for each trait we ranked parents based on their GEBV. Then, we calculated the difference between the average progeny (adjusted) phenotype on the top and the bottom quartile, namely Q1Q4 measure, and compared it to the Q1Q4 measure based on self-performance.

Results

All factors considered (Data Type, YOB, Reference, and Trait), GEBV accuracy was estimated at an average of 0.40 (LSmeans). The ANOVA model explained 68.8% of the variation in the accuracy with the most

Table 1. Summary statistics for adjusted phenotypes, including Birth Weight (BW), Weaning Weight (WW), Yearling Weight (YW) and Mature Cow Weight (MCW), for data collected on the HeiferSELECT animals (self-performance) and the average of their progeny (progeny performance).

	N	Mean	SD	Min	Max
Self-performance					
BW	122,837	0.101	3.886	-23.0	23.4
WW	110,179	2.170	24.037	-182.1	173.9
YW	87,091	2.356	30.349	-201.1	195.0
MCW	7,993	2.662	50.104	-224.3	299.3
Progeny performance					
BW	32,499	0.035	3.084	-19.2	16.9
WW	26,283	0.642	19.567	-113.9	153.2
YW	21,458	0.723	23.867	-127.3	178.1
MCW	4,258	-0.070	41.390	-180.3	244.4

influential factors being Data Type ($P<0.0001$), followed by Trait ($P<0.0001$) and Reference ($P<0.001$). YOB was non-significant ($P>0.10$). Figure1 shows accuracies according to Data Type, Reference and Trait. There was a slight positive tendency for GEBV of sires and dams with a higher number of progeny to have higher accuracies, which was more apparent for BW (data not shown). The phenotypic differences between quartile one and quantile four (Q1Q4) for each trait, based on self and progeny performance, can be seen in Table 2.

Discussion

As expected, accuracies were higher for animals included in the reference populations and accuracies calculated based on self-performance were higher than those based on progeny performance. Accuracies were highest for MCW, followed by BW, YW and WW, which could have been influenced by differences in trait heritability, which follow the same trend (Torrez-Vázquez *et al.* 2018). When the parent was included in the reference population the accuracy was expected to be higher because its own (heritable) phenotype affects that of its progeny. However, no expectation existed about the impact of the number of progeny on the accuracy other than that the accuracy will be more accurately estimated when based on the results of more progeny. However, a slight increase in accuracies was observed for sires and dams with a higher number of progeny.

For the same trait, the Q1Q4 difference based on self-performance was about double that based on progeny performance, meeting inheritance expectations. For instance, animals ranked in Q1 for BW were, on average, 3.53 kg heavier than those ranked in Q4 when considering self-performance, while based on progeny performance the Q1Q4 diff was 1.71 kg. For MCW, Q1Q4 diff was 61 and 25 kg, respectively. Independent from the accuracy value for each trait, the difference between the top and bottom quartile could be translated into productivity differences, demonstrating the use of genomics to inform management decisions.

Figure 1. Accuracies based on self or progeny performance for animals, either in the Angus HeiferSELECT reference population (REF) or not (NOREF), for Birth Weight (BW), Weaning Weight (WW), Yearling Weight (YW) and Mature Cow Weight (MCW).

Table 2. Quartile analyses for Birth Weight (BW), Weaning Weight (WW), Yearling Weight (YW) and Mature Cow Weight (MCW) based on self-performance and average progeny performance when ranking parents by GEBV.

	BW (kg)		WW (kg)		YW (kg)		MCW (kg)	
	N	Mean	N	Mean	N	Mean	N	Mean
Self-performance								
Q1	30,709	1.84	27,544	8.59	21,772	12.20	1,998	33.02
Q2	30,709	0.56	27,545	3.75	21,773	4.80	1,998	9.26
Q3	30,709	-0.31	27,545	0.73	21,773	-0.24	1,998	-3.52
Q4	30,710	-1.69	27,545	-4.39	21,773	-7.34	1,998	-28.11
Q1Q4 Diff		3.53		12.98		19.55		61.13
Progeny performance								
Q1	8,124	0.86	6,570	3.59	5,364	5.45	1,064	12.92
Q2	8,125	0.28	6,571	1.61	5,365	2.04	1,065	2.14
Q3	8,125	-0.16	6,571	-0.36	5,365	-0.51	1,065	-3.28
Q4	8,125	-0.85	6,570	-2.26	5,363	-4.09	1,063	-12.07
Q1Q4 Diff		1.71		5.85		9.55		24.99

Angus HeiferSELECT predicts phenotypic differences in commercial Australian Angus cattle for BW, WW, YW or MCW, which was also reflected in the performance of their progeny. Therefore, this tool can add to the precision of selecting commercial replacement heifers that match the overall breeding objective when compared to the use of phenotypic measurements alone. The additional information provided by Angus HeiferSELECT can positively contribute to the productivity and profitability of commercial Australian Angus herds.

References

Angus Australia, HeiferSELECT: An advanced genomic tool to inform the selection of replacement heifers for commercial Australian Angus breeders. Available at: https://www.angusaustralia.com.au/content/uploads/2021/11/Angus_HeiferSelect_V2.pdf.

Goddard M. and Hayes B. (2009) Nat Rev Genet 10(6):381-91. http://doi.org/10.1038/nrg2575.

Hine B.C., Duff C.J., Byrne A., Parnell P., Porto-Neto L.R. *et al.* (2011) Anim Prod Sci 61(18):1884. https://doi.org/10.1071/AN21051

Meat & Livestock Australia, Cull and replace females. Available at: https://mbfp.mla.com.au/weaner-throughput/4-cull-and-replace-females/.

Neogen, Igenity® Beef. Available at: https://www.neogen.com/categories/igenity-profiles/igenity-beef/.

Torres-Vázquez J.A., van der Werf J.H.J. and Clark S.A. (2018) J Anim Sci 96(11):4521-4531. https://doi.org/10.1093/jas/sky325.

465. Full-scale simulation of the Norwegian red dairy cattle breeding programme to enable testing future breeding directions

A. Ehsani[1*], J. Jenko[2], H. Storlien[2], H. Melbo Tajet[2], I. Pocrnic[1], J. Obsteter[3] and G. Gorjanc[1]

[1]The University of Edinburgh, Roslin Institute, Midlothian, EH25 9RG, United Kingdom; [2]GENO SA, Storhamargata 44, 2317 Hamar, Norway; [3]Agricultural Institute of Slovenia, 1000 Ljubljana, Slovenia; aehsani@ed.ac.uk

Abstract

A full-scale simulation that can cover all aspects of an ongoing breeding programme is a useful tool to test future breeding decisions. For example, dairy breeding is facing new challenges to include feed efficiency and methane emission traits in their breeding goals. To this end we have simulated a real size breeding programme of the Norwegian Red dairy cattle with detailed population structures and breeding decisions over the last 50 years. We simulated traditional progeny test and genomic selection scheme using AlphaSimR package. The simulation involved about 0.5 million animals per year spread across 7,000 herds with variable size, five groups of lactating cows, calves, heifers, and young, waiting, and proven bulls. This In-silico simulation tool will enable Geno SA – Norwegian Red breeding organisation to test future breeding directions, such as, including feed efficiency and methane emission traits into a selection index.

Introduction

Predicting future breeding outcomes is challenging due to the complexity of breeding programmes. Simulations have been widely used to predict the impact of various breeding decisions on future genetic trends. However, most simulation studies such as Wallen *et al.* (2017), often oversimplify breeding programmes by omitting breeding actions, downscaling the size of a breeding programme, and ignoring heterogeneity among different sub-populations. Therefore, results of such simulations must be interpreted with these limitations in mind. Accounting for such complexities in a simulation set-up can help to find better solutions for future challenges.

Dairy cattle breeding programmes are constantly improving. To maximize genetic gain these new findings, need to be optimally integrated in their breeding programs. One example is the integration of expensive and difficult to measure traits, such as feed efficiency and methane emissions. Deterministic prediction equations can be used to decide how many animals should be recorded (e.g. Daetwyler, *et al.* 2008) to maximize return on investment. However, such equations cannot take population and breeding programme structure into account.

Norwegian Red dairy cattle (NR) have a unique population structure and breeding history. NR is growing in popularity due to its high merit for production and functionally balanced cow, which is a product of selection towards such a breeding goal for at least 50 years. To improve this population even more, breeding organization of NR is evaluating data collection strategies to incorporate new traits, such as feed efficiency and methane emission. To support this aim we developed a simulation of a real size breeding programme of NR.

Materials & methods

We developed a simulation of a full population size of NR breeding programme including all live animals along with every single sub-population. We simulated the detailed population structure such as number of the herds, different lactating groups of the cows, the newborn calves, still-born calves, heifers, waiting bulls for progeny test, genotyped male and female calves, progeny and genomic tested bulls, proven bulls, and

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_465, © A. Ehsani *et al.* 2022

heifers for embryo transfer, according to the real population size of NR. Moreover, we accounted for the breeding history of the past 50 years and matched simulated trends in genomic and phenotypic parameters to real trends from the breeding data.

Data. First, we simulated a base population according to the cattle demographic history with effective population size of 250 to reflect the historical NR base population. Next, we simulated the 50 years of breeding under the breeding organization of NR, Geno SA (https://www.norwegianred.com, Table 1). The first 45 years of the simulation spanned years 1971 to 2015, where the NR breeding programme used the conventional progeny test scheme, and years 46-50 of the simulation spanned years 2016 to 2020, where NR breeding programme used the genomic selection scheme. We simulated milk yield trait with heritability of 0.19; additive genetic, permanent environment, herd-year, and residual variances of respectively 0.25, 0.35, 0.25, and 0.45 in an active population of 200,000 cows in five lactating groups producing the same number of calves with 1.5% stillborn, and 48,000 heifers. In the progeny testing scheme, there were five groups of selected young bulls and waiting bulls with 115 individuals each, and 16 proven bulls in service. In the genomic selection scheme 8,000 males were genotyped and 50 bulls with the highest estimated breeding value were selected for mating each year. Moreover, 22,000 female calves were selected for genotyping based on their parent average, and the top 90 heifers were selected for embryo transfer every year. The population consisted of 7,000 herds of variable size, matching the herd size distribution of NR in Norway with a Conway-Maxwell Poisson distribution with the mean of 29 and SD of 23.

Statistical models. We used repeatability animal model (Mrode 2014) for genetic evaluation. The statistical model used for progeny test scheme was:

$$y_{ijk} = \mu + a_i + pe_i + hy_j + e_{ijk} \quad (1)$$

where y_{ijk} is the phenotype of the individuals, μ is the general mean, a_i and pe_i are the random animal and permanent environmental effect for animal i, hy_j is the random herd-year effect for j^{th} herd-year and, e_{ijk} is random environmental effect respectively following $N\sim(0, A\sigma^2_a)$, $N\sim(0, I\sigma^2_{pe})$, $N\sim(0, I\sigma^2_{hy})$, *and* $N\sim(0, I\sigma^2_e)$. Variance components were based on estimated variance components in the NR population for milk yield with heritability of 0.192. Genomic evaluations were performed using the single-step method combining all the phenotypic, pedigree, and genomic information (Legarra *et al.* 2014).

The breeding programme was simulated using the AlphSimR package (Gaynor, *et al.* 2020) in R software (R Core Team, 2014) and genetic/genomic evaluations were performed using MiX99 (MiX99 Development Team, 2019). Ten replicates of the simulation were run to evaluate the genetic progress, as well as other genetic and non-genetic parameters.

Table 1. Simulation parameters.

Object	Value	Object	Value	Object	Value
Lactating cows[1]	200,000	chromosomes	29	herds	7,000
calves	200,000	QTLs	8,700	Mean herd size	29
Progeny test bulls	115	SNPs	58,000	SD herd size	23
Genomic test bulls	8,000	PT years	45	Bulls in service (PT)	16
Genomic test heifers	22,000	GT years	5	Bulls in service (GT)	50
Embryo transfer heifers	90	Ne	200	Lactating groups	5

[1] All the values are for one year simulation.

Results

To start with the simulation for the NR breeding program, we draw the overall picture of the two main breeding schemes, the conventional progeny testing scheme (Figure 1, left) for 45 years from 1971 to 2015 and the genomic selection scheme from 2016 to 2020 (Figure 1, right).

We have simulated every single sub-population stated in the Figure 1 as well as their distribution across the 7,000 herds. As stated earlier, the herd sizes and the spread of the herds over Norway is unique and needs a special care for a proper simulation.

AlphaSimR can incorporate any user defined effects such as herd, herd-year, permanent environmental effects, etc. to individuals or any groups of individuals like herds or sub-population groups such as cows in different lactations.

Figure 2 is a code snippet for assigning cows to herds and associating corresponding herd effects to a group of lactating cows. By grouping cows by lactation (five groups) and heifers inheriting the dams' herd we can simply simulate herd population structure with five AlphaSimR objects.

Evaluated genetic progress and other genetic and population parameters such as the level of inbreeding and heterozygosity, as well as genetic and phenotypic variances matched the observed trends in real breeding programme. Even though we needed to evaluate around 0.5 million individuals every year, the population size was not an issue in terms of computational limitations until the genomic selection scheme started. 30,000 new genotyped calves every year, increases the memory usage quadratically and will soon reach to the limit. We dealt with this issue using singular value decomposition procedure applied in the Mix99 software.

Discussion

We have generated *in silico* breeding program for NR using the AlphaSimR stochastic simulator. This is presenting the fundament for testing the potential changes in the future breeding programme of NR. This simulation will be used in the future to optimize phenotyping for traits that are expensive to record such as

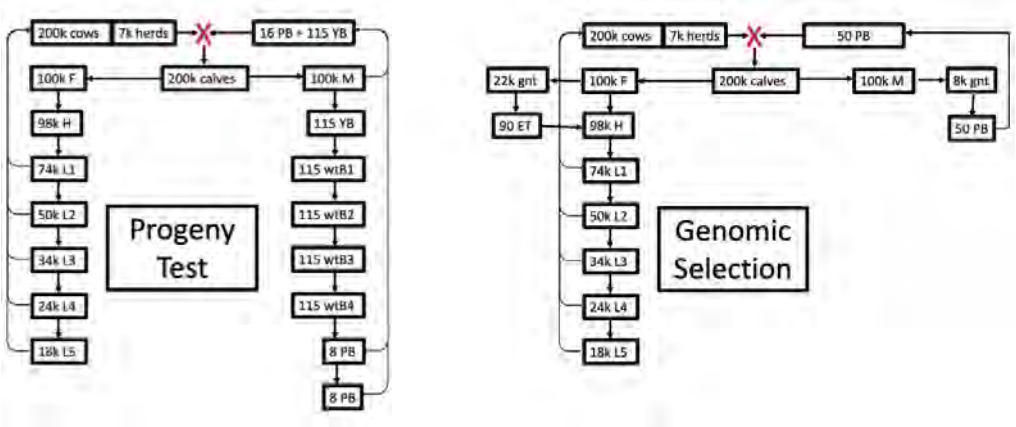


Figure 1. Population structure in the conventional progeny testing scheme (left) and the genomic selection scheme (right) for Norwegian Red dairy cattle breeding programme. PB = Proven Bulls, YB = Young Bulls, F = female calves, M = Male calves, H = Heifers, L1 to L5 refer to cows in lactations 1 to 5 respectively, wtB1 to wtB4 refer to waiting bulls in year 1 to 4 respectively, gnt = genotyped individuals, ET = Embryo transfer heifers.

```
Herd_Size = simcmp(n = (Number_Of_Herds),
                   v = c(Mean_Herd_Size, SD_Herd_Size))
Herd_Effects = rnorm(n = length(Herd_Size),
                     mean = 0,
                     sd = sqrt(Herd_Variance))
Herds_Data = data.frame(
  Herd = 1:length(Herd_Size),
  Herd_Size,
  Herd_Effect = Herd_Effects,
)
Assign_Herd_Effects = function(pop, Herds_Data) {
  pop@misc = vector(length = nInd(pop), mode = "list")
  misc_NULL = list(Herd = NA,
                   Herd_Effect = NA)
  Herd_Allocation = sample(
    x = Herds_Data$Herd,
    size = nInd(pop),
    replace = TRUE,
    prob = Herds_Data$Herd_Size
  )
  for (ind in 1:nInd(pop)) {
    pop@misc[[ind]] = misc_NULL
    pop@misc[[ind]]$Herd = Herd_Allocation[ind]
    pop@misc[[ind]]$Herd_Effect = Herds_Data[pop@misc[[ind]]$Herd, "Herd_Effect"]
  }
  pop
}
```

Figure 2. AlphaSimR code snippet to allocate cows to herds and associated herd effects. The distribution of the herd size and the herd effects for Norwegian red dairy cattle following respectively Conway-Maxwell-Poisson distribution and normal distribution.

feed efficiency and methane emission. The scenario with the maximum return on investment will be then suggested for the implementation in practice. We believe that having such a simulator is essential for every breeding programme, by which, every change in the future could be first tested *in silico* and then only the best solutions would be implemented in practice. After further testing and optimisation, we will share the AlphaSimR code with the research community.

Acknowledgements

Jenko J would like to acknowledge the support from the Research Council of Norway (proj. n. 330747). This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 801215 and the University of Edinburgh Data-Driven Innovation programme, part of the Edinburgh and South East Scotland City Region Deal.

References

Daetwyler, H. D., Villanueva B., and Woolliams J. A. (2008). PloS One 3 (10): e3395.

Gaynor, R. C., Gorjanc G., and Hickey J. M. (2020). AlphaSimR: G3 Genes|Genomes|Genetics 11 (2). https://doi.org/10.1093/g3journal/jkaa017.

Legarra, A., Christensen O. F., Aguilar I., and Misztal I. (2014). Genomics Applied to Livestock Production. 54–65. https://doi.org/10.1016/j.livsci.2014.04.029.

MiX99 Development Team (2019) MiX99: Release XI/2019. Available at: http://www.luke.fi/mix99

Mrode, R. A. (2014). CABI, Oxfordshire, UK.

R Core Team (2021). R Foundation for Statistical Computing, Vienna, Austria. URL:
https://www.R-project.org/.

Wallen, S. E., Lillehammer M., and Meuwissen T. H. E. (2017). Journal of Dairy Science. 100:6327-6336. https://doi.org/10.3168/jds.2016-11458.

466. Improving carcase value by incorporating primal weights into beef breeding objectives

N.R. Sarker*, B.J. Walmsley and S. Hermesch

AGBU, a joint venture of NSW Department of Primary Industries and University of New England, 2351, Armidale, Australia; nipa.bau02@gmail.com

Abstract

The impact selection for increased high-value primal cuts (HVC) has on improving beef profitability is important to explore. This study presented an approach that facilitates the extension of existing beef breeding objectives (BO) by including HVC. Economic values for HVC were derived for two different production systems: one pasture-based for the Australian domestic market (DPS) and the other grain-based for the Japanese market (JPS). The BO incorporated HVC on a percentage basis to maintain independence to carcase weight. The emphasis on HVC in the breeding objective was 8.78 and 10.77%, respectively, with annual predicted genetic gains from a mean of 0.10 and 0.18% in HVC for DPS and JPS, respectively. The addition of HVC increased the index response by 14.95 and 54.22% for DPS and JPS, respectively over the base scenario, which does not include primal cuts, indicating the potential of HVC to enhance selection for improved carcase value.

Introduction

In Australia, and most other countries, beef producers are currently paid on carcase weight, conformation and fat cover. These measurements provide the basis for payment to producers, with financial penalties for carcases that do not conform to desirable characteristics (Kenny *et al.*, 2020). The red meat yield, especially in primal cuts, dictates the economic and production efficiency at the processor level (Drennan *et al.*, 2009). Producing carcases with a higher proportion of red meat yield positively impacts beef processing profitability due to less fat and bone in the carcase. Overall, the proportion of primals, especially the high-value primals, is the main price determiner of a beef carcase for processors. The current beef breeding objectives do not include primal cuts. Therefore, the aim of this study was to derive economic values (EV) for high-value primal cuts (HVC) and predict responses to selection for HVC and other traits incorporated in current breeding objectives (BO) for Australian beef.

Materials & methods

Animal and primal cuts. Primal weights were recorded on 1,014 Angus cattle (876 steers and 138 heifers), from NSW Department of Primary Industries research herds. The animals were slaughtered between 2017 and 2020 at around 600 days of age and in lots of 23 to 54 animals per day. The primal cuts were classified into two groups based on their retail value; the HVC composed of the cube roll, striploin, tenderloin, rump, topside, intercostal and thick flank, and the low-value cuts (LVC) composed of the thin flank, hindquarter shank, forequarter shank, brisket, inside skirt, silverside, blade, rib set, chuck tender and chuck.

The breeding objective. In addition to the current BO traits, the BO included HVC to determine the value of increasing HVC at a fixed carcase weight. Breeding objectives were constructed for two terminal production systems relevant to Australian beef. One system was based on finishing steers off pasture for the Australian domestic market (DPS) with steers slaughtered at a target live weight of 620 kg. Marbling was assumed to have low value in this market. The other system was based on feedlot finishing steers for the high-quality Japanese market (JPS), where marbling was of high importance. JPS steers were slaughtered at a target live weight of 800 kg. Carcase prices at the farm-gate level were assumed to be $AUD5.40/kg and $AUD6.20/kg for the DPS and JPS markets, respectively.

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_466, © N.R. Sarker *et al.* 2022

Derivation of economic values for high-value cuts at farm-gate level. The BO defined HVC on a percentage basis rather than a weight basis to maintain independence from carcase weight. Firstly, the wholesale prices of different primal cuts were converted to farm-gate prices using the conversion factor k, so the derived EV of HVC was also at the farm-gate level. The factor k was calculated with Equation 1, Where P_p was the total farm-gate value of a carcase adjusted to a fixed weight of 305 kg or 350 kg for DPS and JPS, respectively.

$$k = P_p / \text{total weighted price of all primal cuts at the wholesale level} \quad (1)$$

Based on the weighted average of the difference between HVC and LVC, EVs were calculated by multiplying the price difference at farm-gate by 3.05 and 3.50 since a 1% increase in HVC represents 3.05 kg and 3.50 kg increases in HVC for DPS and JPS, respectively.

Beef index description. The BOs were developed with BreedObject (Barwick et al., 2005) and the selection criteria based around BREEDPLAN EBVs that align with terminal systems (Johnston et al., 1999). The EVs and parameters for BO (Table 1) and selection criteria (Table 2) traits used in this study were assembled from published (Archer et al., 2004; Wood *et al.*, 2004) and unpublished sources (pers. comm. Brad Walmsley, 2021). It was assumed that the live animal traits were recorded for the selection candidate, sire, dam and 25 half-sibs, while carcase traits except HVC were recorded for 25 half-sibs only. To investigate the effect of including HVC in beef breeding programmes, two different scenarios were modelled: one in which HVC was not included in the BOs, and another in which HVC was incorporated in the BOs without any relatives information for HVC. The annual response for each trait was predicted assuming an average generation interval of four years. Trait EVs were multiplied by their genetic standard deviation to show the relative importance of traits in BO.

Results

The HVC represented 8.78 and 10.77% of the BO emphasis for DPS and JPS, respectively (Table 3). This demonstrates that HVC could be important to include in future beef BOs. Inclusion of HVC in current

Table 1. Heritabilities (h^2), phenotypic standard deviations (PSD) and economic values for breeding objective traits for two Australian beef markets.

Traits	h^2	PSD	Economic value	
			Domestic	Japanese
Weaning weight – direct (kg)	0.10	25.10	0.02	0.09
Residual feed intake – pasture (kg/day)	0.40	1.10	-22.22	-16.33
Feedlot entry live weight (kg)	0.17	31.15	-	0.04
Residual feed intake – feedlot (kg/day)	0.35	1.32	-	-20.29
Sale weight (kg)	0.31	34.64	0.76	0.61
Dressing percentage (%)	0.33	1.80	11.20	11.80
Saleable meat percentage (%)	0.56	2.00	8.63	9.69
Fat depth at rump (mm)	0.41	1.97	2.79	1.13
Marbling score (score)	0.38	0.71	102.41	135.51
Calving ease – direct (%)	0.10	20.30	3.45	3.66
High-value primal cuts (%)	0.62	0.80	19.05	28.80

Table 2. Heritabilities (h^2) of selection criteria traits and genetic correlations with the breeding objective traits.

Traits[1]	BW	200d WT	400d WT	600d WT	CEMA	P8	RBY	CIMF
h^2	0.39	0.18	0.25	0.31	0.26	0.36	0.58	0.25
WWd	0.45	0.85	0.68	0.65	0.04	-0.11	0.06	0.00
RFI-p	-0.10	-0.30	-0.15	-0.15	-0.10	0.25	-0.07	0.20
EWT	0.40	0.65	0.80	0.65	0.00	-0.10	0.06	0.03
RFI-f	-0.10	-0.30	-0.15	-0.15	-0.10	0.25	-0.04	0.20
SW	0.49	0.62	0.72	0.88	0.36	0.05	0.03	-0.22
DP	-0.01	-0.19	-0.16	-0.18	-0.09	0.50	0.06	0.10
SMP	0.00	-0.06	-0.04	-0.04	0.34	-0.52	0.70	-0.27
FD	-0.05	0.04	0.06	0.06	0.28	0.85	-0.35	0.26
MS	0.00	-0.06	-0.14	-0.22	-0.15	0.23	-0.17	-0.14
CEd	-0.50	-0.20	-0.16	-0.17	-0.10	0.00	0.00	0.00
HVC	0.05	0.05	0.05	0.05	0.38	-0.30	0.71	-0.36

[1] Traits: selection criteria traits, BW: birth weight (kg), 200d WT: 200-day weight (kg), 400d WT: 400-day weight (kg), 600d WT: 600-day weight (kg), CEMA: carcase eye muscle area (cm^2), CFD: P8 fat depth (mm), RBY: retail beef yield (%) and CIMF: carcase IMF (%). Breeding objective traits, WWd: weaning weight- direct (kg), RFI-p: residual feed intake- pasture (kg/day), EWT: feedlot entry live weight (kg), RFI-f: residual feed intake- feedlot (kg/day), SW: sale weight (kg), DP: dressing percentage (%), SMP: saleable meat percentage (%), FD: fat depth on the rump (mm), MS: marbling score (score), CEd: calving ease-direct (%) and HVC: high-value primal cuts (%).

Table 3. Relative contribution of different breeding objective traits and predicted annual responses for different selection index strategies.[1]

Traits	Breeding objectives							
	DPS		JPS		DPS		JPS	
	Relative contribution				Response per year			
HVC	✗	✓	✗	✓	✗	✓	✗	✓
WWd (kg)	0.14	0.13	0.49	0.44	-0.18	-0.10	0.49	0.40
RFI-p (kg/day)	12.32	11.24	7.50	6.69	-0.06	-0.06	-0.01	-0.02
EWT (kg)	-	-	0.31	0.27	-	-	1.22	0.87
RFI-f (kg/day)	-	-	10.51	9.38	-	-	0.00	-0.02
SW (kg)	11.66	10.64	7.78	6.94	0.26	0.31	1.57	1.24
DP (%)	9.26	8.45	8.09	7.22	-0.01	0.00	0.11	0.10
SMP (%)	10.33	9.42	9.62	8.58	-0.06	0.16	0.26	0.36
FD (mm)	2.82	2.58	0.95	0.85	0.05	-0.02	-0.07	-0.10
MS (score)	35.72	32.58	39.18	34.96	0.05	0.03	0.01	0.01
CEd (%)	17.74	16.18	15.56	13.89	0.17	0.20	0.47	0.50
HVC (%)		8.78		10.77	-	0.10	-	0.18
Total response in BO $					7.09	8.15	8.06	12.43

[1] Traits: WWd: weaning weight- direct, RFI-p: residual feed intake- pasture, EWT: feedlot entry live weight, RFI-f: residual feed intake- feedlot, SW: sale weight, DP: dressing percentage, SMP: saleable meat percentage, FD: fat depth on the rump, MS: marbling score, CEd: calving ease-direct and HVC: high-value primal cuts. DPS: domestic production system and JPS: Japanese production system. ✗: without HVC and ✓: with HVC included in BO.

beef BO produced additional responses of 0.10 and 0.18% for DPS and JPS, respectively. Total responses improved when HVC was added to the current BO. The inclusion of HVC improved overall predicted index responses by $1.06 ($7.09 to $8.15) and $4.37 ($8.06 to $12.43) per year, respectively for DPS and JPS

and this equated to increases of 14.95 and 54.22%, respectively. Including HVC also changed the predicted responses of other BO traits favourably for both production systems. For instance, genetic gain for saleable meat percentage increased when HVC were included because of the positive genetic correlation between HVC and retail beef yield, which is one of the main profit drivers in beef terminal indexes. The response in marbling score was slightly decreased for the DPS. Another important trait in beef indexes was calving ease-direct, which also had a favourable response in the extended beef BOs. The inclusion of HVC in BOs trait could therefore be expected to be useful to improve overall selection response in both production systems in beef.

Discussion

Given current pricing systems, the findings of this study show that incorporating HVC into current breeding goals would benefit the beef industry. Selection for increased HVC resulted in increased responses for both production systems; however, in JPS, the relative importance of HVC was higher than DPS. If breeding values could be estimated by multi-trait animal models and made available, the EVs produced with the current model can be utilised to construct selection indices for beef in specific production systems. The model has a large number of traits, but it would be beneficial to include more carcase and meat quality traits. Broadening the breeding goal, particularly by including HVC, should result in more profitable animals; as a result, breeders' emphasis should include valuable primal cuts to increase the total market value of the carcase.

Acknowledgements

Data recording of primals weights was funded by NSW Department of Primary Industries and Meat and Livestock Australia (L.EQT.1627 and P.PSH.0942). Angus Australia are acknowledged for their funding of the Angus Sire Benchmarking Project from which some of the animals in this paper were derived. The first author was supported by an Australian Government RTP International Scholarship. Constructive comments by Peter Wahinya are gratefully acknowledged.

References

Archer J.A., Barwick S.A., and Graser H.-U. (2004) Aust. J. Exp. Agric. 44(5):393-404. https://doi.org/10.1071/EA02054

Barwick S.A., and Henzell A.L. (2005) Aust. J. Exp. Agric. 45(8):923-933. https://doi.org/10.1071/EA05068

Drennan M.J., McGee M., Conroy S., Keane M.G., Kenny D.A., *et al.* (2009) Animal 3(11):1613-1624. https://doi.org/10.1017/S1751731109990395

Johnston D.J., Tier B., Graser H.-U., and Girard C. (1999) Proc. of the AAABG, Armidale, Australia

Kenny D., Murphy C.P., Sleator R.D., Judge M.M., Evans R.D., *et al.* (2020) J. Anim. Sci. 98(7):1-12. https://doi.org/10.1093/jas/skaa191

Wood B.J., Archer J.A., and Van Der Werf J.H.J. (2004) Livest. Prod. Sci. 91(1-2):69-81. https://doi.org/10.1016/j.livprodsci.2004.06.009.

467. A method for partitioning trends in genetic mean and variance to understand breeding practices

T.P. Oliveira[1*], J. Obšteter[2], I. Pocrnic[1] and G. Gorjanc[1]

[1]The Roslin Institute and Royal (Dick) School of Veterinary Studies, The University of Edinburgh, Easter Bush, EH25 9RG, Edinburgh, Scotland, United Kingdom; [2]Agricultural Institute of Slovenia, Hacquetova ulica 17, 1000 Ljubljana, Slovenia; thiago.oliveira@ed.ac.uk

Abstract

Quantifying the sources of genetic change is essential for optimising breeding programmes. However, breeding programmes are often complex because many breeding groups are subject to different breeding actions. Understanding the contribution of these groups to changes in genetic mean and variance is essential to understanding genetic change in breeding programmes. Here we extend the previously developed method for analysing the contribution of groups to changes in genetic mean to analysing changes in genetic variance. We, expectedly, show that the contribution of females and males to change in genetic variance can differ and are not independent, indicating we should not look at the contributions in isolation.

Introduction

Quantifying sources of genetic change is essential for identifying key breeding actions and optimising breeding programmes. Due to the complexity of many breeding programmes, the observed genetic change is a sum of contributions from different groups of individuals (often referred to as selection pathways). These groups vary in their contribution due to different selection intensity, accuracy, genetic variation, generation interval, and dissemination. At the individual level, the breeding value of an individual can be partitioned as $a_i = \frac{1}{2}a_s + \frac{1}{2}a_m + w_i$, where a_i is the breeding values of the individual, a_s and a_m are breeding values of sire and dam, and w_i is the Mendelian sampling term. García-Cortés *et al.* (2008) provided a methodology to allocate these partitions of breeding values into group-specific contributions to the change in genetic mean. To gain insight into time-trends or other data partitions, they summarised group-specific contributions by a variable such as year, gender, population, etc. While García-Cortés *et al.* (2008) addressed the contribution of different groups to changes in genetic mean, breeding programmes should also consider the contribution of different groups to changes in genetic variance to fully understand drivers of genetic change in a population (Sorensen *et al.*, 2001; Lara *et al.*, 2021). Here we extend the method of García-Cortés *et al.* (2008) to analyse changes in genetic variance. We develop the method and demonstrate it with a simulation. We assume throughout we know the true breeding value. Future work will extend the method to estimated breeding values by applying the method on posterior samples of breeding values following the work of Sorensen *et al.* (1994; 2001) and Lara *et al.* (2021).

Materials & methods

Statistical framework. Let a be a vector of breeding values from a normal distribution with mean 0 and covariance $\mathbf{A}\sigma_a^2$. a can be written as a linear combination of breeding values from ancestors and individual's deviation from the ancestral expectation a = Tw, where T is a triangular matrix of expected gene flow between ancestors and individuals and $\mathbf{w} \sim \mathrm{N}(\mathbf{0}, \mathbf{W}\sigma_a^2)$ are Mendelian sampling terms representing deviations, with W being a diagonal matrix of variance coefficients and σ_a^2 the base population additive genetic variance (Henderson, 1976).

Assuming a factor with p groups and for any set $\sum_{j=1}^{p} \mathbf{P}_j = \mathbf{I}$, García-Cortés *et al.* (2008) partitioned the gene flow into contributions of each group by defining $\mathrm{T}_j = \mathrm{TP}_j$, and further partitioned the contribution of each group to breeding values *a priori* using the equality $\mathbf{a} = (\mathbf{T}_1 + \mathbf{T}_2 + \cdots + \mathbf{T}_p)\mathbf{w} = \mathbf{a}_1 + \mathbf{a}_2 + \cdots + \mathbf{a}_p$.

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_467, © T.P. Oliveira *et al.* 2022

They further showed that these partitions can be estimated from data collected in breeding programmes (*posteriori*) by first estimating breeding values $\hat{\boldsymbol{a}} = E(\boldsymbol{a}|\boldsymbol{y})$ from phenotype data (y) and then partitioning $\hat{\mathbf{a}} = (\mathbf{T}_1 + \mathbf{T}_2 + \cdots + \mathbf{T}_p)\hat{\mathbf{w}} = \hat{\mathbf{a}}_1 + \hat{\mathbf{a}}_2 + \cdots + \hat{\mathbf{a}}_p$. By summarising these partitions, they quantified the contribution of each group (for example, males vs females) to the time-trend in genetic mean, where time-trend represents a change in genetic mean over time. This method has been implemented in the AlphaPart R package (Obšteter *et al.*, 2021), which efficiently calculates (via T^{-1}) and summarises such partitions by one or more variables, such as generation, selection status, etc. AlphaPart enables use of any function for the summarisation, that is, $f(a_j)$.

Here we extend the partitioning method to analyse the contribution of groups to genetic variance. Variance of breeding values is, *a priori*, $Var(\boldsymbol{a}) = Var(\boldsymbol{Tw}) = \boldsymbol{TWT}^T\sigma_a^2$. Using the equality $\mathbf{a} = (\mathbf{T}_1 + \mathbf{T}_2 + \cdots + \mathbf{T}_p)\mathbf{w} = \mathbf{a}_1 + \mathbf{a}_2 + \cdots + \mathbf{a}_p$, we can partition genetic variance as:

$$Var(\boldsymbol{a}) = Var\left((\mathbf{T}_1 + \mathbf{T}_2 + \cdots + \mathbf{T}_p)\boldsymbol{w}\right) = Var(\mathbf{T}_1\boldsymbol{w}) + Var(\mathbf{T}_2\boldsymbol{w}) + \cdots + Var(\mathbf{T}_p\boldsymbol{w}) +$$
$$2Cov(\mathbf{T}_1\boldsymbol{w}, \mathbf{T}_2\boldsymbol{w}) + \cdots + 2Cov(\mathbf{T}_p\boldsymbol{w}, \mathbf{T}_{p-1}\boldsymbol{w}) = \sum_{j=1}^{p}\mathbf{T}_j\boldsymbol{W}\mathbf{T}_j^T\sigma_a^2 +$$
$$2\sum_{j=1}^{p-1}\sum_{j'=j+1}^{p}\mathbf{T}_j\boldsymbol{W}\mathbf{T}_{j'}^T\sigma_a^2 = \sum_{j=1}^{p}\sigma_{a_j}^2 + 2\sum_{j=1}^{p-1}\sum_{j'=j+1}^{p}\sigma_{a_j,a_{j'}}.$$

While this 'theoretical' partitioning involves matrix products, we can also summarise partitions $a_1 + a_2 + \ldots + a_p$ (calculated via T^{-1}) by calculating variance of each partition $f(a_j) = Var(a_j)$ and covariance of each pair of partitions $f(a_j,a_{j'}) = \text{Cov}(a_j,a_{j'})$. The partitions can be summarised in many ways, for example, to quantify the contribution of males and females to change in genetic variance over time. We implemented this method in the AlphaPart R package (Obšteter *et al.*, 2021) with the summary method: summary(part, by = 'time', FUN = var), where part is an object with group-partitioned breeding values. This partitioning is for *a priori* or true breeding values. To estimate partitions of genetic variance from data collected in a breeding programme (*posteriori*) we would have to use methods from Sorensen *et al.* (2001) and Lara *et al.* (2021). While this is straightforward, we left this to future work.

Simulation. To illustrate the method, we have simulated one replication of a breeding programme with 1000 individuals over 40 generations. For the burn-in phase, generations -20 to 0, we selected the best 5 males (out of 500) as sires based on their phenotype and mated them with all 500 females from the previous generation and all 500 females from the current generation. These matings produced 1000 selection candidates for the next generation. After the burn-in phase, we tested two selection scenarios over additional 20 generations: we selected 5 best males from 500 male candidates based on their phenotypes (scenario 'phenotype') or true breeding values (scenario 'TBV'). The simulation was done with AlphaSimR version 1.0 (Gaynor *et al.*, 2020) using cattle genome, where we assumed 30,000 quantitative trait loci (QTL) with additive effects sampled from a normal distribution, and phenotype with heritability 0.3. With this simulation, we aimed to show how trends in genetic mean and variance come about due to the selection of males. We split T by using $\mathbf{P}_m + \mathbf{P}_f = \mathbf{I}$ and $\mathbf{P}_m^s + \mathbf{P}_m^n = \mathbf{P}_m$, where $\mathbf{P}_m$ is a diagonal matrix with ones in rows for males and zeros otherwise; $\mathbf{P}_f = \mathbf{I} - \mathbf{P}_m$; $\mathbf{P}_m^s$ is a diagonal matrix with ones in rows for selected males, and $\mathbf{P}_m^n = \mathbf{P}_m - \mathbf{P}_m^s$ is a diagonal matrix with ones in rows for non-selected males.

Results

Figure 1A shows time-trends in the genetic mean and variance and its partitioning by males and females, while Figure 1B represents trends partitions by selected males, non-selected males, and females. We scaled the genetic mean and variance of the base population respectively to zero and one to facilitate interpretation.

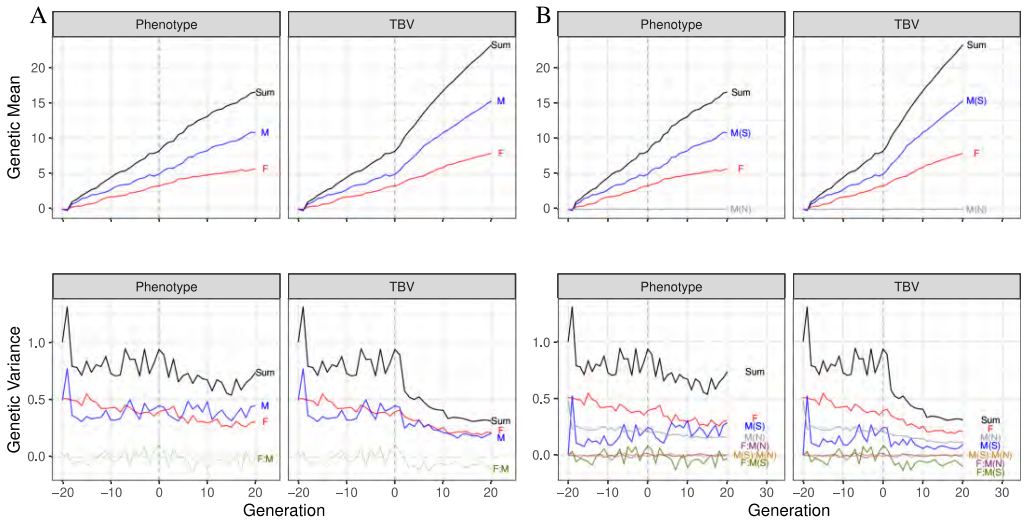


Figure 1. Partitioning of the total genetic mean and variance (Sum) over generation by (A) gender (male (M) and female (F)) and (B) gender and status (selected males (S) and non-selected males (N)). Each sub-figure shows the scenario with selection on phenotype (Phenotype, left) and selection on true breeding value (TBV, right). Covariance between two supposed levels A and B is represented using 'A:B'.

Figure 1A shows that selection of males contributed almost twice as much as females, which contributed only via dissemination of past male selection. However, if only sex is used to investigate the contribution of the males and females to the genetic variance, we can see that the contributions of males and females are very similar for both selection scenarios. In fact, males seem to be contributing a bit more to genetic variance than females in the phenotypic selection scenario. This observation raises a question, how can males contribute similar proportions to the genetic variance as females if we are selecting among males? The answer to this question is demonstrated in Figure 1B. Although non-selected sires do not contribute to the change in the genetic mean, they still contribute to the genetic variance. Selection on true breeding values clearly showed that the contribution of groups to genetic variance is a function of the intensity of selection. Selected males contributed least to genetic variance, followed by non-selected males and females. Partitioning in the scenarios with phenotype selection showed a similar result but selecting animals on their phenotype showed that the true partition for selected males may have a higher contribution to genetic variance than non-selected males.

Splitting the male contribution into selected and non-selected also shows that the negative covariance between male and female contributions in Figure 1A is driven by the covariance between female and selected male contributions (F:M(S)). The latter was consistently negative from generation 2 to 20 in the scenario with selection on true breeding values, resulting in a mean correlation of -0.52 (±0.12). This negative correlation means that the contribution of females and selected males to the genetic variance are not independent. That is, the total genetic variance is smaller than the sum of individual contributions.

Discussion

We presented a simple method for partitioning the trends in genetic mean and variance into contributions from different groups as an extension of the work of García-Cortés *et al.* (2008). We implement this method in the AlphaPart R package available from CRAN. This method can be a powerful and valuable

tool to understand how different breeding groups contribute to genetic mean and variance change. We demonstrated that the choice of groups is essential and that contributions are not independent. Hence, they should not be analysed in isolation from each other. Future work will apply the method to estimate trends in the genetic mean and variance following the work of Sorensen *et al.* (2001) and Lara *et al.* (2021).

References

García-Cortés, L.A., Martínez-Ávila, J.C., Toro, M.A. (2008) Partition of the genetic trend to validate multiple selection decisions. Animal 2:821–824.

Gaynor, R.C., Gorjanc, G., Hickey, J.M. (2020) AlphaSimR: An R-package for Breeding Program Simulations. bioRxiv 21. https://doi.org/10.1101/2020.08.10.245167

Gorjanc, G., Obšteter, J., Oliveira, T.P. (2021) AlphaPart: Partition/Decomposition of Breeding Values by Paths of Information. https://CRAN.R-project.org/package=AlphaPart

Henderson, C.R. (1976) A simple method for computing the inverse of a numerator relationship matrix used in prediction of breeding values. Biometrics 32:69–83.

Lara, L.A.C., Pocrnic, I., Oliveira, T. P., Gaynor, C., Gorjanc, G. (2021) Temporal and genomic analysis of additive genetic variance in breeding programmes. Heredity. https://doi.org/10.1038/s41437-021-00485-y

Obšteter, J., Holl, J., Hickey, J.M., Gorjanc, G. (2021) AlphaPart – R implementation of the method for partitioning genetic trends. Genet. Sel. Evol. 53(30). https://doi.org/10.1186/s12711-021-00600-x

Sorensen, D., Wang, C., Jensen, J., Gianola, D. (1994) Bayesian analysis of genetic change due to selection using Gibbs sampling. Genet. Sel. Evol. 26(333). https://doi.org/10.1186/1297-9686-26-4-333

Sorensen, D., Fernando, R., Gianola, D. (2001) Inferring the trajectory of genetic variance in the course of artificial selection. Genet. Res. 77:83–94. https://doi.org/10.1017/S0016672300004845

468. A practical approach to predicting expected response to selection with a new index

X. Zhang*, P.R. Amer and C. Quinton

AbacusBio Ltd., 442 Moray Pl, Dunedin 9016, New Zealand; lzhang@abacusbio.co.nz

Abstract

Selection indexes need to be periodically evaluated and re-designed to adapt to new market trends and new included traits with economic weightings. Response to selection is one of the statistics used to predict the genetic trend, and to compare old and new indexes. The most popular method for response to selection requires knowledge of the selection intensity and generation intervals, which are hard to obtain. However, the ratio of these often unknow parameters can be assumed to remain constant and can be estimated by observation in an established breeding program. Therefore, we proposed a method to estimate the response to selection using a baseline selection index. The results show how new selection indexes can be characterised for estimated response to selection in trait units and economic units while avoiding making assumptions about breeding scheme design and structure. The new method is practical and easy to calculate, making index comparison straight-forward.

Introduction

Selection index redesign and updates occurs regularly across national breeding organisations as well as in commercial breeding companies. When changing the indexes, we generally evaluate a few statistical comparisons between old and new indexes with the help of a few graphics, such as: (1) degree of re-ranking among individuals displayed by scatterplots, index correlations, and percentage of re-ranking among the top-ranking individuals; (2) breeding objective trait percent emphasis pie charts or bar charts; (3) expected responses to selection in traits which are genetically correlated with the indexes, including non-index traits that have high correlations with the indexes. Among the three statistics, expected responses to selection are typically of most interested to animal geneticists. The objective of this study is to demonstrate a method to calculate the expected response to selection using known industry information without requiring detailed assumptions about the genetic improvement programs that drive genetic progress.

Materials & methods

Method. Response to selection (R) or change in true aggregated merit (ΔI) for an index is commonly calculated using a simple breeders' equation, such as that defined by Rendel and Robertson (1950) for example as:

$$R = \Delta I = \frac{i}{L_m + L_f} \cdot r_{TI} \cdot \sigma_T \quad (1)$$

where i is the average selection intensity of sire and dam, L_m and L_f are the generation intervals of males and females, r_{TI} is the correlation between g, the BV of a trait, and index I, and σ_T is the genetic SD of the trait.

For our example breeders' equation, we define $K = \frac{i}{L_m + L_f}$ as the effort of selection. A similar definition can typically be obtained for many other variations of the breeders' equation. Our method is underpinned by the assumption that selection effort K remains constant, even though the selection index is being changed. In the definition of selection index theory $r_{TI} = \frac{\sigma_I}{\sigma_T}$ (Gjedren, 1976). Combining this property of selection indexes with Equation 1 then $R = K \cdot \sigma_I$. In a situation where we have previously observed a change in a selection index in a particular period, we define this index as the baseline index and its change as R_{base}, i.e. $R_{base} = K_{base} \cdot \sigma_{Ibase}$, and

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_468, © X. Zhang *et al.* 2022

$$K_{base} = \frac{R_{base}}{\sigma_{I_{base}}} \quad (2)$$

We can use K_{base} as a baseline effort of selection to compare to a new index. With the same effort of selection over the same period of time, selection on a new index (I_{new}) will give response R_{new} in the new index as follows:

$$R_{new} = K_{base} \cdot \sigma_{I_{new}} = R_{base} \cdot \frac{\sigma_{I_{new}}}{\sigma_{I_{base}}} \quad (3)$$

And R_{new} is our interest. The response for any genetic trait of interest denoted g in its trait units when achieving R_{new} response in the new index is

$$R_{TU} = R_{new} \cdot b_{g,I_{new}}$$

where R_{TU} is the response in trait units, and $b_{g,I_{new}}$ is the regression coefficient of trait of interest g on the new selection index. Combining

$$b_{g,I_{new}} = \frac{\sigma_{g,I_{new}}}{\sigma^2_{I_{new}}} = \frac{r_{g,I_{new}} \cdot \sigma_g \cdot \sigma_{I_{new}}}{\sigma^2_{I_{new}}} = \frac{r_{g,I_{new}} \cdot \sigma_g}{\sigma_{I_{new}}} \quad (4)$$

where $r_{g,I_{new}}$ is the correlation between g and the new index. With Equation 3, the trait response in g when selecting on the new index is

$$R_{TU} = R_{new} \cdot \frac{r_{g,I_{new}} \cdot \sigma_g}{\sigma_{I_{new}}} = R_{base} \cdot \frac{\sigma_{I_{new}}}{\sigma_{I_{base}}} \cdot \frac{r_{g,I_{new}} \cdot \sigma_g}{\sigma_{I_{new}}} = R_{base} \cdot r_{g,I_{new}} \cdot \frac{\sigma_g}{\sigma_{I_{base}}} \quad (5)$$

The response in economic unit e.g. dollar will be

$$R_{EU} = R_{TU} \cdot a \quad (6)$$

where a is the economic weight of g. And the standardised response (i.e. expressed in SD of the trait of interest) will be

$$R_{STD} = \frac{R_{TU}}{\sigma_g} = r_{g,I_{new}} \cdot \frac{R_{base}}{\sigma_{I_{base}}} = r_{g,I_{new}} \cdot K_{base} \quad (7)$$

In the situation where $R_{base} = \sigma_{I_{base}}$ (i.e. we define the progress in the base index so as to quantify selection effort in index SD units) then K_{base} = 1 and the standardised response in g will be $R_{STD} = r_{g,I_{new}}$.

Materials. The economic weights of 13 traits in 2020 Australian dairy cow Balanced Performance Index (BPI) and health weighted index (HWI) were used here to demonstrate the response in trait unit (R_{TU}) and response in economic unit (R_{EU}) (Axford *et al.*, 2021; DataGene, 2020). The genomic Australian breeding value (ABV) predictions of 7,338 Holstein-Friesian bulls were used, excluding bulls with daughter fertility (FERT) accuracy less than 0.6, and bulls with missing ABV values. The R_{base} for the index was set to \$30 as observed annual response to selection (DataGene, 2021). An index testing app that has been developed by AbacusBio (Dunedin, New Zealand) to deploy the above methodology was used to calculate the two sets of response values and generate reports.

Results

Table 1 shows the R_{TU} and R_{EU} of 14 breeding objective traits and 3 monitor traits in BPI and HWI. Monitor traits had \$0 in economic weight therefore their R_{EU} were also 0. For example, the response to selection on

Table 1. Expected annual response to trait unit (R_{TU}) and to economic unit (R_{EU}) in Australian Balanced Performance Index and Health Weighted Index in year 2020.

			BPI		HWI	
Trait	**Full name**	**Trait unit**	R_{TU}	R_{EU}	R_{TU}	R_{EU}
PROT	Milk protein	kg	0.667	4.510	0.057	0.249
FAT	Milk fat	kg	1.695	3.525	0.510	0.689
MILK	Milk volume	l	3.005	-0.331	-2.634	0.184
SURV	Survival	% survived from one parity to the next	0.783	5.634	0.643	4.631
RESSURV	Residual survival	% deviation[1]	-0.014	0	-0.035	0
FERT	Daughter fertility	42d calving%	1.070	7.429	1.164	16.424
SCC	SCC	count/ml	4.573	3.155	3.721	2.567
MAS	Mastitis resistance	resistance ABV unit	0.515	3.476	0.401	2.707
MSPEED	Milking speed	ABV unit	0.086	0.430	0.043	0.214
TEMP	Temperament	ABV unit	0.133	0.478	0.057	0.205
MAMM	Mammary system	ABV unit	0.356	0.981	0.197	0.706
UDDEP	Udder depth	ABV unit	0.700	0.574	0.611	0
OTYPE	Overall type	% increase in score	0.181	0.246	0.043	0.059
PINSET	Pin set	% increase in score	-0.190	-0.149	-0.139	-0.109
FOREA	Fore-udder attachment	ABV unit	0.327	0	0.266	0
FEEDEF	Feed saved	kg DM saved	0.214	0.041	3.824	1.473
HEAT_T	Heat tolerance	production decline rate[2]	-0.101	0	0.158	0

[1] DataGene., 2016.

[2] Cheruiyot *et al.*, 2021.

BPI for the trait FERT in a year was estimated as an increase of 1.07% in 42-day calving rate, equivalent to $7.43 AUD increase in economic value.

Discussion

The most popular way of calculating response to selection requires the knowledge of selection intensity and generation interval of male and female selection candidates. The standardised response to selection is often used in selection index to predict the annual genetic progress. In this study, we demonstrated a method to calculate annual expected response without the knowledge of selection intensity and generation intervals, assuming their ratio remains constant in an established breeding programme. Then we used the observed baseline response to selection in the estimation of the new index. Our method is simpler and likely more accurate than the traditional method, because it can directly estimate the observed intensity to generation interval ratio based on the actual genetic trend under selection (Equation 2) and can avoid making assumptions about breeding scheme design. The forms of response to selection in trait units and in economic units rather than standardised response are more straight forward for practical use, easing the stakeholders' decision-making process.

When evaluating a national index where many different breeding programs operated by different companies have all contributed to genetic gains, our method eliminates the need for assumptions about each specific breeding structure and ultimate aggregation.

A limitation of our approach is that is we rely on the correlations among trait EBVs and indexes to estimate the response to selection, and this can change with different sets of candidates. These correlations are biased by prior selection of candidates, and this could be a problem when any new index being evaluated weights

traits radically different to the old reference index. However, in most instances, the impact of preselection of candidates based on the old index is consistent across the new and old indexes. This effect gets captured in K, and the regressions $b_{g,I_{new}}$ which are intrinsic to the method should be reasonably robust to prior selection affecting the variance of I_{new} which is the independent variable in the regression. In addition, using sets of estimated breeding values where there are mixtures of very old and much more recently born animals can introduce false correlation into the calculations due to genetic trends being greater in some traits than in others. Therefore, the choice of data input is important when evaluating indexes, although our experience across many different species and industries is that inferences about the future effects of new selection indexes are typically robust to whichever set of candidates estimated breeding values and indexes is chosen to deploy the method.

References

Axford M., *et al.* (2021) Impact of a multiple-test strategy on breeding index development for the Australian dairy industry. Animal Production Science 61(18):1940-1950. https://doi.org/10.1071/AN21058

Cheruiyot E.K., *et al.* (2021) Using selected sequence variants to improve genomic prediction of heat tolerance in dairy cattle. Proc. of the 24th AAABG, online and Australia/New Zealand

DataGene. Survival ABV and Residual Survival ABV Technote 4 2016. Available at https://datagene.com.au/sites/default/files/Upload%20Files/Technote%204%20Survival%20ABV%20and%20Residual%20Survival%20ABV_0.pdf

DataGene. National Breeding Objective 2020 Options paper V2. Available at: https://datagene.com.au/sites/default/files/Upload%20Files/NBO%202020%20Options%20Paper%20May%202020%20GESC.pdf

DataGene. Annual update 2020/21. Available at https://datagene.com.au/sites/default/files/inline-files/Datagene%20Annual%20Update%202021%20-%20FINAL.pdf

Gjedrem T. (1976) Selection Indexes compared with single trait selection. Acta Agri. Scandinavica 17:4, 263-268. https://doi.org/10.1080/00015126709433162

Rendel J.M. and Robertson A. (1950) Estimation of genetic gain in milk yield by selection in a closed herd of dairy cattle. J Genet. 50(1):1-8

469. Theory and tools that are needed for the design and optimization of breeding programs

P.R. Amer

AbacusBio Limited P.O. Box 5585, Dunedin, New Zealand; pamer@abacusbio.co.nz

Abstract

Optimisation of genetic progress is of great interest and relevance to nearly all stakeholders in livestock genetic improvement and the wider supply chains that are impacted. However, many of the methods and tools developed and used by animal breeders assume that optimisation of genetic progress is the sole goal and focus of these stakeholders. In reality, the investor stakeholders in genetic progress are typically more strongly influenced by simple concepts and messages relating to things they can change, rather than complex black box representations of what a perfect breeding program would look like. Further, market share and price of germplasm sold relative to the costs of achieving genetic gains are important consideration. This paper first summarises the key issues facing many key investor stakeholders, and then describes at a high level the tools available to design and optimize breeding programs, and some potential gaps.

Introduction

The magnitude of investment in livestock genetic improvement globally must be at least hundreds of millions of dollars/euros, but the benefits in terms of increased farmer competitiveness, higher quality products, and lower costs of production which often lead also to lower consumer prices will be manyfold greater. Decision makers investing in livestock genetic improvement are predominantly interested in what they can do 'at the margin'. How much more benefit will they get relative to their change in total investment, or perhaps redirection of the investment they are making. Quite often the net benefits of marginal changes in investment in breeding program design are quite modest. From time to time as analysts, we might get the opportunity to design a new breeding program from scratch, but this is often a rarity. In this instance, the process can be more exciting, because of the broad nature of the decisions to be made, and quite a different approach is required. This paper presents the views of the author in terms of challenges and solutions associated with design and optimisation of breeding programs. This is then used as a baseline for commenting on the tools available, and what new tools might be most useful in the future.

Methods

Issues and challenges existing programs. The key decisions faced by investors in existing genetic improvement programs include:

- How much to invest in genetic improvement of new traits where there is currently direct financial incentive along the supply chain for their improvement?
- How much to invest in genetic improvement of new traits where there is currently no direct financial incentive along the supply chain for their improvement, but future incentives are anticipated?
- How much to invest in genomic reference population, how big they should be, and what phenotypes to collect, type of sample, and type of genotyping platform to use?
- How many and which selection candidates to genotype using what genotyping platform?
- What reproductive technologies are available and could be used to introduce new genetic variation, increase the pool of elite candidates and intensity of selection, multiply elite genetics, minimize risk of unfavourable mendelian sample in a single offspring from a mating pair?
- How to change the relative weighting placed on traits being selected now to maximise future industry benefits and/or future market share of gene-stock sales?

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_469, © P.R. Amer 2022

- How to select and mate individuals to optimise the trade-off between short term genetic gain, inbred offspring generation, and the long-term genetic diversity of the population?
- What opportunities are there to source elite genetics and/or introduce new genetic variation and genetic diversity from other countries, populations and breeds?

Issues and challenges new programs. The key issues faced by investors in a new genetic improvement program include;

- What foundation germplasm should I gather from which to build a new and genetically improved population?
- What traits are potentially easy to record and for which genetic gains will sufficiently catch the attention of the target market for improved germplasm to generate sales?
- What traits to record on what selection candidates and their relatives?
- Is a pure genomics approach likely to be superior than a more traditional design relying on information from individuals or their close relatives such as progeny testing?
- How long will it take before my investment starting today will yield a return that will convince key stakeholders that the investment is justified and worth continuing?

Existing tools and enhancement opportunities. A substantial proportion of animal breeding research addresses the need for fast, efficient and robust analysis of phenotypic, pedigree, and genomic data to help identify candidates who are superior for existing traits. In many instances, affordable canned software is available to complete basic analyses. More complex software systems can also be licensed at a much lower cost than individual redevelopment from scratch. Despite this, maintenance of entrenched legacy systems and a desire to completely control one's own destiny mean that custom systems are still common. Developers of these tools could focus more on issues influencing bias and inflation (over dispersed predictions) of estimated breeding values, as opposed to often focusing solely on prediction accuracy. In particular, when there are differences in the degree of over dispersion and/or biases for selection candidates arising from different breeding strategies (e.g. progeny test versus genomic selection), incorrect conclusions about the optimal breeding strategy can arise. These problems are exacerbated when industry loses confidence, for example when young bulls with high ranking from genomics typically drop once they have recorded daughters in the evaluation.

Beyond genomic prediction, design of breeding structures incorporating genomics often seems to be based on winning a large research budget and maximising animals genotyped. Sometimes an excessively high genotyping density is pursued to enable discovery of new variants that to date have made only a disappointing contribution to the rate of genetic gain. In some cases, commercial control of a large genotyping budget can be used as a barrier to market entry for potential new competitors but who cannot match the genotyping budget of the dominant market leaders. Equations which link accuracy of genomic prediction to training population size, prediction reliability of information contributing to training individual merit predictions, and effective population size, have proven a blunt, but useful tool when addressing genomic training population design issues. Unfortunately, the rather esoteric use of effective population size in the prediction formulae initially obscured the key issue of the importance of the degree of relatedness (number of meiosis) between selection candidates, and training population individuals. As more and more attempts are made to improve genomic prediction accuracy via causal variant discover, it may in the future be beneficial to account for the impact of genetic architecture of a trait on accuracy of genomic prediction from a specified training population.

A well-researched area relates to the trade-off between short-term genetic gain and long-term maintenance of genetic variation. Very powerful and useful tools have arisen from this research although their impact is

still limited in many cases. When a single decision-making entity is not responsible for all of the selections relevant for a single population, commercial interests are pushed by the market towards maximising their own short term competitive position. This can come at the expense of genetic diversity. In other situations, the unpredictable nature of the ability of selection candidates to contribute to the next generation becomes a problem. For example, if an optimal contributions selection algorithm identifies a single candidate for a relatively high contribution, but then that candidate fails on some physical or genetic criteria at the last minute prior to mating, the entire analysis may need to be reworked in a very short time frame. Systems that can identify interchangeable contributors to the next generation and/or optimise the contributions of families are important ways of overcoming these practical limitations. Provided all candidates have been genotyped, replacement of pedigree relationships with genomic relationships will be straight forward in these approaches. This will also open the way for sophisticated management of genetic disorders where frequency has increased sufficiently to be of concern to farmers.

In theory, accurate quantification of absolute economic values should be a key ingredient at the outset of designing a new breeding program and when considering reinvestment or redesign of an existing breeding program. In practice, this is seldom the case. Just as one does not usually learn to ride a bicycle by spending a large amount of time pointing the bicycle in some desired long-term direction of travel, the tendency is for breeding programs to initially follow a wobbly trajectory in terms of trait priorities. The direction of travel can be subsequently refined once the breeding program is up and running. Some basic knowledge of the value of genetic progress that is currently, and could potentially, be achieved is still critical to breeding program design. While in some cases, the value of genetic gain is considered from the perspective of the farmer end users of the genetic gain achieved, for competitive breeding companies, value of increased market share, and/or price received for germplasm units are also important. In many situations, the relationship between genetic merit and price received is exponential in nature, with elite offerings achieving very high premiums, while intermediate merit offerings realising much more modest prices. A market research philosophy is often helpful when considering breeding objectives to help drive adoption and market share gains. An accurate understanding of value is particularly important when considering the value from adding new traits into a new or existing breeding program.

Stochastic simulation tools have played a valuable role in uncovering key design nuances. However, they can be cumbersome to build and operate, and commonly fail to accommodate commercial and practical realities that constrain the design options in practice. Maximum benefits can be obtained from these approaches if great attention is paid to understanding the existing breeding program and industry structures, and the inherent constraints within them. Because decision makers are faced with simultaneously addressing many strategic and logistical decisions, results and interpretation of the simulations should be broken into simple step changes to current structures with a clear and compelling description of the benefits that should arise from each change. Because of this, the simulations should simplify as much as possible the aspects of industry structure and breeding scheme design that do not require a complex model representation. A fast turnaround on results, and ease of design modification are critical to facilitate an interactive process with key decision makers and industry representatives. In this way, the industry can build ownership in the results, and future changes in scheme design are more likely to be forthcoming.

Equations that express the annual rate of genetic progress in a trait or index as a function of selection accuracy, genetic standard deviation, and generation interval (sometimes referred to as the 'Breeders Equation'), is a classic and helpful tool for communicating the basic principles of breeding program design. However, there are many limitations to this approach. Firstly, the equations assume static rates of genetic gain, and fail to account for the transition from the status quo to the time any new breeding approach has an impact. This is particularly important when comparing a progeny testing strategy with a genomic

selection strategy. Genomic selection strategies potentially allow superior animals to be marketed within a much shorter time frame (basically at reproductive age), rather than waiting for performance records to be available on progeny, for example with a progeny testing strategy. The delays between selections in higher tiers of multi-tier pyramid breeding structures and commercial impacts can be very large, although the multiplication of benefits across a larger industry share mean that long term sustainability and competitiveness of the breeding program is improved.

Finally, selection index theory seems to be becoming something of a lost art in animal breeding. While the principles of selection index theory are fully embedded in mixed model equation methodology applied in animal breeding, genetic evaluation systems provide little insight into the design of breeding programs. The concepts of having contributing information sources from selection candidates along with their close relatives and recorded animals in genomic training populations to the accuracy of genetic evaluation are key to understanding and optimising breeding scheme design. Several existing tools are available, but there is an opportunity to generate shared libraries of source code that would allow a more flexible and modular approach to their application.

Discussion

Breeding schemes will deliver many billions of benefits to livestock industries around the world with the costs to run these schemes a small fraction of the benefits. Research into breeding scheme design has been instrumental in some step changes in breeding program approaches and has also generated a lot of useful tools to help with further refinements to breeding scheme designs. Stronger dialogue among industry, academia and government is required to make sure that future refinement of the tool set outlined in this paper is fit for purpose going forward. A better understanding of capital budgeting from an investment perspective, and cost benefit analysis, would be very valuable. Perhaps a new on-line forum that would allow new ideas, and past experiences to be shared in an open and constructive way. Internships that facilitate placement of young professionals in a range of different industry settings should continue to be encouraged to strengthen these important links.

470. General properties of the optimum selection index

R. Wellmann

University of Hohenheim, Department of Animal Genetics and Breeding, Garbenstraße 17, 70599 Stuttgart, Germany; r.wellmann@uni-hohenheim.de

Abstract

Selecting animals for breeding based on their selection index should maximize the genetic merit of a population in future generations. Selection index theory assumes traditionally that the vector with breeding values of an animal is a sufficient statistic from which the optimum selection index can be calculated. It is, however, intuitively obvious that selecting outcross animals for breeding can increase the genetic variance in future generations, and can thus accelerate future genetic progress. This calls the traditional assumption into question. It is shown that the optimum selection index for the additive model depends not only on the breeding values, but also on the allele contents of major QTLs, and on the kinship of the individual with the population.

Introduction

Selection index theory was founded by F.H. Smith, and L.N. Hazel in the 1940s. It is the standard tool used in livestock breeding to combine information on different traits into a single value. Selecting animals for breeding based on their index should maximize the genetic merit of the population in future generations. Traditional selection index theory is based on the assumption that genetic correlations and genetic variances do not change over time. Under this assumption, the optimum method for constructing selection indices generalizes the approach of Moav and Hill (1966). The authors assume that the total merit of an individual i is obtained by a non-linear function $\mathrm{TMI}(\mathrm{Ph}_i)$ from the K-vector Ph_i of phenotypes. The total merit function is integrated over the distribution of phenotypes, and is then considered a function $\overline{\mathrm{TMI}}(\mu)$ of the K-vector μ with trait means. For a planning horizon of n generations, the response area is defined as the set of all population means μ_n that could be achieved within n generations of selection. The response area is an ellipsoid with center equal to the current population mean μ_0. For a given breeding policy π, the population mean μ_n^π after n generations of selection belongs to the response area. The optimum breeding policy $\dot{\pi}$ is defined by the straight line from the current population mean μ_0 to the local maximum $\mu_n^{\dot{\pi}}$ of the total merit function. The method of Moav and Hill (1966) thus provides the solution to the problem of finding the optimum breeding policy π that maximizes the objective function $f(\pi) = \overline{\mathrm{TMI}}(\mu_n^\pi)$.

Dekkers *et al.* (1995) generalized the method to the more relevant case that future genetic gains are discounted. This leads to more emphasis being placed on the short-term than on the long-term response. In this case, the objective function has the general form $f(\pi) = \sum_n \zeta_n \overline{\mathrm{TMI}}(\mu_n^\pi)$, where μ_n^π is the expected population mean after n generations of selection under breeding policy π and ζ_n is the discount factor for generation n. The index weights for the optimum breeding policy depend on the population mean in the respective generation.

To our knowledge, all authors assume implicitly that the vectors with breeding values of the selection candidates are sufficient statistics from which the optimum selection index can be calculated. This paper summarizes recent work of the author that shows that this is not the case, and provides an overview on the general form of the optimum selection index for the additive genetic model.

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_470, © R. Wellmann 2022

Theory

The optimization problem. Let TMI(Phi) denote the total merit of an animal i with phenotype vector Ph_i. The objective of the breeding program is assumed to maximize the total merit of the breed in future generations. The expected total merit of the breed in n generations is:

$$\overline{\mathrm{TMI}}_n(\pi) = E_\pi\big(\mathrm{TMI}(\mathrm{Ph}_n)\big)$$

where Ph_n is the K-vector with phenotypic values of an individual that is randomly chosen from generation n, and π is the breeding policy followed for determining the matings in each generation based on the information that is available on the selection candidates. A general form of the objective function is:

$$f(\pi) = \sum_n \zeta_n \overline{\mathrm{TMI}}_n(\pi)$$

where ζ_n discounts future genetic gains. The optimum breeding policy $\dot{\pi}$ maximizes the objective function on a set Π of permissible breeding policies.

A decomposition of the objective function. Let the distribution of the phenotype vectors be indexed by a parameter $\xi \in \Xi$. That is, f_ξ is the density of the phenotype vector Ph_i for animals that are randomly chosen from the population. In the case of a multivariate normal distribution, we have $\xi = (\mu, \Sigma_P)$, where μ is the K-vector with trait means, and S_p is the $K \times K$ phenotypic covariance matrix.

The change of population parameters from one generation to the next is assumed to be deterministic. Let Θ be a set of population parameters such that first, $\xi = \xi(\theta)$ can be estimated from θ, and second, a mapping $\Gamma_\pi: \Theta \to \Theta$ exists that describes the change of the population from one generation to the next under a given breeding policy π. The breeding policy also determines the mating list $L \in \mathcal{L}$ of the selection candidates from the current generation based on the information that is available on the selection candidates. Let the population parameter $\theta_1(\mathrm{L})$ in Generation 1 be defined as a function of the mating list L. The expected population parameter in generation n≥1 equals

$$\theta_n^\pi(L) = \Gamma_\pi^{n-1}\big(\theta_1(L)\big)$$

and the expected total merit of the population in n generations is

$$\overline{\mathrm{TMI}}_n(\pi) = \overline{\mathrm{TMI}}(\xi_n^\pi) = \int \mathrm{TMI}(x) f_{\xi_n^\pi}(x)\, d\, x,$$

where $\xi_n^\pi = \xi\big(\theta_n^\pi(L)\big)$ extracts the parameter of the phenotypic distribution from $\theta_n^\pi(L)$.

An algorithm for finding the optimum breeding policy. The breeding policy affects the objective function in two ways. It determines the mating list of the selection candidates from the current generation, and it determines the change of population parameters from one generation to the next. Solving for the optimum breeding policy could be done with an iterative approach that assumes that the breeding policy used in later generations could be different from the breeding policy used in the current generation. This allows us to consider the mapping Γ_π to be known. The objective function then depends only on the mating list L_π of the selection candidates, so we can write $f(\pi) = f(L_\pi)$. The algorithm is initialized with some prior guess π_0 about the optimum breeding policy. In each iteration j, the breeding policy π_j for a certain generation could be determined that maximizes the objective function under the assumption that breeding policy π_{j-1} will be used thereafter.

The problem of finding the optimum breeding policy could be simplified by restricting the set Π of breeding policies to policies of a certain type. In the following, we assume that all permissible breeding policies implement truncation selection and random mating. The assumption of random mating allows us to compute the mating list from the vector c with genetic contributions of the selection candidates, so we can write $\theta_n^\pi(L) = \theta_n^\pi(c)$, and $f(L) = f(c)$. The index on which selection is based is given by the gradient of the objective function, so an animal is selected for breeding, if the selection index $I_i = \frac{\partial f}{\partial c_i}(\tilde{c})$ is larger than a certain threshold value, where vector $\tilde{c}$ contains some default contributions of the selection candidates.

Assumptions. We assume normally distributed phenotypes, an additive genetic model, and a parameter space Θ that contains all tuples $\theta_n^\pi = (\mu_n, p_n, \bar{f}_n)$, where μ_n is the vector with trait means, p_n is the Q-vector with the frequencies of known QTLs, and $\bar{f}_n$ is the average kinship of the population in generation n. Moreover, we assume a total merit function of the form:

$$\mathrm{TMI}(\mathrm{Ph}_i) = \tau_{\max} - \sum_{k=1}^{K} \omega_k |\mathrm{Ph}_{ik} - \mathrm{Opt}_k|$$

where Ph_i is the K-vector with phenotypes of animal i, τ_{max} is the maximum achievable total merit, Opt_k is the optimum for trait k, and ωk is the weight of trait k in the total merit index.

The selection index. The phenotypic variance of trait k in Generation n is:

$$\sigma_{Pnk}^2 = \mathrm{V}_{Ek} + (1 - \bar{f}_n)\mathrm{V}_{Ak} + \sum_{q=1}^{Q} 2p_{nq}(1 - p_{nq})a_{qk}^2$$

where V_{Ek} is the environmental variance, V_{Ak} is the polygenic additive variance of trait k in a non-inbred population, p_{nq} is the frequency of the alternative allele of QTL q in generation n, and a_{qk} is the additive effect of QTL q on trait k. We can write p_{nq}, μ_{nk} $\bar{f}_n$, and σ_{Pnk}^2 as functions of the vector c with genetic contributions of the selection candidates. It is shown in Wellmann (2022) that the optimum selection index I_i of an animal depends on the changes of these parameters in subsequent generations that would be induced by using that animal for breeding. That is, the selection index is a linear function of:

$$I_\mu^{nki} = \chi_\mu^{nk} \frac{\partial \mu_{nk}}{\partial c_i}(\tilde{c}), \quad I_q^{nki} = \chi_\sigma^{nk} \frac{\partial p_{nq}}{\partial c_i}(\tilde{c}), \quad \text{and} \quad I_f^{nki} = \chi_\sigma^{nk} \frac{\partial \bar{f}_n}{\partial c_i}(\tilde{c})$$

The weight χ_μ^{nk} placed on improving the mean of trait k in generation n is:

$$\chi_\mu^{nk} \propto 1 - 2\Phi(\Delta\sigma_{Pnk})$$

where Φ is the distribution function of the standard normal distribution, and $\Delta\sigma_{Pnk}$ is the number of phenotypic standard deviations, trait k is in generation n above the optimum. The weight χ_μ^{nk} approaches ±1 for large, and 0 for small values of $|\Delta\sigma_{Pnk}|$. That is, traits under directional selection are implicitly given large weights, while traits under stabilizing selection are given small weights.

The weight χ_σ^{nk} placed on adjusting the phenotypic variance of trait k in generation n is:

$$\chi_\sigma^{nk} \propto \sqrt{\frac{2}{\pi}} \exp\left(-\frac{(\Delta\sigma_{Pnk})^2}{2}\right)$$

The weight χ_σ^{nk} approaches 0 for large values, and ≈0.8 for small values of $|\Delta\sigma_{Pnk}|$. That is, for a trait under stabilizing selection, the selection index gives implicitly a high priority on reducing the phenotypic variance of that trait. This is because traits with a small phenotypic variance tend to deviate less from the optimum.

The envisaged reduction of the phenotypic variance could be achieved by driving QTL alleles to fixation or by inbreeding, whereby the latter would have unwanted side effects. We have:

$$\frac{\partial \mu_{1k}}{\partial c_i}(\tilde{c}) = \mathrm{EBV}_{ik},$$
$$\frac{\partial p_{1q}}{\partial c_i}(\tilde{c}) = \frac{X_{iq}}{2},$$
$$\frac{\partial \bar{f}_1}{\partial c_i}(\tilde{c}) = 2\bar{f}_i,$$

where EBV_{ik} is the breeding value of selection candidate i for trait k, X_{iq} is the allele content of the individual for QTL q, and $\bar{f}_i$ is the kinship of the individual with the population. The derivatives $\frac{\partial \mu_{nk}}{\partial c_i}(\tilde{c})$, $\frac{\partial p_{nq}}{\partial c_i}(\tilde{c})$ and $\frac{\partial \bar{f}_n}{\partial c_i}(\tilde{c})$ for $n>1$ are linear functions of these values. In particular, $\frac{\partial \mu_{nk}}{\partial c_i}(\tilde{c})$ depends on the kinship $\bar{f}_i$ of the selection candidate with the population. Hence, if selection candidate i has a small kinship with the population, then the use of that animal for breeding can have a delayed positive effect on the genetic gain in later generations. This delayed positive effect results from a better conservation of the genetic variance.

Discussion

The formulas presented in this paper are examined in greater detail in Wellmann (2022). They show that the optimum selection index for the additive model depends not only on the breeding values, but also on the allele contents of major QTLs, and on the kinship of the individual with the population. The index weights on the allele frequencies have the purpose to drive alleles to fixation when the trait mean approaches the optimum. This reduces the phenotypic variance, and thereby increases the total merit of the population. The index weight on the kinship has the purpose to balance the benefit that arises from using outcross individuals on increasing selection response in future generations, and the benefit that arises for traits under stabilizing selection from reducing the phenotypic variance. In a dominance model with inbreeding depression, the weight put on the kinship could also account for the negative effect of inbreeding depression. Selection in livestock breeding is currently not based on the optimum selection index, which might contribute to the low effective population sizes of various breeds.

References

Dekkers, J. C. M., Birke, P. V., and Gibson, J. P. (1995). Animal Science. 61 (1):165-175. https://doi.org/10.1017/S1357729800013667

Hazel, L. N. (1943) Genetics. 28 (6): 476-490

Moav, R. and Hill, W. G. (1966). Animal Production. 8: 375-390. https://doi.org/10.1017/S000335610003806X

Smith, F. H. (1936). Annals of Eugenics. 7 (3): 240-250. https://doi.org/10.1111/j.1469-1809.1936.tb02143.x

Wellmann, R. (2022). Selection index theory for populations under directional and stabilizing selection. submitted.

471. Advantages of modularisation for industry breeding objective development

C.D. Quinton[1*], P.R. Amer[1], T.J. Byrne[2] and B.F.S. Santos[1]

[1]AbacusBio Limited, 442 Moray Place, Dunedin 9016, New Zealand; [2]AbacusBio International Limited, Roslin Innovation Centre, University of Edinburgh, Easter Bush Campus, Midlothian EH25 9RG, United Kingdom; cquinton@abacusbio.co.nz

Abstract

Bioeconomic models are commonly used in advanced breeding programs to calculate breeding objective economic values and index trait weightings. As industry genetic improvement programs expand to include more traits and diverse farming styles, models necessarily increase in complexity. Here we describe flexible modular approaches to bioeconomic models and breeding objective development that are readily adapted to new traits and farming systems. This approach allows substantial emphasis on user acceptability, transparency, and ultimately adoption of index-based selection.

Introduction

The goal of multi-trait breeding objective development is to derive the relative weightings for various goal traits to generate optimal overall genetic gain. Typically, in advanced animal breeding objective design this means calculating the economic value (EV) of each goal trait and generating selection index weightings. An EV can be defined as the change in profit per 1-unit change in the trait, independent of all other objective traits. While a technically efficient balance of traits is important, the balance of weightings can also have critical implications for index adoption, which can dramatically improve selection efficiency relative to reliance on ad-hoc threshold criteria.

The classic profit function and bioeconomic model approaches to EV calculation have been well described (e.g. Goddard, 1998). In a profit function approach, a simple linear profit equation is built that describes the influences all breeding objective traits have on profitability via revenues and costs. For each trait, the EV is calculated as the partial derivative of the profit function with respect to that trait keeping all other traits constant (at population mean). This process can work well in simple situations, with a small number of traits and straightforward associations with profit, and where genetic parameters among breeding objective traits that have direct economic value and available selection criteria (EBVs) are well established. However, modern large-scale breeding programs have many objective traits spanning production quantity and quality, fertility, survival, health and welfare. Environmental traits are a further new addition, and it is common for a number of these traits to have a complex or obscure relationships with system profitability. For example, in a dairy system, longevity will influence herd dynamics and age structure which in turn affect lactation patterns, fertility, and milk production; these all have revenue and cost, as well as environmental, implications. As stated by Nielsen *et al.* (2014), simple linear profit function approaches are usually inadequate to calculate EVs for advanced breeding programs.

Sophisticated bioeconomic models are a better approach to represent complex herd/flock structures and dynamics which are necessary to value functional traits such as longevity that are increasingly important to breeding programmes (Nielsen *et al.*, 2014). These models usually encompass many traits and accommodate sophisticated interactions among genetic traits and management variables. EVs are calculated from changes in revenues and costs that derive from trait changes. For breeding objective design, bioeconomic models are preferably designed specifically for calculating EVs; those built for other purposes (e.g. nutrition models) tend to contain over-complex interactions among traits, and/or between traits and management that are

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_471, © C.D. Quinton *et al.* 2022

irrelevant to EVs. Embedded genetic correlations/associations that violate the EV definition of change in profitability independent of other objective traits can be a particular concern. A challenge of bioeconomic models is the scale of knowledge needed. Analysts need a comprehensive understanding of the model contents and calculation processes to ensure EVs are correctly calculated and adapted for new knowledge, objectives and genetic evaluation development.

Modular bioeconomic model approach

Nielsen *et al.* (2014) proposed that a trait-by-trait or modular approach to bioeconomic modelling has advantages for calculating EVs in advanced breeding programmes. In this structure, the EV calculations for each trait are built as separate modules. For some traits the EV module can be based on simple profit functions. For traits with greater complexity, EV modules may contain detailed models of long-term growth or other biological change with calculated nutrient requirements, or herd structure models. Separate modules can be built for traits in alternate production systems within an industry and resultant EVs can be proportionately weighted to create an overall trait EV. Where a trait influences profitability at multiple points in time or in different generations, modules can be built for each trait and resultant EVs weighted by discounted genetic expressions (e.g. Amer, 1999; Quinton *et al.* 2021 mature cow body weight) to create an overall trait EV. It is crucial that economic, biological and environmental parameters are consistent, therefore all modules should draw from a common set of core input parameters. While care is required to avoid double-counting trait value across modules, in some instances, a small degree of double counting can great simplify the index calculation, without deviating from a highly efficient and industry accepted outcome. This modular bioeconomic model approach to breeding objective development has been successfully applied in a wide range of animal breeding objectives. Here we describe advantages of this approach with examples.

Transparency. Modularised bioeconomic modules are generally easier for all levels of breeding objective developers and users to understand and interpret. Transparency is key to uptake through stakeholders' understanding how their needs are included in the breeding objective design. In a large integrated model, many trait interactions can obscure the economic consequences of trait change; this can leave users disconnected and potentially sceptical of the methods. The modular approach allows each trait to be valued one at a time, which can be more intuitive for farmers and non-geneticists. The modular approach also allows application of subindexes that users can understand easily and apply for their specific selection targets. For instance in sheep, the New Zealand Maternal Worth (https://www.blnzgenetics.com/ram-selection-tools) standard index is built from a set of 5 subindex modules (lamb growth, adult size, reproduction, lamb survival, wool) that sum to the total index. Here, modularisation provides farmers with a middle ground that is more readily understood than 10 EBVs and a total index value. There is also a level of user customisation with optional add-on subindexes for meat yield, health, and other trait groups for farmers willing to record; EVs for each of these options are derived in specific trait modules.

Adaptable to new evaluations. With a modular structure, it is more straightforward to add or change individual trait EV calculations when genetic evaluations change, without having to re-calculate other goal trait EVs. This adaptability is increasingly important as genomics is rapidly adding new trait evaluations for both nucleus/breeder level and commercial animals. Axford (2018) described the challenge of adding Feed Saved to developed Australian dairy indexes (Byrne *et al.* 2016); here the modular approach allowed this trait to be incrementally added to the index without changing the EVs for other traits. It is also an advantage for genetic evaluation implementation code, minimizing the complexity of changes needed. Similarly, modularity allows exploration of traits where there is less confidence in genetic parameters or genetic relationships without influencing EV calculations of traits whose genetics are well understood.

Non-linear economic value flexibility. Non-linear EV functions are increasingly of interest in breeding objectives (Quinton et al. 2019;). In cases where a linear EV is too simple or does not reflect profitability in diverse populations, non-linear functions can provide more realistic valuation of traits within a breeding objective and are intuitive to users. However, in a fully integrated bioeconomic model where traits interact, non-linear functions can confuse the relationships between other traits and profitability and make it difficult to understand the exact drivers of correlated traits' EVs. Therefore, a modular approach that values each trait independently is critical to implement non-linear EV functions. We have had many favourable experiences with non-linear indexes in practice. For example, farmers generally wish to penalise very difficult calving bulls. However, they do not want a very easy calving bull to rank well if they do not deliver good growth and carcase composition; non-linear EV have therefore been implemented in beef and dairy beef objectives (Berry et al. 2019; Miller et al. 2021). Maternal milk production in extensively reared beef cattle is another trait where low milk is not desirable, but farmers also see little benefit from high milk production. In sheep where there is often a need to compare a large range of biological types using a single index, we have found the non-linear approach to work well for traits such as litter size and wool fibre diameter (Quinton et al. 2019).

Non-economic factor inclusion. Recent trait prioritisation surveys have revealed that functional health and welfare traits tend to be more important to farmers than economic calculations suggest, and intrinsic differences in farmers' trait priorities exist that are independent of economics (Martin-Collado et al., 2015, 2016; Santos et al. 2019, 2022). The modular approach allows pure economics EVs from a profit equation or bioeconomic model to be combined with preference-based weighting, which can increase farmers' acceptance of breeding objectives (Nielsen et al., 2014). This also applies to environmental impact traits such as methane production for which evaluations are now being implemented in several species. Methane production is generally closely associated with production traits and feed intake, but current carbon price-based economics tend not to incentivise emissions reduction. Therefore, some non-economic emphasis should be explored for methane while retaining the EVs for production traits that have clear economic drivers. Modularity allows flexibility in defining these non-economic traits' importance without influencing the balance among other traits that have more concrete economic value.

Transferable to new systems. As genetics providers expand markets, breeding programmes need to address the diversity of production systems and environments that can occur in an industry. A modular approach provides clear options for breeding objectives to address alternate production systems or environments within an industry. In these cases, trait EVs can be calculated separately for each system or environment, then these either proportionately weighted to create an overall trait EV (e.g. Quinton *et al.* 2021) or applied into separate indexes. Modules can be set up to deal with genotype by environment issues which might otherwise need more complex solutions, often involving multiple genetic evaluations.

Testing and implementation

As mentioned above, modularisation can risk double-counting the value of a trait if the modeller is not aware of the full bio-economic model contents. For example, feed intake traits are increasingly being added to breeding objectives, but EV calculations for these must consider the EBV definition (e.g. total intake vs residual intake) and if other traits such as growth have feed costs included in their EV calculation. Developers need a good understanding of all modules to ensure double-counting is avoided, and that correlations with important traits not in objective are included. To combine module EV results and evaluate how derived breeding objectives perform we test by applying EVs to real EBV data. In this way we can observe effects of trait correlations and differences in variance (see Zhang *et al.*, 2022).

Conclusions

Flexible approaches to breeding objective development facilitate the expansion of indexes to accommodate new traits in breeding programmes, allowing for a richer and more socially acceptable outcome from genetic selection. The modular approaches we have used for many different countries and species allow us to place substantial emphasis on user acceptability, transparency, and ultimately adoption of index-based selection and in this way have contributed substantially to the value of genetic progress being achieved.

References

Amer P.R. (1999) NZ J. Agric. Res. 42(3):325-336. https://doi.org/10.1080/00288233.1999.9513382

Axford, M. (2018) Proc. of 11 WCGALP, Auckland, NZ.

Axford, M., Santos, B., Stachowicz, K., Quinton, C., *et al.* (2021) Anim. Prod. Sci. 61:1940–1950. https://doi.org/10.1071/AN21058

Berry, D.P., Amer, P.R., Evans, R.D., Byrne, T., *et al.* (2019) J. Dairy Sci. 102:10056–10072. https://doi.org/10.3168/jds.2019-16912

Byrne, T.J., Santos, B.F.S., Amer, P.R., Martin-Collado, D., *et al.* (2016) J. Dairy Sci. 99:8146–8167. https://doi.org/10.3168/jds.2015-10747

Goddard, M.E. (1998) J. Dairy Sci. 81:6–18. https://doi.org/10.3168/jds.S0022-0302(98)70150-X

Martin-Collado, D., Byrne, T.J., Amer, P.R., Santos, B.F.S., *et al.* (2015) J. Dairy Sci. 98:4148–4161. https://doi.org/10.3168/jds.2014-9194

Martin-Collado, D., Hely, F., Byrne, T.J., Evans, R., *et al.* (2017) Animal 11:318–326. https://doi.org/10.1017/S1751731116001567

Miller, S.P., Archer, J.A., Hely, F., Quinton, C.D., *et al.* (2021) Proc. of 24 Assoc. Advmt. Anim. Breed. Genet.

Nielsen, H.M., Amer, P.R., and Byrne, T.J. (2014) Acta Agric. Scand. A 64:2–12. https://doi.org/10.1080/09064702.2013.827237

Quinton, C.D., Amer, P.R., Byrne, T.J., Archer, J.A., *et al.* (2019) Proc. of 23 Assoc. Advmt. Anim. Breed. Genet. Armidale, Australia.

Quinton, C.D., Archer, J.A., Amer, P.R., Harburg, S., *et al.* (2021) Proc. of 24 Assoc. Advmt. Anim. Breed. Genet.

Santos, B., Archer, J.A., Martin-Collado, D., Quinton, C.D., *et al.* (2019) Proc. of 23 Assoc. Advmt. Anim. Breed. Genet. Armidale, Australia.

Santos, B.F.S., Jenkins, G.M., Kok, J., Stachowicz, K., and Amer, P.R. (2022) Proc. of 12 WCGALP, Rotterdam, NL.

Zhang, X., Amer, P.R., and Quinton, C. (2022) Proc. of 12 WCGALP, Rotterdam, NL.

472. Enhancement of the terminal carcass production index to incorporate birth weight and lambing ease in Australian sheep

E. Dehnavi* and A.A. Swan

AGBU, a joint venture of NSW Department of Primary Industries and University of New England, 2351, Armidale, Australia; edehnavi@une.edu.au

Abstract

Sires ranking highly on Australian terminal sheep indexes tend to have higher birth weight (BWT), which potentially translates to increased incidence of difficult births in commercial flocks. The Terminal Carcass Production industry index was therefore modified to include a penalty targeting a genetic change of -0.005 kg in BWT over 10 years. An alternative scenario was also considered, where an economic value for lambing ease (LE; %) was added to the index. The largest impact of the BWT penalty approach was a reduction in the response for growth traits with little improvement in LE but with minimal impact on carcass yield and eating quality. The economic value for LE was calculated as $1.15/1% improvement in lambing ease/ewe/year. Compared to the first approach, including LE in the index produced is preferred with similar outcomes for birth weight but with less impact on response for the growth traits and also resulted in a positive response in LE.

Introduction

Terminal sheep breeds in the LAMBPLAN genetic evaluation system in Australia have achieved high rates of genetic gain in growth and lean meat yield over an extended time period (Swan *et al.*, 2017). From 2017, new selection indexes were introduced to address declining eating quality, as described by Swan *et al.* (2015), including the Terminal Carcass Production (TCP) index which balances increasing growth and lean meat yield with modest improvements in eating quality.

Genetic trends over the past five years confirm that as selection emphasis has been directed to eating quality, with favourable changes in Australian Sheep Breeding Values (ASBVs) for intramuscular fat, shear force, and consumer eating quality, there have also been increases in birth weight. These increasing genetic trends in birth weight are higher than predicted and have potentially occurred due to breeders placing attention on the new eating quality ASBVs at the expense of birth weight.

Highly ranked sires that have high birth weight ASBVs may lead to a greater incidence of lambing difficulty. A genetic evaluation system for lambing difficulty from data recorded in terminal flocks is available for terminal sire breeds (Li *et al.* 2021), with ASBVs for the trait lambing ease. Therefore, to address this concern at selection, two modifications to TCP were considered: firstly; adding a desired gains penalty on increasing birth weight, and secondly; including an economic value for lambing ease directly in the breeding objective.

Materials & methods

Terminal carcass production breeding objective. The TCP breeding objective places emphasis on increasing post-weaning weight (PWT), carcass eye muscle depth (CEMD), lean meat yield (LMY), dressing percentage (DRESS) and consumer eating quality (MSA), while reducing carcass C-site subcutaneous fat (CCFAT; Swan et al. 2015). Balanced selection to overcome an antagonistic relationship between yield and eating quality attributes has been made possible through the availability of genomically enhanced ASBVs for these traits, along with two eating quality selection criteria traits, intramuscular fat (IMF) and shear force (SF5). Weightings for these traits are included in the TCP index.

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_472, © E. Dehnavi and A.A. Swan 2022

Desired gains index for birth weight. Birth weight (BWT) was added to the TCP index (referred to as TCPB index) using a desired gains approach to hold BWT constant while maintaining gains in other traits. This was essentially done by calculating an index weight that targeted a genetic gain of -0.005 kg for BWT over 10 years. It was assumed that the traits measured to derive the desired gains index included BWT, weaning weight (WWT), PWT, post-weaning eye muscle depth (PEMD) and post-weaning fat depth (PFAT) from ultrasound scanning of live animals. In addition, genomic predictions for PWT, CEMD, CCFAT, LMY, DRESS, IMF, and SF5 were assumed to be available and used in the derivation of the index, following Dekkers (2007). Population structure parameters used to annualise predicted genetic gain were based on data extracted from the LAMPLAN terminal sire analysis (16 August 2021).

Relative economic value for lambing ease. The economic value for lambing ease was calculated assuming an underlying normally distributed liability, partitioned by thresholds (Meijering, 1980) into categories defined as 1 = no assistance (normal birth), 2 = moderate assistance (assisted, or evidence of difficult birth for an unassisted ewe), and 3 = hard assistance (Li *et al.* 2021). Thresholds between categories were derived from the LAMBPLAN terminal sire database, with probabilities for each category of 1=0.9006, 2=0.0775, and 3=0.0219. The lambing ease economic value was calculated by determining the change in profit from a 1% improvement in lambing ease, acting through reduced costs associated with lower incidence of difficult births in categories 2 and 3. Assumed costs associated with additional labour, lamb and ewe mortalities are shown in Table 1. Note that the inclusion of costs for dead lambs assumes that mortalities due to lambing difficulty are considered separately to other causes. Lamb mortality rates relative to unassisted live lambs (Table 1) were derived from observed lambing data in maternal breeds in the genetic evaluation system and represent ram breeder rather than commercial flocks. The 1% improvement in lambing ease was modelled by changing the probabilities for each category proportional to their distribution. The changes in proportions for each category were then multiplied by the associated cost relative to unassisted births and aggregated to calculate the overall economic value. This was expressed in terms of dollars per ewe joined per year through multiplication by 1.2 number of lambs born per ewe joined, representative of the Merino ewes most commonly mated to terminal sires. The TCP index augmented with lambing ease is referred to as the TCPL index.

Comparison of indexes. The TCP, TCPB, and TCPL indexes were calculated using ASBVs for 1,792 sires of Poll Dorset and White Suffolk progeny born in 2020 in the LAMPLAN terminal sire evaluation. An additional index, Lamb 2020 (LP2020) was also included for comparisons. The LP2020 is a legacy (pre-TCP) index which included emphasis on growth and yield, with a small penalty on increasing birth weight, but no emphasis on eating quality (Swan *et al*., 2017). The indexes were compared using 'selection advantage',

Table 1. Cost assumptions for moderate and hard lambing difficulty categories relative to live unassisted lambs used to calculate the economic value of lambing ease.

Item	Unit	Moderate	Hard
Stockman time	hours	0.2	0.85
Stockman cost per hour	$/hour	35	35
Relative percentage of lamb mortalities	%	30	45
Cost per dead lamb	$	250	250
Relative percentage of ewe mortalities	%	0	1
Cost per dead ewe	$	240	240
Total cost relative to no assistance	$	82	144.65

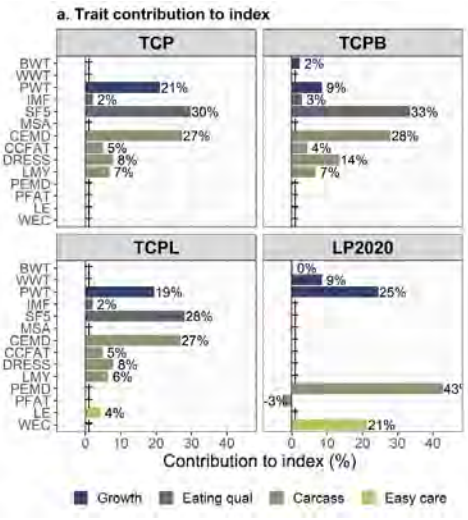
calculated as the trait superiority of the top 10% of sires selected on index, along with correlations between sire index values.

Results & discussion

The relative economic value for a 1% increase in lambing ease was calculated to be $1.15 per ewe per year. Byrne *et al.* (2010) calculated an economic value of -€0.247 for single-lambing difficulty per ewe. The relatively higher value calculated in this study was due to the inclusion of the costs of dead lambs along with higher lamb mortality rates in assisted categories. As noted above, these incidence and mortality rates have been derived from ram breeder flocks in the genetic evaluation system and may differ in commercial flocks. Survey data from commercial flocks would help to refine the calculations.

Comparing selection advantages for the top sires ranked on TCP, TCPB and TCPL indexes indicated that sires selected on TCPB index had lower average ASBVs for BWT by -0.09 kg, at the expense of 29% lower superiority for PWT (1.53 compared to 2.17 kg for TCP, Figure 1). The loss for WWT was relatively higher (0.60 compared to 1.06 kg for TCP). This was expected as birth weight had a 2% contribution to the TCPB index and there were high correlations between BWT and the other growth traits (Figure 1). The top sires remained negative for lambing ease (-0.46 and -0.16%) when ranked on the TCP and TCPB indexes, indicating that direct emphasis is required to achieve positive change in lambing ease.

Compared to TCPB, the TCPL index led to a lower reduction in superiority for weaning weight (0.94 kg for TCPL compared to 0.60 kg for TCPB) and post-weaning weight (2.02 compared to 1.53 kg). This shows that adding lambing ease directly to the TCPL index is beneficial in retaining response in growth traits compared to penalising birth weight through TCPB. While superiority for eating quality and carcass traits was marginally inferior for TCPL, differences with TCP and TCPB for these traits were not great. Importantly, TCPL was the only index to realise a favourable response in lambing ease, while at the same time giving the best outcome for birth weight (equal with TCPB).



b. Superiority of top 10% (n=1792)

trait	TCP	TCPB	TCPL	LP2020
BWT (kg)	-0.02†	-0.09	-0.07†	-0.05
WWT (kg)	1.06†	0.6†	0.94†	1.3
PWT (kg)	2.17	1.53	2.02	2.49
IMF (%)	0.24	0.29	0.23	0.02†
SF5 (N)	-2.85	-3.07	-2.71	-0.67†
MSA (grade)	1.36†	1.42†	1.07†	-0.56†
CEMD (mm)	1.29	1.42	1.28	1.19†
CCFAT (mm)	-0.24	-0.21	-0.24	-0.21†
DRESS (%)	0.83	0.86	0.85	0.87†
LMY (%)	0.68	0.57	0.65	0.86†
PEMD (mm)	1.25†	1.44†	1.26†	1.33
PFAT (mm)	0.15†	0.21†	0.17†	0.22
LE (%)	-0.46†	-0.16†	0.68	-0.25†
WEC (%)	-10.9†	-12.1†	-10.4†	-25.5

Traits that are not in objectives are specified by †.

Figure 1. Trait contribution to the current (TCP), modified (TCPB and TCPL) and previous (LP2020) indexes (a) and, trait selection advantages for the top 10% of current Poll Dorset and White Suffolk sires ranked on indexes (b). Abbreviations are described in the text.

The LP2020 index includes selection emphasis to limit birth weight, and the outcome was superior to TCP for this trait (-0.05 compared to -0.02 kg), while achieving the highest responses for growth traits of any index (1.3 kg for WWT and 2.5 kg for PWT). However, this comes at the cost of declining eating quality (IMF, SF5 and MSA), which is not included in the index. In addition, the response in lambing ease was unfavourable.

The correlation between TCP and TCPB indexes for these sires was 0.98, indicating that the modification has minimal impact on rankings overall, as expected from the small contribution of birth weight in the TCP index (2% of index gains in Figure 1). The correlation between TCP and TCPL was lower at 0.97 consistent with a slightly higher contribution from lambing ease to TCPL (4% in Figure 1). Correlations for LP2020 were lower at 0.86, 0.82 and 0.84 with the TCP, TCPB and TCPL indexes, respectively, due to the lack of any direct emphasis on carcass and eating quality traits.

Birth weight is not currently included in selection indexes for Australian sheep, but positive emphasis on birth weight may be useful where low birth weights are associated with poor lamb survival, whereas a negative penalty can be desirable when high birth weights lead to increased incidence of lambing difficulty. From a breeding objective perspective specifying economic values for survival and lambing ease directly will be preferable, as this study has shown for lambing ease. Accurate genetic evaluations are also required to support lambing ease ASBVs to be used in indexes, and while new genomically enhanced ASBVs will soon be available (Li *et al.* 2021), it is also important for breeders to improve the collection of data on lambing ease, survival, and birth weight to underpin the system.

Acknowledgements

This research was funded by Meat and Livestock Australia (project L.GEN.1704).

References

Byrne T.J., Amer P.R., Fennessy P.F., Cromie A.R., Keady T.W.J. *et al.* (2010) Livest. Sci. 132(1): 135–144. https://doi.org/10.1016/j.livsci.2010.05.013.

Dekkers J. C. (2007) J. Anim. Breed. Genet 124 (6): 331–341. https://doi.org/10.1111/j.1439-0388.2007.00701.x.

Li L., Gurman P. M., Swan A.A., and Brown D.J. (2021) Anim. Prod. Sci. 61(18): 1990-1999. https://doi.org/10.1071/AN21257.

Meijering A. (1980) Livest. Prod. Sci. 7(5): 419-36. https://doi.org/10.1016/0301-6226(80)90081-0.

Swan A.A., Brown D. J., Banks R. G., and Chandler H.R. (2017) Proc. Assoc. Adv. Anim. Breed. Genet. 22:365-368

Swan A.A., Pleasants T., and Pethick D. (2015) Proc. Assoc. Adv. Anim. Breed. Genet 21:29–32.

473. iGENDEC: a web-based decision support tool for economic index construction

M.L. Spangler[1*], B.L. Golden[2], S. Newman[3], L.A. Kuehn[4], W.M. Snelling[4], R.M. Thallman[4] and R.L. Weaber[5]

[1]University of Nebraska-Lincoln, 1400 R St, Lincoln NE, 68583 USA; [2]Theta Solutions, LLC, 4745 Plover St NE, Olympia, WA 98516, USA; [3]Genus-ABS, 1525 River Rd. DeForest, WI 53532, USA; [4]USDA, ARS, Roman L Hruska US Meat Animal Research Center, Clay Center, NE 68933, USA; [5]Kansas State University, 1007 Throckmorton Hall, Manhattan, KS 66506, USA; mspangler2@unl.edu

Abstract

Economic selection indexes are the preferred method of practicing multiple trait selection. In the U.S., the majority of beef breed associations that have genetic evaluations provide generalized indexes. However, utilization of these tools has been less than is desired. IGENDEC, a web-based platform underpinned by a stochastic simulation, was written in the GO programming language, and was developed to enable the construction of user-specific economic selection indexes. The platform made use of user-specified breeding objectives, breeding systems, costs, returns, and phenotypic means. Interface level output included index weights and graphical representation of relative emphasis and Pearson correlations between the custom index and component trait expected progeny differences (EPD). The system has undergone beta testing and source code was placed in a publicly available repository. The USDA is an equal opportunity employer and provider.

Introduction

Germplasm selection at an enterprise level requires contemplating both genetic and economic progress. Selection index theory (Smith, 1936; Hazel and Lush, 1942) enables the construction of multi-trait selection tools towards improvement in a specified breeding goal. Common across animal industries and countries is the use of economic selection indexes which rank selection candidates on their genetic potential to improve net profit, where net profit should be defined at the commercial level. Critical to the formation of such selection tools are the underlying assumptions relative to expected costs and returns, breeding systems and objectives, and current levels of animal performance. Unique enterprise-level circumstances exist that can alter the economic weighting of traits in the decision-making process, and potentially lead to a re-ranking of selection candidates.

The objectives of the current work were to develop and deploy a web-based decision support platform (iGENDEC) that enables the comparison of selection candidates based on economically optimized indexes considering the unique circumstances of the user.

Materials & methods

Use cases. Three use cases were defined, including producers wishing to purchase germplasm; sell germplasm; or make mating decisions based on a custom index. The iGENDEC software can be used to create either generalized (i.e. breed association provided) or customized (i.e. unique to an enterprise) indexes. High-level breeding goals specified by the user included sale endpoints at weaning, after a backgrounding phase, and at slaughter (live- or carcass-basis). The goal considered can be terminal (no female replacements kept) or general purpose (female replacements retained).

Simulation model. The iGENDEC program was written in the GO programming language and has been tested on Linux systems. The stochastic simulation was modeled under the assumptions of a North

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_473, © M.L. Spangler *et al.* 2022

American beef industry setting. A cow herd was simulated and genetic values of bulls were perturbed by one unit, one trait at a time, to determine the Marginal Economic Value (MEV) of each trait in the breeding objective. The model accounted for discounted gene flow, discounted expression rates, and differences in planning horizons. Philosophically, planning horizon was thought of as the length of time (years) that the user wanted to consider in determining the economic impact of a genetic selection decision. Operationally, planning horizon represented the number of years simulated with the improved genetic merit of bulls.

Parameters. Hard-coded parameters included genetic and residual (co)variances, breed contrasts, and heterosis (direct and maternal) values. These parameters were obtained from the scientific literature (Koots et al., 1994a,b) and the U.S. Meat Animal Research Center (USMARC) Germplasm Evaluation (GPE) program (Schiermiester et al., 2015; Ahlberg et al., 2016; Kuehn and Thallman, 2017; Retallick et al., 2017; Freetly et al., 2020; Zimmermann et al., 2021). Breeds included as options were those with current genetic evaluations in the U.S. These parameters were contained in the program's master hjson file. User-defined parameters included value of cattle sold (US$/cwt) with differential values based on weight and sex classes and carcass quality, costs (variable cow costs, cost of feed for growing cattle), death loss rate due to dystocia, pregnancy rate, cow age distribution, breeds (bulls, cows, calves), and current phenotypic means. Default values for user-defined variables were also available. Economic index-specific parameters were specified in separate hjson files.

Web interface. The interface required user accounts that were password protected, allowing input through a series of drop-down menus (i.e. breeding objectives, breed selection) and either click and drag or tabular entry (i.e. cow age distribution, value of sale weight). Jobs saved under user-defined names and comments pertaining to a given job were allowed such that key components were easily remembered by the user for future reference. Previous jobs could be edited and re-run. The capability for user-specific branding existed such that an organization (i.e. breed association) could offer iGENDEC as a service to its membership and retain identity of the organization if iGENDEC were to be accessed through a common server external to the organization and available to numerous such organizations.

Industry engagement. An industry advisory board comprised of beef breed association technical staff, producers (seedstock and commercial), and extension personnel was established. The advisory board provided initial direction on the scope and structure of the project, then served as alpha-version testers. Additional beta-version testers were recruited from industry.

Results

User interface output. User-observed run time was ~2 minutes for most jobs. Run time increased roughly linearly with the length of the planning horizon. Resulting index weights were provided on screen and relative emphasis was depicted graphically using a pie chart. A bar chart depicted the Pearson correlations between the custom index and component EPD among the selection candidates. The user could download a csv file containing the custom index (default) and selected information from the group of individuals the user wished to apply the index to (i.e. animal identification, component EPD, etc.). The interpretation of differences between two animal's index values was the difference in enterprise-level net profit per exposure.

Software release. Source code and documentation were made available on github (https://github.com/blgolden/igendec). Although the current work has been supported by a competitive grant through the USDA-AFRI-CARE program (Award Number: 2018-68008-27888), sustainability of this system can only be achieved by availability to industry and corresponding use. To this end, the Beef Improvement Federation (BIF) is currently exploring a fee-based model for hosting iGENDEC.

Discussion

The development of decision support tools and systems (DSS) is not new (Newman *et al.*, 2000) and indeed current examples of such platforms exist outside of the U.S. (e.g. Barwick and Henzell, 2005; Dunne *et al.*, 2021). Historically such tools have not been maintained due to a lack of funding and less than ideal rates of adoption by end users. Lynch *et al.* (2000) detailed reasons why such decision support tools in agriculture suffer low adoptions rates, including: Lack of field testing; No end user input preceding and during development of the DSS; DSS complexity and possibly considerable data input; No reason seen for changing current management methods; Distrust for the output of a DSS because producers do not understand the underlying theories of the models; Mismatch of the DSS output with the decision-making style of the producer because the producer's conceptual models are excluded; and Unclear definition of the beneficiaries (e.g. scientists, primary producers, and technology transfer agents). To combat these known hurdles to software sustainability, the current project established an industry advisory board described above.

Model considerations such as discounted gene flow and discounted expression rates are not uncommon in the development of economic selection indexes. However, to our knowledge the notion of planning horizon has not been reported previously relative to the construction of economic indexes in beef cattle although this concept has been reported in dairy (McMahon *et al.*, 1985). Given that germplasm purchases (i.e. bull purchase) is an investment, the return on investment should have a specified time horizon. Shorter planning horizons favor traits that are expressed earlier in life (i.e. sale weight) while longer planning horizons tend to favor traits such as sustained fertility that are expressed much later. Pragmatically a user might specify a relatively short (e.g. five-year) planning horizon despite the desire to emphasize sustained fertility in a self-replacing herd simply due to the need for operating capital that can be generated by an increased emphasis on sale weight. Such trade-offs, and more formal comparisons of efficiency between short- and long-term horizons, is a needed area of future research.

The current framework of iGENDEC enables the addition of more breeds if and when additional organizations adopt genetic evaluations and the addition of new traits should genetic predictors become available. The vision for this system is that it has the capability of evolving as the needs of the U.S. beef industry evolves, but this requires that the current system be adopted and supported.

References

Ahlberg, C.M., Kuehn L.A., Thallman R.M., Kachman S.D., Snelling W.M., *et al.* (2016) J. Anim. Sci. 94(5):1857-1864. https://doi.org/10.2527/jas.2015-0161

Barwick S.A., and Henzell A.L. (2005) Aust. J. Exp. Ag. 45:923-933 https://doi.org/10.1071/EA05068

Freetly, H.C., Kuehn L.A., Thallman R.M., and Snelling W.M. (2020) J. Anim. Sci. 98(1) https://doi.org/10.1093/jas/skz394

Hazel L.N., and Lush J.L. (1942) J. Hered. 33:393–399. https://doi.org/10.1093/oxfordjournals.jhered.a105102

Dunne F.L., Evans R.D., Kelleher M.M., Walsh S.W., and Berry D.P. (2021) Animal 15: https://doi.org/10.1016/j.animal.2020.100077

Koots K.R., Gibson J., Smith C., and Wilton J. (1994). Animal Breeding Abstracts 62:309-338.

Koots K.R., Gibson J., and Wilton J. (1994). Animal Breeding Abstracts 62:825-853.

Kuehn, L.A. and Thallman R.M. (2017) Across-breed EPD tables for the year 2017 adjusted to breed differences for birth year of 2015. pp112-144. Available at: https://beefimprovement.org/wp-content/uploads/2013/07/2017BIFProceedings.pdf.

Lynch T., Gregor S., and Midmore D. (2000) Aust. J. Exp. Agric. 40:609-620.

McMahon R.T., Blake R.W., Shumway C.R., Leatham D.J., Tomaszewski M.A. *et al.* (1985) J. Dairy Sci. 68(9):2295-2302. https://doi.org/10.3168/jds.S0022-0302(85)81102-4

Newman S., Lynch T., and Plummer A.A. (1999) J. Anim. Sci. 77 (suppl_E): 1–12 https://doi-org.eres.qnl.qa/10.2527/jas2000.77E-Suppl1e

Retallick, K.J., Bormann J.M., Weaber R.L., MacNeil M.D., Bradford H.L., *et al.* (2017) J. Anim. Sci. 95:1444-1450. https://doi.org/10.2527/jas2016.1260

Schiermiester, L.N., Thallman R.M., Kuehn L.A., Kachman S.D., and Spangler M.L. (2015) J. Anim. Sci. 93(1):46-52. https://doi.org/10.2527/jas2014-8493

Smith F.H. (1936) Ann. Eugen. 7:240–250. https://doi.org/10.1111/j.1469-1809.1936. tb02143.x

Zimmermann, M.J., Kuehn L.A., Spangler M.L., Thallman R.M., Snelling W.M., *et al.* (2021) J. Anim. Sci. 99(8). https://doi.org/10.1093/jas/skab209

474. Analysis of causal structure of traits involved in sow lactation feed efficiency

M. Mora[1*], I. David[2], H. Gilbert[2], G.J.M. Rosa[3], J.P. Sánchez[1] and M. Piles[1]

[1]Institute of Agrifood Research and Technology (IRTA), Animal Breeding and Genetics, Torre Marimon C-59, km 12,1, 08140 Caldes de Montbui, Barcelona, Spain; [2]GenPhySe, Université de Toulouse, INRAE, ENVT, 31320 Castanet Tolosan, France; [3]Department of Animal and Dairy Sciences of University Madison, 1675 Observatory Dr, Madison, WI 53706, USA; monica.mora@irta.cat

Abstract

Feed efficiency during lactation involves a set of phenotypic traits which form a complex system in which some traits exert causal effects on the others. We proposed a structural equation model to investigate the complex network involving the main components of lactation feed efficiency (i.e. daily lactation feed intake, dLFI; daily sow weight balance, dSWB; daily litter weight gain, dLWG; daily back fat thickness balance, dBFTB; sow metabolic body weight, SMBW), and to predict the effect of external interventions on the system using selection indices. Results indicate that dLFI and dLWG affect dSWB, and dSWB affect dBFTB and SMBW. Multiple indirect links contribute to these effects, especially on the latter two associations. Overall, selection strategies holding dLFI constant did not offer any advantage over the others.

Introduction

Improving feed efficiency during lactation (LFE) has many benefits in terms of economic costs and production-related drawbacks, as well as animal wellbeing. This is because milk yield is one of the most energy-demanding processes in the productive life of a sow (Thekkoot *et al.*, 2016). In order to define an optimal genetic selection strategy to improve LFE, knowledge regarding the variances and covariances among target traits is required at both the genetic and environmental levels. Estimates of such parameters are usually obtained by implementing a Multiple Trait Animal Model (MTAM). However, MTAM describes only association among traits, so no information is retrieved in terms of causal relationships. Associations can be due to common factors directly affecting two traits and/or a causal relationship between the two traits. In the latter case, factors affecting one trait have an indirect effect on the other trait through the causal link that exists from the first trait on the second. Structural equation models (SEM) (Gianola and Sorensen, 2004) allow the representation of causal mechanisms between variables, in which the magnitude of the causal relationships are described by model parameters called structural coefficients. Knowledge about such a causal structure allows the prediction of the effect of external interventions in a trait (e.g. management practices such as feed restriction or cross-fostering) on another trait. The objectives of this research were: (1) to investigate potential causal effects underlying the complex relationships among the main traits involved in LFE in sows; and (2) to propose a selection strategy, based on a selection index including or not FI records, for improving sow productivity with no impairment of its body condition, which is the main driver for future reproductive success.

Materials & methods

Animals and data. The data used for this analysis cover the first 10 generations of an experiment of divergent selection for residual feed intake (RFI) in pigs. The selection process is described in depth in Gilbert at al. (2007), and details about the recorded traits are in Gilbert *et al.* (2012). The main components of LFE considered in this study were daily lactation feed intake (dLFI), daily sow weight balance (dSWB), daily litter weight gain (dLWG), daily back fat thickness balance (dBFTB), and sow metabolic body weight (SMBW). The daily traits were computed as the difference between the corresponding trait at weaning (or at day 21 for dLWG) and at farrowing divided by the number of days in lactation, whereas SMBW was

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_474, © M. Mora *et al.* 2022

computed as the average between sow weight at weaning and sow weight at farrowing to the power of 0.75. After removing outliers, the dataset used for the analysis contained information regarding 1,342 farrowings from 576 sows.

Statistical analysis. To assess potential causal relationships between the main components of LFE the following two-steps procedure was implemented:

Step 1: Searching for recursive causal structure

The search of the casual structure was performed on the residual (co)variances obtained from a MTAM, as proposed by Valente *et al.* (2010). These (co)variances provide information from the joint distribution of phenotypic traits conditional to unobservable genetic and permanent environmental effects. The residual covariance matrix obtained from the MTAM was then used for the subsequent analysis of the inductive causation (IC) algorithm using a program written in R (Team, 2020) by Valente and Rosa (2013). This algorithm performs a series of statistical decisions based on partial correlations between pairs of traits based on their highest posterior density (HPD) intervals. This algorithm is well described in Valente *et al.* (2010). The output of the IC algorithm is typically a directed acyclic graph (DAG), which consists of a set of variables (symbolized by nodes) connected by directed edges (symbolized by arrows) representing direct causal relationships.

Step 2: Fitting a SEM

Once the recursive causal structure was established for the traits under study, a SEM with this specific structure was fitted:

$$y = \Lambda y + Xb + Z_1a + Z_2p + Z_3cg + e \quad (1)$$

where Λ is a ($t \times t$) matrix with zeros in the diagonal and with structural coefficients (λ) in the off diagonals (t is the number of traits), y is the vector of phenotypic records, b is the vector of systematic environmental effects including the effects of the farm (with 2 levels), parity order (with 5 levels: 1,2,3,4,>4) and the 2-degree polynomial of litter size at day 21 as a covariate; a, p, cg and e are the vectors of additive genetic effects, permanent effects, contemporary group effects and residual effects, and X, Z_1, Z_2 and Z_3 are incidence matrices relating systematic and random effects to y. Similarly to a MTAM, the SEM provides estimated (co)variance matrices, with the additional information of potential direct and indirect causal effects allowing the prediction of the consequences of external interventions on the system.

Both models were fitted using a Bayesian approach via Gibbs sampling algorithm with a chain of 2,000,000 iterations, a burn-in of 300,000 and an interval between saved samples of 100 using Gibbsf90 and Postgibbsf90 software (Misztal *et al.*, 2018).

Selection strategies. All genetic selection strategies proposed were based on selection indices in which the objectives were always increasing dLWG, reducing or keeping constant dLFI, and increasing or keeping constant dSWB through records on dLWG and dBFTB, and including or not records on dLFI. Since there is no published information regarding the economic weights of the analyzed traits, the selection index coefficients were computed for different combinations of three arbitrary values (0,1,2) with the appropriate sign according to the selection objective. An additional strategy was proposed which consisted in holding constant dLFI for all sows. The idea was to remove this source of variation so that variation in feed efficiency only depends on the variation in dBFTB and on the variation in dLWG, which are easy to measure.

Knowledge of the causal relationship among traits allows predicting the effect on selection response when the external intervention is applied, without the need of a new experiment and data analysis.

Results

Recursive causal structure. The IC algorithm using $HPD_{95\%}$ returned a DAG that indicated that dLFI and dLWG affected dSWB, with the posterior mean of the corresponding structural coefficients (PM) being 0.12 and -0.03 kg/d, respectively. In turn, dSWB influenced dBFTB and SMBW, with PM equal to 0.07 mm/d and -1.22 kg/d, respectively. These structural coefficients express the expected change in the affected trait by an increment of one unit in the causal trait.

Variance components and heritabilities. In SEM, variance components reflect direct effects of genetic and environmental factors on a trait whereas in MTAM they correspond to total effects, i.e. direct plus indirect effects mediated by other traits that have causal effects. Results from both models indicate that the analyzed traits show moderate to high heritabilities with posterior means (posterior SD) ranging from 0.16 (0.03) for dSWB to 0.41 (0.05) for SMBW in the MTAM (total heritability), and from 0.12 (0.03) for dBFTB to 0.36 (0.06) for SMBW in the SEM (direct heritability). For the traits which are phenotypically affected by one or two traits (i.e. dSWB, dBFTB and SMBW), residual and genetic variances were statistically lower in the SEM model compared to MTAM suggesting a relevant contribution of indirect effects to the variance of those traits.

Association between traits. Posterior means of correlations between traits for genetic, permanent, and contemporary group effects were marginally different between the two models (MTAM and SEM) for the traits that were not connected in the DAG, but differences were larger between the other traits. However, HPD95% intervals were relatively large in all cases precluding any claim of statistical difference. Multiple indirect effects contributed to associations between the analyzed traits. The most relevant were those involved in the association between dSWB and dBFTB, and between dSWB and SMBW.

Response to selection. For all selection strategies, responses followed a similar pattern and, for the same economic weights, responses to selection could not be considered statistically different between approaches. In general, it would be hard or not possible to decrease dLFI while increasing or keeping constant dSWB and increasing dLWG and a favorable response in all traits would be hardly achieved.

Discussion

We investigated the causal structure among traits involved in LFE, and estimated the magnitude of putative causal effects implicated in it using data from a population of pigs selected for RFI during growth.

Structural equation models describe the causal relationships that are present in biological systems, but with the specification of a potential causal structure underlying it. Implementing the IC algorithm is one of the many available strategies to search a network structure compatible with the joint probability distribution of the traits considered. However, in this approach, it is necessary to assume that there are no simultaneous relationships (i.e. feedbacks) between these traits. Moreover, in some situations, it may not be possible to find the links and the causal directions using only data information, so the incorporation of biological prior knowledge may be required to complete the causal structure specification. In our study, the data-driven IC algorithm yielded a full DAG indicating that an increase in dLFI would increase dSWB while an increase in dLWG would decrease dSWB and, in turn, an increase in dSWB would increase dBFTB while decreasing SMBW.

Estimated structural coefficients express the strength of each causal link among traits which is essential to quantify the component of every element of variance component matrices (variances and covariances) that comes from indirect sources of (co)variation. In our study, the most important causal effect was the one of dSWB on SMBW (-1.22 kg/d), indicating that dSWB has a negative effect on SMBW. Selection for increased dSWB would reduce requirements for maintenance given the low to moderate total genetic correlation between those traits (-0.28). The second most important causal effect was the one of dSWB on dBFTB (0.07 mm/d). Our results showed that both traits are highly correlated, with a total genetic correlation of 0.69 and direct genetic correlation of 0.31, indicating that indirect effects exerted by other traits are quite relevant in this relationship.

Knowing the proportion of the association between two traits that is due to either direct or indirect effects is important to design new selection strategies when an external intervention is performed. In our study, holding constant dLFI offered no clear advantage over other strategies.

References

Gianola, D., & Sorensen, D. (2004). Genetics 167(3): 1407–1424. https://doi.org/10.1534/genetics.103.025734

Gilbert, H., Bidanel, J. P., Billon, Y., Lagant, H., Guillouet, P., *et al.* (2012). J. Anim. Sci., 90(4), 1097–1108. https://doi.org/10.2527/jas.2011-4515

Gilbert, H., Bidanel, J. P., Gruand, J., Caritez, J. C., Billon, Y., *et al.* (2007) J. Anim. Sci., 85(12), 3182–3188. https://doi.org/10.2527/jas.2006-590

Misztal, I., Tsuruta, S., Lourenco, D., Aguilar, I., Legarra, A., *et al.* (2018) https://www.researchgate.net/publication/331812692%0Ahttp://nce.ads.uga.edu/wiki/lib/exe/fetch.php?media=blupf90_all2.pdf

Team, R. C. (2020). R: A Language and Environment for Statistical Computing. Available at: https://www.R-project.org/

Thekkoot, D. M., Young, J. M., Rothschild, M. F., & Dekkers, J. C. M. (2016). J. Anim. Sci., 94(6), 2317–2331. https://doi.org/10.2527/jas.2015-0258

Valente, B. D., Rosa, G. J. M., De Los Campos, G., Gianola, D., Silva, M. A. (2010) Genetics, 185(2), 633–644. https://doi.org/10.1534/genetics.109.112979

Valente, B. D., Rosa, G. J. M. (2013) Methods Mol Biol. 2013;1019:449-64. https://doi.org/10.1007/978-1-62703-447-0_21. PMID: 23756905.

Part 28

Methods and tools: breeding goals

475. Manipulating response to selection on many traits

J.H.J. van der Werf* and S.A. Clark

School of Environmental and Rural Science, University. of New England, Armidale, NSW 2351, Australia; jvanderw@une.edu.au

Abstract

While selection index theory is a solid framework for a linear optimization of multiple trait selection, it does not always provide desirable selection responses. Responses can be manipulated by varying the economic weights used in a multiple trait index. However, trait responses depend not only on weights, but also on accuracies of the estimated breeding values (EBVs) as well correlations between these EBVs. Predicted selection responses for some traits can be in an undesirable direction if the traits have unfavourable correlations and if there is a large difference in the economic values of the standard deviation of the EBV, as in that case one trait dominates over the other. We extend insight in a two-trait example to a more general case with many traits in the breeding objective. We propose to investigate for each trait its correlation and its dominance relative to the other traits. These parameters can help determine a strategy to avoid undesirable trait responses, whether by changing economic weights or by aiming for more accurate EBVs, or by a combination of these.

Introduction

The balancing of selection and genetic gain among multiple traits is one of the main challenges of breeding programs. Selection index theory proposed by Hazel (1943) provides a simple and solid framework for providing index weights for the different sources of information available for the various traits. Currently, with most traits of relevance having estimated breeding values based on multivariate best linear unbiased prediction (BLUP), the optimal weights used for an index are simply the economic values derived from bioeconomic models of the relevant production system. These models do not require any information about genetic parameters and trait measurement. However, for predicting and understanding the response resulting from multiple trait selection, it is important to keep in mind that the selection response depends on the accuracy of the EBVs and the correlations among those EBVs. Both the accuracies and the correlations among EBVs are determined by the phenotypic information measured in breeding programs. Covariances among EBVs are easy to determine mathematically, and the response to index selection can be predicted as a result. However, when indexes contain many traits it is not always easy to fully comprehend the response and, more importantly, it is not always easy to see what levers can be used to achieve a more desirable response. Economic weights are frequently manipulated for that purpose, using trial and error, but this can be an inefficient approach if there is a lack of data for some traits.

The dynamics of multiple trait selection are easily illustrated by the simplest case of selection for two traits where graphical representation can be provided by the ellipse model (Moav and Hill, 1966). The ellipse is a graph showing all possible selection responses and the optimal response according to selection index theory is at the tangent of the ellipse and line of equal profit – LEP (Figure 1). The shape of the ellipse is determined by the correlation between EBVs, i.e. conditional on the information available in the breeding program, which is then assumed to be the same for all selection candidates. For a two-trait model, a challenge arises when traits have unfavourable correlations and when correlations between EBVs are high, in which the case the ellipse is relatively flat (Figure 1, left pane). A high correlation between EBVs can be due to having limited information available for one of the traits. In that case, the selection response can be very insensitive to economic if one trait is economically dominant, that is when the LEP is close to horizontal or close to vertical. For example, in dairy cattle EBVs for milk yield are often highly accurate, but

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_475, © J.H.J. van der Werf and S.A. Clark 2022

those for fertility are not if it is infrequently recorded. It can be difficult then to improve fertility without heavily compromising response for milk production. Sensitivity of trait response to economic weights is highest when the LEP is parallel to the main axis of the ellipse, which is when the economic weight per standard deviation of the EBV is equal for the two traits (Van der Werf *et al.*, 2018). The desired response in multiple trait selection can be achieved by changing the index weights or by increasing EBV accuracies, or both. Accuracies depend on the information collected in the breeding program, requiring investment, and this can refer to measurement of phenotypes on the selection candidates, on their relatives, or, increasingly, on measurement in reference populations for genomic selection. We will illustrate a method to predict the effect on response when changing economic weights and/or EBV accuracy, first in a two traits model and then suggest a generalization to the case with many traits.

Selection for two traits

We present first a two-trait example, similar to Van der Werf *et al.* (2018), representing the case with an unfavourable correlation (Figure 1). In the example, both traits are standardized to have a phenotypic variance of 1 and a heritability is equal to the genetic variance (σ_a^2) of 0.3 and economic weights (v_i, i = 1, 2) are 0.625 and 0.4. The genetic correlation is -0.5 and the phenotypic correlation is -0.1. Consider unequal information for traits 1 and 2, such that multivariate EBV accuracies (r_i, i = 1, 2) are 0.84 and 0.48. The shape of the ellipse is determined by the correlation between EBVs for the two traits (-0.91, Figure 1 left pane). Note that this correlation is much stronger than the genetic correlation as information from the correlated trait is used to estimate the EBV. The stronger the EBV correlation, the more difficult it is to improve both traits in the desired direction, i.e. into the top-right quadrant in Figure 1. Due to the limited information about trait 2, the selection index seeks an optimal solution by improving the trait with more information and with a higher economic weight. Changing the economic weight on trait 2 has limited impact on the optimal response and the weight for traits 2 would have to double at least to avoid a decline in trait 2. Improving the accuracy of the EBV of trait 2 results in a response such that trait 2 is not declining (Figure 1, right) and the EBV correlation is less strong (-0.61). Changing the economic weight will have a lot more effect in the desired direction of genetic change. It was shown that the point where $r_1.\sigma_{a1}.v_1 = r_2.\sigma_{a2}.v_2$ is where the index solution is most sensitive to economic weights, but it is also where both traits are improved in the desirable direction if EBV accuracies were close to 1 (Van der Werf *et al.*, 2018). The expression $r_i.\sigma_{ai}.v_i$ represents the dollar value of one standard deviation of the EBV of trait i ($v_{\sigma EBVi}$), and

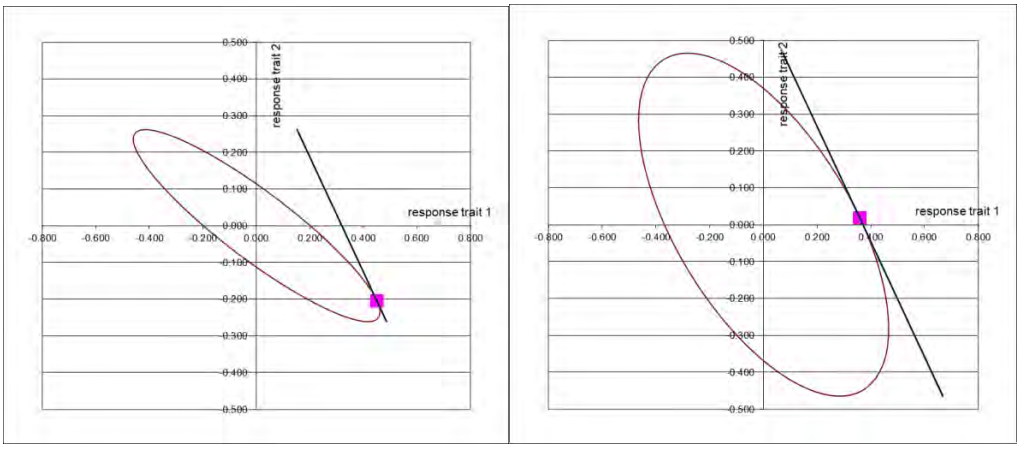


Figure 1. Index responses for two traits (square mark) when the economic values are 0.625 and 0.4 for traits 1 and 2, the accuracy of the EBV for trait 1 is 0.84 and for trait 2 it is 0.48 (left) or 0.84 (right).

these values are 0.23 and 0.10 for traits 1 and 2 in the left pane and 0.29 and 0.19 in the right pane. The expression $r_i.\sigma_{ai}.v_i$ can be used to predict the outcome of increasing the accuracy of the EBV of traits. The most sensitive point is when $r_2 = r_1.(v_{\sigma a1}/v_{\sigma a2})$. Hence, if the economic value per genetic standard deviation ($v_{\sigma ai}$) differs substantially between the traits, increasing the accuracy of traits has limited effect on trait responses to selection. However, if these values are similar, the optimal response more likely improves both traits if the accuracy of the EBVs are more similar. The same holds for working out index weights for selection: $v_1 = v_2\ (r_2\sigma_{a2} / r_1\sigma_{a2})$, i.e. when there is a large difference in EBV accuracy between the traits, trait responses will not be sensitive to changes in economic value, unless the least accurate trait has a much higher economic value per genetic standard deviation.

In conclusion, in the two-trait example, we can compare the economic value of one standard deviation of the EBVs for the traits, and if these differ a lot, we find that selection is dominated by the response for one trait. If the two traits have an unfavourably correlation, i.e. the correlation between EBVs has an opposite sign as the product of the economic values, it is likely that the response for the less dominant trait is not in the desired direction.

Extension to many traits

In a breeding objective with more than two traits it is not possible to rely on graphical representation to evaluate the sensitivity of the response to manipulating economic weights. Instead, we can look at one trait at a time, and look whether this is dominated by the other traits. These other traits can be summarised in a subindex $I_r = v_r'\hat{a}_r$, where v_r is the vector with economic values for the traits in the subindex, $\hat{a}_r$ is the vector of EBVs for these traits, and the genetic variance of this aggregate trait is equal to $\sigma^2_{Hr} = v_r'var(a_r)v_r$. The variance of the subindex is $\sigma^2_{Ir} = v_r'var(\hat{a}_r)v_r$ and the accuracy of the subindex is $r_1 = \sigma_{Ir} / \sigma_{Hr}$. Therefore, we can collapse the subindex into 'trait 1' with genetic standard deviation $\sigma_{a1} = \sigma^2_{Hr}$ and the economic value is $v_1 = 1$ as the index is in dollar values. The trait to be compared with the subindex, i.e. the breeding objective trait not included in the subindex, can be termed 'trait 2' with genetic standard deviation σ_{a2}, accuracy r_2 and economic values v_2. The correlation between the EBV for trait 2 and the subindex I_r ('trait 1') can be calculated using the covariance matrix between EBVs and v_r. If the sign of this correlation is not equal to the sign of v_2, then the trait under consideration is unfavourably correlated with the aggregate genotype of the other breeding objective traits. As in the two-trait example, we can now compare $r_1.\sigma_{a1}.v_1$ and $r_2.\sigma_{a2}.v_2$. If then also the economic value of the EBV standard deviation of the considered trait is much smaller than the standard deviation of the sub index, the considered trait is likely selected in the undesirable direction and either the accuracy or the index weight have to be increased to avoid that. In the case of many traits, the economic value of the subindex will be often much higher than that of an individual trait, so the correlation between the trait and the subindex will be critical in most cases.

Discussion

We have shown that the key parameters that drive response for each trait in multiple trait selection are its accuracy and the relative economic value of one genetic standard deviation of each trait. The product of these parameters, for each trait, determines the relative dominance of a trait in a multiple trait selection index. If a trait is not economically dominant, it is likely to see an unfavourable response when its correlation with the other traits is unfavourable. Whether this can be easily adjusted, either by increasing the accuracy of the trait EBV, or its weight, can be determined by these parameters.

Although selection index is a logical and straightforward framework to optimise selection on multiple traits, their proposed response is not always desirable. As a result, desired gains indexes are commonly used in practice where predicted index solutions based on derived economic weights are unsatisfactory. These can be simply derived mathematically, based on the theory presented. This is due to several factors. Individual

choices of breeders may differ from a formal economic model due to different market expectations, real or perceived competitive positions in the market. Moreover, the breeding objective may not be linear with more value given to a more balanced trait improvement. In fact, there is a well-known dispute between economists and experts in decision theory, the latter noting that negative outcomes are weighted about twice as much as positive outcomes (Kahneman, 2011). This principle could also be applied in index selection, by giving more negative weight to undesirable trait responses.

It is noted that in practice selection decisions are more dynamic, and made across candidates from different age classes. Therefore, selection accuracies differ among selection candidates. Moreover, genetic parameters will change after selection as a result of the Bulmer effect and selection decision are based not just on EBV but also co-ancestry and possibly other practal constraints. Tactical selection decisions can be optimised by decision tools such as MateSel (see Brian Kinghorn's paper at this conference). However, a fundamental understanding of the parameters driving multiple trait selection response is important and the theory discussed in this paper maybe useful to design strategies in breeding programs, in particular in relation to trait measurement.

References

Dekkers, J.C.M. and J.H.J. van der Werf (2014). Proc. of the 10th WCGALP, Vancouver, Canada.

Hazel, L.N. (1943) Genetics 28: 476-490.

Kahneman, D. (2011) Thinking fast and slow. New York: Farrar, Straus and Giroux.

Moav, R. and Hill, W.G. (1966) Anim. Prod. 8: 375-390.

Van der Werf, J.H.J., R.G. Banks and J.C.M Dekkers (2018) Proc. of the 11th WCGALP, Auckland, New Zealand.

476. Selection for birth weight environmental variability in mice as a model to improve animal welfare in livestock species

J.P. Gutiérrez[1*], N. Formoso-Rafferty[2], C. Ojeda-Marín[1], L. El-Ouazizi El-Kahia[1], K.D. Arias[3] and I. Cervantes[1]

[1]Departamento de Producción Animal, Facultad de Veterinaria, Universidad Complutense de Madrid, Avda. Puerta de Hierro s/n, 28040 Madrid, Spain; [2]Departamento de Producción Agraria, E.T.S. Ingeniería Agronómica, Alimentaria y de Biosistemas, Universidad Politécnica de Madrid, 28040 Madrid, Spain; [3]SERIDA-Deva, Camino de Rioseco 1225, 33394 Gijón, Spain; gutgar@ucm.es

Abstract

Several results of a divergent selection experiment for environmental variability of birth weight in mice are presented. Genetic control of environmental variability of birth weight is shown to be possible through selection. It is also shown that selection for homogeneity improved litter size, embryo and fetal survival, but also at birth, at weaning and at reproductive stage. It also improved fertility, reproductive longevity, and the response to feed restriction. Homogeneous animals were individually smaller, but females weaned 42% more total weight than heterogeneous ones. Signs of robustness and welfare were identified related to uniformity, suggesting this kind of selection would be very useful in livestock species.

Introduction

The main objective of animal breeding programs is to increase productivity. However, the focus is shifting to other types of traits. One of these types is trait homogeneity, where selection is for reduced trait variability. Homogeneity has been associated with a reduction in handling and production costs for the animal production industry, thus ensuring homogeneous batches of animals (Hennessy, 2005). It has also been associated with a decrease in mortality in prolific species and has also been related to robustness, improving functional traits. Finally, homogeneity has also been related to animals resistant to environmental challenges, also as a sign of robustness. Minimizing the residual variance is usually the selection criterion to increase homogeneity. In this work, results of a divergent selection experiment for environmental variability of birth weight in mice are presented. It was analysed if genetic control of environmental variability of birth weight was possible through selection and if it would affect other traits of interest.

Materials & methods

A divergent selection experiment was performed based on breeding values for environmental variability of birth weight in mice. Two divergent lines were created: High variability Line (H-line) and Low variability (L- line). These breeding values were predicted using a heteroscedastic model (SanCristobal *et al.*, 1998, Formoso-Rafferty *et al.*, 2016a). Birth weights were recorded for litters from 43 females per line. Each female produced one or two litters and each female was sired by a different male. Four males and four females from the best 11 females were selected as parents for the next generation, and selection was run for 29 generations. A total of 33,245 birth weight records of 3,687 litters after 29 generations and a pedigree of 32,885 animals were used. Genetic and phenotypic trends were analysed by drawing the average within-litter variance by line and generation. The heteroscedastic model included sex (male/female/unknown), parity number (2), litter size and generation as fixed effects and maternal genetic, litter and residual as random effects affecting the mean birth weight and also its environmental variability.

Genetic control of environmental variability of birth weight is possible through selection and also it produces changes in other traits. Then, in this work we include: (1) correlated trends in weights and

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_476, © J.P. Gutiérrez *et al.* 2022

litter size; (2) feed and reproductive efficiency; (3) reproductive longevity; (4) differences in heritabilities between lines; and (5) ovulation rate and embryo and fetal survival.

1. Phenotypic and/or genetic correlated trends were also analyzed for birth weight trait itself, weaning weight, total litter birth and weaning weights, litter size, survival at weaning and variability of both weaning weight and litter size (Formoso-Rafferty *et al.*, 2016b).
2. Animals from the H-line and L-line were used to study feed and reproductive efficiency. Records from a total of 271 females of five generations (10 to 14) were used. From 21 to 56 days, weekly body weight, weight gain, feed intake and cumulated transformation index were analyzed using a statistical model including line and generation as effects. Number of performed parturitions was also compared via a Chi-square test (Formoso-Rafferty *et al.*, 2018). For each line, 120 weaned females were used, that is, four complete sisters from each of 10 random litters from three consecutive generations of selection. Four groups were made within each generation according to combinations of restricted feeding or *ad libitum*, in two different periods, during growth and in reproductive stage. In each generation (12, 13 and 14), different levels of feed restriction were applied at 75, 90 and 85% (Formoso-Rafferty *et al.*, 2019).
3. To compare the reproductive longevity between selected lines, two generations of an initial number of 43 females and 43 males mated one to one per line, remained together to obtain consecutive parturitions. Females were discarded when the time elapsed since the last parturition was greater than 63 days. The time to the end of the reproductive period between both lines was compared by fitting a Cox proportional hazard regression model, and the farrowing rate in both lines was also compared using a Prentice-Willians-Peterson model (Formoso-Rafferty *et al.*, 2022).
4. As residual variance was not unique, estimations of heritabilities differed among levels of systematic effects, and among lines and generations. Data from the first eleven generations were used, totalling 15,431 birth weight records from 959 females and 1,641 litters in combination with 14,786 pedigree records. Estimates of the residual variance across levels of effects, keeping constant the additive genetic variance allowed to estimate different heritabilities (Formoso-Rafferty *et al.*, 2017).
5. Last studies in the mice divergent selection lines compared the fetal survival between 14 days of gestation and at birth, and the embryo survival from fecundation to 14 days assessing also the ovulation rate in both lines. A total of 138 H-line and 240 L-line ultrasound scans were performed at day 14 of the first gestation in order to address fetal mortality as the difference between number of fetuses in the scan and litter size at birth. Also, in order to determinate ovulation rate and embryo survival, a total of 48 H-line and 50 L-line females were sacrificed in their third gestation at day 14 and counted the corpora lutea presented in both ovaries (Formoso-Rafferty *et al.*, 2021).

Results

Average within-litter variance was diverging with variable success across generations and lines. In the last generation the variances were 0.0536 and 0.0096 g^2 for the H-line and L-line. This difference was also observed in the variance across litters being 0.0792 and 0.0149 g^2.

A genetic correlation between the birth weight and its environmental variability was consequently estimated to be 0.30 which led also to divergence in the within litter mean birth weight that reached averages of 1.87 g and 1.33 g for H-line and L-line. Divergence was also observed in the mean weaning weight that was 12.64 g and 8.53 g for H-line and L-line. Total litter weight was also higher in the H-line with an average of 83.25 g in contrast to 69.09 g in the L-line. However, accounting for differences between lines in litter size, survival at weaning and fertility, L-line weaned a total of 5,458 vs 3,829 g in the H-line. Litter size at birth also differed across generations between lines being in the last generation 7.54 and 9.02 pups, with a respective survival of 89 and 92% in the H-line and L-line. In this last generation, only 32 and 42 out of 43

females performed parturitions in H-line and L-line. Fourteen and 37 of the females reached two parities in H-line and L-line.

Regarding feed efficiency, results showed that accumulated transformation index was similar in both lines. However, concerning other traits, for the H-line and L-line, the number of parities was 1.464 vs 1.600, the litter size at the first parturition was 8.76 vs 9.98, the litter size at weaning was 7.61 vs 9.05, and the number of pups stillborn was 0.25 vs 0.11. Similar performance differences were observed at the second parturition, being all highly significant ($P<0.001$). When animals were fed under restriction, fertility was 33% in H-line and 72% in the L-line, survival at weaning was 65% in the H-line and 95% in the L-line, and survival at the reproductive stage was 60% in the H-line and 85% in the L-line, differing significantly between lines, but were not different in those animals fed *ad libitum*. Conversely, diet groups were found significantly different in the H-line but not different in the L-line. In the H-line, *ad libitum* animals survived 100% at weaning and 95% after reproductive stage and 100% performed parturition whilst restricted animals survived 65% at weaning and 60% after reproductive stage and only 33% performed parturition.

The reproductive longevity study revealed that the L-line was associated with almost three times higher delivery rate (2.93) with regard to the H-line. The Cox model showed that homogeneous females also presented 26% of the risk of stopping their reproductive life. The median reproductive days was 55.5 in the H-line, while the median was 252.5 days for the L-line females.

The birth weight heritability was different between lines and also across systematic effects levels. It was higher in L-line across generations (0.06 higher in eleventh generation). The heritability for birth weight ranged from 0.04 to 0.22 across different systematic effect levels. Using this findings, it could be possible to modulate the heritability of this trait by selecting to decrease residual variability.

Regarding reproductive analysis, there were no significant differences in the number of fetuses observed at 14 days between lines, but litter size at birth was significant higher in L-line (9.82 vs 8.36 pups). In the embryo survival analysis, there were no differences in the number of fetuses at 14 days. However, significant differences were found (H-line vs L-line) in: number of hydatidiform moles of the uterus (3.96 vs 1.27), corpora lutea (13.88 vs 11.53), embryo survival (0.53 vs 0.79) and ovules lost from corpora lutea to fetuses at 14 days (6.38 vs 2.59). Then, the higher litter size in L-line showed important advantages in terms of higher embryo and fetal survival.

Discussion

The genetic selection of environmental variability of birth weight was possible in mice. Clear divergent phenotypic trends were observed between lines for litter size, weaning weight and weaning weight standard deviation. This selection criterion directly affected other traits in such as litter size, total weaning weight, fertility and pup survival. The experiment led to two divergent lines with the L-line presenting benefits in production, animal welfare, heritability and traits usually considered indicators of robustness such as feed efficiency or reproductive longevity. This better performance of the L-line was achieved without extra energetic cost. Moreover, despite the selection process was designed and carried out identically, the selection response was asymmetric in favor the low variability line when compared with a control (Formoso-Rafferty *et al.*, 2020), therefore increasing BW homogeneity could be achieved in a few generations. Trends in predicted breeding values diverged according to the selection made, and were in the same direction for birth weight trait, increasing both in the H-line and decreasing in the L-line. In this sense and to study the genetic correlation between the trait and its variability, a simulation study showed that a scale effect (in the sense that the higher the mean of a variable, the higher its variability) would explain many genetic

correlations reported in the bibliography, and specifically that one found in the mice experiment (Tatliyer *et al.*, 2019). This fact should be take into account when applying this type of selection.

The animals of the L-line became homogeneous although smaller due to the genetic correlation between the trait and its variability. However, the selected homogeneity has also been accompanied by signs of greater robustness, with larger litter size, better survival and better reproductive performance without losses in total weaned weight. This contrasts with what has been observed in pigs, where there has been an increase in mortality as a consequence of the selection aimed at increasing litter size, which is linked to a greater heterogeneity of piglet weights at birth (Damgaard *et al.*, 2003). Therefore, it would be interesting to investigate whether selection to reduce birth weight variability would have the same effects in livestock species.

Acknowledgements

This work has been supported by research grants from Ministry of Science, Innovation and Universities (PGC2018-096198-A-I00, RTA2014-00015-C02-02 and AGL2008-00794).

References

Damgaard, L.H., Rydhmer, L., Lovendahl, P., Grandinson, K. 2003. J. Anim. Sci., 81: 604-610. https://doi.org/10.2527/2003.813604x

Formoso-Rafferty N., Cervantes I., Ibáñez-Escriche N., Gutiérrez, J.P. (2016a) J. Anim. Breed. Genet., 133:227-237. https://doi.org/10.1111/jbg.12174

Formoso-Rafferty, N., Cervantes, I., Ibáñez-Escriche, N., Gutiérrez, J.P. (2016b) Animal, 10(11):1770-1777. https://doi.org/10.1017/S1751731116000860

Formoso-Rafferty, N., Cervantes, I., Ibáñez-Escriche, N., Gutiérrez, J.P. (2017) J. Anim. Sci., 95:531-537. https://doi.org/10.2527/jas.2016.1169

Formoso-Rafferty, N., De La Flor, M., Gutiérrez, J.P., Cervantes, I. (2018) J. Anim. Breed. Genet., 135:378-389. https://doi.org/10.1111/jbg.12345

Formoso-Rafferty, N., Cervantes, I., Sánchez, J.P., Gutiérrez, J.P., Bodin, L. (2019) Genet. Sel. Evol., 51:27. https://doi.org/10.1186/s12711-019-0471-9

Formoso-Rafferty, N., Chávez, K.N., Ojeda, C., Cervantes, I., Gutiérrez, J.P. (2020) Animals, 10:920. https://doi.org/10.3390/ani10060920

Formoso-Rafferty, N., Gutiérrez, J.P., Cervantes, I. (2021) Proc. of 72th EAAP, Davos, Switzerland.

Formoso-Rafferty, N. Gutiérrez, J.P., García-Álvarez, A., Pérez, T., Cervantes, I. (2022). J. Anim. Breed. Genet., submitted.

SanCristobal-Gaudy M., Elsen J.M., Bodin L., Chevalet C. (1998) Genet. Sel. Evol., 30:423–51. https://doi.org/10.1186/1297-9686-30-5-423

Hennessy, D. A. 2005. Am. J. Agric. Econ., 87(3): 600–609. https://doi.org/10.1111/j.1467-8276.2005.00750.x

Tatliyer, A., Cervantes, I., Formoso-Rafferty, N., Gutiérrez, J.P. (2019). G3: Genes Genomes Genet., 9:3001-3008. https://doi.org/10.1534/g3.119.400497

477. Sensitivity Analysis for a new Economic Selection Index for Chilean dairy cattle

J.L. Chuma-Alvarez[1], H.H. Montaldo[1*], D.A.L. Lourenco[2], F.J. Ruiz-López[3], M.E. Olivares[4] and C. Lizana[4]

[1]Facultad de Medicina Veterinaria y Zootecnia, Universidad Nacional Autónoma de México, Ciudad Universitaria, 04510, Mexico City, Mexico; [2]College of Agricultural and Environmental Sciences, University of Georgia, 425 River Road, Athens, GA 30602, USA; [3]Centro Nacional de Investigación Disciplinaria en Fisiología y Mejoramiento Animal, Instituto Nacional de Investigaciones Forestales, Agrícolas y Pecuarias, Secretaría de Agricultura y Desarrollo Rural, Ajuchitlán Colón, 76280, Querétaro, Mexico; [4]Cooperativa Agrícola y de Servicios, COOPRINSEM, Manuel Rodríguez 1040, 5290000, Osorno, Región de Los Lagos, Chile; montaldo@unam.mx

Abstract

Contribution of production, functional and feed efficiency traits to total economic response (Δ$) of a new economic index VEL+ for Holsteins in Chile were estimated using the selection index method with local and U.S. economic and genetic parameters. Contribution of the traits and trait groups on Δ$ was estimated as the difference for each evaluated index compared to the base index. We modelled a situation of genomic U.S. sire selection, adjusting the selection accuracy of each trait using estimates of genetic correlations between countries. Contribution of the production, functional, and feed efficiency traits group were 25.9, 38.5 and 4.6%, respectively. When reducing or increasing relative economic values of each trait by multiplying them by a range of values ranging from 0 to 1.5 in the index, protein yield (0.4 to 2.6%), fat yield (0.3 to 2.4%), and longevity (0.2 to 1.9%) were the traits with the greatest contributions to Δ$.

Introduction

Economic selection indices must be updated periodically to include new traits and reflect changing economic conditions and marketing systems, as well as changing genetic parameters among traits (Cole *et al.,* 2021). One approach to estimating these contributions is to assess the effect of modifying the relative economic values (REV) of each trait or group of traits on the total economic response of a selection index (Pryce *et al.,* 2009). This sensitivity analysis allows for an assessment of the robustness of indices against inaccurate REV. This is particularly important in situations where key information for some potentially important traits is not locally available.

This new index was based on a previous index for dairy production and functional traits in Chile. We evaluated the addition of two traits used in the United States to assess feed efficiency (VanRaden *et al.,* 2021) that are of interest to the Chilean dairy sector.

The objective of this work was to conduct a sensitivity analysis by determining the contribution of each trait and group of traits to total economic response (Δ$), changing the REV for each trait in a new economic index VEL+ (economic dairy value) developed in Chile for Holstein cattle.

Materials & methods

VEL+ economic index. An economic index (VEL) was developed based on data for a Holstein herd in southern Chile with a production system based on grazing with supplementation, with prices, milk payment systems, and productive parameters from 2019, which represent an average situation for this region (Cooperativa Agrícola y de Servicios, 2020). Traits considered in the VEL index were milk yield (MY), fat yield (FY), protein yield (PY), longevity (LONG), somatic cell score (SCS), calving interval (CI),

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_477, © J.L. Chuma-Alvarez *et al.* 2022

and age at first calving (AFC). The economic values (Table 1) of the traits in the VEL index were estimated as the first partial derivatives of the economic benefit function with respect to each trait, as a way to calculate profit before and after the genetic change of one unit in a given trait (Weller, 1994).

To form the VEL+ index, body weight composite (BWC) and residual feed intake (RFI) from the U.S. CM$ (Cheese Merit) index were incorporated into the VEL index. The relative economic values (REV) of each trait in the VEL and the VEL+ indices were expressed by taking the economic value of one kg of protein yield as reference (Table 1).

Calculation of economic response. Economic response to selection (Δ$) of the traits included in the VEL+ index was calculated using selection index theory (Weller, 1994) with the Multiple Trait/desired gains spreadsheet developed by Van der Werf (2014) that calculate responses for multitrait selection in animal populations. The number of progeny equivalent to U.S. genomic evaluations were obtained by trait (Council on Dairy Cattle Breeding, 2021) and adjusted to Chile based on estimates of genetic correlations (less than 1) between the two countries obtained from genetic evaluations of bulls (Cooperativa Agrícola y de Servicios, 2020). Heritabilities and phenotypic variances of VEL index traits as well as genetic and phenotypic correlations were estimated using the COOPRINSEM historical archive for the population of Holstein cows from southern Chile that comprised the period 1998-2015. A range of 253,015 to 873,840 records were analysed. Genetic parameters and relative economic values for BWC and RFI for the U.S. population (Li et al., 2020; VanRaden, 2021) were adapted to the Chilean population.

Sensitivity analysis and experimental design. Δ$ was estimated as the sum of the products of trait selection response and corresponding economic values. Sensitivity analysis was made by multiplying the base REV of each trait or group of traits by a range of values from 0 to 1.5. Estimation of the contribution (%) of a given trait or group of traits with a specific modified REV was based on the difference between the Δ$ obtained with the modified REV and the Δ$ for the base index, and then expressed in percentage. This approach allows to estimate the robustness of the index against changes in the REV actually used when departing from the reference in the base index. Three groups of traits were formed, grouping MY, FY and PY as production traits, LONG, SCS, CI and AFC as functional traits, and BWC and RFI as feed efficiency traits.

Table 1. Economic values and economic response to selection (Δ$) by trait for base VEL+ index.

Trait		EV[1]	REV[2]	Δ$ (Chilean Pesos)	Contribution (%)
Milk yield (kg)	MY	-30	-0.01	-508	-0.66
Fat yield (kg)	FY	1,914	0.64	19,166	24.89
Protein yield (kg)	PY	2,981	1.00	8,645	11.23
Longevity (months)	LONG	15,818	5.31	29,194	37.92
SCS (log2)	SCS	-45,949	-15.41	4,952	6.43
Calving interval (days)	CI	-2,362	-0.79	4,536	5.89
Age at first calving (days)	AFC	-174	-0.06	1,053	1.37
Body weight composite (units)	BWC	-11,659	-3.91	8,468	11.00
Residual feed intake (kg)	RFI	-171	-0.06	1,482	1.93
Total				76,991	100.00

[1] Economic values in Chilean pesos.

[2] Relative economic values to 1 kg of protein yield.

Results

Table 1 lists the 9 traits that comprise the base VEL+ index, the units in which they are measured, the economic value of each unit, relative economic value, the economic response to selection (Δ$), and the contribution (%) of each trait.

The base VEL+ index (76,991 Chilean pesos) showed more Δ$ (10.8%) than base VEL (69,464 Chilean pesos), due to the inclusion of BWC and RFI. In the base VEL+ index the most important traits in terms of percent contribution were LONG, FY, PY and BWC; the lowest were AFC and RFI (Table 1).

When setting the REV of the given group of traits to zero in the index, the greatest contribution to Δ$ was for functional traits (38.5%), followed by production traits (25.9%), and feed efficiency traits (4.6%).

In the sensitivity analysis, the contributions of each trait to Δ$ with modified REV with a range of values (0.5-1.5) are shown in Figure 1. The traits that most contributed to Δ$ were PY (0.4 to 2.6%), FY (0.3 to 2.4%) and LONG (0.2 to 1.9%), and the trait with the lowest contributions was AFC (all <0.01%) across the range of levels. Other traits had intermediate contributions (Figure 1). Changing REV by 20% above (1.20) and below (0.80) the base REV caused very similar changes within each trait (Figure 1).

Discussion

In our analysis, exclusion of functional traits from the index would cause the greatest reduction in the index's Δ$ for genomic selection of Holstein bulls in Chile, followed by production and feed efficiency traits. With respect to individual trait sensitivity analysis, PY, FY and LONG represented the greatest contribution to Δ$, and their changes in REV had the greatest economic effects, unlike the three Australian indices evaluated by Pryce *et al.* (2009), where production traits contributed most to the total economic response, despite the fact that fertility and longevity had greater relative weights in these indices than in our indices. This may be because in our data fertility is more unfavourably genetically correlated with production.

In conclusion, the incorporation of BWC and RFI as feed efficiency traits in genomically evaluated bulls increase the expected economic response to selection, although the contribution of RFI is small due to the low accuracy of current genomic evaluations.

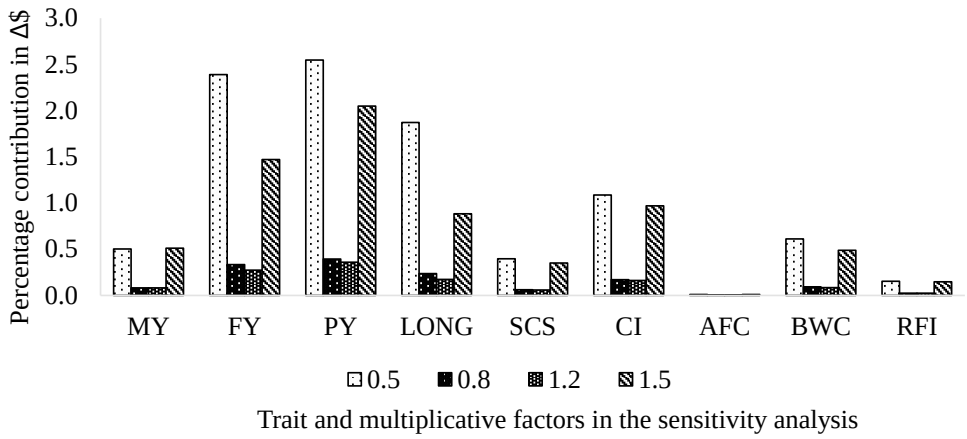


Figure 1. Percent contribution to total economic response (Δ$) for the VEL+ index, by changing the relative economic values of the trait by a multiplicative factor equal to 0.5, 0.8, 1.2, or 1.5.

Emphasis should be placed on obtaining precise REV for all traits, but the impact of inaccurate REV would be greater for PY, FY and LONG. The influence of uncertainty of REV values in BWC and RFI selection is likely to be small.

Maximum efficiency could be obtained by incorporating new traits and forming a new economic index that improves the feed efficiency of cows (VanRaden *et al.*, 2021), which would be advantageous to local dairy producers and those financing the breeding programs.

References

Council on Dairy Cattle Breeding (CDCB) (2021). Comparison of genomic and traditional evaluations. Available at: https://queries.uscdcb.com/eval/summary/comparexml_menu.cfm?R_menu=v_2108.v_Young_Bulls.v_Holstein_wddx#StartBody

Cooperativa Agrícola y de Servicios (COOPRINSEM) (2020). Valor Económico Lechero: Índice económico para la selección de bovinos lecheros en Chile. Available at: http://cooprinforma.cl/indice-economico-para-la-seleccion-de-bovinos-lecheros-en-chile/

Cole J.B., Dürr J.W., and Nicolazzi E.L. (2021) J. Dairy Sci. 104(5):5111–5124. https://doi.org/10.3168/jds.2020-19777

Li B., VanRaden P.M., Guduk E., O'Connell J.R., Null D.J., *et al.* (2020) J. Dairy Sci. 103(3): 2477-2486. https://doi.org/10.3168/jds.2019-17332

Pryce J.E., Van der Werf J.H.J., Haile-Mariam M., Malcolm B., and Goddard M.E. (2009) Anim. Breed. Genet. 18:143-146.

Van der Werf J. (2014). Multiple Trait/desired gains. MT selection index with desired gains option (10 traits max). Available at: https://jvanderw.une.edu.au/software.htm

VanRaden P.M., Cole J.B., Neupane M., Toghiani S., Gaddis K.L, *et al.* (2021). Net merit as a measure of lifetime profit: 2021 revision. AIP- USDA research report.

Weller JI. (1994). Economic aspects of animal breeding. Springer Science and Business Media. London, UK.

478. Opportunities for animal genetics to foster the ecological transition of livestock farming systems

D. Savietto*, F. Douhard, H. Gilbert, R. Rupp, L. Fortun-Lamothe and A. Ducos

GenPhySE, INRAE, Université de Toulouse, ENVT, 31326 Castanet Tolosan, France; davi.savietto@inrae.fr

Abstract

Industrial livestock systems faces a major crisis of legitimacy and animal breeding contributes to this situation. Worldwide, breeding goals focus on economic performance and rarely consider the environmental and social impacts. Agroecology offers a conceptual framework to think alternative solutions to face this crisis and to guide the transition of livestock farming towards nature-based systems. Animal genetics, as an essential component of animal husbandry, should contribute to this evolution. Based on the agroecological principles applied to animal farming and the efficiency-substitution-redesign (ESR) framework, we showed that animal breeding mainly focused on the E-S levels of the ESR framework and few on the redesign of animal farming. We then present some perspectives for the redesign of breeding goals, the use of the G×E interaction information in the evaluation of animals, and the benefits of managing the available genetic diversity of livestock species to fit to the needs of nature-inclusive systems.

Introduction

Over the brief period of the 20th century, agriculture in industrialised countries changed from small and diversified farms to big and specialized enterprises (Horrigan *et al.*, 2002). This change has been driven by the specialization of production (single animal species), the mechanization of the labour (e.g. milking machines) and the introduction of chemical (disinfectants) and pharmaceutical (antibiotics) inputs to assure production in a controlled environment. This industrialization process favoured the emergence of highly productive systems, designed to provide large amounts of low price food for consumers (Carolan, 2018).

Industrial farming, however, faces a major legitimacy crisis. The use of antibiotics in animal husbandry contributes to the emergence of antimicrobial resistant bacteria, the concentration of animals in a single place or region scales up environmental issues, while citizens (at least in Europe) increasingly reject intensive animal production and plead for the end of the cage age (https://www.endthecageage.eu/). In the face of this crisis, the livestock production must evolve towards systems that are more acceptable. In this context, nature-based systems, founded on the principles of agroecology (Altieri, 1989; Gliessman, 2015) need to be developed. Dumont *et al.* (2013) proposed five agroecological principles to initiate the transition of sustainable animal production systems by: (1) promoting integrated management of animal health; (2) reducing inputs needed for production; (3) decreasing pollution; (4) taking advantage of system diversity to strengthen their resilience; and (5) preserving biological diversity. So far, the contribution of animal breeding to those principles is not clear yet, whereas it might represent a key lever to foster the evolution in thinking livestock systems and their technics.

Ducos *et al.* (2021) reviewed the contribution of animal genetics to each agroecological principles. Through its contribution to increase the production performance in breeds reared in highly specialized systems, developments in animal breeding have also embraced the first three principles whereas the last two principles related to diversity to a large extent were neglected. In addition, most works to date have focused on a single principle and thus rarely addressed the necessary complementarity among principles.

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_478, © D. Savietto *et al.* 2022

This lack of complementarity lead us to classify most contributions of animal breeding to agroecology as 'shallow ecological contributions', as described in the efficiency-substitution-redesign (E-S-R) framework (Hill, 1985). For Hill (1985), the industrialization process tends to optimize every single element of the system in a search for more efficient (E) practices. Elements of the system may be substituted (S) by some alternative if their use is no longer efficient, controversial, (e.g. soybean linked to deforestation) or threatens human health (e.g. antibiotics). Hill (1985) argued that solutions based on the E-S approaches 'usually protect and perpetuate the causes of the problem', and pled for the deployment of preventive ecological-based solutions, which will require a deep redesign (R) of systems.

Here, based on the E-S-R framework (Hill, 1985) and on some existent alternative production models, we propose several perspectives for animal genetics to contribute to the ecological transition (the Redesign) of livestock production systems.

The efficiency-substitution-redesign (E-S-R) framework: a new paradigm for animal breeding

Animal genetics have mainly contributed to the E-S levels of the E-S-R framework/gradient (Table 1, adapted from Ducos *et al.*, 2021). For example, the selection for feed efficiency contributed to reduce the amount of cereal-based diets per unit of product (E). It also helped to improve the capacity of non-ruminant animals to grow using fibrous diets to reduce the high costs of cereal-based diets (S). On the other hand, selection of animals for integrated crop-livestock systems, or fed with food waste (from warehouses, restaurants, etc.), or capable of recycling food resources not edible (or not desirable) for human consumptions, are still scarce. Other domains of applied animal breeding, aiming at improving the health or adaptive capacity of animals, for instance, also focused on curative or shallow ecological solutions. Most advances in these domains concern the search for disease-resistant (or tolerant), long-living or robust individuals. Although desirable, the inclusion of these functional traits in the selection indexes still focus on the animal component of the system, leaving the capacity of animals to cope with their environment (i.e. the G×E interaction) unexplored. Indeed, selection populations are generally raised in controlled conditions set to avoid the G×E. Even when the G×E information is available the *E-S* dynamic prevails. As an example in dairy cattle, Interbull structures its evaluations to reduce the G×E across countries, creating a sort of standardization on the ranking of dairy bulls, in the search for the international – *universal* – animal (Chavinskaia *et al.*, 2017). This sort of normalization is not desirable for nature-based systems that usually face a broad range of constraints. Organic (a proxy of agroecology) dairy farmers seems to contribute to the development of agroecological systems through the adoption of 'home based' breeding practices such as the use local breeds, kin breeding, and crossbreeds to produce (Nauta *et al.*, 2005). Some organic farmers use high producing breeds, while the majority of them question the adequacy of specialized milk-type breeds to the organic system. For such farmers, the desired animal should 'take time to mature', 'yield less milk' (mentioned as less adapted to a low concentrate and high grass regime), 'live longer', and have a better 'integrity' and 'vitality' (Nauta *et al.*, 2005; Bieber *et al.*, 2019). This example of organic dairy cattle shed some light on which traits animal breeding may focus on to fit the needs of agroecological systems. This notably include functional traits (fertility, vitality, disease resistance-tolerance, functional longevity) and/or life history traits (age and size at maturity, growth pattern, number and frequency of reproductive events in the lifetime, etc.). The use of data from farms (companies, cooperatives and independent farms), the development of participatory selection programs and the redesign of the academic selection facilities and programs (at national and international level) may all contribute to breed animals capable of live and produce (animal products and ecological services) in a wide range of livestock production systems. This strategy may favour the development of tools and mathematical methods to quantify the potential interest of G×E interactions and promote their use. Other strategies may help to identify adapted genotypes for nature-based systems.

Table 1. Classification of present (and desirable) breeding goals in the efficiency, substitution, redesign framework of Hill (1985). Adapted from Ducos *et al.* (2021).

Domain	Transition gradient (E >> S >> R)		
	Efficiency (E)	**Substitution (S)**	**Redesign (R)**
Feed efficiency	Better efficiency in use of feed concentrates	Improved ability to valorize alternative feed resources	Search for individuals and populations adapted to systems using a wide range of feed resources (raw products, food waste, etc.)
Animal health	Sick and morbid individuals are less productive and not selected	Selection of diseases resistant/tolerant, robust individuals	Breeding within and between populations of animals living in the same ecological habitat. Favor traits related to the use of space and resources for health
Adaptability	Selection for milk persistence or longevity reduce the number of animals needed for production	Animals capable of using their body reserves to sustain (re)productive and immune functions when facing underfeeding or infection	Selection of multiple-traits at both individual and herd levels: able to digest a wide array of resources, docile towards humans, yet alert to predators, body reserves to fuel (re) production and as a safety deposit), immune function co-evolving with the pathogens present in the system, fur (plumage) adapted to the outdoor (color and density), etc.

Which genetics in a perspective of redesign animal breeding systems?

Agroecological (nature-based) systems combine crops and animals, facilitate nutrient cycling and the system's resilience as well as the preservation of biodiversity. A diversity of agroecological systems may emerge, each one combining different species and genotypes at a given place and time. In this aspect, one of the main challenges of animal breeding is to select farm animals economically valuable and helpful (interesting) for the system as a whole, providing ecological and social services to their abiotic and biotic environment. For this, we need to redesign breeding goals, and develop tools for the evaluation and use of the G×E information.

For animal genetics, the key issue relies in the optimization of the breeding goals. Worldwide, breeding goals mostly aim to increase profit by increasing production or reducing costs (feed, animal health) and rarely consider environmental and social impacts. New traits (in the *E-S* level), even those related to resilience, are usually evaluated under the industrialized systems or with limited reference to G×E and long-term climate change. In order to help farmers and breeders to transition towards resilient/agroecological breeding, the methods and models to optimize breeding goals need to evolve. Breeding goals should include multi-criteria assessments and combine targets not only related to economics, but also to labour, greenhouse gas emission, biodiversity, land use, and food competition. In this context, the notion of an 'eco-social' breeding objective that integrates all these aspects might be worth to explore. How to combine those targets with different metrics is an issue. In fact, calculations of greenhouse gas emissions based on unit of products tend to rate intensive livestock systems since they are more efficient at the animal scale, while metrics based on area or resources use tend to favour extensive ones (Garnett, 2011). Furthermore, breeding objectives will need to consider simultaneously the evolution of breeding systems and animals (within or across breeds) toward a balanced solution on the global scale (from the farm to the overall food system).

The agroecological framework also offers the opportunity to better value the genetic diversity of livestock species. First, there is a need to unravel and characterize the genetic basis for resilience and adaptation of commercial as well as hardy and underutilized breeds, with reference to environmental breeding conditions. Then, breeding programs will require to consider a variety of options including, crossbreeding, introgression, managing diversity within farms (within and across breeds and species).

References

Altieri M.A. (1989) Agr Ecosyst Environ 27(1-4):37-46. https://doi.org/10.1016/0167-8809(89)90070-4

Bieber A., Wallenbeck A., Leiber F., Fuerst-Waltl B., Winckler C. *et al.* (2019) J Dairy Sci 102(6): 5330-5341. https://doi.org/10.3168/jds.2018-16147

Carolan M. (2018) The real cost of cheap food, Routledge, Taylor & Francis Group, New York, USA.

Chavinskaia L., Ducrocq V., and Joly. P.-B. (2017) Interbull Bulletin No. 51. Tallinn, Estonia, August 25-28. https://journal.interbull.org/index.php/ib/article/view/1428

Ducos A, Douhard F., Savietto D., Sautier M., Fillon V., *et al.* (2021) INRAE Prod Anim 34(2): 79-96. https://doi.org/10.20870/productions-animales.2021.34.2.4773

Dumont B., Fortun-Lamothe L., Jouven M., Thomas M., and Tichit M. (2013) Animal 7(6):1028-43. https://doi.org/10.1017/S1751731112002418

Garnett T. (2011). Food Policy 36; S23–S32. https://doi.org/10.1016/j.foodpol.2010.10.010

Gliessman S. (2015) Agroecology: the ecology of sustainable food systems, Third Edition. CRC Press, Taylor & Francis Group, Boca Raton, USA.

Hill S. (1985). Alternatives, 12, 32-36. https://eap.mcgill.ca/publications/eap23.htm

Horrigan L., Lawrence R.S., and Walker P. (2002) Environ Health Persp 10(5): 445-456. https://doi.org/10.1289/ehp.02110445

Nauta W.J. Groen A.F., Veerkamp R.F., Roep D., and Baars T. (2005) NJAS-Wagen J Life Sc 53(1): 19-34. https://doi.org/10.1016/S1573-5214(05)80008-9

479. Defining animal breeding objectives and strategies within the framework of responsible research and innovation

I. Olesen[1*], A. Kettunen[1], G.K. Rosendal[2] and H.M. Nielsen[3]

[1]Nofima, P.O. Box 210, 1431 Ås, Norway; [2]The Fridtjof Nansen Institute (FNI), PB 326, 1326 Lysaker, Norway; [3]Center for Quantitative Genetics and Genomics, Aarhus University, Blichers Alle 20, 8830 Tjele, Denmark; ingrid.olesen@nofima.no

Abstract

Rapid changes and innovations in food production are prompted by the increased demand of sustainably produced food, while addressing concerns of climate change and environmental impacts. To address complex societal issues, research and innovation policies, research bodies often request a framework for responsible research and innovation (RRI) to be applied for developments of new science and technologies. In addition to the individual researchers' and key actors' thinking and reflections in farm animal breeding, there is a need to look at mechanisms affecting ethical and social aspects of animal breeding at an institutional level. In this article we exemplify and discuss how breeders can define animal breeding objectives and develop technology strategies within the RRI framework to contribute more efficiently to sustainable animal production.

Introduction

Food production systems need rapid changes and adaptations to respond to the increased demand for sustainably produced food, while addressing the concerns of climate change and environmental and societal impacts. Commercial and idealistic interests push for various short and long-term objectives and strategies using new technologies in food production. We need to produce more food, more effectively, while respecting and responding to international agreements such as UN goals for sustainable development and the UNFCCC and its Paris agreement on climate change. Adopting new breeding technologies, such as gene editing, may contribute to food security. This has prompted a need for a more open dialogue between scientists, policy makers and the public. In response, EU research and innovation policies and research bodies have developed a framework for responsible research and innovation (RRI) to be applied when developing new science and technologies (EC, 2014). In Von Schomberg's (2012) definition, RRI is 'a transparent, interactive process by which societal actors and innovators become mutually responsive to each other with a view on the (ethical) acceptability, sustainability and societal desirability of the innovation process and its marketable products'.

Animal breeding programs can be viewed as continuous innovation processes as breeders are usually in a constant process of applying new (gene) technologies and methods. In addition to a focus on individual researchers' and key actors' thinking and reflections, there is a need to look at mechanisms affecting social and ethical consequences of animal breeding at an institutional level. In line with the framework of RRI, Olesen *et al.* (2000) concluded that 'important prerequisites for breeding programs for sustainable production are appropriate governmental policies, awareness of our way of thinking, and a more communal worldview informed by a subjective epistemology and a holistic ontology.' Hence, breeding objectives and strategies, informed by holistic assessments and more awareness among the actors, can facilitate socially accepted breeding programs contributing more efficiently to sustainable production meeting the current challenges of the environment, climate and the food system.

The objective of this article is to exemplify and discuss how animal breeding objectives and technology strategies can be defined and developed within the RRI framework.

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_479, © I. Olesen *et al.* 2022

The framework and requirements of RRI

RRI is a dynamic and iterative process inviting representatives of 'all' relevant stakeholders into a dialogue regarding the research/innovation process and its outcomes. RRI aims to align technological processes with societal values, expectations, and concerns. RRI can be seen as an additional layer through which societal views can be heard and taken into consideration, particularly in situations where the technological and societal developments move too fast for the electorate to affect legislation through voting and democratic means. The RRI process can be compared to traditional hearing processes related to legal acts and may supplement democratic processes. As such, RRI faces the same conflicts over values and distribution in society – how to deal with: (1) representation and asymmetrical power relations; (2) issues that do not involve win-win solutions; and (3) differences in interpretation of central concepts, as sustainability. In the Research Council of Norway (RCN), RRI includes four dimensions:

1. Anticipation: active engagement to describe, analyse and openly discuss foreseen impacts and uncertainties that might arise from processes or outcomes of the technology application.
2. Reflexivity: holistic reflections of (implicit) assumptions of research and innovation (regarding e.g. environmental and distributional effects), uncertainty and lacking knowledge.
3. Inclusion: invite all relevant stakeholders to contribute with views and opinions on the research trajectory, both on how the research is done and on the outcome.
4. Responsiveness: willingness to change the trajectory of research if results from above (e.g. stakeholders' feedback) reflect that the research is not likely to meet the desired impact.

RRI strategies for ensuring animal breeding within RRI

We discuss how breeders can comply with the four dimensions above expected by RCN to facilitate broader public acceptance of breeding schemes, and meet current complex challenges.

Anticipation. For analysing, and openly discussing the consequences of a breeding strategy or a breeding technology, an ethical matrix may be applied. Forsberg (2004) argued that public legitimacy is enhanced when the matrix method is applied in a participatory process to discuss various ethical concerns and how to ethically justify application of GMO rape seed. Furthermore, different consumer studies or surveys can be useful to anticipate impacts of different breeding objectives or strategies. For instance, Grimsrud *et al.* (2013) carried out an internet survey, which indicated a high willingness among Norwegian households to pay an extra tax to improve the welfare of farmed Atlantic salmon through increased resistance to diseases and salmon lice. The ethical matrix method is useful for broadening the range of concerns considered, but will not necessarily provide an agreed policy direction (e.g. on ethical or economic precedence).

Reflexivity. Breeders may commit to holistic reflections of the ethical, environmental, and political concerns relating to breeding choices to recognize and take social responsibility. Rosendal and Olesen (2022) discussed barriers to breed for increased lice resistance in Norwegian farmed Atlantic salmon, and factors that may explain the untapped potential of breeding for lice-resistance (market and regulation of breeding technologies). Based on interviews with breeders and farmers, it was concluded that market mechanisms were unsuitable, and policy instruments were not aimed at promoting lice resistance breeding. This gives rise to a paradoxical situation, which is allowing a greater number of stressful mechanical treatments of farmed salmon while the sea lice threat to wild salmon remains unchanged at an unacceptable high level in some regions. The study illustrated that there is little political willingness to interfere with the privately controlled breeding programmes, that are not transparent regarding some aspects. Social science findings might still have an impact in raising awareness among breeding actors.

Inclusion. Stakeholder participation is a central element in RRI, although challenging in terms of sufficient representation and asymmetrical power relations (Barnett and Finnemore, 2004). While traditional

stakeholder channels for policy involvement, such as lobbying, have been found to favour powerful actors, the same is largely true for other stakeholder processes (Orsini, 2016).

Effective involvement of some key stakeholders is currently facilitated automatically in cooperative breeding organisations and in integrated companies when key stakeholders (e.g. the farming members' representatives of the cooperative or farming managers of the integrated company) are consulted in decisions on breeding strategies. This illustrates how the organisation structure or business model allows for incorporating views and opinions of farmers and (or) other suppliers in decisions on developments within breeding and their outcome. In Norwegian dairy cattle breeding, the cooperative organisation Geno was particularly successful in implementing health and functional traits in the breeding objective in the early eighties as a response to the farmers' demand for healthy animals (FAO, 2007). Important factors for this success were a close cooperation with veterinarians on health recording and university geneticists as well as early responsiveness to farmers interests. The long-term strategy on sustainable production, rather than a short-term objective on yield only, still makes the Norwegian Red cattle breed competitive nationally as well as internationally.

Nielsen *et al.* (2006) presented a method to engage stakeholders when defining breeding goals for sustainable production. The method was based on how much, e.g. farmers or breeding organizations are willing to lose in gain in milk yield to improve traits related to sustainable production. Accepting a lower gain in milk yield, allows for increasing the value on e.g. mastitis resistance or reproduction.

These examples could speak in favour of RRI in relation to choices within traditional selective breeding where key stakeholders are relatively easy to identify. Similar representation may be challenging in processes linked to gene technology, with a broader range of issues to be raised.

Responsiveness. Responsiveness concerns whether the RRI process affects decision-making processes and technological choices. Stakeholder inclusion may enhance legitimacy but does not resolve which types of knowledge are more valid in the final decision, still to be made by the breeders. Social issues raised during the RRI may be further addressed by a relevant governmental body.

Both citizen panel and stakeholder involvement are examples of active engagement of stakeholders in a research project on gene editing in Cardiomyopathy syndrome in Atlantic salmon. A dynamic dialogue between lay people, industry representatives, NGO representatives, different scientists, and the project group include mutual learning activities with citizen panels and stakeholder workshops to build thrust and facilitate responsiveness.

Breeders can contribute to the public and regulatory discussions as well as to the education of stakeholders about breeding with research-based knowledge on e.g. animal welfare and consumer safety, environmental impacts and need for developing regulations and policies. The prospected opportunities of using gene editing have raised the discussion about softening the strict GMO regulations, encompassing gene edited products. Recent surveys have shown that most Norwegian consumers are positive towards using gene editing in plants and animals, if it is beneficial to society and contributes to sustainable development (Bioteknologirådet, 2020; Bugge, 2020). Therefore, the bulk of consumers tend to demand GMO labelling for gene edited organisms. By involving consumers in the gene technology innovation processes, the breeding organisations may comply to consumer preferences, and benefit from higher consumer acceptance. For example, breeders can respond by focusing on editing genes improving animal health and welfare or reducing unfavourable environmental impact with documented low risk.

Conclusions

Farm animal breeding programs, implementing new technologies and methods, can be viewed as continuous innovation processes calling for a framework of RRI to ensure consumers' and policy makers' acceptance and further licence to breed. Stakeholder engagement alone may not make innovations responsible or reduce the need for regulating technologies. Hence, the actions of anticipation, reflexivity (regarding assumptions, representation, power distribution, uncertainty and lacking knowledge), and finally responsiveness to all of these are also crucial for breeding programs' responsible outcome. By this, breeding programs can succeed in supporting a sustainable food system, that can meet the current and future complex challenges.

References

Barnett, M., & Finnemore, M. (2004). *Rules for the world: International organizations in global politics*. Ithaca, NY: Cornell University Press.

Bioteknologirådet. (2020). Norske forbrukeres holdninger til genredigering i norsk landbruk og akvakultur. p. 6. In Norwegian with English summary.

Bugge, A.B. (2020). GMO-foods or not: Have there been changes in consumers' views on genetically modified foods from 2017 to 2020? In Norwegian with English summary.

EC. (2014). Responsible research and innovation: Europe's ability to respond to societal challenges. Available at: https://op.europa.eu/en/publication-detail/-/publication/2be36f74-b490-409e-bb60-12fd438100fe

FAO. (2007). Norwegian Red Cattle – selection for functional traits. In: Rischkowsky, B. and Pilling, D. (eds) The State of the World's Animal Genetic Resources for Food and Agriculture. FAO, Rome. p. 399. Available at: https://www.fao.org/3/a1250e/a1250e.pdf

Forsberg, E-M. (2004). "The Ethical Matrix – a tool for ethical assessment of biotechnology," in L. Landeweerd, L-M Houdebine, and R. Termeulen (eds.), BioTechnology-Ethics. An Introduction, pp. 263–270.

Grimsrud, K.M., Nielsen, H.M., Navrud, S. and Olesen. (2013). Aquaculture 372-375: 19-27. https://doi.org/10.1016/j.aquaculture.2012.10.009

Nielsen, H. M., Christensen, L. G., and Ødegård, J. (2006). J. Dairy Sci., 89, 3615–3625. https://doi.org/10.3168/jds.S0022-0302(06)72401-8

Olesen, I., Groen, A.F. and Gjerde, B. (2000). J. Anim. Sci. 78, 570–582.

Orsini, A. (2016). Do Non-State Perspectives Matter for Treaty Ratification and Implementation? The case of the European Consultation on the Nagoya protocol. *Environmental Policy and Governance, 26*(5), 377-393.:10.1002/eet.1712

Rosendal, G.K. and Olesen, I. (2022). Aquaculture 548: 737574. https://doi.org/10.1016/j.aquaculture.2021.737574

Von Schomberg, R. (2012). Prospects for technology assessment in a framework of responsible research and innovation. In M. Düsseldorp & R. Beecroft (Eds.), *Technikfolgen abschätzen lehren. Bildungspotenziale transdisziplinärer Methoden*. Berlin and Darmstadt: Springer VS.

480. On farm specific breeding goals for mating lead to higher genetic gains at herd level

M. Bérodier Gomès[1,2,3*], M. Brochard[3,4] and V. Ducrocq[1]

[1]GABI, INRAE, AgroParisTech, Université Paris-Saclay, 78350 Jouy-en-Josas, France; [2]MO3, 01250 Ceyzériat, France; [3]Umotest, 01250 Ceyzériat, France; [4]Institut de l'Elevage, 75012 Paris, France; mberodier@umotest.com

Abstract

Even when dairy cattle breeding organizations have no interest to implement environment-specific breeding programs, farmers may benefit from using herd-specific breeding goals instead of the main breeding goal of their breed. To assess this hypothesis, we constructed and compared four different breeding goals under conventional or organic farming system (two per system). The Total Merit Index derived from the simulated breeding goals were used to optimize the mating of 9,143 Montbéliarde females to 54 bulls. Despite strong correlations between the total merit indices (0.98), more than a quarter of the planned matings were different among the breeding goal considered. Results also showed that farmers would economically benefit from using herd-specific breeding goals instead of the main breeding goal of their breed without the need to create a specific breeding program.

Introduction

At breed's level, Mulder *et al.* (2006) showed that from dairy cattle breeding organizations' point of view (breeding elite sires), there is no need to implement environment-specific progeny testing breeding programs when the genetic correlation between two environments is higher than 0.61. More recently, Slagboom *et al.* (2019) reached similar conclusions under genomic selection. But at herd's level when optimizing the matings of their females, even if there is no need to create a specific breeding program, farmers may benefit from using herd-specific breeding goals instead of the main breeding goal of their breed. To test this hypothesis in the case of the Montbéliarde breed, we studied the situation under four distinct breeding goals (organic-based or conventional) for which different total merit indices were derived.

Materials & methods

Simulated mating plans. To assess the losses/gains related to the optimization of the matings based on a given expected economic score (EES) specific to a farming system (e.g. standard conventional farming system) while the females to be bred (and their future offspring) are raised under a different context (e.g. organic farming system), the average economic results of four mating plans (scenarios) were calculated using the EES values of each of the four farming systems (scales). Therefore, 9,143 genotyped Montbéliarde females from 160 French herds were virtually planned for mating with a set of 54 genotyped bulls, using a linear programming approach. It resulted in 4 mating plans, one per breeding goal tested. Mate allocation optimization was performed based on the EES of each potential mating. This EES accounts for the expected genetic merit of the progeny (expected Total Merit), mates' coancestry, type of semen (sexed or conventional) and risk to conceive an embryo affected by a genetic defect as in Bérodier *et al.* (2020).

Breeding goals definition. To assess the influence of a breeding goal on optimized mating plans, we constructed and compared four different breeding goals, similarly to the work of Slagboom *et al.* (2018): (1) the current breeding goal of the Montbéliarde breed which focuses on standard traits of economic interest but also includes type traits (STD_TYP); (2) an organic breeding goal based on economic optimization at farm level (ORG_ECO scenario); (3) a so-called 'ethic' organic breeding goal based on the four main international principles of organic agriculture set up by IFOAM (International Federation of

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_480, © M. Bérodier Gomès *et al.* 2022

Organic Agriculture Movements), namely Health, Ecology, Fairness and Care (ORG_ETH scenario); (4) a conventional breeding goal based on economic optimization at herd level, consequently, not integrating type traits (STD_ECO scenario). The four main international principles of organic agriculture are not economically orientated. IFOAM asserts that organic farming should sustain and enhance the health of all beings, should be based on living ecological systems and cycles, should ensure fairness with regard to the common environment and life opportunities and should protect the health and well-being of current and future generations and the environment.

Genetic gain and inbreeding depression. The last update of the STD_TYP breeding goal was performed in 2012 based on economic optimization models and including the Montbéliarde breed specific expectations in terms of selection for type. The STD_ECO and ORG_ECO breeding goals were directly derived from economic models specific to the farming system considered (Pinard and Regaldo, 2013), respectively a conventional and an organic farming system. For ORG_ETH system, the challenge was to build a breeding goal that leads to genetic gain on the most important traits according to the four main principles of organic agriculture (and not a purely economic objective). Results from Slagboom *et al.* (2018) were adapted to the traits available for selection in the Montbéliarde breed in France. An in-house software developed by the Institut de l'Elevage was used. Briefly, this software simulates selection in a real population on the animals' estimated breeding values (EBV), provided as an input file with many different weight combinations until a response to selection maximizing economic gain is found. This software also accounts for specific constraints to prevent a negative selection response for a particular trait with no economic value associated (Fuerst-Waltl *et al.*, 2016), for instance traits related to the principles of organic agriculture. The use of data from a real population as an input enables the program to account for observed genetic correlations (positive or negative) between the different traits of the breeding goal. The inputs were: a dataset of 11,400 young Montbéliarde bulls born between 2013 and 2018 with their EBV; economic values of the genetic standard deviation of each trait; if necessary, constraints on responses to selection (e.g. no negative response to selection for dairy production traits). Correlations between the four Total Merit Index EBV (one per breeding goal) were calculated from the EBV of the 54 bulls and 9,143 females considered in this study (Table 1).

Table 2 provides the relative weights per genetic standard deviations unit for the traits included in our four simulated breeding goals. To calculate EES, the economic value associated to one genetic standard deviation of the simulated total merit indices (σ) and to the economic impact of 1% of inbreeding (λ) were calculated as in Bérodier *et al.* (2020) (Table 2).

Economic gain. The 16 economic results obtained from the simulated mating plan were expressed as a deviation from a random mating plan (i.e. where cows were mated at random to one of the 54 bulls available; Bérodier *et al.* (2020)). For instance, the expected average economic gain when optimizing the mate allocation based on STD_TYP information while the herd is under an ORG_ECO farming system, compared to a random mating situation, was calculated as:

$$\frac{\sum_{i=1}^{nb_couples} score_{i,ORG_{ECO}}(STD_TYP) - \sum_{j=1}^{nb_couples} score_{j,ORG_{ECO}}(RAND)}{nb_couples}$$

Table 1. Correlations between Total Merit Index EBV of the four breeding goals.

	STD_ECO	ORG_ECO	ORG_ETH
ORG_ECO	0.98		
ORG_ETH	0.81	0.76	
STD_TYP	0.85	0.88	0.62

Table 2. Traits' relative weights per genetic standard deviation and economic values associated to each breeding goal (= scenario).

	Trait[1] (%)														Value[2] (€)	
Scenario	FY	PY	FR	PR	CS	MS	CF	HF	CI	LO	CE	CM	MS	TT	σ	λ
STD_ECO	12	36			17	1	13	4		17					295	26.1
ORG_ECO	15	39			12	9	9	2		14					295	34.5
ORG_ETH	7	17			13	19	15	4		19	3	3			251	34.5
STD_TYP	1	9.8	4.8	29.4	8.7	5.8	9	4.5	4.5	5			5	12.5	246	26.1

[1] Traits included each breeding goal: Fat yield (FY), Protein yield (PY), Fat rate (FR), Protein rate (PR), Cells (CS), Mastitis (MS), Cow fertility (CF), Heifer fertility (HF), Calving to first insemination interval (CI), Longevity (LO), Calving Ease (CE), Calf Mortality (CM), Milking speed (MS), Type traits (TT).

[2] Economic values associated to one genetic standard deviation (σ) and to 1% of inbreeding (λ).ß

With $score_{i,ORG_ECO}$(STD_TYP) the ORG_ECO EES of couple i which was planned using linear programming mate allocation optimization base on STD_TYP information; $score_{j,ORG_ECO}$(RAND) the ORG_ECO EES of couple j which was planned using random mate allocation; nb_couples the number of couples to plan (9,143, the number of females to mate).

Results

Mate allocation value. Table 3 can be interpreted as follows: if breeders' breeding goal (scale) is ORG_ETH and their breeding plan are based on the STD_TYP information (scenario), they will earn on average 78€ more per mating over the productive life of an offspring by optimizing their matings using the STD_TYP total merit index whereas they could have earned on average 107,5€ more per offspring by optimizing their matings with the ORG_ETH total merit index.

As expected, for each economic score scale (i.e. for each farming system), the best average economic score is obtained with the scenario (i.e. with the mating plan) using the same information (Table 3). For instance, farmers in ORG_ETH farming systems can expect to increase on average its benefit from 29.5€ (=107.5-98) per per female offspring over its carrier when using ORG_ETH instead of STD_TYP breeding goal to optimize matings. In Table 4, the percentage of identical matings between two scenarios indicates that despite strong correlations between the total merit indices (e.g. 0.98 between the STD_ECO and ORG_ECO scenarios), more than a quarter of the planned matings are different. Similarly, the lower the correlations between the total merit indices of the different breeding goals, the lower the correlations between the bulls' contributions (i.e. their number of planned matings).

Table 3. Average economic score deviation (in euros) from random matings.

	Scale[1]			
Scenarios[2]	STD_ECO	ORG_ECO	ORG_ETH	STD_TYP
STD_ECO	84	90.6	94	63.5
ORG_ECO	80.6	93.6	89.7	66.3
ORG_ETH	74.4	81.6	107.5	57.9
STD_TYP	69	81	78	73.9

[1] Farming system of the herd and farmer's breeding goal.

[2] Type of EES used to optimize the matings.

Table 4. Percentage of identical matings between two mate allocation scenarios (lower triangular matrix) and correlations between sires contributions (i.e. number of planned matings) across scenarios (upper triangular matrix).

	STD_ECO	ORG_ECO	ORG_ETH	STD_TYP
STD_ECO	-	91%	78%	65%
ORG_ECO	74%	-	70%	71%
ORG_ETH	58%	52%	-	55%
STD_TYP	53%	57%	45%	-

Discussion

These results confirm our hypothesis that farmers can benefit from using herd-specific breeding goals instead of the main breeding goal of their breed even when there is no need to create a specific breeding program adapted to their within-herd breeding goal. For the four breeding goals simulated, the correlations between the total merit indices are not low enough to justify the creation of new, separate selection schemes, according to Mulder *et al.* (2006) and Slagboom *et al.* (2018). Nevertheless, we have shown that the less farmers' breeding goal is correlated to the one of the breed (due to the specificities of their farming system and/or their own expectations), the more they could improve their within-herd breeding choices by using information specific to their farming system and their breeding goal. Indeed, access to EBV specific to their breeding goal and farming system would allow them to better select sires and make better matings plan choices in order to maximize the genetic and the economic gains for their own breeding goal. We recommend farmers to use the breeding goal that matches their production system.

References

Bérodier, M., Berg P., Meuwissen T., Boichard D., Brochard M., *et al.* (2020) Animal. 15(1): 100016. https://doi.org/10.1016/j.animal.2020.100016

Fuerst-Waltl, B., Fuerst C., Obritzhauser W. and Egger-Danner C. (2016) J. Dairy Sci. 99:9796-9809. https://doi.org/10.3168/jds.2016-11095

Mulder, H.A., Veerkamp R.F., Ducro B.J., van Arendonk J.A.M. and Bijma P. (2006) J. Dairy Sci. 89:1740-1752. https://doi.org/10.3168/jds.S0022-0302(06)72242-1

Pinard D. and D. Regaldo. 2013. Proc. of the 64th Annual Meeting of the European Association for Animal Production, Nantes, France.

Slagboom, M., Wallenbeck A., Hjortø L., Sørensen A.C., Rydhmer L., *et al.* (2018) J. Dairy Sci. 101:11086-11096. https://doi.org/10.3168/jds.2017-14316

Slagboom, M., Kargo M., Sørensen A.C., Thomasen J.R. and Mulder H.A. (2019) J. Dairy Sci. 102: 8197-8209. https://doi.org/10.3168/jds.2018-15939

481. A bio-economic model to estimate the economic value of sow feed intake during lactation

D.M. Thekkoot[1*], R.A Kemp[1] and J.C.M. Dekkers[2]

[1]Genesus Inc., 101 2nd Street, Oakville, R0H 0Y0, MB, Canada; [2]Iowa State University, Ames, IA 50011, USA; dinesh@genesus.com

Abstract

The objective of this paper was to develop a comprehensive bio-economic model to calculate the economic values of sow breeding goal traits. In addition to the regular production and management parameters, the model accounted for sow feed intake during lactation (SLFI) and its interactions with sow body resource mobilization during lactation and piglet weight gain during lactation, and how these traits could influence the successive gestation feed intake. Weighted by the genetic standard deviation of each trait, the number of piglets born had the highest contribution to the breeding goal (48%), followed by preweaning mortality (incl. stillborn) (45%), age of puberty and wean to service interval (both 3%), and SLFI (1%). The results show that incorporating this complex interaction resulted in a very small but negative economic value for SLFI.

Introduction

The goal of any pig breeding program is to increase producer profit without adversely affecting the overall health and welfare of the pigs. This is attained by using a total merit index, which is a linear function of estimated breeding values of economically important traits, weighted by their economic values. Economic values can be calculated from a system of equations known as a bio-economic model, which is a function of phenotypic traits and management factors that contribute to the profit of pork production (Tess *et al.*, 1983). Generally, traits considered for maternal lines include litter size, pre-weaning mortality, litter weight at weaning, sow longevity, etc. Typically, little emphasis is put on traits associated with lactation, such as sow feed intake during lactation (SLFI), likely because of the cost required for accurately measuring relevant phenotypes. Studies have shown that traits associated with lactation have a sizable genetic component and, hence, including them in the selection index should allow concurrent improvement of sow lactation performance, along with reproduction and grow-finish performance (Bergsma *et al.*, 2009, Thekkoot *et al.*, 2016). SLFI is a cost component and including it in the bio-economic model without accounting for other lactation factors such as sow body resource mobilization, piglet weight gain, and the impact of these on successive gestation feed intake, can lead to overestimation of the negative weight of SLFI (3.5% based on previous work (unpublished)), which can result in selection reducing sow feed intake during lactation. The objective of this paper was to accurately estimate the economic value of SLFI using a bio-economic model that takes the complex interactions of SLFI with body resource mobilization, piglet weight gain, and successive gestation feed intake into account.

Materials & methods

A deterministic bio-economic model was developed to describe the production of an F1 sow over 5 litters, covering the birth of the gilt (day 0) to the time of culling of the sow after the 5th parity. The model covers a basic farrow to wean operation and does not attempt to model all complexities of the production system. The time period was divided into 6 phases (the gilt phase, i.e. birth to first insemination, and parities 1, 2, 3, 4, and 5). Income and expenses for these periods were calculated separately and discounted to the time of birth of the sow (day 0). The net profit (NP) generated by the F1 sow was expressed as discounted Total Income (dTI) minus discounted Total Costs (dTC), i.e.

$$NP = dTI - dTC \quad (1)$$

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_481, © D.M. Thekkoot *et al.* 2022

Total income. Total income from the sow was derived from piglet sales at weaning (IWP) from all 5 parities and from the sale of the culled sow after the 5th parity (ICS). There was no income from the gilt phase. Income from parities 1 to 5 was calculated separately for each parity. In parities 1 to 4, income depended on the sale of piglets only, whereas income in parity 5 also included sale of the culled sow. Income from the sale of weaned piglets (IWP) depends on: (1) the total number of piglets weaned per parity (NW), which in turn depends on the total number of piglets born dead or alive (TNB) and total pre-weaning mortality (PWM). PWM depends on the number of stillborn piglets (SB), the number of piglets that died in the first 24 hours (PWMb24), and death after day 1 and before weaning (PWMa24); and (2) the price of a weaned piglet (PrWP). Income from the culled sow (ICW) depends on: (1) the proportion of sows that are culled after parity 5 (PCUS); (2) average weight of a culled sows (WCSS); and (3) price per kg of a culled sows (PrCUS). As a result, total income from all 5 parities before discounting can be written as:

$$\textit{Total income} = [5 \times (TNB - (SB + PWMb24 + PWMa24)) \times PrWP] + [PCUS \times WCSS \times PrCUS] \quad (2)$$

Total costs. Costs were calculated separately for each phase.

Cost of gilt phase. Costs incurred during this period depends on: (1) feed costs for the gilt (FCOG); (2) non-feed costs (NFCOG); and (3) the selection rate of gilts (gSelRate). The total cost in the gilt phase before discounting can then be written as:

$$\textit{Total cost in gilt phase} = (FCOG + NFCOG) / gSelRate \quad (3)$$

Cost in parity 1. This is the cost incurred from the day of the first insemination to the last day of the parity 1 wean to service interval and can be broadly divided into sow costs and piglet costs. Sow costs can then be divided into feed costs and non-feed costs. Each of these costs was further divided into three periods based on the physiological state of the sow (gestation, lactation, and wean to next service interval). Sow costs during gestation depended on: (1) gestation length (GL); (2) average feed intake during first gestation (GGADFI); (3) price per kg of gestation feed (pkgGFeed); and (4) non-feed cost in gestation (NFCG); costs during lactation depended on (5) lactation length (LL); (6) average feed intake during first parity lactation (GLADFI); (7) price per kg of lactation feed (pkgLFeed); and (8) non-feed cost in lactation (NFCL); costs from wean to next service depended on (9) wean to service interval (WSI); (10) average feed intake during wean to next service (WSIADFI); (11) price per kg of sow feed (pkgDFeed); and (12) non-feed cost from weaning to next service ((NFCWSI). Total sow costs in parity 1 before discounting can then be written as:

$$\textit{Total SOW cost parity 1} = [(GL \times GGADFI \times pkgGFeed) + (LL \times GLADFI \times pkgLFeed) + (WSI \times WSIADFI \times pkgDFeed)] + [(GL \times NFCG) + (LL \times NFCL) + (WSI \times NFCWSI)] \quad (4)$$

Piglet costs in parity 1 are divided into two components: (1) cost of raising piglets from birth to weaning; and (2) cost of disposing dead piglets. The cost of raising piglets depends on: (1) lactation length (LL); (2) cost (feed and non-feed) per piglet per day in the farrowing pen (CORPL); and (3) the total number of piglets weaned (TW). TW in turn depends on TNB, SB, PWMb24, and PWMa24. The total cost of disposing dead piglets depends on: (1) pre-weaning mortality (SB, PWMb24, PWMa24); and (2) the cost of disposing one dead piglet (CODPL). As most piglet deaths happen in the first 48 hours of birth, rearing costs of dead piglets can be ignored. Total piglet costs in parity 1 before discounting can then be written as:

$$\textit{Total piglet cost in parity 1} = [(LL \times (TNB - (SB + PWMb24 + PWMa24)) \times CORPL] + [(SB + PWMb24 + PWMa24) \times CODPL] \quad (5)$$

Costs in parities 2 to 5. During parity 2 to 5 gestation and lactation, all cost components are the same as in parity 1, except for feed intake. Sow feed intake during gestation for parities 2 to 5 depends on the energy balance from the previous lactation, which in turn depends on the number of piglets nursed, feed intake during lactation, and the change in body condition during lactation. During lactation, demand for energy is high, as the energy available to the sow has to be divided between its own growth, maintenance, and that required for milk production. Energy requirements during lactation (ERL) can be divided into three parts: (1) energy required for the maintenance of the sow during lactation (ERSM), which in turn depends on sow body weight at farrowing (SBWTF), bodyweight at weaning (SBWTW), and LL; (2) energy required for maintenance of the litter (ERLM), which depends on the average birth weight of the piglets (ABWT), average weaning weight of the piglets (AWWT), number of piglets weaned, and LL; and (3) energy required for growth of the litter (ERLG), which depends on ABWT, AWWT, NW, metabolizable energy content of 1 kg of body fat mass (MEFAT), metabolizable energy content of 1 kg of body protein mass (MEPROT), and average daily gain of piglets from birth to weaning (ADGBW). Each of these cost components can then be written as:

$$ERSM = 0.44 \times ((SBWTF + SBWTW) / 2)^{0.75} \times LL \quad (6)$$

$$ERLM = 0.44 \times ((ABWT + \mathrm{A}WWT) / 2)^{0.75} \times LL] \times \mathrm{NW} \quad (7)$$

$$ERLG = [\{((AWWT - ABWT) \times 0.135 + (0.00014 \times ADGBW))) \times MEFAT\} + \{((AWT - ABT) \times 0.16) \times MEPROT\}] \times NW \quad (8)$$

$$ERL = ERSM + ERLM + ERLG \quad (9)$$

Feed consumed and body resources mobilized are the two sources of energy for a sow during lactation. Energy available from feed intake during lactation (EFEED) depends on: (1) average daily feed intake during lactation (ADFIL); (2) LL; and (3) energy content in lactation feed (MELFEED), and can be written as:

$$EFEED = ADFIL \times LL \times MELFEED` \quad (10)$$

Total energy mobilized from body reserves during lactation (EMOBL) depends on: (1) body fat mobilized during lactation (BFMOBL); (2) body protein mobilized during lactation (PROMOBL); (3) metabolizable energy content in 1 kg of body fat mass (MEFAT); and (4) metabolizable energy content in 1 kg of body protein mass (MEPROT). The amount of fat and protein mobilized during lactation, in turn, depends on body weight loss (WTLOSS) and back fat loss (BFLOSS) during lactation. Total energy mobilized during lactation can then be written as:

$$EMOBL = [(0.1711 \times WTLOSS - 0.3113 \times BFLOSS) \times MEPROT] + (0.1027 \times WTLOSS + 1.904 \times BFLOSS) \times MEFAT]` \quad (11)$$

Eissen *et al.* (2000) stated that during lactation, when milk production is a high priority and nutrient intake is insufficient to meet demands, sows mobilize body reserves to maintain milk production. This will result in a negative energy balance or an energy deficit (ENERDEF) at the end of lactation, which is the difference between the available energy sources and energy demands during lactation. The sow recoups this energy deficit in the following gestation by increasing feed intake during gestation. The amount of additional feed intake depends on the energy deficit from the previous lactation (ENERDEF), the energy density of gestation feed (MEGFEED), the efficiency of converting feed to body reserve (EFFBDR), and the efficiency of converting body reserves to milk (EFBRML), and can be calculated as:

Additional gestational feed = ((*ENERDEF* / *MEGFEED*) / *EFFDBR*) / *EFBRML* (12)

Incorporating these equations into the bio-economic model allows feed intake during gestation in parities 2 to 5 to be calculated to estimate the economic value of SLFI.

Using this bio-economic model, we calculated the net profit for an F1 gilt utilizing trait means from Genesus herds and average North American prices and costs. The economic values were then calculated from the change in predicted profit based on a unit change in the population mean for the trait, keeping all other traits constant (Quinton *et al.*, 2006). The relative contribution of the traits to the breeding goal were calculated by multiplying the economic values for a unit change of the trait by its genetic standard deviation, as described by Amer *et al.* (2014).

Results

The economic values obtained from the bio-economic model resulted in the following relative contribution of traits to the breeding goal: total number of piglets born 48%, total preweaning mortality including stillborn 45%, age at puberty 3%, wean to service interval 3%, and sow lactation feed intake 1%. Litter size at birth was the highest contributor to the breeding objective, and agrees with the results of Quinton *et al.* (2006), Amer *et al.* (2014).

Discussion

The economic value of SLFI, was calculated as the change in profit by increasing the lactation feed intake by 1 kg per day, while keeping everything else constant in the bio-economic model. A bio-economic model that includes SLFI without considering factors such as sow body resource mobilization during lactation, and the impact of SLFI on successive gestation feed intake, resulted in negative relative weight of 3.5% for SLFI. An increase in feed intake during lactation will result in less body reserve mobilization during lactation, and the sow will compensate for this by reducing the feed intake in successive gestation. However, the economic value of SLFI also depends on the difference in energy density and in price between lactation and gestation feed, and on the efficiency with which these feeds are converted to metabolic needs. This complex interaction was incorporated into the bio-economic model that we developed. The results showed that by taking all these factors into consideration, the relative emphasis of SLFI was still negative (1%) but smaller than without accounting for these. Further investigation showed that the economic value of SLFI never became positive, as feed intake during first parity gestation was not affected by an increase in lactation feed intake, and the increase in revenue from a higher body weight of a culled sow was not sufficient to overcome the higher cost of lactation feed. These results show that the relative emphasis of SLFI was negative but very small.

References

Amer P. R., Ludemann C. I., and Hermesch S. (2014) J. Anim. Sci. 2014(92): 5345–5357. https://doi.org/10.2527/jas.2014-7943

Bergsma, R., Kanis E., Verstegen M. W. A., van der Peet-Schwering C. M. C., and Knol E. F. (2009) Livestock Science, 125 (2–3): 208-222 https://doi.org/10.1016/j.livsci.2009.04.011

Eissen J.J., Kanis E., and Kemp B. (2000) Livestock Production Science 64 (2–3): 147-165 https://doi.org/10.1016/S0301-6226(99)00153-0

Quinton V.M., Wilton J.W., Robinson J.A., and Mathur P.K. (2006) Livestock Science 99(1): 1871-1413 https://doi.org/10.1016/j.livprodsci.2005.06.002

Tess, M.W., Bennett, G.L., and Dickerson, G.E., (1983) J. Anim. Sci. 56(2): 336–353. https://doi.org/10.2527/jas1983.562336x

Thekkoot D. M., Kemp R. A., Rothschild M. F., Plastow G.S., and Dekkers J. C. M., (2016) J. Anim. Sci. 94(11): 4516-4529 https://doi.org/10.2527/jas.2015-0255

482. Consequences of using different economic selection index methods on greenhouse gas emissions in beef cattle

B.J. Walmsley

AGBU, a joint venture of NSW Department of Primary Industries and University of New England, 2351, Armidale, Australia; brad.walmsley@dpi.nsw.gov.au

Abstract

The impacts of using partial (sub) indexes on expected greenhouse gas emissions (GHG) from a self-replacing cow herd producing 120-day grain-finished progeny is compared to selection with a full self-replacing BreedObject index. Sub-indexes only containing post-weaning traits of the slaughter progeny (terminal) and a maternal sub-index were developed using BreedObject. The terminal sub-index is expected to increase mature cow weight (MCW) and milk, resulting in increases in GHG emissions. In contrast, the full self-replacing index is expected to produce slight reductions in MCW and negligible change in milk, which results in decreases in GHG emissions of the cow herd and young animal at pasture. The maternal sub-index is expected to result in larger GHG emission reductions due to larger decreases in MCW as well as negligible change in milk. When constructing partial indexes, the implications on GHG emissions of the cow herd and progeny need to be carefully considered.

Introduction

Pressure is increasing to reduce greenhouse gas (GHG) emissions due to societal concerns about climate change. In livestock, attention is particularly focused on beef cattle because of the relatively large contribution made by enteric methane (Edenhofer *et al.* 2014). Blaxter and Clapperton (1965) showed the association between feed intake (FI) and GHG emissions, which Walmsley *et al.* (2019) used to show upward trends in GHG emissions due to increased FI from productivity gains in the absence of feed efficiency improvements (Walmsley *et al.* 2017). Barwick *et al.* (2019) demonstrated GHG emissions improvements could be made using full production system index selection when feed is appropriately costed, while still increasing net returns. Dividing the breeding objective into sub groupings (sub-indexes) has been proposed as a means of aiding adoption of genetic improvement (Dekkers and Gibson 1998; Wilmink 1996). However, using sub-indexes may not result in optimal selection (Dekkers and Gibson 1998; Barwick and Henzell 2005). The utility of any index is a function of its representation of the production system it is applied in. This has important implications for addressing GHG emissions given the cow herd accounts for the majority of feed-related costs (Walmsley *et al.* 2018). This study examines the impact the use of sub-indexes could have on GHG emissions compared to using a full production system (self-replacing) index.

Materials & methods

Breeding objectives. Breeding objectives for net return per cow mated were derived with BreedObject (Barwick and Henzell 2005) following the outline of Schneeberger *et al.* (1992), where EBVs are combined into one aggregate EBV for the breeding objective. The EBVs used were from the December 2021 BREEDPLAN analysis for 636 published Hereford sires. The BreedObject genetic parameters were derived from industry and literature estimates, and relevant for *Bos taurus* breeds in Australia. Parameter details can be requested from the author. Weightings (b) applied to the EBV selection criteria, û, are derived as:

$$b = G_{11}^{-1}G_{12}\nu \quad (1)$$

where G_{11} and G_{12} are the genetic variances and covariances among BREEDPLAN EBVs (G11) and between these and the objective traits (G12), respectively, and ν is a vector of trait economic values.

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_482, © B.J. Walmsley 2022

The example production system included a pasture backgrounding period followed by a 120-day feedlot finishing period producing 620 kg *Bos taurus* steers at 21 months of age from a self-replacing cow herd at pasture. Traits in the breeding objective relating to the growing animal were sale liveweight (kg), dressing %, saleable meat %, rump fat depth (mm), marbling score (AUSMEAT scale, 1-12), feedlot entry liveweight (kg), weaning liveweight (direct; kg), residual feed intake-pasture (kg/day), and residual feed-intake feedlot (kg/day) plus cow herd traits; weaning liveweight (maternal; kg), calving ease (direct and maternal; %), mature cow weight (MCW, kg), cow weaning rate (%), and cow condition score at joining (1-15). The general form of the economic value for traits is change (Δ) in returns – Δ FI cost – Δ non-feed management cost. The FI associated with Δ in each objective trait was estimated using the equations described by Freer *et al.* (2007).

Constructing partial indexes. A full self-replacing index (Base), a terminal sub-index (TSub) focused on slaughter progeny and a maternal sub-index (MSub) were constructed using BreedObject. The TSub index only contained traits of the growing animal postweaning in the breeding objective but used the same cost structure as the base index. The MSub index was constructed to show what would be required to counter-balance outcomes from the TSub index. The MSub index used the same cost structure and contained all breeding objective traits but only took account of the maternal contribution to slaughter progeny post-weaning.

Production, feed intake and methane responses. Anticipated responses to selection using the Base and partial indexes were examined by comparing the selection differential in EBVs between the top 10% of sires and the whole published sire list when ranked on index. Estimates of genetic total FI (excluding any period of surplus feed) per sire were derived as presented by Walmsley *et al.* (2017). The breeding objective did not include GHG emissions and residual GHG emissions were not included as EBVs because they are not currently available and their associations with other traits are still under investigation. Methane production was derived using the phenotypic relationships between FI and methane production of Charmley *et al.* (2016) when animals were at pasture, Equation 7 of Johnson *et al.* (1993) when animals were in the feedlot, and the recommended global warming constant for methane (28; Edenhofer *et al.* 2014) for determining kg of CO_2-equivalent.

Results

Figure 1 shows calving ease, sale liveweight and cow weaning rate have the highest positive relative economic emphasis, while negative economic emphasis is placed on feed efficiency (reducing RFI is desired), MCW and milk, to reduce feed requirements. Selection differentials in Figure 2 show the Base index is expected to result in slight reductions in MCW and negligible change in milk while allowing improvements in the other production EBVs, such as calving ease, growth, carcass quality and fertility (negative desired). This index is also expected to result in reductions in both FI and GHG emissions of the cow herd (Figure 2b) and the young animal at pasture, but small increases for the young animal in the feedlot. In comparison, the TSub index is expected to increase growth capacity, which is associated with increases in MCW and milk EBVs. The TSub index is also expected to result in negligible change in calving ease, small improvements in fertility and carcass EBVs along with expected increases in both FI and GHG emissions across the production system.

The selection differentials in Figure 2 also show the MSub index is expected to reduce MCW which is expected to lead to reductions in FI and GHG emissions in the cow herd. The MSub index is also expected to improve fertility, calving ease and carcass traits, other than retail beef yield. However, growth potential is expected to decrease particularly for 600-day weight which also results in reductions in FI and GHG emissions for the young animal.

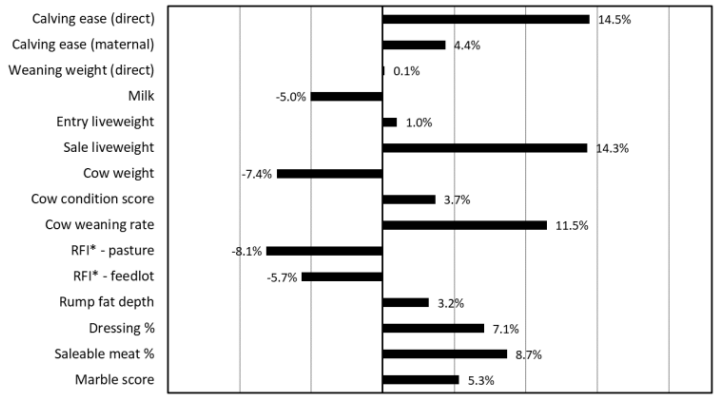


Figure 1. The relative economic emphasis on the breeding objective traits for a self-replacing production system producing 120-day grain-finished steers (Base index). RFI: residual feed intake.

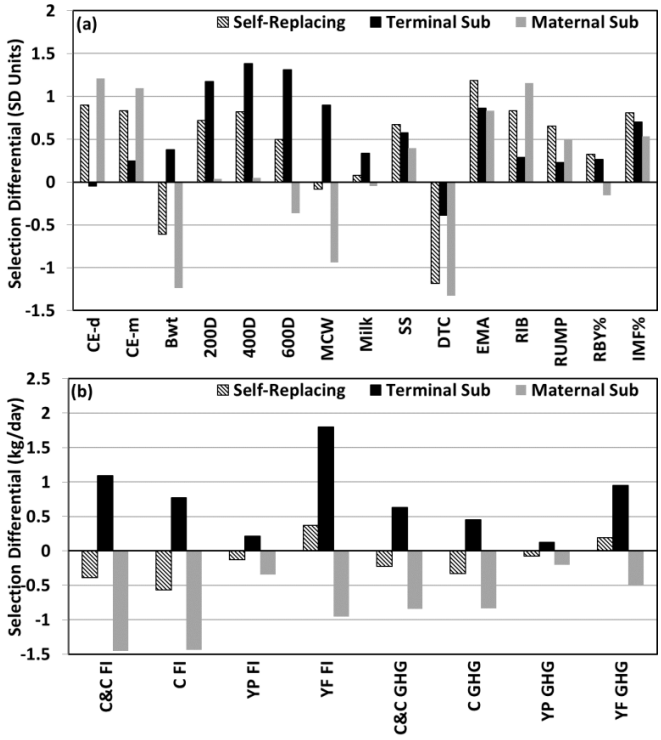


Figure 2. Selection differentials between top 10% and full published Hereford sire list ranked on the self-replacing index (base), the terminal sub-index (TSub) or the maternal sub-index (M-Sub) for (a) the BREEDPLAN EBVs and (b) estimated daily feed intake and greenhouse gas emissions. Traits: CE-d: calving ease-direct, CE-m: calving ease-maternal, Bwt: birth weight, 200D: 200-day weight, 400D: 400-day weight, 600D: 600-day weight, MCW: mature cow weight, Milk: 200-day maternal weight, SS: scrotal size, DTC: days-to-calving, EMA: eye muscle area, RIB: rib fat depth, RUMP: rump fat depth, RBY%: retail beef yield percentage, IMF%: intramuscular fat percentage, C&C: cow and calf at pasture, C: only cow at pasture, YP: young animal at pasture, YF: young animal in feedlot, FI: feed intake and GHG: GHG emissions.

Discussion

The results highlight if partial indexes intended for the post-weaning segment only were inappropriately applied then they should be expected to result in undesirable increases in FI and subsequently GHG emissions in the cow herd. The development of maternal indexes could be an avenue for partially countering such an outcome, but it comes with the added expense of maintaining two breeding nuclei and potentially compromising gains in production traits like retail beef yield or feedlot growth performance thus negatively impacting feedlot profitability. In contrast, a well-constructed self-replacing index is capable of restricting MCW while allowing other production traits to improve and result in a reduction in expected FI and GHG emissions. When constructing partial indexes, the potential implications on GHG emissions need to be carefully considered, particularly if there is scope for these indexes to be inappropriately used when breeding replacement females.

Acknowledgements

The author acknowledges Meat & Livestock Australia (L.GEN.1704 and L.GEN.2204) for financial support and Herefords Australia for BreedObject access to data.

References

Barwick S.A., Henzell A.L., Herd R.M., Walmsley B.J. and Arthur P.F. (2019) Genet. Sel. Evol. 51:18-. https://doi.org/10.1186/s12711-019-0459-5

Barwick S.A., and Henzell A.L. (2005) Aust. J. Exp. Agric. 45(7):923-933. https://doi.org/10.1071/EA05068

Blaxter K.L. and Clapperton J.L. (1965) Br. J. Nutr. 19(1):511-522. https://doi.org/10.1079/BJN19650046

Charmley E., Williams S.R.O., Moate P.J., Hegarty R.S., Herd, R.M., *et al.* (2016) Anim. Prod. Sci. 56(2):169-180. https://doi.org/10.1071/AN15365

Dekkers J.C. and Gibson J.P. (1998) J. Dairy Sci. 81(2):19-35. https://doi.org/10.3168/jds.S0022-0302(98)70151-1

Edenhofer O., Pichs-Madruga R., Sokona Y., Kadner S., Minx J.C., *et al.* (2014) Climate change 2014: Mitigation of climate change. Cambridge University Press, New York.

Freer M., Dove H. and Nolan J.V. (2007) Nutrient requirement of domesticated ruminants. CSIRO Publishing, Collingwood, Victoria, Australia.

Johnson D.E., Hill T.M., Ward G.M., Johnson K.A., Branine M.E., *et al.* (1993) Ruminants and other animals. In: Khalil M.A.K. (ed) 'Atmospheric methane', Springer-Verlag, Berlin, Germany, pp 199-229. https://doi.org/10.1007/978-3-642-84605-2_11

Schneeberger M., Barwick S.A., Crow G.H. and Hammond K. (1992) J. Anim. Breed. Gene. 109(1-6):180-187. https://doi.org/10.1111/j.1439-0388.1992.tb00395.x

Walmsley B.J., Henzell A.L. and Barwick S.A. (2017) Proc. of the Assoc. Adv. Anim. Breed. Gene., Townsville, Australia.

Walmsley B.J., Henzell A.L. and Barwick S.A. (2019) Proc. of the Proc. of the Assoc. Adv. Anim. Breed. Gene., Armidale, Australia.

Walmsley B.J., Lee S.J., Parnell P.F. and Pitchford W.S. (2018). Anim. Prod. Sci 58(1):1-19. https://doi.org/10.1071/AN12428

Wilmink J.B.M. (1996) INTERBULL Annual Meeting, Veldhoven, the Netherlands.

483. Impact of planning horizon length on the relative emphasis of traits in economic breeding goals

H.F. Valasek[1*], B.L. Golden[2] and M.L. Spangler[2]

[1]University of Nebraska-Lincoln, 1400 R St, Lincoln, NE 68583, USA; [2]Theta Solutions, LLC, 4745 Plover St NE, Olympia, WA 98516, USA; hvalasek2@huskers.unl.edu

Abstract

Unique situations exist in beef production herds that can affect the weighting of traits in an index (e.g. breeding system, economic costs/returns, current phenotypic performance). Given genetic purchasing decisions represent capital investments, the time horizon to determine economic impacts also represents a variable that differs between enterprises. The current study used an economic index construction platform (iGENDEC) to investigate the impact of planning horizon length (PH) on the relative emphasis of traits in the breeding goal for beef production systems. A general-purpose index was constructed for three breeding systems with six different PH (2-50 y). As PH increased, relative emphasis for sale weight (weaning or hot carcass) decreased and stayability increased. Rank correlations between indexes with 2 and 50 yr. PH were 0.72 and 0.64 for weaning and slaughter endpoints, respectively. Results illustrate that correctly determining PH can impact the rank of selection candidates and progress towards improved net profit.

Introduction

The concept of selection indexes as a means of multiple-trait selection towards an aggregate goal is not new (Hazel, 1943) and economic selection indexes have become common currency in making selection decisions across many species. The purpose of selection indexes is to select for maximize genetic improvement for the recognized economic selection criterion (Hazel *et al.*, 1994). Traditionally, the construction of selection indexes relied on several components including knowledge of economic, genetic, and phenotypic parameters that aligned with specified breeding goals. The design of a selection index requires the estimation of relative economic value as this derives the relative emphasis assigned to each trait (MacNeil *et al.*, 1997).

In general, selection indexes can be classified as either terminal or maternal depending on the breeding goal. In some species, such as beef cattle, general purpose indexes offer an opportunity for self-replacing commercial herds to place emphasis on both maternal and terminal traits. These general-purpose indexes could be targeted at different sale endpoints (i.e. weaning or slaughter). Regardless of index type, economically optimized indexes should provide an expectation of changes in net profit due to selection.

Enterprises experience unique situations that could change the relative emphasis of traits in enterprise-specific breeding goals. Differences in current levels of performance are known to impact the relative economic importance of traits (Enns *et al.*, 2006). The planning horizon (PH) assumed by an enterprise represents a complex decision that should be contemplated and requires consideration of breeding objectives and the need for revenue at alternative points in the future. Planning horizons are an instrumental aspect of creating a profitable sire selection program and allows a producer to capture monetary and genetic returns over the length of the PH considering the initial financial situation (McMahon *et al.*, 1985). Variation in PH account for the initial investment of a bull and how quickly a producer may need or want to generate revenue with that animal. The objectives of the current study were to investigate the impacts of different lengths of PH on the relative emphasis of traits in the breeding goal and the ranking of selection candidates.

R.F. Veerkamp and Y. de Haas (eds)
CC BY
Proceedings of 12th World Congress on Genetics Applied to Livestock Production (WCGALP) 2007
DOI: 10.3920/978-90-8686-940-4_483, © H.F. Valasek *et al.* 2022

Materials & methods

The selection indexes evaluated were produced by iGENDEC, a web-based platform for economic index construction in U.S. beef production systems (Spangler *et al.*, 2022). In brief, iGENDEC used a stochastic simulation to construct a cow herd and genetic values of bulls were modified by one unit, one trait at a time, to determine the Marginal Economic Value (MEV) of each trait in the breeding objective. The model accounted for discounted gene flow and discounted expression rates (McClintock and Cunningham, 1974). General-purpose selection indexes constructed assumed replacement females were retained and terminal animals were sold either at weaning or at slaughter. Traits included in the index for the weaning endpoint included weaning weight-direct (WW-D), weaning weight-maternal (WW-M), mature cow weight (MW), stayability (STAY), heifer pregnancy (HP), calving ease-direct (CE-D), and calving ease-maternal (CE-M). Traits included in the index when the endpoint was slaughter included the same traits as the weaning endpoint along with hot carcass weight (HCW), ribeye area (REA), fat depth (FAT), marbling score (MS), and feed intake (FI).

For each index described above, three breeding systems were assumed: Angus bulls bred to Angus cows, bulls and cows that were half Simmental and half Angus, and cows that were half Angus and half Hereford paired with Simmental bulls. These breeding systems were investigated to contrast the impact of different levels of direct (initial values of 0, 50, and 100%) and maternal heterosis (initial values of 0 and 100%). The initial phenotypic means were assumed to be the same in all breeding systems. Six PH lengths (2, 5, 10, 20, 30, and 50 yrs.) were evaluated to assess the implications of PH on the relative emphasis for each trait where relative emphasis was defined as the MEV multiplied by the genetic standard deviation of the trait and then divided by the sum of this product for all traits. Planning horizon represented the number of years simulated with the improved genetic merit of bulls. In total, 36 scenarios were investigated. The economic parameters were assumed the same in all scenarios.

In addition, an evaluation of HCW was conducted to determine the change in relative emphasis as hot carcass weight approached a pre-defined discount threshold of 477 kg as might exist in U.S. beef production systems. This evaluation was conducted using the purebred Angus production system at three different time horizons (2, 20, and 50 yrs.). Hot carcass weight started at 295 kg and increased by 45 kg to reach a final weight of 522 kg.

Using the 2 and 50 yr. PH for the purebred Angus system at both sale endpoints (weaning and slaughter), the indexes were applied to a group of selection candidates (n=27,123) to estimate rank correlations between indexes that differed in PH.

Results

For the indexes that assumed a weaning endpoint (Table 1), as the length of PH increased within, the relative emphasis shifted from WW-D to STAY and MW. Likewise, when the endpoint was slaughter (Table 2), a similar trend was observed as the relative emphasis of HCW decreased and the relative emphasis of STAY increased. The relative emphasis of STAY never exceed that of HCW, however, when weaning weight was the sale endpoint the relative emphasis of STAY was larger than the relative emphasis of WW-D at longer PH.

The Spearman's correlation coefficients between the rank of selection candidates based on a 2 and 50 yr. PH were 0.72 and 0.64 when the sale endpoint was weaning and slaughter, respectively.

Table 1. Comparison of relative emphasis for weaning weight and stayability for different Lengths of planning horizon for an index with a weaning sale endpoint.

	Purebred Angus		Angus-Simmental		Three breed cross	
Time horizon (years)	WW-D	STAY	WW-D	STAY	WW-D	STAY
2	0.873	0.000	0.794	0.000	0.911	0.000
5	0.702	0.016	0.501	0.271	0.555	0.262
10	0.440	0.249	0.402	0.264	0.459	0.179
20	0.337	0.329	0.322	0.310	0.313	0.323
30	0.295	0.353	0.284	0.347	0.268	0.385
50	0.276	0.354	0.263	0.359	0.239	0.411

Table 2. Comparison of relative emphasis for hot carcass weight and stayability for different lengths of planning horizon for an index with a slaughter endpoint.

	Purebred Angus		Angus-Simmental		Three breed cross	
Time horizon (years)	HCW	STAY	HCW	STAY	HCW	STAY
2	0.468	0.002	0.448	0.002	0.431	0.001
5	0.457	0.003	0.431	0.043	0.444	0.001
10	0.419	0.040	0.403	0.076	0.400	0.069
20	0.368	0.149	0.373	0.140	0.348	0.192
30	0.353	0.177	0.353	0.182	0.338	0.210
50	0.339	0.205	0.339	0.210	0.324	0.242

As the mean HCW increased towards the discount threshold the relative emphasis decreased (Table 3). Once HCW surpassed the pricing discount threshold, relative emphasis for HCW weight gradually increased.

Discussion

When the targeted sale endpoint was at weaning, the largest changes in the relative emphasis occurred between the 2 and 5 y PH (crossbred systems) or between the 5 and 10 y horizons (purebred system). Between these time points, the relative emphasis of WW-D decreased and the relative emphasis of STAY increased. When slaughter was considered the sale endpoint the changes in relative emphasis between HCW and STAY were more gradual and the largest differences occurred between the 10 and 20 y PH. Again, the trend was for the relative emphasis for HCW to decrease and the relative emphasis for STAY to increase.

Table 3. Relative emphasis of hot carcass weight for different hot carcass weight means and planning horizons when the discount threshold was 477 kg.

	Time horizon (years)		
Weight (kg)	2	20	50
295	0.560	0.477	0.445
341	0.489	0.398	0.368
386	0.433	0.331	0.303
432	0.271	0.232	0.209
477	0.168	0.148	0.127
523	0.282	0.196	0.174

The shift in relative emphasis from the endpoint weight (weaning or hot carcass) to sustained reproductive traits (i.e. STAY) as PH increased was expected. As PH increased, the return on investment of a genetic purchasing decision shifted from traits expressed sooner. This represents an intuitive interaction between the discounted expression rate and the length of the PH. When the PH was short (e.g. 5 y or less), traits such as STAY may not have been fully expressed in all breeding females, whereas traits such as weight at sale had the opportunity to be expressed sooner. The relative emphasis for sale weight and STAY were not as large when the endpoint was slaughter as they were when the endpoint was weaning. This was an artifact of the slaughter index being comprised of more traits than the weaning index. As an example of the consequences of PH length on selection decisions, rank correlations in the purebred scenario were estimated for both sale endpoints between 2 and 50 y PH. The rank correlations were 0.72 and 0.64 for weaning and slaughter endpoints, respectively, suggesting considerable re-ranking among selection candidates.

The decrease in relative emphasis of HCW as the current mean of HCW increased was expected given a pricing discount threshold existed. Interestingly, once the mean HCW exceeded the discount threshold the relative emphasis began to increase suggesting that once the discount was already incurred there was economic incentive to increase sale weight. The relative emphasis of HCW at an assumed mean HCW decreased as length of PH increased. This illustrates a latent interaction between discounted gene flow and length of the PH.

Length of the planning horizon is an important consideration in constructing economic selection indexes. Further work is needed to formally compare the efficiency of indexes constructed under different planning horizon assumptions.

References

Enns, R.M., Garrick, D.J., and Brigham B.W. (2006) Proc. of the 8th WCGALP, BeloHorizonte, Brazil.

Hazel, L.N. (1943) Genetics 28(6):476-490. https://doi.org/10.1093/genetics/28.6.476

Hazel, L.N., Dickerson G.E., and Freeman A.E. (1994) J Dairy Sci. 77(10):3236-3251. https://doi.org/10.3168/jds.S0022-0302(94)77265-9.ht

MacNeil M.D, Nugent R.A., and Snelling W.M. (1997) Proc. of the 15th Range Beef Cow Symposium, Rapid City, United States of America.

McClintock A.E. and Cunningham E.P. (1974) Anim Sci. 18:237-247. https://doi.org/10.1017/S0003356100022418

McMahon R.T., Blake R.W., Shumway C.R., Leatham D.J., Tomaszewski M.A. *et al.* (1985) J Dairy Sci. 68(9):2295-2302. https://doi.org/10.3168/jds.S0022-0302(85)81102-4

Spangler, M.L., Golden B.L., Newman S., Kuehn L.A., Snelling W.M. (2022) Proc. of the 12th WCGALP, Rotterdam, the Netherlands (in press)

Part 29

Society – gene editing: ethics, economics and experiences

484. Effect of alternative synchronisation treatments on the vaginal microbiota of ewes and their impact on pregnancy rate

E.L. Reinoso-Peláez[1*], O. González-Recio[1], C. González[1], M. Saura[1], A. Fernández[1], A. López-García[1], A. Saborío-Montero[1], J.H. Calvo[2], M. Ramón[3] and M. Serrano[1]

[1]Instituto Nacional de Investigación y Tecnología Agraria y Alimentaria (INIA-CSIC), Ctra. de La Coruña, km 7.5, 28040 Madrid, Spain; [2]Centro de Investigación y Tecnología Agroalimentaria de Aragón (CITA-ARAID-IA2), Av. Montañana 930, 50059 Zaragoza, Spain; [3]Instituto Regional de Investigación y Desarrollo Agroalimentario y Forestal de Castilla-La Mancha (IRIAF), Av. del Vino s/n, 13300 Valdepeñas; edgar.reinoso@inia.es

Abstract

The low pregnancy rate by artificial insemination in sheep represents a fundamental challenge for breeding programmes. In this species, estrous synchronization is done by manipulating hormonal regimens through the insertion of progestogens in intravaginal sponges, which may alter the vaginal microbiota and the success of artificial insemination. In this study, we analysed the microbiome of 94 vaginal swabs from 47 ewes with alternative treatments applied to the progesterone-releasing intravaginal devices, in two sample periods (before and after removing the devices) using nanopore-based metagenome sequencing. A lower abundance of the genera *Neisseria* (Proteobacteria) and *Oenococcus* (Firmicutes) was observed in pregnant compared to non-pregnant ewes. Although differences between treatments were not detected, the progestogen of the synchronization devices seemed to increase the alpha-diversity and decrease the abundance of harmful microorganisms belonging to Gammaproteobacteria and Fusobacteriia classes, suggesting that progestogens may have a beneficial effect on vaginal microbiota favouring the artificial insemination success.

Introduction

Ovine production has a great economic value in Spain, which is the second producer in Europe and the fifth worldwide. Artificial insemination (AI) is the key technology in dairy ruminants breeding programmes for progeny testing, connecting herds and disseminating genetic improvement. However, the efficiency of AI in sheep is low as fertility rates range from 30 to 60%. This low efficiency is due to the particular morphology of the reproductive tract of the ewe, the requirement of using fresh semen and the difficulty to determine the exact stage of the ovulatory cycle at the time of AI. A typical practice for oestrus synchronization is the use of progesterone-releasing intravaginal devices (PRID) consisting on a stainless spiral coil coated with an inert silicone rubber matrix impregnated in flugestone acetate (20 mg), allowing the alteration of ewes hormonal regime. However, our understanding of other factors that may affect the success of AI in sheep, such as the genital tract microbiota, is still low.

In recent years, the study of the composition and abundance of microbial communities (microbiota) has been facilitated by the advances of high-throughput sequencing technologies. In particular, metagenomics has emerged as the reference technique to directly analyse the genomes contained in an environmental sample (microbiome). While there is strong evidence of the role of the reproductive tract microbiota on fertility and sexual disease in humans (e.g. Kindinger *et al.* 2017), its study in livestock species is more limited (Quadros *et al.* 2020; Pascottini *et al.* 2021), and very few in sheep (Serrano *et al.* 2020; Koester *et al.* 2021).

The aim of this study was to elucidate the role of the vaginal microbiota in fertility of Assaf sheep by sequencing the microbial metagenome from pregnant and non-pregnant ewes. In addition, we evaluated the effect of PRIDs supplemented with different additives on the microbiota composition, and whether potential changes in that composition were related with the pregnancy rate. Additionally, the advantages of third-generation

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_484, © E.L. Reinoso-Peláez *et al.* 2022

nanopore sequencing technology (Oxford Nanopore Technologies, ONT) were exploited to study microbial genes and functional pathways (KEGGs, Kyoto Encyclopaedia of Genes and Genomes) involved in fertility.

Materials & methods

Experimental design, DNA extraction and sequencing. A total of 94 vaginal exudate (swabs) samples from 47 ewes from an Assaf flock were obtained in two sampling times. A first vaginal swab was collected from each female, immediately prior to insertion of PRIDs (S1 group). Intravaginal devices were removed 14 days later, and AI was conducted 55 hours after PRID removal. Just before insemination, a second sample of vaginal exudate was taken from each ewe (S2 group) in the same way as the first. At the time of PRID administration the ewes were divided in four batches, depending on alternative treatments added to the PRID (200 mg lyophilized): (1) *probiotic* (n=13), containing *Lactobacillus rhamnosus* and maltodextrine (excipient); (2) *maltodextrine*, the probiotic excipient (n=10); (3) *antibiotic*, containing Framitecin (n=13); and (4) *control*, no additive added in the PRID (n=11). Microbial DNA was extracted using the QIAamp ©DNA Microbiome extraction kit (Qiagen Inc., Valencia, CA, USA). Metagenome sequencing was performed in a MinION device (Oxford Nanopore, Oxford, UK) using the ligation sequencing kit (SQK-LSK109, Oxford Nanopore, Oxford, UK). Pregnancy status was assessed using ultrasonography pregnancy testing 42 days after AI. Notice that all samples (n=94) were used for comparison between sampling times, while only samples from the S2 group (n=47) were used for the analysis of pregnancy outcome.

Bioinformatics and statistical analyses. Base calling was performed with the Guppy 4.2.2 software (ONT). A first filtering of the host genome was performed by mapping the reads against *Ovis aries* reference genome (Oar_rambouillet_v1.0, GCA_002742125.1) using the BBMap software. The retained sequences were analysed using the *sqm*_longreads.pl SqueezeMeta 1.1.0 pipeline software (Tamames and Puente-Sánchez, 2019), which runs a Blastx search of the translated nucleotide reads against GenBank nr, and Kyoto Encyclopedia of Genes and Genomes (KEGG) databases obtaining taxa (phylum, class, and genus) and KEGG. Subsequently, quality filters were applied to remove unclassified taxa and those at a relative abundance threshold <0.1%. Imputation of zero values was performed using a Bayesian multiplicative replacement procedure. Normalization was implemented by applying the Centred Log-Ratio (CLR) transformation. Microbial abundance was evaluated using SQMtools and alpha-diversity by the Phyloseq software. The effect of microbiota on pregnancy rate, as well as the effect of PRID were assessed with PCA, PERMANOVA, and a differential abundance analysis, LIMMA R package. Quality filters and analyses were carried out using R-project (https://www.r-project.org).

Results

Of the 47 inseminated ewes, 26 were pregnant (four in the *probiotic*, seven in the *antibiotic*, six in the *maltodextrine*, and nine in the *control* groups) and 21 were not pregnant (nine in the *probiotic*, six in the *antibiotic*, four in the *maltodextrine* and two in the *control* groups).

No significant differences were observed in microbial alpha-diversity when comparing pregnant and non-pregnant groups, although it was higher in pregnant (56±11.6) than in non-pregnant (47.1±17.2) ewes. Differential abundance analysis revealed that *Oenococcus* (Bacilli) and *Neisseria* (Betaproteobacteria) genera were significantly more abundant in non-pregnant ewes. The comparison of microorganism abundance did not reveal significant differences between probiotic, maltodextrine, antibiotic and control groups, both in the analysis of pregnancy outcome and when comparing sampling times S1 and S2.

Regarding sampling times, alpha-diversity was significant higher in S2 (54.1±14) compared to S1 (39±13.2). Also, significant differences between both groups were found in terms of abundance as revealed by PCA, PERMANOVA and differential abundance analysis for genus, phylum and KEGGs (Figure 1).

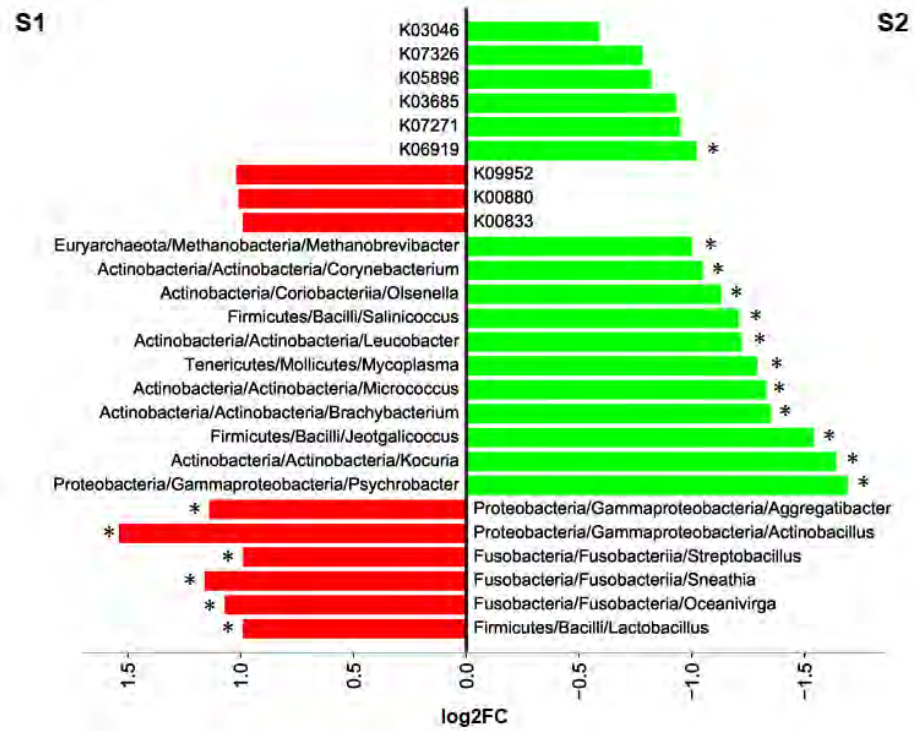


Figure 1. Differential abundance analysis for samples collected before PRIDs placement (S1) and after PRIDs removal (S2), corrected by PRID treatments, at phylum, class, genus and KEGG levels. Asterisks represent adjusted *P*-values at the 5% level. Bars without asterisk correspond to adjusted *P*-values at the 10% level. The x-axis represents the fold-change (FC).

So, Proteobacteria was the most abundant phylum in the S1 group, whereas Firmicutes was the most abundant phylum in S2. At the genus level, *Actinobacillus, Sneathia, Streptobacillus* and *Aggregatibacter* were significantly the most abundant in S1, whereas *Jeotgalicococcus, Kocuria* and *Psychrobacter* were significantly the most abundant in S2 (Figure 1). Concerning functional analysis (genes and genomes), K00800 (L-xylulokinase) and K09952 (CRISPR-associated endonuclease Csn1) KEGGs were significantly more abundant in the S1 group, while K06919 (DNA primase, phage- or plasmid-associated) was significantly more abundant in the S2 group.

Discussion

In this work we have investigated the role of vaginal microbiota in sheep AI pregnancy rate and the effect of different treatments added to PRIDs in its composition and abundance. Our results revealed a significantly higher abundance of the genus *Neisseria* in non-pregnant ewes, and this may be related with the AI failure. Species of this genus (*Neisseria gonorrhoeae*) have been found to trigger ectopic pregnancy and infertility in women (Smolarczyk *et al.* 2021).

In addition, we observed that the progestogen included in the synchronization sponges increased the alpha-diversity of the vaginal microbiota, suggesting a beneficial effect on pregnancy outcome. In line with this, a recent study by Koester *et al.* (2021) evaluating the microbial composition at a pre-breeding time point and after pregnancy testing found increased alpha-diversity associated with pregnancy rate.

Regarding the effect of the PRIDs (and irrespective of the treatment added) on the microbiota composition and abundance, our results showed that the genera *Actinobacillus* (Gammaproteobacteria) and *Sneathia* (Fusobacteriia) decreased in samples collected after the PRID removal (S2). Pathogenic species from these genera (e.g. *Actinobacillus seminis, Sneathia vaginalis* and *Streptobacillus notomytis*) have been associated to infertility in a recent studio by Serrano *et al.* (2020), which may suggest a beneficial effect of the synchronisation treatment for reproductive success. The authors also found that *Mageebacillus* and *Histophilus* were less abundant in the pregnant group, although this discrepancy with our results might be related with the different sequencing methods used (sequencing of the hypervariable regions V3-V4 of the 16S rRNA gene vs metagenome Nanopore sequencing).

From a functional perspective, we detected a higher abundance of sequences corresponding to L-xylulokinase (K00800) in the S2 sampling time. This KEGG is indirectly involved in the process of phage colonization of host bacteria, representing an essential enzyme for phage DNA replication (Wang *et al.* 2020). Its increase after the removal of the synchronization devices could be responsible for the decrease of harmful classes of bacteria such as Gammaproteobacteria and Fusobacteriia. Interestingly, these bacterial classes have more than 10 genes related to helicase and primase activities (K06919), involved in prokaryotic defence against foreign genetic elements, which were found in lower abundance in samples collected after the removal of the sponges (S2).

In conclusion, this study provides new insights about the effect of the progesterone-releasing intravaginal devices used to synchronize estrous on the vaginal microbial composition and abundance. Although preliminary, our results suggest a beneficial effect of the synchronization treatment, contributing to reduce the abundance of some genera associated with reproductive failure. This study also evidences the potential of nanopore sequencing not only to characterise the microbial populations present in a community but also to provide information about genes and pathways involved.

Acknowledgements

This work was funded by Ministerio de Ciencia e Innovación (RTI-2018-096487-R-C33) and the Convenio INIA-GENOVIS CON19-043. The authors gratefully acknowledge the computing time granted by the Centro de Supercomputación de Galicia (CESGA).

References

Kindinger L.M., Bennett P.R., Lee Y.S., Marchesi J.R., Smith A., *et al.* 2017. Microbiome 5:114. https://doi.org/10.1186/s40168-016-0223-9

Koester L.R., Petry A.L., Youngs C.R., and Schmitz-Esser S. 2021. Front. Microbiol. 12:3240. https://doi.org/10.3389/fmicb.2021.745884

Pascottini O.B., Spricigo J.F.W., van Schyndel S.J., Mion B., Rousseau J., *et al.* 2021. PLoS One 16:1–17. https://doi.org/10.1371/journal.pone.0233943

Quadros D.L., Zanella R., Bondan C., Zanella G.C., Facioli F.L., *et al.* 2020. Res. Vet. Sci. 131:1–6. https://doi.org/10.1016/j.rvsc.2020.03.027

Serrano M., Climent E., Freire F., Martínez-Blanch J.F., González C., *et al.* 2020. High-Throughput 9:1–17. https://doi.org/10.3390/ht9030016

Smolarczyk K., Mlynarczyk-bonikowska B., Rudnicka E., Szukiewicz D., Meczekalski B., *et al.* 2021. Int. J. Mol. Sci. 22:2170. https://doi.org/10.3390/IJMS22042170

Tamames J., and Puente-Sánchez F. 2019. Front. Microbiol. 9:3349. https://doi.org/10.3389/FMICB.2018.03349

Wang H., Chan H. H., Ni M.Y., Lam W. W., Chan W.M. M. *et al.* 2020. J. Investig. Dermatol. 140(1): 182-190. https://doi.org/10.1016/j.jid.2019.05.023

485. Modelling of direct-transfer vitrification of bovine embryos

J. Lamy[1*], A. de Wit[1], F. Fotinga-Meiring[2], E. Mullaart[2], L. Zak[3], E. Knol[3], E. Ellen[1] and H. Woelders[1]

[1]Wageningen Livestock Research, Wageningen University & Research, P.O. Box 338, 6700 AH Wageningen, the Netherlands; [2]CRV, Wassenaarweg 20, 6843 NW, Arnhem, the Netherlands; [3]Topigs Norsvin, Research Center, Schoenaker 6, 6641 SZ, Beuningen, the Netherlands; julie.lamy@wur.nl

Abstract

Development of an effective direct transfer vitrification procedure for biopsied *in vitro* derived embryos is necessary for large scale use in the field and increased genetic gain in the cattle industry. One possible reason for insufficient results to date may be the swelling of embryos following warming/transfer. We used mathematical modelling of the volume changes of embryos during and following vitrification protocols, to study the effect of different temperatures and concentrations of cryoprotective agents (CPAs) and non-permeant solutes. The simulations showed no effect of temperature on the level of post-warming swelling at the assumed parameterizations. However, adapted vitrification media with lower concentrations of CPAs and higher non-permeant solutes led to clearly lower swelling after direct transfer, and could potentially be less toxic for the embryos, but also led to a stronger shrinking during vitrification. The alternative media and protocol seem promising, but empirical experiments are needed to validate this approach.

Introduction

Assisted reproductive techniques, such as *in vitro* fertilization and embryo transfer, are used in the cattle industry to decrease the generation interval and increase the genetic gain per generation (Crowe *et al.,* 2021*).* For instance, embryos may be biopsied to obtain a genomic estimated breeding value of each embryo. Embryos can then be frozen and stored in Liquid Nitrogen (LN_2). The use of frozen embryos increased over the past decade and in 2018, a total of 1.1 million of bovine embryos (*in vivo* derived or *in vitro*) were transferred of which 38.2% were frozen (Viana *et al.,* 2019). Direct transfer (DT) cryopreservation protocols allow a straw with an embryo to be thawed and used at the farm without laborious and complicated steps. This is very important for large scale use in the field as it does not require on-site equipment or expertise. Slow-freezing (SF) followed by direct transfer is currently possible and used for bovine embryos (Dochi *et al.,* 2019). Slow freezing at relatively low concentrations of cryoprotective agents (CPAs) allows formation of extracellular ice, which prevents intracellular ice formation. However, survival rates after SF of biopsied *in vitro* produced bovine embryos are lower than with vitrification (Najafzadeh *et al.,* 2021). In contrast, vitrification is fast and doesn't allow any ice formation, but requires high concentrations of cryoprotective agents and high cooling and warming rates. Current vitrification protocols result in good pregnancy rates, but do not allow direct transfer, hence they require one to three washing steps in the field (Ferré *et al.,* 2020). Therefore, DT vitrification methods (no washing steps) need to be developed. Addition of CPAs prior to vitrification, and removal of CPAs after vitrification can lead to stresses caused by massive volume changes, extremely high intracellular concentrations of CPAs and cytoplasmic salts, dehydration of the cellular cytoplasm, and excessive swelling of the cells upon removal of CPAs. We use mathematical modelling to describe transmembrane fluxes of water and CPAs following addition of CPAs prior to vitrification, and removal of CPAs after vitrification. Mathematical modelling allows us to 'look inside cells' as we can accurately calculate not only the volume changes of the cells but also the intra- and extracellular concentrations of water and all solutes at every point in time. The objectives of the present study were to use the mathematical modelling to explore possibilities to mitigate risks of swelling following direct transfer of vitrified bovine embryos. More precisely, we investigated the effect of different concentrations of permeant and non-permeant CPAs in the vitrification solution and of different temperatures after vitrification-warming on post-DT swelling.

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_485, © J. Lamy *et al.* 2022

Materials & methods

General mathematical model. The modeling of osmotic events was done using two parameter (2P) description of the transmembrane fluxes of water and CPA (Kleinhans, 1998). In this description, fluxes of water and CPA are proportional to the transmembrane difference of the chemical potential (ΔM) for water and for CPA, respectively. A full description will be detailed elsewhere (Woelders *et al.,* in preparation). It was assumed that the 'isotonic' (physiological) osmolality is 0.29 Osm/kg of water. Further, in the modelling, the embryo was considered as a single 'cell' that is confined by a single semipermeable outer cell membrane. This assumption is appropriate as the applied membrane permeability coefficients are 'empirical' phenomenological coefficients obtained by fitting actual empirical osmotic volume excursions of embryos. Assumed values for the permeability coefficients and other embryo characteristics were based on reported values for bovine blastocysts, morula's, and earlier embryo stages for various different CPAs (Woelders *et al.*, 2007; Jin *et al.*, 2011). Thus, it was assumed that embryos at isotonic conditions have volume and surface area of a sphere with a diameter (excluding the zona pellucida) of 120 µm, that the embryos surface area remains constant, and that the initial (isotonic) aqueous fraction is 0.8. Further it was assumed that the hydraulic conductivity, or L_p (at 25 °C, in the presence of EG and DMSO) is 1.0 µm/min/atm, that the membrane permeability for permeant CPAs (P_s) is equal for EG and DMSO, and has a value (at 25 °C) of 20 µm/min, and that the activation energy (E_a) of L_p is 6.5 kcal/mol and E_a of P_s is 7.3 kcal/mol. L_p and P_s at other temperatures were calculated as described in Woelders and Chaveiro, 2004.

Simulations. 'Standard' vitrification featured incubation in 'vitrification solution 1' (VS1; base medium with 7.5% ethylene glycol (EG) and 7.5% dimethyl sulfoxide (DMSO)) for 5 minutes, followed by incubation in 'vitrification solution 2' (VS2; base medium with 15% EG, 15% DMSO, 0.69M of sucrose, and 10 mg/ml Ficoll 400,000) for 1 minute, all at room temperature (RT; 22 °C). The osmolality of VS2 was 12 Osmol/kg of water. Direct transfer was then simulated as an immediate return to base medium (isotonic). Subsequently, similar simulations were run using different temperatures during 'direct transfer', as illustrated in Figure 1.

Then, we investigated the effect of the final amount of (permeant) CPAs taken up by the embryos on the level of embryo swelling after 'direct transfer'. This was done by using variations of VS1 and VS2, with reduced concentrations of the permeant CPAs, combined with elevated concentrations of the non-permeant solute sucrose in VS2, with in all cases a final osmolality of VS2 of 12 Osm./kg water, i.e. identical to that of the 'standard' vitrification protocol. In all VS1 and VS2 media the vol% of EG was equal to that of DMSO. Embryos were incubated in VS1 at RT for 10 min, allowing them to approach equilibrium. The molar concentrations of all base medium solutes was kept constant in all VS1 variants. For reasons explained in

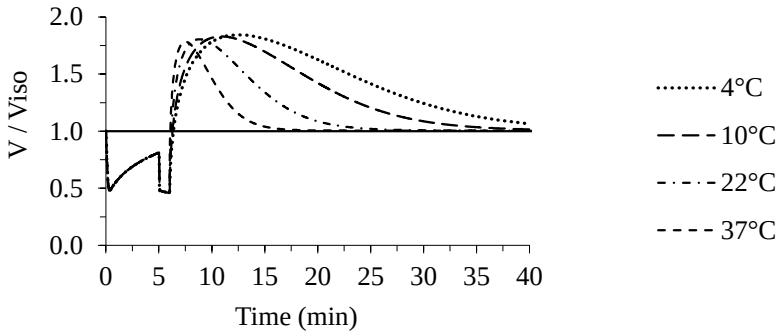


Figure 1. Predicted swelling of bovine embryos in a 'standard' vitrification protocol using different temperatures after warming/transfer. V/Viso is the embryo volume relative to the initial volume (Viso).

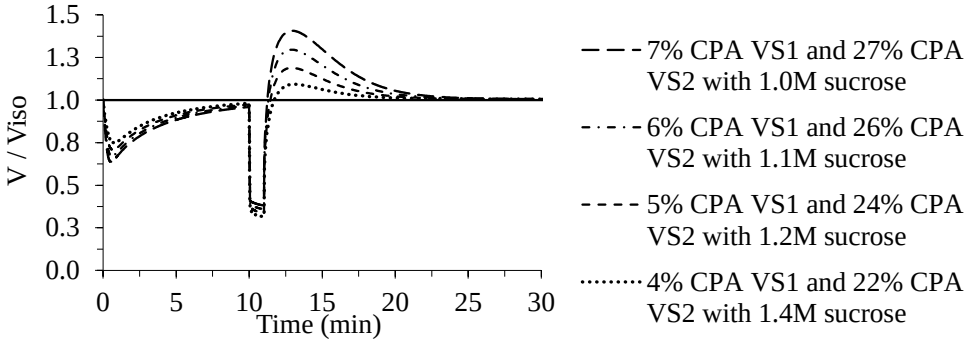


Figure 2. Predicted effect of VS1/VS2 with lower [CPA], with increased sucrose in VS2, all at room temperature. V/Viso is the embryo volume relative to the initial volume (Viso).

the discussion, in all VS1-VS2 combinations the ratio of permeant and non-permeant solutes in VS2 was made equal to that of VS1. VS1 and VS2 concentrations of CPA and sucrose are detailed in Figure 2.

Discussion

In this study, we used a mathematical model to predict volume changes of bovine embryos in different vitrification protocols, with a focus on swelling of the embryos following warming. As is well known, embryos first shrink rapidly after addition of CPA, due to osmotically driven rapid efflux of water, then reswell due to the continued entry of CPA. In two-step vitrification protocols, embryos in VS1 are usually allowed to reswell, i.e. to take up sufficient amounts of CPA approaching equilibrium. In a second (shorter) step in VS2, which has a much higher CPA concentration, the embryos merely shrink due to a rapid efflux of water, which concentrates all solutes inside the embryo. Transfer of embryos after vitrification/warming to a medium with less or with no CPAs is known to lead to strong swelling of embryos above their isotonic volume. Our simulations show that in the 'standard' protocol direct transfer to isotonic medium leads to swelling of the embryos to 1.8 times their initial volume (Figure 1). This represents the 'worst-case scenario', as we assumed immediate and complete mixing with isotonic medium. In reality, embryos placed together with a small volume of VS2 into the uterus will be progressively exposed to isotonic conditions, so swelling will be less. However, the level of swelling is considered to be potentially harmful. We have investigated whether the level of swelling depends on the temperature of the medium to which the embryos are exposed after warming. The rapidity of the swell and shrink cycle strongly depended on temperature, but temperature had only little effect on the peak level of swelling. The latter is due to the fact that at the assumed values of E_a of L_p and E_a of P_s, the rates of water flux and CPA flux had a similar dependence on temperature. If E_a of P_s were substantially higher, peak swelling could be much lower, but one would need to assume an unrealistically high value, four times higher than the value we used, to bring peak swelling down to 1.2 × Viso (data not shown).

Another approach we followed was to reduce the CPA intake of the embryos. In fact, this is what is achieved in protocols with shortened incubation in VS1, with the 'standard' (high) concentration of CPA (Ortiz-Escribano *et al.*, 2018). Here, we used a longer incubation (full equilibration) with a strongly reduced CPA concentration in VS1. The full equilibration allowed us to have control over the amount of CPA taken up by the embryos. The low concentrations of CPA in VS1 (4-7 vol.%, or 0.64-1.1 mol/l) were in the range of what is commonly used in slow freezing protocols, which is considered harmless for embryos during a 10 minute incubation. In VS2, we used elevated sucrose and lower CPA concentrations compared with the 'standard' VS2 such that the ratio of permeant and non-permeant solutes in VS2 was equal to that in VS1. This

ensures that there was no or little net influx or efflux of CPA during the short incubation in VS2. However, a rapid efflux of water was observed in VS2, which brought the intracellular osmolality to the level of the extracellular osmolality (of VS2) within a few seconds. This approach was shown to be highly effective in reducing the peak level of swelling of embryos following direct transfer (to 1.2-1.4× Viso). However, the lower concentration of CPA in VS1 led to stronger shrinking in VS2 (just prior to vitrification), with V/Viso of 0.32-0.38, compared with 0.46 in the standard protocol. The stronger shrinking may pose larger mechanical stress on the embryos, but at the same time, the lower concentrations of CPA used in VS1 and VS2 could be less toxic for the embryo. Empirical experiments are needed to show if these adapted media and protocols will result in adequate embryo survival and whether or not these sufficiently reduce post-warming swelling to allow direct transfer. This study was financially supported by the Dutch Ministry of Economic Affairs (TKI Agri and Food project).

References

Dochi O. (2019) J. Reprod. Dev. 65(5):389-396. https://doi.org/10.1262/jrd.2019-025

Jin B., Kawai Y., Hara T,, Takeda S., Seki S., *et al* (2011) Biol. Reprod. 89(4):87. https://doi.org/10.1095/biolreprod.112.107250

Kleinhans F.W. (1998) Cryobiology 37:271-289. https://doi.org/10.1006/cryo.1998.2135

Najafzadeh V., Secher J.B.M., Pihl M., Ærenlund A., Jørgensen N. *et al.*, (2021) Theriogenology 171:44-54. https://doi.org/10.1016/j.theriogenology.2021.04.020

Ortiz-Escribano N., Bogado Pascottini O., Woelders H., Vandenberghe L., De Schauwer C., (2018) Equine Vet J. 50(3):391-397. https://doi.org/10.1111/evj.12747

Ponsart C., Le Bourhis D., Knijn H., Guyader-Joly C., Otter T. *et al.*, (2013) Reprod. Fertil. Dev.26(1):12-21. https://doi.org/10.1071/rd13328

Viana J. (2020) Embryo Technology Newsletter, v. 38, n.4

Woelders H., Chaveiro A. (2004) Cryobiology 49(3):258-271. https://doi.org/10.1016/j.cryobiol.2004.09.001

Woelders H., Mullaart E., Merton S. (2007) Proc. of 23rd Annual Meeting A.E.T.E. Alghero, Italy.

486. Fertility analysis in stallions: a 35-year retrospective genetic study in the Argentinean Thoroughbred

A. Karlau[1], A. Molina[2], A. Antonini[3] and S. Demyda-Peyrás[1,3*]

[1]CONICET, CCT La Plata, Argetina; [2]Universidad de Córdoba, Córdoba, Spain; [3]Facultad de Ciencias Veterinarias, Universidad Nacional de La Plata, Buenos Aires, Argentina; sdemyda@fcv.unlp.edu.ar

Abstract

Stallions are characterized by large variations in fertility among individuals. Those differences were previously linked, among other causes, to genetic factors. Hereby, we estimated breeding values for two fertility traits (fertility per season (Fert) and matings per foal produced (MpF)) analyzing a large dataset of ≈500,000 matings produced by 5,927 thoroughbred stallions during the last 35 years in Argentina. Pedigree data included 28,724 individuals showing an average Equivalent Complete Generation (ECG) of 6.47. Estimated Breeding Values (EBV) were estimated using a bivariate linear mixed model implemented in the BLUPF90 package. Heritabilities were 0.20±0.008 and 0.17±0.016 in Fert and MpF respectively. The genetic correlation between traits was -0.76. 5,927 individuals showed accuracy values higher than 50% in both traits. Interestingly only 43 individuals were included among the 100 less fertile in both traits, whereas 54 were included among the 100 most fertile. To our knowledge, this is one of the largest studies performed estimating fertility EBV's in stallions.

Introduction

Fertility is one of the key factors in any livestock production system. In horses, this is particularly important, since the species is well known for lower fertility rates in comparison with other livestock species. In addition, stallions are known for a large variability among individuals in terms of fertility, which was linked, among other factors, to genetic causes. However fertility-related traits are particularly difficult to model from a genetic point of view in the species since reliable phenotypic datasets are scarce (Laseca *et al.* 2021).

In stallions, fertility can be analyzed using different parameters. One of them is fertility per season, which is estimated as the number of foals produced by a stallion to the number of mares mated. However, this parameter does not take into account the number of matings necessary to produce a foal, and therefore, it can overestimate the real fertility of the individual. However, it can be solved by estimating the average number of matings necessary to produce each foal. This parameter allows discriminating stallions that need more mating to produce a foal from those that do not; allowing a more accurate value of individual fertility to be estimated. However, to our knowledge, large population studies analyzing stallion fertility are scarce.

The Thoroughbred (TB) is probably one of the oldest horse breeds, being characterized by an impressive racing ability, but also by a reduced genetic variation and increased inbreeding rates (Corbin *et al.* 2010). Furthermore, TBs are clustered from a genetic point of view in maternal families, in which the assignation of each individual to a certain genetic group is determined based on the direct relationship with one of the 43 founder mares of the breed (Hill *et al.* 2002). According to the International studbook, TBs can only be bred by natural mating, without the use of any assisted reproductive technique. But also, all the matings must be reported by the breeder to the breeders association, regardless of whether or not a pregnancy was produced. For that reason, large and reliable reproductive datasets in TBs could be gathered.

Therefore, we modeled and estimated the fertility of 6,713 TB horses using around 509,558 mating records collected during the last 35 years in Argentina.

R.F. Veerkamp and Y. de Haas (eds)
CC BY
Proceedings of 12th World Congress on Genetics Applied to Livestock Production (WCGALP) 2020
DOI: 10.3920/978-90-8686-940-4_486, © A. Karlau *et al.* 2022

Materials & methods

The estimation of fertility phenotypes was made based on 509,558 mating records retrieved from the Studbook Argentino public database. The dataset included matings from 6,713 thoroughbred stallions and 86,106 breeding mares in Argentina, collected from 1986 to 2020. The fertility of the stallions was estimated using two different criteria. First, we determined the fertility per season (Fert) of each stallion as the relationship between the number of foals produced and the number of mares bred during each entire season. In addition, a second estimation of the stallion fertility was determined as the average number of matings per foal produced (MpF) during each season. On average, stallions were active during 5.35 seasons (ranging from 1 to 26).

EBV's for Fert and MpF were estimated using a restricted maximum likelihood (REML) animal bivariate model as follows:

$$y = \mu + age + F + Xb + Zu + e$$

where yi is the vector of each dependent variable, μ is the overall population mean, b is a vector of fixed effects, including *stud*, *size of the stud*, *season*, *number of matings per season* (in quartiles), *age of the stallion* (in yearly classes), and *maternal family* (43 classes); F is the inbreeding of the stallion included as a linear covariate, u is the vector of random effects due to the additive animal genetic effect, and e is the vector of the residual error; whereas X and Z are incidence matrices that relate the fixed and random effects with the dependent variable. Pedigree data of all known relatives (n=32,840) included in the additive genetic relationship matrix was also obtained from the Studbook Argentino. The genealogical information of the individuals averaged 6.47 ECG, with a maximum of 30. The average inbreeding of the individuals was 1.29%, but ranging between 0 and 26.06%. Genetic parameters were estimated in AIREMLF90 software from the BLUPF90 family (Misztal *et al.* 2002). Finally, data clustering and curation, pedigree handling and analysis, as well as the analysis of the results were performed in R, using *optisel*, *data table,* and *tydiverse* packages.

Results

The phenotypic dataset showed an average Fert of 53.15±23.16%. and 2.97±3.51 MpF. In both cases, the standard deviation was large, in comparison with average values. We also have not found any significant correlation (Spearman test, $P>0.05$) between the age of the stallion and Fert and MpF ($P>0.05$, Figure 1a).

However, the effect produced by the maternal family showed significant variations among groups in both traits (Figure 1b).

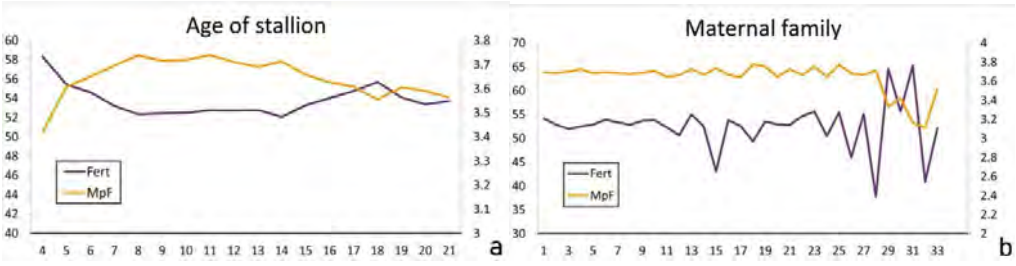


Figure 1. Phenotypic analysis of Fert (left axis in %) and MpF (right axis) clustered by stallion age(a) and family (b).

The genetic analysis showed moderate-to-high heritabilities (in comparison with previous estimations in horse fertility's traits). Fert h^2 was 0.20±0.008 and MpF h^2 was 0.17±0.016. In both cases, the standard deviations were 10 to 15 fold lower than the average values. Moreover, both traits were heavily negatively correlated (-0.76), which is expectable, since fertility is usually reduced when the number of matings per foal increases. It is noteworthy that 5,927 individuals (5,762 with phenotypic values and 165 estimated based on pedigree records), showed an accuracy of 62.87±8.04 for Fert and 61.23±7.53 for MpF.

The average Fert and MpF EVBs per season were estimated during the last 35 years (Figure 2a). the results demonstrated a general increase in the fertility of the breed. Finally, we determined the 1000 individuals showing the highest Fert and lowest MpF EBV's (top fertile individuals). Interestingly, only 660 of them were present in both groups (Figure 2b), showing a negative Spearman correlation of -0.399. However, the same analysis identifying the bottom 1000 less fertile individuals (low Fert and high Mpf) showed only 543 individuals in both groups but the correlation was close to -0.931.

Discussion

The principal challenge in assessing stallions' fertility is to obtain large and reliable phenotypic datasets. However, the precise recording and reproductive management employed in the TB breed make them almost the only exception. In this study, we were able to gather records of 500,000 matings by more than 6,000 individual stallions, occurring in Argentina during the past 35 years. The analysis, performed using a REML linear mixed model, has enabled >5,700 EBV's with an accuracy higher than 50%. To our knowledge, this is one of the biggest and most precise studies carried out to date, to estimate the genetic component of fertility in stallions using field data.

Declining fertility is commonly seen as stallions age (Turner 2019) and is usually due to diminishing testicular function, among other causes. However, we were not able to determine such a relationship in our data. Since TB horses are only bred by natural matings, it is a usual practice that mare and stallions are managed using fine-tuning strategies for improving reproductive outcomes. Among them, the management of aging stallions by reducing their activity and assigning them to mares with proven fertility is one of the most common strategies. Our results suggest the existence of such practices among Argentinean breeders, allowing them to obtain an increased number of foals from racing champions even at older ages.

We also detected differences in fertility of individuals belonging to certain maternal genetic lines. This characterization is considered an important indicator of genetic value in TBs, being commonly used to

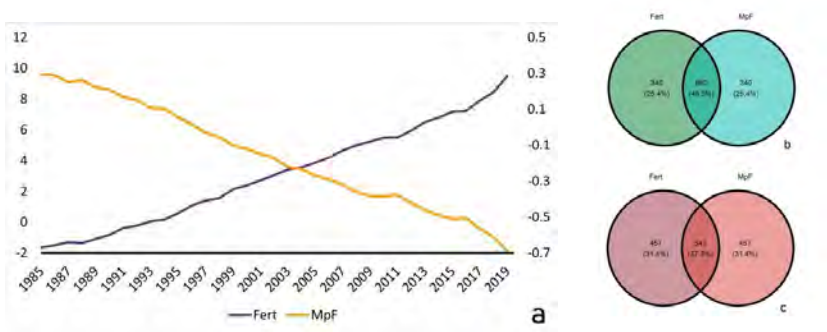


Figure 2. (a) Average EVB estimation per season. Fert (%) in left axis. MpF in number in right axis. On the right, comparison between the top 1000 most (b) and less (c) fertile individuals per trait.

select specific mating combinations. Since the existence of genetic differences among families was largely demonstrated using genomic data (Hill *et al.* 2002), our results suggest that those differences could affect the fertility of the individuals.

Both heritabilities (Fert=0.20 and MpF=0.17)) were high in comparison with previous reports on Spanish breeds (Gómez *et al.* 2020). Such differences could be explained by the differences in reproductive management (fine tunning in TB stallions and extensive breed in PRE), which allows maximizing the expression of genetic differences among individuals. This can also be observed in the genetic progress detected in the EBV's of Argentinean TBs during the last 35 years. Since maternal lineages are important in racing horses, it could be expected that stallions with less fertility and likely to fail to get a valuable mare pregnant are less and less employed, and consequently, the fertility of the breed is improved.

Finally, an interesting finding was the high negative correlation (-0.92) observed between traits in the 543 lowest fertile individuals, compared to a much lower negative correlation (-0.40), in the same analysis of the top 660 most fertile horses. Although further analysis is needed, these results suggest that less fertile stallions are heavily dependent on the number of matings per mare, whereas highly fertile stallions are not.

To our knowledge, this is one of the largest studies assessing stallion fertility from a genetic point of view. We demonstrated the existence of an interesting genetic variability among the Argentinean TB horses in terms of fertility, which showed moderate-to-high heritabilities in comparison to previous reports. Our study will also serve as a reliable comparison point to be employed by further studies estimating the fertility of stallions.

References

Corbin L.J., Blott S.C., Swinburne J.E., Vaudin M., Bishop S.C. *et al.* (2010) Anim. Genet. 41(SUPPL. 2): 8-15. https://doi.org/10.1111/j.1365-2052.2010.02092.x

Gómez M.D., Sánchez M.J., Bartolomé E., Cervantes I., Poyato-Bonilla J. *et al.* (2020) Animal 14(7): 1351-1361. https://doi.org/10.1017/S1751731120000087

Hill E.W., Bradley D.G., Al-Barody M., Ertugrul O., Splan R.K. *et al.* (2002) Anim. Genet. 33(4): 287-294. https://doi.org/10.1046/j.1365-2052.2002.00870.x

Laseca N., Anaya G., Peña Z., Pirosanto Y., Molina A. *et al.* (2021) Animals 11(2). https://doi.org/10.3390/ani11020393

Misztal I., Tsuruta S., Strabel T., Auvray B., Druet T. *et al.* (2002) BLUPF90 and related programs, pp. in *7th World Congress on Genetics Applied to Livestock Production*, Montpellier, France.

Turner R.M. (2019) Anim. Reprod. Sci. 207(171-179. https://doi.org/10.1016/j.anireprosci.2019.06.009

487. Bull effects on *in vitro* embryo production quality traits are repeatable and heritable

G. Gebreyesus[1*], J.B. Secher[2], M.S. Lund[1], K. Kupisiewicz[3], M. Ivask[4] and G. Su[1]

[1]Aarhus University, Blichers Allé 20, 8830 Tjele, Denmark; [2]University of Copenhagen, Hoejbakkegaard Alle 5a, 2630 Taastrup, Denmark; [3]Viking Genetics, Ebeltoftvej 16, 8960 Randers, Denmark; [4]Estonian University of Life Sciences, Kreutzwaldi 62, 51014 Tartu, Estonia; grum.gebreyesus@qgg.au.dk

Abstract

In this study, *in vitro* production of embryos (IVP) was performed using oocytes aspirated from slaughterhouse ovaries and fertilized with ejaculates from 123 bulls aged between 10- and 92 months at semen collection. The restricted maximum likelihood approach was used to estimate genetic parameters for embryo production quality traits including blastocyst rate, hatch rate and an index trait combining Blastocyst rate, Kinetic Score, and Morphology score (BLxMxK). A total of 259 records were available for analysis. The results indicate that bulls at a younger age had better performances in all the studied IVP traits compared to bulls at later ages. The results also show low to high heritability and moderate to high repeatability for the IVP traits, suggesting the possibility of implementing selective breeding of bulls for better IVP performance.

Introduction

Genomic selection (GS) has revolutionized the cattle breeding industry through shortening generation interval, improving the accuracy of selection, reducing cost of breeding programs (Miglior *et al.*, 2017). In this context, combining GS with Assisted Reproductive technologies (ARTs) such as multiple ovulation, embryo transfer and *in vitro* fertilization (IVF), can allow not only improving selection accuracy but also maximizing the intensity of selection (Ponsart *et al.*, 2013; Kasinathan *et al.*, 2015). To the extent that the use of artificial insemination (AI) has allowed rapid and wide-spread use of bulls of superior genetic merit, ARTs can increase the reproduction of superior females (Tonhati *et al.*, 1999)

In vitro production of embryo (IVP) has become an alternative and highly competitive technique for multiple ovulation and embryo transfer (Boni, 2012), leading to increasing interest for harnessing the methodology, alongside GS, for rapid genetic gains in cattle and other livestock breeding. Several factors might affect the success of IVP in cattle including oocyte quality, sperm preparation methods, and media for *in vitro* maturation, fertilization and embryo culture (Akyol *et al.*, 2014). Studies have additionally shown inter-bull differences in IVP performances (Eyestone and First, 1989; Hillery *et al.*, 1990; Shi *et al.*, 1990). However, to our knowledge, no study investigated the extent of genetic control of these differences between bulls in IVP performances. Therefore, the objective of this study was to estimate genetic parameters for bulls' contributions to IVP quality traits.

Materials & methods

IVP experiments. IVP was performed in three different laboratories using the same chemically defined media from IVF Bioscience. All steps were performed according to manufacturer guidelines. In short: for each experiment, oocytes were aspirated from slaughterhouse ovaries, pooled and randomly divided into two or three groups containing at least 50 oocytes. One of the two groups served as control and the same bull was used as a control for all IVP experiments. The other groups were used as test groups, where ejaculates from one of 123 test bulls were used for fertilization. Ejaculates from 123 bulls aged between 10- and 92 months at semen collection were provided by Viking Genetics (Randers, Denmark). For the statistical analysis, age of bulls was categorized into 3 classes: (1) peripubertal (less than 12 mo, mean age =

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_487, © G. Gebreyesus *et al.* 2022

11 months), (2) pubertal (between 12 and 15 mo, mean age = 14 months), and (3) mature (more than 15 mo, mean age = 27 months) bulls. The following traits were assessed at day 7 and 8: (1) Blastocyst rate (Blast) calculated as number of blastocysts over the total number of inseminated oocytes; (2) Hatch rate (Hatch) was defined as the proportion of blastocysts that are hatched. Kinetic obtained by visual classification of each blastocyst as non-expanded, expanded or hatching/hatched, and scored as 1, 2 or 3, respectively; additionally, a Blastocyst rate × Kinetic Score × Morphology score (BL×M×K) (Avery and Greve, 2004) was calculated for each bull. For this calculation, morphology was first defined by scoring blastocysts as poor, good or excellent (scored as 1, 2 or 3). A total of 259 records where available for the analysis. All records for each trait and bull were subsequently normalized to the control bull as:

$$Normalized_{Trait_i} = \frac{Testbull_{trait_i}}{Controlbull_{trait_i}} \quad (1)$$

Where $Testbull_{trait_i}$ is the record for each test bull in trait i while $Controlbull_{trait_i}$ is the corresponding record measured for the control bull from the same experiment and in the same lab as for the test bull.

All bulls in the IVP experiments were genotyped using the Illumina Infinium BovineSNP50 BeadChip (50k). A total of 42,363 SNPs were available after quality control for the analyses.

Statistical analysis. The REML approach in DMU was used to estimate variance components and genetic parameters (Madsen and Jensen, 2013). The statistical model to describe observations of IVP traits is:

$$y = X\beta + Z_1 g + Z_2 c + e \quad (2)$$

Where y is the vector of phenotypes, β is the vector of fixed effects, i.e. bull age class (1-3), semen type (conventional and sexed), as well as month and year of oocyte collection, X is the incidence matrix relating observations with fixed effects, g is the vector of random additive genetic effects of bulls, Z_1 is the incidence matrix relating observations with random genetic effects, c is the vector of permanent environmental effects of bulls, Z_2 is the incidence matrix relating observations with permanent environmental effects, and e is the vector of residual effects. The random effects are assumed to be normally distributed. Thus $\mathbf{g} \sim N(\mathbf{0}, \mathbf{G}\sigma_g^2)$, where G is a genomic relationship matrix and σ_g^2 is the additive genetic variance; $\mathbf{c} \sim N(\mathbf{0}, \mathbf{I}\sigma_c^2)$, where I is an identity matrix and σ_c^2 is the permanent environmental variance; and $\mathbf{e} \sim N(\mathbf{0}, \mathbf{D}\sigma_e^2)$ where D is a diagonal matrix with elements $d_{ij} = \sqrt{(N_{oocyte_{test}} * N_{oocyte_{control}})}$, standardized to have a mean of one, to account for heterogeneous residual variances due to differences in the number of oocytes used in each record in the test and control. Heritability (h^2), was defined as:

$$h^2 = \frac{\sigma_g^2}{\sigma_g^2 + \sigma_c^2 + \sigma_e^2} \quad (3)$$

Repeatability (t) was defined as:

$$t = \frac{\sigma_g^2 + \sigma_c^2}{\sigma_g^2 + \sigma_c^2 + \sigma_e^2} \quad (4)$$

Results

IVP traits across bull age. Figure 1 presents box plots of phenotypic values for the studied IVP traits in the 3 age classes. Bulls at younger age had significantly higher performances in all the studied IVP traits, except hatch rate at day 7 (Hatch_D7), compared to bulls at later ages.

Heritability and repeatability estimates. Estimates of phenotypic variance, heritability, and repeatability in the IVP traits are presented in Table 1. Generally, low to high heritability and moderate

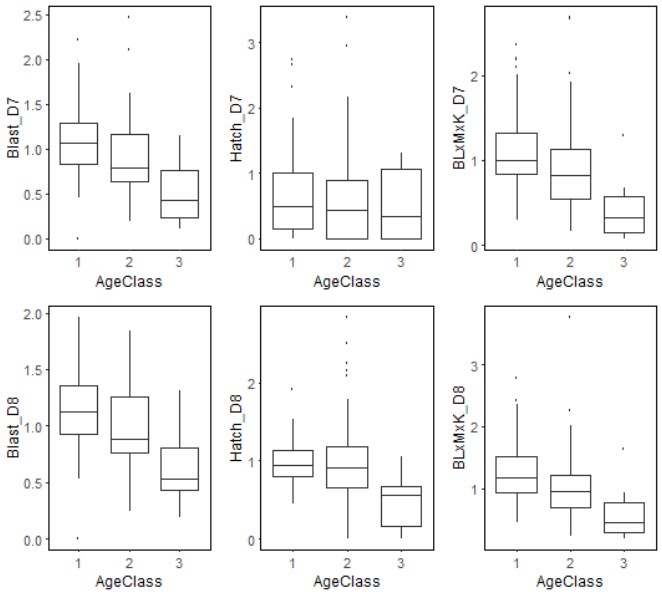


Figure 1. Plots of phenotypic measurements in IVP traits against age of bulls.

to high repeatability were estimated for the IVP traits. However, all estimates showed high standard errors. Compared to the IVP traits measured at day 7 after fertilization, relatively higher heritability and repeatability estimates were observed for the measurements on day 8.

Discussion

Blastocyst rates, morphology and kinetics have long been suggested as reliable indicators of the quality of an IVP system (Avery and Greve, 2004). Understanding the genetic underpinnings of these traits would allow implementation of selective breeding on bulls for enhanced performance in *in vitro* production of embryos. Despite high standard errors across estimates, our study suggests potentially substantial genetic control of the bull's contribution to the indicators of IVP success. Further studies with more sample size might be needed to validate the parameter estimates with probably lower standard errors. The study also reports significant effects of bull age on the IVP quality parameters. Effect of bull age have similarly been reported for male fertility related traits such as semen quality parameters (Olsen *et al.*, 2020; Gebreyesus *et al.*, 2021). Semen from younger bulls appear to result in higher blastocyst and hatch rates as well as

Table 1. Phenotypic variance (σ_p^2), heritability (h^2, with standard errors) and repeatability (t, with standard errors) for IVP traits.

Trait	σ_p^2	h^2 (SE)	t (SE)
Blast_D7	0.14	0.08 (0.25)	0.46 (0.26)
Blast_D8	0.12	0.19 (0.27)	0.54 (0.27)
Hatch_D7	0.44	0.10 (0.32)	0.79 (0.33)
Hatch_D8	0.14	0.19 (0.21)	0.28 (0.22)
BL×M×K_D7	0.20	0.18e-5 (0.26)	0.55 (0.27)
BL×M×K_D8	0.24	0.46 (0.29)	0.61 (0.30)

BL×M×K score at day 7 and 8 after fertilization. These results can provide useful inputs towards planning successful IVP systems.

References

Akyol N., Kizil S.H., Satilmis M. and Karb_Ahin T. (2014) Kafkas Universitesi Veteriner Fakultesi Dergisi 20(4):561-564. https://doi.org/10.9775/kvfd.2013.10582

Avery B., and Grève T. (2004) Reproduction, Fertility and Development 16(2):194-194. https://doi.org/10.1071/RDv16n1Ab144

Boni R. (2012) Animal reproduction 9(3):362-369.

Eyestone W.H., and First N.L. (1989) Theriogenology 31(1):191. https://doi.org/10.1016/0093-691X(89)90599-2

Gebreyesus G., Lund M.S., Kupisiewicz K. and Su G. (2021) J.Dairy Sci. 104(9):10010-10019. https://doi.org/10.3168/jds.2021-20403

Hillery F.L., Parrish J.J. and First N.L. (1990) Theriogenology 33(1):249. https://doi.org/10.1016/0093-691X(90)90673-H

Kasinathan P., Wei H., Xiang T., Molina J.A., Metzger J. *et al.* (2015) Sci Rep. 2(5):8674. https://doi.org/10.1038/srep08674

Madsen P., and Jensen J. (2013) A User's Guide to DMU, pp. Aarhus University.

Miglior F., Fleming A., Malchiodi F., Brito L.F., Martin P.M. *et al.* (2017) J.Dairy Sci. 100(12):10251-10271. https://doi.org/10.3168/jds.2017-12968

Olsen H.B., Heringstad B. and Klemetsdal G. (2019) J.Dairy Sci. 103:545–555. https://doi.org/10.3168/jds.2019-17291

Ponsart C., Le Bourhis D., Knijn H.M., Fritz S.B., Guyader-Joly C. *et al.* (2013) Reproduction, fertility, and development 26(1):12-21. https://doi.org/10.1071/RD13328

Shi D.S., Lu K.H. and Gordon I. (1990) Theriogenology 33:324. https://doi.org/10.1016/0093-691X(90)90748-I

Tonhati H., Lôbo R.B. and Oliveira H.N.D. (1999) Theriogenology 51(6):1151–1156. https://doi.org/10.1016/S0093-691X(99)80018-1

488. Five years of optimizing the assisted reproduction protocol for European eel: what worked and what didn't?

A.P. Palstra[1*], P. Jéhannet[1], L.T.N. Heinsbroek[2] and W. Swinkels[3]

[1]Wageningen University & Research, P.O. Box 338, 6700 AH Wageningen, the Netherlands; [2]Wageningen Eel Reproduction Experts B.V., Mennonietenweg 13, 6702 AB Wageningen, the Netherlands; [3]Palingkwekerij Koolen B.V., Hongarijesedijk 12, 5571 XC Bergeijk, the Netherlands; arjan.palstra@wur.nl

Abstract

The production cycle of the European eel needs to be closed in order to supply aquaculture with juvenile glass eels and to alleviate fishery pressure on the natural population, thereby contributing to its recovery. Currently, we are able to produce larvae batches three times per week. Over the past five years we have executed experiments aiming to condition glass eels into high quality brood stock females and optimize the artificial reproduction protocol. Feminization of young juveniles contributes to shortening the generation time at least 5-fold. Simulated migration induces early sexual maturation. Steroid implants containing 17 methyltestosterone (17MT) and 17β estradiol (E2) induce the more advanced maturation stages and shorten the stressful period of weekly carp pituitary extract (CPE) injections to fully mature females. Instead of using CPE, stable eel-specific recombinant gonadotropins have been successfully applied to produce eel larvae.

Introduction

The life cycle of the European eel is fascinating. Eels can reside in the continental fresh waters for decades. Then, at a largely unspecified moment in autumn, they stop feeding, start swimming and only then does puberty commence. Still in pre-pubertal condition, the eel disappears into the ocean, and swims for up to 6,000 km. When arriving at the spawning grounds in the Sargasso Sea, each female eel consists of approximately 50% of ovarian tissue and is ready to spawn several million eggs. It is generally assumed that parents die shortly after spawning. However, most of the available knowledge comes from the laboratory (e.g. Palstra *et al.*, 2020), as there are no records of maturing or spawning European eels in the ocean, nor are there tracks of their carcasses. Since the eel cannot currently be propagated and raised to the juvenile glass eel stage, each large European eel, including those from the farm, once hatched in the Sargasso Sea.

In 2016, the Eel Reproduction Innovation Centre (EELRIC) was launched, as a collaborative initiative of Wageningen Livestock Research with the Dutch eel sector, united in the sustainable eel foundation DUPAN. The major aim of EELRIC is to close the production cycle of the European eel in order to supply eel aquaculture with juvenile glass eels and to alleviate fishery pressure on the natural population, thereby contributing to its recovery. Now, 5 years later, we are able to produce larvae batches from different females three times per week although larval quality is still often poor as evidenced by high mortalities during the first week and occurrence of deformities. Over the past five years we have executed many experiments aimed to condition glass eels into high quality brood stock females and optimize the artificial reproduction protocol. An overview is presented of the methods that worked and those that didn't.

Materials & methods

Feminization was applied as developed for Japanese eel by Chai *et al.* (2010) and involved feeding with E2 coated pellets for 5-7 months during the early elver stage. The simulated migration procedure was originally described by Mes *et al.* (2016; Figure 1). Eels were subjected to ~3,000 km simulated migration, and in combination with a single CPE injection (20 mg/kg), or a dopamine antagonist implant of eticlopride (Jolly *et al.*, 2016). In contrast to the routine protocol of weekly human chorionic gonadotropin (hCG) injections,

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_488, © A.P. Palstra *et al.* 2022

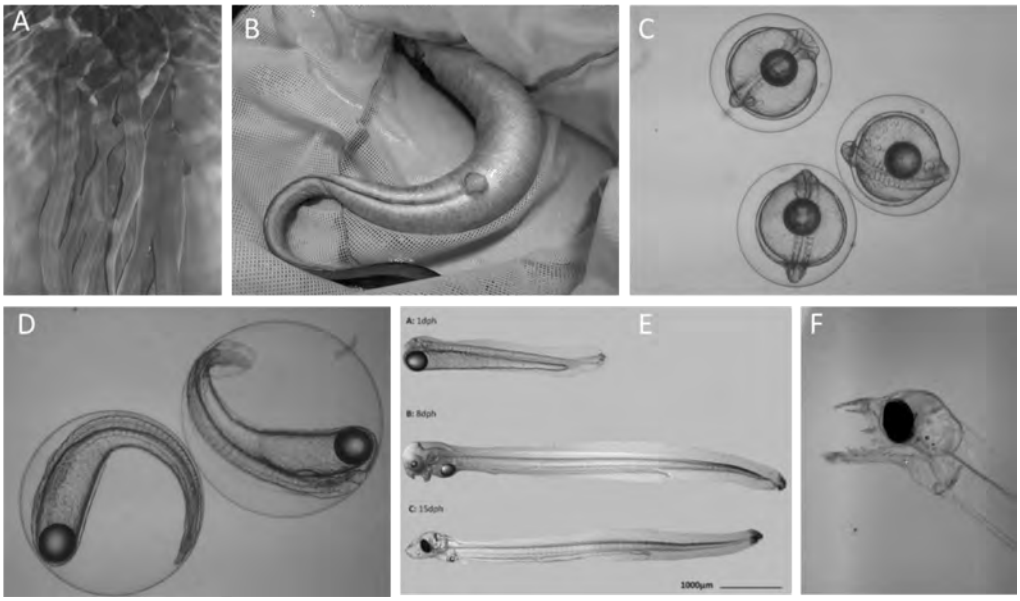


Figure 1. Photo panel showing (A) eels swimming during simulated migration; (B) mature eel with extruding egg bulb; (C) early eel embryos; (D) late eel embryos just before hatching; (E) eel larvae at 1-, 8- and 15-days post hatching (Jéhannet *et al.*, 2021); (F) eel larva at 15 days post hatching ready to initiate exogenous feeding.

males were matured by a single hCG injection (1000 IU; Kahn *et al.*, 1987) followed by a 2nd injection 24 h before stripping. Females were matured by weekly CPE injections (20 mg/kg) followed by a CPE booster at ~10% body weight increase and a single 17α,20β-dihydroxy-4-pregnen-3-one (DHP; 2 mg/kg) injection at ~20% body weight increase (Palstra *et al.*, 2005). The 17MT (5 mg), E2 (2 mg) and the combined 17MT+E2 implants were tested as pre-treatments for the weekly CPE injections (Thomson-Laing *et al.*, 2019). Three experiments have been executed testing weekly injections with recombinant FSH (recFSH: 6 or 12 µg) and LH (recLH: 10 or 20 µg) replacing CPE treatment.

Results

Feminization provides 99% females of which 90% reach 300 g in 12-30 months and can then be used as brood stock.

Simulated migration makes these feminized eels silver (larger eyes reflecting the oceanic phenotype) and induces early sexual maturation (increased Gonadosomatic Index – GSI). Simulated migration in combination with a single CPE injection increased the GSI further but not up to values indicating the onset of vitellogenesis (yolk deposition in the oocytes). Eticlopride was expected to lift the dopaminergic inhibition of gonadotropin production and release but did not show any additional effect on sexual maturation to simulated migration.

Both 17MT and E2 implants significantly increased GSI and oocyte diameter and decreased the number of weekly CPE injections required to mature the eels. The combined implants worked synergistically in advancing vitellogenesis (GSI of 6.6).

Eels that were treated with recFSH followed by recLH matured after 15-22 weeks, similar to eels that were treated with recFSH followed by CPE treatment, but different from eels treated with CPE of which most matured after 7-9 weeks. Larvae were produced from eels of all treatments (Figure 1).

Discussion

The feminization procedure also works for European eel and generation time is shortened by at least 5-fold. Simulated migration can be applied for the natural triggering of early sexual maturation but does not stimulate the more advanced maturation stages, even in combination with a single CPE injection or dopamine antagonist implant. 17MT+E2 implants do stimulate development to more advanced, vitellogenic stages of oocyte development. These implants can be considered as a pre-treatment to reduce the period of stressful weekly hormonal injections although 17MT implant treatment, followed in time by E2 treatment, may result in a more natural progression of steroid-mediated effects. For the first time, European eel larvae were produced with recombinant gonadotropins, but dose and durations still need optimization.

Acknowledgements

The projects received funding from the DUPAN foundation; The Dutch Ministry of Economic Affairs and the European Union; European Maritime and Fisheries Fund. The authors acknowledge students Thijs Böhm, Erwin Choy, Safia Balvet and Lotte Bouwman for their contributions to specific studies in this overview, and value the strong collaborations with Mark Lokman (University of Otago), Ignacio Giménez Nebot (Rara Avis) and Ron Dirks (Future Genomics Technologies BV).

References

Chai Y., Tosaka R., Abe T., Sago K., Sago Y. *et al.* (2010) Aquaculture Sci. 58:269-278. https://doi.org/10.11233/aquaculturesci.58.269

Jéhannet, P., Palstra, A.P., Heinsbroek, L.T.N., Kruijt, L., Dirks, R.P. *et al.* (2021) Animals 11:1710. https://doi.org/10.3390/ani11061710

Jolly, C., Rousseau, K., Prézeau, L., Vol, C., Tomkiewicz, J. *et al.* (2016) Journal of neuroendocrinology 28(9). https://doi.org/10.1111/jne.12411

Khan, I.A., Lopez, E., and Leloup-Hâtey, J. (1987) Gen. Comp. Endocrinol. 68(1), 91-103. https://doi.org/10.1016/0016-6480(87)90064-5

Mes, D., Dirks, R.P., and Palstra, A.P. (2016) Aquaculture 452, 367–372. https://doi.org/10.1016/j.aquaculture.2015.11.020

Palstra, A.P., Cohen, E., Niemantsverdriet, P., van Ginneken, V., and van den Thillart, G.E.E.J.M. (2005) Aquaculture 249 (1-4): 533-547. https://doi.org/10.1016/j.aquaculture.2005.04.031

Palstra, A.P., Jéhannet, P., Swinkels, W., Heinsbroek, L.T.N., Lokman, P.M., *et al.* (2020) Sci. Rep. 10: 2339. https://doi.org/10.1038/s41598-020-59331-6

Thomson-Laing, G., Damsteegt, E.L., Nagata, J., Ijiri, S., Adachi, S., *et al.* (2019) Biol. Reprod. 100, 1319–1332. https://doi.org/10.1093/biolre/ioz007

489. Random regressions for modelling semen production traits in HF purebred and crossbred bulls using Bayesian framework

V.B. Ambike[1*], R. Venkataramanan[2], S.M.K. Karthickeyan[1], K.G. Tirumurugaan[3], K.G. Bhave[4] and M. Swaminathan[4]

[1]Madras Veterinary College, Chennai 600 007, India; [2]The office of Vice-Chancellor, TANUVAS, Chennai 600 051, India; [3]Zoonosis Research Laboratory, Chennai 600 051, India; [4]BAIF, Pune 412 202, India; vrindaambike@gmail.com

Abstract

Semen traits including ejaculate volume (EV, ml), sperm concentration (SC, 10^9/ml), initial sperm motility (ISM, %) and post-thaw motility (PTM, %) were analysed using random regression models. Data pertaining to 105,321 ejaculates from 297 Holstein-Friesian (HF) purebred and crossbred bulls collected from Bharatiya Agro Industries Foundation were utilized. 200,000 Gibbs samples were generated with a burn-in of 20,000 and thinning interval of 50 in a Bayesian framework. Legendre Polynomials with orders of fit up to five for additive and permanent environmental effects were used. Heritability estimates ranged from 0.2 to 0.5 for EV and 0.35 to 0.5 for SC in purebreds while it ranged from 0.05 to 0.45 for EV and 0.2 to 0.7 for SC in crossbreds. Estimates were very low for motility traits in both purebred and crossbred bulls. Random regression models in a Bayesian framework were useful for obtaining solutions for mixed models even with large number of parameters.

Introduction

The HF breed has been extensively used for crossbreeding to increase milk production. Since most of the selection is for milk yield, it is important to assess the semen parameters periodically to maintain optimum fertility. The Bayesian framework provides the advantage of robustness with less stringent assumptions about the distribution of parameters.

Random regression models have the ability to model the covariance of repeated data (longitudinal data) measured on the same animal through covariance structures. This ability to model a trait with age as a control variable helps in understanding the trajectory of various genetic parameters over the age of the bulls. Worldwide there are very few studies conducted using random regression models for semen traits in cattle (Al-Kanaan *et al.* 2015; Carabaño *et al.* 2007). The present study was carried out to understand the genetic variability of various semen traits over age the age of the bulls in HF purebred and crossbred cattle.

Materials & methods

Data pertaining to 105,321 ejaculates from 120 purebred and 177 crossbred HF bulls collected from Bharatiya Agro Industries Foundation (BAIF) were utilized for this study. The BAIF frozen semen stations are located at Uruli Kanchan (Maharashtra, India; 18.5°N and 73.8°E) and Dharouli (Haryana, Jind; 29.2°N and 76.2°E). Semen production data were available for 11 years (2010 to 2020) along with the pedigree information for 1,317 bulls. The founder population was imported between 1967-1974 from Canada, the USA, Denmark, and the Netherlands, where these bulls were proven through progeny testing.

Traits and factors included. Semen production traits included in the study were ejaculate volume (EV, ml), sperm concentration (SC, 10^9/ml), initial sperm motility (ISM, %) and post-thaw motility (PTM, %). The fixed effects included in the model were location (Pune or Jind), breed of bull (purebred or crossbred), order of ejaculate (first or second), the season (summer, monsoon, or winter), the year (2010 to 2020), age

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_489, © V.B. Ambike *et al.* 2022

of the bull (≤2 to >10 years), collection interval (≤2, 3, 4, ≥5 days) and collection time (5 AM to 12 PM), while semen collector and bull were included as random effects.

Statistical analysis. Gibbs sampling using Markov Chain Monte Carlo (MCMC) procedure was implemented to generate these samples which enabled the estimation of parameter estimates based on the given data set. 200,000 Gibbs samples were generated for each analysis with a burn-in of 20,000 and thinning interval of 50 Gibb samples. The stability of the model was ascertained using trace plots, which were obtained from post-Gibbs samples, of the parameters plotted along with the number of iterations.

The age of the bulls in days was taken as the control to obtain the trajectories of traits over the age of the bulls using random regression models. The various combinations of animal and permanent environment random effects were modelled using Legendre polynomials up to order of fit of five with error variances modelled as a homogenous class of random effect. Model selection was done based on Deviance Information Criterion (DIC).

$$Y_{ij} = Xb + \sum_{k=0}^{k_a-1} Z_a a_k + \sum_{k=0}^{k_p-1} Z_p p_k + C_m + e \quad (1)$$

Where Y_{ij}: semen trait of i^{th} animal at j^{th} age, b: vector of significant fixed effects or incidence matrix relating fixed effects with Y, a_k and p_k: set of n values (n = number of bulls) of k random regression coefficients corresponding to animal and permanent environment effects, with the order of fit k_a and k_p respectively, Z_a and Z_p are incidence matrices with dimensions $n \times k_a$ and $n \times k_p$ respectively, C_m: random effect of semen collector with mean zero and variance σ_m^2 and 'e' is the random residual heteroscedastic error variance. The model also included fixed regression for age with the order of that of the animal effect. The genetic (co) variance between ages was estimated from the matrix of random regression coefficients using the general expression

$$G_{xj} = \Phi_{xj} K_{xj} \Phi'_{xj} \quad (2)$$

Where G_x is the (co)variance matrix for x = animal or individual permanent environment and Φ_{xj} is the vector of Legendre polynomials for the random effect of x and j^{th} age group.

All analyses were done in a Bayesian framework using the BLUPF90 family of software (Misztal *et al.* 2018). THRGIBBS1F90 was used for generating Gibbs samples and Post-Gibbs analyses were done with POSTGIBBSF90.

Results

The trajectory of performance for semen production traits in HF purebred and crossbred bulls along with the mean and number of observations are represented in Figure 1 and Table 1. The variation in mean increased in all the traits after 90 months of age. The phenotypic characteristics for all traits in both breeds were comparable. Ejaculate volume showed an increasing trend with age up to 70 months, after which there was a slight decline, while sperm concentration showed a complementary trend in both breeds. Motility traits followed a consistent trajectory, but ISM was more variable than PTM.

Heritability and repeatability were unusually high for the initial and later periods which included ages beyond 36 to 120 months. As seen from Figure 1, the number of observations were lower and random regression models are also susceptible to the effect of extreme ages also known as 'end effect of the polynomials' (Meyer 2005). Hence, all the ages were presented for the period between 36 to 120 months after pruning off the extreme ages to obtain a relatively stable curve. In general, the genetic parameters had

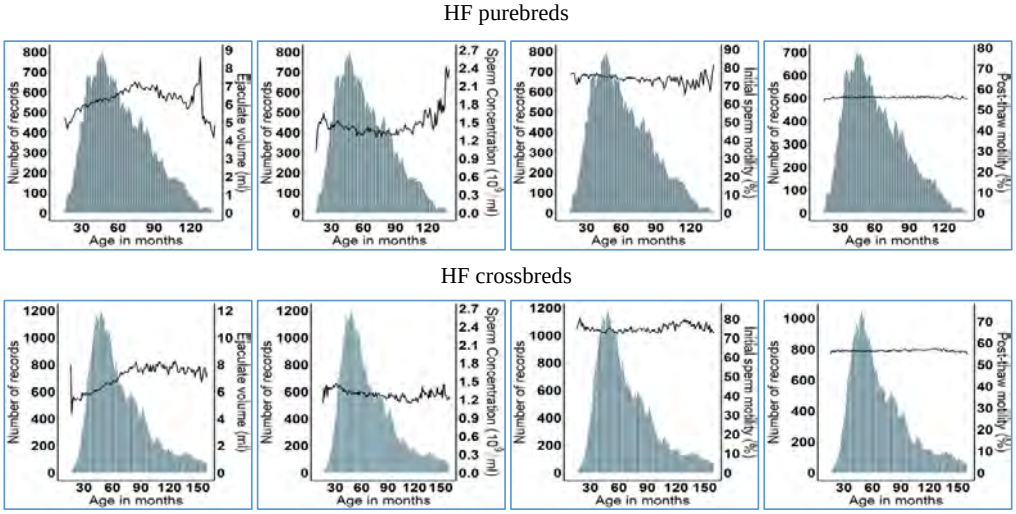


Figure 1. Semen production traits plotted with number of observations and mean across age in months.

Table 1. Range for the number of observations, mean, heritability and repeatability estimates in HF purebred and crossbred bulls.[1]

Trait	Range for n	Range for mean	Order of fit	Range for h^2	Range for r
HF purebreds					
EV	82-797	5.771-7.234	5×5	0.189-0.420	0.384-0.612
SC	82-797	1.261-1.713	5×5	0.344-0.498	0.614-0.827
ISM	81-794	65.423-77.000	5×5	0.035-0.188	0.073-0.339
PTM	66-699	55.160-56.770	5×5	0.000-0.001	0.001-0.002
HF crossbreds					
EV	153-1,192	5.750-8.227	5×5	0.036-0.460	0.113-0.901
SC	154-1,192	1.408-1.142	5×5	0.225-0.700	0.484-0.981
ISM	154-1,172	72.203-78.669	5×5	0.004-0.793	0.392-0.984
PTM	145-1,043	55.972-57.355	5×4	0.000-0.016	0.001-0.022

[1] n = number of records; EV = ejaculate volume (ml); SC = sperm concentration (10^9/ml); ISM = initial sperm motility (%); PTM = post-thaw motility (%); order of fit = legendre polynomial (additive × permanent environmental); h^2 = heritability; r = repeatability.

more uniformity in purebreds while those in crossbreds showed more waviness with a wider range, which could be attributed to less number of observations during nine to ten years of age in crossbreds compared to purebred bulls.

Legendre polynomials of order five for both additive and permanent environmental variance were suitable. Values of random regression models for heritability were more than those obtained through univariate analysis (Figure 2) carried out using a repeatability model with the bull effect included as the random effect. In purebreds, heritability for EV increased until the age of eight years and then decreased again, while in crossbreds, it increased until the end after a plateau in the middle. In purebreds, SC plateaued during the early ages, followed by a slight decline and a short crest, while in crossbreds, the dip and the ascent were more prominent. Motility traits, which are indicators of fertility, had low estimates of heritability. In

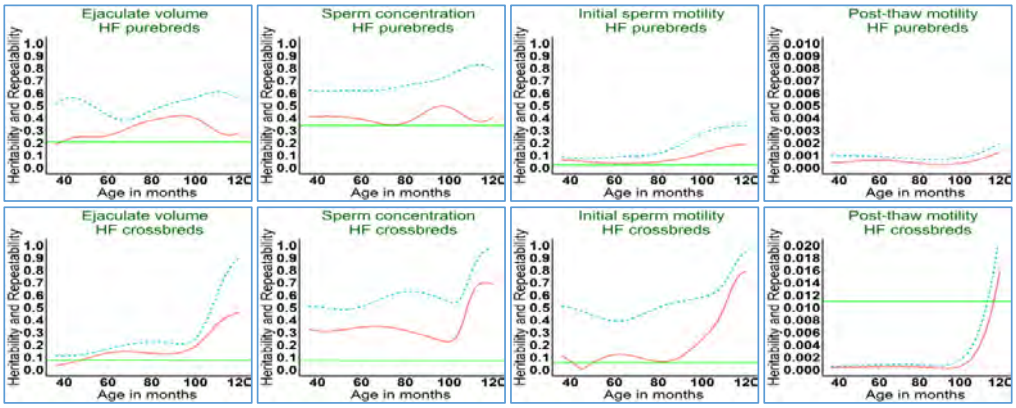


Figure 2. Heritability and repeatability estimates over the ages for semen production traits obtained through random regression models (Red line – Heritability; blue line – repeatability; green line- heritability obtained through univariate analysis).

purebreds, the trajectory was very uniform where repeatability closely resembled heritability. In crossbreds, ISM showed a wide variation in heritability and repeatability, whereas PTM was similar to that of purebreds.

Discussion

The semen production traits for HF purebreds and crossbreds observed in this study were better compared to estimates reported from other parts of the tropics (Gopinathan *et al.* 2018; Kawakami *et al.* 2016). This indicated optimum management practices followed at BAIF.

Precise estimates which were within the range of the estimates obtained in different studies from tropical countries (Atagi *et al.* 2017; Hagiya *et al.* 2017; Kawakami *et al.* 2016) were obtained for the genetic parameters using the Bayesian framework as seen from the low posterior standard deviations. The heritability of motility traits was low which warrants additional information from the relatives in selection programmes. In traits, like SC and ISM, the wide gap between heritability and repeatability could be due to the permanent effect of the environment on these traits. The order of fit used in the best model for the additive and permanent environmental effects were similar to the ones used by Carabano *et al.* (2007) in EV.

Higher estimates of heritability obtained through random regression models compared to univariate analyses could possibly indicate that the modelling of (co)variance functions in the form of random regression coefficients helped in accounting for a greater proportion of the additive genetic effect. As the semen traits are repeatable in nature, pedigree information accumulates over the lifetime of the bulls, which may explain the variation between the early and later stages of a bull's life.

To conclude, random regression models were successfully used to model semen traits over the age of the bull. Purebred HF had higher heritability estimates compared to the HF crossbred animals. The traits such as motility which are directly related to fertility were found to have very low heritability.

References

Al-Kanaan A., König S., and Brügemann, K. (2015) Livest. Sci. 177: 15–24. https://doi.org/10.1016/j.livsci.2015.04.003

Atagi Y., Onogi A., Kinukawa M., Ogino A., Kurogi K., *et al.* (2017) J. Anim. Sci. 95(5): 1900-1912. https://doi.org/10.2527/jas.2016.1186

Carabaño M.J., Díaz C., Ugarte C., and Serrano M. (2007) J. Dairy Sci. 90(2):1044-1057. https://doi.org/10.3168/jds.S0022-0302(07)71591-6

Gopinathan A., Sivaselvam S.N., Karthickeyan S.M., and Kulasekar K. (2018) Indian J. Anim. Reprod. 39(1):24-27 https://doi.org/10.20546/ijcmas.2018.711.370

Hagiya K., Hanamure T., Hayakawa H., Abe H., *et al.* (2018) Animal 12(10): 2027-2031 https://doi.org/10.1017/S1751731117003470

Kawakami J., Hanamure T., Hagiya K., Hayakawa H., *et al.* (2016) Anim. Sci. J. 87:101-106

Meyer K. 2005. Genet. Sel. Evol. 37(6):1-28 https://doi.org/10.1186/1297-9686-37-6-473

Misztal I.S., Tsuruta D.A.L., Lourenco. Y., Masuda I., *et al.* (2018) BLUPF90 http://nce.ads.uga.edu/wiki/doku.php?id=documentation

490. Removing the dopaminergic inhibition with CRISPR-Cas9 to study the progression of sexual maturation in zebrafish

P. Jéhannet[1*], A.A.C. De Wit[1], S. Brugman[1], Y. Zhang[2], J. Bogerd[2] and A.P. Palstra[1]

[1]Wageningen University & Research, P.O. Box 338, 6700 AH Wageningen, the Netherlands; [2]Utrecht University, Heidelberglaan 8, 3584 CS Utrecht, the Netherlands; pauline.jehannet@wur.nl

Abstract

In fish, dopaminergic neurones originating from the preoptic area innerve the region of the pituitary gland that harbors gonadotropin cells. The innervation inhibits the release of gonadotropins that are essential for sexual maturation. The dopaminergic inhibition, mediated by dopaminergic receptors, is a major bottleneck in the propagation of many fish species in aquaculture. In this study, we aimed to knock-out one of the three dopamine receptors paralogue (*drd2b*) that mediates the inhibitory control of gonadotropins in zebrafish *Danio rerio* to study the progression of sexual maturation and potential occurrence of side-effects. Here, *drd2b* heterozygote or mosaic mutants were produced with an efficiency up to 73%. Knocking-out *drd2b* could be lethal since *drd2b* is involved in crucial processes during early ontogeny. If future studies confirm that inactivating *drd2b* is lethal, it may be useful to produce inducible and/or spatially restricted mutants to study sexual maturation.

Introduction

In fish, oogenesis is controlled by two pituitary gonadotropins (Gths); the follicle stimulating hormone (Fsh) and the luteinizing hormone (Lh). Fsh and Lh regulate oogenesis consisting of the succeeding developmental stages of oogonia proliferation, oocyte growth, oocyte maturation and ovulation (reviewed by Lubzens *et al.*, 2010). In fish, Fsh activity is essential for oocyte growth and storage of lipids and yolk. When ending the growth period, Lh induces oocyte maturation and ovulation after which oocytes are released to the external environment and fertilized.

In fish, Gths production and release are under the dual control of gonadotropin releasing hormone (GnRH) and dopamine (DA) (reviewed by Dufour *et al.*, 2009). Dopaminergic neurons originating from the *nucleus anteroventralis* of the preoptic area (NPOav) directly innerve the region of the pituitary gland that harbors the gonadotropic cells, the *proximal pars distalis* (Jolly *et al.*, 2016). The dopaminergic innervation in the pituitary downregulates Gths release by counteracting the stimulatory effects of GnRH in many teleosts (reviewed by Dufour *et al.*, 2009). The effect of DA is mediated by D2 dopaminergic receptors (D2r) (reviewed by Dufour *et al.*, 2009) that are expressed in the pituitary (Jolly *et al.*, 2016). In zebrafish, one of the D2r paralogue, *drd2b*, is involved in mediating the inhibitory control of gonadotropins (Fontaine *et al.*, 2013).

Dopaminergic inhibition of the reproductive axis is present in many fishes belonging to the salmonids, cyprinids, silurids, percomorphs, elopomorphs but also in one chondrostean fish (reviewed by Dufour *et al.*, 2009). The inhibitory role of DA is one of the major bottlenecks in the propagation of many fish species in aquaculture. Treatment with pituitary extracts to circumvent the pituitary, but also with GnRH in combination with D2r antagonists to counteract dopaminergic inhibition, are used to induce sexual maturation and ovulation. Long term hormonal treatment can impact egg quality and thus it is essential to gain insights about the mechanisms of dopaminergic inhibition to gain control over it. Thus, we explored the feasibility of producing *drd2b* knock-out mutants with CRISPR-Cas9 in zebrafish as model species to study the progression of sexual maturation.

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_490, © P. Jéhannet *et al.* 2022

Materials & methods

An overview on how CRISPR-Cas9 induces targeted mutations is presented in Figure 1. In this study, two different protocols, applied by Utrecht University and Wageningen University & Research respectively, were used to explore the feasibility of producing *drd2b* mutants with CRISPR-Cas9. While Utrecht University follows the protocol of Essner *et al.* (2016), Wageningen University & Research used a combination of the Shawn Burgess lab protocol (Varshney *et al.*, 2015) and Hubrecht lab protocol. CRISPR-Cas9 activities were covered by GGO permits (Wageningen University & Research: IG 19-08; Utrecht University: IG 02-115).

sgRNA. At Utrecht University, the functional sgRNA was generated by mixing crRNA that was designed on the target sequence and the generic tracrRNA that anchors to the Cas9. Two crRNAs (AS and AX) were designed on the first exon of *drd2b* by using the IDT (Integrated DNA technologies, Coralville, United States) online tool after entering the *drd2b* sequence of *Danio rerio* (NM_197936.1). At Wageningen University & Research, the sgRNA was designed as follows (Figure 2): (1) the target sequence (20 nt) of the fourth exon of *drd2b* was designed by using the online tool of UCSC Genome Browser (Santa Cruz, United States); (2) the oligo A was ordered after adding the T7 promoter sequence (17 nt) and the first 20 nt of the gRNA to the respective 5' and 3' of the target sequence; (3) the generic oligo B (80 nt) that was complementary (20 nt) to the last part of oligo A was ordered; (4) both oligos were annealed and extended to produce a 117 nt double stranded DNA fragment that was *in vitro* transcribed to generate the sgRNA (Exon4).

Microinjections. At Utrecht University, the CRISPR-Cas9 solution was made by: (1) mixing and heating at 95 °C for 5 min 1 µl crRNA, 1 µl tracrRNA and 31.3 µl buffer to create the 3 µM sgRNA solution; (2) adding 0.5 µl Alt-R *S.p.* Hifi Cas9 Nuclease V3 (IDT, Coralville, USA) diluted in 9 µl Cas9 working buffer and 0.5 µl Phenol Red to a working concentration of 0.5 µg/µl; and (3) combining 3 µl of sgRNA with 3 µl diluted Cas9

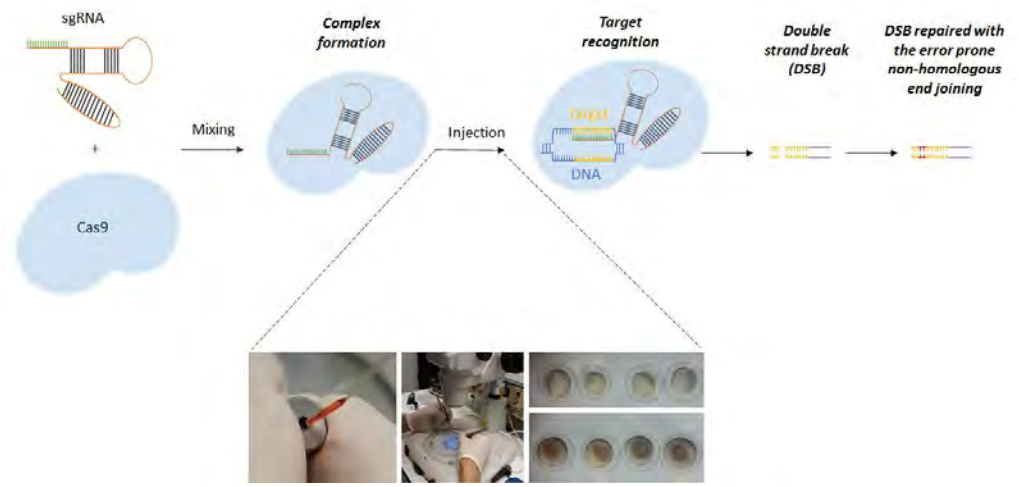


Figure 1. Overview of how the CRISPR-Cas9 system induces targeted mutations. CRISPR-Cas9 consists of two components: (1) a sgRNA that contains a guide sequence (designed by the user) which is complementary to the target sequence and (2) a scaffold sequence that is necessary for the Cas9 protein binding. Once mixed together, the sgRNA and Cas9 form a complex that binds to the target to induce a double strand break (DSB) within the DNA. The DSB will be repaired by the error prone non-homologous end joining (NHEJ) pathway that causes indels at the DSB site.

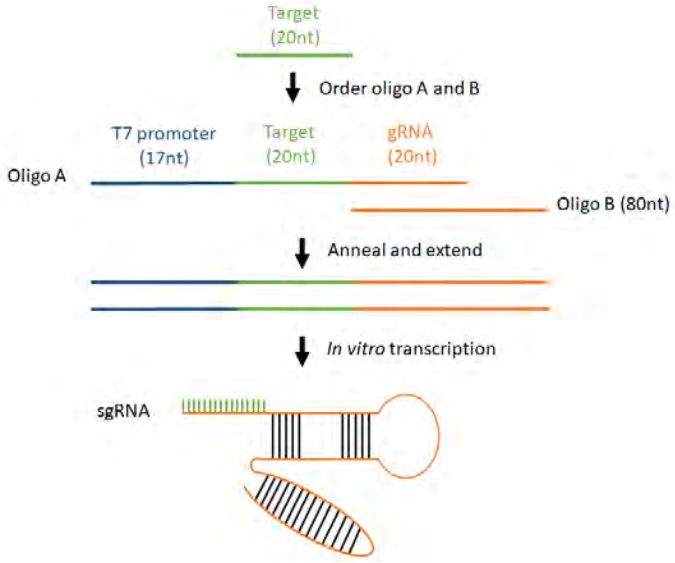


Figure 2. *In vitro* transcription of sgRNA (adapted from Varshney *et al.* 2015).

protein. At Wageningen University & Research, the CRISPR-Cas9 complex was made by mixing 600-900 ng sgRNA, 0.8 µl Alt-R *S.p.* Cas9 Nuclease V3 (IDT) at 5 ng/µl in 400 mM KCl, 0.3 µl Phenol Red and 400 mM KCl to a final volume of 6 µl. Injection mixes were loaded, calibrated under the binocular to inject 1 nl in at the 1-cell stage in zebrafish. Larvae were sampled at 2 days post hatching (dph) for DNA extraction to screen for mutation. Primers, that screened for the mutations were designed and PCR products sent for sequencing to check for mutation efficiency that was calculated as followed: (the number of zebrafish mutants confirmed by sequencing/ the total number of zebrafish sequenced) × 100%.

Results

An overview of the results is presented in Table 1. In zebrafish, ~300 eggs that were injected with CRISPR-Cas9 had a survival rate ranging from 6 to 40%. Screening showed that WT sequences had one clear peak for each nt in the *drd2b* sequence, while mutants had several peaks (two or more) for each nt around and after the protospacer adjacent motif (PAM) sequence (Figure 3); results suggesting that mutants were heterozygotes or mosaics. CRISPR-Cas9 induced mutations with an efficiency up to 73%.

Discussion

In zebrafish, *drd2b* mutants were produced but they were heterozygotes or mosaics. It is possible that homozygote mutants did not survive because knocking-out *drd2b* would be lethal. This hypothesis is

Table 1. Number of eggs injected with CRISPR-Cas9, survival rates and mutation efficiency for each target in zebrafish.

Target	Number of eggs injected	Survival rate (%)	Efficiency (%)
Exon 1, AS	300	6	33
Exon 1, AX	300	10	0
Exon 4	300	40	73

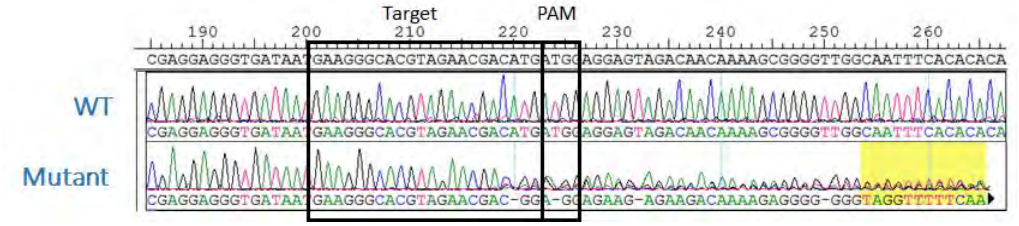


Figure 3. *Drd2b* sequences of the wild type (WT) and mutant in zebrafish *Danio rerio*.

supported by the study of Boehmler *et al.* (2004) who suggested that *drd2b* is involved in crucial processes (e.g. motor behaviour). Unfortunately, we could not confirm that knocking-out *drd2b* is lethal since the controls also died in this experiment. It is important to note that mutation in a seemingly crucial gene is not necessarily lethal (Ziv *et al.*, 2012). Future studies should investigate whether viable homozygous *drd2b* mutants can be produced. If knocking-out *drd2b* turns out to be lethal in zebrafish larvae, it might be useful to produce inducible and/or spatially restricted *drd2b* mutants to study the progression of sexual maturation. Gained knowledge could help to develop methodology to sexually mature fish with natural triggers rather than with artificial stimulators. The authors acknowledge Wageningen Livestock Research for funding this study as 'innovation challenge'.

References

Boehmler W., Obrecht-Pflumio S., Canfield V., Thisse C., Thisse B. *et al.* (2004). Dev. Dyn. 230 (3): 481-493. https://doi.org/10.1002/dvdy.20075

Dufour S., Weltzien F.A., Sébert M.E., Le Belle N., Vidal, B. *et al.* (2005) Ann. N.Y. Acad. Sci. 1040(1): 9–21.https://doi.org/10.1196/annals.1327.002

Essner J. (2016) Zebrafish embryo microinjection: Ribonucleoprotein delivery using the Alt-R CRISPR-Cas9 system. Available at: https://www.idtdna.com/pages/education/decoded/article/protocols-for-crispr-genome-editing-in-your-model-system

Fontaine R., Affaticati P., Yamamoto K., Jolly C., Bureau C. *et al.* (2013) Endocrinology 154(2); 807-818. https://doi.org/10.1210/en.2012-1759

Jolly C., Rousseau, K., Prézeau L., Vol C., Tomkiewicz J. *et al.* (2016). J. Endocrinol. 28, (9). https://doi.org/10.1111/jne.12411

Lubzens E., Young G., Bobe J., Cerda J. (2010) Gen. Comp. Endocrinol. 165(3):367-389. https://doi.org/10.1016/j.ygcen.2009.05.022

Varshney G.K, Pei W., LaFave M.C., Idol J., Xu L. *et al.* (2015) Genome Res. 25, 1030-1045. https://doi.org/10.1101/gr.186379.114

Ziv L., Muto A., Schoonheim P.J., Meijsing S.H., Strasser *et al.* (2013) Mol. Psychiatry. 18, 681-691. https://doi.org/10.1038/mp.2012.64

491. Societal aspects of animal breeding: a bibliometric analysis

M. van der Sluis*, K.H. de Greef and G. Bonekamp

Wageningen University & Research, P.O. Box 338, 6700 AH Wageningen, the Netherlands;
malou.vandersluis@wur.nl

Abstract

Societal and ethical issues addressed in animal breeding studies from 1990 onwards were examined by means of a bibliometric analysis. Using a specific set of search terms in Scopus, 673 publications of potential interest for societal aspects of animal breeding were found. Subsequent manual selection resulted in 270 relevant studies. Across the period from 1990 to the present, the number of publications addressing societal or ethical aspects of animal breeding, mainly focussing on animal welfare and behaviour or genetic modification, increased from the year 2000 and peaked around 2010. Of the 270 publications, 24% explicitly mentioned ethics, mainly in relation to genetic modification, biotechnology or animal welfare/behaviour. The minority of the publications (32%) were experimental studies. To conclude, only a minor part of publications on animal breeding explicitly address societal issues and, of these, the minority explicitly addresses ethics.

Introduction

Societal awareness of animal production has increased (Busch and Spiller, 2018), and animal breeding plays an important role in animal production. Animal breeding activities can be a cause for societal, and ethical, concern. Examples are concerns regarding genetic modification or gene editing, such as animal integrity or naturalness (e.g. Eriksson *et al.* (2018)), and regarding the impact of genetics, or the choices made by breeders, on animal welfare (e.g. Fernyhough *et al.* (2020)). To obtain a broader overview of the degree to which societal and ethical concerns in relation to animal breeding are addressed in the animal breeding literature, we here performed a bibliometric analysis in which we assessed general trends in publications on societal aspects of animal breeding, regarding the species and main societal or ethical issues under study. We specifically focussed on dairy cattle, pigs, and chickens, as these represent the major food producing sectors.

Materials & methods

Bibliometric approach. Scopus was used as the data source for this bibliometric analysis. The selection was performed in three steps (Table 1). The first step comprised the combination of a set of key words regarding: (1) animal; (2) breeding; (3) society; and (4) concerns. In the following two steps, false positive hits were excluded. To reduce subjectivity in the selection process, at least two people decided together which decision should be made in steps 2 and 3.

Data analysis. After selection of the final data set of publications, we assessed for each of the publications, based on the title and key words (and when not conclusive, based on the abstract), what topics were examined (Table 2). Moreover, each abstract was analysed for characterisation of the type of study as either 'experimental' (based on an experiment, a data set, simulation study or similar) or 'descriptive' (reviewing literature, reporting from expertise, et cetera), and to assess whether the term ethics was explicitly mentioned ('ethic', with potential prefixes or suffixes) or not.

Results

In total, 270 papers met all criteria (steps 1, 2 and 3). In total, 86 out of the 270 papers (32%) were experimental studies, with the remainder being mainly overviews, reviews, foresights and state of the

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_491, © M. van der Sluis *et al.* 2022

Table 1. Workflow and selection criteria used in the bibliometric analysis.

Steps in the workflow	
Step 1: Literature search (n=673)	Automated search in titles, abstracts and key words, in publications from 1990 onwards (search performed on the 3rd of December, 2021). Only publications with at least one of the required search terms for each of the categories below were included: - *Animal*: animal AND production; animal AND farm*; dairy AND cattle; poultry; pig* - *Breeding*: breeding AND selection; genetic AND selection; breeding AND genetic - *Society or ethics*: soci*; public; ethic*; moral* - *Concerns or discussion forms*: concern*; issue*; problem*; dilemma*; question*; demand*; worry; worries; debate; discourse; discussion
Step 2: Selection based on titles (n=400)	Manually excluding titles not meeting the criteria, mainly excluding publications on species outside the scope of this study (for example pigeons and guinea pigs, due to the term pig*)
Step 3: Selection based on animal breeding aspect (n=270)	Manually excluding publications that did not have animal breeding or genetic engineering in the species of interest as a topic of focus, based on the title, key words and abstract

Table 2. Structuring of topics examined in the publications.

Species under study	**What societal or ethical issue is explicitly mentioned?**	
	Animal (trait) related	**Methodology related**
Chickens	Animal health	Animal genetic resource management
Pigs	Animal welfare/behaviour	Biotechnology
Dairy cattle	Environment	Breeding goal definition
Across species	Food safety/security	Genetic modification
	Heat stress	General methods
	Meat quality	
	Multiple	
	Resilience/robustness	

art reports. The topics that were studied differed per species (Table 3). For chickens, the vast majority of publications focused on animal welfare and behaviour, while for pigs many studies focused on meat quality, besides animal welfare and behaviour. For cattle, aside from animal welfare and behaviour, many studies focussed on animal health. For the across-species papers, the majority focused on genetic modification. Out of the 270 publications, 65 (24%) mentioned ethics explicitly, resulting in 90 species-topic combinations that included ethics. Most ethics-mentioning papers were related to biotechnology, genetic modification or animal welfare and behaviour (Table 4).

The number of publications on societal or ethical aspects of animal breeding appears to start increasing around 2000, peaking in 2010 (Figure 1). Although this increase was visible for all three species, the topics that were discussed were different for the different species, with mostly trait-related publications for chickens and pigs, and more methodology-related papers for across species studies (Figure 1).

Discussion

In this study, we aimed to provide an overview of societal and ethical concerns that are raised in relation to animal breeding, using a bibliometric approach. Methodology-wise, it became clear that the specificity of the search string was low. After applying the manual exclusion steps in our initial search results, only 40% of the publications remained, highlighting the importance of manual selection steps in bibliometric analyses. The analysis revealed that a small number of topics received substantially more attention than others.

Table 3. Number of publications addressing specific topics by species. Publications addressing multiple topics or multiple of our key species (chickens, pigs and/or dairy cattle) are included multiple times, resulting in n=335. AGRM = animal genetic resource management.

Topic under study		Chickens	Pigs	Dairy cattle	Across species
Animal (trait) related	Animal health	3	5	12	9
	Animal welfare/behaviour	26	15	13	20
	Environment	0	1	4	5
	Food safety/security	2	1	6	3
	Heat stress	1	1	1	0
	Meat quality	1	11	0	1
	Multiple	4	2	2	4
	Resilience/robustness	2	1	1	2
Methodology related	AGRM	2	3	5	14
	Biotechnology	1	0	1	14
	Breeding goal definition	1	0	5	4
	Genetic modification	4	1	7	27
	General methods	1	1	5	7
Other	Other	6	16	26	25

Table 4. Publications explicitly mentioning ethics per species and topic, in percentages of the total number of publications mentioning ethics. Publications addressing multiple topics or multiple of our key species (chickens, pigs and/or dairy cattle) are included multiple times, resulting in n=90. AGRM = animal genetic resource management.

Topic under study		Chickens	Pigs	Dairy cattle	Across species
Animal (trait) related	Animal health	0%	1%	2%	2%
	Animal welfare/behaviour	7%	7%	6%	16%
	Environment	0%	1%	2%	2%
	Food safety/security	0%	0%	1%	0%
	Heat stress	0%	0%	0%	0%
	Meat quality	0%	3%	0%	0%
	Multiple	1%	0%	0%	0%
	Resilience/robustness	2%	0%	0%	1%
Methodology related	AGRM	0%	0%	0%	0%
	Biotechnology	1%	0%	1%	8%
	Breeding goal definition	0%	0%	0%	2%
	Genetic modification	2%	1%	3%	17%
	General methods	0%	0%	0%	1%
Other	Other	0%	3%	1%	4%

Unsurprisingly, animal welfare and behaviour, and genetic modification, were commonly addressed, also within the subset of studies that explicitly mentioned ethics. However, there were differences between species in what topics were mainly studied. This raises the question why, besides differences in the physiology or use of the animals, some topics are potentially considered of greater societal relevance in one species than in another. Future work examining the publications in detail, that is, (a selection of) the publications will

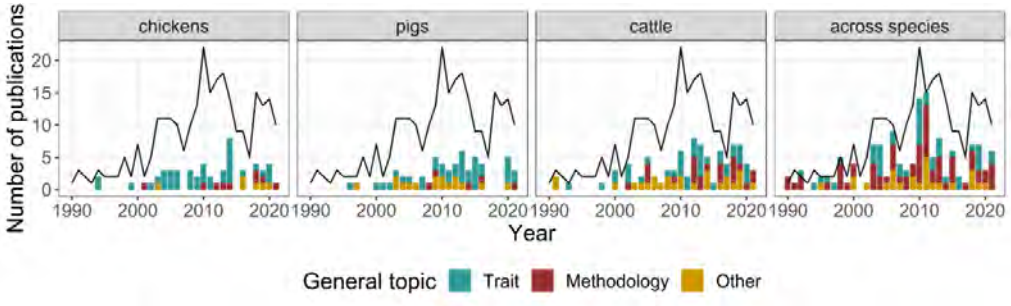


Figure 1. Number of publications on the different general topics and species. The black line represents the total number of publications by year (n=270), whereas the bars represent, per species, the number of studies focussing on certain general topics. Publications addressing multiple topics or multiple species categories are included multiple times, resulting in n=335 for the bar graphs.

be fully read, can possibly provide insight into the context in which the societal issues are raised and the reasoning behind the choice of topic. Even though the number of studies that address societal aspects of animal breeding increased after the year 2000, it appears that only a small proportion of the publications on animal breeding do so. A search for animal breeding studies on our species of interest from 1990 onwards, without the societal aspects constraint, resulted in approximately 24,500 publications (results not shown), of which the selection in this study forms only about 1%. Thus, only few publications on animal breeding address societal issues explicitly and, of these, the minority explicitly addresses ethics, mainly focussing on only a small subset of topics.

Funding

This study is part of the Breed4Food project Ethics and Society, financially supported by the Dutch Ministry of Economic Affairs (TKI Agri and Food project 16022) and the Breed4Food partners Cobb Europe, CRV, Hendrix Genetics and Topigs Norsvin.

References

Busch G., and Spiller A. (2018) Anim. Front. 8:1-3. https://doi.org/10.1093/af/vfx005

Eriksson S., Jonas E., Rydhmer L., and Röcklinsberg H. (2018) J. Dairy Sci. 101:1-17. https://doi.org/10.3168/jds.2017-12962

Fernyhough M., Nicol C.J., van de Braak T., Toscano M.J., and Tønnessen M. (2020) J. Agric. Environ. Ethics 33:15-36. https://doi.org/10.1007/s10806-019-09810-2

492. The need for harmonized, risk-based regulation of gene editing in livestock

A.L. Van Eenennaam* and M.L. Mueller

University of California, One Shields Avenue, Davis, CA 95616, USA; alvaneenennaam@ucdavis.edu

Abstract

One challenge for the incorporation of gene-editing into livestock breeding programs is the lack of international regulatory harmony. At the current time several jurisdictions are treating the risks associated with knockouts and intraspecies allele substitutions that could have been achieved using conventional breeding, in the same way as those associated with the random insertion of exogenous DNA achieved using traditional genetic engineering methods. Many of the potential hazards, such as novel allergens or toxins associated with a transgenic expression product, are not relevant to gene knockouts and cisgenic allele replacements. As a result, several countries are treating such products as analogous to those produced though conventional breeding. Subjecting gene-edited livestock to a unique regulatory burden, irrespective of the nature of the edit, is difficult to justify from a risk perspective. The lack of regulatory harmony with regard to gene-edited animals jeopardizes the global adoption of this technology in livestock breeding programs.

Introduction

In January 2017, the United States Food and Drug Administration (FDA) released a draft guidance signalling its intention to regulate intentional genomic alterations (IGA) in gene-edited animals as new animal drugs (FDA, 2017). The guidance reads, that 'a specific DNA alteration is an article that meets the definition of a new animal drug at each site in the genome where the alteration (insertion, substitution or deletion) occurs', and that 'in general, each specific genomic alteration is considered to be a separate new animal drug subject to new animal drug approval requirements.' By this definition, all gene-edited livestock are unapproved animal drugs and therefore are in general 'deemed unsafe'. As a result, edible products from gene-edited animals are not allowed enter the food or feed supply without prior FDA authorization. Rather, the FDA recommends that investigational animals and their biological products be disposed of by incineration, burial, or composting. To obtain approval for a new animal drug first means applying for an Investigational New Animal Drug (INAD). In this paper I will outline my experience as a university researcher working with gene-edited livestock, and the information that has been requested to comply with the regulatory requirements of the FDA, and contrast this to the regulatory approaches of various countries.

Materials & methods

In February 2021, an application for an INAD for gene-edited *NANOS3*-/- knock-out cattle was submitted to the FDA. These cattle were produced using Cas9 and CRISPR guide RNAs (gRNAs) targeting exon 1 of the bovine NANOS3 gene in bovine zygotes. The cattle were healthy and unremarkable other than the fact they lacked germ cells and were therefore infertile, as hypothesized. The gRNAs used for this experiment were designed and screened for potential off-target sites based on the bovine reference genome using bioinformatic tools, sgRNA Scorer 2.0 and Cas-OFFinder, respectively. Based on a systematic analysis of CRISPR-Cas9 mismatch tolerance only gRNAs that met specific mismatch parameters were selected. A mismatch is defined as a discrepancy between a base of the gRNA and the predicted off-target site. The gRNA selection criterion was: (1) at least 3 total mismatches; and (2) at least 1 of the mismatched bases was located in the seed region (8-10 bp proximal to the PAM site) of the gRNA. Our previous work in bovine zygotes using these design parameters for testing multiple gRNAs and targeting several different loci showed no off-target effects (Hennig *et al.* 2020).

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_492, © A.L. Van Eenennaam and M.L. Mueller 2022

The information required to support this INAD request was an overview of the project objectives, a description of the constructs or genomic alterations in the animal, a description of the method(s) or technologies used to produce and deliver/introduce the genomic alteration, and a description any completed studies including any relevant information from other animals in which the same genes have been knocked out (e.g. references to published articles characterising knockout mice). Additionally, a request for 'a categorical exclusion from the requirements to prepare an environmental assessment' was submitted to accompany the INAD file. When FDA reviews and approves new animal drugs, it has to comply with the requirements of the National Environmental Policy Act, which includes a review of environmental risks, if any, where required. Requesting this exclusion entailed certifying that no extraordinary circumstances existed whereby the gene-edited cattle could significantly affect the quality of the human environment. This required documenting the location of facilities where the animals were housed, containment measures, management practices and conditions for all facilities, animal waste and animal disposal, and a description of standard operating procedures, and submitting this information to allow for a determination to be made.

Results

Upon obtaining this INAD, an investigational food use authorization (FUA) was requested to introduce the knock-out cattle into the food supply at the completion of the experiment, rather than incinerating them. This required providing data outlined in Table 1. The off-target in silico analysis included 10 sequences with three mismatches to the guide sequence in the reference cattle genome. There were no sequences with less than three mismatches other than our target, as per our guide design criteria. Given the expense involved

Table 1. Information requested by the US FDA to consider authorizing food use of experimental gene-edited knockout cattle with Intentional Genomic Alterations (IGA).

A complete description of animals proposed to enter the human food supply:
Details regarding the species, class, and number of requested investigational animals, including the strategy used to produce the offspring.
Characterisation of the intentional genomic alteration (IGA) and any intended effects:
A description of the purpose of the IGA and the intended function (including location in the genome, impact on protein expression, if any, intended sequence, etc.) and details on how the alteration was achieved.
PCR genotyping data and/or sequencing data to demonstrate that the IGA is correctly targeted, as expected.
If available for comparison, you may also provide PCR genotyping data and/or sequencing data from non-edited comparators.
Justification that the IGA is expected to result only in the intended effect
i.e. knockout should not lead to any other changes in animals, which may impact the edible tissues. The justification may rely on available literature, phenotypic data on knockout animal models collected in lab (e.g. biomarker data), and other resources that report on germline abnormalities in knockout animals (e.g. cattle and mice).
Animal health data/information:
Any data/information collected on the health status of the animals that will enter the food supply. This includes visual observations, bloodwork including any hormone measurements, etc.
Compositional analysis:
Justification/explanation that no substantial differences in nutritional or composition content of edible tissues between edited and non-edited animals are expected as a result of the introduction of the IGA, including any unintended alterations. Data from other animals in the lineage or with a similar alteration that will aid in the evaluation can be submitted as justification.
Predicted off-targets:
Option 1: For each edited animal, PCR and/or sequence-verification of the absence of off-target effects in the 10 predicted regions identified via in silico analysis
Option 2: Justification for each of the 10 regions as to why there is no need to provide additional data that verifies the lack of an off-target. For example, justification could include information pertaining to the impact on protein expression (or lack thereof) or known impacts of the gene on human food/meat quality.

with performing this off-target sequence verification for 10 loci for each animal, and the fact that this analysis was unrelated to our scientific interest in these animals (meaning we would be doing this analysis solely for the FUA), the FUA was abandoned in favour of requesting a rendering use authorization. This meant that the animals would be allowed to enter the animal feed chain, rather than the human food chain at the completion of the experiment. To obtain rendering authorization required a separate application to the FDA Division of Animal Feeds. Similar information to that outlined for the FUA was requested but this request did not require off-target sequence verification. In December 2021 we obtained a feed use authorization for these animals to be rendered at the completion of the experiment.

Discussion

These regulatory exchanges involved a substantial investment of time, both for the regulators and investigators. Such regulatory interactions are not a routine part of basic academic research in animal breeding and genetics. Neither is disposing of all animals by incineration, burial, or composting at the completion of a research project. This is particularly expensive when working with large livestock species, and is a unique expense associated with the use of gene-editing in food animals not incurred when researching conventional breeding methods. Further, these regulatory data are typically being requested prior to the conclusion of the experimental work, rather than at the completion of the research when manuscripts describing and detailing the methods and results are typically being compiled and written. As such, preparing these documents required a dedicated time commitment solely for regulatory compliance, and pursuing a FUA would have required further additional experimental work.

Internationally harmonized guidelines for assessing the safety of foods derived from transgenic animals have been developed by Codex Alimentarius, but these do not consider aspects of animal welfare. Other

Table 2. Agricultural gene editing regulation index (Genetic Literacy Project, 2021).

Country/Region	Food/ Crops	Animals	Average ag rating
Argentina	10	10	10
Australia	8	8	8
Brazil	10	10	10
Canada	8	8	8
Central America	6	6	6
Chile	10	1	5.5
China	5	5	5
Colombia	10	1	5.5
European Union	2	2	2
India	6	6	6
Israel	10	5	7.5
Japan	8	8	8
Mexico	1	1	1
New Zealand	4	4	4
Paraguay	10	10	10
Russia	5	5	5
United Kingdom	2	2	2
Ukraine	1	1	1
Uruguay	6	6	6
United States	10	4	7

Regulation Status	Rating
Determined[1]: No Unique Regulations	10
Lightly Regulated[2]	8
Proposed[3]: No Unique Regulations	6
Regulations in Development	5
Highly Regulated	4
Mostly Prohibited	2
No Clear Regulations	1
Prohibited	0

[1] Determined, [3] Proposed: No Unique Regulations: Gene-edited products that do not incorporate DNA from another species will be regulated as conventional products with no additional restrictions.

[2] Lightly Regulated: Some or all types of gene editing are regulated more strictly than conventional agriculture, but not as strictly as transgenic GMOs.

jurisdictions, notably in Europe, explicitly include aspects of animal health and welfare in the premarket assessment of genetically modified animals for food and feed purposes (van der Berg *et al.* 2020). In 2015, Argentina issued a regulation clarifying the regulatory status of gene-edited products (Lema 2019). It keeping with the definition in the Cartagena Protocol, it outlined that for a gene-edited organism to be a genetically modified organism (GMO), it had to 'possess a novel combination of genetic material'. The Argentine regulation calls for gene-edited plants and animals to be presented to the biosafety commission in order to establish, on a case by case basis, whether it is a GMO. An interesting aspect of this regulation is that there is an opportunity to present projects at the 'design stage', whereby a preliminary opinion based on the expected outcome will be issued by the commission.

Later, when the plants or animals have been obtained and fully characterised, applicants must present a follow up report that will be used to establish a final decision. That determination is mostly based on any changes present in the genome of the product intended to be sold commercially. Table 2 shows the evolving regulatory status of gene-edited crops and animals in different countries and regions throughout the world. It highlights the discrepancies between regulations for gene-edited food applications from the plant and animal kingdoms, and the current lack of global harmonization. Additionally, many regions are still developing their regulations. For example, no African nation has passed regulations for gene-edited animals. In practice, gene-edited animals will pose challenges in relation to aspects of traceability in the food chain. International harmonization of risk-based regulatory criteria will be essential to avoid trade obstacles, and in its absence gene-editing will likely play a limited role in livestock genetic improvement programs.

References

Food and Drug Administration (FDA), Regulation of Intentionally Altered Genomic DNA in Animals. 2017. Available at: https://www.fda.gov/media/74614/download.

Genetic Literacy Project (GLP), Global Gene Editing Regulation Tracker 2021. Available at https://crispr-gene-editing-regs-tracker.geneticliteracyproject.org/.

Hennig S.L., Owen J.R., Lin J.C., Young A.E., Ross P.J. *et al.* (2020) Sci. Rep. 10(1): 22309. https://doi.org/10.1038/s41598-020-78264-8

Lema, M. A. (2019) Transgenic Res. 28(2) 147-150. https://doi.org/10.1007/s11248-019-00145-2.

Van Der Berg J.P., Kleter G.A., Battaglia E., Groenen M.A.M. and Kok E.J. (2020) Front. Agr. Sci. Eng. 7(2): 136-147. https://doi.org/10.15302/j-fase-2019306

493. Broadening the ethical debate on breeding innovations, public engagement and the role of the Democs game

F.L.B. Meijboom[1*], D.M. Bruce[2], K. Kramer[1] and A. Bruce[3]

[1]Utrecht University, Faculty of Veterinary Medicine, Sustainable Animal Stewardship, Yalelaan 2, 3584 CM Utrecht, the Netherlands; [2]Edinethics Ltd, View Hill, 1 Peinsoraig, Uig, Isle of Skye IV51 9YD, United Kingdom; [3]University of Edinburgh, School of Social and Political Science, Old Surgeons' Hall, Edinburgh EH1 1LZ, United Kingdom; f.l.b.meijboom@uu.nl

Abstract

As genomic selection becomes increasingly important in livestock breeding, it also entails ethical and societal issues, and we are examining these as part of the BovReg project, funded by the Horizon 2020 programme of the European Union, on advanced cattle genomics. We argue that discussion of these issues must have broad scope, in at least three respects. Firstly, it should go beyond risk and animal welfare. Secondly, it should consider not only what is novel but assess also existing ethical issues in the light of genomic innovations. Lastly, because issues of animal production are of interest to European society, public engagement is an essential element for responsible innovation. To stimulate debate on genomic selection, we have created a Democs card game to engage with general publics and stakeholders. The game explains the context of current production and selection in cattle, examines ethical and social issues raised by breeding techniques, and invites players to form their own views.

The need for ethical reflection

Genomic selection has become an important tool in livestock breeding, with further potential to widen both the traits and the breeds to which it can be applied. It also entails ethical and societal choices, for example in balancing such factors as animal health and welfare, fertility and climate impact alongside improving yields. As part of the BovReg project (www.bovreg.eu) on advanced cattle genomics within the FAANG network, we are examining these ethical questions. In contrast to some other reproductive technologies in animal breeding, such as cloning or genome editing, genomic selection has not raised much ethical attention. This can be explained partly by the characteristics of the technology itself, which is about the use of genome-wide genetic markers to predict the breeding value of selection candidates. Genomic selection therefore does not cross species boundaries, or require applying new invasive techniques on animals. However, that genomic selection is in certain respects continuous with earlier breeding techniques does not imply that it raises no ethical issues. It is also the result of a rather limited view on what counts as relevant ethical considerations. However, genomic selection has not so far resulted in extensive public debate.

An analysis of the academic literature found only few papers pay attention to ethical aspects. One of them is by Mark and Sandøe (2010). They highlighted issues such as unexpected effects on animal welfare due to correlations with non-measured traits, a higher risk of spreading deleterious mutations, the centralisation of capabilities and increasing dominance of specific breeds. They also discuss concerns over increased monopoly within dairy cattle breeding. The paper stresses the importance that stakeholders monitor the effects of genomic selection and that breeding companies act responsibly. This is in line with the development and regular updates of Code-EFABAR which since 2006 has aimed to define and maintain good practices for farm animal breeding (Code-EFABAR 2020).

In this paper, we first argue that the discussion on genomic selection needs to reflect a wider range of ethical issues than is commonly the case. Secondly, we focus on extending discussion beyond the novel concerns, and lastly we argue that the current comparative lack of public debate provides an opportunity to work on

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_493, © F.L.B. Meijboom *et al.* 2022

public engagement. For this purpose, we have created a Democs card game to engage with general publics and stakeholders, on which we will elaborate in the second part of the paper.

Three factors to broaden the debate

A broader ethical scope: beyond risk and animal welfare. It is important to move away from confining ethical discussion to risk and safety. In response to legal frameworks and a no-harm principle, there is a tendency for ethical evaluation to focus primarily on the question to what extent a technology can harm others and what risks are involved (Fleming *et al.* 2018). Notwithstanding the importance of minimizing risks to food safety, genetic diversity and economic risks, an ethical evaluation of breeding innovations should encompass more than this dimension. In addition, given the duties we have towards animals and the environment, the ethical considerations should not be restricted to human-related issues.

There has been attention to the positive and negative impacts the technology can have for animals and their welfare (Mark and Sandøe, 2010, Windig 2012). This is a crucial dimension of the ethical assessment, as animal welfare is a core concept in animal ethics. Focussing ethical and public debates on animal welfare may overload the concept, however, because this invites people to 'translate' their wider concerns into issues of welfare. Therefore, we argue for a broader scope in the discussion on the ethics of genomic selection that includes risk and animal welfare related arguments, but also goes beyond them. Our mapping of the potential ethical issues showed views on genomic selection that cannot be reduced to animal welfare concerns. These included questions raised by more critical voices, of justice, autonomy (Coles *et al.*, 2015) and 'biopower' i.e. changing animal bodies and populations in a direction dictated by particular human interests (e.g. Twine, 2010). There needs to be room to discuss such ethical issues as the instrumentalization of animals, changing power relations among stakeholders in animal breeding, or human-oriented ideas about perfecting animals. Most of the latter concerns are not specific to genomic selection and are relevant in conventional breeding programs or even livestock farming in general. But we argue that this is a reason to take into account a second factor in broadening the ethical discussion.

Beyond what is novel. Especially when changes are mediated by technology the ethical analysis tend to focus what is, ethically speaking, novel about those technologies.' This starts from the assumption that the situation before the innovation was morally justified or acceptable. On that basis, genomic selection should be assessed only by considering whether it introduces novel ethical issues, compared with existing breeding practices. This view is problematic for two reasons, however.

Firstly, adding one element to an existing practice can have far-reaching consequences that expand dimensions that were already present in the existing situation. Not without reason, genomic selection is portrayed as a next step in the development of existing breeding programs and as a paradigm shift at the same time (Meuwissen *et al.* 2016). Secondly, it fails to reflect back on what may be ethically problematic in existing practices. The assumption that livestock breeding is an uncontroversial practice cannot be taken for granted, as the ethics of breeding animals has been widely discussed in various contexts (Olsson *et al.* 2006; Leenstra *et al.*, 2012; Farstad 2018). Therefore, ethical reflection on genomic selection should not only address novel issues of the technique on itself, but should consider it as part of the existing practices of breeding and livestock farming. This does not make the discussion easier, but can do justice to the complexity at stake. It is important that the discussion on the societal and ethical dimensions is not a task for only ethicists or social scientists. It needs the inclusion of stakeholders and the general public. This leads to our third factor.

Public engagement. Genomic livestock breeding is a complex technical field that requires multiple forms of expertise of highly skilled persons and specialised organisations, which tend also to be international.

The infrastructure needed to guarantee high quality in research, product development and implementation also leads to a gap between the experts in genomic selection, the users (farmers) and the general public. The economic relation between breeding companies and their clients, means that farmers may participate in discussions on breeding innovations, or have their views studied (e.g. Lund *et al.* 2021), but the general public is seldom consulted. This is problematic because public engagement is recognised as an essential element in responsible innovation (Owen *et al.* 2021). Only engaging with technical experts risks limiting the debate to dominant perspectives (Kayumova *et al.*, 2019). Two-way public engagement is crucial to invite lay persons to form their own views and enable them to participate in the debate, and to give the general public a voice in making breeding organizations, companies and governance institutions more responsive and accountable. In line with this step, it is important to engage with wider publics, who are showing an interest in animal production but are rarely consulted about the priorities and direction of livestock breeding. By the same token, general public comments need to be informed to have credibility, and communicating a complex subject such as genomic selection is challenging.

The Democs game

An integral part of the European BovReg project is to conduct public engagement as part of our aim for making responsible innovation. For this purpose, we have created a Democs card game, the purpose of which is to stimulate wider debate among stakeholders, and to give small groups of lay people in different European countries the opportunity to discuss and express their views about genomic selection in cattle, including ethical questions about breeding priorities and practices more generally. Democs has been a proven tool for 20 years to enable grassroots engagement on many technological issues in small groups, assuming no prior knowledge. The cards are the 'expert' and provide the basis for group learning and discussion. These cards have been written in a detailed iterative process drawing upon the technical, ethical and Democs game expertise and insights from BovReg partners.

Story cards introduce different aspects of cattle production, genomics and breeding through case studies of imaginary people involved the field or affected by it. These include a breeding company CEO, a genomics researcher, a veterinarian, an Alpine dairy farmer of specialty cheeses from local breeds, a government climate scientist, an environmentalist and an critical consumer. Each one explains their involvement and an ethical dilemma which it presents. Information Cards provide the context of different approaches of cattle production, with which publics may not be familiar. They explain genomic selection and its potential, and give information on animal health and welfare, land use, climate and environmental impacts. Issue Cards then open up relevant ethical and social issues arising from advanced breeding and production, expressing differing standpoints to reflect the range of opinions.

Players are invited to discuss and to form their own views built up from the suite of cards. There are two outputs. One is group opinion statements, written on Cluster Cards, based on the cards the group members have chosen to discuss. The second output is individual votes by each player on priorities and questions in cattle breeding, with their reasons in their own words. These outputs will be analysed, and are expected to provide valuable qualitative information on public views on genomic selection in cattle and related issues.

The game was first produced in a pilot version for beta testing for accuracy, balance and being understandable by lay people, and is now being finalised for distribution, with the expectation of being translated into various languages, so that people can take part in their 'mother tongue'. Covid restrictions on small face-to-face group meetings have delayed playing somewhat. We are investigating the feasibility to adapt the game to be played on-line, which is not a simple task for a tool that is designed to feature interactive discussion.

Conclusions

Breeding innovations raise ethical issues that are not limited to risk and animal welfare, and that go beyond what is novel about such innovations. Responsible innovation requires discussing such ethical issues with a wide range of stakeholders, and we invite engagement with those involved in livestock breeding and innovation. The Democs game that we have created facilitates the engagement of stakeholders, but especially of lay publics in ethical discussions on genomic selection and cattle breeding.

Acknowledgements

This project has received funding from the European Union's Horizon 2020 Research and Innovation programme under grant agreement No 815668.

References

Code-EFABAR (2020). Code of Good Practice for Farm Animal Breeding Organisations. https://www.responsiblebreeding.eu/uploads/2/3/1/3/23133976/01_general_document_2020_final-code_efabar.pdf. (accessed 30-12-2021)

Coles, D., Frewer L.J., and Goddard, E. (2015). Ethical Issues and Potential Stakeholder Priorities Associated with the Application of Genomic Technologies Applied to Animal Production Systems. J Agric Environ Ethics 28:231–253

Farstad, W. (2018). Ethics in animal breeding. Reproduction in domestic animals, 53, 4-13.

Fleming, A., Abdalla, E. A., Maltecca, C., & Baes, C. F. (2018). Invited review: Reproductive and genomic technologies to optimize breeding strategies for genetic progress in dairy cattle. Archives Animal Breeding, 61(1), 43-57.

Kayumova, S., McGuire, C. J., & Cardello, S. (2019). From empowerment to response-ability: Rethinking socio-spatial, environmental justice, and nature-culture binaries in the context of STEM education. Cultural Studies of Science Education, 14(1), 205-229.

Lund, T. B., Gamborg, C., Secher, J., & Sandøe, P. (2021). Danish dairy farmers' acceptance of and willingness to use semen from bulls produced by means of *in vitro* embryo production and genomic selection. Journal of Dairy Science, 104/7; 8023-8038, https://doi.org/10.3168/jds.2020-19210.

Mark, T., Sandøe, P. (2010). Genomic dairy cattle breeding: risks and opportunities for cow welfare. Animal Welfare, 19(S), 113-121.

Meuwissen, T. Hayes, B., Goddard, M. (2016). Genomic selection: A paradigm shift in animal breeding, *Animal Frontiers*, 6/1: 6–14, https://doi.org/10.2527/af.2016-0002.

Olsson, I. A. S., Gamborg, C., & Sandøe, P. (2006). Taking ethics into account in farm animal breeding: what can the breeding companies achieve? Journal of Agricultural and Environmental Ethics, 19(1), 37-46.

Owen, R., von Schomberg, R., & Macnaghten, P. (2021). An unfinished journey? Reflections on a decade of responsible research and innovation. Journal of Responsible Innovation, 8(2), 217-233.

Twine, R. (2010). Animals as biotechnology: ethics, sustainability and critical animal studies. London: Earthscan.

Windig, J.J. (2012). Genomic Selection: Alternatives, Risks and Benefits. Ethical Consideration in Livestock Breeding. in: Leenstra, F. *et al.*,: Ethical Consideration in Livestock Breeding. Proceedings of the first LowInputBreeds Symposium. Research Institute of Organic Agriculture (FiBL), Frick, Switzerland.

494. Perceptions of genome editing in farm animals by livestock stakeholders

R. Duclos[1,4,5], E. Delanoue[2], L. Journaux[3], D. Guéméné[1], M. Sourdioux[1] and J.P. Bidanel[4,5*]

[1]SYSAAF, UMR BOA, INRAE, 37380 Nouzilly, France; [2]Institut de l'Elevage, Ifip-institut du porc, Itavi, 149 rue de Bercy, 75012 Paris, France; [3]France Génétique Élevage, 149 rue de Bercy, 75012 Paris, France; [4]GABI, INRAE, AgroParisTech, Université Paris-Saclay, 78350 Jouy-en-Josas, France; [5]GIS Avenir Elevage, France; jean-pierre.bidanel@inrae.fr

Abstract

The objective of this study was to better understand the positions of the stakeholders of the livestock sector with respect to the use of genome editing techniques. A qualitative survey by semi-directive interviews was conducted with different actors of the livestock sector. It allowed to establish a typology of stakeholders in five categories (opponent, skeptical, cautious, enthusiastic, convinced) and to characterize the main arguments associated with the position of the different stakeholders.

Introduction

Since the development of the Crispr-Cas9 system (Jinek *et al.*, 2012), genome editing (GnEd) staktools have gained a major place in biological research and the prospects for applications are numerous, in medicine for the treatment of many diseases, medical diagnosis or eradication of disease vectors (Ashmore-Harris and Fruhwirth, 2020), in bioindustry (biofuels; Lu *et al.*, 2022) or in agriculture. The application of GnEd in agriculture is nevertheless controversial. For some, its use is essential to be able to feed 9 billion people in 2050 and to succeed in the agroecological transition (Fan *et al.*, 2021). For others, there is a risk that genetic engineering is used to perpetuate unsustainable agricultural systems (Jacobsen *et al.*, 2013) and its use involves risks that are not yet fully controlled (Norris *et al.*, 2020). However, the controversy is currently limited to the scientific and professional communities directly concerned.

On the other hand, the status of genome-edited organisms varies around the world. In Europe, they fall within the scope of the regulation on Genetically Modified Organisms (Directive 2001/18/EC) since the July 25, 2018 ruling of the Court of Justice of the European Union. In the USA, the regulatory approach in animals is also very strict, based on the food, drug and cosmetic act. In other American countries (Argentina, Brazil, etc.), GnEd organisms produced by cisgenesis or intragenesis are not considered GMOs. In some European countries (United Kingdom, Norway), which have regulations similar to those of the EU, changes in regulations are being considered. In the EU, the question of regulatory changes is also being raised, at least for plants, following a recent report by the European Commission (Commission, 2021).

In this context (controversy, regulatory uncertainty), it seemed relevant to us to better identify the positions of the various actors in the field of livestock selection with regard to the potential use of genome editing techniques in livestock breeding. To this end, a qualitative survey was carried out by means of semi-directive interviews with people involved in the breeding sector. The objective of this article is to present the main results of this survey.

Materials & methods

The semi-structured interview is an approach widely used in sociology. It consists of a series of open-ended questions, prepared in advance by the interviewer, which allows the respondents to organize their answers as they wish. It aims to be representative of the diversity of opinions, without any quantitative objective (Delanoue *et al.*, 2018). In the present case, the question grid was completed by cards presenting typical

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_494, © R. Duclos *et al.* 2022

cases of genome editing use (hornless cattle, resistance to PRRS, hypoallergenic milk, myostatin) presented during the interview.

The interviews took place from October 2020 to February 2021. Most of the interviews were conducted by video conference, due to the Covid-19 pandemic. Respondents were mostly alone, but some interviews were conducted with 2 or even 3 respondents. They were recorded once the respondents had given their consent. The average duration of the interviews was 1 hour and 25 minutes. A total of 48 people were interviewed. Their distribution by activity sector is given in Table 1.

The respondents belonged to four sectors (ruminants, pigs, poultry and aquaculture) and, for some, to the plant sector. Two interviews were withdrawn from the survey because answers were vague an unclear. The final corpus studied was 46 respondents in 36 interviews. The interviews were analyzed by viewing the videos using a standardized analysis grid.

Results & discussion

The analysis of the surveys quickly revealed that knowledge of the subject was still limited to a very narrow sphere (mainly research and breeding sectors), who were familiar with the subject (specialists). The other respondents had varying degrees of knowledge of the subject and were grouped into 3 categories, layman (did not know), novice (superficial knowledge) and expert (followed the subject).

The controversy is built around uncertainties. We have grouped together in the diagram below (Figure 1) the main uncertainties expressed by the respondents. They were linked to the technology, its impact on the environment and/or biodiversity, on the animal and its socio-political impact. An uncertainty related to human health impact was also mentioned by a small number of respondents and was therefore classified as secondary. The uncertainties were linked to the major cross-cutting issues raised almost systematically by the respondents, namely: (1) the type of agricultural model in which GnEd tools is used; (2) the impacts of their use, particularly their irreversible nature; (3) ethical issues (link to living organisms, man's right to modify nature) and ownership of living organisms. The technological uncertainties concerned the mastery of the tool and the maturity of the related scientific knowledge, with references to older techniques and the 'Recombinetics' affair (Norris *et al.*, 2020), and fears about collateral effects on animal welfare or the environment.

Once the uncertainties had been identified, we looked to see if a typology of actors could be identified within our sample. As already indicated, one of the first differentiating factors was the level of knowledge of the subject. We also tried to get the respondents to express their position or that of their organization with regard to the use of GnEd (are you in favor or against the use of GnEd in livestock farming?). The individual

Table 1. Distribution by activity sector of the interviews.

Activity sector	Number of interviews
Research (livestock and crops)	12
Public authorities	4
Breeding	8
Livestock industry (outside breeding)	7
Retailing	3
Non-governmental organisations (NGO)	14
Total	48

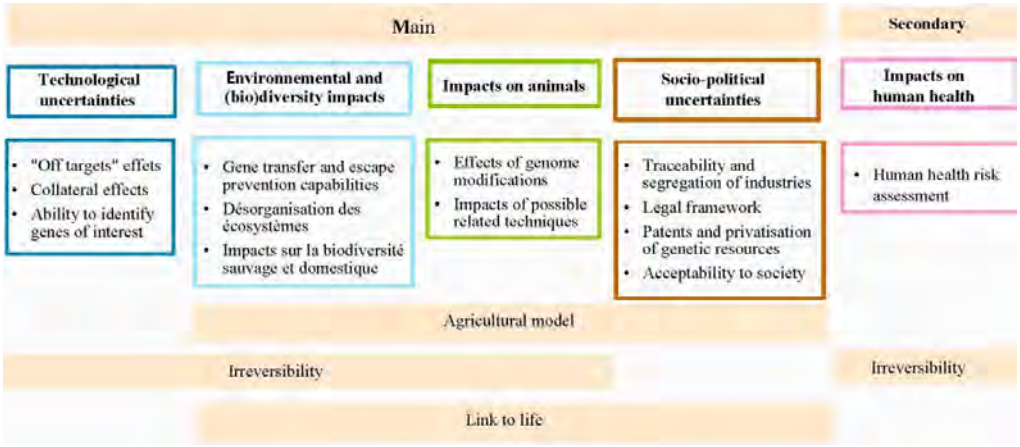


Figure 1. Uncertainties at the origin of the controversy on genome editing in livestock.

positions were analyzed, as the structures very rarely had any official or unofficial position on the subject. This analysis made it possible to construct a typology in five categories (Figure 2).

Two categories, located at the ends of the graph, represented the committed actors. They had a good or even very good knowledge of the subject and had a definite position on the use of GnEd. The convinced, who were closely linked to the R&D sector, were in favor of one or more types of use and considered that GnEd was compatible with all farming systems. Opponents often had a long-standing commitment against the use of genetic engineering in agriculture.

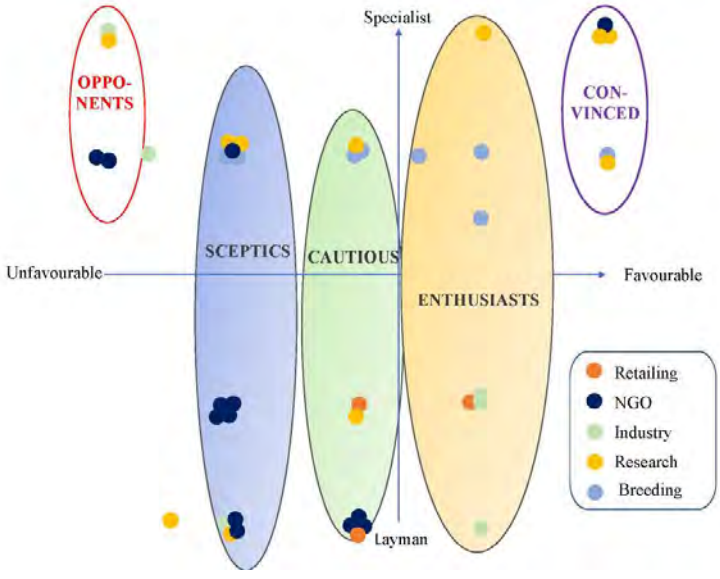


Figure 2. Typology of actors.

The other categories represented the bulk of the respondents (37/46). They expressed a less assertive position, but with differences in viewpoints that made it possible to group them into three categories: the sceptics, the cautious and the enthusiasts. Enthusiasts differed from the convinced in their lack of commitment and in their level of knowledge of the subject, from the specialist researcher who did not wish to get involved in the debate to the layman who saw many possibilities linked to the use of GnEd. Many in the agricultural sector hoped that GnEd will be an additional solution to the challenges facing their industry. The sceptics included respondents who were very skeptical about the use of GnEd, expressing doubts about the technological interest of GnEd and the possibility of using these tools in animal welfare friendly systems. This category included a larger number of NGOs and a researcher specializing in the subject. The cautious were a more difficult category to characterize. It included more actors with little knowledge of the subject and who had often not expressed a clear position on the use of GnEd.

Beyond the controversy over GnEd, the question of the farming system in which these techniques are used was a central issue for many stakeholders. This was often equated with highly intensive and unsustainable systems. The question of their use in more agro-ecological and sustainable systems remained highly debated.

Acknowledgements

This study was financed by the GIS Avenir Elevage, L'institut de l'élevage, IFIP and SYSAAF.

References

Ashmore-Harris C. and Fruhwirth G. O. (2020) The clinical potential of gene editing as a tool to engineer cell-based therapeutics. Clinical and Translational Medicine 9(1):e15. https://doi.org/10.1186/s40169-020-0268-z.

European Commission (2021) Study on the status of new genomic techniques under Union law and in light of the Court of Justice ruling in Case C-528/16.

Delanoue E., Dockes A. C., Chouteau A., Roguet C., Philibert A. (2018) Social acceptability of French livestock production: debated issues and controversies on livestock production, points of view of multiple stakeholders. Inra Productions Animales 31(1):51-67. https://doi.org/10.20870/productions-animales.2018.31.1.2203.

Fan, Z. Mu Y., Li K., Hackett P.B. (2021) Safety evaluation of transgenic and genome-edited food animals. Trends in Biotechnology https://doi.org/10.1016/j.tibtech.2021.10.012

Jacobsen S. E., Sorensen M., Pedersen S.M., Weiner J. (2013) Feeding the world: genetically modified crops versus agricultural biodiversity. Agron. Sustain. Dev. 33(4):651-662. (Review) https://doi.org/10.1007/s13593-013-0138-9

Jinek M., Chylinski K., Fonfara I., Hauer M., Doudna J.A., Charpentier A. (2012) A Programmable Dual-RNA-Guided DNA Endonuclease in Adaptive Bacterial Immunity. Science 337(6096):816-821. https://doi.org/10.1126/science.1225829

Lu H., Yadav V., Zhong M., Bilal M., Taherzadeh M.J., Iqbal H.M.N. (2022) Bioengineered microbial platforms for biomass-derived biofuel production – A review. Chemosphere 288:132528. https://doi.org/10.1016/j.chemosphere.2021.132528

Norris A. L., Lee S.S., Greenlees K.J., Tadesse D. A., *et al.* (2020) Template plasmid integration in germline genome-edited cattle. Nat. Biotechnol. 38(2):163-164. https://doi.org/10.1038/s41587-019-0394-6.

495. Production of a germline ablated bull using CRISPR/Cas9 to target *NANOS3* in bovine zygotes

M.L. Mueller[1*], B.R. McNabb[2], J.R. Owen[3], S.L. Hennig[1], P.J. Ross[1] and A.L. Van Eenennaam[1]

[1]University of California, Davis, One Shields Avenue, Davis, CA 95616, USA; [2]University of California, Davis, School of Vet. Med., 944 Garrod Dr, Davis, CA 95616, USA; [3]University of California, San Diego, 9500 Gilman Drive, La Jolla, CA 92093, USA; macmueller@ucdavis.edu

Abstract

NANOS3 is known to be critical for germline development of both sexes in several organisms, but the role has not yet been characterized for male bovines. This study optimized a gene knockout (KO) approach using CRISPR/Cas9 in bovine zygotes to produce a live *NANOS3* KO bull. The co-injection of two selected guide RNA/Cas9 ribonucleoprotein complexes at 6 hours post fertilization achieved a high *NANOS3* KO rate in developing embryos. Subsequent embryo transfers resulted in a 25% pregnancy rate. A bull with no wildtype *NANOS3* alleles was produced and at sexual maturity he was found to have normal libido, an anatomically normal reproductive tract, and microscope evaluation of an ejaculate revealed seminal plasma only with no spermatozoa present, suggesting ablation of the germline. This study provides insights into bovine germline development and suggests the potential of *NANOS3* KO bulls to serve as hosts for germline complementation to enable donor-derived gamete production.

Introduction

NANOS3 is known to be critical for normal germline development in several organisms. Murine *Nanos3* is expressed by migrating primordial germ cells (PGCs) of both sexes, and the elimination of this gene results in the complete loss of male and female germ cells, but otherwise normal gonadal development (Tsuda *et al.* 2003). In humans, mutations in *NANOS3* are associated with germ cell loss and infertility (Julaton and Reijo Pera 2011). Also, a *NANOS3* knockout (KO) bovine female 200d-fetus was reported to phenocopy *Nanos3* KO mice (Ideta *et al.* 2016). However, the effect of knocking-out *NANOS3* on male bovine germline development has not yet been reported and no live *NANOS3* KO cattle have been produced to our knowledge. Characterization of the role of *NANOS3* in bulls and production of live *NANOS3* KO cattle will be important for further understanding bovine germline development. We hypothesized there would be a complete loss of germ cells in *NANOS3* KO bulls, making them ideal hosts for germline complementation to enable donor-derived gamete production. Germline complementation consists of using donor cells from one genetic background to complement or replace the germline of an otherwise sterile host of a different genetic background (Giassetti *et al.* 2019). Ultimately, germline complementation technology could enable more efficient and wide-spread dissemination of superior cattle genetics and/or traits improved via gene editing to enhance production efficiency. The objective of this study was to generate and characterize the germline development and reproductive capability of a CRISPR/Cas9 *NANOS3* KO bull.

Materials & methods

Guide RNA (gRNA) design and *in vitro* testing. Guide RNAs (gRNAs) targeting the bovine *NANOS3* exon one were designed and screened for potential off-target sites based on the bovine reference genome (ARS-UCD1.2) using sgRNA Scorer 2.0 (Chari *et al.* 2017) and Cas-OFFinder (Bae *et al.* 2014), respectively. Selected gRNAs had three or more total mismatches compared to the sequence for predicted off-target sites, and had at least one mismatch located in the seed region (8-10 bp proximal to the protospacer adjacent motif (PAM) site) of the gRNA (Hennig *et al.* 2020). Synthetic gRNA (Synthego, Menlo Park, CA, USA) cleavage efficiency was assessed by an *in vitro* cleavage assay, as previously described (Hennig *et al.* 2020).

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_495, © M.L. Mueller *et al.* 2022

Bovine embryo production. Bovine ovaries were obtained from a local processing plant. Cumulus-oocyte complexes (COCs) were aspirated from follicles and transferred to maturation media (IVF Bioscience, Falmouth, UK) for 21-24h. COC's were fertilized in SOF-IVF (Bakhtari and Ross 2014) with 2×10^6 sperm per ml for 6h. *In vitro* maturation and fertilization incubations were at 38.5 °C in a humidified 5% CO_2 incubator. Six hours post fertilization (hpf), presumptive zygotes were denuded by light vortex in SOF-HEPES (Bakhtari and Ross 2014) for 5 min. Zygotes were *in vitro* cultured (IVF Bioscience, Falmouth, UK) at 38.5 °C in a humidified atmosphere of 5% CO_2, 5% O_2, and 90% N_2 for 7-8 days.

gRNA *in vivo* testing. To determine the mutation rate for each individual gRNA and for dual guide RNA (dgRNA) combinations, 6-hpf zygotes were microinjected (i.e. laser-assisted cytoplasmic injection (Bogliotti *et al.* 2016)) with 6 pl of a solution containing 167 ng/µl of Cas9 protein (PNA Bio, Inc., Newbury Park, CA, USA) and 67 ng/µl of an individual gRNA or 67 ng/µl each of a combination of 2 gRNAs. The Cas9 protein and gRNA(s) were incubated together at room temperature for 30 minutes to form a ribonucleic protein (RNP) prior to microinjection. Microinjected embryos were incubated for 7-8 days and on day 7 embryos were scored for developmental stage reached. Blastocysts were individually collected in lysis buffer for DNA extraction. The *NANOS3* target region was amplified by two rounds of nested PCR using custom primers (Eurofins Genomics, Louisville, KY, USA). PCR products were visualized on a 1% agarose gel, purified using the QIAquick Gel Extraction Kit (Qiagen, Inc., Valencia, CA, USA) and Sanger sequenced (GeneWiz, South Plainfield, NJ, USA).

Embryo transfer (ET) & pregnancy monitoring. All animal procedures were approved by the Institutional Animal Care and Use Committee (IACUC) of the University of California (UC), Davis (#21513). Recipient females were synchronized in a standard fashion using a 7-day CoSynch + CIDR® protocol. Bovine embryos were *in vitro* produced and microinjected at 6-hpf with a dgRNA system, as described above. On day 7, microinjected blastocysts were selected and loaded individually into ET straws and transferred (one per recipient) into the recipients via a transcervical technique, placing the embryo in the uterine horn ipsilateral to the corpus luteum. Pregnancy was diagnosed initially via blood pregnancy-associated glycoproteins and followed by confirmatory transrectal ultrasound examinations. Pregnancies were routinely monitored with transrectal ultrasound.

Live animal evaluation. Blood samples were collected from new-born calves, DNA was extracted using a DNeasy® Blood & Tissue Kit (Qiagen, Inc., Valencia, CA, USA), and the *NANOS3* target region was amplified by PCR. PCR products were visualized on a 1% agarose gel, purified, and Sanger sequenced. Once the confirmed *NANOS3* edited males reached reproductive age (~12 months), a breeding soundness exam was conducted by UC Davis veterinarians. The breeding soundness exam followed the standards set forth by the Society for Theriogenology and included a general physical examination, inspection of reproductive organs, and semen collection via electroejaculation.

Results

Generation of CRISPR/Cas9 bovine NANOS3 KO embryos. Bovine *NANOS3* is a 2,633 bp gene with two exons (Figure 1A). Seven gRNAs were designed to target the larger exon 1 (451 bp), as it contains the coding region for the critical zinc finger binding domain. Four gRNAs (#1, #4, #5, #7) successfully cut the target region *in vitro*, so they were further tested *in vivo.* To determine blastocyst development rate and mutation efficiency, each gRNA was independently microinjected alongside Cas9 protein in groups of 30 zygotes, 6-hpf. Groups of 30 non-injected embryos were cultured as controls. All injected groups resulted in a blastocyst development rate of ≥20%, which was similar to other microinjected bovine embryo experiments (Hennig *et al.* 2020). Three individual gRNAs (#4, #5 and #7) resulted in >60% total mutation rate.

Figure 1. Bovine *CRISPR/Cas9 NANOS3* guide RNA design. (A) Diagram of bovine *NANOS3* gene and (B) guide RNA locations; checkmarks indicate successful *in vitro* cleavage assay and circles indicate guide RNAs selected for the dual guide RNA system.

Next, a dgRNA system, which is an efficient method for complete gene disruption in livestock species (Vilarino *et al.* 2017), was tested. gRNA #4 and #7 were selected based on their high individual mutation efficiencies (83 and 63%, respectively), and genomic locations, so that when co-injected they could introduce a deletion of 297 bp (Figure 1). The dgRNA_4+7 microinjections resulted in a similar blastocyst development rate to the individual gRNAs (19%) and >75% KO rate (i.e. 0 wild-type alleles) (n=28, 4 replicates).

Production of a CRISPR/Cas9 NANOS3 KO bull. For ET, *in vitro* produced bovine embryos were microinjected at 6-hpf with the dgRNA_4+7. On day 7 of *in vitro* culture, eight presumptive *NANOS3* KO blastocysts were selected and transferred into synchronized recipients. Twenty-one days after ET (28-day gestation), initial pregnancy rates were determined via a blood progesterone test (n=3/8, 38%). Ultrasound was used approximately one week later (35-day gestation) to confirm pregnancies (n=2/8, 25%), and around 60-days of gestation to determine foetal sex (one male and one female).

Two healthy calves one male (#838) and one female (#854) were born unassisted at the UC Davis Beef Barn in the fall of 2020; however, this paper will only focus on the resulting male (Figure 2). PCR and Sanger Sequence analysis of DNA extracted from blood indicated that #838 is a mosaic *NANOS3* KO, carrying at least four different KO alleles (Figure 2C). One allele has a large (298 bp) deletion and >3 alleles have small indels at one or both gRNA cut sites. All mutations present in #838 were predicted to disrupt *NANOS3* function.

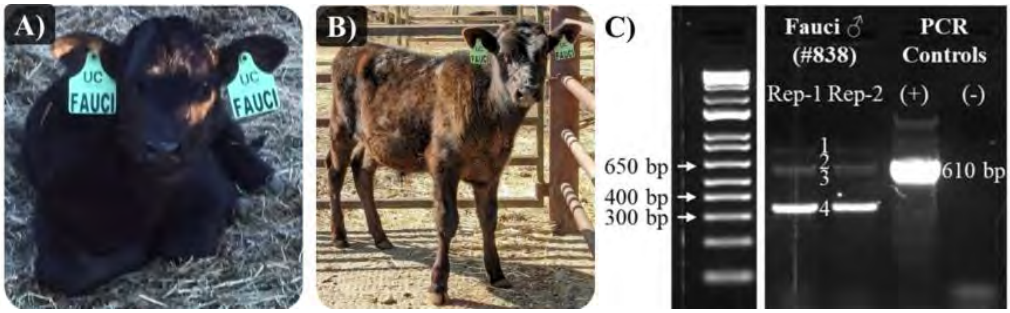


Figure 2. Production of a CRISPR/Cas9 *NANOS3* KO bull. Images of (A) 1-day-old #838 (named 'Fauci'), (B) 4-month-old #838, and (C) *NANOS3* PCR on DNA extracted from #838's blood. Numbers (1-4) represent different alleles visualized on a 1% agarose gel.

Phenotype of a CRISPR/Cas9 NANOS3 KO bull. To evaluate the potential for *NANOS3* KO cattle to serve as hosts for donor-derived gamete production, the reproductive development and capability of bull #838 was documented. The bull demonstrated normal libido and a 12-month breeding soundness exam found that #838 had an anatomically normal reproductive tract (i.e. accessory sex glands and penis) and normal testicular development, although the scrotal circumference (27 cm) was smaller than expected for age and breed matched controls. Microscope evaluation of an ejaculate obtained via electroejaculation revealed seminal plasma only with no spermatozoa present.

Discussion

We successfully produced a live bull calf with a targeted gene KO of *NANOS3* using a dgRNA gene editing approach in developing bovine zygotes. Although the bull was mosaic at the *NANOS3* target locus, no wildtype sequences remained, and all mutations resulted in *NANOS3* loss of function. The lack of spermatozoa in the ejaculate of this bull at sexual maturity supports the hypothesis that inactivation of *NANOS3* in male cattle will result in complete germline ablation. Importantly the bull had normal libido and an anatomically normal reproductive tract. This study demonstrates that *NANOS3* KO bulls have a phenotype that would be well suited to serve as hosts for germline complementation studies.

Acknowledgements

This work was supported by Select Sires, Inc. (Plain City, OH), and the National Institute for Food and Agriculture (NIFA) Predoctoral Fellowship (no. 2021-67034-35150) from the U.S. Department of Agriculture (USDA).

References

Bae S., Park J. and Kim J.S. (2014) Bioinformatics 30(10): 1473-1475. https://doi.org/10.1093/bioinformatics/btu048

Bakhtari A., and Ross P.J. (2014) Epigenetics 9(9): 1271-1279. https://doi.org/10.4161/epi.32087

Bogliotti Y.S., Vilarino M. and Ross P. (2016) J. Vis. Exp. 5(116): e54465. https://doi.org/10.3791/54465

Chari R., Yeo N.C., Chavez A. and Church G.M. (2017) ACS Synth Biol 6(5): 902-904. https://doi.org/10.1021/acssynbio.6b00343

Giassetti M.I., Ciccarelli M. and Oatley J.M. (2019) Annu. Rev. Anim. Biosci. 7(1): 385-401. https://doi.org/10.1146/annurev-animal-020518-115239

Hennig S.L., Owen J.R., Lin J.C., Young A.E., Ross P.J. *et al.* (2020) Sci. Rep. 10(1): 22309. https://doi.org/10.1038/s41598-020-78264-8

Ideta A., Yamashita S., Seki-Soma M., Yamaguchi R., Chiba S. *et al.* (2016) Sci. Rep. 6(1): 24983. https://doi.org/10.1038/srep24983

Julaton V.T., and Reijo Pera R.A. (2011) Hum. Mol. Genet. 20(11): 2238-2250. https://doi.org/10.1093/hmg/ddr114

Tsuda M., Sasaoka Y., Kiso M., Abe K., Haraguchi S. *et al.* (2003) Science 301(5637): 1239-1241. https://doi.org/10.1126/science.1085222

Vilarino M., Rashid S.T., Suchy F.P., Mcnabb B.R., Van Der Meulen T. *et al.* (2017) Sci. Rep. 7(1): 17472. https://doi.org/10.1038/s41598-017-17805-0

Part 30

Challenges – use of multi-omics data

496. Using allele-specific expression to uncover *cis*-regulation in bovine muscle

J.J. Bruscadin[1,2], T.F. Cardoso[2], W.J.S. Diniz[3], J. Afonso[2], M.M. de Souza[4], J. Petrini[5], V.H. da Silva[5], A. Zerlotini[6], L.L. Coutinho[5] and L.C.A. Regitano[2*]

[1]Center of Biological and Health Sciences. Post-Graduation Program of Evolutionary Genetics and Molecular Biology, Federal University of São Carlos, São Carlos, SP, Brazil; [2]Embrapa Pecuária Sudeste, São Carlos, SP, Brazil; [3]Department of Animal Sciences, Auburn University, AL, USA; [4]Department of Animal Science, Iowa State University, Ames, IA, USA; [5]Department of Animal Science, University of São Paulo/ESALQ, Piracicaba, SP, Brazil; [6]Embrapa Informática Agropecuária, Campinas, SP, Brazil; luciana.regitano@embrapa.br

Abstract

Allele-specific expression (ASE) analysis improves the understanding of transcription's *cis*-regulation. Herein, we used imputed SNPs along with RNA-Seq data from the *Longissiumus thoracis* muscle of 190 Nelore steers to identify functional *cis*-regulatory variants from ASE analysis. Using a Binomial Test, we identified 38,177 SNPs in ASE regions (ASE SNPs; FDR ≤0.05). We then searched for aseQTLs (SNPs potentially regulating the ASE) by comparing their heterozygosity to the measured allelic ratio under a Wilcoxon Rank Sum test. We identified 21,543 aseQTLs potentially regulating a total of 430 ASE SNPs (FDR ≤0.05). Based on a linear model, ASE SNPs and aseQTLs were associated with transcript abundance. We identified 3,333 SNPs acting as *cis*-eQTLs (FDR≤0.05). Results were integrated with previous ASE, functional regions, and meat quality-related differentially expressed genes data. This study described novel SNPs potentially regulating the transcription of genes that may affect beef traits.

Introduction

Genotypic and RNA sequencing (RNA-seq) data enable the identification of SNPs showing allele-specific expression (ASE SNPs), which consists in the imbalanced expression between alleles from each haplotype. Describing ASE regions improves the causal *cis*-regulatory variants screening (Guillocheau *et al.*, 2019). The ASE expression pattern responds to cumulative *cis*-regulatory mechanisms affecting the allele (Castel *et al.*, 2020), e.g. transcription factors (TFs) binding in accessible chromatin regions can be modulated by DNA methylation and histones modifications (Calo and Wysocka, 2013). These mechanisms can act on regulatory variants as aseQTLs, i.e. SNPs associated with the allelic imbalance (Battle *et al.*, 2014), and *cis*-eQTLs, i.e. SNPs regulating the expression of nearby transcripts (Gilad, Rifkin and Pritchard, 2008). In this study, we used a large dataset of imputed SNP genotypes to increase the power in detecting *cis*-regulatory variants in bovine muscle by using ASE approaches.

Materials & methods

Animal population, RNA, and DNA processing. The Nelore steers population was described elsewhere (Tizioto *et al.*, 2013). Here we used *Longissimus thoracis* muscle samples of a subset of this population (n=190) for which the RNA-Seq and high-density genotypic data were available. RNA was extracted from muscle tissue per sample, as previously reported (Cesar *et al.*, 2015). The RNA-Seq data were mapped to the bovine genome assembly ARS_UCD1.2 using the STAR software (Dobin *et al.*, 2013) with the default parameters. The WASP software (Van De Geijn *et al.*, 2015) was used to find and remove duplicated reads and those showing mapping bias. DNA extraction, quality control, and genotype calling were executed as described by Tizioto *et al.* (2013). BovineHD genotypes of the steers and a reference panel constructed from whole-genome sequence data of the 26 progenitors were used for imputation as previously described (Cardoso *et al.*, 2021). The imputed genotypes (4,812,912 SNPs) were phased using Beagle v.5.227 (Browning and Browning, 2008).

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_496, © J.J. Bruscadin *et al.* 2022

ASE-related SNPs identification. To identify ASE SNPs, we measured the allelic counts from heterozygous SNPs using ASE Read Counter from GATK v.4.2 (https://github.com/broadinstitute/gatk/). Ten reads mapping a SNP were required for ASE analysis. The allelic counts for each SNP were compared per animal using the Binomial Test against the expected value of 0.5. We used the Benjamini-Hochberg method to correct the False Discovery Rate (FDR) and significant SNPs were defined by FDR≤0.05. For the aseQTL analysis, we tested SNPs located up to 200 Kb upstream and downstream to each ASE SNPs position. Tests were performed as in a previous study (Bruscadin *et al.*, 2021), being significant those with FDR≤0.05. For the *cis*-eQTL screening, the RNA-Seq transcripts were quantified using Salmon (Patro *et al.*, 2017). The associations were conducted between transcripts abundance and either the respective ASE SNPs or the aseQTLs using a linear model implemented by MatrixEQTL (Xia *et al.*, 2017). The model included sequencing lane and the contemporary group as fixed effects, and age at slaughter as a covariate. Significant *cis*-eQTLs were the SNPs with associated FDR≤0.05.

Functional annotation and data integration. Annotation of ASE-related SNPs (ASE SNPs, aseQTLs, and *cis*-eQTLs) was conducted using the SNPEff software (Cingolani *et al.*, 2012). ASE genes were functionally annotated in the ClueGo plugin for Cytoscape (Bindea *et al.*, 2009). All ASE-related SNPs and ASE genes were compared with ASE data from *Limousin* cattle (Guillocheau *et al.*, 2019), Nelore cattle (de Souza *et al.*, 2020) and the GeneImprint database (https://www.geneimprint.com/). We also integrated with differentially expressed genes (DEGs) from the same population (Cesar *et al.*, 2016; Silva-Vignato *et al.*, 2017; Afonso *et al.*, 2019) and with Functional Annotation of Animal Genomes (FAANG) epigenetics data (Kern *et al.*, 2021).

Results

ASE SNPs in bovine muscle. Based on our criteria and quality control, 40,969 SNPs out of 4,812,912 SNPs were kept for further analyses. Significant ASE SNPs were frequent, being 60.9% of the tests significant (FDR≤0.05), comprising 38,177 ASE SNPs. Monoallelic expression was found for 37,227 SNPs in at least one sample. ASE SNPs were located in 7,448 genes (ASE genes), corresponding to 52.38% of all muscle-expressed genes of the bovine genome. Some of these ASE SNPs were predicted to have high impact effects as stop lost and gain, splice donor, and acceptor variations. The KEGG pathway enrichment retrieved 151 significant metabolic pathways (FDR≤0.05), including autophagy, endocytosis, focal adhesion, MAPK signaling pathway and proteasome.

aseQTLs were associated with the allelic imbalance of the identified ASE SNPs. We identified 21,543 aseQTLs associated with 430 ASE SNPs (FDR≤0.05). Forty aseQTLs located from 106,229,134 bp to 106,243,323 bp of BTA3 were associated with six ASE SNPs within the *NT5C1A* and *HEYL* genes. Most of the aseQTLs were located in intronic regions (50.4%), and some were located in high impact regions such as splice acceptor, start lost and stop gained regions. We found eight aseQTLs previously identified in the same population from HD Chip genotypes (Bruscadin *et al.*, 2021).

ASE SNPs and aseQTLs can affect transcripts' abundance. We identified 3,333 *cis*-eQTLs associated with the abundance of 758 transcripts. These *cis*-eQTLs included 2,098 ASE SNPs and 1,075 aseQTLs. Beta values (β, effects sizes) ranged from -823.08 to 862.80. The following *cis*-eQTLs presented the higher effects sizes: The A-allele of the BTA23:27522058 *cis*-eQTL was associated with the increased abundance of transcript *HSPA1A-201* (β=862.80, FDR=0.013; Figure 1A). *CKMT2-201* was positively regulated by the C-allele of the BTA7:81274253 *cis*-eQTL (β=98.88, FDR=0.008) (Figure 1B). Thirteen *cis*-eQTLs with identical genotypes located in the BTA29:49576764-49594787 window were associated with two different transcripts of the *TNNT3* gene, with its alternative allele upregulating the *TNNT3-205* (β=111.44, FDR=0.003; Figure 1C) and downregulating *TNNT3-206* (β=-823.08, FDR=0.010; Figure 1D)

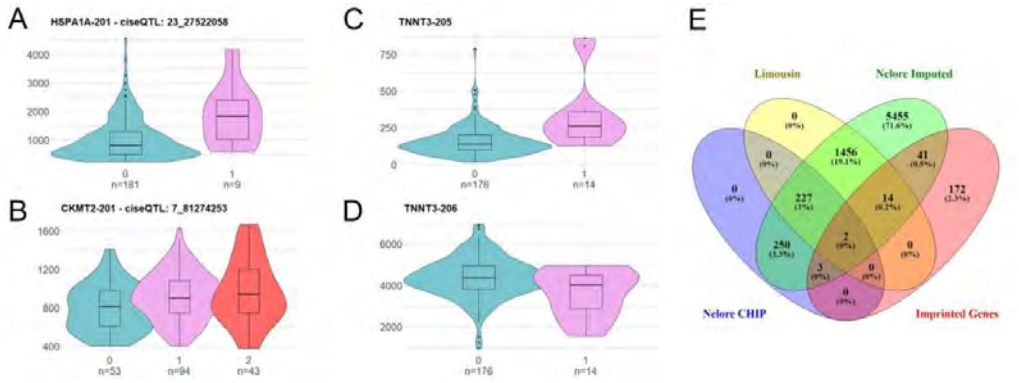


Figure 1. Violin plots showing the abundance of each transcript according to individual genotypes. The X-axis contains the genotypes, being '0' the reference homozygotes, '1' the heterozygotes and '2' the alternative homozygote. Y-axis contains the expression counts. (A) *HSPA1A-201*; (B) *CKMT2-201*; (C) *TNNT-205*, (D) *TNNT-206*. (E) Venn diagram of overlapping ASE genes.

Data integration showed functional evidence of ASE-related SNPs. A total of 482 ASE genes integrated with a previous study in Nelore (de Souza *et al.*, 2020), and 1,699 with *Limousin* cattle, corresponding to 90.94 and 80.17% of the ASE genes originally described, respectively (Figure 1E). We identified 5,495 novel ASE genes in bovine muscle. Furthermore, 51 ASE genes were previously reported as imprinted. Among them, seven in *Bos taurus,* of which five genes showed ASE SNPs with monoallelic expression in all samples of this study. We observed 1,450 ASE genes overlapping with DEGs previously identified in the muscle of the same Nelore population for fatty acid content (Cesar *et al.*, 2016), mineral content (Afonso *et al.*, 2019), and carcass quality traits (Silva-Vignato *et al.*, 2017). Additionally, we observed overlap with SNPs in accessible chromatin regions (ACR) or ChIP-Seq data from FAANG (Kern *et al.*, 2021), as for BTA23:27522058 *cis*-eQTL (Figure 1A) integrating with ACR, H3K27ac, H3K27me3, H3K4me1, and H3K4me3. BTA29:49580452 and BTA29:49580473 are *cis*-eQTLs for *TNNT3-205* (Figure 1C) and integrated with ACRs.

Discussion

Our results were consistent with known bovine ASE genes and indicated conservation of the ASE pattern in cattle muscle. The previous study in Nelore cattle identified 820 ASE SNPs (de Souza *et al.*, 2020), and 479 of them overlapped with our data (WASP filtered out the remaining). With the same population, but using imputation data, it was possible to identify 38,177 ASE SNPs. Additionally, we identified monoallelic expression in various ASE genes that overlapped with genes that are known to be imprinted in bovine. ASE genes are related to metabolic pathways underlying muscle homeostasis and were previously associated with meat-quality phenotypes. For example, *TNNT3* (Troponin 3) is related to muscle contraction (Wei and Jin, 2016) and *HSPA1A* (Heat Shock Protein Family A (Hsp70) Member 1A) was previously associated with ribeye area, water holding capacity, and cooking loss in the Nelore population used here (Tizioto *et al.*, 2013). The overlap of ASE-related SNPs and functional data from FAANG suggests potential mechanisms acting in transcript regulation, as the *cis*-regulation resulting from allele-specific permissive chromatin states. These epigenetic mechanisms need to be further explored in within-population experiments.

References

Afonso, J., Coutinho, L.L., Tizioto, P.C., Diniz, W. J. S., de Lima, A. O., *et al.* (2019). Sci Rep. 9(1):12715–12715. https://doi.org/10.1038/s41598-019-49089-x.

Battle, A., Mostafavi, S., Zhu, X., Potash J. B., Weissman, M. M., *et al.* (2014). Genome Res. 24(1):14–24. http://www.genome.org/cgi/doi/10.1101/gr.155192.113.

Bindea, G., Mlecnik,B., Hackl, H., Charoentong, P., Tosolini, M., *et al.* (2009). Bioinformatics 25(8):1091–1093. https://doi.org/10.1093/bioinformatics/btp101

Browning, B. L. and Browning, S. R. (2008). Am J Hum Genet. 84(2):210–223. https://doi.org/10.1016/j.ajhg.2009.01.005.

Bruscadin, J. J., de Souza, M. M., de Oliveira K. S., Rocha, M. I. P., Afonso, J., *et al.* (2021). Sci Rep. 11(1):7321–7321. https://doi.org/10.1038/s41598-021-86782-2.

Calo, E. and Wysocka, J. (2013). Mol Cell. 49(5):825–837. https://doi.org/10.1016/j.molcel.2013.01.038.

Cardoso, T. F., Coutinho, L. L., Bruscadin, J. J., Diniz, W. J. S., Petrini, J., *et al.* (2021). Genes, 12(1):67. https://doi.org/10.3390/genes12010067.

Castel, S. E., Aguet, F., Mohammadi, P., GTEx Consortium, Ardlie, K. G., *et al.* (2020). Genome Biol. 21(1):1–12. https://doi.org/10.1186/s13059-020-02122-z.

Cesar, A. S. M., Regitano, L. C. A., Koltes, J. E, Fritz-Waters, E. R., Lanna, D. P. D., *et al.* (2015). PLoS ONE. 10(6):e0128350. https://doi.org/10.1371/journal.pone.0128350.

Cesar, A.S.M., Regitano, L. C. A., Poleti, M. D., Andrade, S. C. S., Tizioto, P. C., *et al.* (2016). BMC Genomics. 17(1):961. https://doi.org/10.1186/s12864-016-3306-x.

Cingolani, P., Platts, A., Wang, L. L., Coon, M., Nguyen, T., *et al.* (2012). Fly, 6(2): 80–92. https://doi.org/10.4161/fly.19695.

Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., *et al.* (2013). Bioinformatics. 29(1):15–21. https://doi.org/10.1093/bioinformatics/bts635.

Gilad, Y., Rifkin, S.A. and Pritchard, J.K. (2008). Trends Genet. 24(8):408–415. https://doi.org/10.1016/j.tig.2008.06.001.

Guillocheau, G. M., Hou, A. E., Meersseman C., Esquerré D., Rebours E., *et al.* (2019). Sci Rep. 9(1):4297–4297. https://doi.org/10.1038/s41598-019-40781-6.

Kern, C.,Wang, Y., Xu, X., Pan, Z., Halstead, M., *et al.* (2021). Nat Commun. 12(1):1821–1821. https://doi.org/10.1038/s41467-021-22100-8.

Patro, R., Duggal, G., Love, M., Irizarry, R. A. and Kingsford, C. (2017). Nat Methods. 14(4):417–419. https://doi.org/10.1038/nmeth.4197.

Ren, G., Jin, Wenfei., Cui, K., Rodrigez, J., Hu, G., *et al.* (2017). Mol Cell. 67(6):1049-1058.e6. https://doi.org/10.1016/j.molcel.2017.08.026.

Silva-Vignato, B., Coutinho, L. L., Cesar, A. S. M., Poleti, M. D., Regitano, L. C. A., *et al.* (2017). BMC Genomics. 18(1):506–506. https://doi.org /10.1186/s12864-017-3897-x.

de Souza, M. M., Zerlottini, A., Rocha, M. I. P., Bruscadin, J. J., Diniz, W. J. S., *et al.* (2020). Sci Rep. 10(1):10204–10204. https://doi.org/10.1038/s41598-020-67089-0.

Tizioto, P.C., Decker, J. E., Taylor J. F., Schnabel R. D., Mudadu M. A. *et al.* (2013). Physiol Genomics. 45(21):1012–1020. https://doi.org/10.1152/physiolgenomics.00066.2013.

Van De Geijn, B., McVicker, G., Gilad, Y., and Pritchard, J. K., *et al.* (2015). Nat Methods. 12(11):1061–1063. https://doi.org/10.1038/nmeth.3582.

Wei, B., and Jin, J.-P. (2016). Gene, 582(1):1–13. https://doi.org/10.1016/j.gene.2016.01.006.

Xia, W., Zhu, X-W., Mo, X-B., Wu, L-F., Wu, J., *et al.* (2017). Hum Genet. 136(4): 451–462. https://doi.org/10.1007/s00439-017-1771-1.

497. Enrichment of causative variants in tissue-specific and shared ATAC-Seq peaks in cattle

C. Yuan[1], L. Tang[1], T. Lopdell[2], C. Oget-Ebrad[1], G. Costa Monteiro Moreira[1], J.L. Gualdron[1], Z. Cheng[3], M. Salavati[3,4], D.C. Wathes[3], M.A. Crowe[5], GplusE consortium[5], W. Coppieters[6], C. Charlier[1], T. Druet[1], M. Georges[1] and H. Takeda[1*]

[1]Unit of Animal Genomics, University of Liège, 4000 Liège, Belgium; [2]Research and Development, Livestock Improvement Corporation, 3240 Hamilton, New Zealand; [3]Royal Veterinary College, AL9 7TA Herts, United Kingdom; [4]Roslin Institute, Royal School of Veterinary Studies, Midlothian, United Kingdom; [5]School of Veterinary Medicine, University College Dublin, Dublin 4, Ireland; [6]GIGA Genomics platform, University of Liège, 4000 Liège, Belgium; htakeda@uliege.be

Abstract

To profile putative regulatory variants in the bovine and test their validity, we first characterized ~1 million putative regulatory elements by 'Assay for Transposase Accessible Chromatin using Sequencing (ATAC-Seq)' in 77 distinct types of tissues and cells. The elements collectively accounted for 10.3% of the genome, from which we identified 938,374 common DNA variants. We then analyzed bovine blood and liver RNA-Sequencing data and identified 5,336 and 3,613 *cis*-expression Quantitative Trait Loci (eQTL), respectively. Credible sets of DNA variants that drive these *cis*-eQTL were enriched in tissue-specific as well as tissue-shared putative regulatory elements identified by ATAC-Seq. These results pave the way to utilizing ATAC-Seq information to improve genomic selection.

Introduction

Genomic selection has had a tremendous impact on livestock selection in the last ten years (e.g. Georges *et al.*, 2019). The accuracy of selection nonetheless remains inferior to what may be expected given the heritability of the selected traits. This could be due to a number of factors including that all variants are generally given an equivalent weight in computing the additive relationship between animals for Restricted Maximum Likelihood analyses or equivalent prior probabilities of variant effects in Bayesian approaches. Only a minority of variants are indeed causative, with the remainder being (at best) passenger variants in linkage disequilibrium (LD) with one or more of the causative variants. It is generally believed that knowing the causative variants, or at least those that are more likely to be, may help further improve the accuracy of genomic selection.

Amongst causative variants, coding variants (including missense, nonsense, frameshift and splice site variants) are easily recognized. However, these account for only a limited part of the genetic variance for complex phenotypes. It is increasingly apparent that most of the genetic variation for complex traits is due to regulatory variants that act either by perturbing the expression profile of genes located in *cis*, or by perturbing the gene regulatory network and affecting the expression profile of a restricted number of core genes in *trans*. Active regulatory elements can be recognized by virtue of epigenetic features including chromatin accessibility, histone marks, transcriptional activity, their participation in loop structures, and transcription factor occupancy. However, regulatory variants remain difficult to identify, as the effect of polymorphisms on the functionality of proximal and distant *cis*-acting regulatory elements is difficult to predict.

In an effort to identify putative regulatory variants in the bovine, we herein report: (1) the generation of a comprehensive catalogue of bovine putative regulatory elements identified by Assay for Transposase Accessible Chromatin using Sequencing (ATAC-Seq); (2) the generation of a catalogue of common DNA

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_497, © C. Yuan *et al.* 2022

variants that map to identified proximal and distal putative regulatory elements; (3) the demonstration that variants that drive expression Quantitative Trait Loci (eQTL) in liver and blood are more likely to map to regulatory elements that are active in the relevant tissue, and hence that variants in these regulatory elements are more likely to be causative.

Materials & methods

ATAC-Seq. Libraries were constructed following Corces *et al.* (2017) using 78 tissues (65 tissue-types) collected from two juvenile males and 11 archived frozen samples (4 tissue types). Additionally, fifteen public ATAC-Seq data sets (eight tissue/cell-types) were included (Fang *et al.*, 2019, Foissac *et al.*, 2019, Halstead *et al.*, 2020, Johnston *et al.*, 2021). The ATAC-Seq reads were mapped to the bovine reference genome (ARS-UCD1.2) using Bowtie2 (Langmead and Salzberg, 2012) and peaks called with Macs2 (Zhang *et al.*, 2008). Peaks were merged across samples to build a reference chromatin accessibility map following Meuleman *et al.* (2020). The complexity of peak patterning was decomposed using non-negative matrix factorization (NMF) (Meuleman *et al.*, 2020). We used transcriptome data from Fang *et al.* (2020) for a correlation analysis.

eQTL analysis. 224 blood and 176 liver RNA-Seq samples collected from Holstein-Friesian cows were analyzed (Wathes *et al.*, 2021, Lee *et al.*, 2021). The reads were mapped to the reference genome using HISAT2 (Kim *et al.*, 2015). Transcript abundances were estimated using StringTie (Pertea *et al.*, 2015) based on a reference gene annotation (Bos_taurus.ARS-UCD1.2.105.gtf). The data were normalized within sample using DESeq2 (Anders and Huber, 2010) and across samples using inverse normal transformation. Genotypes obtained with 770 K single nucleotide polymorphism (SNP) arrays were further imputed to whole genome using Minimac4 (Das *et al.*, 2016) with the Damona population as a reference, yielding 8.4 million SNPs and 1.3 million insertion-deletions (INDELs) with minor allele frequency (MAF)>0.02. *Cis*-eQTL analyses were conducted using residuals corrected for country and hidden PEER factors (Stegle *et al.*, 2010) under an additive model using QTLtools (false discovery rate (FDR)<0.001) (Delaneau *et al.*, 2017). Enrichment analyses of eQTL variants in ATAC-Seq peaks were conducted following Trynka *et al.* (2017).

Results

Generating a catalogue of bovine cis-acting gene regulatory elements. 89 in-house and 15 publicly available ATAC-Seq data sets (including 77 tissue/cell types) were analyzed and identified on average 166,999 ATAC-Seq peaks per sample. To build a chromatin accessibility reference map, overlapping peaks were merged across the samples, yielding 977,261 chromatin accessible sites. 16 components (representing immune, liver, muscle, central nervous system, etc.) were identified from the 977,261 chromatin accessibility × 104 sample matrix by NMF (Figure 1). Public RNA-Seq data for 58 tissue types were used to search for correlations between chromatin accessibility and gene expression levels to connect regulatory elements with target genes. Highly significant enrichment of positive correlations for genes ≤250 Kb from ATAC-Seq peaks were observed, i.e. potential enhancer effects, without any enrichment for negative correlations, which would be expected for silencers.

Generating a catalogue of common variants mapping to cis-acting regulatory elements. A catalog of DNA variants called from 264 Holstein-Friesian animals containing ~7.6 million SNPs and ~1.2 million INDELs with MAF>0.05 were used. Of these 29,494 (0.34%) mapped to proximal and 909,074 (10.4%) to distal putative regulatory elements accounting respectively for 0.37 and 9.9% of sequence space, hence providing no strong evidence for detectable purifying selection.

Identifying bovine cis-eQTL in blood and liver. To identify putative causative variants affecting gene expression, eQTL analyses using 224 blood and 176 liver RNA-Seq data were performed. 5,336 (37% of

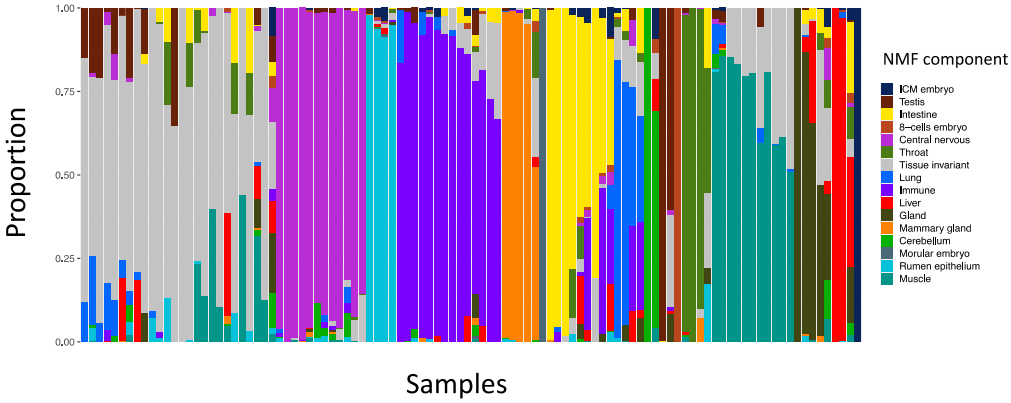


Figure 1. Identification of 16 components by NMF-decomposition of a 977,261 chromatin accessibility × 104 sample matrix and importance of the components in the bovine tissues.

genes) and 3,613 (23%) significant eQTL in the blood and liver, respectively, were obtained. The proportion of blood-eQTL that would also operate in liver (π_1) was estimated by following Storey and Tibshirani (2003) at 74%, while the proportion of liver-eQTL that would also operate in blood was estimated at 81%. These unexpectedly high values could in part be due to blood irrigating the liver.

eQTL-driving variants are enriched in variants mapping to ATAC-Seq peaks. It was assumed that, if variants mapping to ATAC-Seq peaks were indeed enriched in causative variants, they should be enriched in credible sets of variants driving *cis*-eQTL effects (i.e. the lead *cis*-eQTL SNPs and syntenic variants in LD with it). The significance of the overlap between *cis*-eQTL credible sets and ATAC-Seq peaks was evaluated by NMF components (by assigning each peak to its dominant NMF component). Credible variant sets ($r^2 \geq 0.8$) for 3,336 blood-specific *cis*-eQTL were significantly ($P<10^{-5}$) enriched in variants overlapping ATAC-Seq peaks that were assigned to the immune (1.21-fold enrichment), tissue-shared (1.17-fold) and pulmonary (1.15-fold) components. Credible sets for 1,613 liver-specific *cis*-eQTL were enriched in variants mapping to ATAC-Seq peaks assigned to the hepatic (1.42-fold), tissue-shared (1.21-fold) and pulmonary (1.15-fold) components. Permutation testing conditional on proximal vs distal status of the ATAC-Seq peaks indicated that the observed enrichment was not due to an enrichment of variants mapping to promotor regions irrespective of ATAC-Seq chromatin accessibility.

Discussion

These results indicate that genetic variants causing *cis*-eQTL are enriched in variants that map to ATAC-Seq peaks marking accessible chromatin regions, and hence likely to active regulatory elements, in the corresponding tissue. This suggests that sets of variants overlapping ATAC-Seq peaks are enriched for regulatory variants, and hence potentially driving organismal phenotypes. Assigning such variants with extra weight in genomic selection may therefore improve its accuracy. However, this putative improvement will dependent on: (1) the proportion of variants mapping to ATAC-Seq peaks that are indeed causative for a trait of interest; and (2) the proportion of causative regulatory variants that map to ATAC-Seq peaks. Further analyses are in progress to estimate these features, and to evaluate the impact of the identified variants sorted by the NMF component on genomic selection.

References

Anders S., and Huber W. (2010) Genome Biol 11(10):R106. https://doi.org/10.1186/gb-2010-11-10-r106

Corces M.R., Trevino A.E., Hamilton E.G. Greenside P.G., Sinnott-Armstrong N.A. *et al.* (2017) Nat Methods 14(10):959-962. https://doi.org/10.1038/nmeth.4396

Das S., Forer L., Schönherr S., Sidore C., Locke A.E. *et al.* (2016) Nat Genet 48(10):1284-1287. https://doi.org/10.1038/ng.3656

Delaneau O., Ongen H., Brown A.A., Fort A., Panousis N.I. *et al.* (2017) Nat Commun 8:15452. https://doi.org/10.1038/ncomms15452

Fang L., Liu S., Liu M., Kang X., Lin S. *et al.* (2019) BMC Biol 17(1):68. https://doi.org/10.1186/s12915-019-0687-8

Fang L., Cai W., Liu S., Canela-Xandri O., Gao Y. *et al.* (2020) Genome Res 30(5):790-801. https://doi.org/10.1101/gr.250704.119

Foissac S., Djebali S., Munyard K., Vialaneix N., Rau A. *et al.* (2019) BMC Biol 17(1):108. https://doi.org/10.1186/s12915-019-0726-5

Georges M., Charlier C., Hayes B. (2019) Nat Rev Genet 20(3):135-156. https://doi.org/10.1038/s41576-018-0082-2

Halstead M.M., Ma X., Zhou C., Schultz R.M., Ross P.J. (2020) Nat Commun 11(1):4654. https://doi.org/10.1038/s41467-020-18508-3

Halstead M.M., Kern C., Saelao P., Wang Y., Chanthavixay G. *et al.* (2020) BMC Genomics 21(1):698. https://doi.org/10.1186/s12864-020-07078-9

Johnston D., Kim J., Taylor J.F., Earley B., McCabe M.S. *et al.* (2021) BMC Genomics 22(1):14. https://doi.org/10.1186/s12864-020-07268-5

Kim D., Langmead B., and Salzberg S.L. (2015) Nat Methods 12:357-360. https://doi.org/10.1038/nmeth.3317

Langmead B., and Salzberg S.L. (2012) Nat Methods 9:357-359. http://doi.org/10.1038/NMETH.1923

Lee Y.L., Takeda H., Costa Monteiro Moreira G., Karim L., Mullaart E. *et al.* (2021) PLoS Genet 17(7):e1009331. http://doi.org/10.1371/journal.pgen.1009331

Meuleman W., Muratov A., Rynes E., Halow J, Lee K. *et al.* (2020). Nature 584:244-251. https://doi.org/10.1038/s41586-020-2559-3

Pertea M., Pertea G.M., Antonescu C.M., Chang T., Mendell J.T. *et al.* (2015) Nat Biotechnol 33(3):290-295. https://doi.org/10.1038/nbt.3122

Stegle O., Parts L., Durbin R., Winn J. (2010) PLoS Comput Biol 6(5):e1000770. https://doi.org/10.1371/journal.pcbi.1000770

Storey J.D. and Tibshirani E. (2003) Proc Natl Acad Sci USA 100(16):9440-9445. https://doi.org/10.1073/pnas.1530509100

Trynka G., Westra H., Slowikowski K., Hu X., Xu H. *et al.* (2015) Am J Hum Genet 97(1):139-152. http://doi.org/10.1016/j.ajhg.2015.05.016

Wathes, D.C., Becker F., Buggiotti L., Crowe M.A., Ferris C. *et al.* (2021) *Ruminants* 1(2):147-177. https://doi.org/10.3390/ruminants1020012

Zhang Y., Liu T., Meyer C.A., Eeckhoute J., Johnson D. *et al.* (2008) Genome Biol 9: R137. https://doi.org/10.1186/gb-2008-9-9-r137

498. Preliminary investigation into the association of members of the rumen biome with production traits in Afrikaner cattle

E.D. Cason[1*], J. Alom[2], J. Barnard[1], M. Fair[1], P.D. Vermeulen[3] and F.W.C. Neser[1]

[1]Department of Animal Science, University of the Free State, Bloemfontein, South Africa; [2]Department of Microbiology and Biochemistry, University of the Free State, Bloemfontein, South Africa; [3]Office of the Dean: Natural and Agricultural Sciences, University of the Free State, Bloemfontein, South Africa; casoned@ufs.ac.za

Abstract

Rumen microbiota plays an important role in ruminant metabolism as well as immunity development. As such, the microbiome serves as a link between diet and animal productivity. The aim of this study was to examine fecal samples of Afrikaner cattle at weaning age using 16S rRNA targeted metagenomics to determine if there are any association with above average performing animals and the rumen microbiome. Rumen microbiomes were similar with no statistically significant differences observed between above and below average performing animals. However, various individual microbial genera were found to be associated performance, such as the lactic acid bacteria (LAB) *Lactobacillus* and *Enterococcus.* This preliminary study highlights both the importance of the transmission of important maternal rumen microbiota to the offspring and the importance of host genetics for the colonization of these organisms.

Introduction

Feed costs can represent more than half of the total cost of production in the beef cattle industry (Mathews and McBride, 2011). This has led to various novel methods to improve feed efficiency and nutrient utilization. Rumen microbial communities are directly correlated to several host phenotypes, including feed efficiency (Myer *et al.*, 2017). These microorganisms produce glucogenic, lipogenic and aminogenic precursors that partake in energy metabolism (Hungate, 1975). This can represent up to 70% of the metabolic energy utilized by the animal (Bergman, 1990). As such, the microbiome may serve as a link between diet and animal productivity (Mullins *et al.*, 2013).

In general, it has been commonly accepted that diet is the main driving force shaping the gut microbial diversity (Spor *et al.*, 2011). However, more and more research are starting to indicate host genetics is also an important factor in determining the composition of gut microbiota. In mice, 18 quantitative trait loci (QTLs) were found to be associated with the relative abundance of microbial taxa in the gut (Benson *et al.*, 2010). More recently, 58 single nucleotide polymorphisms (SNPs) were also identified in reference to 33 fecal microbial taxa in humans (Turpin *et al.*, 2016). In ruminants it has been shown that rumen microbial diversities could be influenced by breed. For example, differences in the composition of rumen microbiota between Holstein and Jersey was detected while being fed on the same diet (Paz *et al.*, 2016). Of note are the recent reports of heritability estimates for rumen bacteria based on 750 lactating Holstein cows (Difford *et al.*, 2018). Based on all these finding, it is becoming more evident that host genetics are influencing the rumen microbiota, however, even though much research has been done on the influence of diet on the rumen microbiome, the connection between the microbiome, production traits and genetics in beef cattle remain unclear.

The aim of this preliminary study was to determine whether the rumen microbiome can serve as a predictor of performance in beef cattle.

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_498, © E.D. Cason *et al.* 2022

Materials & methods

Sample and data collection. A total of 32 Afrikaner cattle were raised under the same feedlot conditions at the Paradys Experimental Farm facility of the University of the Free State, South Africa. When the animals were 257±27 days of age, approximately 10 g of fecal matter was collected from each animal. Fresh fecal matter was collected from the interior of the produced cowpat, in an effort to minimize environmental contamination, in a 50 ml Falcon tube. All samples were stored at -20 °C until further processing. Weaning weights of all animals were weighed at 223±16 days and all weights were adjusted to 205 days as well as for age of the dam.

DNA extraction and targeted 16S metagenomic sequencing. DNA was isolated from each fecal samples using ZymoBIOMICS DNA Miniprep Kit (Zymo Research). To assess the bacterial profiles, the V3-V4 region were amplified using the primers 341F (5'-CCTACGGGNGGCWGCAG-3') and 805R (5'-GACTACHVGGGTATCTAATCC-3'). Paired-end sequencing (2×300 bp) of amplicons was performed using the Illumina MiSeq at mrDNA (Shallowater, TX, USA).

Microbial composition analysis. Read data was analyzed using QIIME 2 (https://qiime2.org) (Caporaso *et al.*, 2010). Paired-end sequences imported into QIIME 2 were quality-controlled and combined using DADA2. The settings for quality control were based on the reads' quality distribution along the length of the sequence. This grouped sequences into amplicon sequence variants (ASVs) based on 99% sequence similarity. After building the ASV table and removing chimeras, taxonomy was assigned using the SILVA 132 database (SILVA).

Statistical analysis. The processed gene sequence data was further analyzed using RStudio, an integrated development environment (IDE) for the programming language R. The following R packages were used during the processing of the taxonomic data: Phyloseq (McMudie and Holmes 2013), Tidyverse (Wickham 2017), Vegan (Dixon, 2003).

To facilitate comparative analysis, animals were divided into three groups; one standard deviation above the mean weaning weight of all sampled animals (Above), one standard deviation below the mean weaning weight of all sampled animals (Below) and between one standard deviation above and below (Average). Permutational multivariate analysis of variance (PERMANOVA), using the adonis function in the R Vegan package, was used to determine significance in dissimilarity matrices across samples in these three categories. Differential abundance analysis between the three groups were performed using DeSeq2 (v 1.22.1) (Love *et al.*, 2014).

Results

Microbial diversities across all groups were similar (PERMANOVA, $P>0.05$). The microbial consortia were dominated by the phyla Actinobacteria, Bacteroidetes, Euryarchaeota and Firmicutes (Figure 1). DeSeq2 analysis identified various microbial genera associated with above and below average weaning weight performance. *Lactobacillus*, *Enterococcus* spp. and members of a Lachnospiracaea group were all associated with improved performance while *Bacillus* sp., *Psychrobacillus* and *Cellulosilyricum* sp. were associated with lower performance (Figure 2).

Discussion

In cattle, Firmicutes and Bacteroidetes are usually the most dominant phyla, however, there is no known 'normal' rumen microbiome (Henderson *et al.*, 2015). In this study, all animals were from the same breed and subjected to the same environmental conditions and diet. Although no statistically significant differences, for microbiome communities as a whole, were observed between above and below average performers

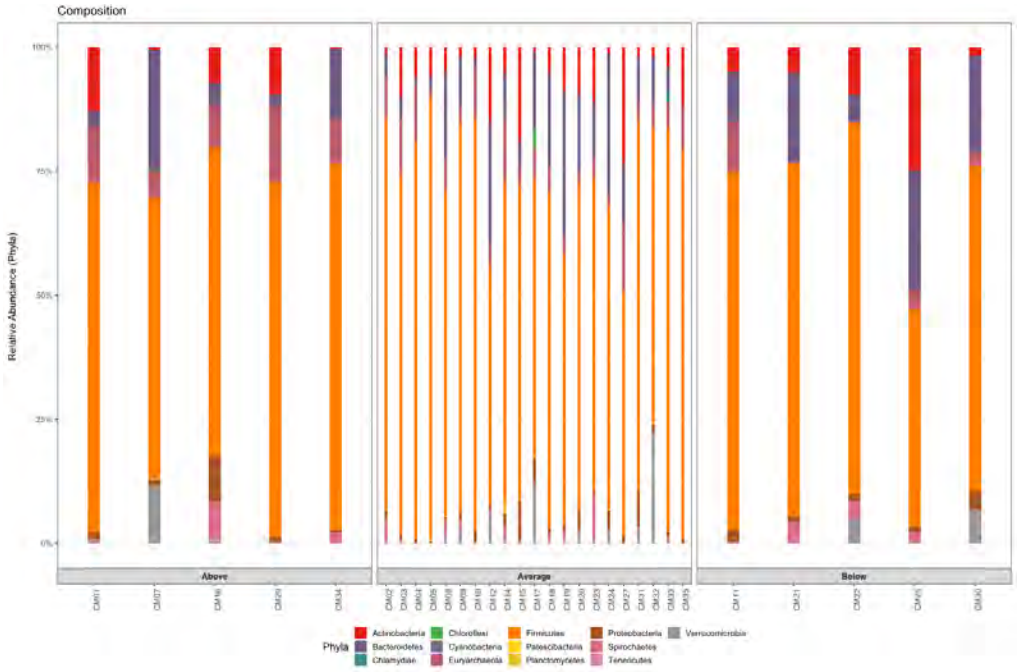


Figure 1. Phyla level bacterial diversity of fecal samples. Samples above and below one standard deviation from the mean were grouped together.

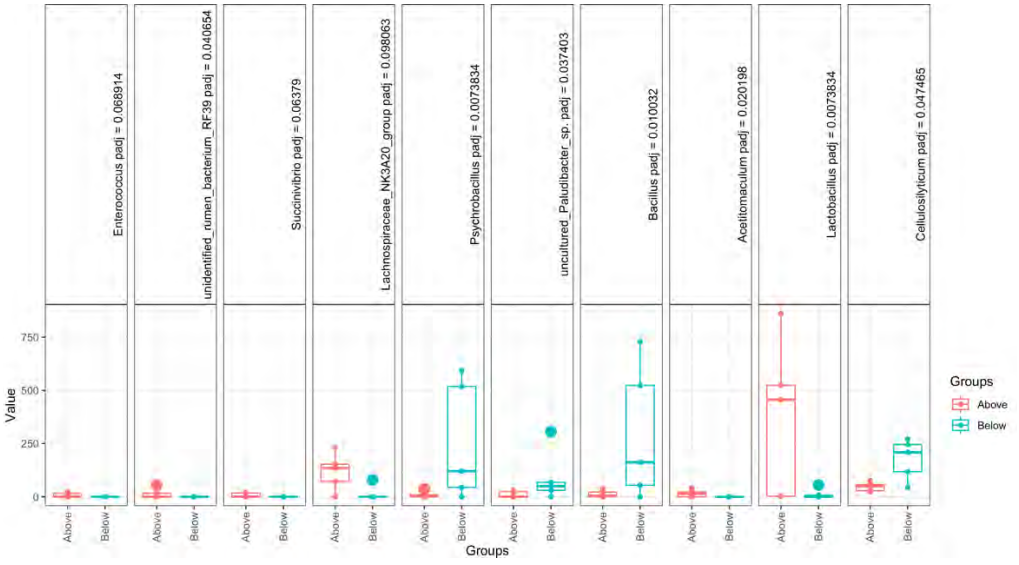


Figure 2. Differentially abundant genera between above and below average weaning weight performing samples.

for weaning weight (PERMANOVA, $P>0.05$), important differences were determined on a community member level. *Lactobacillus* and *Enterococcus* spp. were significantly (DeSeq2, $P<0.05$) more abundant in the samples from above average performing animals. A stable load of lactic acid bacteria (LAB), such as *Lactobacillus* and *Enterococcus,* has been shown to improve weight gain and immunocompetence in young calves as well as improve intestinal health and feed conversion efficiency (Al-Saiady, 2010). Also, early colonization by LAB in the intestine may decrease the adherence of pathogens to the intestinal mucosa.

Most studies on cattle production and LAB focus on feed additives. In this preliminary study the LAB was naturally occurring and appear to have contributed to the increased performance of the animals, however, a larger study with more animals would be required to validate these findings. The intestine of a newly born calf is sterile, thereafter a complex microbial ecosystem is established in the large intestine as the animal grows to maturity. The initial microbiota of the calf is thus influenced by the microbiota of the birth canal, the dam's fecal microbiota and the colostrum fed. As such, this preliminary study appears to hint that some part of the heritable maternal microbiome could result in improved production in the progeny and that host genetics might play an important role in improving colonization rates of LAB in the intestinal track.

References

Al-Saiady M.Y. (2010) J Anim Vet Adv 9(3):604–609. http://doi.org/10.3923/javaa.2010.604.609

Benson A.K., Kelly S.A., Legge R., Ma F., Low S.J., *et al.* (2010) Proc Natl Acad Sci USA 107(44):18933–18938. https://doi.org/10.1073/pnas.1007028107

Bergman E.N. (1990) Physiol Rev 70(2):567–590. https://doi.org/10.1152/physrev.1990.70.2.567

Caporaso J.G., Kuczynski J., Stombaugh J., Bittinger K., Bushman F.D., *et al.* (2010) Nat Methods 7:335–336. https://doi.org/10.1038/nmeth.f.303

Difford G.F., Plichta D.R., Lovendahl P., Lassen J., Noel S.J., *et al.* (2018) PLoS Genet 14:e1007580. https://doi.org/10.1371/journal.pgen.1007580

Dixon, P. (2003) J Veg Sci 14:927-930. https://doi.org/10.1111/j.1654-1103.2003.tb02228.x

Henderson G., Cox F., Ganesh S., Jonker A., Young W., *et al.* (2015) Sci Rep 5:14567. https://doi.org/10.1038/srep14567

Hungate R. (1975) The rumen microbial ecosystem. Annu Rev Ecol Syst 6:39–66. https://doi.org/10.1146/an nurev.es.06.110175.000351

Love M.I., Huber W., Anders S. (2014) Genome Biol 15(12):550.
https://doi.org/10.1186%2Fs13059-014-0550-8

Mathews K., and McBride W. Economic Information Bulletin March 2011, Washington, DC. Available at: https://www.ers.usda.gov/publications/pub-details/?pubid=44532

McMurdie P.J., and Holmes S. (2013) PLoS ONE 8(4): e61217. https://doi.org/10.1371/journal.pone.0061217

Mullins C.R., Mamedova L.K., Carpenter A.J., Ying Y., Allen M.S., *et al.* (2013) J. Dairy Sci. 96(9):5872–5881. https://doi.org/10.3168/jds.2013-6775

Myer P.R., Wells J.E., Smith T.P.L., Kuehn L.A., Freetly H.C. (2017) J Anim Sci 95 (7):3215-3224. https://doi.org/10.2527/jas.2016.1059

Spor A., Koren O., Ley R. (2011) Nat Rev Microbiol 9:279–90. https://doi.org/10.1038/nrmicro2540

Turpin W., Espin-Garcia O., Xu W., Silverberg M.S., Kevans D., *et al.* (2016) Nat Genet 48(11):1413–7. https://10.1038/ng.3693

Wickham H., Averick M., Bryan J., Chang W., McGowan L.D., *et al.* (2019) J Open Source Softw, 4(43)1686. https://doi.org/10.21105/joss.01686

499. Maternal effect on the metagenomic composition determining the intramuscular fat content in rabbits

A. Zubiri-Gaitán[1*], M. Mora[2], C. Casto-Rebollo[1], M.A. Santacreu[1], A. Blasco[1], P. Hernández[1] and N. Ibañez-Escriche[1]

[1]*Institute for Animal Science and Technology, Universitat Politècnica de València, 46022 Valencia, Spain;* [2]*Institute of Agrifood Research and Technology (IRTA), Animal Breeding and Genetics, Torre Marimon C-59, km 12.1, 08140 Caldes de Montbui, Barcelona, Spain;* agzugai1@upvnet.upv.es

Abstract

The intramuscular fat (IMF) content is one of the main parameters influencing meat quality and its deposition is influenced by the gut microbiome composition. This composition is determined by several effects including environmental, genomic, and maternal. A divergent selection experiment for IMF has been developed in rabbits. Embryos from the high-IMF (H) and low-IMF (L) lines were randomly transferred to dams of both lines. To evaluate the maternal effect, kits from the same genetic line but born from dams of the H and L lines were compared. The analysis performed confirmed the existence of an important maternal effect that affected the metagenome composition of the lines and, ultimately, their IMF content. The microbial genes discriminating between the groups were different when the genetic line of the kits varied. However, there were no important differences in their metabolomic pathways.

Introduction

The intramuscular fat (IMF) content is one of the main parameters influencing meat quality. The genetic and environmental factors influencing this trait have been extensively studied and, over the last few years, the gut metagenome has gained importance. Numerous studies suggest that the former plays an important role in the host metabolism affecting, among others, the energy harvest and fat deposition (Bäckhed *et al.* 2004; Khanal *et al.* 2020; Turnbaugh *et al.* 2006). The microbiome composition, and its metagenome, can be easily modified by environmental factors such as diet, and is also determined by the host's genome (Wen and Duffy 2017). The mother plays an important role in determining the microbiome composition as well. During labor, the microbiome composition of the kits is influenced due to the exposure to the vaginal microbiota of the mother and, after that, due to the microbiota of the milk ingested during breastfeeding (Wen and Duffy 2017). A better understanding of the maternal effects on the microbiome composition is key to optimize the phenotypic improvement of the populations.

A divergent selection experiment for IMF content in the *Longissimus thoracis et lumborum* muscle of rabbits has been developed at the Universitat Politècnica de València. The selection originated two lines, one with high-IMF (H) content and the other with low-IMF (L) content. These lines are exceptional to study the effects influencing the trait. The complete selection experiment is described in Martínez-Álvaro *et al.* (2016).

The objective of this study was to analyze the importance of the maternal effect in the metagenome composition of the divergent lines and, ultimately, on their intramuscular fat content.

Materials & methods

To evaluate the maternal effect, embryos from the H and L lines were randomly transferred to dams of both lines from the 10th generation of divergent selection. To do so, 14 dams from the H and 9 dams from the L line were used as embryo donors, and 8 dams from the H line and 6 from the L line were used as embryo recipients. The embryos were transferred to the recipients 72 h after the fertilization. Genotype information from a previous generation of selection was used to select fixed SNPs in each line, which would

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_499, © A. Zubiri-Gaitán *et al.* 2022

allow identifying the genetic line of the kits. After the kits were born, DNA was extracted from an ear tissue sample, and the genotype of the selected SNPs was determined by HRM (high-resolution melting). Finally, 4 groups were obtained according to the genetic line of the recipient dam and of the kit. The groups were identified as HH, HL, LH, and LL. The first letter of each group indicates the genetic line of the dam and the second one is the genetic line of the kit. For this experiment, 73 rabbits from the 4 groups were used, distributed as follows: 20 HH, 19 HL, 19 LH, and 15 LL.

The kits were weaned at 28 days, housed collectively with their own litter, and fed *ad libitum* with a commercial diet until the slaughter at 9 weeks of age. Animals were deprived of food 4 h before the sampling. After slaughter, cecum content samples were collected, homogenized, and stored at -80 °C. The metagenome of the samples was sequenced at Sistemas Genómicos (Valencia, Spain). Sequencing was performed on NextSeq 550 sequencer (Illumina) with 150 bp paired-end chemistry. The data obtained were processed using the SqueezeMeta fully automatic metagenomics analysis pipeline (Tamames and Puente-Sánchez 2019), using the coassembly mode. The microbial genes and their metabolic routes were identified by homology to the Kyoto Encyclopaedia of Genes and Genomes (KEGG).

The dataset obtained had 7,195 microbial genes (MG). After exploratory analyses, all MGs that were undetected (i.e. 0 counts) in at least one sample were discarded. The final dataset had 4,572 MGs that accounted for more than 99.8% of the total counts of each sample. To deal with the compositional nature of the metagenomic data (Gloor *et al.* 2017), the counts was transformed using additive log-ratio (ALR) transformation. The reference gene for the ALR was selected according to the criteria established by Greenacre *et al.* (2021).

The maternal effect was evaluated by performing the following group comparisons: HH vs LH to evaluate the effect when the kits were from the H line, and HL vs LL to evaluate it when the kits were from the L line. The MGs discriminating between groups were identified by Projection to Latent Structures – Discriminant Analysis (PLS-DA), using the program SIMCA (Sartorius; Umea; Sweden). The model for each comparison was adjusted and the MGs responsible for the discrimination between maternal lines were identified using the variable importance in projection (VIP) value. A first variable selection step was performed to reduce the noise of the data, in which all the MGs with a VIP<1 were discarded. The models were then re-estimated and their adjustment was assessed with the coefficient of determination (R^2). The predictive capacity of the models was estimated with the Q^2, which is equivalent to a R^2 after cross-validation (CV). A misclassification table was elaborated for each model, based on the CV results. Finally, the MGs selected for both models were compared.

The same group comparisons were made to study the maternal effect on the IMF content. The differences were estimated as the phenotypic differences between groups. The model included the fixed effect of group and sex. Bayesian inference was applied using the program Rabbit (Institute for Animal Science and Technology, UPV, Valencia, Spain). The marginal posterior distributions of the phenotypic differences were obtained. The parameters of the posterior distributions taken into consideration account were the median of the phenotypic difference and the probability of the absolute value of the difference of being greater than zero.

Results

Figure 1 shows the projection of the samples into the space formed by the first two latent variables of the PLS-DA models (score plot). The number of MGs selected after the selection step were 1,763 for the model comparing HH vs LH, and 1,677 for the one comparing HL vs LL. Both models presented good adjustment (R^2>97%) and predictive capacity (Q^2>85%). Also, they were able to correctly classify all the samples after the cross-validation procedure, as evidenced by the percentage of classification (Table 1). The clear discrimination between the maternal lines in both models (i.e. independently of the genetic line of the kits) confirms the existence of a maternal effect that affected the metagenome composition of the lines.

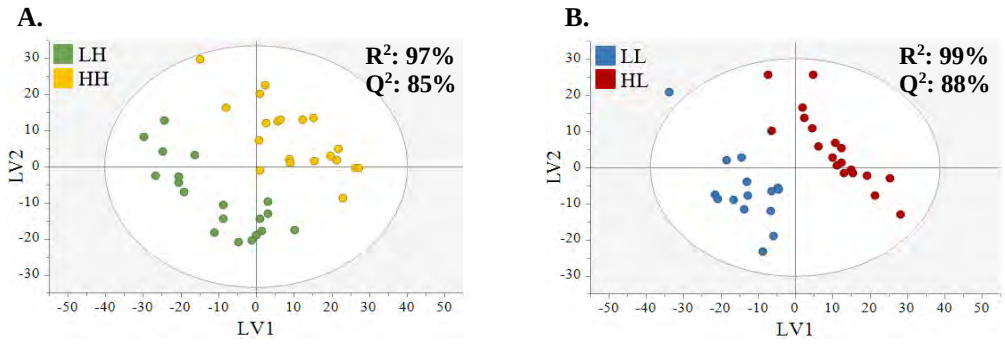


Figure 1. Projection of the samples in the space formed by the latent variables of each PLS-DA. HH, LH, HL, LL: the first letter indicates the genetic line of the dam and the second the genetic line of the kit. H: high-IMF; L: low-IMF. A. Model 1: maternal effect when the kits are from the H line. B. Model 2: maternal effect when the kits are from the L line.

Once the influence of the maternal effect was confirmed, the MGs responsible for the discrimination in both models were compared. Most of the MGs discriminating between the groups were different in both models (around 10% of overlap, Venn diagram not shown), meaning that the maternal effect on the metagenome composition is different according to the genetic line of the kit. However, when comparing the pathways in which those MGs were involved, common subgroups were found. The most affected pathways were carbohydrate transport and metabolism, amino acid transport and metabolism, energy production and conversion, cell envelope biogenesis, and signal transduction mechanisms, among others.

Differences were found in the IMF content of the kits, depending on the genetic line of their mother. The phenotypic difference between HH and LH was 0.03 g IMF/100 g of *LTL*, with a probability of 76% of being greater than 0. On the other hand, the difference between HL and LL was 0.04 g IMF/100 g of *LTL*, with an 82% probability of being greater than 0.

Discussion

The metagenome composition is influenced by several factors, including environmental, genetic, and maternal (Wen and Duffy 2017). The divergent lines used for this experiment were contemporarily reared, under the same environmental conditions, and were fed the same diet. Hence, no difference in their metagenome composition is expected due to environmental factors. Martínez-Álvaro *et al.* (2021) confirmed the influence of the metagenome composition on their IMF content, and the link between the host's genome and its metagenome, in a previous metagenomic analysis performed on these lines. However,

Table 1. Misclassification tables of the models comparing HH vs LH and HL vs LL.

Model 1: HH vs LH[1]			**Model 2: HL vs LL[2]**		
Misclassification table	**HH[3]**	**LH[3]**	**Misclassification table**	**HL[3]**	**LL[3]**
HH	19	0	HL	19	0
LH	0	20	LL	0	15
% classification	100%	100%	% classification	100%	100%

[1] Model 1: maternal effect when the kits are from the high-IMF line (H).

[2] Model 2: maternal effect when the kits are from the low-IMF line (L).

[3] HH, LH, HL, LL: the first letter indicates the genetic line of the dam, and the second one is the genetic line of the kit.

they could not differentiate between the genetic and the maternal effects. The experimental design of the present study and the comparisons made ensured the possibility of separating the maternal and the genetic effects, avoiding their confusion.

The mother alters the metagenome composition of the kit through several mechanisms, such as feeding. The microbiome composition of the milk, but also its nutrients, could influence the core metagenome of the kits. An analysis performed on the milk composition of these lines found important differences in their fat (higher in the H line) and lactose (higher in the L line) composition (results not shown). The former could explain the higher IMF content found on the kits born from a high-IMF line dam, and it provides evidence of a maternal environmental effect influencing the metagenome composition. Besides, the probabilities of difference in the IMF content, mentioned in the results section, give enough confidence to support the relevance of the maternal effect. Finally, El Nagar *et al.* (2021) also found a relevant maternal effect for IMF in a previous quantitative genetic study, reinforcing the results found in this experiment.

Several studies have suggested the importance of the maternal effect on the microbiome composition. However, this is the first study that allows to measure it directly because of the exceptional animal material used and the experimental design. Hence, the differences found between the groups can be directly attributed to the maternal effect on the gut metagenome composition. The results of this study proved that the mother plays an important role in determining the mentioned composition of the kits. Even though the results showed that the maternal effect was different according to the genetic line of the kit, a preliminary analysis suggested that the pathways affected were the same. A thorough biological analysis is still needed to elucidate the mechanisms by which the genetic and maternal factors influence on the gut metagenome composition and, ultimately, on the IMF deposition of these lines.

Acknowledgements

This work was funded by the PID2020-115558GB-C21 project of the Spanish National research plan, and the AICO/2020/349 project of the Consellería of Innovation, Universities, Science and Digital Society.

References

Bäckhed F., Ding H., Wang T., Hooper L.V., Gou Y.K., *et al.* (2004) Proc Natl Acad Sci U S A. 101(44):15718–15723. https://doi.org/10.1073/pnas.0407076101

El Nagar A.G., Sosa-Madrid B.S., Hernández P., Blasco A., and Ibañez-Escriche N. (2021) Proc. of XIX Jornadas sobre Producción Animal – On-line. ISBN: 978-84-09-30674-9

Gloor G.B., Macklaim J.M., Pawlowsky-Glahn V., and Egozcue J.J. (2017) Front Microbiol. 8(NOV):2224. https://doi.org/10.3389/fmicb.2017.02224

Greenacre M., Martínez-Álvaro M., and Blasco A. (2021). Front Microbiol. 12(OCT):2625. https://doi.org/10.3389/fmicb.2021.727398

Khanal P., Maltecca C., Schwab C., Fix J., Bergamaschi M., *et al.* (2020) Genet Sel Evol. 52(1):1–13. https://doi.org/10.1186/s12711-020-00561-7

Martínez-Álvaro M., Hernández P., and Blasco A. (2016) J Anim Sci. 94(12):4993–5003. https://doi.org/10.2527/jas.2016-0590

Martínez-Álvaro M., Zubiri-Gaitán A., Hernández P., Greenacre M., Ferrer A., *et al.* (2021) Commun Biol. 4(1):1–10. https://doi.org/10.1038/s42003-021-02784-w

Tamames J., and Puente-Sánchez F. (2019) Front Microbiol. 10(JAN):1–10. https://doi.org/10.3389/fmicb.2018.03349

Turnbaugh P.J., Ley R.E., Mahowald M.A., Magrini V., Mardis E.R., *et al.* (2006) Nature. 444(7122):1027–1031. https://doi.org/10.1038/nature05414

Wen L., and Duffy A. (2017) J Nutr. 147(7):1468S-1475S. https://doi.org/10.3945/jn.116.240754

500. The gut microbiome profile varies among two maternal rabbit lines with different longevity

I. Biada*, N. Ibáñez-Escriche, A. Blasco and M.A. Santacreu

Institute for Animal Science and Technology, Universitat Politècnica de València, 46022 València, Spain; ibiada@posgrado.upv.es

Abstract

Analyses of the gut microbiome profile using bacterial 16S rRNA gene between two maternal rabbit lines (A and LP) and two groups of animals from line LP with different longevity (LLP and HLP, females up to four parities and with more than four parities, respectively) were carried out. For that, 298 soft faeces samples collected from 191 does were sequenced. After filtering and quality control, the sequences were analysed using QIIME2. Differences between lines A and LP (DLINES) and between groups LLP and HLP (DLP GROUPS) in both alpha and beta diversities indexes at first parity were observed. The PLS-DA analysis allowed us to identify 45 and 37 Amplicon Sequence Variants (ASVs) that were able to discriminate between lines A and LP, and between the two groups within line LP. *Eubacteriacea* and *Ruminococcus* were the main taxa presented in the final ASVs for both DLINES and DLP GROUPS comparison. Preliminary results support that gut microbiome profile is related to longevity and it could be used to improve resilience in rabbits commercial lines.

Introduction

Resilience and longevity are important traits on rabbits' production but are difficult as criteria of selection. Indirect selection for resilience has been proposed in rabbits by Argente *et al.* (2019) using a low heritability criterion (residual litter size variance), and line LP of this experiment was founded using longevity as a criterion for better resilience, but longevity shows also low heritability (Piles *et al.* 2006) and is recorded late in the animal's life when they die or are culled (Sánchez *et al.* 2008). An alternative criterion to increase longevity and resilience in rabbit production can be based on knowledge on the rabbit's gut microbiome. The gut microbiome is important in regulating traits related to health and welfare and it plays a key role in immune system development (Round and Mazmanian 2009). Therefore, the aim of this research was to compare the microbiota of two lines with different longevity to study the association between gut microbiota profile and longevity and resilience. Results from this study would allow stablishing a selection criteria based on the gut microbiome composition for improving the rabbit resilience.

Materials & methods

Animals. Two contemporaneous rabbit lines, line LP and A, with different criteria of foundation and currently selected for litter size at weaning were used in this experiment. Particularly, females of the 15th generation of line LP ('resilient' line) and of the 50th generation line A (standard line). Line LP was founded with females with a minimum of 25 parities and having an average prolificacy per parity close to the average of the Spanish commercial population in 2002 (Sánchez *et al.*, 2008). Line A has a New Zealand White origin. Females were housed at the farm of the Universitat Politècnica de València.

Sample collection and 16S rRNA gene sequencing. From both lines, soft faeces samples were taken from the anus after their first parities and the females were allowed to reach the maximum number of parities possible. The complete data set included 298 samples records, involving 191 does (117 of line LP, and 74 of line A), and the data set was recorded from 23/04/2019 to 20/07/2020. The 16S rRNA gene profile of the samples were analysed using amplicon sequencing protocol targeting the V3 and V4 hypervariable regions (459bp) of bacterial 16S rRNA gene with the primers designed surrounding conserved regions. The

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_500, © I. Biada *et al.* 2022

microbial 16S rDNA profiles of the DNA extracts were analysed using MiSeq 16S rRNA gene-sequencing protocols (Illumina).

Bioinformatic analyses. The sequences of the amplicons were analysed using QIIME2 software version 2021.11 (Bolyen 2019) and sequences were clustered into ASVs using DADA2, the final table had 59072 ASVs. For Taxonomic annotation we used a Naive Bayes classified. This classifier was pre-trained on the SILVA database ('silva-138-99-nb-classifier').

Statistical analyses. Two comparisons were carried out: (1) the difference of the gut microbiota profile between lines A and LP at first parity (DLINES); (2) the difference of the gut microbiota profile at fist parity between animals of line LP reached five parities or more and the ones who died or were sacrificed before (DLP GROUPS).

Diversity measures. Alpha and beta diversity metrics were calculated after samples were subsampled without replacement (rarefied) to 24,000 sequences per sample. Four α-diversity metrics were computed: Shannon's diversity index (Shannon and Weaver 1949), Observed features (DeSantis *et al.* 2006), Evenness (Pielou 1966) and Faith's Phylogenetic Diversity (Faith 1992). A permutational multivariate analysis of variance (PERMANOVA) statistical test was applied on four dissimilarity matrixes (Bray-Curtis, Jaccard, Unweighted and Weighted Unifrac) to determine gut microbiome profile differences. To rule out differences due to a high degree of dispersion (within-group variance), we made a permutation test (PERMDISP).

PLS-DA (partial least squares discriminant analyses). The PLS-DA was performed at ASVs level after filtering samples <100 reads, ASVs<10 and not present in at least 20% of samples. Additive log ratio transformation was applied to the filtered abundance table (Greenacre 2018). The reference variable was those with the lowest coefficient of variation and a high mean of counts. Procrustes analysis was performed to verify the quality of the transformation (Greenacre *et al*., 2021). The Procrustes correlation was 0.98. PLS-DA was performed to address differences in the microbiota profiles with 'mixOmics' package of R (Lê Cao *et al.* 2011). Those ASVs with balance error rate (BER) of the classification between groups for each analysis (DLINES and DLP GROUPS) reaching the minimum value were identified. Finally, a permutation test and a confusion matrix were computed to check the robustness of PLS-DA for both analyses.

Results

Diversity measures. Table 1 showed differences on the Shannon and Observed Features α-diversities indexes in the two comparisons performed (DLINES and DLP GROUPS). The difference in Pielou's Evenness α-diversities index was found for the DLINES comparison, whereas for the DLP GROUPS comparison was in Faith paid index (Table 1). The Bray-Curtis, Jaccard Weighted and Unweighted Unifrac Matrixes β-diversities indexes showed differences between DLINES groups for the PERMANOVA test (P=0.013, P=0.001, P=0.005 and P=0.005). However, the weighted and unweighted Unifrac results could be due to high degree of dispersion according to the permutation test (P=0.004 and P=0.012). For the DLP GROUPS only the unweighted Unifrac showed clear differences between groups with a P-value of 0.013 for the PERMANOVA test and a P-value of 0.57 for the permutation test.

PLS-DA. PLS-DA models with the minimum BER identified 45 ASVs able to discriminate between groups for the DLINES comparison and 37 ASVs for the DLP GROUPS comparison (Figure 1).

The confusion matrix showed 80% true positives (sample correctly predicted to the line) and 86.5% true negatives (correctly predicted that the sample did not belong to the assigned line) for DLINES comparison. In the case of DLP GROUPS comparison these values were 79% true positives and 72.4% true negatives.

Table 1. Kruskal-Wallis test of Observed Features, Pielou Evenness, Shannon and Faiths pd α-diversity indexes for the line A and LP (DLINES) and line LP with less than 5 and more than 5 parities (DLP GROUPS).

Comparison	DLINES		DLP GROUPS	
	α-diversity	*P*-value	α-diversity	*P*-value
Pielou evenness	6.73	0.009	0.999	0.317
Faiths pd	1.91	0.166	6.19	0.013
Observed features	6.79	0.009	3.82	0.051
Shannon	4.25	0.039	3.09	0.079

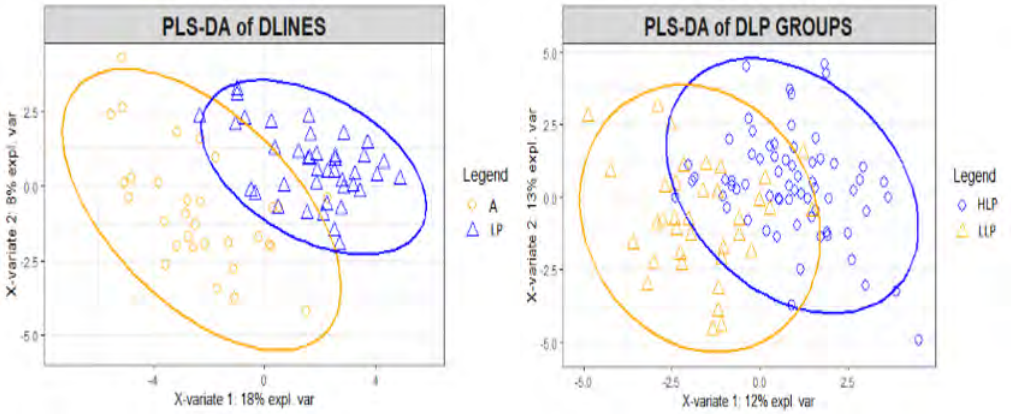


Figure 1. Representation of individual plots of first and second components of partial least squares discriminant analysis (PLS-DA) between line A and LP (DLINES, left plot) and between line LP with less than 5 (LLP) and more than 5 parities (HLP, right plot), each line is represented by a different colour, contoured by 95% confidence ellipses, n=71 for DLINES and n=103 for DLP GROUPS.

The permutation test showed ~50% of true positives and negatives for DLINES and DLP GROUPS. These results would indicate the ASVs identified in the PLS-DA not randomly discriminate between groups. After taxonomy annotation, *Eubacteriaceae* (35.5%), *Ruminococcus* (20%) *and Akkermansiaceae* (15.5%) were the main families of the 45 ASVs classifying between line A and LP (DLINES). For the DLP GROUPS comparison the main families identified were *Oscillospiraceae* (35.13%), *Eubacteriaceae* (16.2%), *Lachnospiraceae* (10.8%) and *Ruminococcus* (10.8%).

Discussion

The results of this study showed that the observed operational taxonomic units (OTUs) and Shannon α-diversities indexes representing community richness were greater in line A (less longevity) than in line LP. This will be in disagreement with the common assumption that animals with the greater diverse microbial community are potentially have better gut health (Fouhse *et al.*2016). However, in our study also the individuals of line LP with less longevity (<5 parties) showed a greater diversity than those with more longevity for the Faith pd index. Similar findings were presented in a recent study (Savietto *et al.* 2020). In addition, beta diversity metrics and PLS-DA analysis evidenced also differences between the gut microbiota profile of lines and individuals of different longevity. The ASVs identified in DLINES's PLS-DA belong mainly to Akkermansiaceae family. In rabbits Akkermansiaceae, especially Akkermansia muciniphila, plays a protective role in diet induced obesity and other diseases (Li *et al.* 2021, unpublished

data) and is also considered as beneficial microbe to immunity (Naito *et al.* 2018). The Eubacteriaceae and Ruminococcaceae families were present in DLINES and DLP GROUPS's PLS-DA. Eubacteriaceae and three groups from the family Ruminococcaceae belong to the bacterial taxa that are most common across the entire rabbit gastrointestinal tract (Cotozzolo *et al.* 2020) Ruminococcaceae are also a butyrate-producing bacteria which are associated with a healthy gut microbiome (Funosas *et al.* 2021).

Funding

This research was supported by MEC (AGL2014-55921-C2-1-P and PID2020-115558GB-C21) and Generalitat Valenciana (AICO/2020/349).

References

Argente M.J., García M.L., Zbynovska M., Petruska P., *et al.* (2019) Anim. 13 (10): 2348–55. https://doi.org/10.1017/S1751731119000302.

Bolyen E., Rideout J.R., Dillon M.R., Bokulich N.A., *et al.* (2019) Nat. Biotechnol 37 (8): 852–57. https://doi.org/10.1038/s41587-019-0209-9.

Camarinha-Silva, A., Maushammer, M., Wellmann, R., Vital, M., *et al.* (2017). Genetics, 206, 1637-1644. https://doi.org/10.1534/genetics.117.200782

Cotozzolo E., Cremonesi P., Curone G., Menchetti L., *et al.* (2020) Animals 11 (1): E31. https://doi.org/10.3390/ani11010031.

DeSantis T.Z., Hugenholtz P., Larsen N., Rojas M., *et al.* (2006) Appl. Environ. Microbiol. 72 (7): 5069–72. https://doi.org/10.1128/AEM.03006-05.

Faith D.P. (1992) Biol. Conserv 61 (1): 1–10. https://doi.org/10.1016/0006-3207(92)91201-3.

Fouhse J., Zijlstra R.T., and Willing B. (2016) Animal Frontiers 6 (January): 30. https://doi.org/10.2527/af.2016-0031.

Funosas G., Triadó-Margarit X., Castro F., Villafuerte R., *et al.* (2021) Sci. Rep 11 (1): 766. https://doi.org/10.1038/s41598-020-80782-4.

Greenacre M. (2018) Compositional Data Analysis in Practice. CRC Press.

Greenacre M., Martinez Álvaro M., and Blasco A. (2021). Front. microbiol 12 (October): 727398. https://doi.org/10.3389/fmicb.2021.727398.

Lê Cao K., Boitard S., and Besse P. (2011) BMC Bioinform. 12 (1): 253. https://doi.org/10.1186/1471-2105-12-253.

Martínez-Álvaro M., Zubiri-Gaitán A., Hernández P., Greenacre M., *et al.* (2021) Commun. Biol. 4 (1): 1246. https://doi.org/10.1038/s42003-021-02784-w.

Naito Y., Uchiyama K., and Takagi T. (2018) J Clin Biochem 63 (1): 33–35. https://doi.org/10.3164/jcbn.18-57.

Pielou E. C. (1966) J. Theor. Biol. 13 (December): 131–44. https://doi.org/10.1016/0022-5193(66)90013-0.

Piles M., Garreau H., Guarro O., Larzul C., *et al.* (2011) Livest. Sci. 136 (2): 201–6. https://doi.org/10.1016/j.livsci.2010.09.009.

Round J.L., and Mazmanian S.K. (2009) Nat. 9 (5): 313–23. https://doi.org/10.1038/nri2515.

Sánchez J.P., Theilgaard P., Mínguez C., Baselga M. (2008) J. Anim. Sci. 86 (3): 515–25. https://doi.org/10.2527/jas.2007-0217.

Savietto D., Paës C., Cauquil L., Fortun-Lamothe L. et al (2020) Anim. 14 (11): 2253–61. https://doi.org/10.1017/S1751731120001305.

Shannon C.E., and Weaver W. (1949) The Mathematical Theory of Communication. Champaign, University of Illinois Press, IL, US.

501. Comparing methods to summarize gut microbiota composition in estimating microbiability of host phenotypes in swine

Y. He[1], C. Maltecca[1*], J. Howard[2], Y. Huang[2], K. Gray[2] and F. Tiezzi[1,3]

[1]Department of Animal Science, North Carolina State University, 120 W Broughton Dr, Raleigh, NC 27607, USA; [2]Smithfield Premium Genetics, Rose Hill, NC 28458, USA; [3]Department of Agriculture, Food, Environment and Forestry (DAGRI), University of Florence, Piazzale delle Cascine 18, 50144 Firenze, Italy; cmaltec@ncsu.edu

Abstract

This study aimed to investigate eight approaches (four kernel functions, two distance methods, and two ordination methods) for creating covariance matrices to summarize microbiome information among animals and assess their performance in estimating trait microbiability in three commercial swine breeds. We collected rectal swabs and measured several growth and carcass composition traits on 651 pigs (Duroc: n=205; Landrace: n=226; Large White: n=220) at market weight. Based on the matrix used, microbiability estimates ranged from 0.07 to 0.21 and 0.12 to 0.53 for Duroc, 0.03 to 0.21 and 0.05 to 0.44 for Landrace, and 0.02 to 0.24 and 0.05 to 0.52 for Large White pigs averaged over traits in the model with sire, pen, and microbiome, and model with only microbiome, respectively. We observed differences in the contribution to trait microbiability estimation across the eight matrices.

Introduction

Gut microbiota have recently received great attention in livestock research. In swine, the gut microbiome composition has been associated with variations in host phenotypes, such as meat quality and carcass traits (Khanal *et al.*, 2021), feed efficiency (Camarinha-Silva *et al.*, 2017), and feeding behaviour (He *et al.*, 2022). This proportion of phenotypic variance accounted by the microbiota composition was termed 'microbiability' by Difford *et al.* (Difford *et al.*, 2016). The microbial information of individuals in a group can be summarized in a square matrix M and inform a mixed linear model. Most studies used the method introduced by Ross *et al.* (2013) to construct the microbial relationship matrix for microbiability estimation and phenotype prediction (Camarinha-Silva *et al.*, 2017; Difford *et al.*, 2016; Khanal *et al.*, 2021). However, due to the complexity and special features of the microbiome data (Coenen *et al.*, 2020), other methods, such as distance matrix and ordination based on the sample covariance, could be considered as alternative approaches to summarize the microbial information. In recent studies, Saborío-Montero *et al.* (2021) evaluated different matrices based on ruminal microbiota in estimating microbiability for methane emission in cattle. Unfortunately, there is still a lack of relevant research in swine to the best of our knowledge. Thus, we applied eight approaches to build the microbial matrices and compare their contributions to estimating the trait microbiability in three pig breeds.

Materials & methods

Animal and data. This study included 651 boars from either three breeds: Duroc (DR; n=205), Landrace (LR; n=226), and Large White (LW; n=220). After weaning, pigs were kept by single-breed groups in 64 pens and fed the same pelleted feed throughout the fattening period. Growth and carcass composition traits, including body weight (BW), and ultrasound backfat thickness (BF), muscle depth (MD), and intramuscular fat (IMF) content, were measured on live animals when they reached the market weight at an average age of 156±3 days. Rectal swabs were taken from each pig at 158±4 days of age and subjected to 16S rRNA gene sequencing. The procedures for sequencing and processing the sequence data are described in our previous studies (He *et al.*, 2022; Lu *et al.*, 2018). The amplicon sequence variant (ASV) table was subjected to quality control. The ASVs were removed if the overall counts were less than 1000 across all

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_501, © Y. He *et al.* 2022

samples and with a prevalence rate less than 0.05. A total of 824 ASVs passed the quality control and was used in the analyses.

Microbial similarity matrices construction. We used eight methods from three types to create the M matrix, which captures the resemblance among individuals in a population based on the abundance of their gut microbiota. First, we used four general kernel functions described in (Montesinos-López *et al.*, 2021): Linear Kernel (LK), Polynomial Kernel (PK), Gaussian Kernel (GK), and Arc-cosine Kernel with one hidden layer (AK1) for creating the M matrix. The ASV counts were log-transformed, then centred and scaled before being used in the kernel functions provided in (Montesinos-López *et al.*, 2021). Then, we created two distance matrices: Bray-Curtis (BC) and Jaccard (JA), on the log-transformed ASV count data using the *vegdist* function of *vegan* package (Dixon, 2003) in R. The resulting distance matrices were square with zero on the diagonal and values ranging from zero to one on the off diagonal. The M matrix was then calculated by one minus each distance matrix. Lastly, two ordination methods: Metric Multidimensional Scaling (MDS; known as Principal Coordinates Analysis) and Detrended Correspondence Analysis (DCA) were used to construct the M matrices from the BC distance matrix. Ordinate vectors for MDS and projections for DCA were obtained using the *ordinate* function of the *phyloseq* package (McMurdie and Holmes, 2013). The procedures to build the M matrices from vectors and projections are described in (Saborío-Montero *et al.*, 2021).

Microbiability estimation. We fitted the following two models for each breed group to investigate the proportion of phenotypic variation explained by the gut microbiota composition, environment, and genetic background.

$$\text{Model_S_P_M: } y_{ijkl} = u + \text{sire}_i + \text{pen}_j + m_k + e_{ijkl} \quad (1)$$

$$\text{Model_M: } y_{kl} = u + m_k + e_{kl} \quad (2)$$

Where y is the trait measures, u was the overall intercept, sire_i was the random effects of i^{th} sire, pen_j is the random effect of j^{th} pen, m_k contains the random effects of gut microbiota for each individual with $m \sim N(0, M\sigma_m^2)$, where M is the microbial relationship matrix created by one of the eight methods described in the previous section and σ_m^2 is the microbial variance, e is the residual.

We fitted the Bayesian ridge regression models as implemented by *BGLR* package (Pérez and De Los Campos, 2014) in R. A Markov chain Monte Carlo algorithm was used to perform the analyses with 150,000 iterations, 50,000 iterations discarded as burn-in, and 10 iterations set as the thinning interval. The microbiability was calculated as the σ_m^2 over the total phenotypic variance.

Results

Pearson's correlation coefficients between the off-diagonal elements (beta-diversity) of the microbial similarity matrices are depicted in Figure 1. Overall, the patterns of correlations among matrices were similar across breeds. The highest correlations (r=0.99 to 1.00) were found between two distance-based matrices JA and BC in three breeds. Strong positive correlations (r=0.60 to 0.97) were found between LK and MDS, LK and AK1, BC and GK, JA and GK, and MDS and AK1 matrices. However, the beta-diversity among animals captured by JA and BC matrices was distinct from that captured by DCA, AK1, and PK, with negative to small correlations between the matrices.

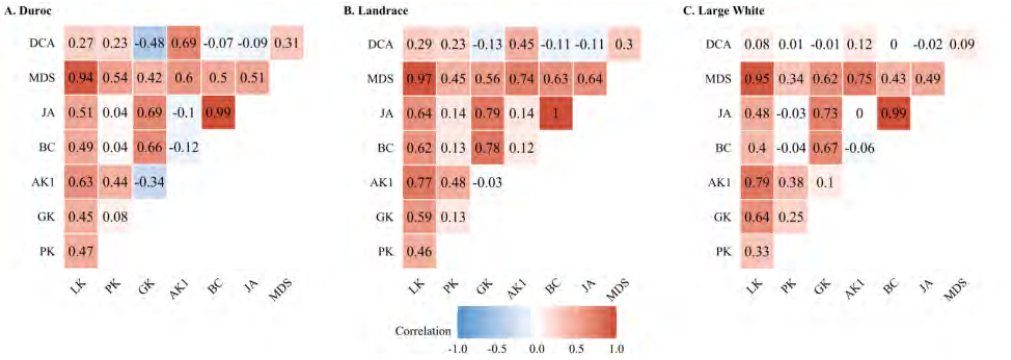


Figure 1. Pearson's correlation between off-diagonal elements of microbial similarity matrices in each breed. Correlations that are negative are shown in blue, while those that are positive are shown in red.

We estimated the variance components and calculated the proportion of host phenotypic variance due to pen, sire, or gut microbiota using the Model_S_P_M and Model_M (Figure 2). The pen and sire factors accounted for 6 to 46% and 6 to 20% of the phenotypic variance. The microbiability ranged from 0.03 to 0.32 and 0.02 to 0.52 for BW, from 0.02 to 0.29 and 0.02 to 0.64 for BF, from 0.02 to 0.18 and 0.08 to 0.69 for LD, and from 0.02 to 0.24 and 0.03 to 0.46 for IMF obtained by the Model_S_P_M and Model_M, respectively. In both models, the GK, JA, BR, and AK1 matrices had greater microbiability than the other matrices, while the DCA estimated the least microbiability. Moreover, the microbiability estimates were higher obtained by the Model_M than Model_S_P_M. There were also differences in the microbiability estimates among traits or breeds.

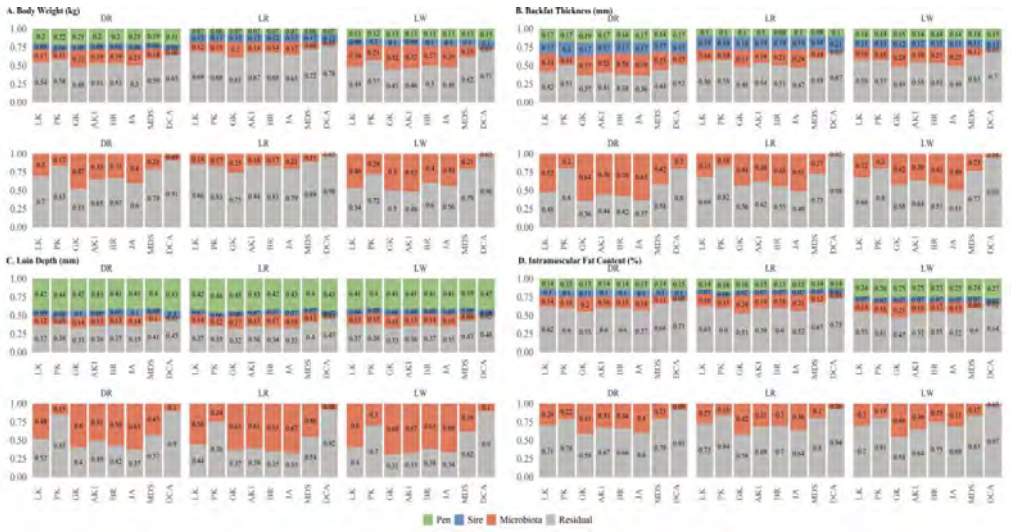


Figure 2. Proportions of variance component estimates for growth and carcass traits by breed. Plots for Model_S_P_M are in the odd row, while for Model_M are in the even row. The x axis represents eight matrices. The proportion value is indicated on the y axis.

Discussion

Our study used alternative approaches to measure resemblance in the gut microbiota composition among pigs and estimate microbiability in three western breeds. We observed differences in the beta-diversity estimated by the eight approaches by computing the correlation between the off-diagonal elements of paired matrices. The outcome was expected since different methods emphasize components of microbiome data in different ways when calculating the beta-diversity between pairs of samples (Tuomisto, 2010). Furthermore, we identified a close relationship between the JA and BC matrices in all three breeds, which agrees with the results reported by Saborío-Montero *et al.* (2021) based on the ruminal microbiota in dairy cattle.

We calculated the proportions of phenotypic variance explained by pen, sire, or gut microbiota in each breed population. The microbiability estimates varied depending on the matrix used, trait, and breed. Additionally, the total proportions of phenotypic variance due to the pen, sire, and microbiota in the Model_S_P_M and the microbiota in the Model_M were comparable in most cases. These findings suggest that the gut microbiota could be a valuable source of information for understanding performance traits in swine, especially in situations when systematic information is unavailable. However, more research is needed to validate our findings in larger populations.

References

Camarinha-Silva, A., Maushammer, M., Wellmann, R., Vital, M., Preuss, S. *et al.* (2017) Genetics. 206:1637–1644. https://doi.org/10.1534/genetics.117.200782

Coenen, A. R., Hu, S. K., Luo, E., Muratore, D., and Weitz, J. S. (2020) Front. Genet. 0:310. https://doi.org/10.3389/FGENE.2020.00310

Difford, G., Lassen, J., and Løvendahl, P. (2016) In: Book of Abstracts of the 67th Annual Meeting of the European Federation of Animal Science. p. 285.

Dixon, P. (2003) J. Veg. Sci. 14:927–930. https://doi.org/10.1111/j.1654-1103.2003.tb02228.x

He, Y., Tiezzi, F., Howard, J., Huang, Y., Gray, K. *et al.* (2022) BMC Microbiol. 22:1. https://doi.org/10.1186/s12866-021-02409-6

Khanal, P., Maltecca, C., Schwab, C., Fix, J., and Tiezzi, F. (2021) J. Anim. Breed. Genet. https://doi.org/10.1111/JBG.12504

Lu, D., Tiezzi, F., Schillebeeckx, C., McNulty, N. P., Schwab, C. *et al.* (2018) Microbiome. 6:4. https://doi.org/10.1186/s40168-017-0384-1

McMurdie, P. J., and Holmes, S. (2013) PLoS One. 8:e61217. https://doi.org/10.1371/JOURNAL.PONE.0061217

Montesinos-López, A., Montesinos-López, O. A., Montesinos-López, J. C., Flores-Cortes, C. A., de la Rosa, R. *et al.* (2021) Heredity (Edinb). 1–20. https://doi.org/10.1038/s41437-021-00412-1

Pérez, P., and De Los Campos, G. (2014) Genetics. 198:483–495. https://doi.org/10.1534/genetics.114.164442

Ross, E. M., Moate, P. J., Marett, L. C., Cocks, B. G., and Hayes, B. J. (2013) PLoS One. 8:e73056. https://doi.org/10.1371/JOURNAL.PONE.0073056

Saborío-Montero, A., Gutiérrez-Rivas, M., López-García, A., García-Rodríguez, A., Atxaerandio, R. *et al.* (2021) Livest. Sci. 250:104538. https://doi.org/10.1016/J.LIVSCI.2021.104538

Tuomisto, H. (2010) Ecography (Cop.). 33:23–45. https://doi.org/10.1111/J.1600-0587.2009.06148.X

502. Organoids as a research tool in animal breeding and nutrition

L.M.G. Verschuren[1*], D. Schokker[2], S. Verstringe[3], L. Kruijt[2], A.A.C. de Wit[2], E.G.T. van der Valk[2], S.K. Kar[2] and E.D. Ellen[2]

[1]Topigs Norsvin Research Center B.V., Schoenaker 6, 6641 SZ, Beuningen, the Netherlands; [2]Wageningen University & Research, P.O. Box 338, 6700 AH Wageningen, the Netherlands; [3]Nutrition Sciences N.V., Booiebos 5, 9031, Drongen, Belgium; lisanne.verschuren@topigsnorsvin.com

Abstract

In vitro miniaturized and simplified three-dimensional versions of an organ, commonly known as organoids, are an important innovative key technology in human and, more recently, animal research. So far, the *in vitro* organoid system has been applied on a small number of animals in a research setting. However, to apply the *in vitro* organoid system on a larger scale for industrial purposes, a high throughput system needs to be developed. The major aim of this research is to pave the way for a high throughput *in vitro* organoid system of the pig intestinal tract to be used in animal breeding, nutrition, and health. We developed high throughput tools to generate organoids from pig intestines on a large scale. We also looked at different applications addressing intestinal functionality, which allow to accurately phenotype animals for complex traits. Organoid research enables animal breeding companies to switch more quickly towards less animal testing.

Introduction

Developments in cell biology show that it is possible to grow and maintain three-dimensional (3D) organ-like structures in culture for a virtually indefinite time. These structures, called organoids, contain multiple cell types, resemble the *in vivo* tissue, and recapitulate key function(s) of the organ. It has been demonstrated that organoids can be dissociated and seeded onto Matrigel- or collagen-coated Transwell membranes to form two-dimensional (2D) epithelial monolayers (Foulke-Abel *et al.*, 2014; Van der Hee *et al.*, 2020). These 2D cultures allow the study of interactions with apically applied exogenous stimuli, such as pathogens or compounds, i.e. feed ingredients (Kar *et al.*, 2020), nutrients, drugs, and chemicals. In previous studies, the transport and reaction on different agents has been investigated using 2D monolayers (Van der Hee *et al.*, 2020). Hence, the 2D organoid structure could potentially be used to test differences in responses to various exogenous stimuli between organoids derived from different animals. Organoids can then be used as a phenotyping tool in animal breeding. So far, porcine intestinal organoids have been used to study host-pathogen interaction in relation to feed efficiency on a small number of pigs (Ellen *et al.*, 2018). However, to apply the *in vitro* organoid system on a larger scale for industrial purposes, like phenotyping of animals, a high throughput system needs to be developed. The major aim of this research was to pave the way for high throughput *in vitro* organoid systems of the pig intestinal tract to use as a key technology in animal breeding, nutrition, and health. To develop an *in vitro* high throughput organoid system, several steps within the organoids system need to be optimized: (1) collection of tissue samples; (2) production of organoids; (3) storing of the tissue and organoids; and (4) testing. Furthermore, the resemblance between the *in vitro* system and the *in vivo* system needs to be validated.

Materials & methods

Collection of tissue samples. Tissue samples have been collected from the small intestine (jejunum and ileum) and large intestine (colon). These locations have been selected because the jejunum and ileum are important for nutritional research, whereas the colon could be beneficial in future research to collect tissue samples without sacrificing the animals. Tissue samples have been collected from the slaughterhouse. So far, tissue samples have been collected from 26 three-way crossbred (Synthetic boar × (Landrace × Large

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_502, © L.M.G. Verschuren *et al.* 2022

White)) pigs raised at research farm 'Laverdonk' of Agrifirm Innovation Center (Heeswijk-Dinther, the Netherlands).

Establishing a cryo-stock for pig organoids. Adult stem cells (ASCs) from the collected tissue have been used to derive 3D organoids. In our study, intestinal crypts have been used. Under appropriate conditions using different growth factors and Matrigel, ASCs can grow, and their daughter cells can differentiate in multiple cell types, resulting in 3D structures (Kar *et al.*, 2021). We can produce large units of 'mother' organoids from a large number of animals to use as a phenotyping tool in animal breeding. We can also produce a large number of 'daughter' organoids to perform multiple tests on tissue from the same animal, for example for multiple phenotypes per pig or as a screening tool in animal nutrition. To validate the resemblance between the organoids and the respective tissue, gene expression measurements using RNA-Sequencing technology has been used as a read-out parameter. To trim low-quality data, we used Trim Galore version 0.6.6 (Krueger, 2018) and Cutadapt version 1.16 (Martin, 2016) with default settings except for length 35 (default 20) and stringency 6 (default 1). This resulted in paired-end reads if both reads were ≥35 bp. These trimmed reads were aligned against the pig reference genome (Ensembl *Sus scrofa* 11.1.103, (Howe *et al.*, 2021) using STAR version 2.7.3a with settings at default (Dobin *et al.*, 2013). Gene expression was quantified using RSEM version 1.3.0 (Li and Dewey, 2011) with default settings, except for the strand specific protocol, which was set to 0 to derive all upstream reads from the reverse strand. RSEM expected counts and Transcripts per Million (TPM) values were quantified. Principal component analyses have been performed to compare the tissue and organoids.

Optimization of the 2D monolayer Transwell system. Single cells, obtained after dissociation of 3D intestinal 'daughter' organoids, were seeded on Transwells to obtain 2D monolayers for further testing. We determined the optimal conditions to obtain stable and reproducible trans-epithelial electrical resistance (TER) values in 2D organoids, and to confirm the integrity and permeability of the monolayer. For this, the cellZscope (NanoAnalytics GmbH, Münster Germany) was used to measure the TER every half an hour to investigate the optimal Matrigel concentration, cell density, and timestamp.

Results

Gene expression from organoids act as a proxy for the respective tissue. We successfully produced organoids from different pigs and different locations. By measuring gene expression of both the tissue of interest and its derived organoids, we established if a relationship exists. Here, we targeted the colon as an example. When focusing on the moderate and highly expressed genes, i.e. transcripts per million above 100, we observed an overlap of 1,539 genes (75%) between tissue and organoids, 242 genes (12%) unique for tissue, and 276 genes (13%) unique for organoids. This shows that many genes are expressed in both the *in vivo* and *in vitro* situation. To assess the functionality of these genes we subsequently performed a Gene Ontology (GO) enrichment analysis for the genes in 'unique tissue', 'overlap', and 'unique organoids'. As we were interested in the biological processes that could be affected, we solely focused on the GO biological processes. Furthermore, we focused on the top 10 biological processes based on the false discovery rate (FDR<0.01, Table 1).

2D Transwell system. Figure 1 shows preliminary results of the TER measurements. So far, the optimal Matrigel concentration we observed is 1% with a cell density of 10^5 cells. This resulted in a window-of-opportunity to perform functional studies between approximately 80-90 hours after seeding, because the TER values were above 1000 Ωcm^2.

Table 1. Top 10 significantly enriched[1] Gene Ontology biological processes.

Ranking	'Unique tissue'	'Overlap'	'Unique organoid'
1	Immune response	Organonitrogen compound biosynthetic process	Small molecule metabolic process
2	Complement activation	Cellular amide metabolic process	Localization
3	Humoral immune response	Organonitrogen compound metabolic process	Macromolecule localization
4	RNA splicing	Peptide metabolic process	Organic substance transport
5	Iron ion transmembrane transport	Amide biosynthetic process	Actin filament organization
6	Immune system process	Translation	Cellular localization
7	RNA splicing, via transesterification reactions	Peptide biosynthetic process	Oxidation-reduction process
8	RNA splicing, via transesterification reactions with bulged adenosine as nucleophile	Generation of precursor metabolites and energy	Protein localization
9	MRNA splicing, via spliceosome	Oxidation-reduction process	Actin filament depolymerization
10	Tricarboxylic acid cycle	Protein metabolic process	Protein transport

[1] False discovery rate <0.01.

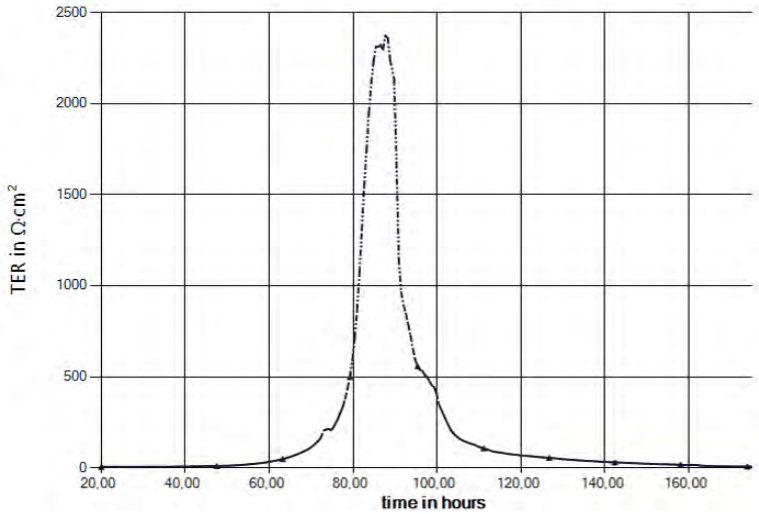


Figure 1. Trans-epithelial electrical resistance (TER) measurements with 1% Matrigel and 10^5 cells.

Discussion

Organoids have been identified as a key *in vitro* tool to reduce animal testing and to define new phenotypes for animal breeding. For pigs, especially for the intestinal tract, this *in vitro* system has been optimized. Results from the gene expression study of the colon indicate that organoids resemble the *in vivo* tissue to a large extend, which is in line with results for jejunum organoids (Van der Hee *et al.*, 2020). The main difference between tissue and organoids is the lack of immune cells in organoids, which our study showed by the genes from 'Unique tissue', indicating multiple processes involved in immunity. The GO terms of 'Unique organoid' genes showed more processes such as localization and actin organization, which was expected because the organoids grew in an artificial environment, i.e. Matrigel. The genes from 'Overlap' were by far the greatest number, i.e. 1,539 genes among the comparison. These 'Overlap' genes were more involved in generic biological processes, i.e. metabolic processes. As we aim to use the organoids as phenotyping and screening tool, similarity in metabolic processes between organoids and tissue is highly

important. Hence, the organoids are likely suitable for comparing responses of animals to exogenous stimuli on the organoids when this is related to metabolic processes, but not when the aim is to test the immune response.

Preliminary results of the 2D system show that a Matrigel concentration of 1% with 10^5 cells resulted in the most optimal TER values. For transport studies, TER values of >750 Ωcm^2 results in low variability between replicates (Van der Hee *et al.*, 2020). Future work will focus on extending the window-of-opportunity when the TER values are above a certain threshold, e.g. 1000 Ωcm^2. The latter will benefit the application for investigating differences between organoids derived from individual animals.

Currently, a high-throughput *in vitro* screening system is under development that will enable breeding and nutrition companies to test and phenotype pig intestinal organoids on a larger scale. Furthermore, simultaneously a biobank will be developed, including (intestinal) organoids and tissue from several pigs, for future new research questions.

References

Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, *et al.* (2013) Bioinformatics 29(1): 15-21. https://doi.org/10.1093/bioinformatics/bts635

Ellen, E. D., N. Taverne, J. J. Taverne-Thiele, O. Madsen, H. Woelders, *et al.* (2018) Proc. of the 69th EAAP, Dubrovnik, Croatia.

Foulke-Abel, J., J. In, O. Kovbasnjuk, N. C. Zachos, K. Ettayebi, S. E. Blutt, *et al.* (2014) Exp Biol Med (Maywood) 239: 1124-1134. https://doi.org/10.1177/1535370214529398

Howe, K. L., Achuthan, P., Allen, J., Allen, J., Alvarez-Jarreta, *et al.* (2021). Nucleic acids res 49(D1): D884-D891. https://doi.org/10.1093/nar/gkaa942

Kar, S. K., B. van der Hee, L. M. P. Loonen, N. Taverne, J. J. Taverne-Thiele, *et al.* (2020) J. Anim. Sci 11: 51. https://doi.org/10.1186/s40104-020-00443-4

Kar, S. K., J.M. Wells, E.D. Ellen, M.F.W. te Pas, O. Madsen, *et al.* (2021) Vet. Res. 52: 43. https://doi.org/10.1186/s13567-021-00904-2

Krueger, F. Trim Galore! Available at: http://www.bioinformatics.babraham.ac.uk/projects/trim_galore

Li, B. and Dewey, C. N. (2011) BMC bioinform 12(1): 1-16. https://doi.org/10.1186/1471-2105-12-323

Martin, M. (2011). EMBnet. journal 17: 10-12. https://doi.org/10.14806/ej.17.1.200

Van der Hee, B., O. Madsen, J. Vervoort, H. Smidt, and J. M. Wells. (2020) Front. Cell Dev. Biol. 8. https://doi.org/10.3389/fcell.2020.00375

503. Simulation of host-microbiome evolution throughout a divergent selection experiment

C. Casto-Rebollo[1*], I. Pocrnic[2], G. Gorjanc[2] and N. Ibáñez-Escriche[1]

[1]Institute for Animal Science and Technology, Universitat Politècnica de València, C/Camino de Vera s/n, 46022 València, Spain; [2]The Roslin Institute, The University of Edinburgh, Easter Bush Campus, Edinburgh, EH25 9RG, United Kingdom; cricasre@posgrado.upv.es

Abstract

The microbiome composition influences the host response to selection and shapes complex phenotypes. It is a multifactorial complex trait in which the microbial inheritance, the host-genome, and the own microbial interactions influence its variability. The expensive sequence-based techniques limit the availability of empirical data. Thus, other approaches are necessary to evaluate strategies for microbiome studies. This study aimed to develop a simulator following a long-term phenotype selection experiment in rabbits, considering both genome and microbiome inheritance. The result showed genetic, microbiome, and phenotypic trends across 13 generations. Differences in the microbiome contribution to the phenotype were shown only in the scenario with a high host genetic effect of the microbiome. This is a preliminary work and further improvements to the simulator will be shared with the research community.

Introduction

Each year, new research underpins the importance of the effect of microbiota on animals' phenotypes suggesting that this contribution is on par with the host genome (Camarinha-Silva *et al.*, 2017; Difford *et al.*, 2018). However, how the microbiome composition shapes complex phenotypes and mediates the host response to selection remains unclear (Henry *et al.*, 2021).

Animal's microbiome composition is established by both the transmission mode (microbial inheritance) and the host genome (microbial heritability). Microbial inheritance could be vertical or horizontal (Koskella *et al.*, 2017), depending on whether and in what quantity the microbiota in offspring came from their parents (Parental microbiota; PM) or their environment (Environmental microbiota; EM). Thus, the transmission mode plays a crucial role in the individual pool of available microbes (Dominguez-Bello *et al.*, 2010). On the other hand, microbial heritability could influence the dynamic nature of the microbiome in the host, so microbial composition can be considered as another quantitative trait of an animal (Pérez-Enciso *et al.*, 2021). Furthermore, other factors such as the microbial genome and the microbial interactions also shape the microbiome composition, which is clearly a multifactorial complex trait.

The expensive sequence-based techniques limit the availability of empirical data to study microbiome composition at scale. Moreover, the lack of consensus on theoretical models and analytic methodologies constrains comparisons between experiments and learning how various microbiome-related factors impact animal's phenotypes. Here we take a simulation approach to gain understanding of the key principles in studying the effect of the microbiome on phenotypes and selection response. To this end we have developed a simulator following a long-term phenotype selection experiment in rabbits (Blasco *et al.*, 2017), considering both genome and microbiome inheritance. The overall aim is to develop a tool for testing different data collection strategies and statistical models to further the study of microbiome in animal breeding.

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_503, © C. Casto-Rebollo *et al.* 2022

Materials & methods

The developed simulator allows users to simulate a host genome, microbiome composition, and animal's phenotype as a function of genome, microbiome, and environment. In this study, we used this simulator to mimic a phenotypic divergent selection experiment modulated by the host-microbiome (Blasco *et al.*, 2017). Moreover, three scenarios were implemented to study microbial heritability and its effect on the phenotype. The simulation was repeated 10 times.

Genome simulation. The simulation of 1000 diploid genomes in the base population was performed with the AlphaSimR R package (Gaynor *et al.*, 2021), using the genome and demographic history of rabbits (*Oryctolagus cuniculus*). We simulated 21 chromosomes with a length of 1.28^{10} base pairs, 10,000 segregating sites, and 100 Quantitative Trait Loci (QTL) per chromosome. The allele substitution effects of loci QTLs (α) were sampled from gamma distribution, $\alpha \sim \Gamma(k=0.2, \theta=1)$. In the base population, the genetic values (gv) were distributed with mean 0 and genetic variance of 0.81 with a targeted heritability of 0.13 (for litter size from Blasco *et al.* (2017)).

Microbiome simulation. We simulated a parental (PM) and an environmental microbiome (EM) to model both aspects of microbiome inheritance. The species abundances (x_k) in PM and EM were sampled from a negative binomial distribution, $x_k \sim NB(r=2, q=0.0001)$. The total number of species was 1000 in EM and 600 in PM (with overlap between them). Both, PM, and EM, were simulated with a total species abundance of 10^8. The average of species abundance (μ_k) varied depending on the combination of the species in the individuals due to the microbiome inheritance. Hence, a simulation of 100,000 microbiome profiles was done to obtain μ_k in the founder population, considering that the microbiome inheritance was 100% of the microbial species from the PM (assuming a vertical transmission from the dam due to the delivery mode) and 15% from the EM. The stability of microbial species was computed as a coefficient of variation (CV), sampled from a uniform distribution U(0.01, 2). This CV allowed us to compute the species variance ($\sigma^2_{m_k}$) and simulate the microbiome for each animal in the base population (PM_i), sampling the abundance of each microbial species k from a normal distribution $N(\mu_k, \sigma^2_{m_k})$. A total of 35 microbial species were assigned an effect on the phenotype, following a gamma distribution, $\omega_k \sim \Gamma(k=0.2, \theta=5)$ (Pérez-Enciso *et al.*, 2021). The individual's microbiome value (mv) in the base population was distributed with mean 0 and microbiome variance (σ^2_m) of 0.81, to match the genetic variance (Pérez-Enciso *et al.*, 2021).

Phenotype simulation. The phenotype was simulated as:

$$y_i = \mu_P + \sum_{j=1}^{n} z_{ij}\alpha_j + \sum_{k=1}^{m} x_{ik}\omega_k + e_i \quad (1)$$

where y_i is the phenotype of individual i; μ_P is the phenotypic mean in the base population; $\sum_{j=1}^{n} z_{ij}\alpha_j$ is the gv_i of individual i, where z_{ij} is the genotype of individual I for SNP j and α_j is the allele substitution effect of the SNP j; $\sum_{k=1}^{m} x_{ik}\omega_k$ is the mv_i of individual i, where x_{ik} is the abundance of species k in individual i (Equation 2) and ω_k is the effect of species k on the host phenotype; and e_i is the residual of individual i distributed as $N(0, \sigma^2_e)$, were σ^2_e is the residual variance.

Inheritance. A divergent selection based on the female's phenotypes was simulated according to Blasco *et al.* (2017). In the base population, 125 individuals with the highest phenotype and 125 with the lowest phenotype were selected to initiate divergent populations. Breeding males were full-sibs of the best 25 breeding females from each divergent population. Each breeding male was mated with five breeding females, avoiding a close relationship among them. Thirteen discrete generations of selection were simulated simultaneously for the two divergent populations. In each generation, 125 breeding females and 25 breeding males were selected in each population. The breeding females were selected as in the

base population, and the breeding males were males from the offspring of the best female for each sire (best mating). Mating and genome inheritance were simulated using the AlphaSimR package (Gaynor *et al.*, 2021). Microbiome inheritance was 100% from the PM and 15% from the EM. The transmission EM was the same for all animals and the two divergent populations, but different between generations (EM_g). Hence, the microbiome composition in the offspring was computed using Equation 2.

Scenarios considered. Three scenarios were simulated to study how microbial heritability affected response to selection. The first scenario assumed the absence of microbial heritability (NMH), so there was no host genome effect on its microbiome. In the second and third scenarios, microbial heritability was included in the simulation. The second scenario assumed intermediate microbial heritability (IMH) from 0.25 to 0.50 and the third scenario assumed high microbial heritability (HMH) from 0.5 to 0.9. Consequently, the species abundance was considered as another quantitative trait simulated following Pérez-Enciso *et al.* (2021):

$$x_{ik} = \mu_k + \sum_{j=1}^{n} z_{ij}\beta_{jk} + e_{ik} \quad (2)$$

where x_{ik} is the abundance of the specie k in the individual i; μ_k is the average abundance of the specie k in the base population; $\sum_{j=1}^{n} z_{ij}\beta_{jk}$ is the genetic value (gv_{ik}) of individual I for the abundance of specie k, z_{ij} is the genotype of SNP j for individual i, β_{jk} being the substitution allele effect of SNP j for the species k; and e_{ik} is the residual of individual i for the species k distributed as $N(0, \sigma^2_{e_k})$, where $\sigma^2_{e_k}$ is residual variance for the species k. The gv_{ik} were distributed with mean 0 and microbial genetic variance ($\sigma^2_{m,g}$) according to each microbial heritability. We assumed that 10% of the total number of species (100) were under the genetic control of 10% of the QTLs (210). From these 100 species, 18 of them influence the phenotype (50% of total species influencing the phenotype).

Results & discussion

The simulation showed expected divergent trends throughout the 13 generations of selection (Figure 1). Although there were 35 microbial species with effect on host phenotype, and 18 of them under host genetic control, no differences were observed in phenotype trends between the scenarios NMH and IMH (Figure 1C). Differences in the phenotype trends were only shown with the scenario HMH (Figure 1C), that is when microbial heritability was high, from 0.5 to 0.9. In this scenario, microbiome value trend diverged more than with the NMH and IMH scenarios (Figure 1B). These preliminary results suggested that, under the assumed simulation, a high microbial heritability is necessary to observe a significant contribution of microbiome to response to selection. Likewise, the microbiome composition is multifactorial, and it has a huge variability within and between individuals. Another scenario considering only the effect of the host genome on the phenotype is necessary to fully assess the degree of microbiome contribution on the phenotype.

Conclusions

In this study, a simulator of the modulation of a host-microbiome evolution has been developed. The result showed genetic, microbiome, and phenotypic trends across 13 generations. Differences in the microbiome effect on the host phenotype were shown only in the scenario with high microbial heritability. This could be due to the reduction of the impact of the environmental microbiome. This is a preliminary work, and further improvements to the simulator will be shared with the research community.

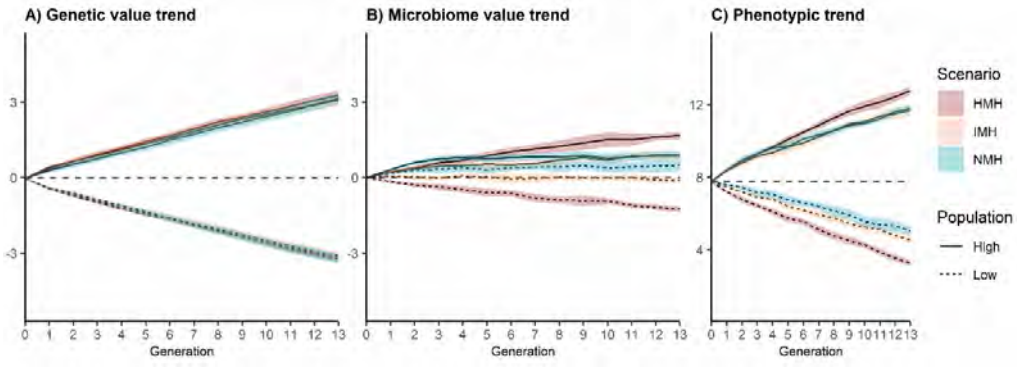


Figure 1. Genetic, microbiome, and phenotypic trends throughout a divergent selection experiment. Averages and 95% quantiles over replicates are shown for (A) genetic value, (B) microbiome value, and (C) phenotype value. HMH: High microbial heritability, IMH: Intermediate microbial heritability, NMH: No microbial heritability. High (dashed line) and Low are the two divergent populations.

References

Blasco A., Martínez-Álvaro M., García ML., Ibáñez-Escriche N., and Argente A. (2017) Genet Sel Evol 49, 48. https://doi.org/10.1186/s12711-017-0323-4

Camarinha-Silva A., Maushammer M., Wellmann R., Vital M., Preuss S., *et al.* (2017) Genetics 206:1637–1644. https://doi.org/10.1534/genetics.117.200782

Difford G. F., Plichta D.R., Løvendahl P., Lassen J., Noel S.J., *et al.* (2018) PLoS Genet 14(10):e1007580. https://doi.org/10.1371/journal.pgen.1007580

Dominguez-Bello M.G., Costello E.K., Contreras M., Magris M., Hidalgo G., *et al.* (2010) Proc Natl Acad Sci U S A 29;107(26):11971-5. https://doi.org/10.1073/pnas.1002601107.

Gaynor R.C., Gorjanc G., and Hickey J.M. (2021) G3 (Bethesda) 9;11(2):jkaa017. https://doi.org/10.1093/g3journal/jkaa017.

Henry L.P., Bruijning M., Forsberg S.K.G., and Ayroles J.F. (2021) Nat Commun 12, 5141 (2021). https://doi.org/10.1038/s41467-021-25315-x

Koskella B., Hall L.J., and Metcalf C.J.E. (2017) Nat Ecol Evol 1:1606–1615. https://doi.org/10.1038/s41559-017-0340-2

504. Rumen bacteria do not provide improved genetic evaluation of dairy traits in sheep

G. Martinez Boggio[1*], O.F. Christensen[2], A. Legarra[1], C. Allain[3], A. Meynadier[1] and C. Marie-Etancelin[1]

[1]GenPhySE, Universite de Toulouse, INRAE, ENVT, 24 Chemin de Borde Rouge, 31326 Castanet Tolosan, France; [2]Center for Quantitative Genetics and Genomics, Aarhus University, Blichers Alle 20, 8830 Tjele, Denmark; [3]INRAE, Experimental Unit of La Fage, 12250 Saint-Jean et Saint-Paul, France; guillermo.martinez-boggio@inrae.fr

Abstract

It is now widely accepted that animals' phenotypes are determined by the host genetic and the microbiome composition. For the genetic evaluation purpose, our objectives are to estimate the heritability and microbiability, and the additive metagenomic effect, assuming that these two components have additive effects. In this study were included, as dairy traits, milk production, fat and protein content; and 16S rRNA gene of rumen bacteria of 795 ewes. Two genomic models were used, a first model to predict individual microbiome contribution to phenotypes; and a second model to predict genetic effects on this predicted microbiome contribution. Results showed that the microbiabilities for milk traits were close to zero, and the residual heritabilities were not different from the trait heritabilities. By disentangling the milk traits' variance explained by the microbiome, we conclude that rumen bacteria abundances do not provide improved genetic evaluation of dairy traits in Lacaune sheep.

Introduction

The animal phenotypes are determined by the genetic component, the environmental conditions, and also by the digestive microbiome which is partially controlled by the host (Pérez-Enciso *et al.*, 2021). To disentangle the holobiont impact on phenotypes, Christensen *et al.* (2021) and Weishaar *et al.* (2020) proposed models to account for intermediate -OMICs data (i.e. metabolomics, transcriptomics, and metagenomics) in genetic evaluation. In polygastrics, the rumen microbiome is crucial, as it allows ruminants to digest plant fibre. Bacteria are able to degrade and ferment fibrous feedstuffs into nutrients, directly usable by the animal for maintenance, growth and lactation. Considering the relevance of rumen fermentation products in the mammary gland, and having demonstrated that rumen bacteria are controlled by host genome, we propose to apply the Christensen *et al.* (2021) methodology on a Lacaune dairy sheep dataset including metabarcoding (16S rRNA gene) as intermediate -OMIC trait, and dairy traits as phenotypes. The objectives are to estimate variance components, as heritability (h^2) and microbiability (m^2), and the additive part of the predicted metagenomic effect.

Materials & methods

Animals phenotyping. Data from 795 multiparous Lacaune dairy ewes were obtained from INRAE Experimental Unit of La Fage (UE 321 agreement A312031, Roquefort, France) between 2015 and 2019. All ewes (weighing 77 kg on average) raised indoors, and fed 93% meadow hay and silage plus 7% of concentrates diet (on dry matter basis). Daily milk production (MP), milk fat and protein contents (FC and PC, respectively) as weighted averages, were recorded as part of one official milk recording of the flock. In brief, two milk samples per animal for morning and afternoon milking were sent for analysis at the Interprofessional Milk Analysis Laboratory (Aurillac, France). Milk FC and PC were recorded for 777 ewes, and analysed with mid-infrared techniques with a Milko-Scan™ FT6000 instrument (Nanterre, France). Rumen sampling was performed within 3 days on morning or afternoon around the official milk recording of the flock. Ruminal content was sampled from each ewe using a vacuum pump and a medical gastric tube.

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_504, © G. Martinez Boggio *et al.* 2022

The protocol received approval from the French Ministry of Higher Education, Research and Innovation – Animal Ethics Committee (approval number: APAFIS#6292-2016080214271984 v8). After extraction, the DNA strands of the 795 samples were sequenced using v3-v4 region of 16S rRNA gene and processed with the FROGS 3.0 pipeline (Escudié *et al.*, 2018), removing operational taxonomic units (OTUs) with average abundances below 0.005%. As result, we obtained an abundance table with 2,059 OTUs.

Animals genotyping. DNA extraction from blood samples and genotyping were performed for 795 Lacaune ewes. From those ewes, 743 were genotyped using a medium-density SNP chip (Illumina Ovine SNP50 BeadChip, 54,241 SNPs), 314 at Laboratoire d'Analyses Génétiques pour les Espèces Animales (Jouy-en-Josas, France) and 429 at Aveyron-Labo (Rodez, France). The remaining 52 ewes were genotyped with a low-density SNP chip (Illumina Ovine SNP15, 16,681 SNPs) at Neogen (Lansing, USA) and imputed to a medium-density SNP chip genotypes in the framework of the Lacaune dairy sheep genomic selection program (Larroque *et al.*, 2017). SNPs were removed using a call rate <0.99 for SNPs, and <0.95 for individuals, and minor allele frequency <0.05. In total, 35,825 SNPs remained for the analyses.

Statistical analyses. The methodology used was that proposed by Christensen *et al.* (2021).

1. The first model for predicting individual contributions of metagenomics to phenotypes:

$$y = Wb + m + a_r + e \quad (1)$$

where; y is the vector of observations for the dairy traits; b is the vector of fixed effects; m is the vector of random rumen bacteria effect (metagenomic effect); a_r is the vector of random additive effects; and e is the vector of random residual effects. W is the incidence matrix for b.

2. Second model used estimates of m to predict the additive part of the predicted metagenomic effect:

$$\hat{m} = W\beta + g + e \quad (2)$$

where $\hat{m}$ is the predicted effects m in Equation 1; β is the vector of fixed effects; g is the vector of random additive effects; and ε is the vector of random residual effects. W is the incidence matrix for β. The same fixed effects (P<0.05) were included in both models: sampling year, days in milk (DIM), and litter size. The distributional assumptions are $m \sim N(0, M\sigma^2_m)$, $a_r \sim N(0, G\sigma^2_{a,r})$, $e \sim N(0, I\sigma^2_e)$, $g \sim N(0, G\sigma^2_g)$, $\varepsilon \sim N(0, I\sigma^2_\varepsilon)$, where M is the microbial similarity matrix, G is a genomic relationships matrix (VanRaden, 2008), I is an identity matrix; σ^2_m is the metagenomic effect variance, σ^2_a and σ^2_g are additive variances, and, σ^2_e and σ^2_ε are residual variances. The microbiability was estimated as $m^2 = \sigma^2_m/\sigma^2_y$, the residual heritability as $h^2_r = \sigma^2_{a,r}/\sigma^2_y$, and the heritability of metagenomic effect as $h^2_m = \sigma^2_g/(\sigma^2_g + \sigma^2_\varepsilon)$. These variance components were estimated using BLUPF90test (Misztal *et al.*, 2002) with OPTION method VCE. Thus, following the formula $h^2 = m^2h^2_m + h^2_r$ presented by Christensen *et al.* (2021), the heritability of the phenotype is decomposed into a metagenomic mediated heritability ($m^2h^2_m$), and residual heritability (h^2_r).

3. M matrix construction

The matrix was computed as a variance-covariance matrix from rumen bacteria abundances as follows:

$$M = XX'/n \quad (3)$$

where; X is the abundance matrix with n=2,059 OTUs with zeros corrected with the geometric Bayesian-multiplicative (GBM) method (Martín-Fernández *et al.*, 2015), and centred log-ratio (CLR) transformed, for the 795 ewes. OTU abundances were pre-corrected for the significant fixed effects (for more than 10% of OTUs), such as lactation number, sequencing run, and sampling time and order, all of them nested on sampling year, and total number of sequences per rumen sample. Then, the similarity among animals in rumen bacteria composition was estimated by constructing a microbial similarity matrix M as in Ross *et al.* (2013).

Results

The decomposition of the parameter estimation using the two proposed models is shown in Table 1, where we see that the microbiability is only different from zero for milk protein content. Although, the heritability of metagenomic effect is different from zero for MP and PC, but only the heritability of PC is slightly higher than the estimated residual heritability. For MP and FC, we obtained a null microbiability leading to heritability of phenotypes equal to residuals ones.

Table 1. Estimates of residual heritability (h^2_r), microbiability (m^2), heritability of metagenomic effect (h^2_m), and heritability of phenotype (h^2).[1]

Trait	h^2_r (SE)	m^2 (SE)	h^2_m (SE)	h^2
Milk production (ml)	0.250 (0.068)	0.00	0.083 (0.058)	0.250
Fat content (g/ml)	0.533 (0.066)	0.00	0.00	0.533
Protein content (g/ml)	0.515 (0.070)	0.040 (0.037)	0.046 (0.052)	0.517

[1] SE = standard error.

Discussion

The heritability estimates obtained were similar to Rupp *et al.* (2003) with 0.28 to 0.32 for annual MP, and 0.41 and 0.51 for annual FC and PC, using all Lacaune dairy ewes (up to 90,000 animals). This is the first estimate of the rumen microbiome contribution to the variability of milk traits, using the largest sample size from a single population of dairy ruminant. The variances explained by rumen bacteria (m^2) were close to zero for most dairy traits, and the genetic variances (h^2_m) were lower than 9%, which is consistent with the weak phenotypic correlations between milk traits and rumen bacteria demonstrated by Martinez Boggio *et al.* (2021). The higher m^2 of PC could be explained by the key role of rumen bacteria in the production of essential amino acids needed for milk proteins (Xue *et al.*, 2019). Our estimates were similar to those presented for dairy cows by Buitenhuis *et al.* (2019) for milk fat (0.08±0.07) and protein (0.08±0.09) contents, and Zhang *et al.* (2020) for methane emissions (0.07). Despite this, it is interesting that with the method proposed by Christensen *et al.* (2021) we can decompose the heritability, and identify which part of the genetic variance was explained by the rumen microbiome, and which was directly explained by the host genetic (residual heritability). In our study, as shown by the low microbiabilities obtained, most of the variance is not explained by the rumen microbiome, even if the M matrix is constructed by omitting OTUs with zero, or considering those present in at least 50% of the animals. But we could further study other traits such as milk fatty acids and proteins, for which preliminary results showed positive microbiabilities. So far, we can conclude that with our dataset, and the low genetic variance explained by metagenomic data, the rumen bacteria abundances do not provide improved genetic evaluation for dairy traits in sheep.

References

Buitenhuis B., Lassen J., Noel S.J., Plichta D.R., Sørensen P., *et al.* (2019). Genet Sel Evol 51:23. https://doi.org/10.1186/s12711-019-0464-8

Christensen O.F., Börner V., Varona L., Legarra A. (2021). Genetics 219(2):iyab130. https://doi.org/10.1093/genetics/iyab130

Escudié F., Auer L., Bernard M., Mariadassou M., Cauquil L., *et al.* (2018). Bioinformatics 34:1287–1294. https://doi.org/10.1093/bioinformatics/btx791

Larroque H., Chassier M., Saintilan R., Astruc J.M. (2017). Proc. of EAAP, Tallinn, Estonia.

Martinez Boggio G., Meynadier A., Daunis-i-Estadella P., Marie-Etancelin C. (2021). PLoS ONE 16:e0254874. https://doi.org/10.1371/journal.pone.0254874

Martín-Fernández J.-A., Hron K., Templ M., Filzmoser P., Palarea-Albaladejo J. (2015). Statistical Modelling 15:134–158. https://doi.org/10.1177/1471082X14535524

Misztal I., Tsuruta S., Strabel T., Auvray B., Druet T., *et al.* (2002). Proc. of 7th WCGALP, Montpellier, France.

Pérez-Enciso M., Zingaretti L.M., Ramayo-Caldas Y., de los Campos G. (2021). Genet Sel Evol 53:65. https://doi.org/10.1186/s12711-021-00658-7

Ross E.M., Moate P.J., Marett L., Cocks B.G., Hayes B.J. (2013). J. Dairy Sci. 96:6030–6046. https://doi.org/10.3168/jds.2013-6766

Rupp R., Lagriffoul G., Astruc J.M., Barillet F. (2003). J. Dairy Sci. 86:1476–1481. https://doi.org/10.3168/jds.S0022-0302(03)73732-1

VanRaden P.M. (2008). J. Dairy Sci. 91:4414–4423. https://doi.org/10.3168/jds.2007-0980

Weishaar R., Wellmann R., Camarinha-Silva A., Rodehutscord M., Bennewitz J. (2020). J Anim Breed Genet 137:14–22. https://doi.org/10.1111/jbg.12447

Xue M.Y., Sun H.Z., Wu X.H., Guan L.L., Liu J.X. (2019). J. Dairy Sci. 102:5031–5041. https://doi.org/10.3168/jds.2018-15974

Zhang Q., Difford G., Sahana G., Løvendahl P., Lassen J., *et al.* (2020). ISME J 14:2019–2033. https://doi.org/10.1038/s41396-020-0663-x

505. Phenotypic characterization of organoids derived from pig intestine segments

F. Blanc*, M. Mongellaz, F. Pepke, C. Bevilacqua, J. Riviere, M. Vilotte, E. Giuffra* and G. Egidy

Université Paris-Saclay, INRAE, AgroParisTech, GABI, 78350 Jouy-en-Josas, France; fany.blanc@inrae.fr; elisabetta.giuffra@inrae.fr

Abstract

Organoids ('mini-organs') derived from adult stem cells residing in tissues can provide near-physiological *in vitro* 'Reduction, Refinement, and Replacement (3Rs)' models for genotype-to-phenotype research in animal sciences. Our aim was to assess the effective ability and reliability of pig intestinal organoids to reflect the phenotypes of the specific gut portion they are derived from. To this issue we generated organoids from the small intestine (duodenum, jejunum, ileum) and colon from cryopreserved biopsies of four Large White pigs after adapting the protocol to commercial slaughter conditions and further standardization of lab culture and phenotyping procedures. Here we report the comparison of organoids cultured in three and two dimensions (3D and 2D) vs the original animal tissue in terms of morphology, number and growth rate, cell composition and barrier tissue integrity, and highlight the main organoid vs tissue specificities.

Introduction

Organoids ('mini-organs') can be obtained from embryonic cells, induced pluripotent stem cells and adult stem cells (from tissue biopsies of almost any organ and animal species). They are easily biobanked and amenable to many biochemical and genetic perturbations (Clevers, 2016). Gut organoids derived from intestinal crypts retain some of the key morphological and molecular features of the specific gut segment they derive from and are increasingly used in several farmed species, including pigs (Li *et al.*, 2019; Van der Hee *et al.*, 2020).

Organoids are regarded as promising, near-physiological *in vitro* '3Rs' models for genotype-to-phenotype research, e.g. to study in highly reproducible conditions how genetic variation may affect key cellular processes underlying complex traits relevant for sustainable productions (Clark *et al.*, 2020). However, little is known so far on the relative influence of experimental and biological factors on organoid phenotypes. Specific culture conditions, passaging and culturing time can change the cellular state within the organoids. Recent analyses of transcriptional variance in organoids derived from human biopsies pointed to the heterogeneity of cohorts, organoids production and growth (Criss *et al.*, 2021; Gehling *et al.*, 2021, Mohammadi *et al.*, 2021).

Our global aim is to assess to what extent porcine gut organoids derived from different segments of the small intestine (duodenum, jejunum, ileum) and colon recapitulate the phenotypic variability of the tissue they are derived from, with a focus on phenotypes involving gut innate immunity responses. We report here the morphology, number and growth rate, cell composition and barrier tissue integrity phenotypes of these multiple organoids in comparison to the original tissue as a result of their production and culture in our laboratory standardized conditions. These results paved the basis of a transcriptome study in progress.

Materials & methods

Generation and culture of pig 3D and 2D organoids from frozen tissues. Pig intestinal organoids were generated from cryopreserved intestinal biopsies of four 5-month-old biobanked Large White pigs as described (Beaumont *et al.*, 2021). Cell number was adjusted at each passage (10,000 cells seeded in 50 µl droplets with 6-8 mg/ml of Matrigel (Corning, ref 356231) in 24-well plates) with basal culture medium

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_505, © F. Blanc *et al.* 2022

(BCM) (Van der Hee *et al.*, 2018). Conditioned media were from the same batches and their optimal concentrations were experimentally tested. To obtain 2D layers, organoids were processed to obtain single cell suspensions, 30,000 cells were seeded on Transwells with 0.4 µm pore polycarbonate membrane inserts (Corning) or 96-well plates (Corning) pre-coated with 0.5% (v/v) Matrigel, and cultured in BCM supplemented with 10 µM Y27632 at first. BCM was then refreshed every 2 or 3 days until cells reached around 75% of confluence. Medium was then replaced by differentiation medium (BCM without Wnt3A conditioned medium) for 3 days. The Transepithelial electrical resistance (TEER) was assessed using the Epithelial Volt/Ohm Meter EVOM2 (WPI) on Transwells at confluence.

Evaluation of 3D & 2D organoid growth rate and count. Phase contrast microscopy images of developing 3D organoids at 5, 6 and 7 days of culture were obtained with a PlanApoN 2× objective. Organoids area and count were estimated by image analysis using Fiji (particles >2,000 µm^2 and circularity between 0.2 and 0.9 were kept for the analysis). Confluence of 2D layers was also estimated by image analysis with Fiji from daily-acquired phase contrast microscopy images.

Histology and immunohistochemical staining. Tissues and organoids in 3D and 2D-layers (insert membranes) were fixed in paraformaldehyde 4% for 1h and paraffin-embedded; 5 µm-sections were stained by hematoxylin-eosin-saffron (HES) or alcian blue/periodic acid Schiff, or with antibodies to detect chromogranin A (Novus biological, ref NBP2-34240), villin-1 (Novus biological, ref NBP1-85335), β-catenin (Novus biological, ref NBP2-34240), and occludin (ThermoFisher, ref 71-500). Detection was then performed with goat anti-rabbit IgG-AF647 (ThermoFisher, ref A-21245) and goat anti-mouse IgG-AF555 (ThermoFisher, ref A-21422) and DAPI as nuclear counterstaining. Images acquired with a digital slide scanner (Pannoramic SCAN, 3DHISTECH), objective 20× were analyzed using CaseViewer and QuantCenter softwares (3DHISTECH).

Statistical analyses. Data analyses were performed using R (v4.1.0). Principal Component Analyses (PCA) were performed using FactoMineR (v2.4) package using the imputePCA function of missMDA (v1.18) package to handle missing values, and visualized with the factoextra R package (v1.0.7).

Results & discussion

Optimized protocols for porcine tissue collection & 3D/2D gut organoid culture. We have optimized existing protocols to directly cryopreserve human intestinal tissues allowing to generate organoids after thawing (Tsai *et al.*, 2018) for the four gut regions (duodenum, jejunum, ileum and colon) from 4 Large White pigs at slaughter age. This method allows sampling and storing of a large number of animal samples at once, which would not be compatible with immediate labor-intensive organoid culture. In addition, this procedure has guaranteed organoid culture generation free from bacterial and fungal contaminations. We then progressed in the establishment of standardized protocols to grow organoids optimizing reproducibility. We found that mechanical and enzymatic digestions performed at each passage reduced the heterogeneity in organoid stages within the same culture. We also set up a fixed number of cells to seed at each passage and a weekly passaging frequency. For 2D gut organoid cultures, we have standardized the number of cells seeded in each well and evidenced a variability concerning the time needed to reach 75% of confluence (ranging from 5 to 9 days).

In those conditions, area of organoids doubled (mean × 2.1) from 5 to 7 days while their circularity decreased (means 0.679 after 5 days vs 0.618 after 7 days). Differences between the original gut segment organoids were derived from, could be revealed in terms of morphology (colon organoids were bigger and duodenum ones less circular) and density in cultures (higher for duodenum organoids and lower for ileum ones). We also observed a variability of organoid density within pigs. Morphology and structure of

organoids in both 3D and 2D resembled that of gut tissues. Labelling of goblet cells, mature enterocytes and neuroendocrine cells was in accordance with the original tissues (Figure 1).

Pig gut organoids retain original tissue characteristics. Gut tissues clearly clustered by the corresponding gut regions in a PCA analysis performed with the different histological and immunohistochemical phenotypes measured (Figure 2A). The four gut regions spread along the first dimension in the gut tract order, clearly opposing duodenum and colon. A thinner layer ($P<0.01$) was observed in duodenum and a higher cell density in jejunum ($P<0.05$). Villin-1 expression intensity was lower in colon ($P<0.05$) and neuroendocrine cell density was higher in duodenum ($P<0.001$). Goblet cell density increased along the digestive tract (lower for duodenum ($P<0.05$) and higher in colon ($P<0.01$)), while the proportion of acid vs neutral mucins decreased (higher in duodenum ($P<0.001$) and lower in colon ($P<0.05$)). These observations are in accordance with published data on these pig tissues (Święch *et al.*, 2019).

In our culture conditions, ileum organoids appeared different from all other gut-derived organoids (Figure 2B) in showing higher cell density ($P<0.01$) and higher goblet cell and neuroendocrine cell density ($P<0.001$ and $P<0.05$, respectively). Jejunum organoids showed also lower goblet cell density ($P<0.05$). Goblet cells were found at high levels in ileum organoids compared to jejunum and duodenum organoids. In colon organoids, goblet cells were also found at high levels although at a lower level than in ileum. Interestingly, in the latter, mucins were more acid than in colon organoids, as observed in the original tissues.

We are currently gaining insights into this comparison using transcriptomics to decipher to what extent pig organoids have retained transcript expression and functional pathways of the epithelia of the different gut segments.

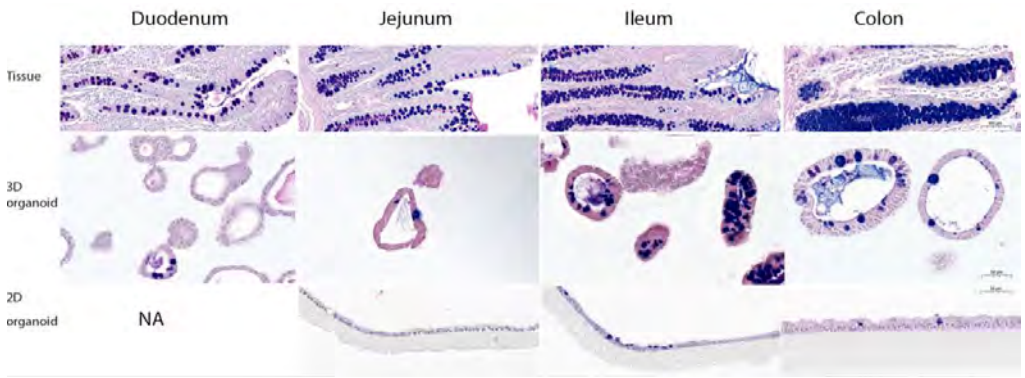


Figure 1. Representative alcian blue/periodic acid Schiff staining of gut tissues, 3D and 2D organoids for duodenum, jejunum, ileum and colon. NA: not yet available.

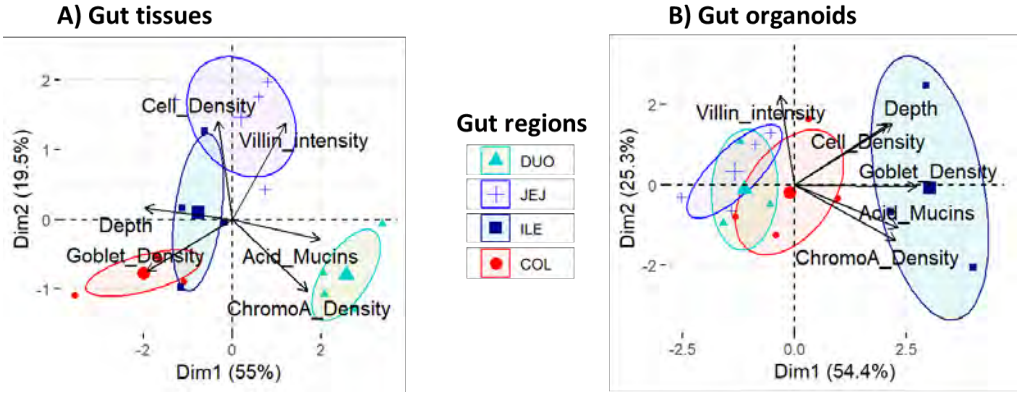


Figure 2. Principal component analyses of (A) gut tissues from different gut regions (duodenum, jejunum, ileum and colon), and (B) their derived organoids after 7 days of culture. Histological and immunohistochemical parameters included: depth of the layer (Depth), cell density (Cell_Density), goblet cell density (Goblet_Density), neuroendocrine cell density (ChromoA_Density), villin-1 expression intensity (Villin_Intensity) and percentage of acid mucins (Acid_Mucins).

References

Beaumont M., Blanc F., Cherbuy C., Egidy G., Giuffra E. *et al.* (2021) Vet. Res. 52(1):1-15. https://doi.org/10.1186/s13567-021-00909-x

Clark E.L., Archibald A.L., Daetwyler H.D., Groenen M.A.M., Harrison P.W. *et al.* (2020) Genome Biol. 21(1):285. https://doi.org/10.1186/s13059-020-02197-8

Clevers H. (2016) Cell 165(7):1586-1597. https://doi.org/10.1016/j.cell.2016.05.082

Criss Z.K., Bhasin N., Di Rienzi S.C., Rajan A., Deans-Fielder K. *et al.* (2021) Physiol. Genomics 53(11):486-508. https://doi.org/10.1152/physiolgenomics.00061.2021

Gehling K., Parekh S., Schneider F., Kirchner M., Kondylis V. *et al.* (2021) bioRxiv 2021.11.22.469588. https://doi.org/10.1101/2021.11.22.469588

Li L., Fu F., Guo S., Wang H., He X. *et al.* (2019) J Virol 93(5):e01682-18. https://doi.org/10.1128/JVI.01682-18

Mohammadi S., Morell-Perez C., Wright C.W., Wyche T.P., White C.H. *et al.* (2021) Stem Cell Reports 16(9):2364-2378. https://doi.org/10.1016/j.stemcr.2021.07.016

Święch E., Tuśnio A., Barszcz M., Taciak M., and Siwiak E. (2019) J Anim Physiol Anim Nutr. 103(3):894-905. https://doi.org/10.1111/jpn.13086

Tsai Y.H., Czerwinski M., Wu A., Dame M.K., Attili D. *et al.* (2018) Cell. Mol. Gastroenterol. Hepatol. 6(2):218-222. https://doi.org/10.1016/j.jcmgh.2018.04.008

Van der Hee B., Loonen L.M.P., Taverne N., Taverne-thiele J.J., Smidt H. *et al.* (2018) Stem Cell Res. 28:165-171. https://www.doi.org/10.1016/j.scr.2018.02.013

Van der Hee B., Madsen O., Vervoort J., Smidt H., and Wells J.M. (2020) Front. Cell Dev. Biol. 8. https://www.doi.org/10.3389/fcell.2020.00375

506. Can microbial data improve prediction of breeding values of efficiency traits in pigs fed conventional or fiber diets?

V. Déru[1,2], F. Tiezzi[3,4], C. Carillier-Jacquin[1], B. Blanchet[5], L. Cauquil[1], O. Zemb[1], A. Bouquet[6], C. Maltecca[3] and H. Gilbert[1*]

[1]GenPhySE, INRAE, 31320 Castanet-Tolosan, France; [2]France Génétique Porc, 35651 Le Rheu, France; [3]Department of Animal Science, North Carolina State University, Raleigh, NC, USA; [4]Department of Agriculture, Food, Environment and Forestry, University of Florence, 50144, Firenze, Italy; [5]UE3P, INRAE, 35590 Saint-Gilles, France; [6]IFIP-Institut du Porc, 35650 Le Rheu, France; helene.gilbert@inrae.fr

Abstract

Recently, digestive efficiency (DE) was proposed as a trait of interest to improve feed efficiency (FE) in pigs, especially when they are fed with alternative feeding resources. Both are influenced by the host genetics, and also by the gut microbiota composition. The goal of this study was to quantify the impact of faecal microbial information on the prediction accuracies of genomic estimated breeding values (GEBVs) of FE and DE traits for pigs fed conventional or fiber diets. For DE traits, gains in prediction accuracy of GEBVs were increased by about 18% when microbial information was included in linear mixed models. In addition, these gains of prediction accuracy were very similar in both diets. For FE traits, no improvement was observed. Thus, the addition of microbial information in breeding programs is promising to better estimate GEBVs for DE traits.

Introduction

Feed costs in pig production currently represent between 60 to 70% of the total cost of pork production (Patience *et al.*, 2015). Feeding pigs with by-products from the agri-food and biofuel industry is an option to reduce feed costs and feed-food competition. In this context, feed efficiency (FE) remains the major objective of selection, but recent research suggested that digestive efficiency (DE) can contribute to FE (Harris *et al.*, 2012; Mauch *et al.*, 2018), and more especially with alternative feeding stuff (Déru *et al.*, 2021a). The DE is influenced by host genetics (Déru *et al.*, 2021a), and it has significant genetic correlations with some faecal microbiota traits (Déru *et al.*, 2021b). In addition, faecal microbiota can explain more than 44% of the phenotypic variation for DE traits (Verschuren *et al.*, 2020; Déru *et al.*, 2021c), and a moderate part, around 20%, for FE traits variance (Camarinha-Silva *et al.*, 2017, Aliakbari *et al.* 2021). The goal of the present study was to evaluate if adding microbial information in the prediction models of breeding values can improve their prediction accuracy for DE and FE traits.

Materials & methods

Animals and diets. Data comprised Large White male pigs reared at INRAE UE3P – France Génétique Porc phenotyping station. Couples of full-sibs were raised in post-weaning facilities until nine weeks of age. When moved to growing-finishing pens, one of the siblings was fed with a two-phase conventional (CO) European diet and the other one a two-phase high-fiber (HF) diet. Diets differed in net energy (NE) (9.6 MJ/kg of NE for the CO diet and 8.2 MJ/kg of NE for the HF diet), and neutral detergent fiber (from 13.90 to 15.12% for the CO diet; from 23.82 to 24.46% for the HF diet), as described in Déru *et al.* (2020).

Data. Individual average daily gain (ADG) and daily feed intake (DFI) were measured between 35 and 115 kg live weight, with GenStar single place electronic feeders (Genstar, Skiold Acemo, Pontivy, France). Feed conversion ratio (FCR) was calculated as DFI/ADG. Residual feed intake (RFI) was determined as a multiple linear regression of DFI on ADG, average metabolic body weight, lean meat percentage and carcass yield (Déru *et al.*, 2020). Microbiota composition and DE were obtained from a unique faecal spot

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_506, © V. Déru *et al.* 2022

sampling at 16 weeks of age. The DE was predicted as digestibility coefficients (DC) of energy, nitrogen and organic matter using near infrared spectrometry (Déru *et al.*, 2021a). Microbiota composition was obtained from sequencing the V3-V4 regions of the 16S rRNA (Déru *et al.*, 2021b), and 14,366 operational taxonomic units (OTU) were identified. Sequence reads were rarefied at 10,000 sequences per sample.

Genotyping was carried out using 70K SNP GeneSeek GGP Porcine chip. Quality control (QC) was carried out: the call rate per individual, i.e. the percentage of genotype present per individual, with a threshold set at 95%, and the SNP call rate, i.e. the percentage of genotype by SNP with a threshold set at 95%. The SNPs with minor allele frequencies lower than 5%, with a significant deviation from Hardy-Weinberg equilibrium (P<0.000001) were also deleted. After QC, 48,919 markers were available for further analyses.

Genomic estimated breeding values (GEBVs) for FCR and RFI (FE traits), ADG, DFI, and DE traits were obtained for pigs that had all microbiota, genomic and digestive efficiency data, i.e. 1,082 pigs raised in 32 batches.

Statistical analyses. The microbial relationship matrix M was computed after filtering out the OTUs present in less than five samples and with an average abundance lower than 0.001% (2,399 OTUs kept), as $M = TT'/k$, where T is the $n \times k$ matrix containing the centred and standardized log abundancies of each OTU, n the number of pigs and k the number of OTUs (Ross *et al.*, 2013). The genomic relationship matrix G was constructed using the first VanRaden method (2008). Predictions were obtained with a Bayesian model fitted with the BGLR package in R (Pérez *et al.*, 2014). Two models were constructed, $y=X\beta +Zu+ Wm + e$ (model M+G) and $y=X\beta +Zu+ e$ (model G), with y the vector of phenotypes for a given trait, X the incidence matrix relating observations to fixed effects, β the vector of fixed effects, Z the incidence matrix for the genetic effects, $u \sim N(0, G\sigma^2_u)$ the vector of additive genetic effects, W the incidence matrix for the microbiota effects, $m \sim N(0, M\sigma^2_m)$ the vector of microbiota effects, and $e \sim N(0, I\sigma^2_e)$ the random residuals.

Gains of prediction accuracies. The advantage of including the microbial information to predict GEBVs was quantified via changes in prediction accuracies of GEBVs from model G to model M+G. A forward cross-validation approach was used, considering the first 25 batches of the population (876 pigs) as a reference population and the last seven batches (205 'candidate' pigs) as a validation population. The reference GEBV, $GEBV_{C,M+G,}$ were obtained with model M+G on the complete population (i.e. phenotypes available for the reference and validation sets). Then, to assess the gains in prediction accuracy with models M and M+G compared to this reference, the criteria ρ_M and ρ_{M+G} were computed as the Pearson correlation coefficients computed between $GEBV_{C,M+G}$ and GEBVs obtained with model G ($GEBV_{P,G}$) or M+G ($GEBV_{P,M+G}$) with a partial dataset, i.e. where phenotypes were available for the reference population and not available for the validation population (Legarra and Reverter, 2018). First, gains of prediction accuracies were obtained for the full validation set. Next, to test for differences in prediction accuracies between diets, ρ_M and ρ_{M+G} were computed separately for the CO (96 pigs) and the HF diet (109 pigs).

Results

The gains of prediction accuracy combining the microbiota and genomic effects were evaluated using the predictions of the M+G model for the complete dataset as reference GEBVs. Indeed, assuming that microbiota composition affects the trait variance components, it should be the model providing the most accurate GEBVs (David and Ricard, 2019). It should be noted that with the complete dataset, correlations between GEBVs obtained with model G and those with model M+G were high, ranging from 0.97 to 0.99 for ADG, DFI and FE traits, and from 0.80 to 0.85 for DC.

Table 1. Correlation coefficients ρ [95% confidence intervals] between estimated breeding values of the validation population obtained with the complete ($GEBV_{C,M+G}$), and with the partial datasets predicted without ($GEBV_{P,G}$) and with ($GEBV_{P,M+G}$) microbial information, considering both diets in the validation population, and separating the validation population according to the diet.[1]

	ρ ($GEBV_{P,G}$,$GEBV_{C,M+G}$)			ρ ($GEBV_{P,M+G}$,$GEBV_{C,M+G}$)		
Traits	**Both diets**	**CO diet**	**HF diet**	**Both diets**	**CO diet**	**HF diet**
FCR	0.60 [0.51;0.68]	0.62 [0.47;0.73]	0.58 [0.45;0.70]	0.59 [0.49;0.67]	0.59 [0.44;0.71]	0.58 [0.44;0.69]
RFI	0.62 [0.53;0.70]	0.72 [0.60;0.80]	0.56 [0.42;0.68]	0.61 [0.52;0.69]	0.70 [0.58;0.79]	0.55 [0.40;0.67]
DFI	0.60 [0.51;0.68]	0.72 [0.61;0.80]	0.57 [0.43;0.68]	0.62 [0.51;0.69]	0.71 [0.60;0.80]	0.56 [0.41;0.68]
ADG	0.70 [0.62;0.76]	0.70 [0.58;0.79]	0.70 [0.58;0.78]	0.70 [0.62;0.76]	0.69 [0.57;0.78]	0.71 [0.60;0.79]
DC of energy	0.55 [0.45;0.64]	0.56 [0.40;0.68]	0.55 [0.40;0.67]	0.67 [0.58;0.74]	0.65 [0.52;0.76]	0.69 [0.57;0.77]
DC of OM	0.53 [0.42;0.62]	0.55 [0.39;0.67]	0.52 [0.37;0.64]	0.64 [0.55;0.71]	0.65 [0.51;0.75]	0.66 [0.54;0.75]
DC of nitrogen	0.51 [0.40;0.61]	0.48 [0.31;0.62]	0.64 [0.39;0.66]	0.62 [0.53;0.70]	0.61 [0.47;0.72]	0.64 [0.51;0.74]

[1] FCR = feed conversion ratio; DFI = daily feed intake; ADG = average daily gain; RFI = residual feed intake; DC = digestibility coefficient; OM = organic matter; CO = conventional; HF = high fiber.

For FE traits, DFI and ADG, ρ were not improved with the model M+G in comparison to the model G (Table 1). In addition, no difference was observed between ρ computed separately for pigs fed the CO and the HF diets, except for DFI and RFI that had slightly higher gains of prediction accuracy in the CO than in the HF diet (for instance 0.72 vs 0.57 for DFI with CO and HF, respectively) with both models. For DE traits, ratios of prediction accuracies tended to be higher in model M+G (~0.65) than in model G (~0.53). There was no difference in prediction accuracy ratios based on the diet for DC, except for DC of nitrogen with model G that had lower ρ in CO than in HF diet. However, because 95% confidence intervals were large due to limited number of pigs in these sub-groups, these differences between diets were not significant.

Discussion

Based on these first results, the joint analysis of microbiota and genomic information does not appear to increase prediction accuracies of GEBVs for FE and growth traits. However, the addition of microbial information to genomic models seemed to improve prediction accuracies for DE traits. A previous study highlighted that a higher proportion of phenotypic variance was explained by gut microbiota for DE traits (≈ 50-60%) than for FE traits (≈ 20%) in this dataset (Déru *et al.*, 2021c). This could explain why adding microbial information in the linear mixed models has more impact for the former traits. However, we also previously found higher proportions of variance due to microbial information in the HF diet than in the CO diet for DE traits, which did not seem to affect the ratios of prediction accuracies obtained when microbial and genomic information were both included in the model. In addition, higher prediction accuracies were observed in the HF diet compared to the CO diet for DC of nitrogen. Previously, heterogeneity of residual variances was observed between diets for this trait (Déru *et al.*, 2021a), which could not be accounted for in the BGLR models, and could explain the higher prediction accuracies in the HF diet. Finally, the GEBV dispersions, appreciated via the regression of GEBVs with the complete dataset on GEBVs of the partial dataset with model G or model M+G (not presented), were also very similar in the two diets.

We can thus conclude that there is an advantage of adding microbial information in linear mixed models to improve the accuracy of breeding values only for DE traits. Furthermore, the accuracy of GEBV predictions was not impaired with the alternative high fiber diet.

References

Aliakbari A., Zemb O., Barilly C., Billon Y. and Gilbert H. (2021). Proc. of the 72nd EAAP, Davos, Switzerland.

Camarinha-Silva A., Maushammer M., Wellmann R., Vital M., Preuss S. *et al.* (2017) Genetics 206(3):1637-1644. https://doi.org/10.1534/genetics.117.200782

Déru V., Bouquet A., Hassenfratz C., Blanchet B., Carillier-Jacquin C. *et al.* (2020) Animal 14:2236–45. https://doi.org/10.1017/S1751731120001275

Déru V., Bouquet A., Labussière E., Ganier P., Blanchet B. *et al.* (2021a) J Animal Breed Genet 138:246-58. https://doi.org/10.1111/jbg.12506

Déru V., Bouquet A., Zemb O., Blanchet B., De Almeida M.L. *et al.* (2021b) Preprint bioRxiv. https://doi.org/10.1101/2021.11.15.468583

Déru V., Tiezzi F., Carillier-Jacquin C., Blanchet B., Cauquil L. *et al.* (2021c) Proc. of the 72nd EAAP, Davos, Switzerland.

Harris A.J., Patience J.F., Lonergan S.M., Dekkers J.M.C., Gabler N.K. (2012) J. Anim. Sci. 90(4):164-166. https://doi.org/10.2527/jas.53855

Mauch E.D., Young J.M., Serão N.V.L., Hsu W.L., Patience J.F. *et al.* (2018) J. Anim. Sci. 96:1221-1236. https://doi.org/10.1093/jas/sky065

Patience J.F., Rossoni-Serão M.C. and Gutiérrez N.A. (2015) J Anim Sci Biotechnol 6(1):33. https://doi.org/10.1186/s40104-015-0031-2

Pérez P. and De los Campos G. (2014) Genetics 198(2):483-495. https://doi.org/10.1534/genetics.114.164442

Ross E.M., Moate P.J, Marett L., Cocks B.G. and Hayes B.J. (2013) J Dairy Sci 96(9):6030-6046. https://doi.org/10.3168/jds.2013-6766

VanRaden P.M. (2008) J Dairy Sci 91(11):4414-23. https://doi.org/10.3168/jds.2007-0980

Verschuren L.M.G., Schokker D., Bergsma R., Jansman A.J.M., Molist F. *et al.* (2020) J Anim Breed Genet 137(1):23-35. https://doi.org/10.1111/jbg.12433

507. Investigating new technologies for on-site real-time sequencing for any animal scientist

H. Ostrovski*, R.P. Savegnago, W. Huang and C. Gondro

Michigan State University, 474 S. Shaw Ln, East Lansing, MI 48824, USA; wackelha@msu.edu

Abstract

Genomic technologies have revolutionized animal breeding but their adoption in commercial animal production systems and throughout the supply chain has been limited due to the complexity of the technology that requires specialized laboratories and trained personnel. Now however, even untrained personnel can potentially easily and rapidly acquire genomic data directly at the sampling site by using small portable real-time sequencing devices. These new sequencing technologies can open a wealth of opportunities for livestock production systems including parentage testing, determination of breed composition, and product certification. This study aims to be the initial exploration into mobile sequencing devices to obtain whole genome information from an Akaushi bull with the Oxford Nanopore MinION. Results of this initial study are the first steps in providing a roadmap for practical adoption of on-site sequencing applications in agricultural production systems which should enable wider industry adoption of genomics to improve food security and livestock production efficiency.

Introduction

Whole genome sequencing has become a widespread practice in most biology disciplines. Both animal breeding and genetics research and the animal industry currently use this kind of data in genomic prediction, whole genome association to identify important QTL (Quantitative Trait Loci), to describe population structures and to understand how selection is being driven on different livestock species. The most recent generation of sequencers, i.e. the third generation, are defined by sequencing single molecules without PCR amplification (Dijk *et al*., 2018). These devices are the most user-friendly, with the products from Oxford Nanopore Technologies (ONT) boasting a small sequencer and straight-forward protocol for any person at various levels of experience. The MinION from ONT (Jain *et al*., 2016), is to be the focus of this study, as we explore the full range of sequencing abilities as a novice in this area by way of obtaining one whole genome sequence.

Understanding the full capabilities of the MinION is crucial in understanding breed composition, parentage or even tracing animal products through the system with an on-site, real-time protocol. Traceability of animal products via some type of genomic testing (DNA barcoding) has already been proposed (Galimberti *et al*., 2013) and studies have shown that even processed food (e.g. smoked fish and cooked meat) still has enough DNA present to be traceable (Aslan *et al*., 2009; Smith *et al*., 2008). The ability to ensure that the composition of a food product matches what is marketed on the label can increase transparency and trust in the system and help reduce the impact of fraudulent food (Spink and Moyer, 2011).

These new ONT technologies bypass sending samples to external labs and long wait times for final genomic information for obtaining real-time genomic sequencing. Establishing usable sequence protocols for all users of genomic information can rapidly increase the usage of this technology in the animal industry, which leads to an increase in the use of genomic data.

The goal of this study was to describe the best protocol to obtain whole genome sequencing of an Akaushi bull using the MinION from ONT. This study aims to be a helpful guide to current molecular geneticists as well as those who have no prior knowledge in the field of molecular genetics by outlining a better use of

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_507, © H. Ostrovski *et al.* 2022

this device to get whole genome sequence data without requiring any difficult lab skills. The methodologies of obtaining a DNA library and sequencing with the MinION will be the focus of this study.

Materials & methods

The subject of interest for obtaining whole genome sequence was a 6-month-old Akaushi bull. Whole blood was collected from this calf and DNA was extracted using the QIAamp blood kit. All experiments run on the MinION were done using the LSK-109 protocol, which is the ligation protocol provided by ONT. This protocol was laid out as an easy-to-use process, which included three steps: (1) DNA repair and end-prep; (2) adapter ligation and clean-up; and (3) priming and loading the flow cell. Seven runs of the same animal were done; each consisted of creating a library, loading it onto a flow cell and running the flow cell for approximately 72hrs.

Guppy basecaller was used to convert the raw signal into fastq files. Post-basecalling steps included removal of chemical adapters with porechop (Wick *et al*., 2017), reads alignment onto the reference genome from 1000 bull Genomes Project (Hayes and Daetwyler, 2019) using minimap2 (Li, 2018), sorting, indexing and merging of bam files with SAMtools (Li and Barrett, 2011; Li *et al.*, 2009), and variant calling with longshot (Edge and Bansal, 2019). BCFtools was used to access the statistics of the final vcf file.

Results

The sequencing protocol that was generated through all runs in this experiment differed from the original protocol taken from the Nanopore website. Many diagnosable issues may arise such as low-quality DNA, not enough or too much extracted DNA used in library prep or even low incubation temperatures with not enough time. Most of the diagnoses were taken from the Nanopore community page, which filled in the gaps of knowledge where the protocol was lacking. The protocol outlined here is aligned with many of the suggestions shared online, as well as trial and error from the runs done.

The final whole genome sequence obtained from this method had 5.2× coverage depth and 2,329,110 variants called. These variants called by longshot only include one sample, which may explain the low number of variants called. Of these variants called, there were 856,924 homozygous (1/1) calls and 926,780 heterozygous (1/0) calls. The output from each run on the MinION is in Table 1. The first runs had the worst outcome due to the learning curve of the protocol techniques that required some troubleshooting. By the final run, the protocol was working more efficiently, and the researcher had more practice with the laboratory techniques. This is still a low amount of coverage for producing a full genome sequence, with a more acceptable coverage hovering from 10-30x. Utilizing low-pass sequencing techniques may be able to bypass this issue and fill in the areas of lower coverage with reference genomes (Snelling *et al.*, 2020).

Table 1. Output of each MinION run at 72 hours.

Run number	Total Gb	Total reads	Mean length	Max length	Ultra-long reads
1	0.13	12,000	11,011	66,867	0
2	1.08	136,000	7,993	117,208	4
3	2.61	262,516	9,945	126,941	6
4	1.79	244,000	7,340	107,872	4
5	3.36	372,000	9,035	112,098	10
6	0.93	112,000	8,334	99,962	0
7	5.53	644,000	8,597	129,442	11

Discussion

The new frontier in animal sequencing is to get it out of the lab, speed up the results collected, and to make it available to all types of researchers and producers. The MinION has the possibility to do all these things, as the size and protocol are easily picked up and easy to adjust to any type of experiment with an experimenter of any skill level. The protocol provided by ONT for producing sequence could be more in-depth, as many steps were simply glazed over, and inexperienced lab users could easily be lead astray. Producing good results of each experiment run with the MinION depends on the user tweaking the amount of reagent used, the time spent incubating or even the time/speed of mixing. Understanding ONT protocols for a specific experiment for a beginner requires reliance on problem solving with the Nanopore community, which can make the sequencing process more labor intensive, with a constant 'debugging' of sorts of the MinION and library preparation. The bioinformatic pipeline produced for alignment of these reads consists of 3rd party software that is free to use except the base calling guppy software and the MinKNOW sequencing software. This could lead to issues for future users who may be using ONT products but do not have access to download this software.

The real-world utilization of this technology is immediately relevant for establishing a traceability system for premium and/or branded products subject to counterfeiting (e.g. wagyu beef or jamon Iberico) to build consumer trust and to detect mislabeled products that tarnish the product's reputation. Practical implementation of on-site, real-time sequencing is not a trivial task but there is a pressing need to improve data connections between farmers and customers, not only for the obvious reasons of improved food safety, but also to capture commercial opportunities through increased exports to other states and countries that already (or will soon) require traceability (Schroeder and Tonsor, 2012). This is a numbers game which, similarly to genomic selection, requires buy-in from the industry, but by implementing these technologies, databases of traceable products would make identification and certification as simple as a Google search.

The MinION has a long laundry list of pros and cons, but the novelty of this technology is important for use in the animal industry. This is because the idea of mobile sequencing, which can be used by any person of any skill level, is unheard of in the sequencing space. This is the only sequencer on the market that can be taken out of the laboratory, which opens the door to the opportunity of on-farm sequencing, at the right price. The practical application of this aspect allows for on-site diagnosis of disease, quick genomic sequencing of animals for quick parentage testing and even traceability of animal products that will become commonplace for the everyday scientist.

References

Aslan, Ö., Hamill, R. M., Sweeney, T., Reardon, W., and Mullen, A. M. (2009). Integrity of nuclear genomic deoxyribonucleic acid in cooked meat: Implications for food traceability. Journal of Animal Science, 87(1), 57–61. https://doi.org/10.2527/jas.2008-0995

Dijk, E. van, Jaszczyszyn, Y., Naquin, D., and Thermes, C. (2018). The third revolution in sequencing technology. Trends in Genetics, 34(9), 666–681.

Edge, P., and Bansal, V. (2019). Longshot enables accurate variant calling in diploid genomes from single-molecule long read sequencing. Nature Communications 2019 10:1, 10(1), 1–10. https://doi.org/10.1038/s41467-019-12493-y

Galimberti, A., De Mattia, F., Losa, A., Bruni, I., Federici, S., *et al.* (2013, January 1). DNA barcoding as a new tool for food traceability. Food Research International. Elsevier. https://doi.org/10.1016/j.foodres.2012.09.036

Hayes, B. J., and Daetwyler, H. D. (2019). 1000 Bull Genomes Project to Map Simple and Complex Genetic Traits in Cattle: Applications and Outcomes. Annual Review of Animal Biosciences, 7, 89–102. https://doi.org/10.1146/annurev-animal-020518

Jain, M., Olsen, H. E., Paten, B., and Akeson, M. (2016). The Oxford Nanopore MinION: Delivery of nanopore sequencing to the genomics community. Genome Biology, 17(1). https://doi.org/10.1186/s13059-016-1103-0l

Li, H. (2018). Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics, 34(18), 3094–3100. https://doi.org/10.1093/BIOINFORMATICS/BTY191

Li, H., and Barrett, J. (2011). A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. Bioinformatics, 27(21), 2987–2993. https://doi.org/10.1093/BIOINFORMATICS/BTR509

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., *et al.* (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics (Oxford, England), 25(16), 2078–2079. https://doi.org/10.1093/BIOINFORMATICS/BTP352

Schroeder, T. C., and Tonsor, G. T. (2012). International cattle ID and traceability: Competitive implications for the US. Food Policy, 37(1), 31–40. https://doi.org/10.1016/j.foodpol.2011.10.005

Smith, P. J., McVeagh, S. M., and Steinke, D. (2008). DNA barcoding for the identification of smoked fish products. Journal of Fish Biology, 72(2), 464–471. https://doi.org/10.1111/j.1095-8649.2007.01745.x

Snelling, W. M., Hoff, J. L., Li, J. H., Kuehn, L. A., Keel, B. N., *et al.* (2020). Assessment of Imputation from Low-Pass Sequencing to Predict Merit of Beef Steers. Genes, 11(11), 1312. https://doi.org/10.3390/genes11111312

Spink, J., and Moyer, D. C. (2011). Defining the Public Health Threat of Food Fraud. Journal of Food Science, 76(9), R157–R163. https://doi.org/10.1111/j.1750-3841.2011.02417.x

Wick, R. R., Judd, L. M., Gorrie, C. L., and Holt, K. E. (2017). Completing bacterial genome assemblies with multiplex MinION sequencing. Microbial Genomics, 3(10). https://doi.org/10.1099/MGEN.0.000132

508. Impact of the bovine reference genome (ARS-UCD1.2) on functional annotation at genome, transcriptome and methylome level

A. Cánovas[1*], P.A.S. Fonseca[1], S. McKay[2] and J.F. Medrano[3]

[1]Department of Animal Biosciences, Centre for Genetic Improvement of Livestock, University of Guelph, N1G 2W1, Guelph, Ontario, Canada; [2]Department of Animal and Veterinary Sciences, University of Vermont, USA; [3]Department of Animal Science, University of California-Davis, Davis, CA 95616, USA; acanovas@uoguelph.ca

Abstract

The ARS-UCD 1.2 genome assembly was released with the expectation of improving continuity and accuracy of the assembly for the bovine reference genome compared with the previous assembly, UMD 3.1. Here, the impact of both reference genomes was compared at the genome, transcriptome and methylome level over the results obtained using different high-throughput OMICs technologies. In general, the ARS-UCD 1.2 outperformed the UMD 3.1 genome assembly. For genome wide association studies, a larger detection of positional and functional candidate genes was observed. In addition, a larger number of uniquely mapped reads, expressed genes, and transcripts was observed in transcriptome analyses using RNA-Sequencing. Interestingly, for the genome wide methylation analysis, the chance of annotating methylated sites in genic regions was ~40% higher than in intergenic regions. The results obtained allowed a proof-of-concept study to be performed regarding the advantages of the ARS-UCD 1.2 over the UMD 3.1 bovine genome assembly.

Introduction

The availability of a bovine reference genome substantially improved the elucidation of genetic mechanisms associated with complex traits. For more than a decade, the bovine reference genome UMD 3.1 was used as the main reference genome for genomic and transcriptomic analyses (Zimin *et al.*, 2009). However, mainly due to the limitations of the sequencing technologies, gaps in the annotation and poorly assembled regions are frequent in this version. Recently, the ARS-UCD 1.2, was released based on a *de Novo* assembly using single-molecule real-time sequencing (Rosen *et al.*, 2020). Theoretically, the use of longer reads in the process of genome assembly can result in improvements of continuity, accuracy, and completeness of assembly. Therefore, the objectives of this study were: (1) to evaluate the differences between the bovine reference assemblies UMD 3.1 and ARS-UCD 1.2 and to compare the results obtained at the genome, transcriptome, and methylome level using different high-throughput -OMICs technologies and approaches (genome wide association study (GWAS), RNA-Sequencing (RNA-Seq), and whole genome bisulfite sequencing (WGBS)); (2) to estimate the real impact of the new genome assembly for these technologies for functional annotation at the genome, transcriptome and methylome level.

Materials & methods

Phenotypes and genotypes used for genome-wide association studies. The phenotype evaluated was meat tenderness in beef cattle defined using Warner-Bratzler shear force to measure the *longissimus dorsi* muscle (LMD) that had been aged for 7-, 14- and 21-days post-mortem (LMD7, LMD14 and LMD21), respectively. A total of 1,740, 1,295, and 1,288 crossbred beef cattle were sampled for LMD7, LMD14 and LMD21, respectively. The mean (and standard deviation) breed composition of the six major breeds for these animals was: Angus (0.43±0.02), Simmental (0.33±0.17), Charolais (0.05±0.10), Piedmontese (0.05±0.15), Gelbvieh (0.02±0.07), and Limousin (0.01±0.05). The animals were genotyped using the Illumina BovineSNP 50K. A total of 37,564 autosomal markers were used after quality control. The weighted single-step method (Wang *et al.* 2012) was used to estimate the SNP effects, where a single-

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_508, © A. Cánovas *et al.* 2022

trait model was considered with fixed effects including sex, herd, year, month of birth, expected breed composition, and slaughter age (Fonseca *et al.* 2022). The candidate windows associated were defined as those windows explaining more than 1% of the total genetic variance. The positional candidate genes within the candidate windows were annotated for both UMD 3.1 and ARS-UCD 1.2 assemblies using the GALLO package in R (Fonseca *et al.* 2020). The proportion of positional candidate genes (mapped within candidate windows) shared between each time-point and assemblies were compared. Additionally, the pairwise linkage disequilibrium (LD) pattern within the candidate windows was calculated using plink v1.9 (Purcell *et al.* 2007).

Comparison groups for transcriptomic analysis using RNA sequencing. A total of 12 milk somatic cell samples from six Holstein dairy cows diagnosed with mastitis according to the California Mastitis Test were collected and used in transcriptomic analyses. Immediately after diagnosis, two milk samples were taken from each cow, one sample from the mastitic udder quarter (n=6), and one from the udder quarter diagonal to the mastitic udder quarter and classified as healthy based on having a somatic cell count <100,000 cells/ml (Asselstine *et al.*, 2019; Guerrero *et al.*, 2015). Sequencing was completed with an Illumina HiSeq 2000 analyzer that yielded 100-bp single-read sequences. Sequence reads were assembled to the annotated UMD 3.1 and ARS-UCD 1.2 bovine reference genome using CLC genomics workbench software (CLC Bio, Aarhus, Denmark). Transcript levels were quantified in reads per kilobase per million mapped reads (RPKM). Analysis of differentially expressed (DE) genes and mRNA isoforms was performed between healthy and mastitic quarters by t-test. Genes and mRNA isoforms were considered significantly DE between healthy and mastitic udder quarters when they met the following significance thresholds: a false discovery rate (FDR) of $q<0.05$ and a fold-change (FC)$>|2|$. The percentage of uniquely mapped reads, total number of expressed genes (RPKM$\geq$0.2) and DE genes, total number of expressed mRNA isoforms, and number of DE mRNA isoforms between healthy and mastitic samples was compared between assemblies.

Whole genome bisulfite sequencing (WGBS) data processing, mapping and methylation calling. The samples used in this analysis were derived from biopsies of two brain areas, *amygdala* (*amy*) and cingulate gyrus (*gyrus*), from eight animals. The extracted DNA was used to prepare 150 bp paired end libraries and the WGBS was performed using the Illumina HiSeq X Ten platform (Illumina, San Diego, CA, USA). The methylation calling was performed using BSseeker2 software with default parameters. The percentage of methylated cytosine distribution was compared between assemblies for *amy* and *gyrus* in all methylation contexts (CG, CHG, and CHH). The bases with a frequency of methylated cytosine higher than zero were annotated based on their genomic function (i.e. intergenic, promoter, exon, and intron). An ordinal logistic regression was performed using the genomic function as the response variable, and the different assemblies were treated as fixed effects. The magnitude of the effect was estimated using the odds ratio between the reference genomic function (intergenic) and other genomic functions as the response variable. The significance of the odds ratio was assessed based on the 95% confidence interval (CI95%).

Results

A total of 119 and 103 genes were annotated within the candidate windows identified associated with beef meat quality from GWAS analysis using ARS-UCD1.2 and UMD3.1, respectively. For LMD7, the window explaining most of the total genetic variance was mapped on BTA29 for both assemblies, harbouring 52 (ARS-UCD1.2) and 40 (UMD 3.1) genes. In total, 29 genes were shared between these windows. The percentage of the genetic variance explained by the top window between assemblies for each meat quality trait varied only for LMD7 (4.63% in ARS-UCD1.2 and 4.58% in UMD3.1). The LD pattern across the markers mapped within these windows helped to explain this difference. The markers present uniquely in the UMD3.1 assembly within the candidate window have no significant LD with the markers present in both assemblies. On the other hand, the majority of the markers uniquely mapped in the candidate window

identified using the ARS-UCD 1.2 assembly show a moderate to strong LD with the makers mapped within the candidate window in both assemblies. These results suggest that the markers mapped uniquely within the candidate window identified using the UMD3.1 assembly mapped incorrectly due to possible assembly errors. At the transcriptome level, the RNA-Seq analysis revealed that 81.0 and 86.5% of the reads were uniquely mapped to the UMD3.1 (24,596 genes) and ARS-UCD1.2 (27,270 genes) reference genomes, respectively. The total number of genes expressed in healthy and mastitic groups was higher using the ARS-UCD1.2 (14,096 and 12,820, respectively) when compared to UMD 3.1 (13,356 and 12,145, respectively). In total, 449 genes were DE using the UMD 3.1, while 413 were identified as DE using the ARS-UCD 1.2. Of these, 353 genes were shared between both assemblies. The total number of transcripts expressed in the healthy and mastitic groups was 14,725 and 13,334 (UMD 3.1) and 21,808 and 19,173 (ARS-UCD1.2). In total, 516 DE transcripts were identified in both UMD 3.1 and ARS-UCD 1.2 assemblies, corresponding to 502 genes for UMD 3.1 and 469 genes for ARS-UCD 1.2. Of these 516 DE transcripts found using the UMD 3.1 and ARS-UCD 1.2 reference genomes, 322 DE transcripts were shared between the reference genomes, while 180 and 147 were unique to the UMD 3.1 and ARS-UCD 1.2 reference genomes, respectively. The comparison between assemblies at the methylome level suggested that there is not a different pattern of mean methylation percentage: UMD 3.1 (CHG: 0.67%±0.14, CHH: 0.76%±0.16, CG: 77.31%±1.99) and ARS-UCD 1.2 (CHG: 0.67%±0.14, CHH: 0.76%±0.16, CG: 76.88%±1.70). The methylated sites called tend to be mapped more frequently in intergenic and genic areas (exons, introns, and promoters) using the UMD 3.1 and ARS-UCD1.2, respectively. Significant odds ratio of 0.708 (CI95%= 0.707-0.708) and 0.716 (CI95%= 0.715-0.716) were obtained comparing the annotation records between assemblies. These results suggest that the use of the new bovine reference genome, ARS-UCD 1.2, results in 41.3 and 40% higher annotation of methylated sites in genic regions than intergenic regions, for *amy* and *gyrus* tissues, respectively.

Discussion

The functional analysis including the biological processes associated with the 23 genes annotated exclusively in the candidate window on BTA29 for the ARS-UCD1.2 assembly indicated that those genes are associated with cell-cell adhesion (*BAD*, *CCDC88B*, and *FERMT3*), response to lipids (*BAD*, and *ESSRA*), and regulation of proteolysis (*BAD*, and *PRDX5*), which are associated with meat quality characteristics, such as shear force, color, water retention, cocking loss, etc. (Zhang *et al.* 2013). At the transcriptome level using RNA-Seq technology, the better annotation of the new reference genome and the more complete coverage of ARS-UCD 1.2 can help to explain the smallest number of DE genes. The results obtained at the methylome level using WGBS reinforced better gene annotation available on ARS-UCD 1.2 reference genome due to the ~40% increase in probability of annotating methylated sites in genic regions. The number of methylation calls did not change between assemblies, suggesting that the differences in the percentage of methylation calls per genomic locations are explained by differences in gene annotation between the genomes. Novel genes were annotated by ARS-UCD 1.2 in regions that were previously being considered intergenic, which can be explained by the improvements of continuity, accuracy, and completeness of the ARS-UCD 1.2 assembly. These help to increase the functional interpretation of the results and to improve the detection of candidate genes associated with epigenetic markers. The results obtained here reinforce the advantages of the new bovine reference genome. The ARS-UCD 1.2 assembly is >100× more continuous than UMD3.1 and has complete coverage of the genome (Rosen *et al.*, 2020). Using a multi-OMICs approach, it was possible to perform a proof-of-concept study regarding the advantages of the ARS-UCD 1.2 over UMD 3.1 bovine genome assemblies.

References

Asselstine V., Miglior F., Suárez-Veja A., Fonseca P.A.S., Mallard B., *et al.* (2019). Genetic mechanisms regulating the host response during mastitis. J. Dairy Sci. 102:9043-9049. https://doi.org/10.3168/jds.2019-16504.

Fonseca P.A.S., Caldwell T., Mandell I., Wood K., and Cánovas A. (2022). Genome-wide association study for meat tenderness in beef cattle identifies patterns of the genetic contribution in different post-mortem stages. Meat Science, 108733. https://doi.org/10.1016/j.meatsci.2022.108733

Fonseca P.A., Suárez-Vega A., Marras G., and Cánovas A. (2020). GALLO: An R package for genomic annotation and integration of multiple data sources in livestock for positional candidate loci. GigaScience, 9(12), giaa149. https://doi.org/10.1093/gigascience/giaa149

Guerrero A., Dallas A.C., Contreras S., Bhandari A., Canovas A., *et al.* (2015). Peptidomic analysis of healthy and subclinically mastitic bovine milk. Int. Dairy J. 46:46–52. https://doi.org/10.1016/ jidairyj.2014.09.006

Purcell S., Neale B., Todd-Brown K., Thomas L., Ferreira M.A.R., *et al.* (2007). PLINK: A tool set for whole-genome association and population-based linkage analyses. Am. J. Hum. Genet. https://doi.org/10.1086/519795.

Rosen B.D., Bickhart D.M., Schnabel R.D., Koren S., Elsik C.G., *et al.* (2020). De novo assembly of the cattle reference genome with single-molecule sequencing. GigaScience 9, giaa021. https://doi.org/10.1093/gigascience/giaa021

Wang H., Misztal I., Aguilar I., Legarra A., and Muir W.M. (2012). Genome-wide association mapping including phenotypes from relatives without genotypes. Genet. Res. (Camb). https://doi.org/10.1017/S0016672312000274

Zhang M., Wang D., Huang W., Liu F., Zhu Y., *et al.* (2013). Apoptosis during postmortem conditioning and its relationship to duck meat quality. Food Chem. https://doi.org/10.1016/j.foodchem.2012.10.142

Zimin A.V., Delcher A.L., Florea L., Kelley D.R., Schatz M.C., *et al.* (2009) A whole-genome assembly of the domestic cow, Bos taurus. Gen. Bio. 10, R42. https://doi.org/10.1186/gb-2009-10-4-r42

Part 31

Functional annotation of genomes: causal variants & gene discovery

509. Detection of genes influenced by environmentally responsive *cis*-regulatory variation in porcine immune cells

E. Murani*, F. Hadlich, S. Ponsuksili, and K. Wimmers

Research Institute for Farm Animal Biology (FBN Dummerstorf), Institute of Genome Biology, Wilhelm-Stahl-Allee 2, 18196 Dummerstorf, Germany; murani@fbn-dummerstorf.de

Abstract

Genotype-by-environment interactions play an important role in adaptation. In spite of the growing relevance and interest in improving resilience of farm animals studies exploring functional genomic background of G×E interactions are scarce. Here we harnessed RNA-Seq data from an *in vitro* challenge experiment (using dexamethasone, lipopolysaccharide and their combination) in mononuclear leukocytes to detect genes influenced by environmentally responsive *cis*-regulatory variation in pigs using the GeneiASE method. We identified 82 genes showing condition-dependent allele specific expression in at least one of a total of four treatment comparisons. Several of the identified genes are involved in the regulation of response to stress, including *IDO1*, showing allele specific expression depending on LPS application. Our results demonstrate that the analysis of conditional ASE in cell culture models is a promising way to unravel the molecular background of G×E interactions as demonstrated previously in humans.

Introduction

Changes in gene expression facilitate rapid and quantitative responses and thus allow fine-tuned adaptation to environmental stimuli (López-Maury *et al.* 2008). Genetic variation in environmentally modulated transcriptional regulators and/or in their genomic response elements can lead to genotype-by-environment (G×E) interactions in the expression of the corresponding gene(s) and eventually to G×E variation in the phenotype. Gene expression can be considered as a quantitative trait and analysed using the same statistical genomics methods as quantitative phenotypic traits (eQTL analysis). However, this requires measurement of gene expression in large cohorts. The advent of next-generation sequencing for gene expression analysis (RNA-Seq) enables genome-wide analysis of expression imbalance of transcribed 'reporter' SNPs (allele-specific expression – ASE). ASE analysis is a sensitive approach to detect regulatory genetic variation acting in *cis*, i.e. directly on the linked gene(s), and performs with reasonable power also in relatively small cohorts. Comparison of ASE between different environments can be used to detect genes influenced by environmentally responsive *cis*-regulatory variation (Esgärd *et al.* 2016, Knowles *et al.* 2017). Impact of G×E interactions on ASE can be particularly well modelled and explored in cultured cells *in vitro* as demonstrated in humans (Moyerbrailean *et al.* 2016). Here we report analysis of ASE in RNA-Seq data from porcine peripheral blood mononuclear cells (PBMC) cultured in four different environments (Li *et al.* 2021). The four conditions included, besides control (C), stimulation with dexamethasone (DEX), lipopolysaccharide (LPS), and their combination (LPS+DEX). These conditions mimic the impact of neuroendocrine stress response (i.e. glucocorticoids), bacterial infection, and neuroimmunomodulation, on immune cells, respectively. The aim of this study is to identify genes whose expression is influenced by G×E interactions and thus to provide basis for genetic improvement of resilience in pigs.

Materials & methods

Cell Culture experiment and RNA-Seq. The cell culture experiment and data generation are described in detail in Li *et al.* (2021). Briefly, PBMC from 24 slaughter pigs (both sexes equally represented) were harvested for RNA extraction after stimulation for 2 h with either vehicle, DEX, LPS, or LPS+DEX. RNA-Seq was performed on an Illumina HiSeq 2500 sequencing platform with 2 × 101 bp paired-end reads. The resulting average sequencing depth was >20 M read pairs with an alignment rate of ~98%.

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_509, © E. Murani *et al.* 2022

SNP Variant calling and counting. SNP calling and counting of reads per allele was based on pipeline described by Tomlinson *et al.* (2021). Parameter settings were taken from https://github.com/CRG-CNAG/CalliNGS-NF. The RNA-Seq reads were aligned to Sus.Scrofa 11.1 reference genome (annotation information from release 103) using STAR (v.2.7.8a) and filtered to remove duplicates and reads with ambiguous mapping. Variants were called using GATK (4.2.0.0) and used to make a masked genome sequence. Mapping and variant calling were then repeated using the masked genome to remove mapping bias.

ASE analysis using GeneiASE. Prior to ASE analysis, SNPs covered by less than 30 reads in total, less than 3 reads per either allele (in heterozygous samples), and with lower than 1% contribution to the total read count per allele, were discarded to increase reliability of the analysis. SNPs located on sex chromosomes, unmapped SNPs, as well as SNPs with ambiguous gene assignment were omitted. In addition, four samples (including data from all four treatments) were excluded based on principal-component analysis (Li *et al.* 2021), leaving 80 RNA-seq datasets for ASE analysis (20 samples × 4 conditions). ASE analysis was conducted using the GeneiASE method developed by Esgärd *et al.* (2016). GeneiASE performs gene-wise ASE analysis for genes with ≥2 SNPs within each informative sample. We implemented GeneiASE within each treatment (called static ASE), and between pairs of treatments in order to detect G×E interactions (called condition-dependent cd-ASE). The individual ASE results were combined across samples using Fisher's method and adjusted using Benjamini-Hochberg (BH) correction.

Functional annotation of ASE genes. Functional annotation of genes showing ASE was performed using the ToppGene online suite (https://toppgene.cchmc.org/enrichment.jsp). Overlapping QTLs were searched in the PigQTLdb (www.animalgenome.org).

Results

In total, we identified >4.4 M transcribed SNPs. After filtering, 75,620 informative SNPs remained for the ASE analysis. Out of 6,010 analysed genes we found significant static ASE (BH-FDR<0.05) in 1,184 genes in C, 1,198 genes in DEX, 1,068 genes in LPS, and 1,072 genes in LPS+DEX. Genes showing static ASE in each of the four treatments (n=509) were enriched for genes involved in the innate immune system (n=57). As previously reported (Li *et al.* 2021), DEX, LPS and their combination induced vast transcriptional changes compared to control, proving suitability of the treatments to study G×E interactions. Results of the cd-ASE analysis (genes with BH-FDR<0.05) are summarized in Table 1. The highest number of cd-ASE was detected in the comparison of the DEX with the LPS treatment, likely because these represent the most contrasting regulatory environments (anti-inflammatory vs inflammatory). In total, 82 different genes showed cd-ASE in at least one treatment. Functional annotation showed that these genes are mainly involved in the regulation of response to stress (GO:0080134, 17 genes). One of the most functionally relevant genes is *IDO1*, showing cd-ASE in all comparisons including LPS (i.e. excluding DEX vs C). Another gene with a similar pattern is ENSSSCG00000020967 (*CD300H* according to NCBI annotation). Interestingly, while *IDO1* is upregulated by LPS, ENSSSCG00000020967 shows the opposite response to LPS. No relevant overlapping QTLs were found in the PigQTLdb.

Table 1. Overview of the results of the cd-ASE analysis.

Comparison	# of analysed genes	# of genes showing cd-ASE
DEX vs C	3,649	13
LPS vs C	3,463	23
LPS+DEX vs C	3,478	26
DEX vs LPS	3,441	35

Discussion

In this study, we applied the GeneiASE method to detect genes influenced by environmentally responsive *cis*-regulatory variation. We identified 82 genes showing cd-ASE. *IDO1*, which showed ASE depending on LPS stimulation, is a rate-limiting enzyme of tryptophan catabolism along the kynurenine pathway. By controlling tryptophan availability and producing bioactive metabolites, IDO1 plays an important immunomodulatory role (Pallotta *et al.* 2021). Unfortunately, no relevant QTLs at the *IDO1* locus were found in PigQTLdb calling for further investigation of the causal regulatory variation and linking it with the immune response. The other promising candidate, ENSSSCG00000020967 (*CD300H*), is less well characterised, but is likely involved in innate immunity (Niizuma *et al.* 2015). The observation that genes showing conditional ASE are enriched for regulation of response to stress is in line with the view that stress-related genes are prone to environmentally responsive *cis*-regulatory variation which is beneficial for adaptation to changing environment. In addition to genes exhibiting conditional ASE, we identified several hundred genes showing static ASE. These represent additional potential candidates with impact on the immune and stress response. By performing gene-wise analysis we likely detected only a part of the present *cis*-regulatory variation. Our further investigation will therefore involve analysis of ASE for individual SNPs.

References

Edsgärd D., Iglesias M.J., Reilly S.J., Hamsten A., Tornvall P. *et al.* (2016) Sci Rep 18;6:21134. https://doi.org/10.1038/srep21134

Knowles D.A., Davis J.R., Edgington H., Raj A., Favé M.J. *et al.* (2017) Nat Methods 14(7):699-702. https://doi.org/10.1038/nmeth.4298

Li Z., Trakooljul N., Hadlich F., Ponsuksili S., Wimmers K. *et al.* (2021) Virulence 12(1):1808-1824. https://doi.org/10.1080/21505594.2021.194827

López-Maury L., Marguerat S., Bähler J. (2008) Nat Rev Genet 9(8):583-593. https://doi.org/10.1038/nrg2398

Moyerbrailean G.A., Richards A.L., Kurtz D., Kalita C.A., Davis G.O. *et al.* (2016) Genome Res 26(12):1627-1638. https://doi.org/10.1101/gr.209759.116

Niizuma K., Takahara-Hanaoka S., Noguchi E., Shibuya A. (2015) J Biol Chem 290(36): 22298-22308. https://doi.org/10.1074/jbc.M115.643361

Pallotta M.T., Rossini S., Suvieri Ch., Coletti A., Orabona C. *et al.* (2021) FEBS J. Online ahead of print. https://doi.org/10.1111/febs.16086

Tomlinson 4th M.J., Polson S.W., Qiu J., Lake J.A., Lee W. *et al.* (2021) Sci Rep 11(1):3968. https://doi.org/10.1038/s41598-021-83459-8.

510. Insight into the most significant quantitative trait locus for calving traits on BTA18 in Holstein dairy cattle

N. Dachs[1,2], M. Upadhyay[1], E. Hannemann[1], A. Hauser[3], S. Krebs[3], H. Blum[3], D. Seichter[2], I. Russ[2], L.J. Gehrke[4,5], G. Thaller[4] and I. Medugorac[1*]

[1]Population Genomics Group, Department of Veterinary Sciences, LMU Munich, Lena-Christ-Str. 48, 82152 Martinsried, Germany; [2]Tierzuchtforschung e.V. München, Senator-Gerauer-Str., 23, 85586 Poing, Germany; [3]Laboratory for Functional Genome Analysis (LAFUGA), Gene Center, LMU Munich, Feodor-Lynen-Straße 25, 81377 Munich, Germany; [4]Institute of Animal Breeding and Husbandry, Christian-Albrechts-University Kiel, Olshausenstraße 40, 24098 Kiel, Germany; [5]Vereinigte Informationssysteme Tierhaltung w.V. (vit) Verden, Heinrich-Schröder-Weg 1, 27283 Verden (Aller), Germany; i.medugorac@lmu.de

Abstract

While the quantitative trait locus (QTL) on BTA18 associated with paternal calving ease and still-birth in Holstein Friesian cattle and its cross is known for over 20 years now, to our knowledge, no study has yet identified the exact causal mutations. Here, we re-examined the region of this QTL on BTA18 and investigated possible reasons behind elusiveness of the exact causal mutations or genes. Using genotyping data of 2,697 German Holstein cattle, we identified the QTL in the region between 58 and 60 Mbp on BTA18. Subsequently, using WGS data, four SNPs in perfect LD with the QTL haplotype were also identified. Our analyses also revealed the abundance of segmental duplications (SDs) within and around the region of QTL. We hypothesize that the presence of complex SDs could have been the major factor contributing to the alignment-difficulty that we observed in our data and thus, concealing the underlying causal mutations.

Introduction

Holstein Friesian (HF) and its derived breeds are amongst the most economically important dairy cattle all over the world. However, certain pathological conditions in these animals, such as low fertility, dystocia or the production of weak calves, which die during the first 24 hours of birth, are some major issues which have significant economic impact on dairy industry. According to one estimate (Bellows *et al.*, 2002), these conditions could lead to a total loss of about 53.20 USD (~EUR 46.74-47.81) per dairy cow per year. It is noteworthy that a part of this financial losses can be attributed to genetic factors such as QTL affecting fertility in cattle. Indeed, for over 20 years now, QTL, having by far the most significant impact on fertility in HF and its derived breeds, is known. In fact, many studies (e.g. Cole *et al.*, 2009, Müller *et al.*, 2017) have mapped this QTL in the distal end (between 50 and 60 Mbp) of BTA18. Further, genes which are close to this region such as *SIGLEC5* and *ZNF613* are proposed as the candidate genes harboring the causal mutations (Cole *et al.*, 2009, Fang *et al.*, 2019). Despite the economic and ethical impact of this QTL none of the existing publications successfully pinpointed the causal genes or mutations. Therefore, the aim of the present study was to confirm the position of the significant QTL on BTA18 (ARS-UCD1.2) in an extended data set to Müller *et al.* (2017) and to investigate the region of interest in a subsequent mapping and WGS analyses.

Materials & methods

Animals, their phenotypes, and SNP array genotypes. The previous study by Müller *et al.* (2017) had a sample size of 2,572 daughter-proven Holstein sires. The data set was extended to 2,697 animals separated into 2,525 males and 172 females. Pedigree records and EBV for all animals were obtained from the official breeding value evaluation. The calving traits, paternal calving ease, and paternal stillbirth were the traits that we investigated here. Calving ease is segmented into the following four classes: Easy, normal, heavy and with veterinary assistance. Stillbirth is specified as an 'all-or-nothing' trait and defined as calving

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_510, © N. Dachs *et al.* 2022

where the calf was born dead or died within 48 hours. All 2,697 animals were genotyped as part of the projects mentioned above with the Illumina BovineSNP50 BeadChip version 1 or version 2. Additionally, 256 of the 2,697 animals were also genotyped with the BovineHD BeadChip. Subsequently, the physical positions of the SNPs from these SNP-chips were converted to the latest cattle assembly (ARS-UCD1.2). Next, genotypes were filtered using the standard criteria and phasing was carried out using *Beagle 5.0* (Browning and Browning, 2009, Browning *et al.*, 2018). To confirm the results by Müller *et al.* (2017) and to identify the position of the most significant QTL on the new assembly ARS-UCD1.2, Combined Linkage Disequilibrium and Linkage Analysis (cLDLA) analysis was performed. This approach was based on the well described method by (Meuwissen *et al.*, 2002) and is already elaborated in detail in our previous study (Gehrke *et al.*, 2020).

Whole genome sequencing (WGS) and targeted SNP genotyping. Four HF (2: heterozygous for the QTL, 1: homozygous for reference-QTL alleles, 1: homozygous for the QTL alleles), one Fleckvieh (FV) and one Kärntner Blondviehwere (KBV) animals were sequenced with long read Oxford Nanopore Technology (ONT). After standard filtering and correction procedure, these data were used to *de-novo* assemble BTA18 chromosome and for the calling of structural variations (SVs). We also downloaded publicly available whole-genome Illumina paired end sequences from additional 24 cattle samples (21 HF and 3 non-HF). After mapping these data against ARS-UCD1.2 cattle assembly, SNPs and indels were called using GATK haplotypecaller and SVs were called using *Lumpy* v0.3.1 (Layer *et al.*, 2014), *BREAKDANCER* v.1.4.5 (Chen *et al.*, 2009) and *MANTA* v1.6.0 (Chen *et al.*, 2016). Additionally, segmental duplications (SDs) were also identified from BTA18 using the approach as described in Dallery *et al.* (2017). Additionally, target SNP genotyping was performed by PCR-RFLP and PCR-based KASP technology for the SNPs that we identified as linked with the QTL. The genotyping was carried in 427 animals belonging to eight different breeds (Table 1).

Results

The *cLDLA* pipeline using 1,144 SNPs (from 50K SNP array) on BTA18 identified a region between 54,233,746 and 63,206,119 bp as having the genome-wide significant LRT values (Figure 1a). It indicates that this region harbored the QTL for the traits of paternal calving ease. After comparing the maternal and paternal haplotypes within the 40-SNP window harboring the LRT_{max} in the middle marker interval, we were able to identify a common putative harmful haplotype (Q) in 608 individuals that showed increased calving difficulties and stillbirth. Furthermore, using HD marker data of animals carrying Q/Q and q/Q genotypes, the putative harmful haplotype could be narrowed down to a region from 57,922,208-60,057,741 bp. Interestingly, in ONT and Illumina sequences of HF cattle, we observed a chromosomal segment of about 16,000 bp (58,846,130-58,861,709 bp), where no read could be mapped to the reference genome and therefore, it appeared as a widespread gap. However, this gap was also observed in zebu and non-HF taurine cattle. Because of this, we excluded this gap as a causal deletion inducing dystocia and stillbirth.

By analyzing publicly downloaded Illumina WGS data, we observed that 2 of the 21 HF samples were heterozygous (Q/q) for the QTL; therefore, we investigated whether we could find variants (SNPs, indels or SVs) from WGS data (Illumine and ONT) that are in perfect LD with the casual QTL. Our analysis revealed that of all the genomic variants that we detected (SNPs, indels or SVs), only four SNPs (Table 1) fulfilled this criterion. Further, except, rs381577268 (57,816,137 bp), which was located downstream of the zinc finger protein 613 (ZNF613) gene, all SNPs were intergenic. Additionally, target SNP genotyping in more than 400 animals of 8 taurine cattle breeds (Table 1) revealed that alternative alleles for these SNPs were only present in HF and HF-derived animals having the QTL.

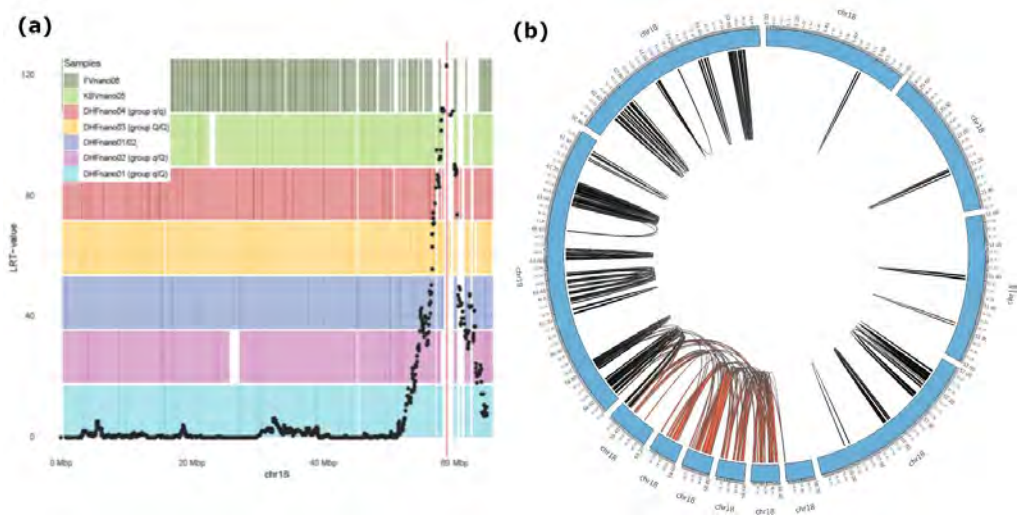


Figure 1. (a) distribution of the likelihood ratio test values (LRT) of paternal calving ease on BTA18 with the maximum peak (LRT_{max}=122.9) at 58,860,538 bp and coverage of BTA18 by contigs from *de novo* assemblies (from the top row: (1) FV; (2) KBV; (3) HF_qq; (4)HF_QQ; (5) HF_Qq: assembly of two heterozygous HF samples combined; (6) HF_Qq_1; (7) HF_Qq_2) (b) Distribution of the segmental duplications (SD) on BTA18 in the ARS-UCD1.2 assembly. Note, that chromosomal segments without any SD were not present in the circular figure. The links within the circle marks the detected SD. At the distal end of the chromosome and in the region between 58-60 Mbp we recognized an increased number ($P<1\times10^{-4}$) of SD compared to the residual chromosome. The red curves represent SD within the CI95% of the QTL (58,343,346-59,432,662 bp).

Table 1. Results of validation of candidate variants using PCR Genotyped breeds are Belgian Blue (BB), Braunvieh (BV), Charolais (CH), Fleckvieh (FV), Gelbvieh (GV), Jersey (JY), Limousin (LM) and Holstein Friesian (HF). Holstein Friesian samples are divided in four groups (Q/Q,Q/q,q/q,N/N); N/N group represented animals that are not used for mapping directly (no the official breeding values) but are ancestors or relatives usefully for derivation of haplotypes and imputation.

rs-ID / Position (bp) / Gene	Genotype*	BB	BV	CH	FV	GV	JY	LM	HF	HF q/q	HF q/Q	HF Q/Q	HF N/N
rs381577268	CC	24	23	9	22	22	17	13	166	141			25
57,816,137	CT				1**				98		86		12
ZNF613	TT								33			31	2
rs381878735	AA	23	23	9	21	22	16	13	141	131			10
59,574,329	AT				1**				86		83		3
ZNF665, ZNF677-like	TT								34			34	
rs464221818	CC	24	23	9	22	22	16	12	146	135			11
59,329,176	CT				1**				86		83		3
ZNF665-like	TT								34			34	
rs472502785	TT	24	23	9	20	22	16	13	142	131			11
59,345,689	TC				1**				86		83		3
ZNF665-like	CC								34			34	

** According to the pedigree records, this Fleckvieh bull contain 1.7% Holstein ancestry.

To avoid the reference bias, we also attempted to *de-novo* assemble the entire BTA18 using ONT sequences, however, the QTL region was marked by the lack of contiguity (Figure 1a) which was more pronounced in the region of QTL. Interestingly, investigation of the QTL region revealed that that this region (between 58 and 60 Mbp) present SV hotspot and contained higher number of SDs (Figure 1b) compared to the remaining chromosome ($P<1\times10^{-4}$).

Discussion

In agreement with the previous studies (Cole *et al.*, 2009: Müller *et al.*, 2017), we identified in HF a major QTL for calving traits at the interval from 58,343,346 to 59,432,662 bp with the peak value located at 58,860,538 bp. Further, the four SNPs that we identified in population wide LD with 'causal haplotype' could be used for monitoring of harmful QTL allele in HF as well as in HF upgraded breeds. Our results also suggested the presence of abundance of SDs in the critical region which could negatively affect the sequencing and *de novo* assembly in this region. Moreover, it is well known that SDs lead to more frequent non-allelic homologous recombination, and this is an essential mechanism for the formation of SV-hotspots leading to pathogenic structural changes (Lin and Gokcumen 2019). Consequently, the QTL allele itself could be an additional large pathogenic SV located in the major SV hotspots of the bovine genome.

References

Bellows D. S., Ott S. L., and Bellows R. (2002). The Professional Animal Scientist 18(1):26-32. https://doi.org/10.15232/S1080-7446(15)31480-7.

Browning B. L., and Browning S. R. (2009). Am J Hum Genet 84(2):210-223. https://doi.org/10.1016/j.ajhg.2009.01.005.

Browning B. L., Zhou Y., and Browning S.R. (2018.) Am J Hum Genet 103(3):338-348. https://doi.org/10.1016/j.ajhg.2018.07.015.

Chen K., Wallis J. W., McLellan M.D., Larson D. E., Kalicki J.M. *et al.* (2009). *Nature methods* 6(9):677–81. https://doi.org/10.1038/nmeth.1363.

Chen X., Schulz-Trieglaff O., Shaw R., Barnes B., Schlesinger F. *et al.* (2016). *Bioinformatics* 32(8):1220–1222. https://doi.org/10.1093/bioinformatics/btv710.

Cole J. B., VanRaden P. M., O'Connell J. R., Van Tassell C. P., Sonstegard T. S. *et al.* (2009). J Dairy Sci 92(6):2931-2946. https://doi.org/10.3168/jds.2008-1762.

Dallery J.-F., Lapalu N., Zampounis A., Pigné S., Luyten I. *et al.* (2017). BMC genomics 18(1):667. https://doi.org/10.1186/s12864-017-4083-x.

Fang L., Jiang J., Li B., Zhou Y., Freebern E. *et al.* (2019). Communications biology 2:100. https://doi.org/10.1038/s42003-019-0341-6.

Gehrke L. J., Capitan A., Scheper C., Konig S., Upadhyay M. *et al.* (2020). Genet Sel Evol52(1):6. https://doi.org/10.1186/s12711-020-0525-z.

Layer R. M., Chiang C., Quinlan R. A., and Hall M. I. (2014). Genome biology 15(6):R84. https://doi.org/10.1186/gb-2014-15-6-r84.

Lin Y. and Gokcumen O. (2019) Genome Biol. Evol. 11(4):1136–1151. https://doi.org/10.1093/gbe/evz058.

Meuwissen T. H., Karlsen A., Lien S., Olsaker I., and Goddard M. E.(2002). Genetics 161(1):373-379. https://doi.org/10.1093/genetics/161.1.373.

Müller M. P., Rothammer S., Seichter D., Russ I., Hinrichs D. *et al.* (2017). J Dairy Sci 100(3):1987-2006. https://doi.org/10.3168/jds.2016-11506.

511. Finding potentially causal genetic factors for meat tenderness using eQTL study in Hanwoo cattle

Y. Chung[1], K. Chung[2], P. Dinh[1], I. Choi[1] and S. Lee[1*]

[1]Chungnam National University, 99, Daehak-ro, Yuseong-gu, 34134, Daejeon, Republic of Korea; [2]Korea National College of Agriculture and Fisheries, 1515, Kongjwipatjwi-ro, Deokjin-gu, 54874, Jeollabuk-do, Republic of Korea; slee46@cnu.ac.kr

Abstract

Meat tenderness is a significant trait influencing eating quality of beef, which could affect consumer satisfaction and preference level. The discovery of causal genetic factors, including regulatory elements on tenderization is needed to improve beef tenderness efficiently. We found 166 putative core genes from an association study between Warner-Bratzler shear force (WBSF) and gene expression. Expression quantitative trait loci (eQTL) analysis with core genes was conducted, and 16 cis-eQTL SNPs for six core genes, *ASAP1*, *CAPN5*, *ELN*, *SUMF2*, *TTC8*, and *MGAT4A*, were identified. The major allele of cis-eQTLs for six core genes was correlated with gene expression, and increased expression level led to meat toughness. In particular, the major allele of rs133370240 in upstream region of *ELN* gene was overlapped with MZF1 binding sites involved with meat tenderization, suggesting that selection of the minor allele of SNPs will be effective in improving beef tenderness.

Introduction

Meat quality is determined by sensory traits including flavor, juiciness, and tenderness. Among them, meat tenderness has a significant influence on eating quality that is related to consumer's perception of beef quality (Koohmaraie *et al.* 2002). Therefore, the improvement of meat tenderness is prioritized on beef cattle. It is practical to utilize the causal factors to improve meat tenderness.

Many studies have been conducted to identify genetic factors such as protein biomarkers or differentially expressed genes (DEGs) (Diniz *et al.* 2019; Leal-Gutiérrez *et al.* 2020). Although functional studies have shown that genetic variants can significantly contribute to regulating gene expression (GE) as well as phenotype (Shi *et al.* 2021), there are limited studies on finding causal genetic variants using omics data.

The objective of this study is to explore causal genetic factors (genes and variants) that explain a path to phenotypic changes caused by regulatory elements.

Materials & methods

Animal data. We used the samples from *longissimus dorsi* muscle of 20 Hanwoo cattle (sire) with 30 months raised in beef feedlots of Hanwoo Experimental Station of National Institute of Animal Science (NIAS), Rural Development Administration (RDA), Republic of Korea (ethics committee approve number: 2015-150).

Association study for detecting core genes. mRNA samples were extracted from *longissimus dorsi* muscle of Hanwoo cattle. We mapped our data to the bovine reference genome (Bos taurus.UMD3.1), and obtained GE normalized using trimmed mean of M-value (TMM) method in edgeR package. An association study between phenotype (WBSF) and GE was conducted to identify core genes (P<0.01). The linear model to find core genes for WBSF is indicated as

$$y = u + e \quad (1)$$

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_511, © Y. Chung *et al.* 2022

where y is a vector of WBSF on 20 Hanwoo cattle, and u is a vector of the normalized GE for i^{th} gene (i = 1, …, n). e is a vector of residual effects. We detected the genes with significant regression coefficient (P<0.01), called as 'core genes'.

eQTL analysis. Single nucleotide polymorphism (SNP) chip data was available from the Illumina Bovine SNP777 BeadChip (777K) platform (UMD3.1), and we used 576,549 SNPs for 20 cattle. The cis-eQTLs were defined as SNPs located within 5 kb of transcription start site (TSS) or transcription termination sites (TTS) of the gene and detected with statistical significance (P<0.01). The linear model used is represented as:

$$y = m + e \quad (2)$$

Where y is a vector of GE of core genes, and m is a vector of the genotype of SNPs (0, 1, 2).

Transcription factor binding site (TFBS) enrichment analysis. To confirm if cis-eQTLs are on transcription factor binding sites (TFBS), we detected TFBSs using TFBSTools R package with JASPAR2020 database. The sequence information of mRNAs was obtained from ENSEMBL database, and one thousand base-pairs (1Kb) upstream from the transcription start site (TSS) were considered TF binding sites. The TFs with the accuracy score more than 0.9 and significantly P-value (Benjamini-Hochberg P<0.05) were selected.

Results

Putative core genes detection. An association study as preliminary analysis was conducted to find putative core genes with significant linear regression coefficient. 166 core genes were detected, and among them, 135 and 31 genes have positive and negative coefficients, respectively.

Cis-eQTLs detection. 16 significant cis-eQTL SNPs modulating GE of six core genes were detected from the association study between WBSF and GE of 166 core genes, *ASAP1* (ArfGAP With SH3 Domain, Ankyrin Repeat And PH Domain 1), *CAPN5* (Calpain 5), *ELN* (Elastin), *SUMF2* (Sulfatase Modifying Factor 2), *TTC8* (Tetratricopeptide Repeat Domain 8), and *MGAT4A* (Alpha-1,3-Mannosyl-Glycoprotein 4-Beta-N-Acetylglucosaminyltransferase A). 13 and 3 variants are in intron and upstream region, respectively (Table 1).

For five genes (*ASAP1*, *ELN*, *CAPN5*, *SUMF2*, and *TTC8*), expression levels tend to decrease in minor allele, leading to a reduction in WBSF, except for *TTC8* (Table 1, Figure 1A and 1B). On the other hand, the expression of *MGAT4A* tends to increase in animals with minor allele, and WBSF level tends to decrease.

TF binding site detection. A total of 26 TFs were significantly detected for six core genes, and among them, 13 TFs were common for all genes; Arnt, Ahr-Arnt, MZF1 (var.2), Pax2, SOX9, Sox17, SRY, Sox5, Hand1-Tcf3, HLTF, Nobox, ZNF354C, and Arid3a. We found that a cis-eQTL SNP, rs133370240 (25:33,825,442), in upstream region of *ELN* gene is located within a TF binding site, MZF1 (Figure 2).

Table 1. Significant cis-eQTLs for six putative core genes.

Gene name	rs id	Major/minor	Chr	Position (bp)	Region	Slope	*P*-value
ASAP1	rs110751858	G/A	14	11,330,635	Intron	-0.31	7.96E-03
	rs135397766	A/G		11,333,062	Intron	-0.31	7.96E-03
ELN	rs111004978	G/A	25	33,824,086	Upstream	-0.61	9.48E-03
	rs133370240	G/A		33,825,442	Upstream	-0.61	9.48E-03
CAPN5	rs41772707	C/A	15	57,261,958	Intron	-0.33	5.78E-03
SUMF2	rs110465445	G/A	25	27,990,277	Upstream	-0.34	1.70E-03
TTC8	rs135354635	A/C	10	101,628,209	Intron	0.26	7.42E-03
	rs136767380	G/A		101,628,784	Intron	0.26	7.42E-03
MGAT4A	rs109601924	G/A	11	3,828,573	Intron	0.29	1.03E-04
	rs110541800	G/A		3,830,575	Intron	0.29	1.03E-04
	rs137222564	A/G		3,832,150	Intron	0.29	1.03E-04
	rs109048556	A/G		3,834,852	Intron	0.17	3.93E-03
	rs136765473	A/G		3,835,362	Intron	0.17	3.93E-03
	rs109687823	A/C		3,833,726	Intron	0.19	3.22E-03
	rs134437911	G/A		3,817,768	Intron	0.26	3.89E-03
	rs134260466	G/A		3,889,038	Intron	0.15	8.54E-03

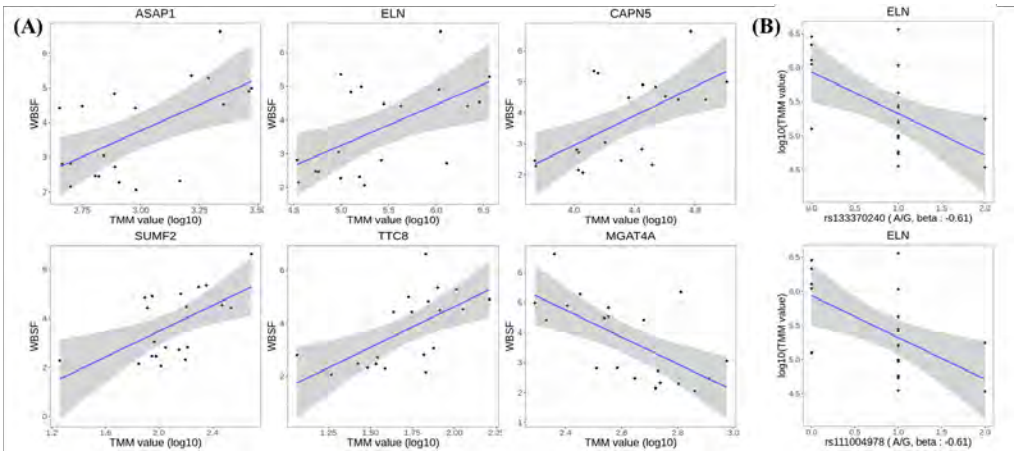


Figure 1. Gene expression and eQTL plot. Figure 1A indicates a change of WBSF depending on GE of six putative core genes, and Figure 1B shows a reduction of *ELN* expression in animals with minor allele of a variant.

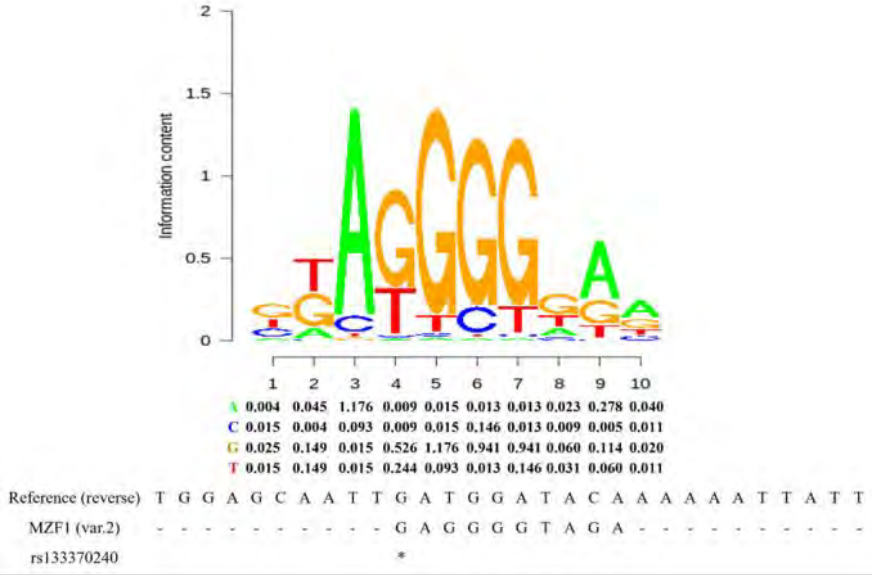


Figure 2. cis-eQTL SNP overlapping MZF1 binding site. Figure 2 indicates rs133370240 of *ELN* gene is in MZF1 binding site. The information content matrix (ICM) represents motif in the biological sequence and contains the weights associated with the occurrence of each nucleotide at the given position in a pattern.

Discussion

Our results showed 6 putative core genes and 16 cis-eQTL SNPs associated with meat tenderness. Major allele of cis-eQTL for *ASAP1*, *CAPN5*, *ELN*, and *SUMF2* was correlated with GE, and increasing GE led to meat toughness. The tendency shows that selection of the minor allele of the variants will be effective in improving beef tenderness.

Interestingly, the function of core genes is related to tenderness or intramuscular fat (IMF). For example, Tizioto *et al.* (2012) reported that a variant located in intron of *ASAP1* was associated with meat quality (tenderness). In terms of IMF, GE of *SUMF2* gene is associated with IMF deposition through glucose metabolism in pig (Ma *et al.* 2014). *ELN* expression is involved in extracellular component mass process for adipogenesis (Cesar *et al.* 2015). Finding genes related to IMF deposition could be due to high-concentration feeding system of Hanwoo cattle (Jo *et al.* 2012).

We identified that a cis-eQTL SNP (rs133370240) in upstream region of *ELN* was overlapped in MZF1 binding site. Compared to the minor allele of this variant, the major allele of this variant has MZF1 binding affinity, affecting GE and beef tenderization (Johnston *et al.* 2019). 13 common TFs including MZF1 for six core genes were found and the TFs have a role for muscle development (Patel *et al.* 2007; McKinnell *et al.* 2008).

In this study, although there are some limits such as a small sample sizes and validation using loss- and gain-of-function, our finding would serve as a basis for further studies to identify causal factors on meat tenderness.

Acknowledgements

This work was supported by Korea Institute of Planning and Evaluation for Technology in Food, Agriculture, Forestry and Fisheries (IPET) funded by Ministry of Agriculture, Food and Rural Affairs (MAFRA) (321082-3).

References

Cesar A.S., Regitano L.C., Koltes J.E., Fritz-Waters E.R., Lanna D.P. *et al.* (2015) PLoS One 10(6)**:** e0128350. https://doi.org/10.1371/journal.pone.0128350

Diniz W.J., Mazzoni G., Coutinho L.L., Banerjee P., Geistlinger L. *et al.* (2019) Frontiers in genetics 10(210. https://doi.org/10.3389/fgene.2019.00210

Jo C., Cho S., Chang J. and Nam K. (2012) Animal frontiers 2(4)**:** 32-38. https://doi.org/10.2527/af.2012-0060

Johnston A.D., Simões-Pires C.A., Thompson T.V., Suzuki M. and Greally J.M. (2019) Nature communications 10(1)**:** 1-16. https://doi.org/10.1038/s41467-019-11412-5

Koohmaraie M., Kent M.P., Shackelford S.D., Veiseth E. and Wheeler T.L. (2002) Meat science 62(3)**:** 345-352. https://doi.org/10.1016/S0309-1740(02)00127-4

Leal-Gutiérrez J.D., Elzo M.A. and Mateescu R.G. (2020) BMC genomics 21(1)**:** 1-15. https://doi.org/10.1186/s12864-020-6520-5

Ma J., Yang J., Zhou L., Ren J., Liu X. *et al.* (2014) PLoS genetics 10(10)**:** e1004710. https://doi.org/10.1371/journal.pgen.1004710

Mckinnell I.W., Ishibashi J., Le Grand F., Punch V.G., Addicks G.C. *et al.* (2008) Nature cell biology 10(1)**:** 77-84. https://doi.org/10.1038/ncb1671

Patel S.R., Kim D., Levitan I. and Dressler G.R. (2007) Developmental cell 13(4)**:** 580-592. https://doi.org/10.1016/j.devcel.2007.09.004

Shi L., Wu X., Yang Y., Ma Z., Lv X. *et al.* (2021) Journal of Animal Science and Biotechnology 12(1)**:** 1-19. https://doi.org/10.1186/s40104-020-00540-4

Tizioto P., Meirelles S., Veneroni G., Tullio R., Rosa A. *et al.* (2012) Meat science 92(4)**:** 855-857. https://doi.org/10.1016/j.meatsci.2012.05.018

512. The application of mixed linear models for the estimation of functional effects on bovine stature based on SNP summary

K. Kotlarz[1], B. Kosinska-Selbi[1], Z. Cai[2], G. Sahana[2] and J. Szyda[1,3*]

[1]Biostatistics Group, Department of Genetics, Wroclaw University of Environmental and Life Sciences, Kozuchowska 7, 51-631 Wroclaw, Poland; [2]Center for Quantitative Genetics and Genomics, Aarhus University, Blichers Allé 20, 8830 Tjele, Denmark [3]National Research Institute of Animal Production, Krakowska 1, 32-083 Balice, Poland; joanna.szyda@upwr.edu.pl

Abstract

Genome-Wide Association Studies (GWAS) help identify polymorphic sites or genes linked to phenotypic variance, but a few genes / SNPs are unlikely to explain the phenotypic variability of complex traits. In the study, the focus was moved from single loci to functional units, expressed by the metabolic pathways (KEGG). Consequently, this study aimed to estimate KEGG effects on stature in three Nordic dairy cattle breeds using SNPs effects from GWAS as the dependent variable. The SNPs were annotated to genes, then the genes to KEGG pathways. The effects of KEGG were estimated separately for each breed using a mixed linear model incorporating the similarity between pathways expressed by common genes. The analysis yielded two conclusions: a single gene in a pathway's component may have a large impact on the pathway's overall effect or the high pathway effect may be due to the influence of many genes constituting the pathway.

Introduction

Genome-Wide Association Studies (GWAS) are very useful for the identification of polymorphic sites, typically Single Nucleotide Polymorphisms (SNPs), or sometimes genes associated with a phenotypic variation or with a disease. Nowadays, the common availability of SNPs obtained based on whole-genome sequencing allows for a very good resolution for the estimation of those associations. However, in the context of phenotypes undergoing a complex mode of inheritance, it is not expected that a few genes / SNPs suffice to explain the variability on a phenotypic level. As a consequence, we often manage to identify loci with a very high effect on the phenotypic variation, but still, a predominant proportion of this variation remains unexplained (Manolio *et al.* 2009), since it is often due to a combined effect of many loci, each with a moderate or small impact. Therefore, in our study, we moved the focus from individual loci to functional units, here expressed by the metabolic (KEGG) pathways, to better understand the physiological mechanisms underlying complex phenotypes. For this purpose, we use SNP summary statistics originating from the GWAS conducted for stature and based on whole-genome sequence data of three Nordic dairy cattle breeds.

Materials & methods

Material. The analysed data comprised SNP summary statistics from GWAS performed on 5,062 Danish Holstein bulls, 924 Danish Red Dairy Cattle bulls, and 2,122 Finnish Red Dairy Cattle bulls (Bouwman *et al.* 2018). The association was calculated for 25.4 million variants imputed with Minimac2 (Fuchsberger *et al.* 2015) from 630,000 SNPs using the 1000 Bull Genomes reference population from Run4, consisting of 1,147 individuals. SNP additive effects were estimated for deregressed EBVs serving as pseudophenotypes, separately for each breed with a single SNP, a mixed linear model including a polygenic effect described by a genomic relationship matrix implemented in the EMMAX software (Kang *et al.* 2010).

Statistical model. Based on their IDs, SNPs were annotated to genes corresponding to the ARS-UCD1.2 reference genome using BIOMART (Smedley *et al.* 2009), next KEGG manually drawn reference pathways (map) were annotated to genes using the David software (Huang *et al.* 2009). The effects of KEGG pathways

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_512, © K. Kotlarz *et al.* 2022

on stature were estimated separately for each breed using the following mixed linear model that accounted for the similarity between pathways:

$$\boldsymbol{y} = \boldsymbol{\mu} + \boldsymbol{Zt} + \boldsymbol{e} \tag{1}$$

where y is the vector of absolute values of SNP additive effects on stature from GWAS, μ represents the general mean, t is the random effect of KEGG pathways with a preimposed normal distribution defined by $N(0, \boldsymbol{V}\sigma_t^2)$, e is a vector of residuals distributed as $N(0, \boldsymbol{I}\sigma_e^2)$, Z is an incidence matrix for t. The similarity between KEGGs i and j, was introduced into the model by incorporating a nondiagonal KEGG covariance matrix V. This covariance was expressed by the Jaccard similarity coefficient:

$$J(i,j) = \frac{M}{N} \tag{2}$$

Where M represents the number of genes shared between KEGG i and j, while N represents the total number of genes involved in KEGG i and j. Variance components were assumed as known, amounting $\sigma_t^2 = 0.3\sigma_y^2$ and $\sigma_e^2 = 1 - \sigma_t^2$. To avoid effect confounding due to high linkage disequilibrium, only one SNP per gene (with the highest effect) was included in y.

Solutions. The mixed model equations (Henderson 1984) were used to obtain solutions for μ and t:

$$\begin{bmatrix} \hat{\boldsymbol{\mu}} \\ \hat{\boldsymbol{t}} \end{bmatrix} = \begin{bmatrix} \boldsymbol{1}^T\boldsymbol{R}^{-1}\boldsymbol{1} & \boldsymbol{1}^T\boldsymbol{R}^{-1}\boldsymbol{Z} \\ \boldsymbol{Z}^T\boldsymbol{R}^{-1}\boldsymbol{1} & \boldsymbol{Z}^T\boldsymbol{R}^{-1}\boldsymbol{Z} + \boldsymbol{G}^{-1} \end{bmatrix}^{-1} \begin{bmatrix} \boldsymbol{1}^T\boldsymbol{R}^{-1}\boldsymbol{y} \\ \boldsymbol{Z}^T\boldsymbol{R}^{-1}\boldsymbol{y} \end{bmatrix}, \text{ where } \boldsymbol{R} = \boldsymbol{I}\hat{\sigma}_e^2 \text{ and } \boldsymbol{G} = \boldsymbol{V}\hat{\sigma}_t^2 \tag{3}$$

To maximise the computational performance of the estimation/prediction process, a custom Python program implementing the NumPy library (Harris *et al.* 2020) was used. Since all calculations were carried out on a high-performance server, the NumPy library was also used to set the array indexing and order what further improved the computing time compared to a native Python application.

Results

Effects of 180 pathways were estimated. Table 1 summarises the top 10 pathways for each breed. Because NumPy and SciPy APIs are implemented with LAPACK and BLAS, which require Fortran memory layout, all input matrices were transformed to Fortran order to avoid costly transpose. In comparison to a fixed matrix input, this approach results in 10 times faster estimation process.

Discussion

While interpreting KEGG pathways effects two scenarios emerge. On the one hand, the overall high effect of a pathway may be driven by a high effect of a single gene that is this pathway's component – a situation that could have been detected in a conventional GWAS. On the other hand, the high pathway effect may be due to the influence of many genes constituting this pathway – a situation that may easily be missed in GWAS due to the small or moderate effects of particular genes. In our study, an emerging pattern is the high effect of pathways associated with the metabolism of xenobiotics (map00363, map00364, map00626, map00642, map00791). Already over 20 years ago Cheriathundam *et al.* (1998) indicated that the Growth hormone regulates the metabolism of enzymes from the cytochrome family which are responsible for the metabolism of a large number of xenobiotics. Furthermore, tetracycline (map00253) induces expression of the SHOX gene, which results in altered human growth (see e.g. Marchini *et al.* 2007). Bile acid synthesis (map00121) disorder in humans is known to influence weight and height (see e.g. Heubi and Setchell 2020).

Table 1. The top 10 KEGG pathways, based on their effect on stature estimated by model (1).

Danish Holstein			Danish Red Dairy Cattle			Finnish Red Dairy Cattle		
KEGG ID	**KEGG name**	**Higher metabolic hierarchy**	**KEGG ID**	**KEGG name**	**Higher metabolic hierarchy**	**KEGG ID**	**KEGG name**	**Higher metabolic hierarchy**
00966	Glucosinolate biosynthesis	Biosynthesis of other secondary metabolites	00253	Tetracycline biosynthesis	Metabolism of terpenoids and polyketides	00253	Tetracycline biosynthesis	Metabolism of terpenoids and polyketides
00642	Ethylbenzene degradation	Xenobiotics biodegradation and metabolism	00642	Ethylbenzene degradation	Xenobiotics biodegradation and metabolism	00945	Stilbenoid, diarylheptanoid and gingerol biosynthesis	Biosynthesis of other secondary metabolites
00333	Prodigiosin biosynthesis	Biosynthesis of other secondary metabolites	00945	Stilbenoid, diarylheptanoid and gingerol biosynthesis	Biosynthesis of other secondary metabolites	01051	Biosynthesis of ansamycins	Metabolism of terpenoids and polyketides
00281	Geraniol degradation	Metabolism of terpenoids and polyketides	01051	Biosynthesis of ansamycins	Metabolism of terpenoids and polyketides	00642	Ethylbenzene degradation	Xenobiotics biodegradation and metabolism
00363	Bisphenol degradation	Xenobiotics biodegradation and metabolism	00364	Fluorobenzoate degradation	Xenobiotics biodegradation and metabolism	00966	Glucosinolate biosynthesis	Biosynthesis of other secondary metabolites
00253	Tetracycline biosynthesis	Metabolism of terpenoids and polyketides	00966	Glucosinolate biosynthesis	Biosynthesis of other secondary metabolites	00364	Fluorobenzoate degradation	Xenobiotics biodegradation and metabolism
00626	Naphthalene degradation	Xenobiotics biodegradation and metabolism	00903	Limonene and pinene degradation	Metabolism of terpenoids and polyketides	04916	Melanogenesis	-
00121	Secondary bile acid biosynthesis	Lipid metabolism	01040	Biosynthesis of unsaturated fatty acids	Lipid metabolism	00903	Limonene and pinene degradation	Metabolism of terpenoids and polyketides
00572	Arabinogalactan biosynthesis	Glycan biosynthesis and metabolism	00571	Lipoarabinomannan biosynthesis	Glycan biosynthesis and metabolism	00791	Atrazine degradation	Xenobiotics biodegradation and metabolism
00650	Butanoate metabolism	Carbohydrate metabolism	04916	Melanogenesis	-	01040	Biosynthesis of unsaturated fatty acids	Lipid metabolism

References

Bouwman A.C., Daetwyler H.D., Chamberlain A.J., Ponce C.H., Sargolzaei M., *et al.* (2018) Nat. Genet. 50(3):362–367.

Cheriathundam E., Doi S.q., Knapp J.R., Jasser M.Z., Kopchick J.J. (1998) Biochem. Pharmacol. 55(9): 1481–1487.

Fuchsberger C., Abecasis G.R., Hinds D.A. (2015) Bioinformatics. 31(5):782–784.

Harris C.R., Millman K.J., van der Walt S.J., Gommers R., Virtanen P., *et al.* (2020) Nature, 585:357–362.

Henderson C.R. (1984) University of Guelph.

Heubi J., Setchell K. (2020) J. Pediatr. Gastroenterol. Nutr. 70(4):423–429.

Huang D.W., Sherman B.T., Lempicki R.A. (2009) Nature Protoc. 4(1):44–57.

Kang H.M., Sul J.H., Service S.K., Zaitlen N.A., Kong S.Y., *et al.* (2010) Nat. Genet. 42(4):348–354.
Manolio T.A., Collins F.S., Cox N.J., Goldstein D.B., Hindorff L.A., *et al.* (2009) Nature. 461(7265):747–753.
Marchini A., Häcker B., Marttila T., Hesse V., Emons J., *et al.* (2007) Hum. Mol. Genet. 16(24):3081–3087.
Smedley D., Haider S., Ballester B., Holland R., London D., *et al.* (2009) BMC Genom. 10:.22

513. Predicted isoforms of a Brahman cow revealed by full-length transcript sequencing

L.T. Nguyen[1*], Y. Cheng[2], R. Kuo[3], B.J. Hayes[1] and E.M. Ross[1]

[1]Centre for Animal Science, Queensland Alliance for Agriculture and Food Innovation, the University of Queensland, QLD 4072, Brisbane, Australia; [2]School of Life and Environmental Sciences, The University of Sydney, Sydney, NSW, Australia; [3]The Roslin Institute and Royal School of Veterinary Studies, University of Edinburgh, EH25 9RG, United Kingdom; t.nguyen3@uq.edu.au

Abstract

Understanding transcriptome complexity is critical in connecting genomic sequence to gene function and phenotype. However, a complete catalogue of transcript isoforms in bovine species is still lacking. This study first processed PacBio isoform sequencing from eight tissues from a *Bos indicus* cow and two fetal tissues. The predicted transcript isoforms were compared to current bovine transcriptome annotation. In total, 149,793 isoforms were obtained from 31,363 genic loci, reporting an average of 4.7 transcripts per gene. Additionally, a large proportion (69%) of predicted transcripts were identified as novel isoforms, providing evidence and valuable datasets for improving the annotation of the bovine reference genome, which could be incorporated into genomic prediction models using appropriate methods to improve prediction accuracy.

Introduction

Precise gene annotations are essential to understanding the complexity of a species transcriptome and connecting genomic sequence to gene function and ultimately phenotype. To date, bovine transcriptome annotations are mostly based on short-read RNA sequencing platforms. However, short-read sequencing faces the challenge of lengthy transcripts, repetitive regions, and transposable elements. Long read sequencing technologies, represented by PacBio Sequencing (McCarthy, 2010; Rhoads and Au, 2015) and Oxford Nanopore Technologies (Loman and Watson, 2015; Bayega *et al.*, 2018), have overwhelmed this challenge by sequencing of full-length transcripts. In these recent years, PacBio single-molecule real-time (SMRT) isoform sequencing (ISO-seq) has been utilised to improve transcriptome annotations in humans (Tilgner *et al.* 2014), pig (Beiki *et al.* 2019), rabbit (Chen *et al.* 2017), chickens (Thomas *et al.* 2014) and Hereford cattle (Rosen *et al.* 2020; Halstead *et al.* 2021).

As a first step towards cataloguing *Bos indicus* transcript isoforms, this study analysed PacBio isoform sequence data from ten tissues from a *B. indicus* cow (Blood, Liver, Thyroid, Spleen, Kidney, Ovary, Uterus and Muscle) and two fetal tissues (Liver and Lung) using poly(A) selected PacBio Isoseq technology. The predicted transcript isoforms were then compared to current bovine transcriptome annotation.

Materials & methods

Animals and phenotypes. Animal use in this study was approved under animal ethics number QAAFI/270/17.

Sample collection and RNA extraction. The transcriptomes of eight tissues (whole blood, kidney, liver, muscle, spleen, ovary, uterus, and thyroid) from a Brahman cow and two tissues (liver and lung) from a fetus of the studied Brahman cow were sequenced by the Ramaciotti Centre for Genomics (UNSW Sydney, Australia) using PacBio Sequel system. Total RNA from each tissue was extracted using mirVAna miRNA isolation kit (Ambion) following the manufacturer's instruction. RNA purity was evaluated with a Nanodrop ND-1000 spectrophotometer (v.3.5.2, Thermo Fisher Scientific). QubitTM 4.0 Fluorometer with

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_513, © L.T. Nguyen *et al.* 2022

Qubit RNA BR (Broad-range) assay kit (Thermo Fisher Scientific) was used to quantify RNA concentration. The assessment of RNA integrity was performed using Agilent 2100 Bioanalyser (Agilent Technologies). Only RNA with an integrity number greater than 8.0 was used for library preparation.

cDNA synthesis and sequencing. The RNA samples were converted to cDNA using the Lexogen TeloPrime kit (Lexogen) following the manufacture protocol. The cDNA libraries were sent to the Ramaciotti Centre for Genomics (UNSW Sydney, Australia) for sequencing on V3.0 sequencing chemistry using the PacBio Sequel system with 10-hour movies.

Sequencing data processing. The circular consensus sequences for each ZMW generated from each ISOseq library were only kept if they had at least two full-length reads longer than 100bp and predicted accuracy of 0.99. These ccs then went through the adapter removing, and poly(A) tails trimming steps to create full-length non-concatemer (FLNC) reads for further analysis.

Long read transcriptome annotation processing. Full length non-concatemer reads were mapped to the *Bos taurus* reference genome (ARS-UCD1.2) using Minimap2 (Li 2018) with parameters '*--secondary=no -ax splice -uf -C5*'. After mapping, resulting sam files from each tissue were then processed using TAMA collapse (Kuo *et al.* 2020) with parameters '*-x capped -a 100 -z 100 -sj sj_priority -lde 3 -sjt 10*'. The collapsed transcripts from different tissues were then merged using TAMA merge (Kuo *et al.* 2020) with parameters '*-a 300 -z 300 -m 30 -d merge_dup*' to create an Isoseq based transcriptome annotation.

Comparison of transcript structures across current NCBI RefSeq annotation (release 106). We used gffcompare v0.12.6 (Pertea *et al.* 2016) to compare NCBI RefSeq annotation gtf file to Isoseq transcriptome annotation file. Based on gffcompare class codes, predicted transcripts were classified as known isoforms (class code '='), novel isoforms (class codes 'c', 'k', 'j', 'm', 'n', or 'o'), novel loci (class code 'i', 'u', 'y' or 'x') or potential artifacts (class codes 'e', 's' or 'p').

Results

Across tissue, a total of approximately 5.0 M Iso-seq reads were observed, with an average of 502,712 Iso-seq reads (Table 1). The longest reads were observed from the Kidney library (length = 10.3 kb). Using the TAMA pipeline, this approach identified a total of 149,793 transcript isoforms at 31,363 genic loci, revealing an average of 4.7 transcripts per gene (tpg). The highest number of transcript isoforms was identified in the uterus tissue, followed by blood tissue, agreeing with the animal physiological stage (pregnant cow). Comparing predicted transcript isoforms with known transcripts in the current bovine genome annotation (NCBI RefSeq release 106) resulted in a total of 23,514 annotated transcripts exactly matching previously annotated transcripts (15.7% of all transcripts, Figure 1). The remaining predicted transcripts were classed into novel isoforms (69.6%), novel loci (12.7%) and potential artifacts (2%).

Discussion

This study revealed predicted transcript isoforms from a Brahman cow using long-read PacBio Isoseq technology. On average, 4.7 transcripts per gene (tpg) were identified closer to the ratio reported in human genome NCBI annotation (4.2tpg, NCBI release 109), one of the most complete transcriptome annotations. Compared to the bovine annotation, this number was higher than bovine NCBI annotation release 106 (1.59 tpg) and Oxford Nanopore Technologies bovine annotation (3.77 tpg). Out of 149,793 predicted transcript isoforms, over 69% were novel compared to current bovine NCBI annotation, revealing the complexity of the bovine transcriptome. A high percentage of novel isoforms has also been identified in other transcriptome annotation studies. For example, the percentage of novel transcript isoforms from improved pig and bovine transcriptome annotations were accounted for 80 and 60%, respectively (Beiki *et al.* 2019; Halstead *et al.* 2021).

Table 1. Summary of full length reads, genes and isoforms identified across tissues.

Sample	CCS[1] (%)	Full-length reads (%)	Mapped reads (%)	Mean / max. length (bp)	Genes / isoforms
Blood	456,914 (79.47)	424,389 (92.88)	422,078 (99.46)	1,651/6,696	11,270/40,311
Kidney	412,798 (79.16)	384,185 (93.06)	384,150 (99.99)	2,058/10,305	8,651/30,791
Liver	378,461 (82.36)	350,830 (92.69)	350,751 (99.98)	1,441/6,284	7,538/20,288
Muscle	249,806 (79.06)	217,085 (86.90)	217,017 (99.97)	1,377/5,508	5,843/16,933
Thyroid	448,207 (78.52)	418,889 (93.46)	418,872 (99.99)	2,144/7,474	8,865/35,238
Spleen	416,450 (85.63)	388,072 (93.18)	388,056 (99.99)	1,952/8,202	9,638/34,062
Uterus	437,153 (82.44)	405,460 (92.75)	405,366 (99.98)	1,622/10,096	16,659/46,311
Ovary	412,081 (83.99)	373,082 (90.54)	373,005 (99.98)	1,383/4,557	8,715/23,985
Fetal lung	416,226 (80.15)	406,425 (97.64)	406,403 (99.99)	1,850/6,997	8,313/34,862
Fetal liver	326,736 (58.57)	212,534 (65.05)	212,453 (99.96)	1,564/6,216	7,252/13,469
Total					31,363/149,793

[1] CCS means circular consensus sequence.

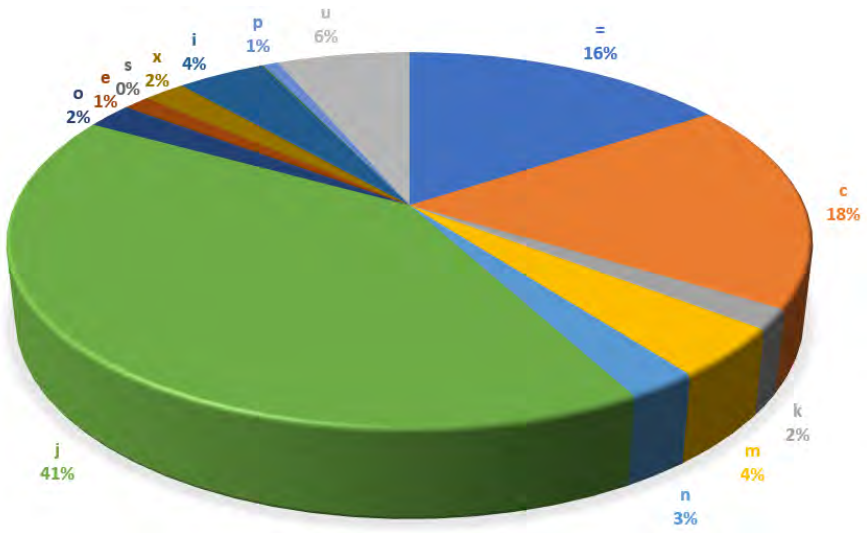


Figure 1. Comparison of predicted transcripts structure with known transcripts in NCBI genome annotation (release 106). Predicted transcripts were classified as known isoforms (class code '='), novel isoforms (class codes 'c', 'k', 'j', 'm', 'n', or 'o'), novel loci (class code 'i', 'u', 'y' or 'x') or potential artifacts (class codes 'e', 's' or 'p').

Compared to the current bovine transcriptome, this dataset has two significant differences. First, we sequenced each tissue in each SMRT cell, making our dataset valuable for tissue-specific transcripts. Second, this is the first Isoseq dataset generated for a *Bos indicus* breed (Brahman), whereas all current bovine transcriptome annotations were built from a *Bos taurus* breed (Hereford). This data will contribute to the improvement of the bovine transcriptome annotation and provides an opportunity for future comparative and evolutionary genomics studies. Also, an improved bovine transcriptome annotation for *Bos indicus* species could be useful for incorporating into genomic prediction models to improve prediction accuracy, as suggested by MacLeod *et al.* 2016 and Xu *et al.* 2020.

References

Bayega, A., Fahiminiya, S., Oikonomopoulos, S. And Ragoussis, J. (2018) Gene Expression Analysis: Methods and Protocols. Springer New York, New York, NY.

Beiki, H, Liu, H, Huang, J, Manchanda, N, Nonneman, D *et al.* (2019) *BMC Genomics* 20, 344. https://doi.org/10.1186/s12864-019-5709-y

Chen, S-Y, Deng, F, Jia, X, Li, C, Lai, S-J. (2017) *Scientific Reports* 7, 7648. https://doi.org/10.1038/s41598-017-08138-z

Halstead, MM, Islas-Trejo, A, Goszczynski, DE, Medrano, JF, Zhou, H, Ross, PJ. (2021) *Front Genet* 12 (621). https://doi.org/10.3389/fgene.2021.664260

Kuo, RI, Cheng, Y, Zhang, R, Brown, JWS, Smith, J, *et al.* (2020) *BMC Genomics* 21, 751. https://doi.org/10.1186/s12864-020-07123-7

Li, HJB (2018). *Bioinformatics* 34, 3094-3100. https://doi.org/10.1093/bioinformatics/bty191

Loman, N.J. and Watson, M. (2015) *Nat. Methods* 12(4):303-304. https://doi.org/10.1038/nmeth.3327

MacLeod, IM, Bowman, PJ, Vander Jagt, CJ, Haile-Mariam, M, Kemper, KE, *et al.* (2016) *Nat. Protoc* 11(9), 1650-1667. https://doi.org/10.1038/nprot.2016.095

McCarthy, A. 2010. *Chem. Biol* 17(7):675-676. https://doi.org/10.1016/j.chembiol.2010.07.004

Rhoads, A. and Au, K.F. (2015) *Genomics, Proteomics & Bioinformatics* 13(5): 278-289. https://doi.org/10.1016/j.gpb.2015.08.002

Rosen, BD, Bickhart, DM, Schnabel, RD, Koren, S, Elsik, CG, *et al.* (2020) *GigaScience* 9, giaa021. https://doi.org/10.1093/gigascience/giaa021

Thomas, S, Underwood, JG, Tseng, E, Holloway, AK, (2014). *PLoS One* 9, e94650. https://doi.org/10.1371/journal.pone.0094650

Tilgner, H, Grubert, F, Sharon, D, Snyder, MP (2014) *Proc. Natl. Acad. Sci. U. S. A.* 111, 9869. https://doi.org/10.1073/pnas.1400447111

Xu, L, Gao, N, Wang, Z, Xu, L, Liu, Y, *et al.* (2020). *Front Genet* 11, 481-481. https://doi.org/10.3389/fgene.2020.00481

514. Meta-analysis of genome-wide association studies for litter size in pigs indicates new causal variants

E. Sell-Kubiak[1*], M.F.L. Derks[2], M.S. Lopes[2,3], J. Dobrzański[1] and T. Szwaczkowski[1]

[1]Poznań University of Life Sciences, Department of Genetics and Animal Breeding, Wołyńska 33, 60-637 Poznań, Poland; [2]Topigs Norsvin Research Centre, Schoenaker 6, 6641 SZ Beuningen, the Netherlands; [3]Topigs Norsvin, R. Visc. do Rio Branco, 1310 Centro, Curitiba, PR, 80420-210, Brazil; ewa.sell-kubiak@puls.edu.pl

Abstract

The aim of this study was the meta-analysis (MA) of five GWAS for total number born (TNB) coming from five pig populations. MA was performed in METAL based on weighted Z-scores. As a result, two new SNPs were discovered that were not reported before as associated with TNB. First, on *Sus scrofa* chromosome (SSC) 14 located in intron of *FAM13C* gene. Second, on SSC9 located within intron of *AGMO gene*. pCADD analysis revealed a strong candidate causal gene underlying the QTL on SSC9. The third best hit and the most promising candidate gene for litter size was found within the *SOSTDC1* gene associated with lower male fertility in rats.

Introduction

Since the discovery of the first genomic association between litter size and estrogen receptor in 1996 (Rothschild *et al.*, 1996), there were high hopes in finding more genomic regions for reproduction traits. Currently, one of the most popular methods for uncovering the associations between traits and genes are genome-wide association studies (GWAS) using SNP BeadChip technology. In Pig QTL database already 1,625 QTLs are reported for different litter traits. Among them at least 356 are associated with total number born (TNB).

One of the main approaches to integrate and pool the results from GWAS is a meta-analysis (MA). MA could be used to increase the power of association studies combining datasets from different sources and to reduce false-positive associations (e.g. Sauvant *et al.*, 2020). Therefore, the aim of this study was to combine the full GWAS results for TNB from five studies and to perform MA in hopes of detecting new associations.

Materials & methods

The data for this study came from five full GWAS generated by: Uimari *et al.*, 2011, Sell-Kubiak *et al.*, 2015, Ma *et al.*, 2018, Balogh *et al.*, 2019, Zhang *et al.*, 2019. Detailed description of each study is presented in Table 1. The final dataset used for MA consisted of 63,531 SNPs.

Meta-analysis. METAL software (Willer *et al.*, 2010) was used to performed MA based on weighted Z-scores. This approach considers the *P*-value, direction of effect, and number of individuals present in each study. The method is most preferable when MA combines GWAS for differently defined phenotypes (Van den Berg *et al.* 2016) as in this study. The Bonferroni correction (less severe than using a threshold of $-\log_{10}(P\text{-value}) \geq 6$) was applied to established statistically significant associations. Post MA the search for candidate genes was done with Ensembl BioMart.

Candidate genes and causal variants. To assess possible candidate genes and variants underlying GWAS peaks from MA also the pCADD pipeline (Derks *et al.*, 2021) was used. The pipeline evaluates the presence of identified variants in a whole genome sequence (WGS) dataset in four elite breeds from Topigs Norsvin (WGS data description in Derks *et al.,* 2021). In short, we tested whether there are sequence variants in LD with the significant variant from the MA, and whether causal variants are likely to be identified. The

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_514, © E. Sell-Kubiak *et al.* 2022

Table 1. Overview of phenotypic and genomic data available for meta-analysis of GWAS results for total number born.

Source of data	**Uimari *et al.***	**Sell-Kubiak *et al.***	**Ma *et al.***	**Zhang *et al.***	**Balogh *et al.***
Population	Finnish Landrace	Large White	Erhualian	Duroc	Hungarian Large White
Individuals	328	2,351	48	1,067	290
Phenotype	deEBV[1]	deEBV[1]	EBV[2]	deEBV[1]	TNB
BeadChip	Porcine SNP60 BeadChip				
Available SNPs	57,868	40,969	28,020	32,147	56,592
Method	single-SNP mixed model with pedigree relationship matrix	single-SNP mixed model with genomic relationship matrix	single-SNP mixed model with pedigree relationship matrix	single-SNP mixed model with genomic relationship matrix	multi-SNP mixed model with genomic relationship matrix
Detected SNPs	5	0	5	7	3
Threshold	*P*-value $\leq 2.0*10^{-6,3}$	$-\log_{10}(P\text{-value}) \geq 5$	*P*-value $\leq 1.78*10^{-6,3}$	$-\log_{10}(P\text{-value}) \geq 4$	$-\log_{10}(P\text{-value}) \geq 5$

[1] deEBV i.e. deregressed breeding values of total number born; used to remove parents average.

[2] EBV i.e. direct breeding values of total number born.

[3] Bonferroni corrected *P*-value.

candidate variants in LD with the top SNP are ranked according to their pCADD score. The pCADD provides a per base impact score (Gross *et al.*, 2020) to distinguish between variants that likely have an impact (positive or negative) and variants that are benign. The pCADD pipeline additionally provides gene annotation and functional genomic information to further finemap the QTL region.

Results

Two SNPs were found to be significantly associated with TNB. First, *rs81300422* located on SSC9:84696142, which was, unfortunately, present only in one population of Finnish Landrace pigs (Uimari *et al.*, 2011). According to Ensembl Sscrofa11.1, it is located within an intron of a gene *AGMO*. Second, *rs80945731* located on *Sus scrofa* chromosome (SSC) 14:62633073, which was genotyped in four out of five populations (Uimari *et al.*, 2011; Ma *et al.*, 2018; Balogh *et al.*, 2019; Zhang *et al.*, 2019). Based on Ensembl Sscrofa11.1, this SNP is located in the intron region of the *FAM13C* gene.

Candidate genes and causal variants. To assess candidate causal genes at the two QTL loci we ran the pCADD-GWAS pipeline in four different breeds from Topigs Norsvin (Derks *et al.*, 2021). The *rs81300422* SNP on SSC9 was only present in a synthetic boar line (with 7% allele frequency). The third best hit with the pCADD pipeline (after two intergenic SNPs) is upstream of the *SOSTDC1* gene.

The *rs80945731* SNP on SSC14 was common in all four commercial breeding populations. The results yield SNPs within and close to the *PHYHIPL* and *FAM13C* genes. Thus, partially overlapping with the results of candidate gene search applying the classical approach.

Discussion

The meta-analysis of GWAS results coming from five different populations indicated two new SNPs associated with litter size. Even though our MA produced far less significant results than each of the GWAS separately this is the first time those SNPs are reported as associated with litter size.

Candidate genes on SSC9. The first of the SNPs *rs81300422* (genotyped only in one population used in MA) is located in the intron of not very well studied swine gene *AGMO* encoding enzyme Alkylglycerol monooxygenase. It is involved in the degradation of ether lipids (Watschinger and Werner, 2013) and is the only enzyme that can breakdown alkylglycerols and lysoalkylglycerophospholipids (Sailer *et al.*, 2021). *AGMO* was shown to be associated with many disorders in humans (Okur *et al.*, 2019), e.g. autism (Sebat *et al.*, 2007), type 2 diabetes (e.g. Fontaine-Bisson *et al.*, 2010) and immune defense (Sailer *et al.*, 2021). However, no link with fertility was reported.

Thus, far more interesting and promising as a candidate gene is *SOSTDC1*, indicated by pCADD analysis, which has been found to affect fertility in male rats (Pradhan *et al.*, 2019). *SOSTDC1* is a negative regulator of spermatogenesis, and downregulation of this gene during puberty is essential for quantitatively and qualitatively normal spermatogenesis governing male fertility. Further analysis comparing litter size between pigs with different genotype for *rs81300422* confirmed that the detected SNP is affecting TNB in pig population available for pCADD.

Candidate genes on SSC14. The second of the significant SNPs *rs80945731* (present in four populations used in MA) is located within a gene *FAM13C*, which was also confirmed by pCADD analysis. However, it does not have an immediate link with fertility in any specie. Nonetheless, it was expressed in swine adipose tissue and tissues of the nervous system (e.g. amygdala or medulla oblongata). In humans, *FAM13C* was shown to be associated with endometrial mixed adenocarcinoma (Stelzer *et al.*, 2016) as well as it has been used as a marker for prostate cancer (Burdelski *et al.*, 2017) or rectum adenocarcinoma survival predictor (Huang *et al.*, 2019). The second gene indicated by pCADD for *rs80945731* was *PHYHIPL* that is differentially expressed between early, full, and extended blastocysts in bovine IVP embryos (Georges *et al.*, 2019). Both *FAM13C* and *PHYHIPL* in humans were expressed (among others) in the tissue of the ovaries, uterus as well as prostate and testis (Stelzer *et al.*, 2016).

Conclusions

Presented results indicate that MA based on GWAS from five populations provided new insights into discovering genomic background of the litter size in pigs. Moreover, pCADD analysis brought us even closer to finding causal variants for TNB. Further studies should focus on *SOSTDC1* gene as its confirmed function has the strongest link with fertility.

Financial support

This project was financed by the National Science Centre, Poland (NCN SONATA grant no. 2016/23/D/NZ9/00029; ESK is the grant holder).

Acknowledgements

The authors would like to thank five scientific groups that generously provided GWAS results from their studies for meta-analysis: National Engineering Research Center for Breeding Swine Industry, and Guangdong Provincial Key Lab of Agro-Animal Genomics and Molecular Breeding, College of Animal Science, South China Agricultural University (China); Institute of Swine Science, Nanjing Agricultural University (China); Agrifood Research Finland, MTT, Biotechnology and Food Research (Finland); Topigs Norsvin Research Centre, Beuningen (the Netherlands); and NARIC Research Institute for Animal Breeding, Nutrition and Meat Science (Hungary).

References

Balogh E.E., Gábor G., Bodó S., Rózsa L., Rátky J. *et al.* (2019) Acta Vet Hung 67:256–273. https://doi.org/10.1556/004.2019.027.

Burdelski C., Borcherding L., Kluth M., Hube-Magg C., Melling N. *et al.* (2017) Oncotarget 8:31494–31508. https://doi.org/10.18632/oncotarget.16357.

Derks M.F.L., Gross C., Lopes M.S., Reinders M.J.T. *et al.* (2021) Genomics 113(4):2229-2239. https://doi.org/10.1016/j.ygeno.2021.05.017.

Fontaine-Bisson B., Renström F., Rolandsson O., Payne F., Hallmans G. *et al.* (2010) Diabetologia 53:2155–2162. https://doi.org/10.1007/s00125-010-1792-y.

Georges H., Bishop J., Van Campen H., Barfield J., and Hansen T. (2020) Reprod Fert Develop 32(2):177. https://doi.org/10.1071/RDv32n2Ab102.

Huang H., Cao J., Guo G., Li X., Wang Y. *et al.* (2019) J Anim Sci 97:1133–1142. https://doi.org/10.1093/jas/skz031.

Ma X., Li P.H., Zhu M.X., He L.C., Sui S.P. *et al.* (2018) Animal 12:2453–2461. https://doi.org/10.1017/S1751731118000332.

Okur V., Watschinger K., Niyazov D., McCarrier J., Basel D. *et al.* (2019) Hum Genet 138:1259–1266. https://doi.org/10.1007/s00439-019-02065-x.

Pradhan B.S., Bhattacharya, I., Sarkar R., and Majumdar S. (2019) Sci Rep 9(1):1-11. https://doi.org/10.1038/s41598-019-47930-x.

Rothschild M., Jacobson C., Vaske D., Tuggle C., Wang L. *et al.* (1996) Proc Natl Acad Sci USA 93:201–205. https://doi.org/10.1073/pnas.93.1.201.

Sailer S., Keller M.A., Werner E.R., and Watschinger K. (2021) Life 11:1–15. https://doi.org/10.3390/life11020088.

Sauvant, D., M. P. Letourneau-Montminy, P. Schmidely, M. Boval, C. Loncke, *et al.* (2020) Animal 14:s207–s222. https://doi.org/10.1017/S1751731120001688.

Sebat, J., B. Lakshmi, D. Malhotra, J. Troge, C. Lese-Martin, *et al.* (2007) Science 316:445–449. https://doi.org/10.1126/science.1138659

Sell-Kubiak E., Duijvesteijn N., Lopes M.S.,. Janss L.L.G, Knol E.F. *et al.* (2015) BMC Genomics 16:1049. https://doi.org/10.1186/s12864-015-2273-y.

Stelzer G., Rosen R., Plaschkes I., Zimmerman S., Twik M. *et al.* (2016) Curr Protoc Bioinforma 54:1.30.1-1.30.33. https://doi.org/10.1002/cpbi.5.

Uimari P., Sironen A., and Sevón-Aimonen M.L. (2011) Genet Sel Evol 43:42. https://doi.org/10.1186/1297-9686-43-42.

van den Berg I., Boichard D., and Lund M.S. (2016) J Dairy Sci 99:8932–8945. https://doi.org/10.3168/jds.2016-11073.

Watschinger K., andWerner E.R. (2013) Life 65:366–372. https://doi.org/10.1002/iub.1143.

Willer C.J., Li Y., and Abecasis G.R. (2010) Bioinformatics 26:2190–2191. https://doi.org/10.1093/bioinformatics/btq340.

Zhang Z., Chen Z., Ye S., He Y., Huang S. *et al.* 2019. Animals 9. https://doi.org/10.3390/ani9100732.

515. Selection signatures analyses revealed genes associated with adaptation in selected goat breeds in Kenya

R.W. Waineina[1,2*], T.O. Okeno[1], E.D. Ilatsia[2] and K. Ngeno[1]

[1]Department of Animal Sciences, Animal Breeding and Genomics Group, Egerton University, P.O. Box 536-20115, Egerton, Kenya; [2]Dairy Research Institute, Kenya Agricultural and Livestock Organization, P.O. Box 25-20117, Naivasha, Kenya; waineinaruth@gmail.com

Abstract

Artificial and natural selection in livestock is expected to leave unique footprints on their genomes. Goat breeds in Kenya have evolved for survival, breeding and production in various harsh ecological areas, and their genomes are likely to have acquired unique alleles for adaptation to such diverse production environments. To investigate signals of selection, we applied Integrated Haplotype Score (iHS) methods to 50K SNP genotype data from four selected goat breeds in Kenya. Following the analyses, a total of 171 candidate selection regions were detected in the studied breeds. Annotations of these regions yielded interesting genes potentially under strong positive selection including *SYT11, CCDC71, ALG14* and *UBQLN4* that are associated with multiple biological pathways underlying adaptation processes. Our results provide knowledge for the conservation and utilization of the goat genetic resources.

Introduction

Goats are spread all over the world, Africa included, following the pattern of human migrations with their livestock and trade routes (Luikart *et al.* 2001; Taberlet *et al.* 2008). In the process of migrations, they adapted to a wide range of various geographical conditions in the mountains to the desert and tropical regions of Africa. Goat populations were managed differently, with both natural and artificial selection playing a major role in their genetic variation due to the allowance of important gene flows among them, thus maintaining a wide range of genetic diversity (Taberlet *et al.* 2008). In the process of domestication, selection and breed formation, some regions within the goat genomes may have developed special patterns to mirror such development.

In Kenya, the indigenous goats (for example, Galla and small East Africa) and the crossbred of exotic and indigenous are known to be well adapted in various ecological areas (Ndeke *et al.* 2015; Kiura *et al.* 2020). It can therefore be assumed that these goats have evolved for survival, breeding and production in various ecological areas and their genome has footprints for adaptation to such environments and other traits of economic importance.

In goats, selection signature analyses using genome-wide SNPs have been done (Brito *et al.* 2017; Onzima *et al.* 2018). The approach has identified several putative genomic candidate regions and genes under selection that influenced some traits such as, thermo-tolerance, nervous and autoimmune response among others,

In this study, we used iHS method to identify footprints of selection. The objective of this study was to identify distinctive selection signatures in the genome of four breeds; Galla, Alpine, Saanen and Toggenburg goat breeds in Kenya.

Materials & methods

Animals and genotype control. Data used was obtained from 96 samples across four goat breeds; Galla (n=12), Alpine (n=29), Saanen (n=24) and Toggenburg (n=31). Animals were genotyped with the Illumina GoatSNP50 BeadChip (Tosser-Klopp *et al.* 2014) containing 53,347 single nucleotide polymorphisms

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_515, © R.W. Waineina *et al.* 2022

(SNPs). Quality control (QC) on genotype data were performed using PLINK v1.90 (Chang *et al.* 2015). The genotyping rate was 0.98. Out of the 96 samples, two samples failed due to missing genotype data (--mind 0.05); 1,079 variants were removed due to missing genotype data (--geno 0.1); 527 variants were removed due to Hardy-Weinberg exact test (--hwe 0.001), and 1,305 variants removed due to minor allele frequency lower than 0.05 threshold (--maf 0.05).

Identifying selection signatures. An iHS was estimated using *Fast*PHASE and performed using the *rehh* package (Gautier and Vitalis 2012) in R v. 3.4.4. Following analyses, the |iHS| scores were transformed into -log 10 [1 - 2| Φ iHS - 0.5|] in which Φ iHS is the cumulative Gaussian distribution function for better visualization and assessment of selection signals. According to Gautier and Naves (2011), and assuming a normal distribution, |iHS| scores can be interpreted as log10(1/P), where P is a one-sided *P*-value associated with the neutral hypothesis of no selection. In this study, all the SNPs ranking above 0.1 percentile of the |iHS| distribution were selected as candidates for selection.

Candidate genes associated with selection signature identification. Based on the above results, for all the SNPs passing the top 0.1% percentile threshold, we extracted goat genes mapping to 10 kb up-downstream of each SNP using BioMart tool based on the bovine reference genome assembly (ARS-UCD1.2).

Results

In Figure 1, the principal component 1, which accounts for 15.2% of the total variance separated Galla breed from the other three breeds. The principal component 2 accounts for 14.1% of the total variance, split the goat breeds into four clusters 1.

Figure 2 shows the genome-wide distribution of iHS values across the genome for each goat breed. A strong selective sweep was observed on chromosome 3 for both Galla and Saanen breeds.

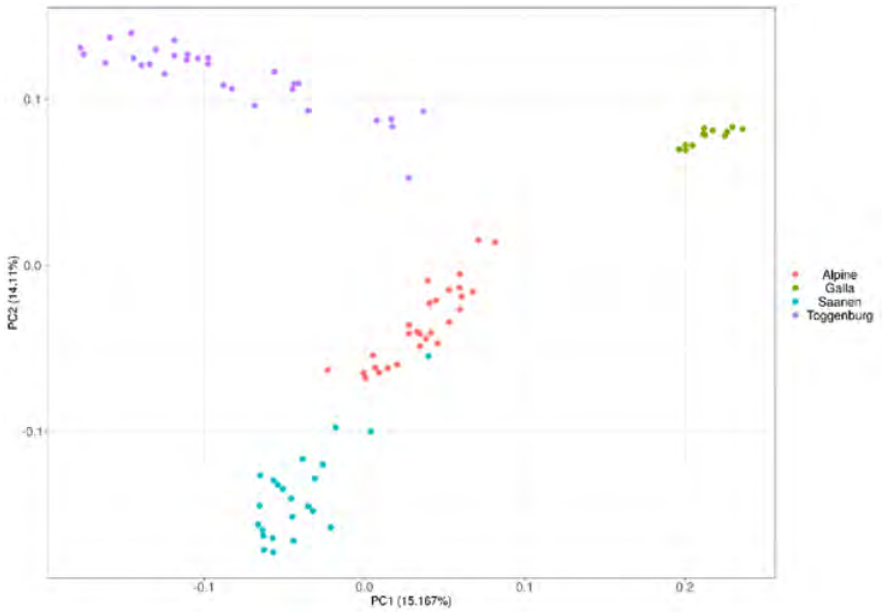


Figure 1. Principal components analysis plot based on SNP array data of goat breeds.

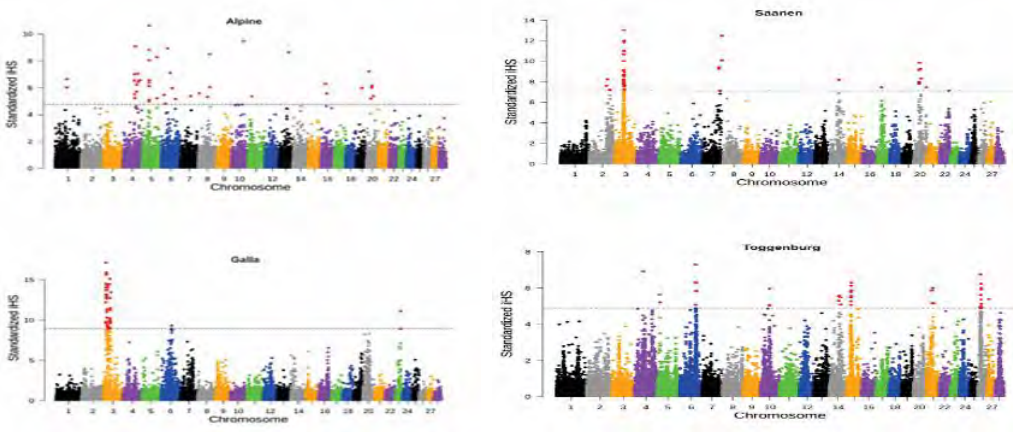


Figure 2. Genome-wide distribution of iHS values for Alpine, Galla, Saanen and Toggenburg goat breeds. The dashed line represents the top 0.01 percentile distribution of iHs values.

A total of 23, 28, 22 and 24 candidate genes overlapped with significant iHS genomic regions detected in Alpine, Galla, Saanen and Toggenburg goat, respectively. Some of the genes that overlapped with the topmost significant iHS genomic regions per breed include; *ALG14, ABCA4* in Saanen breed on Chr3(48504005 – 51372381), *SYT11,UBQLN4* on Chr3(14002666 – 46355720) in Galla breed.

Discussion

In this study, several top putative selective sweeps regions were detected by iHS analyses. A strong signal with the highest number of top candidates sweep regions was detected in Saanen and Galla breeds on chromosomes 3. The reasons why strong signals were detected in chromosomes 3 for the two breeds might be that their alleles have almost reached the point of fixation, this is in agreement with Waineina *et al.* (2021) findings that the two breeds were less admixed compared to the other breeds and are managed in seclusion from other breeds in this study.

Using the top SNPs identified, on chromosome 5 in Alpine breed, *CPNE8* was mapped, which is believed to cause Arthrogryposis multiplex congenita (*AMC*). Overexpression of *UBQLN4* gene, which has been detected in our study in Galla breed, has been associated with poor disease outcomes in various cancer diseases. Copine 8 *(CPNE8)* is a top candidate gene located on chromosome 5 in the Alpine breed which is a candidate quantitative trait gene for prion disease and incubation time in the mouse. The Synaptotagmin-11 *(SYT11)* gene is associated with pains in goats (Deng *et al.* 2018) which was detected on chromosome 1 in the Galla breed. *ABCA4* gene was detected in the Saanen breed, its mutations are responsible for Stargardt mascular degeneration, and related retinal degenerative diseases that cause significant loss in vision (Molday *et al.* 2009). The *CCDC71* is involved in innate and adaptive immunity in mammals (Gautier *et al.* 2009), while *ALG14* conditional mutants are sensitive to heat with reduced function mutants showing increased competitive fitness.

Conclusions

In this study, we investigated the signatures of selection of the four goat breeds in Kenya. The study identified several putative genomics regions. We found that *SYT11, CCDC71, ALG14, UBQLN4* genes may play an important role in underlying adaption to various environmental conditions. Our results provide knowledge for the conservation and utilization of the goat genetic resources in Kenya.

References

Bolormaa, S., *et al.* (2011) Journal of animal science 89(7): 2050-2060. https://doi.org/10.2527/jas.2010-3363

Brito, Luiz F., *et al.* (2017) BMC genomics 18(1): 1-20. https://doi.org/10.1186/s12864-017-3610-0

Burren, Alexander, *et al.* (2016) Animal genetics 47(6): 727-739. https://doi.org/10.1111/age.12476

Chang, Christopher C., *et al.* (2015) Gigascience 4(1): s13742-015. https://doi.org/10.1186/s13742-015-0047-8

Deng, Xiuling., *et al.* (2018) Biological research 51. http://dx.doi.org/10.1186/s40659-018-0174-7

Gautier, Mathieu., *et al.* (2009) BMC genomics 10(1): 1-18. https://doi.org/10.1186/1471-2164-10-550

Gautier M. and Vitalis R. (2012). *Bioinformatics* 28, 1176-7. https://doi.org/10.1093/bioinformatics/bts115

Kiura, J.N., *et al.* (2020) Journal of Agricultural Science and Technology A 10 128-137. https://doi.org/10.17265/2161-6256/2020.03.003

Luikart, Gordon, *et al.* (2001) Proc. of National Academy of Sciences 98(10): 5927-5932. https://doi.org/10.1073/pnas.091591198

Molday R.S., *et al.* (2009). Biochimica et Biophysica Acta (BBA)-Molecular and Cell Biology of Lipids 1791, 573-83. https://doi.org/10.1016/j.bbalip.2009.02.004

Ndeke, A. N., *et al.* (2015) Journal of Agricultural Science and Food Technology 1(6): 78-83.

Onzima, Robert B., *et al.* (2018) Frontiers in Genetics: 318. https://doi.org/10.3389/fgene.2018.00318

Sherman B.T. and Lempicki R.A. (2009) Nature protocols 4, 44.-57

Smith J.M. and Haigh J. (1974) Genetics Research 23(1): 23-35.

Taberlet P., *et al.* (2008) Molecular Ecology 17(1): 275-284. https://doi.org/10.1111/j.1365-294X.2007.03475.x

Tosser-Klopp G., et al (2014) Design and characterization of a 52K SNP chip for goats. *PloS one* 9, e86227. https://doi.org/10.1371/journal.pone.0086227

Waineina R.W., *et al.* (2021) Genetic diversity and population structure among indigenous and imported goat breeds in Kenya. Genetic Resources 2(3): 25-35. https://doi.org/10.46265/genresj.EQFQ1540

Weigand H. and Leese F. (2018) Zoological Journal of the Linnean Society 184(2): 528-83. https://doi.org/10.1093/zoolinnean/zly007

516. Investigation of host genetic role in PCV2 infections

H.R. Wijesena[1], K.M. Sutton[1], S.D. Kachman[2] and D.C. Ciobanu[1*]

[1]Animal Science Department, University of Nebraska-Lincoln, A218K Animal Science Complex, Lincoln, NE 68583-0908, USA; [2]Department of Statistics, University of Nebraska-Lincoln, 354B Hardin Hall, Lincoln, NE 68583-0963, USA; dciobanu@unl.edu

Abstract

A custom swine genotyping array (SowPro91, 111,141 SNPs) was designed to fine map two major QTL (SSC7 and SSC12) associated with PCV2b viral load. SowPro91 was enriched with SNPs in the SLA region of SSC7 QTL, and included the QTN for SSC12 QTL (*SYNGR2 p.Arg63Cys*). The tails of the distribution for PCV2b viral load (n=215) from the PCV2-dataset (n=974) were genotyped with SowPro91. SSC7 SNPs density varied substantially (1-463 SNPs/100 kb). As a result, BayesIM mapping precision was evaluated at multiple prior haplotype sizes (15-240 kb). *SYNGR2 p.Arg63Cys* effect was successfully mapped with shorter haplotypes (<120 kb). Multiple SSC7 QTLs mapped with short haplotypes harbored SNPs with large allelic frequency differences between pigs with extreme GPV for PCV2b viral load. Two closely located QTLs, mapped with short haplotypes, were simply a result of long-distance LD between regions. Further research is needed to genetically characterize this region for its role in immune response.

Introduction

Porcine Circovirus 2 (PCV2) is the etiological agent of a systemic disease that impacts growth and could lead to mortality. Our previous genome-wide association studies (GWAS) using Porcine SNP60 BeadArray (53,529 SNPs) identified two major genomic regions on *Sus scrofa* chromosome (SSC) 7 and 12 associated with viremia and PCV2-specific antibodies (Walker *et al.*, 2018). This study employed one of the largest PCV2 experimental infection data sets (n=974) in which host genetics represented by Porcine SNP60 BeadArray explained 64% of the phenotypic variation for PCV2b viral load. Gene annotation integrated with RNA/genome sequencing followed by *in vitro* gene editing models uncovered the *SYNGR2 p.Arg63Cys* as the source of the SSC12 QTL. The genetic basis of the QTL mapped on SSC7, next to Swine Leukocyte Antigen complex II locus (SLAII) is unknown. The SLA region is characterized by extended linkage disequilibrium (LD) and includes highly polymorphic genes involved in antigen recognition and immune response. Since commercial genotyping arrays are relatively limited in number of SNPs located in SLA region, a novel Affymetrix SNP array was designed (SowPro91) which saturated the SLA complex with SNPs discovered *via* genome and transcriptome sequencing. This array also included SNPs in genes with ontologies related to innate and adaptive immunity and causal variants known to affect viral disease susceptibility (e.g. *SYNGR2 p.Arg63Cys*). The objective of this study was to employ SowPro91 and fine map genomic regions and genetic variants located on SSC7 and associated with PCV2b viral load.

Materials & methods

Populations and phenotypes. Experimental design, procedures and animal use were approved by the IACUC of the University of Nebraska – Lincoln. Experimental PCV2b infection was carried out in 9 batches of maternal crossbred pigs (Large White × Landrace) generated by 14 genetic lines and 7 genetic programs (n=974). PCV2b viremia, specific antibodies and body weights were obtained at 0, 7, 14, 21 and 28 days post-infection (dpi). Experimental details were previously described (Walker *et al.*, 2018).

Sequencing and genotyping. SNPs discovered from RNA and genomic sequencing were used in the development of SowPro91 using Affymetrix Axiom myDesign technology. Transcribed SNPs were obtained from various RNAseq data sets and tissues, including peripheral blood from pigs that expressed extreme

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_516, © H.R. Wijesena *et al.* 2022

levels of viremia (high and low) following an experimental infection with PCV2b (n=8, Walker *et al.*, 2018). Whole genome sequencing and SNP data was obtained from Landrace sires (n=20), a genetic line that contributed to 4 of the 9 batches of PCV2b experimental infections (Wijesena *et al.*, 2019). Genomic DNA was isolated from tail snips using DNeasy tissue kit (Qiagen). The DNA quantity was assessed by Qubit fluorometer (Thermo Fisher Scientific Inc.). A subset of animals representing the tails of the distribution for the residuals for PCV2b viral load (n=215) from the PCV2 dataset (n=974), was genotyped with SowPro91 (111,141 SNPs). The genotype data was filtered at ≥93% sample and ≥97% SNP call rates as described (Wijesena *et al.*, 2019; 2020).

DQB1 haplotyping. The densely polymorphic *DQB1* gene was genotyped by sequencing. Briefly, specific amplicons of the *DQB1* gene were sequenced and compared to full-length reference *DQB1* haplotypes (IPD, https://www.ebi.ac.uk/ipd/mhc). Reference SLAII haplotypes were assigned to pigs with extreme residuals for PCV2b viral load (n=136). Observed and imputed *DQB1* SNP genotypes were combined with SowPro91 in GWAS.

Statistical analyses. Bayes Interval Mapping (BayesIM) was implemented to estimate haplotype effects across the genome on PCV2b viral load. Based on SowPro91 and *DQB1* genotypes, a hidden Markov model was used to generate 8 haplotype states. QTLs were placed every 50 kb across the genome while prior average haplotype size was evaluated at 240, 120, 60, 30 and 15 kb. Model frequencies, genetic variances and haplotype effects were estimated at each locus across the genome. The statistical model included experimental infection batch, litter and pen as random effects, passive IgG and piglet age at infection as covariates.

Results

Recent studies on Porcine Reproductive and Respiratory Virus (PRRSV) and PCV2 showed that host genetics plays an important role in viral disease susceptibility in swine (e.g. Boddicker *et al.*, 2012; Walker *et al.*, 2018). Some of the QTLs detected were likely pathogen-specific such as the ones located on SSC5 (PRRSV) or SSC12 (PCV2). The broad location of the QTL mapped on SSC7 next to SLAII locus, appears to be common in both viruses. While the causal variant of the PCV2-specific SSC12 QTL was functionally demonstrated (Walker *et al.*, 2018) the genetic source of the SLA QTL is unknown.

Development of custom genotyping array. The SowPro91 custom array included 111,141 SNPs and it was designed to aid in fine mapping the QTL mapped near *SLA* locus (SSC7) for PCV2 related traits and other complex traits. Approximately 58% of the SowPro91 SNPs were identified using transcriptomic and genomic sequencing, while 39% of the array includes scaffold SNPs from Porcine SNP60 BeadArray. SowPro91 includes 4,579 SNPs in the broad *SLA* region (22.5-25.3 Mb). The vast majority of the SNPs were from our sequencing data (4,174 SNPs), supplemented by SNPs from Axiom Pig HD array (345 SNPs) and Porcine SNP60 BeadArray (41 SNPs). Across SSC7 (12,249 SNPs) there is an average of 10.9 SNPs/100 kb window. The density varies greatly across the chromosome since it is enriched in variants in the SLA region (Figure 1). While in the broad SLA region (22.5-25.3 Mb) there is an average of 163.5 SNPs/100 kb, outside this region there is an average of 7 SNPs/100 kb. The largest SNP density is located between 24.8-24.9 Mb with 463 SNPs. The average distance between the SNPs in the SLA region is 611.6 bp, while outside the region is 15,488.1 bp.

DQB1 haplotype diversity. Arrays are not effective in genotyping SNPs in dense polymorphic regions. As the result, the polymorphic *DQB1* gene, part of the SLAII, was genotyped by sequencing. The *DQB1* gene was characterized by substantial genetic diversity with 17 observed haplotypes in pigs with extreme residuals for PCV2 viral load. With the exception of a novel haplotype, all the other haplotypes were part

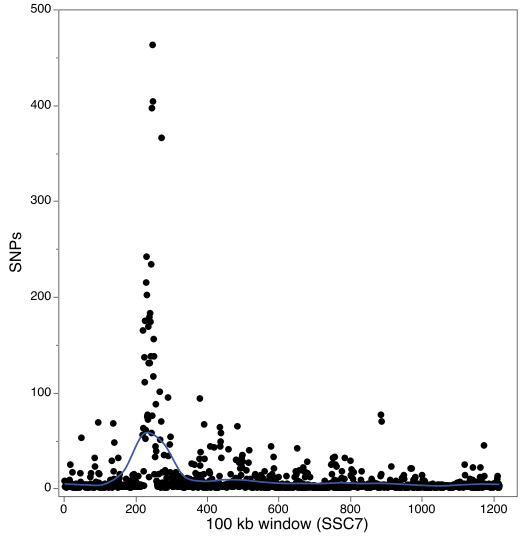


Figure 1. SNP density per 100 kb window on SSC7.

of the reference database. The observed and imputed genotypes of 119 SNPs in *DQB1* gene were combined with SowPro91 in GWAS.

Genome-wide association analyses. Since SNP density varied across SSC7, first we evaluated if prior haplotype size in BayesIM could impact QTL location and effect. While the SSC12 QTL chromosomal position is quite different from SSC7 QTL in terms of recombination rates and SNP density, we hypothesized that the location of the QTN on SSC12 (*SYNGR2 p.Arg63Cys*; 3,797,516 bp) could broadly inform some level of mapping precision for SSC7 QTL. QTL mapping was evaluated by BayesIM using prior average haplotype size of 15-240 kb. We expected that with longer prior haplotype size, the QTN effect will be distributed across different haplotypes sharing functional alleles. The QTL 'window' that included *SYNGR2 p.Arg63Cys* was detected in all models when haplotype size was ≤120 kb (Table 1); the actual distance of the mapped QTL from the QTN was -2,484 bp. At the longest prior haplotype size (240 kb) the QTL was placed farther apart from the QTN (-52, 484 bp). Also, the SSC12 QTL was associated with the largest model frequency at shorter prior haplotype sizes (15-60 kb). In contrast, the SSC7 QTL was the major haplotype at the longer haplotype sizes (120-240 kb). The difference in model frequency between top SSC7 QTL decreased with shorter haplotype size models. Consistent SSC7 QTLs were mapped at 22.1-22.2, 24.5 and 24.65 Mb (Figure 2).

Table 1. Model frequency of the major QTLs for PCV2b viral load using BayesIM.

Prior haplotype size (kb)	Major SSC7 QTL		Major SSC12 QTL	
	Location (Mb)	Model frequency	Location (Mb)	Model frequency
15	24.65	0.021	3.8	0.027
30	24.65	0.019	3.8	0.023
60	24.65	0.024	3.8	0.036
120	24.65	0.030	3.8	0.026
240	24.5	0.063	3.85	0.028

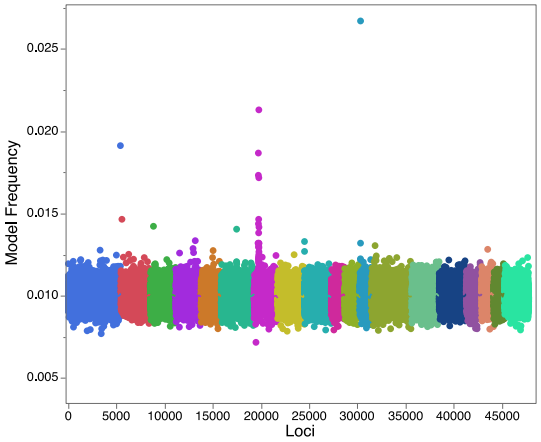


Figure 2. Genome-wide association between SowPro91/DQB1 genotypes and PCV2b viral load using BayesIM. Each dot represents the model frequency associated with 50 kb QTL.

To determine if SNPs located in two closely located major SSC7 QTLs (24.5 and 24.65 Mb) play a role in PCV2 susceptibility, allelic frequency differences were estimated between subsets of pigs representing the tails of the distribution for local genomic prediction values (GPV, 24-25 Mb) for PCV2b viral load. SNPs located in the QTL vicinity (24.5 and 24.65 Mb) displayed the largest differences in allelic frequency. Eleven SNPs had large differences in allelic frequency (>0.8; 24.47-24.65 Mb; Figure 3) with two SNPs having alternate alleles fixed in the subsets of pigs with extreme GPV. LD was estimated between these major candidate SNPs to assess if long LD was the source of these observed QTL. The QTLs mapped at 24.5 and 24.65 Mb are most likely a result of long-distance LD between major SNPs in these regions (r^2=0.61-1.0) (Figure 4). The proximal QTL (22.1/22.2 Mb) appeared to be independent due to low LD with major SNPs from 24.5 and 24.65 Mb QTL ($r^2 \leq 0.20$).

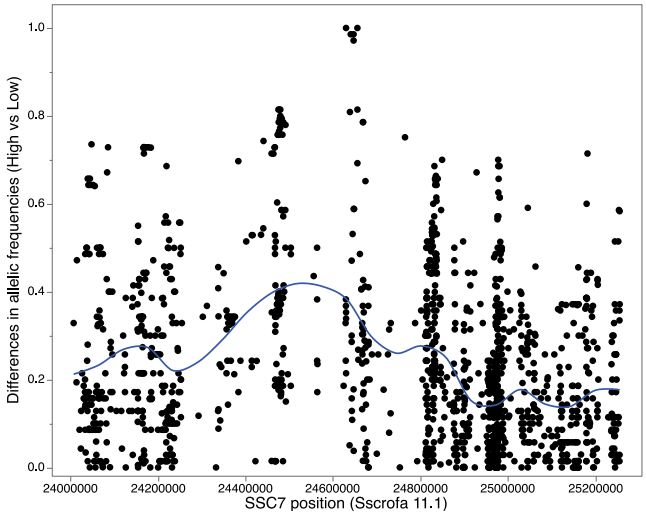


Figure 3. Allelic frequency differences between piglets with extreme local GPV across major SSC7 QTLs (24-25 Mb).

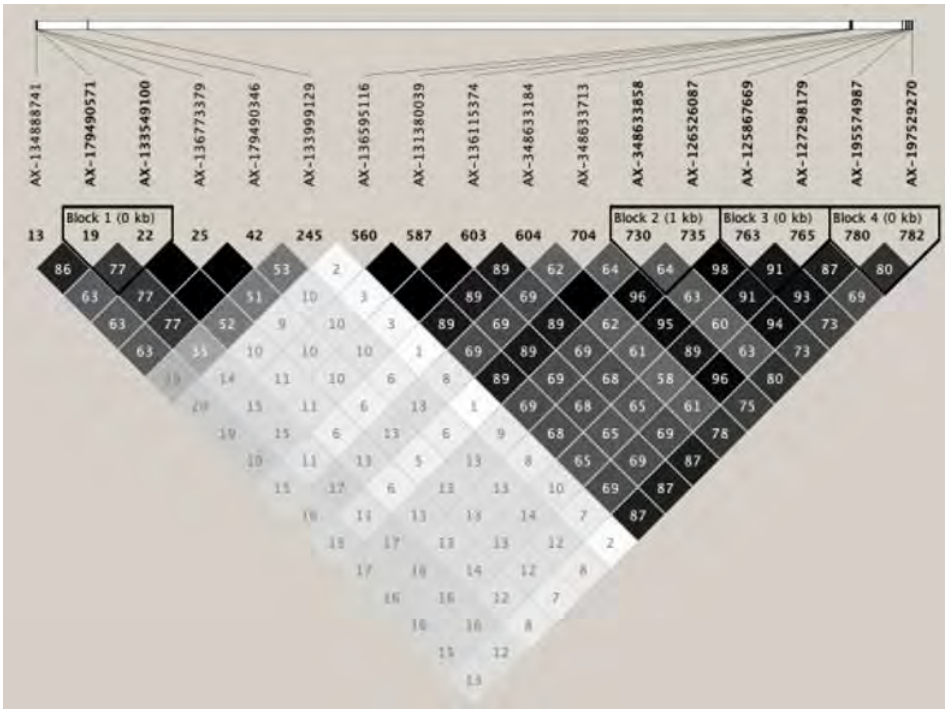


Figure 4. LD (r^2) between major candidate SNPs in the three SSC7 QTLs

Discussion

Detection of causal variants influencing viral disease could impact genetic improvement of modern genetic lines and animal welfare. Mapping precision is a key factor in QTL detection and discovery of functional variants. Previously, we showed that BayesIM was precise in fine mapping QTN, even at lower levels of local LD (Walker *et al.*, 2018). Using SowPro91 data, BayesIM and shorter prior haplotype models fine mapped a known SSC12 causal variant for PCV2b viral load. The short haplotype-based models appeared to detect multiple potentially distinct QTLs in the broad SLA region (SSC7). These regions harbor SNPs that displayed large differences in allelic frequencies between pigs with extreme PCV2b viral load. However, long-distance LD appeared to be the source of two closely located QTLs (24.5 and 24.65 Mb) that initially seemed distinct, especially when mapped with shorter haplotypes models. Further research is needed to characterize the genetic determinism in this highly complex SLA region, with potential broad implications in immune response and antigen recognition.

References

Boddicker N., E. H. Waide, R. Rowland, J. K. Lunney, D. J. Garrick *et al.* (2012) J Anim Sci. 90(6):1733-1746. https://doi.org/10.2527/jas.2011-4464

Walker L.R., Engle T.B., Vu H., Tosky E.R., Nonneman D.J. *et al.* (2018) PLoS Genet. 14(10):e1007750. https://doi.org/10.1371/journal.pgen.1007750.

Wijesena H.R., Rohrer G.A., Nonneman D.J., Keel B.N., Petersen J.L., *et al.* (2019) J Anim Sci. 97(8):3262-3273. https://doi.org/10.1093/jas/skz185

Wijesena H.R., Kachman S.D., Lents C.A., Riethoven J.J., Trenhaile-Grannemann M.D. *et al.* (2020) J Anim Sci. 98(10):skaa293. https://doi.org/10.1093/jas/skaa293

517. Olfaction: a critical force driving adaptive evolution in yaks

A. Kour[1*], S.K. Niranjan[2], S. Jayakumar[3], M. Pukhrambam[1] and M. Sarkar[1]

[1]ICAR-National Research Centre on Yak, Dirang, Arunachal Pradesh 790101, India, [2]ICAR-National Bureau of Animal Genetic Resources, Karnal, Haryana 132001, India, [3]ICAR-Directorate of Poultry Research, Hyderabad, Telangana 500030, India; aneet.kour@icar.gov.in

Abstract

Olfaction is genetically one of the highly evolved senses driving evolution for adaptation in mammals. The present study intended to highlight genomic regions under putative selective sweeps responsible for high-altitude adaptation in the genome of *Arunachali* yak, the only registered yak breed of India (till date). The data of *Arunachali* yak animals obtained through Double digest restriction-site associated DNA (ddRAD) sequencing was processed and 99,919 SNPs were filtered for further analysis. Genetic diversity elucidated based on nucleotide diversity (π=0.041 in 200 bp windows) and effective population size (Ne=83) was found to be optimum. Further, 207 regions across the genome were found to be putatively selected through De-correlated Composite of Multiple Signals (DCMS) statistic. Gene ontology analysis highlighted olfactory sensory perception as significantly overrepresented biological process. These findings reiterate evolutionary role of olfaction in environmental adaptation in one of the most unique germplasm.

Introduction

Sensory adaptations have long been postulated as an effective hypothesis underpinning divergent selection in heterogenous environments. Olfaction is phylogenetically one of the oldest senses helping organisms to interact fruitfully with their respective environments (Yohe and Brand, 2018). Besides imparting cognitive ability for food and life security, olfactory signals have also served as potential catalysts for social interaction and mating success. Thus on an evolutionary time scale, olfaction has immensely contributed to the overall fitness of an individual (Hoover, 2010).

Yak (*Bos grunniens*) is one of the evolutionarily unique species naturally selected to adapt in some of the harshest environmental conditions found on the planet (Qiu *et al.*, 2012). Geographical barriers along with cold and hypoxic conditions at high altitudes have posed a major challenge to the gene pool diversification in the species. It is also expected that certain regions in yak genome including olfactory perception must have been positively selected to ensure its viability and sustainability amidst environmental extremes (Wilkins *et al.*, 2014).

The present investigation was directed at peeking into the evolutionary history of *Arunachali* yak (the only registered yak breed of India till date) population in a holistic manner by estimating divergence in its gene pool and identifying signatures of selection in the breed. Since the breed is reared by tribal pastoralists for sustenance, artificial selection for performance traits has been negligible. Our findings would help us to outline new targets for selection and genetic improvement of the breed.

Methods

ddRAD sequencing was performed for 20 distantly related animals of *Arunachali* yak and the subsequent data generated (BioProject PRJNA577203) was analyzed for genomic diversity estimation and selection signatures identification.

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_517, © A. Kour *et al.* 2022

Pre-processsing of the data. The paired-end FASTQ files were trimmed for the adapter sequences using PRINSEQ lite v0.20.4 (Schmieder and Edwards, 2011) and sequences with PHRED quality score less than 30 were discarded using Stacks (Catchen *et al*., 2011). The processed reads were aligned to *Bos taurus* reference genome assembly (ARS-UCD1.2) using Bowtie 2 (Langmead and Salzberg, 2012). Subsequently, VCFtools (Danecek *et al*., 2011) was used to generate SNPs at read depth 10 (RD 10) and finally, SNPs were filtered out based on quality control (QC) criteria like minor allele frequency (MAF) lesser than 0.05, Call rate (CR) lesser than 90% and *P*-value for Hardy-Weinberg Equilibrium (H-WE) lesser than 0.0001.

Genomic Diversity estimation. In order to assess genomic variability in the population, three indicators were used viz. nucleotide diversity (π), effective population size (N_e) and Runs of Homozygosity (ROH). π for the population was estimated in 200 bp windows of the genome using VCFtools (Danecek *et al*., 2011). N_e was determined using SNeP package in R (Barbato *et al*., 2015) whereas ROH patterns in 0-2 Mb, 2-4 Mb, 4-8 Mb and >8 Mb classes in the genome were assessed using detectRUNS package in R (Marras *et al*., 2015).

Selective sweep identification. Selection signatures in the population were identified by combining the allele-frequency spectrum based and haplotype based statistics like π, Tajima's D and Integrated Haplotype Score (iHS) using a composite measure of selection i.e. DCMS. Subsequently, the DCMS scores for 100 kb windows across the genome were calculated using MINOTAUR package (Lotterhos *et al*., 2017). SNP windows with significant *P*-values ($P \leq 0.01$) were examined for statistical overrepresentation of Gene Ontology (GO) terms in PANTHER (Mi *et al*., 2009) by providing the gene list from the putative regions under selective sweep. False Discovery Rate (FDR) value of ≤0.0001 was set as threshold for significance of GO terms.

Results

96.46 million paired-end raw reads were generated for the analysis. The average alignment rate was 97.71% and it was >85% for all the samples. After filtration, only 99,919 SNPs at RD10 passed the QC criteria and were comsidered for further analysis.

Genomic variability in the population. The average genomic divergence in the population measured using π across 200 bp windows in the genome was 0.041. N_e for the most recent generation (i.e. 13 generations ago) was estimated to be 83. ROH analysis revealed that a major proportion of ROHs (94.33%) were 0-2 Mb class followed by 4.66% in 2-4 Mb and 0.96% in 4-8 Mb category. In longer length category (>8 Mb), a negligible proportion of ROHs was observed (0.03%).

Regions under positive selection. DCMS scores were calculated in non-overlapping 100 kb SNP windows across each chromosome and subsequent *P*-values were calculated. 207 regions across the genome were found to be putatively under selection ($P<0.01$). These regions encompassed 611 protein-coding genes which were functionally annotated to identify the significant GO terms related to biological processes. Sensory perception of smell (GO:0007608) and its associated GO term (GO:0050911) were statistically overrepresented in the analysis (FDR<0.0001). Important genes identified in relation to these processes included: olfactory receptor (OR) genes like *OR5K3, OR5H6* and *OR1E1.*

Discussion

Arunachali yak population was found to be genetically diverse considering the narrow window size used for estimating π. Also, N_e estimates highlight that the number is still viable for long-term survival of the population (Meuwissen, 2009) although the small sample size considered may cloud the reliability of results. Higher prevalence of short and medium length ROHs was reflective of distant demographic and

selection events which resulted in repetitive shortening of chromosome segments due to recombination. Thus, DCMS was used to shed light on the putative regions under positive selection in *Arunachali* yak genome. Being a composite measure of selection, DCMS can perform with greater power and accuracy to detect selection signatures in sample sizes as low as 15 (Ma *et al.*, 2015).

GO terms related to sensory perception of smell were significantly overrepresented in the analysis highlighting the importance of olfaction in evolutionary history of yaks. Olfactory sensation is critical to population fitness as it regulates behaviours necessary for survival and reproduction of a species. It enables the organism to process and interact with various environmental cues and signals so as to facilitate activities like searching for food, attracting mates for procreation, maternal behaviour and resisting attack by predators (Meister 2015). It has also played an important role in speciation events occurring in response to changes in the environmental conditions through adaptation and natural selection of mates (Boughman 2002). As a result of this, each species has evolved its own olfactory strategies and is conditioned to recognize olfactory signals specific to its own evolutionary ecology (Hoover, 2010).

For ruminants residing at high altitude, sensation of olfaction is genetically one of the most evolved physiological adaptations (Friedrich and Weiner, 2020). Olfactory cognition is regulated by a major olfactory system involved in detecting different odorant molecules (or chemical compounds) and an accessory system sensing pheromones or chemical communication between organisms of same species, further eliciting a behavioural response (Hoover, 2010). In order to adapt to difficult environmental conditions of high altitudes and in response to the diversity in the distribution of vegetation at different altitudes, olfactory receptor (OR) genes in yaks have been highly selected and rapidly evolved during domestication (Zhang *et al.*, 2020). Also, pheromone communication may have played a major role in encouraging population fitness and gene pool diversity in the geographically isolated milieu of high altitudes. OR genes have also been reported to promote growth and development of hair follicles, further contributing to yak adaptation at high altitude (Busse *et al.*, 2014). Thus, development of olfactory sense in yaks may have been an important underlying factor ensuring natural selection for 'survival of the fittest'. Our findings shed light on the evolutionarily significant mechanism of olfaction underlying adaptation in yaks. Further, absence of artificial selection signatures for performance traits opens the prospects of genetic improvement in this unique germplasm.

References

Barbato M., Orozco-terWengel P., Tapio M., and Bruford M.W. (2015) SNeP: A tool to estimate trends in recent effective population size trajectories using genome-wide SNP data. Front Genet 6,109. https://doi.org/10.3389/fgene.2015.00109

Boughman J.W. (2002) How sensory drive can promote speciation. Trends Ecol Evol 17:571–577. https://doi.org/10.1016/S0169-5347(02)02595-8

Busse D., Kudella P., Grüning N. M., Gisselmann G., Ständer S. *et al.* (2014) A synthetic sandalwood odorant induces wound-healing processes in human keratinocytes via the olfactory receptor OR2AT4. J Invest Dermatol 134(11):2823–2832. https://doi.org/10.1038/jid.2014.273

Catchen J.M., Amores A., Hohenlohe P., Cresko W., and Postlethwait J.H. (2011) Stacks: Building and genotyping loci de novo from short-read sequences. G3: Genes, Genom Genet 1(3):171–182. https://doi.org/10.1534/g3.111.000240

Danecek P., Auton A., Abecasis G., Albers C. A., Banks E., *et al.* (2011) The variant call format and VCFtools. Bioinformatics *27*(15):2156–2158. https://doi.org/10.1093/bioinformatics/btr330

Friedrich J., and Wiener P. (2020) Selection signatures for high-altitude adaptation in ruminants. Anim Genet 51(2):157–165. https://doi.org/10.1111/age.12900

Hoover K.C. (2010) Smell with inspiration: the evolutionary significance of olfaction. Am J Phys Anthropol 143 (51):63-74. https://doi.org/10.1002/ajpa.21441

Langmead B., and Salzberg S.L. (2012) Fast gapped-read alignment with Bowtie 2. Nat Methods 9(4):357–359. https://doi.org/10.1038/nmeth.1923

Lotterhos, K. E., Card, D. C., Schaal, S. M., Wang, L., Collins, C., & Verity, B. (2017). Composite measures of selection can improve the signal-to-noise ratio in genome scans. Methods Ecol Evol 8(6):717–727. https://doi.org/10.1111/2041-210X.12774

Ma Y., Ding X., Qanbari S., Weigend S., Zhang Q., *et al.* (2015) Properties of different selection signature statistics and a new strategy for combining them. Heredity, 115(5):426–436. https://doi.org/10.1038/hdy.2015.42

Marras G., Gaspa G., Sorbolini S., Dimauro C., Ajmone-Marsan P., *et al.* (2015) Analysis of runs of homozygosity and their relationship with inbreeding in five cattle breeds farmed in Italy. Anim Genet 46(2):110–121. https://doi.org/10.1111/age.12259

Meister M. (2015) On the dimensionality of odor space. Elife 4:e07865. https://doi.org/10.7554/eLife.07865

Meuwissen T. (2009) Genetic management of small populations: A review. Acta Agric Scand A Anim Sci. 59(2):71–79. https://doi.org/10.1080/09064700903118148

Mi H., Dong Q., Muruganujan A., Gaudet P., Lewis S., *et al.* (2009) PANTHER version 7: Improved phylogenetic trees, orthologs and collaboration with the Gene Ontology Consortium. Nucleic Acids Res 38(SUPPL.1):D204-210. https://doi.org/10.1093/nar/gkp1019

Qiu Q., Zhang G., Ma T., Qian W., Ye Z., *et al.* (2012) The yak genome and adaptation to life at high altitude. Nat Genet 44(8):946–949. https://doi.org/10.1038/ng.2343

Schmieder R., and Edwards, R. (2011) Quality control and preprocessing of metagenomic datasets. Bioinformatics *27*(6):863–864. https://doi.org/10.1093/bioinformatics/btr026

Wilkins A.S., Wrangham R.W., and Fitch W.T. (2014) The 'domestication syndrome' in mammals: A unified explanation based on neural crest cell behavior and genetics. Genetics, 197(3):795–808. https://doi.org/10.1534/genetics.114.165423

Yohe L.R., and Brand P. (2018) Evolutionary ecology of chemosensation and its role in sensory drive. Curr Zool 64(4): 525–533. https://doi.org/10.1093/cz/zoy048

Zhang S., Zhang K., Peng X., Zhan H., Lu J., *et al.* (2020) Selective sweep analysis reveals extensive parallel selection traits between large white and Duroc pigs. Evol Appl 13(10): 2807–2820. https://doi.org/10.1111/eva.13085

518. Functional analysis of selective sweeps to identify phenotypic traits in South African Nguni cattle

A.A. Zwane[1*], K.S. Nxumalo[1], M.L. Makgahlela[1,3], E. Van Marle-Koster[2] and N. Maiwashe[1,3]

[1]Animal Breeding and Genetics, Agricultural Research Council-Animal Production, Private Bag X2, Irene, 0062, South Africa; [2]Department of Animal and Wildlife Sciences, University of Pretoria, Private Bag, X20, Pretoria, 0028, South Africa; [3]Department of Animal, Wildlife and Grassland Sciences, University of the Free State, P.O. Box 339, Bloemfontein, 9300, South Africa; zwanea@arc.agric.za

Abstract

Next generation sequencing have made it possible to study positive selection footprints and adaptation traits in livestock. South African (SA) Nguni cattle are well known for their adaption to harsh environments and can produce well. A total of 264 selective sweeps regions identified in Nguni were used to perform functional analysis of genes associated with phenotypic and adaptive characteristics. About 107 genes were identified across chromosomes with 58 genes associated with eight selected phenotypes queries. The $CRHR_2$ gene was associated with meat quality, *IRAK3* with decreased body size, feed intake and fatness, and *CARD15* gene was associated with disease resistance. Four genes (*SPI*, *YWHAZ*, *RGS4* and *RGS5*) were associated with myometrial relaxation in cattle. The *NOD2* and *IL21R* genes were associated with inflammatory bowel diseases whereas *CPLS* gene was associated with fat content in cattle. These genes were significant to the phenotypic and adaptive features present in SA Nguni cattle and will assist in breeding.

Introduction

Indigenous cattle breeds are important to human communities in Africa. South Africa (SA) is the home of five indigenous cattle breeds, Afrikaner, Bonsmara, Drakensberger, Nguni and Tuli. These breeds played a very important role in the history of the country, can survive in harsh local conditions, and are disease resistant. Nguni is a well-known cattle breed in SA due to its adaptive features, hardy and possess excellent resistant to parasites with natural immunity to tick-borne diseases (Mapholi *et al.*, 2015). It generally has a small body size, mostly horned with beautiful colors which makes the breed unique and easily distinguishable (Scholtz and Theunissen, 2010). Recent genomic technologies have made it possible to study the genomes of large animals to detect selection footprints to understand genetic architecture of local adaptation, phenotypic variation and gene function (Ramey *et al.*, 2013). Gene enrichment and pathway analyses can explore polygenic background of complex traits including leucosis, bull fertility, and meat quality on beef cattle. These analyses focus from a single gene to a group-based gene analysis and, findings of these analyses can help understand the genetic and biological architecture of complex traits (Visscher *et al.*, 2017). The aim of this study was to conduct functional enrichment analysis to identify genes pertaining phenotypic variation in SA Nguni cattle using selective sweeps. This information is important in breeding and improvement of these breeds.

Materials & methods

The Nguni selective sweeps data was obtained from the study of Zwane *et al.* (2019). Selective sweeps were identified using the Z-transformed pooled heterozygosity (ZHp) scores computed from a 50% overlapping sliding window approach, with SNP distribution of 150 kb windows. A total of 264 candidate selective sweep (ZHp Z-scores ≤-4) were used for gene enrichment analysis. The QTL information within the selective sweep regions was retrieved and visualized using Animal quantitative trait loci database (QTLdb). A database for annotation, visualization and integrated discovery (DAVID) v6.8 tool was used to convert the Gene ID list into gene names and symbols. In DAVID, the genes were examined to identify significant (P<0.05) gene ontology (GO) terms and for enrichment in molecular functional categories using Kyoto

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_518, © A.A. Zwane *et al.* 2022

Encyclopaedia of Genes and Genomes (KEGG) pathways, using selected phenotypes queries such as fat content, milk production, walking ability, heat tolerance, meat production, reproduction, bone and muscle development. The *P*-values were tested and adjusted by Benjamini correction, and the fold enrichment was used to select the top listed GO terms with statistical significance of $P<0.05$. The false discovery rate (FDR) was used to determine the significance of enrichment or overrepresentation of terms for each annotation (Lewin and Grieve, 2006).

Results

Figure 1 shows 264 candidate selective sweeps regions used for gene enrichment analysis while Table 1 indicate the top 10 genes identified in DAVID.

Discussion

From 264 candidate selective sweeps, the most significant selective sweep regions were identified in chromosome 8 and 19 (Figure 1). The sweeps identified 107 genes located in different chromosomes in

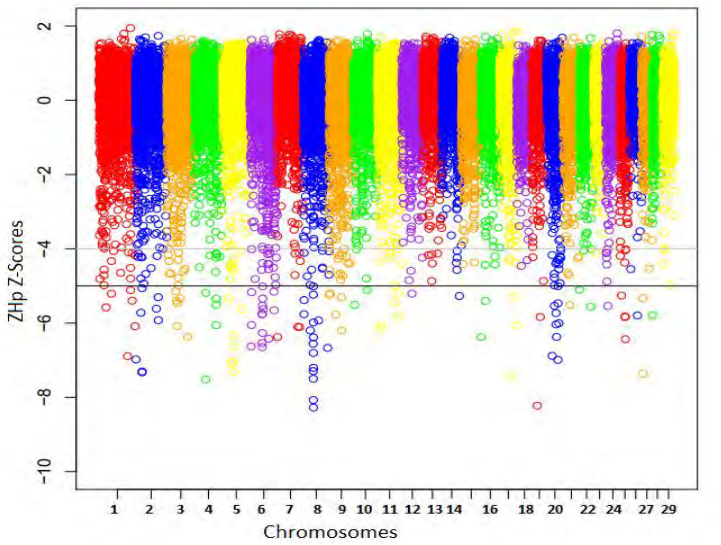


Figure 1. Candidate selective sweep regions (ZHp Z-Scores ≤-4) identified in South African Nguni cattle.

Table 1. List of top 10 genes identified within genomic location (strand) and their associated phenotypes/functions.

Gene symbol	Genomic location	Phenotype/function	Associated trait in cattle
CRHR2	4:66002072-66006203	Increased insulin sensitivity	reproduction, production
NOD2	18:19157447-19166134	Decreased tumour necrosis factor secretion	meat and carcass, health
YWHAZ	14:65584487-65617329	Abnormal skin morphology	exterior, milk
IRAK3	5:47796256-47839675	Decreased body size	meat and carcass, milk, exterior
IL21R	5:25233186-5:268874	Decreased T cell proliferation	milk, health
IMPDH1	4:66002072-66006203	Developmental growth, increased rate	reproduction, production
RGS4	3:6287012-6292402	Decreased systemic arterial diastolic blood pressure	milk, meat and carcass
PRKAA1	20:33688291-33716677	Slow postnatal weight gain	meat, milk, exterior
RGS5	3:6228349-6426528	Decreased total body fat amount	milk, meat and carcass
SP1	5:26753574-26759487	Decreased body size	milk, exterior

Nguni. These genes were previously associated with different functions in mice, human and cattle. Table 1 shows 10 significant genes identified on DAVID with $P<0.25$). Genes such as *CRHR2*, *NOD2*, *YWHAZ*, *IRAK3* and *IL21R* are among the top genes identified in Nguni. The *CRHR2* gene in chromosome 4 was associated with insulin sensitivity. This gene has been associate with juiciness and flavor in cattle (Jiang *et al.*, 2009). The NOD_2 gene, also known as *CARD15*, is a pattern recognition receptor for bacterial lipopolysaccharides which release inflammatory mediators that recognize the muramyl dipeptide and activate NF-kb that controls transcription of tumor necrosis factor-α, interferon-γ, interleukin-1β and interleukin-12 (Wang *et al.*, 2015). *CARD15* protein contributes to immune response of cells and plays an important role in antibodies production. The *CRHR2* gene has been associated with meat quality on Duroc × Pietrain populations (Casiro *et al.*, 2017) and the *CARD15* gene polymorphisms have been associated with tuberculosis susceptibility in Chinese Holstein cows (Wang *et al.*, 2015). Nguni cattle are known to be less susceptible to parasites disease affecting livestock in SA such as ticks and tick borne diseases. The presence of *IL21R* gene in Nguni could assist in the enhancement of their innate immunity.

The GO enrichment analysis have revealed the three molecular functions associated with kinase activity (GO:0004674~protein serine/threonine kinase activity; GO:0004672~protein kinase activity; GO:0016301~kinase activity). The protein kinase activity has been negatively associated with intramuscular fat content in longissimus dorsi muscle of beef cattle (Underwood *et al.*, 2008). This could justify the lean meat grades produced by Nguni cattle (Chulayo *et al.*, 2015). Enrichment analysis of genes was done using selected phenotype queries. Genes were mostly associated with walking ability, fat digestion, bone and muscle development, reproduction and heat tolerance. Genes identified in this study were highly significant to the molecular functions associated with phenotypic characteristics in Nguni cattle. This information is crucial and needed in changing environments.

Acknowledgements

The author would like to acknowledge the funding by the Red Meat Research and Development South Africa and the National Research Foundation. The contribution by the team members of the Agricultural Research Council-Animal Production, Animal Breeding and Genetics, and the South African Journal of Animal Science work dissemination.

References

Casiro S., Velez-Irizarry D., Ernst C.W., Raney N.E., Bates R.O., *et al.* (2017) J. Anim. Sci. 95: 545-558. https://doi.org/10.2527/jas.2016.1003

Chulayo A. and Muchenje V. (2015) SA J. Anim. Sci. 45: 1-19. http://dx.doi.org/10.4314/sajas.v45i5.2

Jiang Z., Michal J.J., Chen J., Daniels T.F., Kunej T., *et al.* (2009) Int. J. Biol. Sci. 5: 528-542. https://doi.org/10.1186/1471-2105-7-426

Lewin A. and Grieve A.C. (2006) BMC Bioinf. 7: 1-9. https://doi.org/10.1186/1471-2105-7-426

Mapholi N.O. (2015) Exploring Genetic Architecture of Tick Resistance in South African Nguni Cattle. Doctoral dissertation, Department of Animal Science, Stellenbosch University, South Africa, pp. 1-120.

Ramey H. R., Decker J. E., McKay S. D., Rolf M. M., Schnabel R. D., *et al.* (2013) BMC Genom. 14: 1-18. https://doi.org/10.1186/1471-2164-14-382

Scholtz M.M. and Theunissen A. (2010) Anim. Genet. Res. 46: 33-39. https://doi.org/10.1017/S2078633610000676

Underwood K.R., Means W.J., Zhu M.J., Ford S.P., Hess B.W., *et al.* (2008) Meat Sci. 79: 394-402. https://doi.org/10.1016/j.meatsci.2007.10.025

Visscher P.M., Wray N.R., Zhang Q., Sklar P., McCarthy M.I., *et al.* (2013) Am. J. Hum. Genet. 101: 5-22. https://doi.org/10.1016/j.ajhg.2017.06.005

Wang Y., Wang S., Liu T., Tu W., Li W., *et al.* (2015) PloS One 10: 1-10. https://doi.org/10.1371/journal.pone.0135085

Zwane A.A., Schnabel R.D., Hoff J., Choudhury A., Makgahlela L.M., *et al.* (2019) Front. Genet. 10: 1-16. https://doi.org/10.3389/fgene.2019.00273

519. Fine-mapping of candidate causal genes for tail length in the Merinolandschaf breed

D.K. Lagler[1,2], E. Kunz[1], K. Eck[1,2], D. Seichter[2], I. Russ[2], C. Mendel[3], G. Lühken[4], M. Upadhyay[1] and I. Međugorac[1*]

[1]Population Genomics Group, Department of Veterinary Sciences, LMU Munich, Lena-Christ-Str. 48, 82152 Martinsried, Germany; [2]Tierzuchtforschung e.V. München, Senator-Gerauer-Str. 23, 85586 Poing, Germany; [3]Institute for Animal Breeding, Bavarian State Research Center for Agriculture, Prof.-Dürrwaechter-Platz 1, 85586 Poing, Germany; [4]Department of Animal Breeding and Genetics, JLU Gießen, Ludwigstr. 21a, 35390 Gießen, Germany; i.medugorac@lmu.de

Abstract

In the face of subsequently tightening animal welfare laws, genetic alternatives for docking tails of lambs in long-tailed sheep breeds are urgently needed. This study aimed to analyse the natural tail length variation in the Merinolandschaf breed and to identify possible causal genes for the short tail phenotype. Combined linkage disequilibrium and linkage mapping based on Illumina OvineSNP50 genotypes of 362 Merinolandschaf lambs resulted in two maximum likelihood ratio test statistic (LRT) peaks which were above the genome-wide significance threshold of LRT=16.63. *HOXB13* was identified as the most relevant candidate gene due to its proximity to the LRT maximum of 29.11 at position 37,311,843 bp on sheep chromosome 11 and its known function as an inhibitor of posterior axial growth. Further studies are needed for an Identification of the causal mutations, linked variants or more precise haplotypes to enable gene- or marker-assisted selection for breeding towards shorter tails

Introduction

Sedentary human communities began sheep management as early as 10,000-11,000 years BP in the Fertile Crescent (Zeder, 2008). It is thought that the Asiatic mouflon (*Ovis orientalis*) is the wild ancestor of domestic sheep. Like other wild sheep subspecies, it is a short-tailed hair sheep. The systematic production and processing of wool did occur several millennia later, leading to the '*Secondary Product Revolution*' (Sherratt, 1983) and the worldwide spread of wool sheep. Selection for fine wool fibers culminated in the economically most important sheep breed, the Merino. The common occurrence of fine wool and long tail in Merinos led to the frequent opinion that these phenotypes are genetically coupled or the result of the same artificial selector (Shelton, 1977). This assumption could not be proven directly. In today's sheep husbandry systems, the long woolly tail comes with several problems. Therefore, most lambs of long-tailed breeds are docked (Shelton, 1977). With the gradual introduction of restrictions on practices that cause suffering to animals, tail docking has come under scrutiny. This means that a long-term non-invasive alternative for the painful practice of tail docking is needed. Here, a genetic solution offers itself and medium to high heritability of tail length (TL) in different sheep breeds provides the genetic basis for breeding for shorter tails (see Greeff *et al.* (2015) and Branford Oltenacu and Boylan (1974)). Therefore, breeders in Australia, New Zealand and other countries attempted shortening of the tail by phenotypic selection within individual breeds. Unfortunately, several authors have observed severe dysfunction in animals homozygous for short-tailed alleles. This phenomenon has been observed in sheep (Carter (1976); James *et al.* (1991); Zhi *et al.* (2018)) as well as in dogs (e.g. Hytönen *et al.*, 2009), cats (e.g. Buckingham *et al.*, 2013) and mice (e.g. Beddington *et al.*, 1992). These undesirable negative side effects slowed down an active breeding programme against overlong tails in sheep breeds. Moreover, there have been no successful genetic mapping studies in Merino sheep breeds, and the possible relationship between this phenotype and dysfunctions has never been investigated on a genetic basis in Merinos. The aim of the present study was therefore to investigate the phenotypic and additive genetic variance in TL in *Merinolandschaf* and to map

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_519, © D.K. Lagler *et al.* 2022

the position of the mayor QTL(s) affecting TL. Together these objectives will enable more efficient breeding towards the short-tailed phenotype and thus improve animal welfare in sheep production without negative side effects.

Materials & methods

Animal samples, phenotypes and genotypes. Phenotyping of 362 lambs was conducted according to the method proposed by Eck *et al.* (2019). To correct the TL for body size and environmental factors, body weight (BW) and height at withers (WH) were measured and sex, age and litter size were recorded. To improve haplotype inference, we sampled and genotyped 22 fathers of above lambs. All sampled sheep were genotyped using Illumina's OvineSNP50 BeadChip, which contains 54,241 almost evenly distributed SNPs (Illumina, 2015). After quality filtering 45,114 markers remained. Haplotype phasing and imputation was conducted using a Hidden Markov Model implemented in Beagle version 5.0 (Browning *et al.*, 2018). To improve the accuracy of haplotyping and imputation, genotype and pedigree information from all available ovine data from the in-house database were added to the haplotyping design (Browning and Browning, 2011). Positions on the reference genome discussed below correspond to the Oar_4.0 reference genome

Estimation of heritability and Mixed linear model-based association analysis (MLMA). To estimate the proportion of variance in our phenotype explained by all SNPs we used the software GCTA v1.93.2, (Yang *et al.*, 2011). To map a putative TL locus, we performed a MLMA analyses with a leave-one-chromosome-out approach using the following model:

$$y = X\beta + Z_a a + Z_u u + e \quad (1)$$

Where y is the vector of TL (cm) and β a vector of fixed effects including the mean, sex, age, BW and WH at phenotyping. a is the vector of the additive effect (fixed) of the candidate marker to be tested for association, u is the vector of polygenic effect (random or accumulated) of all markers, excluding those on the chromosome which contains the candidate SNP. e is a vector of residuals. X, Z_a and Z_u are the incidence matrices for β, a and u, respectively. We considered 45,114 markers resulting in the suggestive *P*-value at 2.22×10^{-5} and genome-wide at 1.11×10^{-6} (Xiong *et al.* (2015)).

Combined linkage disequilibrium and linkage analysis (cLDLA). We also performed a haplotype-based cLDLA analyse. To account for linkage disequilibrium in the form of local haplotype relationships, the Locus IBD was estimated according to the method of Meuwissen and Goddard (2001) using sliding windows of 40 SNPs. Variance component analyses were carried out with the programme ASReml (Gilmour *et al.* (2009)) according to the method of Meuwissen *et al.* (2002). ASReml estimated the maximum likelihood, variance components, fixed and random effects simultaneously by considering the inverse of the genome-wide unified additive (see Powell *et al.* (2010)) and locus specific (see Lee and Van der Werf (2006) relationships matrices in the following mixed linear model:

$$y = X\beta + Z_1 u + Z_2 q + e \quad (2)$$

Where y is again the vector of TL and β is the vector of fixed effects (BW, WH, sex, age and the overall mean μ). u is the vector of random polygenic effects, q is the vector of random additive-genetic QTL effects and e is the vector of random residual. X, Z_1, Z_2 are incidence matrices linking the observed values with the fixed and random QTL effects. The logarithm of the likelihood estimated by ASReml for the model with ($log(L_1)$) and without QTL ($log(L_0)$) effects were then used for a likelihood ratio test statistic (LRT). The LRT values follow a χ^2 distribution with one degree of freedom (Heuven *et al.*, 2005) and were calculated as:

$$LRT_p = -2 \times (log(L_0) - log(L_{1p})) \quad (3)$$

where p refers to the position in the genome where the QTL effects are estimated (i.e. the centre of the sliding SNP window p). Bonferroni correction was carried out to account for multiple testing (Bland and Altman, 1995). This resulted in a corrected P-value of $<4.545\times10^{-5}$ and a corresponding LRT value with genome-wide significance is equal to 16.629. For all maxima of the LRT curve (*LRTmax*) that exceeded the genome-wide significance threshold, the 2-LOD (logarithm of the odds) criterion was used to determine the associated confidence intervals (van Ooijen, 1992).

Results

The GCTA-GREML resulted in SNP-based heritability of 0.992 with relatively low standard error of 0.12. Despite heritability close to 1 the association analysis revealed no genome-wide significant association between SNPs and TL. On the other side the haplotype based cLDLA mapping resulted in two genome wide significant QTLs associated with TL in Merinolandschaf (Figure 1). The most prominent and narrow peak is on the sheep chromosome 11 (OAR11) at position 37,111,462 bp with *LRTmax* = 29.46 corresponding to $P=5.71\times10^{-8}$. The second was mapped to the OAR2 at position 94,538,115 bp with *LRTmax* = 19.356 corresponding to $P=1.08\times10^{-5}$. The confidence interval for the *LRTmax* on OAR11 was set between positions 37,000,925 bp and 37,521,490 bp and for OAR2 between positions 93,441,900 bp and 96,402,884 bp. These intervals were then considered in the UCSC Genome Browser. *HOXB13* (37,290,203-37,292,513) as the most prominent candidate lies only 179-Kb proximal to the *LRTmax*.

Discussion

The results presented here suggest that TL inheritance is largely dependent on additive genetic variance (h^2 =0.992), but most likely due to the small sample size and the small number of markers, GWAS does not yield a significant relationship. The haplotype-based mapping resulted in one highly significant and fine mapped QTL on OAR11 and less sharp mapped but genome-wide significant QTL on OAR02. In our study, we could not detect any increase in the LRT curve in the candidate genes or regions detected in mapping studies of dogs, cats and mice. The confidence interval on OAR11 include complete ovine HOXB cluster and further seven genes. A corresponding literature search (see Diaz-Cuadros *et al.* (2021) for a review) clearly indicates *HOXB13* as the most likely gene causing the large effect in our design. The literature search for possible plausible candidates at OAR02 yielded hardly any useful results. We tried to improve our search with MGI Mammalian Phenotype level 4 (MMP4) ontology (66), but still we see no plausible link to our trait of interest. The haplotype analyses of the presented QTL regions will be used to select animals for capture sequencing of the confidence interval on OAR11 and OAR2. These analyses should yield promising causal candidate mutations, hopefully without the negative side effects mentioned above.

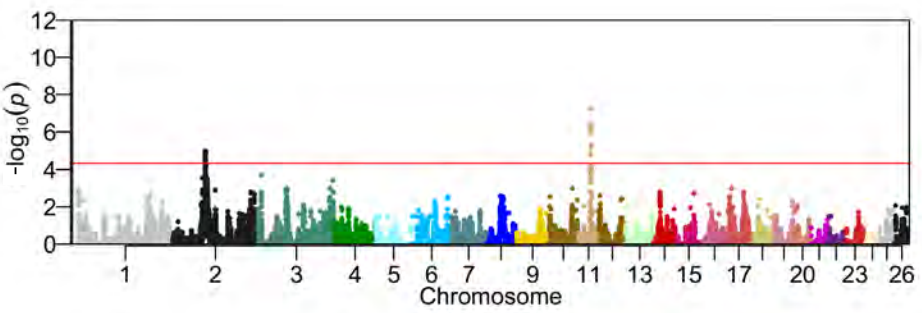


Figure 1. Manhattan Plot of the cLDLA mapping results.

References

Beddington R. S. P., Rashbass P. and Wilson V. (1992). *Development*, 157-65.

Bland J. M. and Altman D. G. (1995). *BMJ, 310*, 170. https://doi.org/10.1136/bmj.310.6973.170

Branford Oltenacu E. A. and Boylan W. J. (1974). *Journal of Heredity, 65*(6), 331-4. https://doi.org/10.1093/oxfordjournals.jhered.a108543

Browning B. L., Zhou Y. and Browning S. R. (2018). *American journal of human genetics, 103*(3), 338-48. https://doi.org/10.1016/j.ajhg.2018.07.015

Browning S. R. and Browning B. L. (2011). *Nature Reviews Genetics, 12*, 703-14.

Buckingham K. J., McMillin M. J., Brassil M. M., Shively K. M., Magnaye K. M., *et al.* (2013). *Mamm Genome, 24*(9-10), 400-8. https://doi.org/10.1007/s00335-013-9471-1

Carter A. H. (1976). *New Zealand Ministry of Agriculture and Fisheries. Annual report of the research division.*, 44-5.

Diaz-Cuadros M., Pourquié O. and El-Sherif E. (2021). *PLoS Genet, 17*(10), e1009812. https://doi.org/10.1371/journal.pgen.1009812

Eck K., Kunz E., Mendel C., Lühken G. and Medugorac I. (2019). *Small Ruminant Research*. https://doi.org/10.1016/j.smallrumres.2019.04.007

Gilmour A. R., Gogel B. J., Cullis B. R., Thompson R. and Butler D. (2009). *ASReml User Guide Release 3.0*. Retrieved from Hemel Hempstead:

Greeff J. C., Karlsson L. J. E. and Schlink A. C. (2015). *Proc. Assoc. Advmt. Breed. Genet., 21*, 237-40.

Heuven H. C. M., Bovenhuis H., Janss L. L. G. and van Arendonk J. A. M. (2005). *Genetics Selection Evolution, 37*, 635-55. https://doi.org/10.1186/1297-9686-37-7-635

Hytönen M. K., Grall A., Hédan B., Dréano S., Seguin S. J., *et al.* (2009). *Journal of Heredity, 100*, 236-40. https://doi.org/10.1093/jhered/esn085

Illumina. Data Sheet: Agrigenomics. OvineSNP50 Genotyping BeadChip. Available at: https://www.illumina.com

James P., Ponzoni R., Gare D. and Cockrum K. (1991). Proceedings of the Association for the Advancement of Animal Breeding and Genetics.

Lee S. H. and Van der Werf J. H. J. (2006). *Genetics, 174*, 1009-16. https://doi.org/10.1534/genetics.106.060806

Meuwissen T. H. and Goddard M. E. (2001). *Genet Sel Evol, 33*(6), 605-34. https://doi.org/10.1186/1297-9686-33-6-605

Meuwissen T. H. E., Karlsen A., Lien S., Olsaker I. and Goddard M. E. (2002). *Genetics, 161*, 373-9. https://doi.org/10.1093/genetics/161.1.373

Powell J. E., Visscher P. M. and Goddard M. E. (2010). *Nat Rev Genet, 11*(11), 800-5. https://doi.org/10.1038/nrg2865

Shelton M. (1977). *The Journal of heredity, 68*(2), 128-30. https://doi.org/10.1093/oxfordjournals.jhered.a108788

Sherratt A. J. W. a. (1983). *15*(1), 90-104. https://doi.org/10.1080/00438243.1983.9979887

Van Ooijen J. W. (1992). *Theor Appl Genet, 84*(7-8), 803-11. https://doi.org/10.1007/BF00227388

Xiong X., Liu X., Zhou L., Yang J., Yang B., *et al.* (2015). *Mamm Genome, 26*(3-4), 181-90. https://doi.org/10.1007/s00335-015-9558-y

Yang J., Lee S. H., Goddard M. E. and Visscher P. M. (2011). *American journal of human genetics, 88*(1), 76-82. https://doi.org/10.1016/j.ajhg.2010.11.011

Zeder M. A. (2008). *Proceedings of the National Academy of Sciences of the United States of America, 105*(33), 11597-604. https://doi.org/10.1073/pnas.0801317105

Zhi D., Da L., Liu M., Cheng C., Zhang Y., *et al.* (2018). *G3 (Bethesda), 8*(2), 377-83. https://doi.org/10.1534/g3.117.300307

520. Massive detection of cryptic recessive genetic defects in livestock mining millions of life trajectories

A. Guintard[1,2], F. Besnard[1,3], M. Cano[1,2], C. Hozé[1,2], H. Leclerc[1,2], M. Boussaha[1], C. Grohs[1], A. Barbat[1], S. Fritz[1,2], C. Escouflaire[1,2], J. Rivière[1,4], C. Péchoux[1], C. Danchin-Burge[3], G. Foucras[5], A. Relun[6], V. Plassard[7], M.-A. Arcangioli[8], S. Mattalia[1,3], D. Boichard[1] and A. Capitan[1,2*]

[1]Université Paris-Saclay, INRAE, AgroParisTech, GABI, 4 av. Jean Jaurès, 78350 Jouy-en-Josas, France; [2]ALLICE, 149 rue de Bercy, 75012 Paris, France; [3]IDELE, 149 rue de Bercy, 75012 Paris, France; [4]Université Paris-Saclay, INRAE, AgroParisTech, MICALIS, 4 av. Jean Jaurès, 78350 Jouy-en-Josas, France; [5]Université de Toulouse, ENVT, IHAP, 23 ch. des Capelles, 31300 Toulouse, France; [6]INRAE, Oniris, BIOEPAR, La Chantrerie, 44307 Nantes, France; [7]ENVA, 7 av. du Général de Gaulle, 94700 Maisons-Alfort, France; [8]VetAgro Sup, University of Lyon1, 1 av. Bourgelat, 69280 Marcy-l'Etoile, France; aurelien.capitan@inrae.fr

Abstract

A data mining method applied to large-scale genotyping data is proposed to detect recessive loci responsible for increased mortality in cattle and that have remained undetected by previous approaches. It is based on a screen for homozygous haplotype enrichment/depletion in groups of females with different life trajectories. After validation of the results in at risk and control mating, 34 deleterious haplotypes (13 in Holstein, 11 in Montbéliarde, and 10 in Normande) were identified, with frequencies ranging from 1.5 to 7.6%. Profiles of survival curves and causes of mortality differed greatly between loci, with early juvenile, late juvenile and evenly distributed death events. Candidate causal variants were found for fifteen haplotypes. A frameshift mutation of *NOA1* and a disruptive inframe deletion of *RFC5*, affecting two genes with no previous record of live homozygous mutants in mammals, were subject to phenotypical characterization.

Introduction

Due to their limited effective size and intense selection, cattle breeds experience a gradual increase in inbreeding, favouring the emergence of recessive defects. While a few of them are highly visible, many remain overlooked because they induce mortality during gestation or after birth but with non-specific symptoms. With the tremendous development of genomic selection and genotyping, new methods have been proposed to detect these defects such as the deficit in homozygotes (VanRaden *et al.*, 2011), reverse genetics (e.g. Michot *et al.*, 2016), or GWAS accounting for dominance (Reynolds *et al.*, 2021). In this paper, we propose another approach combining searches for enrichment/depletion in homozygotes and analyses of life trajectories. We demonstrate its detection power with the identification of several dozens of new defects responsible for increased mortality rate.

Materials & methods

Mapping of recessive loci and analysis of life trajectories. Information on pedigree and life trajectory (date of birth; lifespan; cause of death; and ages at first insemination, calving and lactation) was extracted from the French National database for 5.96 million Holstein, 1.63 million Montbéliarde, and 1.24 million Normande females with Illumina Bovine SNP50 array genotypes available for their sire and maternal grand sire (MGS). Among them, 8,203, 6,198 and 2,254 cases (heifers that died of natural cause before 3 years and were never inseminated) and 291,529, 141,343 and 56,095 controls (females that calved and started a first lactation) from the three breeds, respectively, were also genotyped with various Illumina arrays and their genotypes were imputed and phased to the Bovine SNP50 as described in Mesbah-Uddin *et al.* (2019). For sliding haplotypes of 20 markers we counted the number of homozygotes (hmz) observed (Nobs) within each group of genotyped individuals and in parallel we estimated the expected number of hmz (Nexp) using within-family transmission probability. We filtered haplotypes satisfying the following criteria: Nobs

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_520, © A. Guintard *et al.* 2022

≥10 in cases, increase in hmz (i.e. (Nobs-Nexp)/Nexp)) ≥25% in cases and ≤-25% in controls. Among stretches of consecutive haplotypes, we selected the one showing the highest increase in hmz in cases as the 'peak haplotype'. For validation we compared the proportions of animals belonging to three categories (case, control and others) using a *X* test with Benjamini-Hochberg correction; *P*-value≤5%) among the descendants of at-risk mating ('1'; carrier sire and carrier MGS) or control mating ('0'; noncarrier sire and MGS) born between 2000 and 2015 (and that were mostly not genotyped themselves). Then, for 34 validated haplotypes, we calculated the daily proportion of animals that died of a natural death (D), were slaughtered (S) or were still alive (A) over a period of 6 years for mating types 0 and 1. We also calculated the D0-D1, S0-S1 and A0-A1 differences in proportions on a daily basis and scored the days for which 25, 50, 75 and 100% of the maximum deviation between each difference in proportions was reached. Subsequently, we used these 12 parameters to perform a Principal Component Analysis (PCA) and a Hierarchical Clustering (HC) using RStudio package Factoshiny, v 1.2.5033.

Estimation of the effect of the haplotypes on various traits. For each of the 34 selected haplotypes we estimated the effects of the three genotypes on 11 traits (see results). We used yield deviation data corrected for environmental effects for each animal in a model including the haplotype as a fixed effect and a polygenic effect through a pedigree relationship matrix. Analyses were performed with BLUPF90 (Misztal *et al.*, 2014).The size of genotype groups ranged from 4 (minimum set value) to 346,210 animals.

Analysis of whole genome sequences. The genomes of 611 cattle from 22 breeds (including 190 Holstein, 72 Montbéliarde and 97 Normande individuals) were sequenced with Illumina technology and processed as described in Boussaha *et al.* (2016). Breed specific SNPs or short Indels carried by at least 50% of the haplotype carriers, predicted to be deleterious and located within ±10 Mb from the haplotype were considered as candidate causative mutations.

Clinical examination. Five *NOA1* and five *RFC5* homozygous mutants were subject to clinical examination ante and post mortem including blood dosage and histological analyses.

Results & discussion

Our screen for homozygous haplotype enrichment or depletion (HHED) in groups of females with different life trajectories identified numerous regions, which led us to consider the top 20 peak haplotypes per breed (Figure 1a). In total, 34 haplotypes (13 in Holstein, 11 in Montbéliarde, and 10 in Normande) with frequencies comprised between 1.5 and 7.6% were confirmed to increase juvenile mortality in validation populations. Among them only one haplotype colocalized with a deleterious locus previously reported, namely cholesterol deficiency in Holstein (CDH; e.g. Menzi *et al.*, 2016). The proportion of mating at risk between sire and MGS for one or more validated haplotypes reached 4.7, 7.1 and 8.4% for the three breeds, respectively. The subsequent analysis of the life trajectories of animals from at risk vs control mating using PCA and HC (Figure 1 b-f) distinguished three categories of haplotypes depending on the age and causes of death: early juvenile mortality (d), progressive mortality in juveniles and subadults (e) and increased mortality and premature culling throughout life (f).

For 11 traits routinely collected for evaluation purpose, high and negative effects were observed for the hmz carrier genotype while the heterozygous one showed low or no effect as compared with the noncarrier genotype (Figure 2a). This suggest that these deleterious haplotypes owe their presence in the populations mostly to genetic drift rather than hitchhiking or balancing selection.

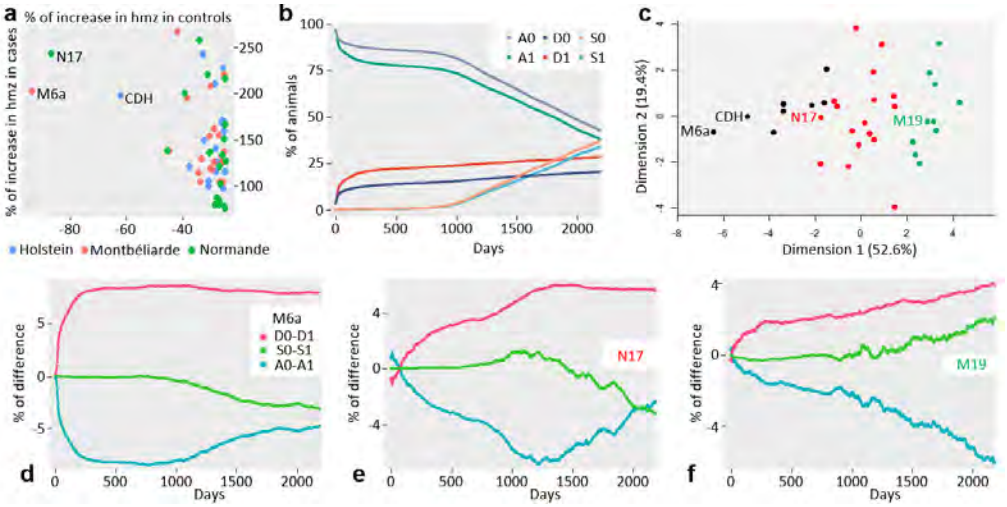


Figure 1. Mapping and characterization of life trajectories. a) increase in hmz in the case and control groups for the 60 selected haplotypes; c) daily proportion of animals from at risk (1) or control (0) mating for haplotype M6a that died (D), were slaughtered (S) or are alive (A); c) Results of PCA and HC for 34 haplotypes × 12 parameters. d-f) daily differences between D,S and E categories for mating types 0 and 1 for haplotypes M6a, N17 and M19, respectively.

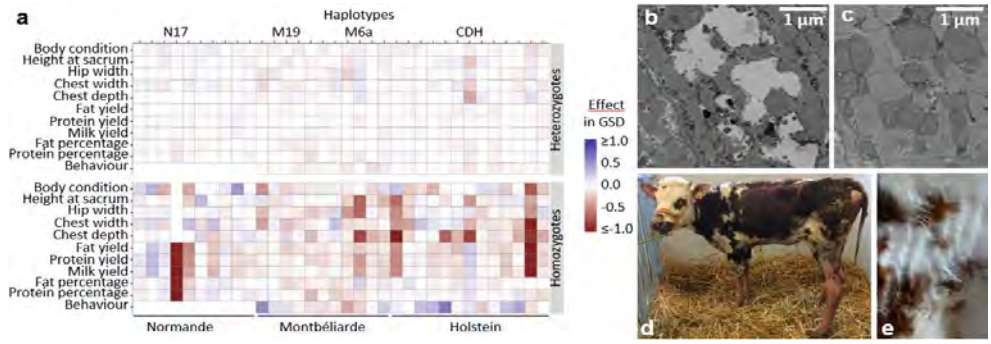


Figure 2. Effects of the 34 haplotypes on 11 traits (a) and results of clinical examinations (b-g). Electron microscopy of myocardium from NOA1-/- (b) and control (c) animals; General view (d) and detail of the fur of a RFC5-/- animal (e).

From whole genome sequences of 611 animals, we retained 22 promising candidates for 15 haplotypes, with predicted consequences including recessive metabolic, immune and neurological defects. Among them, candidate variants for haplotypes M6a in Montbéliarde and N17 in Normande affected genes with no previous record of live homozygous mutants in mammals, and were subject to phenotypical characterization. M6a candidate was a frameshift mutation of NOA1 (p.D400Rfs9) which is involved in the regulation of mitochondrial protein translation and respiration. Based on information of at risk mating data 50% of NOA1-/- individuals died before 41 days of life. The calves we examined were cachetic, had diarrhea and were unable to stand. Biochemical analyses (glycemia>1 g/l, creatine kinase>200 UI/l and K<2.5 mmol/l of serum) suggested that the organism responded to metabolic demand by lytic activity

of muscles, or even autophagy. After euthanasia, electron miscrocopy of myocardium revealed a nearly complete mitochondrial apoptosis and increased storage of glycogen (Figure 2b,c). Finally, N17 candidate was a disruptive inframe deletion of the RFC5 subunit (p.E369del) of a protein complex involved in DNA replication. RFC5-/- animals showed retarded growth, thin and curly hair, and alopecia of body extremities (Figure 2d,e). They died after recurrent episodes of diarrhea between six months and three years of age.

Conclusions

In conclusion, we propose an innovative strategy to detect deleterious recessive loci that remained overlooked by previous approaches. Taken as a whole, the dozen of loci validated within each breed is responsible for the death of 0.5 to 1% of the dairy females born each year in France. The management of all these new defects in selection will contribute to improve both animal welfare and the income of breeders.

Acknowledgements

This work was funded by APIS-GENE (grant Effitness). We thank A. Remot, M. Vilotte, S. Barbey (INRAE), T. Buronfosse, M. Bouchier, and E. Contat (VetAgroSup) for their help.

References

Boussaha M., Michot P., Letaief R., Hozé C., Fritz S., *et al.* (2016) Genet Sel Evol. 48(1):87. https://doi.org/10.1186/s12711-016-0268-z

Menzi F., Besuchet-Schmutz N., Fragnière M., Hofstetter S., Jagannathan V., *et al.* (2016) Anim Genet. 47(2):253-7. https://doi.org/10.1111/age.12410.

Mesbah-Uddin M., Hoze C., Michot P., Barbat A., Lefebvre R., *et al.* (2019) J Dairy Sci. 102(7):6340–6356. https://doi.org/10.3168/jds.2018-16100

Michot P., Chahory S., Marete A., Grohs C., Dagios D., *et al.* (2016) Genet Sel Evol 48, 56. https://doi.org/10.1186/s12711-016-0232-y

Misztal, I., Tsuruta S., Lourenco D.A.L., Aguilar I., Legarra A., *et al.* (2014). http://nce.ads.uga.edu/wiki/lib/exe/fetch.php?media=blupf90_all2.pdf

Reynolds E.G.M., Neeley C., Lopdell T.J., Keehan M., Dittmer K. et al (2021) Nat Genet 53(7), 949-+. https://doi.org/10.1038/s41588-021-00872-5

VanRaden P.M., Olson K.M., Null D.J., Hutchison J.L. (2011) J Dairy Sci 94:6153–6161. https://doi.org/10.3168/jds.2011-4624

521. A *de novo* missense mutation of COL1A1 causes osteogenesis imperfecta type 2 and premature delivery in Normande Cattle

C. Grohs[1*], J. Corbeau[2], J. Jourdain[3], M. Boussaha[1], V. Plassard[4], A. Barbat[1], C. Hamelin[5], J. Mortier[4], D. Boichard[1], R. Guatteo[2] and A. Capitan[1,3]

[1]Université Paris Saclay, INRAE, AgroParisTech, GABI, 78350, Jouy-en-Josas, France; [2]BioEpAR, INRAE, Oniris, CS, 40706 Nantes, France; [3]ALLICE, 75012 Paris, France; [4]Ecole Nationale Vétérinaire d'Alfort, Maisons-Alfort, France; [5]INNOVAL, 35538 Noyal sur Vilaine, France; cecile.grohs@inrae.fr

Abstract

Seventeen male and female calves showing shortened gestation length and multiple fractures at birth were reported to the French National Observatory of Bovine Abnormalities (ONAB) among the descendants of a single Normande sire, suggesting autosomal inheritance with mosaicism. Clinical examination revealed clinical signs consistent with type II osteogenesis imperfecta (OI). Detailed analysis of gestation length and perinatal mortality supported a moderate proportion of affected calves (7.7%; 107 animals) and highlighted the underreporting of this congenital defect to the dedicated observatory. Genetic mapping, whole genome sequencing, and subsequent genotyping of cases and control animals identified a *de novo* missense substitution within the NC1 domain of *COL1A1* (Chr19 g.36473965G>A; p.D1412N). The same mutation has been reported several times to cause type II OI in human. In conclusion, we report the fourth mutation of *COL1A1* associated with type II OI in cattle and an interesting model for a syndrome also observed in human.

Introduction

In vertebrates, osteogenesis imperfecta (OI) constitutes a group of at least 15 genetic disorders characterized by bone deformity and fragility with various degrees of severity and possible addition of other symptoms (Sillence *et al.*, 1979; Besio *et al.*, 2019). Studies in humans have demonstrated that only two genes coding for the alpha 1 and alpha 2 chains of collagen type I account for 77% of cases, the rest being caused by mutations of at least 17 other genes (e.g. Marini *et al.*, 2007; Bardai *et al.* 2016). To date only three independent cases of OI have been characterized at the molecular level in cattle, all caused by *COL1A1* mutations (Bourneuf *et al.*, 2017; Jacinto *et al.*, 2021; Petersen *et al.*, 2019). From 2017 to 2019, 17 purebred Normande calves (9 males and 8 females) with symptoms consistent with OI were reported to the French National Observatory for Bovine Abnormalities (ONAB; doi.org/10.15454/BRKOV3). All cases were direct descendants from the same healthy Normande AI bull (called 'Ly.') suggesting an autosomal dominant inheritance with somatic or germline mosaicism. The purpose of this study was to characterize the clinical and genetic features of this new syndrome.

Materials & methods

Phenotypes. One affected calf and an unrelated control that died of natural death at one week of age were frozen at -20 °C and subjected to various analyses (CT scan, X-ray, necropsy, reassembly of limb bones and histology of bones and tendons). Gross clinical description and DNA samples were also available for 16 additional affected half-sibs. In addition, information on the date of birth, survival within 48h after delivery, and gestation length were recovered for 1,387 progeny of Ly. and 161,413 Normande calves from 63 control bulls born the same year as Ly. (mean=2,439±1,747 progeny per sire).

Genetic analyses. The 17 cases, 84 half-sib controls and Ly. were genotyped with various SNP arrays (Illumina, San Diego, USA) and then phased and imputed to the Bovine SNP50 in the framework of the French genomic evaluation as described in Mesbah-Uddin *et al.* (2019). Then, paternal phases were screened for

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_521, © C. Grohs *et al.* 2022

sliding haplotypes of 50 markers that were carried by all cases and by less than 50% of their half-sib controls. We subsequently sequenced the genome of one case using the Illumina NovaSeq platform (150 bp paired-end reads), processed the data in accordance with the guidelines of the 1000 Bull Genomes Project (Hayes and Daetwyler), and used 5,116 genomes from run 9 of the latter project as controls. A candidate mutation in the *COL1A1* gene was genotyped by PCR amplification with primers GCCTCCCAGAACATCACCTA and CTTTTCGGGGGTTTCAGTTT using Go Taq Flexi DNA polymerase (Promega), followed by Sanger sequencing (Eurofins MWG, Ebersberg, Germany).

Results

At necropsy, the frozen specimen presented light blue sclerae and dentinogenesis imperfecta with grey and friable teeth (not shown). CT scan and preparation of the skeleton revealed the existence of ancient fractures associated with bone loss or abnormal repair mainly located on the diaphysis of long bones, as well as recent limb and rib fractures probably due to the mechanical constraints of calving (Figure 1a-d).

This calf also suffered from arthrogryposis associated with a thickening of the joint capsules, and from a generalized osteopenia as revealed by a comparative x-ray analysis with a control calf (not shown). These symptoms fitted perfectly with the clinical features of the most severe form of OI, namely type II. The 17 cases all died at birth or soon after, and displayed twisted or broken limbs, as well as reduced gestation length (Figure 1e-g). Ly. was found to be the worst Normande bull for perinatal mortality with 15.9% of

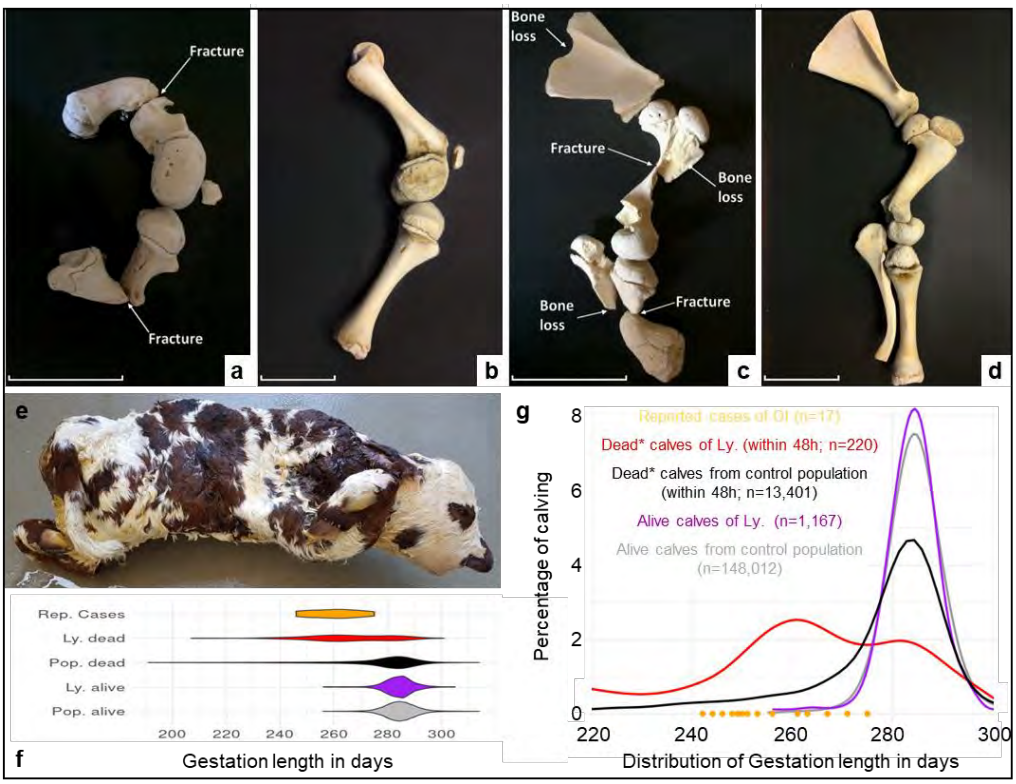


Figure 1. Clinical manifestation of OI. Details of bones from posterior (a,b) and anterior (c,d) limbs of case (a,c) and control (b,d) calves; general view of an OI calf (e); statistics on gestation length and perinatal mortality within the descendants of Ly. and control bulls (f) and related distributions (g).

calves declared as dead within 48h after birth versus 8.2±1.5% for 63 control sires. Assuming that this excess of perinatal mortality was due to OI, we estimated the proportion of affected calves to be 7.7% (i.e. 15.9-8.2%) among its progeny. This would represent 107 cases, which contrasts with the only 17 reported to us and highlights the underreporting of congenital defects to heredosurveillance platforms, even for textbook genetic syndromes. The gestation lengths of the perinatally dead calves of Ly. were significantly shorter than those of matched control calves ($P<2.2\times10^{-16}$; T test; Figure 1f) and showed a bimodal distribution with an additional subpopulation centered around 260 days that comprised the reported cases of OI (Figure 1g). This observation is in agreement with previous reports of premature rupture of membranes in sporadic cases of OI due to collagen defects, leading to preterm delivery (e.g. Cole *et al.*, 1992).

To gain insights into the genetic architecture of OI in this family, we analysed phased Bovine SNP50 genotypes from the 17 cases and 84 half-sib controls. Assuming autosomal dominant inheritance, we mapped the locus to chromosome 19 between positions 35,459,825 and 41,972,985 bp (ARS-UCD1 assembly). Then, by comparing whole-genome sequences of one case and 5,116 controls representing the worldwide cattle diversity, we identified a single heterozygous candidate variant in this interval: a missense variant affecting an aspartic acid residue within the NC1 domain of COL1A1 that is entirely conserved among vertebrates (Chr19 g.36473965G>A; p.D1412N; SIFT score = 0; Figure 2).

Interestingly, the same substitution has been reported three times (e.g. Bodian *et al.*, 2009) to be responsible for type II OI in human which supports its causality. As a further verification we genotyped the sire, 3 affected calves and 3 controls carrying the same paternal haplotype (but in the non mutated version) using PCR and Sanger sequencing and confirmed the *de novo* nature of this mutation. Finally, based on the small deviation of the 1:1 ratio for each paternal haplotype among the 84 control animals (i.e. 40 vs 44) we estimated a proportion of affected calf of 4.5% (4/88) which is consistent with our previous estimation of 7.7% based on the excess of perinatal mortality.

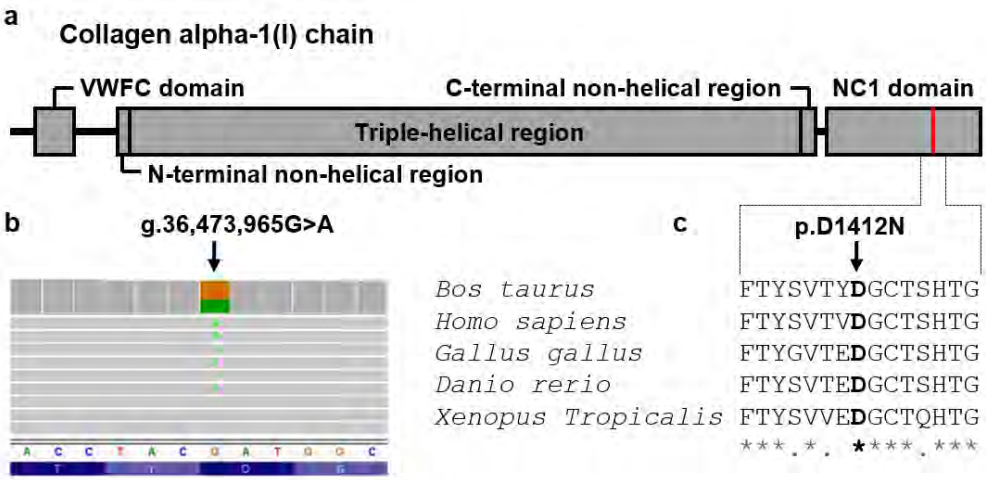


Figure 2. Details on the candidate variant. Functional domain of the bovine COL1A1 protein (a), Integrative genomic viewer screenshot showing the substitution at the DNA level (b), and multiple alignment of COL1A1 ortholog proteins from various vertebrate species (c).

Conclusions

In conclusion, we identified a *de novo* deleterious substitution of COL1A1 that: (1) is responsible for type II OI associated with preterm delivery in the progeny of a mosaic bull; and (2) corresponds to a mutation previously described in human patients suffering from the same syndrome. This study highlights the interest of large data sets available in livestock species to characterize genetic defects and in parallel raises the concern of underreporting of genetic defects to dedicated observatories.

Acknowledgements

This research was funded by APIS-GENE Effitness grant. We thank M. Philippe, S. Lesellier (EVOLUTION), and the veterinarians from Andouillé for providing samples and phenotypic information.

References

Bardai G., Moffatt P., Glorieux F.H., and Rauch F. (2016) Osteoporos Int. 27(12):3607–3613 https://doi.org/10.1007/s00198-016-3709-1

Besio R., Chow C-W., Tonelli F., Marini J.C., and Forlino A. (2019) FEBS J. 286(15):3033-3056. https://doi.org/10.1111/febs.14963

Bodian D.L., Chan T-F., Poon A., Schwarze U.,Yang K. *et al.* (2009) Hum Mol Genet. 18(10):1893-1895. https://doi.org/10.1093/hmg/ddp083

Bourneuf E., Otz P., Pausch H., Jagannathan V., Michot P. *et al.* (2017) Sci Rep. 7(1):11466. https://doi.org/10.1038/s41598-017-11523-3

Cole W.G., Patterson E., Bonadio J., Campbell P.E. and Fortune D.W. (1992) J Med Genet. 29(2):112-8. https://doi.org/10.1136/jmg.29.2.112

Hayes B.J., Daetwyler H.D. (2019) Ann Rev Anim Biosci. 7(1): 89-102. https://doi.org/10.1146/annurev-animal-020518-115024

Jacinto J.G.P., Häfliger I.M., McEvoy F.J., Drögemüller C., and Agerholm J.S. (2021) Animals 11(2):561. https://doi.org/10.3390/ani11020561

Marini J.C., Forlino A., Cabral W. A., Barnes A.M., San Antonio J.D. *et al.* (2007) Hum Mutat. 28(3):209-21. https://doi.org/10.1002/humu.20429

Mesbah-Uddin M., Hoze C., Michot P., Barbat A., Lefebvre R., *et al.* (2019) J Dairy Sci. 102(7):6340–6356. https://doi.org/10.3168/jds.2018-16100

ONAB https://doi.org/10.15454/BRKOV3 Available at: www.onab.fr

Petersen J.L., Tietze S.M., Burrack R.M., and Steffen D.J. (2019) Mamm Genome 30(3-4):81-87. https://doi.org/10.1007/s00335-019-09794-4

Sillence D.O., Senn A., and Danks D.M., (1979) J Med Genet. 16(2):101-16. https://doi.org/10.1136/jmg.16.2.101

522. CRISPR-Cas9 gene editing enables functional evaluation of a host *Synaptogyrin-2* missense variant in PCV2b infection

L.R. Walker*, H. Vu and D.C. Ciobanu

Animal Science Department, University of Nebraska-Lincoln, 3940 Fair St, Lincoln, NE 68583, USA; lianna.walker@huskers.unl.edu

Abstract

Porcine circovirus type 2 (PCV2) is a prevalent DNA virus affecting swine populations worldwide. Previously, we identified a missense variant within the host *Synaptogryin-2* gene (*SYNGR2 p.Arg63Cys*) significantly associated with PCV2b viral load. In this study, we successfully generated a porcine kidney 15 (PK15) clone homozygous for the 'favourable' *SYNGR2 p.63Cys* allele using *in vitro* CRISPR-Cas9 gene editing. Infection of this edited PK15 clone (*emSYNGR2*$^{+p.63Cys}$), demonstrated a significant reduction in total PCV2b replication starting at 48 hours post infection compared to wildtype PK15 cells (*wtSYNGR2*$^{+p.63Arg}$, $P<0.01$). Investigation into the fractional distribution (cell/supernatant) of total PCV2b copies revealed differences in infection dynamics between the variant cell lines across timepoints post infection. Based on these findings, we propose that the *SYNGR2 p.Arg63Cys* variant functions in the early stages of PCV2b infection, specifically in the endocytosis and/or translocation of virions to the nucleus for replication.

Introduction

Viral diseases continuously threaten the sustainability of livestock industries worldwide due to negative effects on growth and animal welfare. As our ability to investigate complex traits at the genomic level has improved, interest in identifying host variants involved in disease resistance across livestock species has also grown. Previously, we identified a missense variant within the host *Synaptogryin-2* gene (*SYNGR2 p.Arg63Cys*) significantly associated with susceptibility to Porcine Circovirus Type 2b (PCV2b), a prevalent DNA virus affecting swine populations across the globe (Walker *et al.* 2018). In this large-scale genome wide association analysis, the *SYNGR2 p.63Cys* allele was deemed 'favourable' and associated with lower PCV2b viral load following *in vivo* experimental challenge (Walker *et al.* 2018). Generation of a Porcine Kidney 15 (PK15) *SYNGR2* knock-out cell line using *in vitro* CRISPR-Cas9 gene editing validated the function of this gene in PCV2b infection (Walker *et al.* 2018). Specifically, *SYNGR2* knock-out resulted in a significant reduction in PCV2b replication by 24 hours post infection (hpi) compared to wildtype PK15 cells ($P<0.05$, Walker *et al.* 2018). However, this does not provide the necessary evidence to validate the function of the associated *SYNGR2 p.Arg63Cys* variant in PCV2b infection. In order to provide direct experimental evidence for the function of this missense variant, cells with alternate homozygous genotypes need to be simultaneously infected with PCV2b. The wildtype PK15 cell line is homozygous for the *SYNGR2 p.63Arg* allele. Therefore, the objective of this research was to generate PK15 clones homozygous for the alternate *SYNGR2 p.63Cys* allele via CRISPR-Cas9 gene editing. By infecting different *SYNGR2* genotypes within the same PK15 genetic background, we can not only validate the *in vivo* association, but better quantify the individual effect of the *SYNGR2 p.Arg63Cys* variant on PCV2b replication.

Materials & methods

PCV2b isolate and cell culture. The PCV2b strain (UNL2014001) utilized in this study was the same isolate used in the prior *in vivo* experimental infection (Walker *et al.* 2018). The PK15 cell line was grown in DMEM high glucose media supplemented with 10% FBS and 1% Penicillin-Streptomycin (5,000 U/ml) at 37 °C with 5% CO_2.

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_522, © L.R. Walker *et al.* 2022

CRISPR-Cas9 gene editing. To generate *SYNGR2 p.Arg63Cys* variant PK15 clones, a single guide RNA was utilized targeting a PAM site overlapping the *SYNGR2 p.Arg63Cys* locus (IDT). The guide RNA was hybridized with fluorescently labelled Alt-R CRISPR-Cas9 tracrRNA ATTO 550 and Alt-R S.p. Hifi Cas9 Nuclease V3 following the manufacturer's protocol to form Ribonucleoprotein (RNP) complexes (IDT). A single stranded DNA (ssDNA) oligonucleotide (80bp) homologous to the non-targeting strand and encoding the alternate *SYNGR2 p.63Cys* allele was used as a template for homology directed repair (IDT). The ssDNA template (3nM) and RNP complexes (10nM guide RNA, 20nM cas9) were reverse transfected into PK15 cells using Lipofectamine RNAiMAX transfection reagent (Invitrogen). After 24 hours post transfection, the cells were collected and sorted using fluorescence-activated cell sorting into 96 well plates to generate single cell clones. Genomic DNA from each clone was extracted using QuickExtract DNA extraction solution (Lucigen) and genotyped for the *SYNGR2 p.Arg63Cys* variant by KASPar assay (LGC). RNA was isolated from selected clones using the RNeasy mini kit (Qiagen) and the full *SYNGR2* transcript was sequenced via Dideoxy sequencing to confirm genotypes.

PCV2b infection and quantification. Wildtype and edited PK15 clones were cultured in 12-well plates (4 cm^2) with 5.0×10^5 cells per well and infected with the PCV2b strain when 80-100% confluent at a multiplicity of infection (MOI)=0.00025. After one hour adsorption, cells were washed with PBS and fresh media was added (DMEM high glucose with 2% FBS). The cells were incubated at 37 °C with 5% CO_2 for up to three days. Control cells were maintained the same way and mock-inoculated with plain DMEM high glucose media. Supernatant and cell samples were collected at specific time points post infection and frozen at -80 °C. Viral DNA was extracted from samples using the QIAamp DNA mini kit (Qiagen). The amount of virus in each sample was determined via real-time quantitative-PCR (RT-qPCR) by plotting the cycle crossing thresholds obtained for technical triplicates on a standard curve generated by serial dilution of a PCV2b standard. The total number of viral copies per well for each time point post infection was calculated by summing the values obtained for the corresponding supernatant and cell fractions.

Results

Following CRISPR-Cas9 gene editing, more than 240 PK15 clones were genotyped for the *SYNGR2 p.Arg63Cys* variant. Of these, ~30 were identified as possible homozygotes for the *SYNGR2 p.63Cys* allele and four were confirmed to have clean sequences with precise edits at only the *SYNGR2 p.Arg63Cys* locus. One of these clones (*emSYNGR2*$^{+p.63Cys}$) was selected for initial comparison of PCV2b replication between *SYNGR2* genotypes. Specifically, wildtype PK15 (*wtSYNGR2*$^{+p.63Arg}$), the *SYNGR2* knock-out PK15 clone (*emSYNGR2*$^{-del}$) previously generated in our lab (Walker *et al.* 2018), and the newly generated *emSYNGR2*$^{+p.63Cys}$ PK15 clone were simultaneously infected with PCV2b in three independent replicates. As expected, the *emSYNGR2*$^{-del}$ cell line exhibited significantly less total viral copies per well than both the *wtSYNGR*$^{+p.63Arg}$ and *emSYNGR2*$^{+p.63Cys}$ cell lines starting at 24 hpi (P<0.01; Figure 1). A significant reduction in PCV2b replication was present in *emSYNGR2*$^{+p.63Cys}$ compared to *wtSYNGR2*$^{+p.63Arg}$ cells starting at 48 hpi (P<0.01; Figure 1). After 72 hpi, the amount of total virus was 140x the starting quantity (0 hpi) in *wtSYNGR2*$^{+p.63Arg}$ cells compared to only 32x in *emSYNGR2*$^{+p.63Cys}$ (P<0.05).

Analysis of the supernatant and cell fractions individually revealed interesting differences in infection profiles. Specifically, in the cell fraction a significant reduction in viral copies is observed at 24 hpi in *emSYNGR2*$^{+p.63Cys}$ compared to *wtSYNGR2*$^{+p.63Arg}$ (P<0.01; Figure 1), whereas, in supernatant the amount of virus is actually slightly higher at 24 hpi in *emSYNGR2*$^{+p.63Cys}$ than *wtSYNGR2*$^{+p.63Arg}$ (P=0.058; Figure 1). In order to further evaluate these differences, the proportion of total virus in each fraction was determined across time points. At 0 hpi, [3]80% of total viral copies are contained within the cell fraction for all three clones (Figure 2). While the proportion of total viral copies in the cell fraction remains fairly constant (>85%) for the *wtSYNGR2*$^{+p.63Arg}$ cells until 72 hpi, both the *emSYNGR2*$^{+p.63Cys}$ and *emSYNGR2*$^{-del}$ clones exhibit

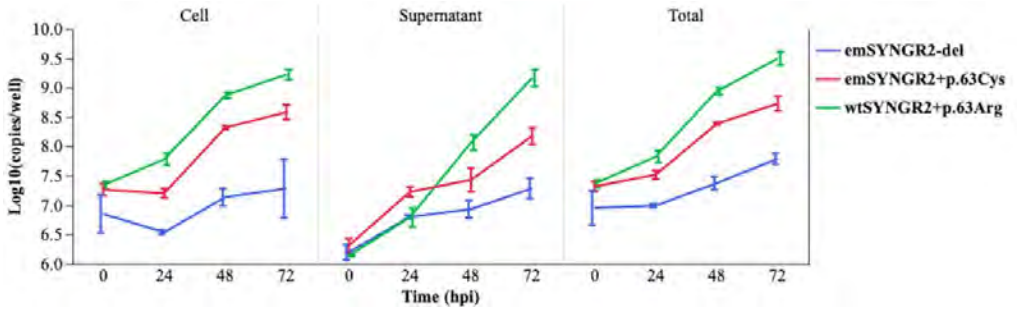


Figure 1. PCV2b copy number across timepoints post infection in PK15 cells with different *SYNGR2* genotypes. The collected cell and supernatant samples from each well were analysed individually (left and middle) and together to assess the total viral copy number per well (right).

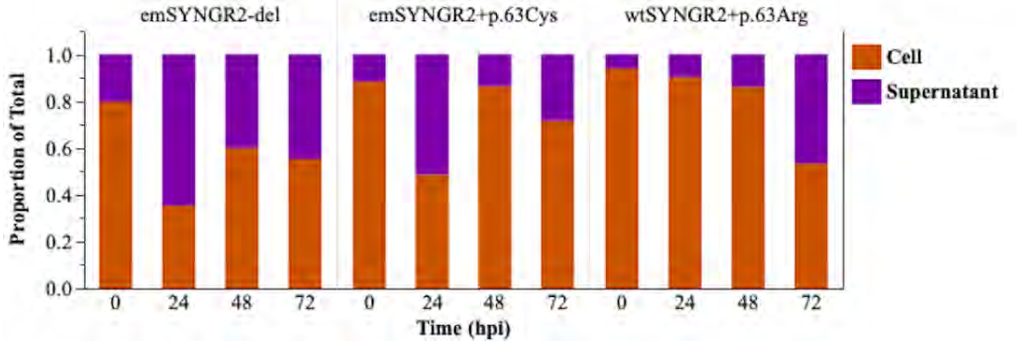


Figure 2. Proportion of total PCV2b copies in cell and supernatant fractions across timepoints post infection in PK15 cells with different *SYNGR2* genotypes.

dramatic changes in the fractional distribution of PCV2b copies (Figure 2). At 24 hpi, only 35 and 49% of total viral copies are present within the cell fraction for the *emSYNGR2*$^{-del}$ and *emSYNGR2*$^{+p.63Cys}$ clones, respectively (Figure 2). Interestingly, the proportion of total virus in each fraction for the *emSYNGR2*$^{+p.63Cys}$ clone 'recovers' back to that exhibited by the *wtSYNGR2*$^{+p.63Arg}$ clone by the 48 hpi (Figure 2), suggesting a potential lag in the early stages of PCV2b infection.

Discussion

In this study, we successfully implemented *in vitro* gene editing and obtained PK15 clones homozygous for the favourable *SYNGR2 p.63Cys* allele. Simultaneous infection of wildtype and edited cell lines provided direct experimental evidence for the role of the *SYNGR2 p.Arg63Cys* variant in PCV2b propagation and validation for the *in vivo* association with PCV2b viral load (Walker *et al.* 2018). Specifically, CRISPR-Cas9 mediated substitution of the *SYNGR2 p.63Arg* allele (*wtSYNGR2*$^{+p.63Arg}$) for the *SYNGR2 p.63Cys* allele (*emSYNGR2*$^{+p.63Cys}$) resulted in a significant reduction in PCV2b replication in PK15 cells. Further investigation into the fractional distribution of total PCV2b copies per well across time points post infection revealed differences in infection dynamics between the variant cell lines. PCV2b appears to still be able to attach and gain cellular entry into *SYNGR2* knock-out cells (*emSYNGR2*$^{-del}$), however, to a drastically lesser degree than in either of the *SYNGR2 p.Arg63Cys* variant cell lines with functional *SYNGR2* proteins. Interestingly, the prominent decrease in proportion of viral copies in the cellular fraction observed for the

emSYNGR2^{-del} clone at 24 hpi is mirrored by the *emSYNGR2*$^{+p.63Cys}$ clone, but not by the *wtSYNGR2*$^{+p.63Arg}$ cell line. Given that PCV2b replication is diminished in both the *emSYNGR2*$^{-del}$ and *emSYNGR2*$^{+p.63Cys}$ clones, the similar decrease in internalized virions observed at 24 hpi may reflect the underlying mechanism in which this gene and variant function in PCV2b infection.

Based on the current understanding and timeline of PCV2 morphogenesis in cell culture, the first 24 hours post infection predominantly corresponds to the first intracytoplasmic phase, in which PCV2 virions are internalized via endocytosis into early endosomes, followed by endosomal escape and translocation of cytoplasmic virions to the nucleus for replication (Rodriguez-Carino *et al.* 2011). *SYNGR2* has been identified as a component of cytoplasmic vesicles (ex. early endosomes) as well as involved in vesicle biogenesis to promote replication of a tick-borne human RNA virus, Severe Fever with Thrombocytopenia Syndrome virus (Sun *et al.* 2016). Additionally, *SYNGR2* has been further characterized as a necessary factor for internalization of a bacterial cytotoxin (Boesze-Battaglia *et al.* 2017, 2020). Therefore, a function of *SYNGR2* predominantly in the early stages of PCV2b infection is a logical conjecture. Specifically, we propose that the decrease in PCV2b replication as a result of *SYNGR2* knock-out or the *SYNGR2 p.63Cys* allele is due to impaired endocytosis and/or slower translocation of virions to the nucleus for replication. Further research will be required to identify the specific underlying mechanism in which the *SYNGR2 p.Arg63Cys* variant functions in PCV2b infection.

References

Boesze-Bataglia K., Walker L.P., Dhingra A., Kandror K., Tang H., *et al.* (2017) Front. Cell. Infect. Microbiol. 7:469. https://doi.org/10.3389/fcimb.2017.00469

Boesze-Bataglia K., Dhingra A., Walker L.M., Zekavat A., Shenker B.J. (2020) Front. Immunol. 11:1262. https://doi.org/10.3389/fimmu.2020.01262

Rodriguez-Carino C., Duffy C., Sanchez-Chardi A., McNeilly F., Allan G.M., *et al.* (2011) Journal of Comparative Pathology 144(2-3):91-102. https://doi.org/10.1016/j.jcpa.2010.07.001

Sun Q., Xian Q., Zhang Y., Wu X., Liang M., *et al.* (2016) JBC Microbiology 291(31):16138-16149. https://doi.org/10.1074/jbc.M116.715599

Walker L.R., Engle T.B., Vu H., Tosky E.R., Nonneman D.J., *et al.* (2018) PLoS Genet 14(10): e1007750. https://doi.org/10.1371/journal.pgen.1007750

523. Sequence-based GWAS meta-analyses for beef production traits

M.P. Sanchez[1*], T. Tribout[1], N. Kadri[2], P.K. Chitneedi[3], S. Maak[3], C. Hozé[1,4], M. Boussaha[1], P. Croiseau[1], R. Philippe[5], M. Spengeler[6], C. Kühn[3], Y. Wang[7], H. Pausch[2] and D. Boichard[1]

[1]Université Paris Saclay, INRAE, AgroParisTech, GABI, 78350 Jouy-en-Josas, France; [2]ETH, 8092 Zürich, Switzerland; [3]Research Institute for Farm Animal Biology (FBN), 18196 Dummerstorf, Germany; [4]Allice, 75595 Paris, France; [5]INRAE, USC1061 GAMAA, Université de Limoges, 87060 Limoges, France; [6]QualitasAG, 6300 Zug, Switzerland; [7]University of Alberta, Edmonton, Canada; marie-pierre.sanchez@inrae.fr

Abstract

Five partners of the H2020 BovReg project have performed sequence-based GWAS for 28 beef production traits (4 growth, 9 morphology, and/or 15 carcass traits) using 54,782 animals from 15 different purebred or crossbred cattle populations. These results were herein combined to conduct 16 different meta-analyses (MA) with both the z-score and the fixed effects MA methods. We identified QTL in 15 MA on BTA2, 5, 6, 7, 10, 11, 13, 14, 15, 17, and 20, most of them being common to several MA. Overall, the fixed effects method outperformed the z-score method in terms of significance level and number of QTL detected. Compared to within-population GWAS, MA found a higher number of QTL in which variants were more frequently located in genes (e.g. *MSTN*, *LCORL, ARRDC3, PLAG1, COL3A1*).

Introduction

When performed at the whole genome sequence (WGS) level, the meta-analysis (MA) of within-population GWAS results can be powerful and accurate to identify causal variants for complex traits (Bouwman *et al.*, 2018). One of the objectives of the H2020 BovReg project is to perform MA at the sequence level for various dairy and beef cattle traits. For beef production, five partners from France (INRAE/ULIM), Switzerland (ETH), Germany (FBN), and Canada (UAL) contributed with 54,782 animals from 15 purebred populations Charolais (CHA), Montbéliarde (MON), Normande (NOR), Limousine (LIM), Blonde d'Aquitaine (BLA), Brown Swiss (BSW), Original Braunvieh (OBR) or crossbred Charolais × Holstein (CH), and Angus, Charolais and beef composite (BP). Each partner conducted sequence-based within-population GWAS for 4 growth, 9 morphology, and/or 15 carcass traits. Here, we combine these GWAS results to conduct 16 MA with fixed effects and z-score methods.

Materials & methods

Within-population GWAS. Within-population GWAS were conducted on 54,782 steers, cows or bulls (19,656 females and 35,126 males) of 15 various populations. Depending on the population, 4 growth, 9 morphology, and/or 15 carcass traits were expressed as yield deviation (YD), daughter YD (DYD), deregressed proof (DRP) or adjusted performance (AP). All partners applied similar imputation and GWAS workflows. Genotypes, aligned on the ARS-UCD1.2 reference genome, were imputed using a stepwise approach: (1) 777K (HD) genotypes were imputed from 50k genotypes with Beagle (Browning and Browning, 2016) or FImpute (Sargolzaei *et al.*, 2014) using animals with HD genotypes as a reference; and (2) sequence variants were imputed using 372 to 3093 animals from the RUN7 or RUN8 reference panel of the 1000 Bull Genomes consortium (Bouwman *et al.*, 2018) with Minimac (Howie *et al.*, 2012) or Beagle. After filtering, allele dosages of between 12.9 and 20.6 million variants were tested for association with different traits in each population separately using GCTA software (Yang *et al.*, 2011) accounting for a polygenic effect estimated from 50k or HD SNPs. DRP and DYD were weighted to account for heterogeneous accuracy (Vandenplas, pers. comm.).

Meta-analyses. We then conducted 16 MA – 2, 5 and 9 for growth, morphology and carcass traits, respectively – combining GWAS results of 1 to 5 traits measured in 3 to 10 different populations from 2 to 5 partners (Table 1). We considered 29.6 million variants imputed by at least two different partners with concordant REF/ALT alleles. For each partner × trait × population combination, we retained variants with a MAF≥0.005 (0.02 for FBN which had a smaller sample size) and with an imputation R^2≥0.20, i.e. 17.9 to 24.9 million variants depending on the MA. Variant effects were standardized by the genetic standard deviation of the trait. The z-score (ZSc) and fixed effects (FE) MA methods implemented in the METAL software (Willer *et al.*, 2010) were applied. For each variant, the ZSc method converts the *P*-value in z-score $Z = S_i z_i w_i / (S_i w_i^2)$ where w_i is the square root of the sample size for study *i* and $z_i = F^{-1}(1 - p_i / 2)\times$(*effect direction for study i*), with p_i the *P*-value of the i^{th} study. The FE method assumes that the true effect of each allele is the same across the different studies and combines effects by weighting them by the inverse of their error variance. Therefore, both MA methods weight the different studies by their sample size. For all GWAS or MA results, we considered an uniform threshold ($-\log_{10}(P) = 8.7$) corresponding to the 5% genome-wide threshold of significance after Bonferroni correction for ~25 million variants.

Results

For 15 of the 16 MA conducted in this study, we found variants with significant effects with both ZSc and FE methods (Table 2). These variants defined QTL regions on BTA2, 5, 6, 7, 10, 11, 13, 14, 15, 17, and 20. The most significant QTL were located on BTA2 and 6 (Figure 1). In 12 of the 15 MA, the FE method identified a higher number of variants with significant effects than the ZSc method. In addition, QTL on BTA6 (C7), 13 (C5), 14 (M2), and 17 (C5) were detected only by the FE approach. In each of the 15 MA, variants with significant effects were also identified in 1 to 5 within-population GWAS. In 7 MA (G2, M1, M2, M5, C4, C5, and C8), the number of variants with significant effects was higher in at least

Table 1. Features of beef production meta-analyses (MA).

MA	Trait group	Traits[1]	Breeds (# populations)	#animals
G1	growth	W15/W18/ADG	CH/CHA/BP/LIM/BLA (7)	18,774
G2	growth	BW	CH/CHA/LIM/BLA (5)	2,720
M1	morphology	MS30/THIGHS/CC	CHA/MON/NOR/LIM/BLA (6)	17,418
M2	morphology	MS30/WITHER/CC	CHA/MON/NOR/LIM/BLA (6)	17,418
M3	morphology	LL	CH/CHA/LIM/BLA (5)	3,695
M4	morphology	WT	CH/CHA/LIM/BLA (5)	3,695
M5	morphology	SS30/SD	CHA/LIM/BLA (4)	12,140
C1	carcass	CW	CH/OBR/BSW/CHA/MON/NOR/BP (7)	19,989
C2	carcass	AS	INRAE/CHA/LIM/BLA (6)	12,208
C3	carcass	CY	CH/CHA/LIM/BLA (5)	3,694
C4	carcass	CG/LMY/MT/CC	CH/OBR/BSW/MON/NOR/BP/CHA/LIM/BLA(10)	25,367
C5	carcass	FS/ABT/FC6/FCU/CF	CH/OBR/BSW/NOR/BP/CHA/LIM/BLA (8)	14,622
C6	carcass	WS	CH/CHA/LIM/BLA (5)	2,636
C7	carcass	ALT	CH/CHA/LIM/BLA (5)	3,692
C8	carcass	IFW	CH/CHA/LIM/BLA (5)	3,686
C9	carcass	REA	CH/BP (3)	4,453

[1] Weight at birth (BW), at month 15 (W15), at month 18 (W18); average daily gain (ADG); muscularity score (MS30), skeletal (SS30) score, thickness of bones (TB30) at month 30; THIGHS; WITHER; fat score (FS); leg length (LL); maximum width of the thigh (WT); skeletal development (SD); carcass conformation (CC), weight (CW), yield (CY), grade (CG) and fat score (CF); age at slaughter (AS); weight at slaughter (WS); lean meat yield (LMY); meatiness (MT); average backfat thickness (ABT); fat content of the 6th rib (FC6); ultrasound fat content (FCU); area of *longissimus thoracis* (ALT); internal fat weight (IFW); rib eye area (REA).

Table 2. Within population GWAS and MA results.

	GWAS			Z-score MA			Fixed effects MA		
MA	# GWAS with sign. variants (min-max)	# BTA with sign. variants[1]	-log$_{10}$(P) max	# sign. variants	-log$_{10}$(P) max	# BTA with sign. variants	# sign. variants	-log$_{10}$(P) max	# BTA with sign. variants
G1	2 (300-1,006)	6	35.7	1,656	42.0	6	2,456	45.5	6
G2	1 (1,354)	1	18.7	314	17.8	1	1,247	22.0	1
M1	4 (116-19,058)	1	154.0	6,384	165.5	2	5,667	165.1	2
M2	4 (58-9,529)	1	154.0	5,710	165.5	2	5,197	165.1	3
M3	4 (88-1,027)	1	15.8	2,334	31.9	1	3,198	37.1	1
M4	4 (36-508)	2	15.6	1,953	23.4	2	2,269	22.7	2
M5	3 (46-3,527)	4	27.9	2,957	28.2	4	3,249	29.4	4
C1	1 (581)	3	23.5	461	29.4	4	740	31.4	4
C2	0	0	-	0	-	0	0	-	0
C3	4 (201-2,575)	2	30.4	3,977	48.5	1	4,899	48.0	1
C4	5 (88-1,120)	3	32.8	249	33.8	4	1,004	42.0	4
C5	3 (21-446)	4	13.8	138	15.8	3	192	14.7	5
C6	3 (57-741)	1	13.4	1,308	25.8	1	1,852	29.2	1
C7	3 (9-110)	1	11.8	1,275	24.9	1	1,724	25.8	2
C8	1 (189)	1	9.7	28	11.2	1	8	9.3	1
C9	3 (24-436)	2	15.0	722	19.8	2	753	22.8	2

[1] In at least one GWAS.

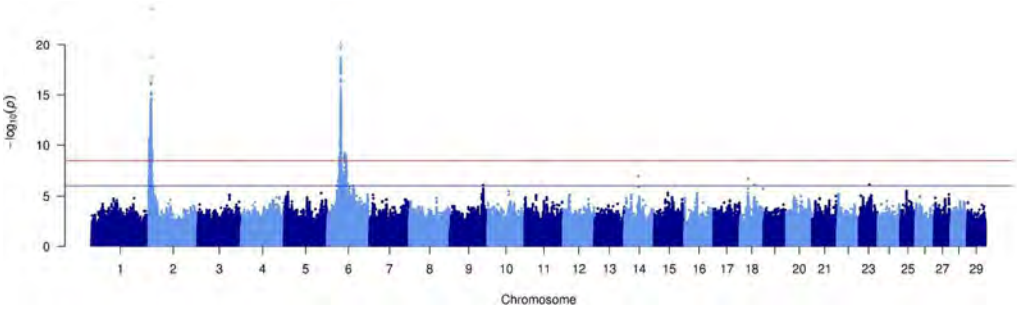


Figure 1. Manhattan plot obtained with M4 fixed effects MA results.

one elementary GWAS than in any of the MA methods used. However, compared to within-population GWAS, MA revealed: (a) more significant effects; (b) novel QTL on BTA2 (C1), 5 (G1), 6 (C7), 7 (M1 and M2), 10 (C1 and C4), 14 (M2), and 17 (C5); and (c) variants with the most significant effects in the QTL peaks (TOP1) more frequently located in genes (72% in fixed effects MA vs 62% in GWAS). For the most significant QTL identified on BTA2 in M1, M2, M4, and C4 MA, the TOP1 variant is a stop gained variant (rs110344317, Q204x) within the *MSTN* gene. In 10 MA (G1, G2, M3, M4, M5, C1, C4, C5, C6, and C9), we found a QTL on BTA6 with 7 unique TOP1 variants located within or near to the *LCORL* gene. For

other QTL, the variants with the most significant effects were located in *ANKAR*, *ASNSD1*, *SLC40A1*, and *COL3A1* (BTA2); *CCND2* (BTA5); *MLLT1*, *PIAS4*, and *ARRDC3* (BTA7); *GLDN* and *CYP19A1* (BTA10); *PLAG1* (BTA14); and *CHST1* and *SLC35C1* (BTA15).

Discussion

This study, conducted on imputed WGS of 54,782 animals from 15 populations of various breeds, is the first meta-analysis of this magnitude dedicated to cattle beef production. We show the value of MA, in complement to within-population GWAS, in identifying: (1) a larger number of QTL; (2) a lower number of variants in QTL; and (3) candidate variants located more frequently in genes. The use of more animals and the conservation of LD over shorter distances across breeds may explain the superiority of MA over GWAS in terms of power and mapping precision. By applying here the most commonly MA methods used for GWAS, i.e. ZSc and FE approaches, we confirm that the FE method appears more powerful in detecting QTL (Begum *et al.*, 2012), although MA combine substantially different traits in the present study. In several regions, MA directly pointed out variants in genes, including *MSTN*, *LCORL*, *ARRDC3*, and *PLAG1*, previously associated with morphology and carcass traits in various studies (e.g. Wang *et al.*, 2020). For example, the Q204X mutation, ranked 1st in the QTL peaks at the proximal end of BTA2 and causing a premature stop codon in the gene encoding myostatin (*MSTN*), was reported as one of the polymorphisms responsible for the double-muscled phenotype in several cattle breeds (Grobet *et al.*, 1998). We also identified dozens of other variants located in genes having a function that may be related to meat production traits (e.g. *COL3A1 collagen type III alpha 1 chain*). By better identifying genes and candidate causative variants associated with beef production traits in cattle, MA appears to be of great interest to decipher the biological mechanisms underlying these traits.

Acknowledgements

The authors thank Dr J. Vandenplas (WUR) for the modification of GCTA to account for weights. The BovReg project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 815668.

References

Begum F., Ghosh D., Tseng G., and Feingold E. (2012) Nucleic Acids Res 40(9):3777-3784. https://doi.org/10.1093/nar/gkr1255

Bouwman A. C., Daetwyler H. D., Chamberlain A. J., Ponce C. H., Sargolzaei M. *et al.* (2018) Nat Genet 50(3):362-367. https://doi.org/10.1038/s41588-018-0056-5

Browning B. and Browning S. (2016) Am J Hum Genet 98(1):116-126. https://doi.org/10.1016/j.ajhg.2015.11.020

Grobet L., Poncelet D., Royo L., Brouwers B., Pirottin D. *et al.* (1998) Mamm Genome 9(3):210-213. https://doi.org/10.1007/s003359900727

Howie B., Fuchsberger C., Stephens M., Marchini J., and Abecasis G. R. (2012) Nat Genet 44(8):955-959. https://doi.org/10.1038/ng.2354

Sargolzaei M., Chesnais J., and Schenkel F. (2014) BMC Genomics 15(1):478. http://doi.org/10.1186/1471-2164-15-478

Wang Y., Zhang F., Mukiibi R., Chen L., Vinsky M. *et al.* (2020) BMC Genomics 21(1). https://doi.org/10.1186/s12864-019-6273-1

Willer C., Li Y., and Abecasis G. (2010) Bioinformatics 26(17):2190-2191. https://doi.org/10.1093/bioinformatics/btq340

Yang J., Lee S., Goddard M., and Visscher P. (2011) Am J Hum Genet 88(1):76-82. https://doi.org/10.1016/j.ajhg.2010.11.011

524. Deciphering genetic variants from whole genome affecting duodenum, liver and muscle transcriptomes in pigs

D. Crespo-Piazuelo[1*], O. González-Rodríguez[1], M. Mongellaz[2], H. Acloque[2], M.-J. Mercat[3], M.C.A.M. Bink[4], A.E. Huisman[4], Y. Ramayo-Caldas[1], J.P. Sánchez[1] and M. Ballester[1]

[1]Animal Breeding and Genetics Program, IRTA, Torre Marimon, 08140 Caldes de Montbui, Spain; [2]INRAE GABI, Domaine de Vilvert, 78350 Jouy-en-Josas, France; [3]IFIP-Institut du porc and Alliance R&D, La Motte au Vicomte, 35651 Le Rheu, France; [4]Hendrix Genetics, P.O. Box 114, 5830 AC Boxmeer, the Netherlands; daniel.crespo@irta.cat

Abstract

With the aim of deciphering genetic variants affecting gene expression levels of duodenum, liver and muscle, the DNA and RNA of 300 pigs from three different breeds (Duroc, Landrace and Large White) was extracted and sequenced. After filtering and normalizing both the genomic and the transcriptomic datasets, expression genome-wide association studies (eGWAS) were conducted between 25,315,878 polymorphisms and the expression of 16,753 genes in duodenum, 15,710 genes in liver, and 13,887 genes in muscle. In total, more than 1.17×10^{12} combinations were performed, obtaining 19,926,590 significantly associated polymorphisms, which were grouped in 148,639 expression quantitative trait locus (eQTL) regions. Half of the polymorphisms were located in close proximity to their associated gene, but only 7,773 *cis*-eQTL regions were defined, the remaining being *trans*-eQTLs regions. Out of the cis-eQTL regions, 18 shared the same top polymorphism among the three tissues; thus, they were considered the strongest candidates for gene expression regulation.

Introduction

Rather than the usual measured phenotypes, gene expression acts as an 'intermediate phenotype', which is expected to be closer to the genomic information than a conventional end-trait phenotype (Stranger *et al.*, 2007). In this regard, genetic variants affecting gene-expression phenotypes are an important source of phenotypic variation. Over the last decade, expression genome-wide association studies (eGWAS) have become one of the most used methods to study the association of the polymorphisms distributed across the genome and the expression levels of a gene. At the time, this method used a limited dataset of genetic markers, but nowadays all the genetic variants found in the genomes of a population can be included, thanks in particular to the progress made to improve computing efficiency. In addition, RNA sequencing approaches have become cheaper and can be applied to larger datasets. This is crucial to understand the relationship between gene regulation mechanisms and the traits of interest, since more than 90% of the phenotype-associated polymorphisms are located within intergenic and intronic regions, rather than on the coding regions of a gene (Maurano *et al.*, 2012). When a location in the genome is found in association with the expression of a certain gene, an expression quantitative trait locus (eQTL) region is defined. Classically, *cis*-eQTL regions have been defined at 1 Mb of the gene analysed; otherwise, they are considered *trans*-eQTL regions. In this study we will aim to decipher potential genetic variants that could modify gene expression levels in duodenum, liver and muscle.

Materials & methods

Duodenum, liver and muscle samples were collected at slaughter from 300 pigs of three different breeds (n=100 Duroc, n=100 Landrace and n=100 Large White). Genomic DNA was extracted from blood (Duroc and Landrace) and liver (Large White) samples using NucleoSpin Blood kit (Macherey-Nagel, Düren, Germany). RNA from muscle samples was extracted using RiboPure™ RNA Purification Kit (Invitrogen, Carlsbad, CA, USA) and RNeasy Fibrous Tissue Mini Kit (Qiagen, Hilden, Germany), and RNA from

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_524, © D. Crespo-Piazuelo *et al.* 2022

liver and duodenum samples was extracted using a chemagic™ 360 instrument with RNA Tissue10 Kit H96 (PerkinElmer, Baesweiler, Germany). The whole genome of the 300 pigs and the transcriptome from the three tissues (duodenum, liver and muscle; n=900) were paired-end (2×150bp) sequenced in an Illumina NovaSeq6000 platform (Illumina, San Diego, CA, USA). DNA sequences were mapped against the reference genome (Sscrofa11.1 assembly) with BWA-MEM/0.7.17 (H. Li, 2013) and the genetic variants were obtained with GATK/4.1.8.0 HaplotypeCaller (McKenna *et al.*, 2010). Then, the genetic variants were filtered (minor allele frequency below 5% and/or more than 10% missing genotype data) for downstream analyses.

RNA sequences were mapped against the reference genome (Sscrofa11.1 assembly) with STAR/v2.5.3a (Dobin *et al.*, 2013), counts were quantified by RSEM/1.3.0 (B. Li and Dewey, 2011) and normalized by TMM (trimmed mean of M-values). Genes with low expression (counts per million below 10/minimum library size in millions) and those that did not show expression in at least 5% of the animals were removed. Thereafter, eGWAS were conducted between the filtered polymorphisms and the normalized expression data using the fastGWA tool from GCTA/1.93.2 (Yang *et al.*, 2011) using the following model:

$$y_{hijk} = sex_{hj} + breed_{hk} + u_{hi} + s_{il} \cdot a_{hl} + e_{hijk}$$

where y_{hijk} corresponds to the expression of the hth gene in the ith individual of sex j and belonging to the kth breed; sex_{hj} corresponds to the jth sex effect (two levels); $breed_{hk}$ to the kth breed effect (three levels); u_{hi} is the infinitesimal genetic effect of the individual i, with u~MVN(0,G·σ^2_u), where G is the genomic relationship matrix calculated using the filtered autosomal polymorphisms and σ^2_u is the additive genetic variance to be estimated; s_{il} is the genotype (coded as 0, 1 or 2) for the lth polymorphism; and a_{hl} is the allele substitution effect of the lth polymorphism on the expression level of the hth gene; finally, e_{hijk} is the residual term. Bonferroni correction was applied to assess the statistical significance for the association studies at genome-wide level with the p.adjust function of R. Only those associations with an adjusted *P*-value<0.05 were considered significant.

Results

The whole genome sequence of the 300 pigs resulted in 44,127,400 genetic variants, which after the filtering step, got reduced to a total of 25,315,878 polymorphisms that were kept for the association analyses. Regarding the type of genetic variant, the polymorphisms were classified as SNPs (74.92%), insertions (13.88%) and deletions (11.20%). After normalizing and filtering the three transcriptomic datasets, 12,892 genes were expressed in all the three tissues, whereas 18,097 genes were expressed in total. For each tissue, eGWAS were conducted for 16,753 genes in duodenum, 15,710 genes in liver, and 13,887 genes in muscle. The eGWAS resulted in a total of 19,926,590 significant associations among the three tissues, which are detailed in Table 1. From all the significantly associated polymorphisms in the three tissues, 29.9-30.2% were novel variants and the remaining 69.8-70.1% were already described in the Ensembl database. In this regard, a polymorphism could be associated with the expression of more than one gene. Thus, in duodenum, 2,813,177 unique polymorphisms were associated with the expression of 6,551 genes; in liver, 4,187,249 unique polymorphisms were associated with the expression of 7,433 genes, and in muscle, 4,814,732 polymorphisms were associated with the expression of 7,496 genes. Out of these genes with any significant association 1,730 were expressed in the three tissues.

Table 1. Number of significantly associated variants and eQTL regions per tissue.

	Significantly associated variants		eQTL regions	
Tissue	**No.**	***cis*-regulatory elements[1]**	**No.**	***cis*-eQTL regions[2]**
Duodenum	4,802,045	2,260,300 (47.1%)	53,236	1,866 (3.5%)
Liver	7,024,941	3,756,049 (53.5%)	36,835	2,704 (7.3%)
Muscle	8,099,604	4,461,375 (55.1%)	58,568	3,203 (5.5%)
Total	19,926,590	10,477,724 (52.6%)	148,639	7,773 (5.2%)

[1] A *cis*-regulatory element was considered if the polymorphism was located at a maximum distance of 1 Mb from the associated gene.

[2] A *cis*-eQTL region was considered if the gene was located within it.

An eQTL region was defined ±1 Mb from a significantly associated polymorphism, which were merged into the same eQTL region if they intersected. Thus, 148,639 eQTL regions were defined (Table 1). However, this number of eQTL regions should be carefully considered as 43.52-59.82% were defined exclusively by a single polymorphism: 23,167 in duodenum, 21,828 in liver, and 35,033 in muscle. In proportion, the ratio of *cis*-eQTL regions defined by a single polymorphism was much lower (5.62-7.61%): 142 in duodenum, 152 in liver, and 191 muscle. For each eQTL region, most of the significant polymorphisms were located within 1 Mb of the top polymorphism of the region (Figure 1).

Regarding the location of the polymorphisms in respect to the expression of their associated gene, 47.1-55.3% of the polymorphisms were closer than 1 Mb to their gene. Despite so, only ~5% of the 148,639 eQTL regions defined were in *cis* (Table 1). When comparing the top polymorphisms of the eQTL regions found in all the three tissues, 332 genetic variants where shared between the three of them, being 18 of them cis-regulatory elements.

Discussion

In this study, more than 1.17×10^{12} combinations were performed through eGWAS in duodenum, liver, and muscle, obtaining 19,926,590 associated polymorphisms. These polymorphisms were grouped in 148,639 eQTL regions, but despite half of the polymorphisms being located in close proximity to their associated gene, only ~5% of the eQTLs regions were in *cis*. A possible explanation of this result is due to the fact that most of the *trans*-eQTL regions were defined by a single polymorphism. Hence, to avoid such artifacts, filters to define eQTL regions with 2 or more polymorphisms may be required.

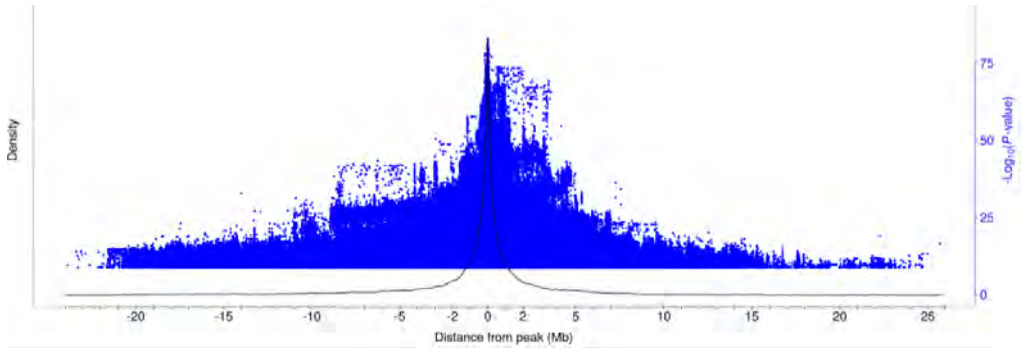


Figure 1. Density plot of the distance of the significant polymorphisms associated with liver gene expression to the peak of their eQTL region (sexual and mitochondrial chromosomes excluded). The significance of the polymorphisms is provided in blue dots.

Considering the distribution of the *P*-values due to linkage disequilibrium, the definition of *cis* elements if they were located at less than 1 Mb from their associated gene seemed appropriate, as they were rarely surpassing the 2 Mb. However, this definition may change depending on the structure of the population used.

One of the top 18 *cis*-regulatory elements of the eQTL regions found in common in the three tissues was a polymorphism located upstream the *SUPT3H* gene, a gene already identified by our group as a potential key regulator for health-related traits in pigs (Crespo-Piazuelo *et al.*, 2021). Five out of the 18 *cis*-regulatory elements were located in long non-coding RNAs (lncRNAs). Of interest was the polymorphism located upstream the *ENSSSCG00000041410* gene, which encodes for a lncRNA that was positively correlated in all the three tissues (r=0.590-0.732) with the expression of *TRIM39*, a nearby gene involved in the immune response and cell cycle progression (Suzuki *et al.*, 2016; Zhang *et al.*, 2012), and that has been already described as participating in the same processes regulated by lncRNAs (Zhou *et al.*, 2019). In conclusion, the identification of genetic variants associated with gene expression levels in the three tissues will contribute to shed light on the molecular mechanisms of regulatory variations to shape end-trait phenotypes.

Acknowledgements

This project is part of GENE-SWitCH (https://www.gene-switch.eu) and has received funding from the European Union's Horizon 2020 Research and Innovation Programme under the grant agreement no. 817998. It is also part of EuroFAANG (https://eurofaang.eu), a synergy of five Horizon 2020 projects that share the common goal to discover links between genotype to phenotype in farmed animals and meet global Functional Annotation of ANimal Genomes (FAANG) objectives.

References

Crespo-Piazuelo D., Ramayo-Caldas Y., González-Rodríguez O., Pascual M., *et al.* (2021). Front. Immunol. 12:784978. https://doi.org/10.3389/fimmu.2021.784978

Dobin A., Davis C.A., Schlesinger F., Drenkow J., *et al.* (2013). 29(1):15–21. https://doi.org/10.1093/bioinformatics/bts635

Li B., and Dewey C.N. (2011). BMC Bioinformatics 12(1):323. https://doi.org/10.1186/1471-2105-12-323

Li H. (2013). https://arxiv.org/abs/1303.3997

Maurano M.T., Humbert R., Rynes E., Thurman R.E., *et al.* (2012). Science 337(6099):1190–1195. https://doi.org/10.1126/science.1222794

McKenna A., Hanna M., Banks E., Sivachenko A., *et al.* (2010). Genome Res. 20(9):1297–1303. https://doi.org/10.1101/gr.107524.110

Stranger B.E., Nica A.C., Forrest M.S., Dimas A., *et al.* (2007). Nat. Genet. 39(10):1217–1224. https://doi.org/10.1038/ng2142

Suzuki M., Watanabe M., Nakamaru Y., Takagi D., *et al.* (2016). Cell. Mol. Life Sci. 73(5):1085–1101. https://doi.org/10.1007/s00018-015-2040-x

Yang J., Lee S.H., Goddard M.E., and Visscher P.M. (2011). Am. J. Hum. Genet. 88(1):76–82. https://doi.org/10.1016/j.ajhg.2010.11.011

Zhang L., Mei Y., Fu N. -y., Guan L., *et al.* (2012). Proc. Natl. Acad. Sci. 109(51):20937–20942. https://doi.org/10.1073/pnas.1214156110

Zhou Y., He L., Liu X.-D., Guan H., *et al.* (2019). Biomed Res. Int. 2019:6305065. https://doi.org/10.1155/2019/6305065

525. Detailed analysis of mortality rates in the female progeny of Holstein bulls allows the discovery of new dominant genetic defects

F. Besnard[1,2*], H. Leclerc[2,3], N. Cesbron[4], M. Boussaha[2], C. Grohs[2], N. Jewell[5], A. Pinton[6], H. Barasc[6], M. Femenia[2], L. Dorso[4], B. Strugnell[7], T. Floyd[8], C. Danchin[1], R. Guatteo[4], X. Hubin[9], S. Mattalia[1,2], D. Boichard[2] and A. Capitan[2,3]

[1]IDELE, 75012 Paris, France; [2]Université Paris-Saclay, INRAE, AgroParisTech, GABI, 78350 Jouy-en-Josas, France; [3]ALLICE, 75012 Paris, France; [4]INRAE, Oniris, BIOEPAR, 44307 Nantes, France; [5]Animal and Plant Health Agency, Carmarthen Veterinary Investigation Centre, Carmarthen, SA31 3EZ, United Kingdom; [6]GenPhySE, Université de Toulouse, INRAE, ENVT, 31326 Castanet Tolosan, France; [7]Farm Post Mortems Ltd., Hamsterley, Bishop Auckland, DL13 3QF, United Kingdom; [8]Animal and Plant Health Agency, Weybridge, KT15 3NB, United Kingdom; [9]AWE, 5590 Ciney, Belgium; florian.besnard@idele.fr

Abstract

We analyzed juvenile mortality rates per sire family for four different periods in a population of Holstein heifers born between 2017 and 2020 in France and Wallony. The global mortality rate from birth to one year of life was 11.8% in the population of 1,001 sires, while it reached 30.6 and 19.9% in the progeny of two bulls, Mo and Pa, respectively. Various investigations were carried out on these two bulls and their progeny, including clinical examination, karyotyping, genetic mapping, and whole genome sequencing. Mo was found to carry a partial translocation from BTA29 to BTA26 while Pa was mosaic for a dominant nonsense mutation in the *GATA6* gene causing cardiac outflow tract defect in its progeny.

Introduction

Reducing juvenile mortality in cattle is important for both economic and animal welfare reasons. Previous studies have revealed a large variability between breeds and sire progeny groups (e.g. Leclerc *et al.* 2016), with some extreme cases due to dominant mutations causing various syndromes among the descendants of mosaic bulls (Bourneuf *et al.* 2017). The purpose of this study was twofold: to fine monitor calf mortality at the level of sire families within the French and Walloon Holstein populations, and to use this information to detect genetic defects that might have been overlooked by lack of specific symptoms.

Materials & methods

Analysis of mortality rates at different ages in the progeny of individual sires. Information on the pedigree, sex, date of birth, date of death, and cause of death (either natural or slaughter) of Holstein animals were recovered from the French and Walloon databases. To focus on the most reliable data, only female calves born between 2017 and 2020, which remained on their farm of birth during their first year of life, were selected. The final data set comprised 2.25 million daughters from 1,001 sires. The size of the half-sib families ranged from 100 to 37,839 individuals, with a mean of 2,246. Natural mortality rate was computed during the first year of life and for four sub-periods (perinatal, postnatal, pre-weaning and post-weaning) corresponding to distinct causes of death (Table 1; Santman-Berends *et al.*, 2019).

Clinical examination. Nine calves (both sexes) sired by two half-brother bulls (Mo: n=3 and Pa: n=6) that showed very high mortality rates were necropsied. Gross phenotypic description was also available for eight and seven additional presumably affected calves of each sire, respectively.

Molecular characterization of the genetic defects of sires Mo and Pa. Mo was karyotyped to screen for chromosomal abnormality while Pa was dead at time of the study and thus not available for blood sampling. Illumina Bovine SNP50 phased and imputed genotypes from Pa and 203 of its progeny were

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_525, © F. Besnard *et al.* 2022

extracted from the French genomic evaluation database. Assuming a dominant inheritance with somatic mosaicism in Pa, we performed transmission disequilibrium tests for 16,487 informative markers, 14 progeny that died before 50 days of age (incl. four necropsied), and 189 half-sib controls still alive at 2 years. We then sequenced the genome of one case with Illumina technology (150 bp paired-end reads), processed the data in accordance with the guidelines of the 1000 Bull Genomes Project (Hayes and Daetwyler 2019), and we used 5,116 genomes from run 9 of the latter project as controls.

Results & discussion

The natural mortality rate of heifers during their first year of life was of 11.8% on average, with 4.2% for perinatal, 2.9% for postnatal, 3.1% for pre-weaning and 3.2% for post-weaning mortalities (Table 1, Figure 1). These rates were lower than most of those reported in the literature (e.g. Leclerc *et al.* 2016), probably because we considered only females and sex is known to have a significant effect on juvenile mortality. Interestingly, natural mortality rates per period and per half-sib family showed approximately normal distribution suggesting quantitative inheritance. Yet we observed outlier families with possible mono or oligogenic inheritance of excess mortality and focused on the worst sire per category.

Among them, two bulls with high mortality rates (Mo and Pa), sired by the most popular bull sire of the cohort (Mogul) particularly retained our attention and were selected for further analysis. Karyotyping of Mo revealed a translocation of approximately the first half of bovine chromosome 29 (BTA29) onto

Table 1. Mortality rates in % for the worst bull per category (bolded) vs population means.

Bull	Bull sire	NB of calves	Perinatal	Postnatal	Pre-weaning	Post-weaning	Global mortality
			0-2 days	3-14 days	15-55 days	55-365 days	1-365 days
Mean of 1,001 bulls	2,250	4.2	2.9	3.1	3.2	11.8	
Mo	Mogul	108	**16.7**	**6.7**	2.4	**7.3**	**30.6**
Pa	Mogul	2,297	5.3	6.1	**6.3**	4.8	19.9

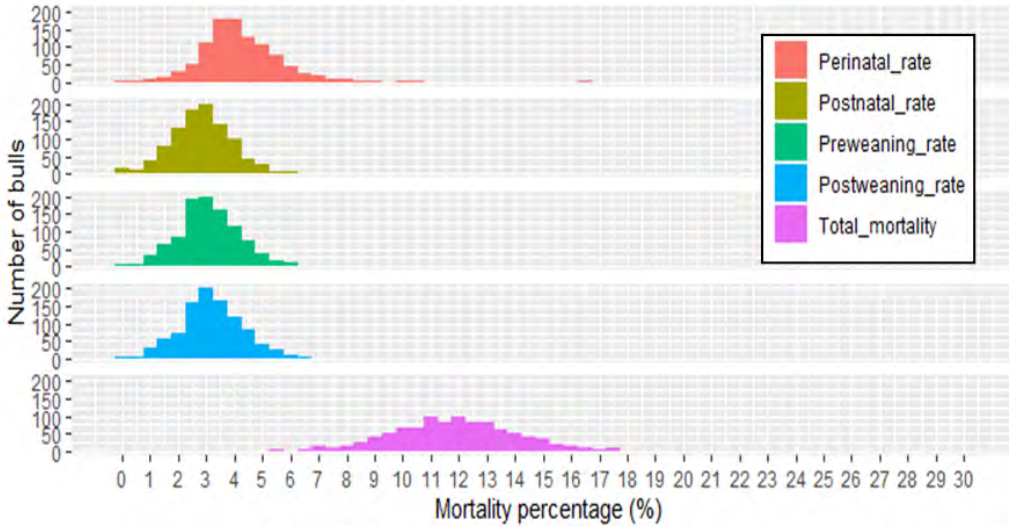


Figure 1. Histogram of distribution of mortality rates for each period of death.

BTA26 (Figure 2A). A survey by French and Belgian veterinarians enabled us to gather phenotypic information for six affected calves. The main clinical features included symptoms reminiscent of Jacobsen syndrome in human (i.e. cranial dimorphism, seizure and heart defects), a rare genetic defect caused by haploinsufficiency for more than 100 genes (Rodríguez-López *et al.* 2021) located in a region of HSA11q orthologous to the translocated segment of BTA29 (Figure 2B). In the near future, the genotyping of the affected calves with SNP arrays should allow us to confirm the diagnostic of partial monosomy of BTA29 and to precise the breakpoints of the translocation.

A different etiology was suspected for the excess of mortality observed among Pa's progeny, since the peak of mortality was later than for Mo's daughters and the symptoms were different. Indeed, necropsy of six affected calves revealed persistent *truncus arteriosus* (Figure 3A) and symptoms compatible with heart defects in seven additional cases. To gain insights into the molecular basis of this defect we analyzed Illumina Bovine SNP50 genotypes from 14 calves that died prematurely and 189 half-sib controls that had been genotyped for genomic evaluation. In doing so, we mapped the locus on BTA24 between positions 19,770,957 and 37,706,580 on ARS-UCD1 assembly. Subsequent sequencing of the genome of a necropsied calf and filtering for variants observed in 5,116 control genomes revealed a single heterozygous candidate variant within the interval: a thymine-to-adenine substitution in exon 2 of *GATA6* predicted to introduce a premature stop codon (chr24: g.34,187,181T>A; GATA6 p.417K>X). If translated the mutant protein would lack about 30% of amino acid sequence including important functional domains (ZF2, NLS and TAD3; Figure 3). Interestingly the loss of GATA6 heterozygosity has been associated with various defects of differentiation, development, and function of the heart in vertebrates, including abnormal septation of the aortopulmonary trunk (e.g. Kodo *et al.* 2009 and Figure 3B).

PCR and Sanger sequencing of the sire Pa, three affected calves and three control carrying the same paternal haplotype but in the non mutated version, confirmed the de novo nature of the nonsense variant. Finally, analyzing segregation distorsion of adjacent markers within the control calves (57:132 ratio) we estimated a proportion of 28.4% of mutants in the progeny of Pa.

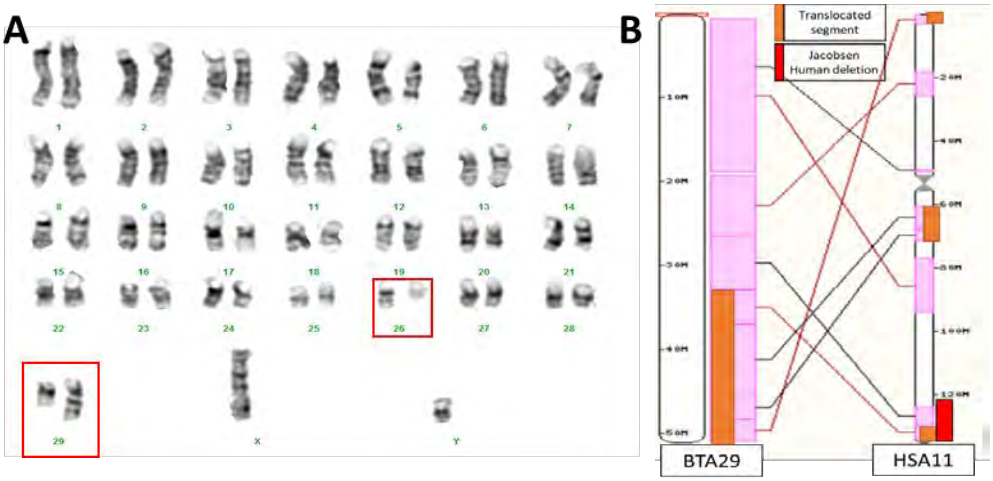


Figure 2. Karyotype of sire Mo (A) and synteny between BTA 29 and HSA11 (B).

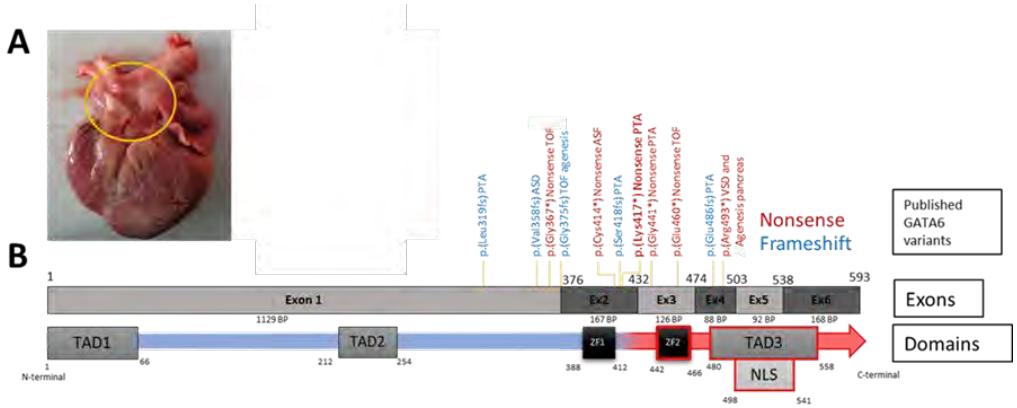


Figure 3. Bovine heart with persistent *truncus arteriosus* (A) and scheme of GATA6 with bovine mutation p.417K>X and deleterious variants reported in human (B).

Conclusions

By focusing on the progeny of extreme bulls for juvenile mortality, we have described bovine models for two syndromes also observed in humans. Our results demonstrate the suitability of our method to reveal genetic defects that are hardly detectable with traditional heredo-surveillance. Beyond this proof of concept, the calculation of mortality rates at different ages for the whole population of bulls pave the way for future detection of QTL associated with juvenile mortality.

Acknowledgements

F. Besnard is supported by a CIFRE PhD grant from IDELE, with the financial support of the Association Nationale de la Recherche et de la Technologie and APIS-GENE (Paris, France).

References

Bourneuf E., Otz P., Pausch H, Jagannathan V., Michot P., *et al.* (2017) *Sci Rep* 7 (1): 11466. https://doi.org/10.1038/s41598-017-11523-3

Hayes B.J.and Daetwyler H.D. (2019) *Ann Rev Anim Biosci* 7 (1): 89–102. https://doi.org/10.1146/annurev-animal-020518-115024

Kodo K., Nishizawa T., Furutani M., Arai S., Yamamura E., *et al.* (2009). *Proc Nat Acad Sci USA*, 106(33): 13933–38. https://doi.org/10.1073/pnas.0904744106

Leclerc H., Lefèbvre R., Douguet M., Phocas F.and Mattalia-Elie S. (2016). *23rd Renc Rech Rum, Paris*. https://hal.archives-ouvertes.fr/hal-01608259/

Rodríguez-López, R., Gimeno-Ferrer F., Montesinos E., Ferrer-Bolufer I., Luján C.G., *et al.* (2021) *Genes* 12(8): 1197. https://doi.org/10.3390/genes12081197

Santman-Berends I.M.G.A., Schukken Y.H.and van Schaik G. (2019) *J Dairy Sci* 102(7): 6404–17. https://doi.org/10.3168/jds.2019-16381

526. Candidate genes for milk, growth and immune system traits in Brazilian Iberian-derived locally adapted cattle breeds

L.L. Verardo[1*], M.A. Machado[2], J.C.C. Panetto[2] and M.V.G.B. Silva[2]

[1]Universidade Federal dos Vales do Jequitinhonha e Mucuri, Campus JK, 39100-000 Diamantina, Brazil; [2]Brazilian Agricultural Research Corporation (EMBRAPA), Rua Eugênio do Nascimento, 610, 36038-330 Juiz de Fora, MG, Brazil; lucas.verardo@ufvjm.edu.br

Abstract

Along decades of breeding in a variable ecosystem throughout the country, the first taurine animals arrived in Brazil became adapted to a wide range of environments with different levels of improved fitness. We analyzed three cattle breeds representative of Brazilian Iberian-derived Locally Adapted cattle (Curraleiro Pé-Duro – CUR with 17 animals, and both Caracu lineages, Caracu Caldeano selected for milk – CCD with 55 animals, and Caracu selected for beef – CCB with 24) aimed to evaluate runs of homozygosity (ROH), identify ROH islands and functionally analyse the identified genes. We observed the occurrence of short ROH islands in all breeds, suggesting a successful matting scheme. Genes located in ROH islands were evaluated and explored throughout their biological processes (e.g. *PRLR* related with milk and growth traits in CCD and CCB; and *CAMKK2* related with immune system in CUR) providing information about the genetic architecture of the breeds.

Introduction

The first taurine animals were introduced in Brazil between sixteenth and seventeenth centuries. Along decades of breeding, these animals adapted to environments with different levels of improved fitness. They become recognized as Iberian-derived locally adapted breeds (Egito *et al.*, 2007; McManus *et al.*, 2009), such as the Caracu lineages (Caracu Caldeano, selected for dairy purpose, and Caracu, selected for beef) and Curraleiro Pé-Duro.

Runs of homozygosity (ROH) patterns may be helpful in understand the history of the population, breeding system, and pattern of geographic subdivision. ROH are continuous homozygous segments of DNA (Gibson *et al.*, 2006), and calculation of that is essential to identify islands and genes with a role in the population. The ROH islands shared among individuals, may be used to identify homozygous regions that result from selection events (Zhang *et al.*, 2015). Moreover, genes identified in those regions can be analysed through gene networks (Verardo *et al.*, 2021) aiming to a better understanding of their biological role in each population.

In this study, we assessed three cattle breeds representative of Brazilian Iberian-derived Locally Adapted cattle (Curraleiro Pé-Duro – CUR and both Caracu lineages, Caracu Caldeano selected for milk – CCD, and Caracu selected for beef – CCB) aimed to evaluate ROH across populations, identify ROH islands and functionally analyse the identified genes.

Materials & methods

Biological samples were obtained from Brazilian commercial herds, and DNA was extracted at the Embrapa Dairy Cattle Research Center. All animals were chosen according to their population representativeness in each breed, to avoid bias due to the use of a limited number of families in the population (lineages). Animals were genotyped using the Illumina Bovine HD Chip (777k SNPs). SNP genotypes from sexual chromosomes were excluded. Also, genotypes were excluded when the call rate <0.95 or minor allele frequency <0.01. In addition, samples showing overall call rates <0.90 were excluded from further analysis.

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_526, © L.L. Verardo *et al.* 2022

A total of 55, 24 and 17 animals genotyped with 624,609, 674,260 and 612,345 markers from CCD, CCB and CUR, respectively, were analysed after quality control.

The ROH were calculated using the software PLINK v1.07 (Purcell *et al.*, 2007) for each population with the following parameters: minimum window length of 120 SNPs, maximum gap size between two SNPs of 1000 kb, minimum ROH length of 1000 kb, minimum number of potential marked SNPs of 50, one heterozygote allowed per window, maximum of five missing calls per window, sliding window length of 50 SNPs and proportion of overlapping windows that must be a homozygous >0.05.

To identify ROH islands, the most observed homozygous segments shared by individuals in each breed were considered as an indication. The '--homozyg-group' function implemented in PLINK was used to assess ROH islands shared among individuals. Principal component analysis (PCA) was used to identify structure in the distribution of genetic variation across breeds. The GenBank annotation based on the ARS-UCD1.2 assembly of the bovine genome was used to identify genes. Gene networks highlighting biological processes among the gene sets of each breed were generated using the ClueGO plugin for Cytoscape (Bindea *et al.*, 2009). This procedure allowed the creation of gene networks highlighting biological roles, and the comparison of gene clusters by visualizing their functional differences or similarities.

Results

PCA analyses discriminated the three populations (Figure 1). The pattern of ROH differed across breeds (Figure 2A). CUR breed showed higher percentage of long-range ROH (>31 Mb; ~5%), while CCB showed the higher percentage of short range ROH (<5 Mb; ~90%). Based on ROH islands, regions were shared by more than 75% of the individuals in each breed (Table 1). Gene networks highlighting biological associations were generated based on genes found on ROH islands (Figure 2B). Several biological processes were highlighted.

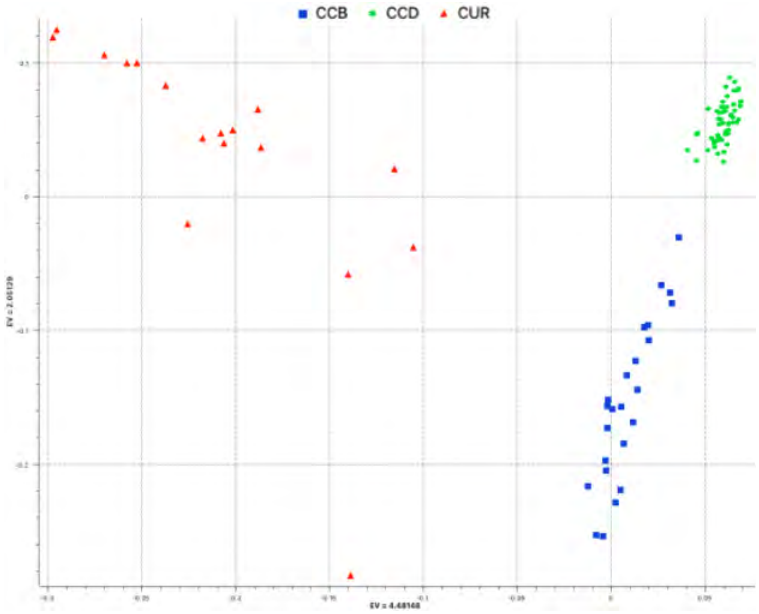


Figure 1. PCA analyses of three three Brazilian Iberian-derived naturalized cattle breeds. CCA: Caracu Caldeano (green); CCB: Caracu (blue); CUR: Curraleiro Pé-Duro (red).

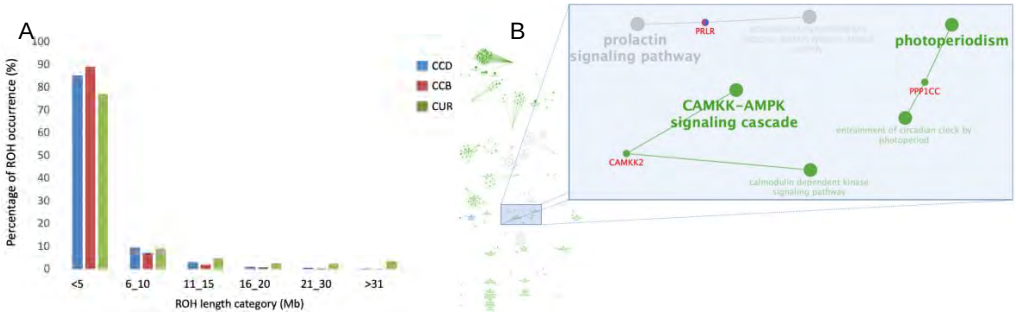


Figure 2. (A) Percentage of runs of homozygosity (ROH) per length categories in three Brazilian Iberian-derived naturalized cattle breeds. The sum of ROH was calculated per animal within each ROH length category. (B) Gene networks highlighting biological associations of genes found on ROH islands in three Brazilian Iberian-derived naturalized cattle breeds. Blue and red nodes are genes identified in CCA and CCB, respectively. Green nodes are biological processes related with genes identified in CUR breed. Grey nodes are biological processes shared between two breeds genes. Enriched terms are in bold.

Table 1. Chromosomal position, length in base pairs (bp), number of markers and gene content of runs of consensus homozygosity (ROH) islands most shared (ROH freq.) identified in three Brazilian Iberian-derived naturalized cattle breeds.

Breed	Chr.	Begin	End	Length (bp)	Nº SNP	ROH freq. (%)	Annotated genes
CCD	20	38,905,848	40,040,897	1,135,049	305	75.55	*PRLR, RAI14, RAD1, C1QTNF3, LOC112443006, SLC45A2, ADAMTS12, BRIX1, TTC23L, AMACR, RXFP3, AGXT2, DNAJC21, LOC107131570, LOC112443059, LOC112443042*
	20	40,248,889	40,961,988	713,099	186	75.55	*NPR3, TARS1, ADAMTS12, LOC107131571*
CCB	20	38,893,253	40,623,091	1,729,838	477	75.00	*PRLR, RAI14, RAD1, C1QTNF3, LOC112443006, SLC45A2, ADAMTS12, BRIX1, TARS1, TTC23L, AMACR, RXFP3, AGXT2, DNAJC21, LOC107131571, LOC107131570, TRNAT-UGU, LOC112443059, LOC112443042*
	20	38,040,564	38,084,918	44,354	10	75.00	*SKP2, NADK2*
CUR	17	53,133,247	54,716,438	1,583,191	181	82.35	*KDM2B, RHOF, CAMKK2, MORN3, DIABLO, MLXIP, HVCN1, ANAPC5, IFT81, TCTN1, RAD9B, ATP2A2, TMEM120B, RNF34, GPN3, PPP1CC, ORAI1, P2RX7, HPD, P2RX4, FAM216A, VPS29, MYL2, BCL7A, SETD1B, VPS33A, ARPC3, PSMD9, LRRC43, CFAP251, ANAPC7, B3GNT4, PPTC7, CCDC63, LOC100300938, LOC104974641, LOC104968522, LOC101905029, LOC101904098, LOC101903644, LOC100335936, LOC783255, TRNAE-UUC, LOC112442115, LOC112442104, LOC112442039, LOC112442038, LOC112442004*

[1] CUR: Curraleiro Pé-Duro; CCD: Caracu Caldeano; CCB: Caracu selected for beef.

Discussion

We evaluated the run of homozygosity in three Brazilian Iberian-derived Locally Adapted cattle. Since a run of homozygosity is the probability that all consecutive markers on a pair of homologous chromosome segments, in the same or different individual(s), display identical alleles (Hayes *et al.*, 2003), the extent and frequency of ROH may provide information about the ancestry of an individual and its population. Moreover, inbreeding may be inferred from the presence of long ROH, with longer segments indicating recent inbreeding within a population (Kirin *et al.*, 2010). We observed that the three breeds showed a similar occurrence of short ROH. This indicates a proper matting scheme to prevent higher inbreeding.

RHO islands were identified in the three breeds from which genes were retrieved and their biological processes analysed. In Caracu lineages (CCD and CCB), we identified islands in chromosome 20 with a gene associated, for example, with Prolactin signalling pathway (*PRLR*). *PRLR* encodes for prolactin receptor protein and was previously reported to be a useful genetic marker for milk production (Zhang *et al.*, 2008) and a promising candidate gene for growth traits in cattle (Lü *et al.*, 2011). This gene was also cited to present 'slick coat' mutations that confer thermotolerance to tropically adapted *Bos taurus* cattle (Littlejohn *et al.* 2014; Florez Murillo *et al.* 2021). Since CCD and CCB has been selected for dairy and beef, respectively, it makes sense to be fixing this gene, with a possible role in milking, growth and heat resilience. ROH island was observed on chromosome 17 of CUR breed. The biological process network hilighted a gene (*CAMKK2*) related with CAMKK-AMPK signalling pathway. This gene encodes for the calcium/calmodulin dependent protein kinase kinase 2 protein and have been cited to be involved in regulating important physiological and pathophysiological processes, including energy balance, adiposity, glucose homeostasis and inflammation (Raciopi and Means, 2012). Thus, it is possible that this gene may play a role in some of these processes in CUR breed.

We observed the occurrence of short ROH islands in all breeds, suggesting a successful matting scheme, also highlighted by a lower occurrence of long ROH. Genes located in ROH islands were explored via biological processes, providing genetic insights about the breeds.

Acknowledgements

The authors would like to thank the *Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq)/ Instituto Nacional de Ciência e Tecnologia – Ciência Animal (INCT-CA)* for their financial support.

References

Bindea G., Mlecnik B., Hackl H., Charoentong P., Tosolini M., *et al.* (2009). Bioinformatics, 25(8), 1091-1093. https://doi.org/10.1093/bioinformatics/btp101

Egito A.A., Paiva S.R., Maria do Socorro M., Mariante A.S., Almeida L.D., *et al.*, (2007). BMC genetics, 8(1), 1-14. https://doi.org/10.1186/1471-2156-8-83

Gibson J., Newton E.M. and Collins A. (2006). Human Molecular Genetics, 15, 789–95. https://doi.org/10.1093/hmg/ddi493

Hayes B.J., Visscher P.M., McPartlan H.C. and Goddard M.E. (2003). Genome Res. 13, 635-643. http://www.genome.org/cgi/doi/10.1101/gr.387103

Kirin M., McQuillan R., Franklin C. S., Campbell H., McKeigue P. M., *et al.* (2010). PLoS One 5:e13996. https://doi.org/10.1371/journal.pone.0013996

Littlejohn M.D., Henty K.M., Tiplady K., Johnson T., Harland C. *et al.*, (2014). Nature communications, 5(1), 1-8. https://doi.org/10.1038/ncomms6861

Lü A., Hu X., Chen H., Dong Y., and Pang, Y. (2011). Molecular biology reports, 38(1), 261-266. https://doi.org/10.1007/s11033-010-0103-5

McManus C., Prescott E., Paludo G.R., Bianchini E., Louvandini, H., *et al.*, (2009). Livestock Science, 120(3), 256-264. https://doi.org/10.1016/j.livsci.2008.07.014

Murillo J.M.F., Landaeta-Hernández A.J., Kim E.S., Bostrom J.R., Larson S.A., *et al.*, (2021). Animal genetics, 52(1), 132-134. https://doi.org/10.1111/age.13027

Peripolli E., Munari D.P., Silva M.V.G.B., Lima A.L.F., Irgang R., *et al.* (2017). Animal genetics, 48(3), 255-271. https://doi.org/10.1111/age.12526

Purcell S., Neale B., Todd-Brown K., Thomas L., Ferreira M.A., *et al.* (2007). The American journal of human genetics, 81(3), 559-575. https://doi.org/10.1086/519795

Racioppi, L., and Means, A. R. (2012). Journal of Biological Chemistry, 287(38), 31658-31665. https://doi.org/10.1074/jbc.R112.356485

Verardo L.L., e Silva F.F., Machado M.A., do Carmo Panetto J.C., Faza D.R.D.L.R., *et al.* (2021). Frontiers in genetics, 12: 702822. https://doi.org/10.3389/fgene.2021.702822

Zhang J., Zan L., Fang P., Zhang F., Shen G., *et al.*, (2008). Canadian journal of animal science, 88(1), 33-39. https://doi.org/10.4141/CJAS07052

Zhang Q., Guldbrandtsen B., Bosse M., Lund M.S., and Sahana, G. (2015). BMC genomics, 16(1), 1-16. https://doi.org/10.1186/s12864-015-1715-x

527. Copy number variation analysis in Nguni and Bonsmara crossbred cattle

B. Bhika Kooverjee[1*], P. Soma[1], F.W.C Neser[2], M.A. van der Nest[3] and M.M Scholtz[1,2]

[1]Agricultural Research Council – Animal Production, Irene, South Africa; [2]Department of Animal, Wildlife and Grassland Sciences, University of the Free State, Bloemfontein, South Africa; [3]Agricultural Research Council – Biotechnology Platform, Onderstepoort, South Africa; kooverjeeb@arc.agric.za

Abstract

Crossbreeding is a useful tool in livestock systems in mitigating the effects of climate change. The aim of this study was to perform a copy number variation analysis in indigenous Nguni and Bonsmara crossbred cattle. Four crossbred individuals were sequenced at 10× coverage. Following quality control, sequence data were interrogated for CNVs using the panelcn.MOPS tool. CNVs detected harboured genes related to important biological processes. Genes relating to fertility (MTERF2), Heat tolerance (LOC109577026), Stress response (HTRA2) and carbohydrate metabolism (LOC109573033) was found in the Bonsmara-sired crossbreds. While genes relating to embryogenesis (*ARMCX1*) and lipid metabolism (*ENHO*) were found in the Nguni-sired crossbreds. Results of this study suggested these crossbred cattle display potential for inclusion in various crossbreeding programs as they harbour genetic resources that may enhance the presence of significant traits such as meat quality, heat tolerance and adaptation, while simultaneously limiting the adverse effect on the environment.

Introduction

Copy number variations (CNVs) is defined as the deletion or repetition of a genome copy number (Goshu *et al.*, 2018), and may have a functional and evolutionary impact (Keel *et al.*, 2016). Changes in gene copy number also influence gene dosage, unjustified gene fusion, gene interruption, and position effects (Liu *et al.*, 2010). Several studies attempted to increase the understanding why crossbred cattle differ in their response to environmental changes such as heat stress. Since CNVs are an important source of genetic variability associated with phenotypic variation, the aim of this study was to examine genome-wide copy number differences between F1 progeny sired from Nguni and Bonsmara dams and with F1 progeny sired from Bonsmara and Nguni dams, which are both indigenous breeds. For this purpose, the genomes of F1 progeny from Bonsmara × Nguni reciprocal crosses were sequenced and panelcn.MOPS, a read depth approach based on next-generation sequencing, was used to identify genes that differ in copy number variation between crossbred progeny from Bonsmara and Nguni sires.

Materials & methods

Ethical approval and animal selection. Ethical approval was obtained from the Agricultural Research Council, Animal Production Ethics Committee (APIEC21/06). Hair samples from two Nguni-sired (N×B) and two Bonsmara-sired (B×N) crossbred animals were collected from a crossbreeding project in the Northern Cape province.

DNA extraction and sequencing. DNA was extracted from hair samples using the Qiagen DNeasy Blood and tissue kit as per manufacturer's protocol. For whole genome sequencing of the four crossbreds, libraries were prepared according to the Illumina TruSeq Nano DNA Library Prep Kit sample preparation guide. Sequence reads were filtered for base quality and adapter trimming using Trimmomatic v0.36. After trimming, only pairs of DNA sequences with reads exceeding 36 bp were retained for analysis. BWA-MEM v0.7.17 software (Li and Durbin, 2009) was used to align sequences with the reference genome (*Bos_indicus*_1.0, Canavez *et al.*, 2012). The *Bos indicus* reference (GCF_000247795.1_*Bos_indicus*_1.0) was

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_527, © B. Bhika Kooverjee *et al.* 2022

chosen due to African Sanga cattle owning unique indicine ancestry that is highly distinguished from that found in modern cattle populations such as the Brahman and Nellore (van Marle-Koster *et al.,* 2021). The sequencing depth determined with Samtools depth command for the crossbred ranged from 14.7 to 19.8.

CNV analysis and gene annotation. The GUI version of PanelcnMOPS program, CNV Detective (https://www.bioinf.jku.at/software/panelcnmops/) was used to identify copy number variations in the crossbreds. The two Nguni-sired crossbreds were used as the control group while the two Bonsmara-sired crossbreds were used to represent the test group. Through the advanced option setting, the duplication threshold was set at 1.46, the deletion threshold was set at 0.57, along with minimum median RC/ROI ratio = 30 and the sex = male. Quality control (QC) involves the minimum median RC/ROI ratio, where, samples with a median RC across all ROIs that is lower than 0.55 times the median of all samples fail the first step of the sample QC. For each ROI, the ratio between the normalized RCs of each sample and the median across all remaining samples is calculated, when samples show a high variation in RC ratios it fails the second QC step. CNVs labelled with 'LowQual' were regarded as insignificant and removed. The gene content of significant CNV regions were assessed based on the gene annotation of the UMD3.1 genome assembly using Ensembl (Ensembl Genes104). The gene ontology terms (for biological processes, cellular component and molecular function) identified using PANTHER with Bonferoni correction with significance of $P<0.05$.

Results

The results are indicated in Figure 1 and Table 1. Figure 1 displays the contribution of each copy number class to the total number of CNV calls per chromosome. In Table 1 the genes associated with different biological traits are listed.

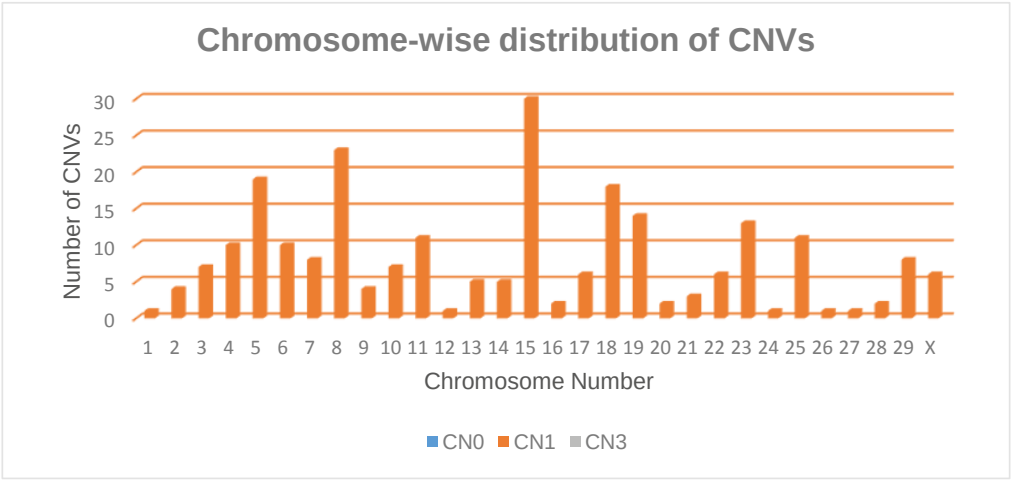


Figure 1. Bar plot displaying the contribution of each copy number class to the total number of CNV calls per chromosome. CN0 (blue) and CN1 (orange) represents the double and single copy number deletion events, respectively. CN3 (grey) represents the single duplication copy number events.

Table 1. CNV genes identified in the Nguni-sired (N×B) and Bonsmara-sired (B×N) crossbreds detected by panelcn.MOPS.

Breed	Trait	Chr	Start	End	Gene
N×B	Embryogenesis	X	30880957	30882318	*ARMCX1*
N×B	Lipid metabolism	8	80116078	80116308	*ENHO*
B×N	Female fertility	5	75275935	75277092	*MTERF2*
B×N	Heat tolerance	23	27266576	27268501	*LOC109577026*
B×N	Stress response	11	10412204	10415095	*HTRA2*
B×N	Methane Production	20	41718533	41742666	*RAI14*
B×N	Taste	4	1,09E+08	1,09E+08	*TASR38*
B×N	Visual Perception	6	1,21E+08	1,21E+08	*HMX1*
B×N	Carbohydrate metabolism	18	2243625	2244812	*LOC109573033*

Discussion

In this study, two hundred and thirty-nine CNVs regions involving 2.29 Mbp (~0.092%) of the reference genome were identified in four animals from two crossbreds. CNVs ranged from 230 bp to 153,354 bp in size with an average length of 7,718 bp. Among these, 185 CNVs were deletions (losses) and 54 CNV duplications (gains). CNV analysis in numerous studies showed great variation in the number of CNV regions reported in cattle (Keel, *et al.,* 2016). A gene in B×N identified as LOC109577026 located on Chromosome 23 was shown to be involved with cellular response to heat stress (Sakatani *et al.*, 2012). Another gene identified that plays a role in adaptation, includes HtrA serine peptidase 2 (HTRA2) and also involved in cellular response to heat (Skorko-Glonek *et al.*, 1999) (Table 1). Reproduction is one of the key factors driving the economic efficacy through sustainable meat production in the beef industry. In this study, genes relating to animal development and fertility were detected. This include mitochondrial transcription termination factor 2 (MTERF2) located on Chromosome 5 in the B×N crossbreds and was previously identified as a candidate gene associated with reproductive traits in Nellore cattle (Oliveria Junior *et al.,* 2019). The armadillo repeat containing X-linked 1 gene located on the X Chromosome of the N×B crossbreds plays a role in embryogenesis during the development of embryos in humans (Kusuma *et al.,* 2010), while its function in cattle is still unclear. Additionally, genes relating to metabolism was detected in the both crossbreds. The carbohydrate sulfotransferase 6 (LOC109573033) located on Chromosome 18 in the B×N crosses is involved in carbohydrate metabolism and inflammatory response to pathogens (Tetas *et al.*, 2016)). While in the N×B crossbreds, the ENHO gene also known as adropin, located on Chromosome 8, plays a role in lipid metabolism and regulation of glucose homeostasis (Jasaszwili *et al.*, 2020). This is further supported by the GO terms cellular macromolecule biosynthetic process (GO: 0009059) and regulation of biological process (GO: 0050789). Overall, the functional genes identified were found to play a role in reproduction and metabolism in both crossbreds. It is therefore possible that these genes will play a role in adaptation to heat and other environmental stresses in both the Nguni-sired and Bonsmara-sired crossbreds.

Acknowledgements

The support of the National Research Foundation under grant UID 135438 is acknowledged.

References

Canavez, F.C., Luche, D.D., Stothard, P., Leite, K.R., Sousa-Canavez, J.M., *et al.* (2012) Journal of Heredity 103(3):342-348. https://doi.org/10.1093/jhered/esr153

Cho, S.J., Kang, M.J., Homer, R.J., Kang, H.R., Zhang, X., *et al.* (2006) Journal of Biological Chemistry, 281(12):8161-8168. https://doi.org/10.1074/jbc.m506770200

Goshu, H.A., Chu, M. and Yan, P. (2018) African Journal of Biotechnology 17(43):1313-1323. https://doi.org/10.5897/AJB2018.16464

Jasaszwili, M., Billert, M., Strowski, M.Z., Nowak, K.W. and Skrzypski, M., (2020) Molecules, 25(3):549 https://doi.org/10.3390/molecules25030549

Keel, B.N., Lindholm-Perry, A.K. and Snelling, W.M. (2016) Frontiers in genetics 7:207. https://doi.org/10.3389/fgene.2016.00207

Kusama, Y., Takayanagi, S., Tategu, M. and Yoshida, K., (2010) Experimental and Therapeutic Medicine *1*(2):395-399. https://doi.org/10.3892/etm_00000062

Liu, M., Fang, L., Liu, S., Pan, M.G., Seroussi, E., *et al.* (2019) BMC genomics 20(1):181. https://doi.org/10.1186/s12864-019-5552-1

Van Marle-Köster, E., Visser, C., Sealy, J., & Frantz, L., (2021) Sustainability, 13(8): 4388. http://dx.doi.org/10.3390/su13084388

Oliveira Júnior, G.A., Santos, D.J.A., Cesar, A.S.M. *et al.* (2019) J Animal Sci Biotechnol 10, 97. https://doi.org/10.1186/s40104-019-0403-0

Sakatani, M., Alvarez, N.V., Takahashi, M. and Hansen, P.J., (2012).. Journal of Dairy Science, 95(6):3080-3091. https://doi.org/10.3168/jds.2011-4986

Skorko-Glonek, J., Zurawa, D., Kuczwara, E., Wozniak, M. Wypych, Z. *et al.* (1999) Molecular and General Genetics 262(2):342-350. https://doi.org/10.1007/s004380051092

Stothard, P., Choi, J.W., Basu, U., Sumner-Thomson, J.M., Meng, Y., *et al.* (2011) BMC genomics 12(1):1-14. https://doi.org/10.1186/1471-2164-12-559

Tetas Pont, R., Downs, L., Pettitt, L., Busse, C. and Mellersh, C.S., 2016. Veterinary ophthalmology, 19(6):488-492. https://doi.org/10.1111/vop.12332

Zhan, B., Fadista, J., Thomsen, B., Hedegaard, J., Panitz, F. *et al.* (2011) BMC genomics 12(1):1-20. https://doi.org/10.1186/1471-2164-12-557

528. Resequencing of casein genes in Swedish Red Cattle with varying milk coagulation properties

T.K. Vanhala[1*], F. Lewerentz[2], L. Buhelt Johansen[3], C. Sibersen[3], M. Paulsson[2], M. Glantz[2] and D.J. de Koning[1]

[1]Department of Animal Breeding and Genetics, Swedish University of Agricultural Sciences, Box 7023, 750 07 Uppsala, Sweden; [2]Department of Food Technology, Engineering and Nutrition, Lund University, P.O. Box 124, 221 00 Lund, Sweden; [3]Arla Foods amba, Agro Food Park 19, 8200 Aarhus N, Denmark; tytti.vanhala@slu.se

Abstract

The ability of milk to coagulate is crucial for cheese production. Poorly or non-coagulating milk will affect industrial cheese production with lower cheese yield and longer processing time. We have previously shown that up to 19% of milk from Swedish Red Dairy Cattle is poorly coagulating and 18% is non-coagulating. We have also identified specific milk proteins and genetic variants as potential biomarkers for selective breeding. However, these variants are not fully characterized and show conflicting results between presumed genetic variants and actual protein masses. We re-sequenced the casein genes in 30 cows with varying milk coagulation properties and detailed analyses of their milk proteins. The analyses of three genes confirmed existing variation in the casein gene, and identified 77 single nucleotide polymorphisms in introns as well as at least ten insertions or deletions. No new coding variants were detected so far.

Introduction

Rennet induced milk coagulation is a crucial part of cheese production. Poorly and/or non-coagulating milk could lead to increased processing times as well as production losses. Earlier studies in Sweden showed that up to 37% of cows from the Swedish Red Dairy Cattle (RDC) produce poor or non-coagulating milk (Gustavsson *et al.*, 2014; Nilsson *et al.*, 2019). Heritability estimates for non-coagulating milk in RDC was estimated to be 0.45 in an earlier study of 395 cows (Gustavsson *et al.*, 2014). More recently, Duchemin *et al.* (2020) showed a heritability of 0.28 for non-coagulating milk with estimates of heritability for other coagulation measurements ranging between 0.32 and 0.56 in a larger study with 600 RDC cows. Duchemin *et al.* (2020) also showed moderate to strong phenotypic and genetic correlations between coagulation traits and the relative proportions of different casein proteins in the milk. The casein proteins in cows' milk are α_{S1}-, β-, α_{S2}- and κ-casein with corresponding genes CSN1S1, CSN2, CSN1S2, and CSN3. The casein genes are all clustered on bovine chromosome six. Using the same samples as Duchemin *et al.* (2020), Nilsson *et al.* (2020) studied different milk protein variants and their post-transcriptional modifications (PTM) in relation to coagulations properties. Nilsson *et al.* (2020) determined the different genetic variants and PTM of the casein proteins via LC-HRMS (liquid chromatography high resolution mass spectrometry). Nilsson *et al.* (2020) showed a significant effect of some casein variants on coagulation properties. These effects were mainly indirect and due to changes in the relative amounts of different caseins (Nilsson *et al.,* 2020). In order to better understand the functional role of caseins in coagulation properties a new project was initiated to follow a subset of samples from DNA sequence, to amino acid sequence and post-transcriptional changes of the resulting caseins. Here we present the initial phase of the project where we re-sequenced the casein genes in a subset of cows with different coagulation properties.

Materials & methods

From our earlier studies (Nilsson *et al.*, 2020; Duchemin *et al.,* 2020), 30 cows were selected based on their coagulation properties, protein genetic variants and PTM, availability of frozen milk samples for further analysis, limited relatedness between animals, and spread across the different farms. All cows had DNA available that was bio-banked at SLU.

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_528, © T.K. Vanhala *et al.* 2022

Re-sequencing of casein genes. Primer sequences from the exons of all four casein genes were taken from Gallinat *et al.* (2013) resulting in 34 sequencing fragments. Sanger sequencing was done using the BigDye® Direct Cycle Sequencing kit (Applied Biosystems, Thermo Fisher). We followed the sequencing protocol exactly as described in the BigDye® manual (Thermo Fisher, online). Purification of the sequencing reactions was performed with BigDye® XTerminator kit (Applied Biosystems, Thermo Fisher). The sequencing reactions were then run in the Genetic Analyzer 3500xL (Applied Biosystems, Thermo Fisher). Sequence data was checked and edited manually using Chromas 2.5 (Chromas, Online) and BioEdit 7.2. (Hall, 1999).

Results

At the time of writing, the sequence analysis of CSN2, CSN3 and CSN1S1 was completed while the analysis of CSN1S2 was ongoing. In the sections below we will discuss our findings for each casein gene.

CSN2 (β-casein). The most notable result for CSN2 is that the cows were segregating for the proline to histidine amino acid substitution at exon 7 (His82Pro, Gallinat *et al.*, 2013) in the DNA sequence. This histidine substitution is present in several variants of β-casein (A1, B, and F/G). The A1 variant of β-casein has been associated to type I diabetes in humans (Elliott *et al.*, 1999), leading to a preference for milk with the A2 variant (Pro^{67}, in the mature protein, Caroli *et al.*, 2009) in several countries. Of the 30 cows, 6 were homozygous for the A1 variant, 11 were homozygous for the A2 variant, while the remaining 13 cows were heterozygous. There was no additional sequence variation in any of the exons in CSN2. This appears in disagreement with the protein mass data that showed additional variation related to exon 7 corresponding to β-casein variant B or F/E in one of the samples (Nilsson *et al.* 2020). Additional single nucleotide polymorphism (SNP) variation was detected in introns 2, 4, 5, and 6 as well as insertion/deletion (indel) variants in introns 2, 7, and 8. Exon 8, covering a single amino acid in mature protein (Val^{209}) appears to have a mutation in the primer site for 9 samples.

CSN3 (κ-casein). The analyses showed known amino acid changes in the DNA sequence (which includes the signal peptide) for exon 4 at amino acid positions 157 (Thr/Ile), 169 (Asp/Ala) and 176 (Ser/Gly) which, in combination, give rise to haplotypes A, B and E of κ-casein (Gallinat *et al.*, 2013). Fourteen cows had haplotypes AA, 7 had AB, 4 had AE, 2 had EE, 2 had BB and 1 cow had haplotype BE. For all cows the DNA derived haplotype agrees with the protein mass analysis. Two additional SNP were detected in exon 4: one synonymous mutation at amino acid position 189 just before a stop codon, and one non-synonymous mutation just after the stop codon. In addition, we found 3 SNP in intron 1, 11 SNP in intron 2, and 3 SNP plus 1 indel in intron 3.

CSN1S1 (α_{S1}-casein). In exon 17 we detected the Glu/Gly substitution at amino acid position 207 (Gallinat *et al.*, 2013) in the DNA sequence, corresponding to position 192 in the mature protein (Caroli *et al.*, 2009). This mutation gives rise to the C variant of αS1-casein. Two animals were heterozygous for the C variant and carried the common B variant as the other allele. The remaining animals were homozygous (BB) for the B variant and we did not observe the E variant in this sample. The sequencing analysis showed 47 SNP in introns and at least 6 indels. The detection of further indels was complicated due to extensive sequence slippage after a mononucleotide T repeat in the data for some primers.

Discussion

The coding variants detected in this study so far are in agreement with those previously described (e.g. Gallinat *et al.*, 2013), which means that we could not identify any new coding variants that could explain the differences in coagulation properties. For κ-casein, all the variants presented by Nilsson *et al.* (2020) were present in our subset of 30 samples. However, for β-casein, we only detected the A1 and A2 variants.

While the current study focused on the exons of the casein genes, many SNP and several indels were detected in the process. The potential role of these SNPs and indels in relation to both casein content and coagulation properties could be studied further by checking against a database of regulatory sequences and conservation across species. In the next phase of this project, milk samples will be subjected to protein sequencing, which will provide the amino acid sequences of the casein proteins, which can be compared to the amino acid sequences derived from the DNA data. By closing the gap between DNA sequence and functional protein properties we hope to understand more about the cheese coagulation properties of milk from RDC.

References

Caroli, A.M., Chessa, S., Erhardt, G.J. (2009) Invited review: Milk protein polymorphisms in cattle: Effect on animal breeding and human nutrition. J. Dairy Sci. 92:5335-5352https://doi.org/10.3168/jds.2009-2461. https://doi.org/10.3168/jds.2009-2461

Chromas, available at: https://chromas.software.informer.com/2.5/

Duchemin, S.I., Nilsson, K., Fikse, W.F., Stålhammar, H., Buhelt Johansen, L. *et al.*(2020). J. Dairy. Sci. 103(9): 8830-8342. https://doi.org/10.3168/jds.2020-18315.

Elliott R.B., Harris D.P., Hill J.P., Bibby N.J., and Wasmuth H.E. (1999) Diabetologia 42, 292–296.

Gallinat, J.L., Qanbari, S., Drögemüller, C., Pimentel, E.C.G., Thaller, G. *et al.* (2013). J.Dairy.Sci. 96(1): 99-709. https://doi.org/10.3168/jds.2012-5908.

Gustavsson, F., Glantz, M., Poulsen, N.A., Wadsö, L., Stålhammar, H. *et al.* (2014) J. Dairy Sci. 97(8): 5219-5229. https://doi.org/10.3168/jds.2014-7996.

Hall TA (1999). Nucl. Acids. Symp. Ser. 41:95-98. Available at: https://bioedit.software.informer.com/7.2/

Nilsson, K., Stålhammar, H., Stenholdt Hansen, M., Lindmark-Månsson, H., Duchemin, S. *et al.* (2019). Int. Dairy J. 95: 50-57. https://doi.org/10.1016/j.idairyj.2019.03.006

Nilsson, K., Buhelt Johansen, L., de Koning, D.J., Duchemin, S.I., Stenholdt Hansen, M. *et al.* (2020). J. Dairy Sci. 103(8): 6858-6868. https://doi.org/10.3168/jds.2020-18357.

Thermo Fisher Scientific, BigDye® Direct Cycle Sequencing Kit manual, available at: https://www.thermofisher.com/document-connect/document-connect.html?url=https%3A%2F%2Fassets.thermofisher.com%2FTFS-Assets%2FLSG%2Fmanuals%2Fcms_091370.pdf

529. Major histocompatibility complex class II DRB gene Exon-2 of Indian camel breeds show limited genetic diversity

V. Prakash*, B. Jyotsana and S. Suthar

ICAR, National Research Centre on Camel, Bikaner, P.O. Box 07, 334001 Bikaner, India;
ved.prakash4@icar.gov.in

Abstract

Major histocompatibility complex (MHC) genes are the most polymorphic genes in vertebrates and play a crucial role in immunity and pathogen resistance. The blood samples were collected from six camel breeds namely Jaisalmeri, Jalori, Bikaneri, Kachchhi, Kharai and Malvi. The DNA was isolated from the blood, primers were synthesised and PCR amplification of MHC class II DRB gene fragment spanning exon-2 region was done. After sequencing of the PCR product, sequence variation, allele diversity and haplotype transmission pattern were analysed. The optimum annealing temperature was 57.5 °C. A 496 bp long MHC DRB gene region was amplified in six camel breeds. Camel MHC-DRB exon-2 region comprised of 270 bp and spanned from 116 to 385 bp of the amplified sequence. The variation at only four loci in exon-2 (NC_044530.1:g.[21049396A>G:21049451A>G:21049517A>G:21049598C>T]) contributed to allelic diversity. The sequence comparison revealed lower genetic variation in Indian camel compared to that observed in livestock species like cattle, buffalo, sheep and goat.

Introduction

In India, camel is an important animal specially in Rajasthan which have almost 60% of its area as desert land. According to 20th livestock census (2019) camel population of India is 0.25 million. India has 9 recognized camel breeds (NBAGR,2021). The camel has high heritage as well as cultural value. Today camel milk is in high demand because of its therapeutic properties (Singh *et al.,* 2017). Immune system of camel is very unique that impart strong resistance to many viral and microbial infections. It harbours unique antibodies that lacks light chain and have only heavy chain so their antibodies are several times smaller (nanobodies) compared to antibodies produced by other animals including humans. They can even interfere with enzyme binding sites, contributing to camels' remarkable disease resistance (Jassim *et al.,*2001). The mammalian Major Histocompatibility Complex (MHC) is a genetic region containing highly polymorphic genes with immunological functions. The antigen binding site is encoded by the second exon of genes encoding antigen presenting molecules. The exon 2 sequences of these MHC genes have evolved under the selective pressure of pathogens and the majority of the functionally important polymorphisms are concentrated in exon 2, which encodes the antigen-binding site of the molecule (Plasil *et al.*, 2016). Study on MHC class II genes in camel revealed low polymorphism in exon-2 region of DRB gene (Plasil *et al.,*2016). However, the study covered limited number of individual. The present study was planned to study genetic diversity in MHC-DRB class II exon-2 region of six Indian dromedary camel breeds.

Materials & methods

Animals DNA isolation and PCR amplification. About 10 ml of blood samples was collected aseptically through jugular vein puncture into vacutainer tube containing anticoagulant EDTA from 203 camels from six breed namely Jaisalmeri (35), Jalori(27), Bikaneri(45), Kachchhi(44), Kharai(30) and Malvi(22). Approval of Institutional Animal Ethics Committee (IAEC) was taken for the collection of blood from selected animals. Genomic DNA was isolated from the blood samples by 'Phenol: chloroform extraction' method as described by Sambrook *et al.* (2001) with slight modification. The integrity and the quality of extracted DNA was checked on 0.8% agarose. Only the DNA samples having concentration greater than 50 ng/µl with optical density of 1.6-1.8 was used for amplification of desired DNA fragment through PCR

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_529, © V. Prakash *et al.* 2022

technique. Primers reported by Plasil *et al.* (2016) was utilised for amplification of targeted fragment. The PCR reactions were carried out in programmable thermocycler (Applied Biosystem gradient thermocycler) using PCR master mix (Promega). The quality and size of the PCR amplicons for studied MHC class II gene region were assessed on 1.5% agarose gel containing ethidium bromide (1% solution). PCR products were checked for amplification by electrophoresis on 2.0% agarose gel (Himedia), in parallel with 100 bp DNA marker (Thermo scientific).

Sequence based typing, deductional allele assignment and haplotype transmission. After purification of the amplified PCR products bidirectional sequencing using forward and reverse primers was done using Sanger Dideoxy Chain termination method (Eurofins Genomics). The forward and reverse sequences obtained for each fragment were edited using Codon Code Aligner software (USA) to arrive at resultant nucleotide sequences. Sequence based typing (SBT) of amplified product was done and homozygous and heterozygous genotype were identified. Cloning of heterozygous sample was done using pGEM®-T Easy Vector system II and sequencing of plasmid was done. Deduction allele assignment was done for allele identification. The sequence pattern of progeny and parents were analysed to understand haplotype transmission pattern.

Results

A 496 bp long MHC DRB gene spanning exon-2 region was amplified in six camel breeds (Figure 1). The amplified fragments were sequenced and analysed to identify the polymorphic locus. All six breed were found polymorphic in the amplified region. Exon-2 comprised of 270 bp and spanned from 116 to 385 bp of the amplified sequence. The sequence variation at positionNC_044530.1:g.[21049396A>G:21049451A>G:21049517A>G:21049598C>T were found in the exon-2 region (Figure 2, 3, 4 and 5). Frequency of the heterozygote and homozygotes genotype at different loci varied between breeds. Twelve different haplotype patterns were observed (Table 1). Kachchhi and Jalori breeds had the highest haplotype diversity. Transmission pattern analysis showed that alleles at different polymorphic loci of parents combined to generate different genotype in the progeny (Table 2).

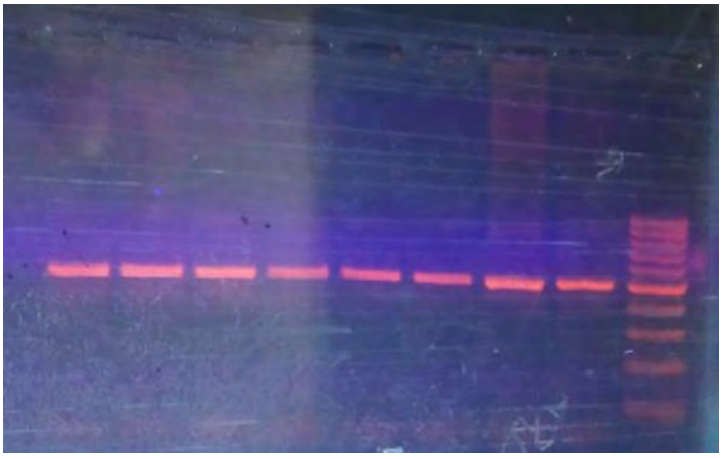


Figure 1. 496 BP fragment of MHC DRB gene spanning exon -2 region.

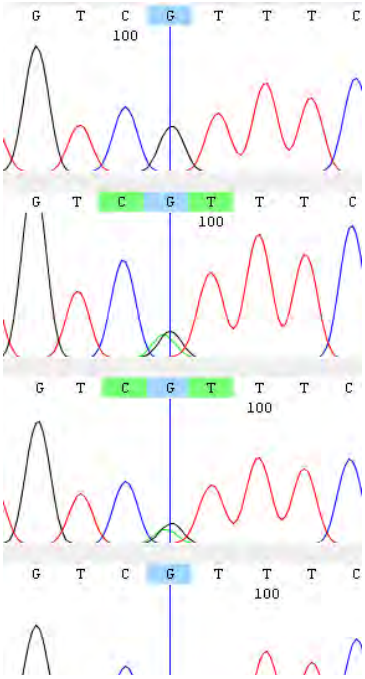


Figure 2. Heterozygosity at position 149 (NC_044530.1:g.21049396) of amplified fragment.

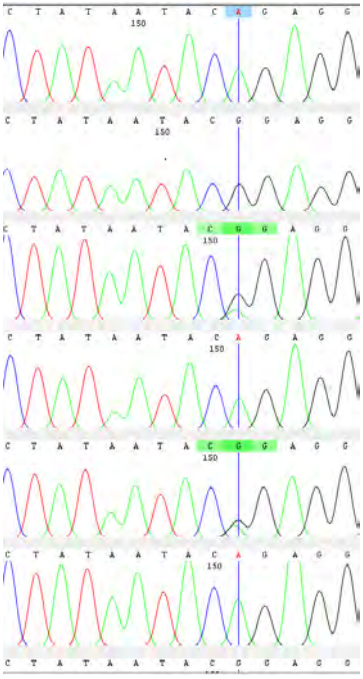


Figure 3 Heterozygosity (A/G) at position 204 ((NC_044530.1:g.21049451), of amplified fragment.

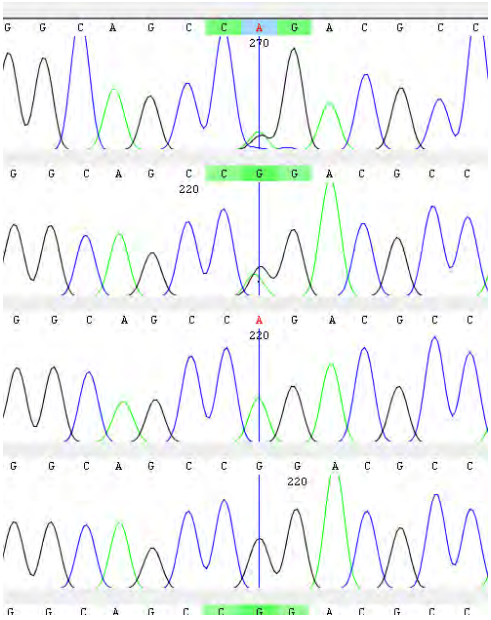


Figure 4. Heterozygosity (A/G) at 270th (NC_044530.1:g.21049517) position of amplified fragment.

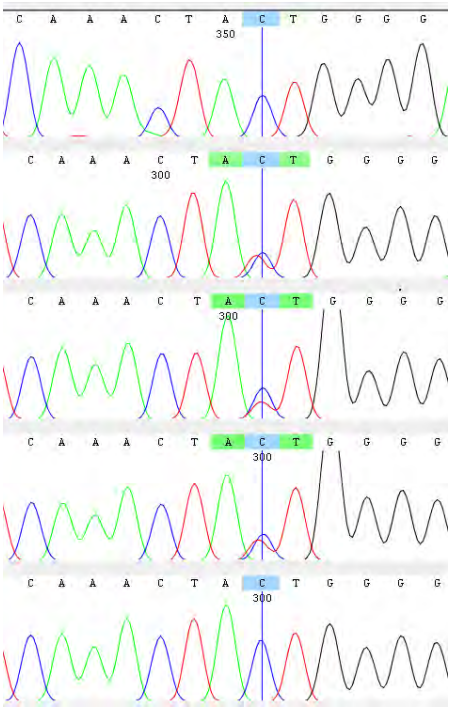


Figure 5. Heterozygosity(C/T) at position 351(NC_044530.1:g.21049598) of amplified fragment.

Table 1. Haplotype pattern found in the MHC DRB exon-2 region across six camel breeds.

S.N	Haplotype	Bikaneri	Jaisalmeri	Kachchhi	Kharai	Malvi	Jalori	Total
1	GAAC	3	4	4	7	5	4	27
2	GAAT	-	-	1	-	-	-	01
3	GAAY	2	1	1	8	-	7	19
4	GGGC	20	9	12	4	7	2	54
5	GRRC	16	14	10	5	8	4	57
6	GRRT	-	-	1	-	-	-	01
7	GRRY	4	7	8	4	1	1	25
8	RAAT	-	-	1	-	-	1	02
9	RAAY	-	-	1	-	-	3	04
10	RRRY	-	-	5	2	1	3	11
11	AAAT	-	-	-	-	-	1	01
12	AAAY	-	-	-	-	-	1	01
		45	35	44	30	22	27	203

Table 2.Transmission pattern analysis between parent and progeny.[1]

Family	Locus with genotype				Transmitted haplotype			
	149	204	270	351	149	204	270	351
Family-1 (Jaisalmeri breed)								
Progeny J 418	GG	AA	AA	CT				
Sire J 388	GG	AG	A/G	CC	G	A	A	C
Dam J 233	GG	AG	A/G	CT	G	A	A	T
Family 2 (Bikaneri breed)								
Progeny B 637	GG	AG	AG	CC				
Sire B 620	GG	AG	AG	CC	G	A	A	C
Dam B 543	GG	GG	GG	CC	G	G	G	C
Family -3 (Bikaneri breed)								
Progeny B 793	GG	AG	AG	CC				
Sire B 764	GG	GG	GG	CC	G	G	G	C
Dam B 697	GG	AG	AG	CT	G	A	A	C

[1] 149 = (NC_044530.1:g.21049396), 204 = (NC_044530.1:g.21049451), 270 = (NC_044530.1:g.21049517), 351 = (NC_044530.1:g.21049598).

Discussion

The first time Indian Dromedary MHC DRB gene exon-2 sequences were identified and polymorphism in it was detected. In the exon-2 region of DRB only limited variation was identified compared to that observed in livestock species like cattle, buffalo, sheep and goat. The findings are similar to reports of Plasil *et al.* (2016). The Kachchhi and Jalori breed had highest diversity. The exon-2 is transmitted as whole to progeny without any recombination.

References

Jassim, SA. And Naji, MA (2001). The Desert Ship heritage and science. Biologist (London); 48(6):268-72. PMID: 11740078.

Plasil, M., Mohandesan, E., Fitak, R. R., Musilova, P., Kubickova, S., Burger, P. A., and Horin, P. (2016). The major histocompatibility complex in Old World camelids and low polymorphism of its class II genes. BMC genomics, 17:167. https://doi.org/10.1186/s12864-016-2500-1

Sambrook, J. and Russel, D. W. (2001). Molecular cloning-A Laboratory manual. 3rd Edition Volume 1 Cold Spring Harbor Laboratory Press, New York, USA.

Singh, R., Mal, G., Kumar, D., Patil, N. V., Pathak, K. M. L., 2017. Camel milk: An important natural adjuvant. Agric. Res. 6(4), 327–340. https://doi.org/10.1007/s40003-017-0284-4

530. Transcriptome profile of liver from pigs fed with two different levels of soybean oil revealed the NF-κB1 modulation

S.L. Fanalli[1], B.P.M. da Silva[1], F.N. Ciconello[2], J. Reecy[3], J. Koltes[3], D. Koltes[3], L.C.A. Regitano[4], J.C.C. Balieiro[5], L.L. Coutinho[2], G.B. Mourão[2], G.C.M. Moreira[6], V.V. de Almeida[7], A. Luchiari Filho[2] and A.S.M. Cesar[1,2*]

[1]Faculty of Animal Science and Food Engineering, University of São Paulo, Pirassununga, SP, Brazil; [2]Luiz de Queiroz College of Agriculture, University of São Paulo, Piracicaba, SP, Brazil; [3]Iowa State University, Department of Animal Science, Ames, IA, USA; [4]Embrapa Pecuária Sudeste, São Carlos, SP, Brazil; [5]College of Veterinary Medicine and Animal Science, University of São Paulo, Pirassununga, São Paulo, Brazil; [6]University of Liège, GIGA Medical Genomics, Unit of Animal Genomics, Liège, Belgium; [7]Federal University of Goiás, Department of Animal Science, Goiânia, GO, Brazil; alinecesar@usp.br

Abstract

Nuclear factor-kappa light chain enhancer of activated B-cells (*NF-κB*) is a transcription factor (TF) considered to be a crucial element in several physiological processes, including fatty acid (FA) metabolism. This study aimed to identify differentially expressed genes (DEG) in liver of pigs fed with two different proportions of the soybean oil (SOY), which could be related to the *NF-κB* network. A total of 281 DEG between SOY (1.5%) and SOY (3.0%) diets (log2fold-change ≥1 or ≤-1; FDR-corrected *P*-value<0.1) were identified, in which 129 were down-regulated and 152 were up-regulated in SOY1.5 group. The gene-network enriched from DEG in response to an increase in the SOY in the liver transcriptomic profile provides us mechanisms to understand the regulatory effects of FAs, which may alter the behaviour of TFs and be modulated according to diet.

Introduction

Nuclear factor-kappa light chain enhancer of activated B-cells (*NF-κB*) is a protein complex of transcription factors, which plays critical roles in immunological processes, inflammation, proliferation, survival, and cell differentiation (Chen *et al.,* 1999). Moreover, *NF-κB* has been demonstrated to be involved in several physiological and pathological processes as well as in cell survival processes (Oeckinghaus and Ghosh, 2009).

The NF-κB protein complex are considered fundamental regulator of inflammation in liver cell populations and are critical for hepatocyte survival (Patterson and Georgel, 2014). Although not directly involved in fatty acid (FA) metabolism, *NF-κB* may act by activating the transcription of a series of enzymes that are related to inflammatory processes and signalling molecules synthesis (Patterson and Georgel, 2014). For instance, cyclooxygenases (COX) and lipoxygenase (LOX), which act in the conversion of members of omega-3 (n-3) and omega-6 (n-6) fatty acid families into signalling molecules that may activate or inhibit pathways related to inflammation processes (Schmitz and Ecker 2008). Then, the use of n-3 and n-6 fatty acids rich oils, such as soybean oil, in the diet reduce the levels of inflammatory mediators (Calder, 2010).

Therefore, maintaining an adequate regulation of this complex of transcription factors is considered by the literature as a critical factor for the maintenance of immune homeostasis of several tissues, besides regulating the expression of several genes related to immunological, inflammatory, as well as cell differentiation and survival processes. However, the transcription factors involved in key processes through the modification of gene expression in pigs are still poorly understood. This study aimed to identify differentially expressed genes (DEGs) in liver of immunocastrated male Large White pigs fed different levels of soybean oil, which could be related to the *NF-κB* network.

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_530, © S.L. Fanalli *et al.* 2022

Materials & methods

All animal procedures were approved by the 'Luiz de Queiroz' College of Agriculture Animal Care and Use Committee (University of São Paulo, Piracicaba, Brazil, protocol number 2018.1787.11.6 and number CEUA 2018-28). In a 98-day study, 35 immunocastrated male pigs (halothane homozygous-negative (NN); average initial body weight of 28.44±2.95 kg) were assigned to one of two treatments in a randomized complete block design with six replicate pens per treatment. Dietary treatments consisted of corn-soybean meal diets supplemented with 1.5% SOY (SOY1.5) or 3% SOY (SOY3.0). There were 17 pigs in the SOY1.5 group and 18 in SOY3.0 group. Diets were formulated to meet or exceed Rostagno *et al.* (2011) recommendations for growing-finishing pigs, comprising 3.28 and 3.36 Mcal of metabolizable energy/kg of diet (as-fed basis) for the SOY1.5 and SOY3.0 groups, respectively. The details of animal housing and experimental diet compositions were mentioned by Almeida *et al.* (2021). All pigs were slaughtered with average final body weight of 133.9±9.4 kg and liver samples were collected, snap-frozen in liquid nitrogen, and then stored at -80 °C until analysis. Total RNA was extracted from liver tissue samples using commercial RNA extraction kits (RNeasy® Mini Kit, Qiagen), according to the manufacturer instructions. The mRNA library preparation and sequencing (RNA-Seq) were described in Fanalli *et al.* (2021). The RNA-Sequencing (RNA-Seq) data quality was checked using the FastQC, v. 0.11.8 software [http://www.bioinformatics.bbsrc.ac.uk/projects/fastqc/]. Adapters and bases with low PHRED score were removed using the Trim Galore v. 0.6.5. Reads with a minimum length of 70 bases were aligned and mapped to the reference pig genome (*Sus Scrofa*11.1) using the assembly available at Ensembl Release 102 [http://www.ensembl.org/Sus_scrofa/Info/Index]. Alignment and mapping were performed using the STAR v. 2.7.6a. The differential expressed genes (DEG) were identified using the DESeq2 package available at Bioconductor open-source software for bioinformatics (Love *et al.*, 2014), using multi-factor design. The functional enrichment analysis of DEG was performed to obtain (P<0.01) comparative networks by analyzing the single network (transcription factor) and process networks using the default parameters of MetaCore software (Clarivate Analytics) v. 21.3 build 70600, which were filtered for liver tissue and *Homo sapiens* species dataset to show activating and inhibiting effects. For the functional analysis to identify important transcription factor (TF) related to lipid homeostasis and immune response, the construction of the gene network by the TF mechanism was performed. For each TF of the master list, the algorithm generated a subnetwork with all the shortest paths to which the TF from the nearest recipient with direct ligands on the list.

Results

On average 78% of total paired reads aligned against the reference genome. After filtering, 19,250 genes were considered for differential gene expression analysis between the SOY1.5 and SOY3.0 group. A total of 281 DEG (log2fold-change ≥1 or ≤-1; FDR-corrected P-value<0.1) were identified, where 129 were down-regulated (log2FC ranging from -3.0 to -0.20) and 152 up-regulated (log2FC ranging from 4.8 to 0.24) in the SOY1.5 group by DESeq2 (Fanalli *et al.*,2022). We detected in the functional enrichment the networks related to inflammation such as Inflammation_IL-6 signaling with DEG *ORM1*, *CDKN1A* and Inflammation_Kallikrein-kinin system with DEG *KNG1* both networks involved with the *NF-κB* transcription factor (P<0.1). Among the total of DEG identified, we showed in the Figure 1 those 23 down-regulated (blue circles) and 21 up-regulated (red circles) in the SOY1.5 group, where the color shade represented the log2fold-change value. These DEG are target genes for the main TF network detected by MetaCore, which identified the *E2F1, SOX9, SMAD3, STAT3, ETS1, Tcf, PR (nuclear), Androgen receptor, TSHZ1, NF-κB and NF-κB1 (p150)* as TF (Figure 1).

Discussion

Previously, we demonstrated that feeding pig diets with SOY3.0 increased the proportion of oleic acid (OA, C18:1 n-9) and decreased the percentages of linoleic acid (LA, C18: 2 n-6) and alpha-linoleic acid (ALA, C18:3 n-3) in the loin intramuscular fat compared to those fed SOY1.5 (Almeida *et al.,* 2021). These

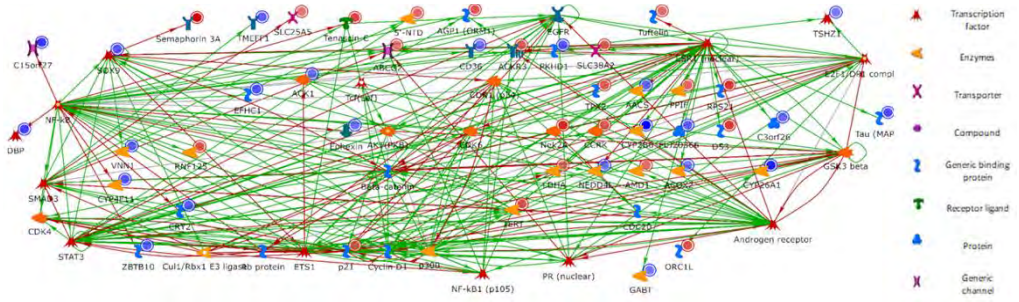


Figure 1. Network built based on the analysed networks (transcription factor) from the list of differentially expressed genes in the liver samples from pig fed with diets with 1.5 and 3.0% of soybean oil groups. Red circles represent up-regulated genes, whereas down-regulated genes are marked with blue circles. The color shade of the circles represents the log2fold-change value. Image created by MetaCore (Clarivate Analytics) [https://portal.genego.com/].

FA have been associated with the lower occurrence of some inflammatory diseases such as obesity, type 2 diabetes, and atherosclerosis (Jung *et al.*, 2008; Poudyal *et al.,* 2011). In the current study, we observed a significant (FDR 10%) number of DEGs in liver of pigs fed SOY1.5 vs SOY3.0, which were related to the network of important transcription factors linked to inflammatory response in liver tissue, being identified the *NF-κB*. *NF-κB* is related as a mediator in the process of vascular inflammation and one of the process networks enriched with DEGs is Inflammation_IL-6 signaling. Vascular inflammation is detected by the increase in circulating Interleukin-6 (IL-6), produced by inflammatory cells and secreted during inflammation by macrophages, endothelial cells, and fibroblasts, important in inducing B lymphocytes to differentiate into antibody-forming cells (plasma cells), in addition, IL6 signaling can lead to anti-apoptotic effects (Brasier, 2010). Another process network that *NF-κB* together DEG *KNG1* has been enriched with is the Inflammation_Kallikrein-kinin system related to inflammation, blood pressure control, coagulation (Peyrou, 2020).

A previous study carried out by Cai *et al.* (2005) reported that *NF-κB* is activated by genetic obesity and may be induced by diet (Chen *et al.,* 1999). The authors also observed that the lipid accumulation in the liver results in liver inflammation through *NF-κB* activation and cytokine production, causing insulin resistance both locally in the liver and systemically. The TF *NF-κB* has an important relationship to immunological and inflammation processes and can be activated by various stimuli, such as oxidizing free radicals and cytokines (Schmitz and Ecker 2008). As a key element of classical immunity, the proinflammatory *IKKβ / NF-κB* pathway modulates a wide range of cellular processes, including proliferation, immunity, inflammation, and apoptosis, from transcription of target genes. Furthermore, pro-inflammatory cytokine responses to dietary changes are induced via signalling pathways that mediate *NF-κB* activation (Bu *et al.*, 2021).

According to Rizzo and Laganà (2020), *NF-kB*, nuclear factor erythroid 2-related factor 2 (Nrf2), peroxisome proliferator-activated receptor gamma *(PPARγ)* and heat shock proteins (HSP), are affected by electrophilic long-chain FA n-3 derivatives and show effects in reducing inflammatory processes and oxidative stress. Using rat L6 myoblasts, Poletto *et al.* (2015) reported that FAs stimulated *NF-kB* expression. The authors demonstrated that gene expression may be regulated by OA and AL and 200 μM of either FA (OA or LA) is enough to increase *Rela* expression and protein binding to the *Slc2a4 – NF-kB* binding site in the skeletal muscle (Poletto *et al.,* 2015).

The network enriched from DEG in response to an increase in the soybean oil in the liver transcriptomic profile provides us with mechanisms to help understand the regulatory effects of fatty acids that can be altering the behaviour of transcription factors, which could be modulated according to diet. Further study to evaluate the effect of the two diets on the inflammatory response of the liver of the pigs used herein is necessary. Immunoassay tests from proteins extracted of the liver of the samples of this study are being performed by our group.

References

Almeida V.V. Silva, J.P.M., Meira, A.N., Moreira, G.C.M. *et al.* (2021) Livestock Science 248:104515. https://doi.org/10.1016/j.livsci.2021.104515

Brasier A.R. (2010). Cardiovasc Res 86(2): 211–218 https://doi.org/10.1093/cvr/cvq076

Bu J., Zhang M., Wu, Y., Jiang, N., *et al.* (2021) Arvo Journal 62(10):13 https://doi.org/10.1167/iovs.62.10.13

Cai D., Yuan M., Frantz D., Melendez P.A. *et al.* (2005) Nature Medicine 11(2). https://doi.org/10.1038/nm1166

Calder P.C. (2010) Nutrients 2(3):355-374. https://doi.org/10.3390/nu2030355

Chen F., Castranova V., Shi X., and Demers L.M. (1999) Clinical Chemistry 45(1):7-17. https://doi.org/10.1093/clinchem/45.1.7

Dobin A., Davis C.A., Schlesinger F., Drenkow J. *et al.* (2013) Bioinformatics 29(1):15–21. https://doi.org/10.1093/bioinformatics/bts635

Fanalli S.L. *et al.* (2021) Proc of the 26th Annual Meeting of the RNA Society, Virtual Meeting.

Fanalli, S.L.; da Silva, B.P.M.; Gomes, J.D.; Ciconello, F.N., *et al.* (2022) *Preprints* https://doi.org/10.20944/preprints202202.0149.v1.

Jung U.J., Torrejon C., Tighe A.P., and Deckelbaum R.J. (2008) The American Journal of Clinical Nutrition 87(6):2003S–2009S. https://doi.org/10.1093/ajcn/87.6.2003S

Love, M.I., Huber, W., Anders, S. (2014). *Genome Biology* 15, 550. https://doi.org/10.1186/s13059-014-0550-8.

Oeckinghaus A. and Ghosh S. (2009) 1(4):1-14. https://doi.org/10.1101/cshperspect.a000034

Patterson W.L. and Georgel P.T. (2014) Biochemistry and Cell Biology 92(5):321-328. https://doi.org/10.1139/bcb-2013-0127

Peyrou M., Cereijo R., Quesada-Lopez, Campderros L. *et al.* (2020) *Nat Commun* 11, 2132 (2020). https://doi.org/10.1038/s41467-020-16009-x

Poudyal H., Panchal S.K., Diwan V., and Brown L. (2011) Progress in Lipid Research 50(4):372-387. https://doi.org/10.1016/j.plipres.2011.06.003

Poletto A.C. (2015) Elsevier 65-72 https://doi.org/10.1016/j.mce.2014.12.001

Rizzo G., and Laganà A.S. (2020) Biomolecules 10(2):219. https://doi.org/10.3390/biom10020219

Rostagno, H.S., Albino, L.F.T., Donzele, J.L., Gomes, P.C., *et al.* (2011). 3rd edition, Viçosa, MG: UFV, 252 p.

Schmitz G. and Ecker J. (2008) Progress in Lipid Research 47(2):147-155. https://doi.org/10.1016/j.plipres.2007.12.004

531. SPP1 gene polymorphisms within Intron-IV and Exon-IV region and their association with milk traits in Gaolao cattle

R.R. Kate, D.S. Kale*, J. Singh and D.V. Patil

Department of Animal Genetics and Breeding, Nagpur Veterinary College, Maharashtra Animal and Fishery Sciences University, Nagpur-06, Maharashtra, India; deepakkaleccmb@gmail.com

Abstract

Secreted phosphoprotein 1 (SPP1) is an acidic, phosphorylated glycoprotein exhibited impact on milk composition and yield. The aim was to detect SPP1 gene polymorphisms and reveal its association with milk traits in Gaolao Cattle. SPP1G1-*Bsr*I locus revealed frequency of TT genotype as 0.99 & CT as 0.01 respectively. The association of SPP1G1-*Bsr*I polymorphism with milk traits revealed significant difference in Milk yield and SNF % for CT genotype as compared to TT genotype. Milk yield (7.33±0.67) and SNF% (9.03±0.08) traits differed significantly for CT genotype as compared to TT genotype. Sequencing of SPP1G1-*Bsr*I locus-based TT and CT genotypes revealed G-A SNP at 32nd position. Significant differences were revealed for SNF% and Lactose% for SPP1G5-SSCP patterns A. SPP1G5-SSCP revealed SNPs A-G at 71st, A-C at 109th and A-C at 208th position. Identified significant association needs validation in large data sets with records for development of markers for bovine milk traits.

Introduction

Currently vast amount of animal genomic research work is being carried out to understand genetic mechanism behind milk production traits. Candidate gene analysis studies in dairy cattle will aid in development of gene level information database for its utilisation in further genetic improvement plans for the breed. Bovine SPP1 consists of six exons spanning about 7 kb of genomic DNA and encodes a 278 amino acid protein (Leonard *et al.* 2005). Due to significant impact of SPP1 gene on milk composition, it has become economically more relevant. Previous studies relating to the polymorphism of the Osteopontin gene analysed its effects on milk yield and milk composition. Genetic variants of SPP1 have been proposed to be associated with milk fat (Khatib *et al.* 2007), milk yield, and protein content in dairy cattle. In view of above, the present study was planned with the objectives to detect SPP1 gene polymorphisms and reveal association with milk traits in Gaolao Cattle using PCR-RFLP, PCR-SSCP and sequencing marker techniques.

Materials & methods

Experimental animals and DNA isolation. Blood was collected from 280 unrelated adult milking Gaolao cattle from the breeding tract & Farm. Test day milk was collected from the cows for estimation of Milk components like Fat %, Protein %, SNF% and recorded in the excel sheet.

PCR-RFLP & PCR-SSCP, DNA sequencing, bioinformatics and statistical analysis. Reported and designed primers amplifying variable regions of SPP1 gene were selected for their use in PCR-RFLP, PCR-SSCP, Sequencing and association analysis (Table 1).

Results

SPP1G1-BsrI PCR-RFLP Analysis. The 290 bp amplified products of Intron-IV region of SPP1 gene were digested using 5 Units of *BsrI* restriction enzyme. SPP1G1-*BsrI* locus was found polymorphic (Figure 1) with genotype frequency for TT genotype as 0.99, CT=0.01 & CC=0.00. The allele frequency for the Talleles was found 0.9946 and for allele C was 0.0054 at SPP1G1-*BsrI* locus. Hardy-Weinberg principle showed that the population of Gaolao cattle has reached a genetic equilibrium with respect to SPP1G1-*BsrI* locus.

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_531, © R.R. Kate *et al.* 2022

Table 1. Name of primer, region & product size for SPP1 gene polymorphism study in Gaolao cattle.[1]

Sr. no.	Primer/region	Primer sequence	Product size (bp)/ Reference
1.	SPP1G1/ Intron-4	F:5′ GCAAATCAGAAGTGTGATAGAC3′ R:5′ CCAAGCCAAACGTATGAGTT3′	290bp/ Leonard *et al.* (2005)
2.	SPP1G5/ Exon-4	F:5′ACAGTAAGTGCTGTTAAATGAATCT3′ R:5′GCCTGCTTGGTGGCTATTTTT3′	362bp/ Designed (AY878328.1)

[1] F = forward primer; R = reverse primer.

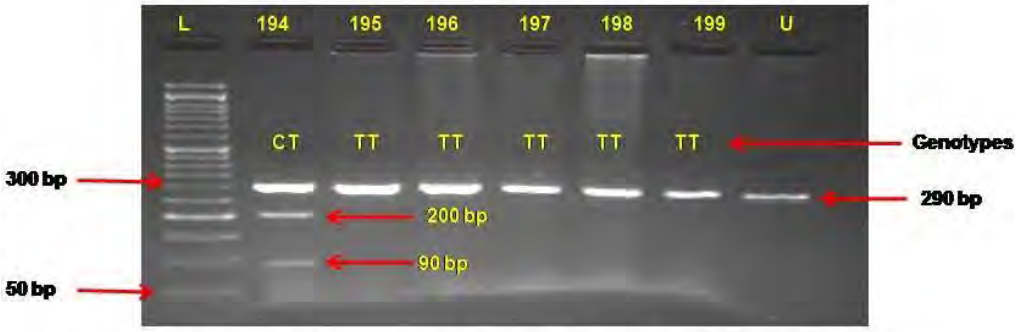


Figure 1. Polymorphic PCR-RFLP (SPP1G1–*Bsr*I) Intron-4 region of SPP1G1 gene resolved in 3% agarose gel electrophoresis in Gaolao Cattle.
Where, L= 50bp DNA Ladder,
GB 194-199= Gaolao Cattle Numbers, U- Uncut
TT=Genotypes (290 bp), CT=Genotype (290,200 & 90 bp)

Figure 1.

SPP1G1-BsrI polymorphism and association Analysis. The association study of identified SPP1G1-*Bsr*I polymorphism with milk traits using SPSS revealed that there was significant difference in Milk yield and SNF % traits for CT genotype as compared to TT genotype at SPP1G1-*BsrI* locus (Table 2) in 208 experimental Gaolao animals. It was found that the milk yield (7.33±0.67) and SNF% (9.03±0.08) traits differed significantly at 5% level of significance for CT genotype as compared to TT genotype.

In SPP1G1-*Bsr* Locus, the PCR products of the Intron-IV region of SPP1gene representing TT and CT genotypes were sequenced which revealed G-A SNP at 32nd position in 236 bp SPP1 gene sequence (Figure 2).

Table 2. Average effect of milk production traits at polymorphic SPP1G1-*Bsr*I locus in Gaolao Cattle.[1]

Genotypes	N	Fat % ± SE	SNF % ± SE	Lactose% ± SE	Protein% ± SE	MY ± SE	LN± SE
TT	205	4.26±0.21	8.64±0.05	4.56±0.03	3.15±0.02	4.72±0.11	1.41±0.10
CT	03	4.73±0.23	9.03±0.08*	4.40±0.25	2.90±0.10	7.33±0.67*	1.15±0.67
Total	208	4.27±0.21	8.65±0.04	4.56±0.03	3.15±0.02	4.76±0.11	2.82±0.10

[1] * indicates statistically significant at 0.05 (P<0.05) level.

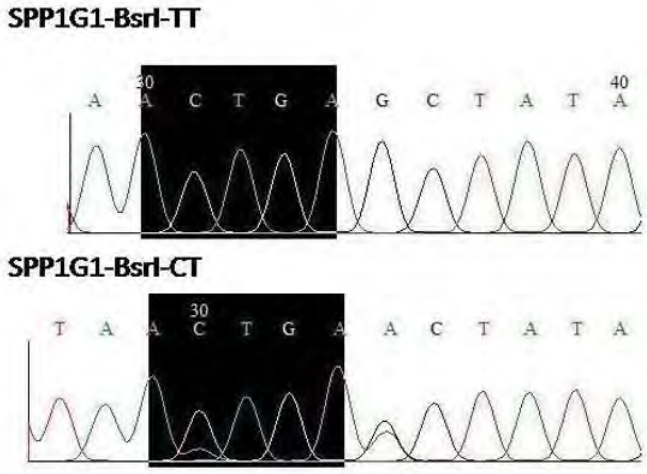


Figure 2. SNPs at G-A at 32nd position in 236 bp sequence of SPP1G1 amplicon in polymorphic SPP1G1-BsrI PCR-RFLP in Gaolao Cattle Population.

Figure 2.

SPP1G5-SSCP, association and DNA sequencing analysis of 362 bp fragment of SPP1 gene. The silver staining using 8% PAGE (49:1) for 6 hours (SPP1G5) revealed polymorphism with two patterns at SPP1G5-SSCP locus (Figure 3).

The frequency of these SPP1G5-SSCP patterns was (A=0.95 and B=0.05) in 64 Gaolao cattle. The comparison in SSCP Pattern A & B revealed significant differences for SNF% and Lactose% traits with SSCP patterns A exhibited as compared to patterns B (Table 3).

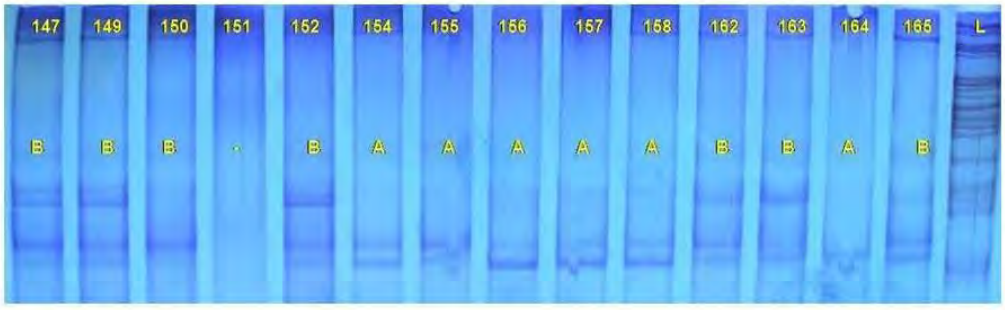


Figure 3. Polymorphic SPP1G5-SSCP patterns in Exon-4 region of 362 bp gene fragment of SPP1 gene (SPP1G5 primer) in Gaolao Cattle & visualised using 8% non-denaturing PAGE stained by silver staining.
Where; A – SSCP Pattern-1, B- SSCP Pattern-2 & L- Ladder
GB147- GB165 – Gaolao Cattle sample numbers

Figure 3.

Table 3. Average effect of milk traits at SPP1G5-SSCP in Gaolao cattle.[1]

SSCP pattern	N	Fat% ± SE	SNF% ±SE	Lactose% ± SE	Protein % ± SE	Milk yield ± SE	LN ± SE
A	61	4.80± 0.67	8.66±0.05*	4.41±0.04*	3.15±0.03	4.16±0.25	2.56±0.18
B	03	4.13±0.24	7.96±0.43	4.07±0.09	2.95±0.19	5.00±0.57	3.67±0.88
Total	64	4.77±0.64	8.62±0.06	4.39±0.04	3.14±0.03	4.20±0.24	2.61±0.18

[1] * indicates statistically significant at 0.05 (P<0.05) level.

The edited sequence of 322 bp was analysed using bioinformatics softwarewhich revealed SNPs A-G at 71st, A-C at 109th and A-C at 208th position in 322 bp (Figure 4). The alignment of 322 bp SPP1G5 amplified DNA sequence of SPP1 gene of Gaolao cattle with *Bos taurus* reference sequence (AY878328.1) revealed 02 computational SNPs viz; C-A at 109th position and C-A at 208th position.

Discussion

In present study the frequency of TT genotype at SPP1G1-*BsrI* locus was high as compared to CC genotype. The other researchers, Sharma *et al.* (2019) reported fixation of TT genotype in 147 animals of Sahiwal and Haryana cattle. The present study revealed significant difference in Milk yield and SNF % traits for CT genotype as compared to TT genotype (Table 3) in 208 Gaolao animals. It was found that the milk yield (7.33±0.67) and SNF% (9.03±0.08) traits differed significantly for CT genotype as compared to TT genotype. Leonard *et al.* (2005) screened 214 Holstein cows & reported association of C allele with increase in milk protein % & fat %. Khatib *et al.* (2007) reported association of C allele with increase in milk protein % & fat%. In conclusion, the findings of current study will fill the information gap about gene polymorphism at Intron-IV & Exon-IV region of SPP1 gene. The SPP1G1-*BsrI* locus was found significant for Milk yield and SNF % traits for CT genotype as compared to TT genotype. SPP1G5-SSCP polymorphism significant differences for SNF% and Lactose% traits with SSCP patterns A exhibited as compared to SSCP patterns B. The molecular genetic tools like PCR-RFLP, PCR-SSCP & SNP revealed genetic variation thus proving them as effective tools for marker trait association. The identified associations after validation in bigger database can be useful for selection and breeding decisions of indigenous dairy cattle.

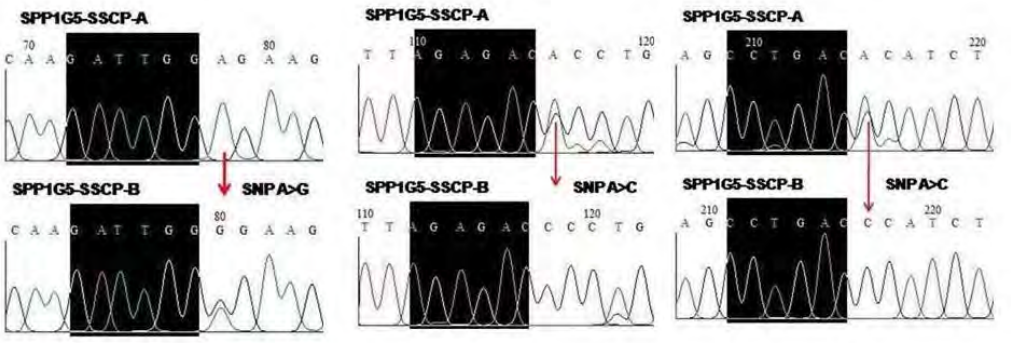


Figure 4. SNPs A-G at 71st, A-C at 109th and A-C at 208th position in 322 bp sequence of SPP1G5 amplicon in polymorphic SPP1G5-SSCP pattern-A & B in Gaolao Cattle Population.

Figure 4.

References

Khatib H., Zaitoun I., Wiebelhaus-Finger J., Chang Y.M. and Rosa G.J.M. (2007) Journal of Dairy Science 90(6):2966-2970.

Leonard S., Khatib H., Schutzkus V., Chang Y.M., and Maltecca C. (2005) Journal of Dairy Science 88(11): 4083-4086.

Sharma M., Singh S.P., Tiwari M., Sharma D., Goel R. *et al.* (2019) Journal of Animal Research 9(2):243-246.

532. Building *in vitro* tools for livestock genomics: chromosomal variation within the porcine PK15 cell line

M. Johnsson[1], J.M. Hickey[2] and M.K. Jungnickel[2*]

[1]Department of Animal Breeding and Genetics, Swedish University of Agricultural Sciences, Box 7023, 750 07 Uppsala, Sweden; [2]The Roslin Institute and Royal (Dick) School of Veterinary Studies, The University of Edinburgh, Midlothian, EH25 9RG, Scotland, United Kingdom; melissa.jungnickel@roslin.ed.ac.uk

Abstract

In this work, we called variants and used read coverage in combination with within-sample allele frequency to detect potential aneuploidy in two samples of an immortalise Porcine Kidney Epithelial (PK15) cell line. Porcine cell culture and cell lines are powerful tools for functional genomics and *in vitro* phenotypic testing of candidate causal variants. However, to be utilised for any type of genomic or variant interrogation assay, the genome sequence and structure of cell lines must be realised. Read coverage and within-sample allele frequencies suggest that several chromosomes are fully or partially aneuploid in PK15, including potential triploidy of chromosome 4 and tetraploidy of chromosome 17. We compared two PK15 cultured cells samples: a new American Type Culture Collection (ATCC) sample and one that has been utilised within the laboratory for an extended period (>10 years). The older PK15 sample showed evidence of additional structural variation and potentially clonal variation. We discuss implications for livestock functional genomics.

Introduction

Cell lines are an essential tool for functional genomics and functional studies of potential causative genetic variants. Large-scale epigenomic projects like ENCODE rely heavily on cell lines as models of different cell types, and functional work on causative genes in livestock make use of the relatively few livestock cell lines that exist (e.g. Walker *et al.*, 2018).

Most CRISPR screens are conducted *in vitro*. This approach requires an exquisitely precise and detailed knowledge of the genome sequence of target cell lines. It is well established that such sequence details are essential prerequisites to design guide RNAs and direct Cas9/dCas9 function in CRISPR manipulation and perturbation assays. Less well known is the contribution of genome structure and cell line aneuploidy to CRISPR-genome interrogation. Aneuploidy, in which cells have an abnormal number of chromosomes, has been shown to have a devastating impact on cell phenotype and gene expression (Sheltzer *et al.*, 2012). To our knowledge, an impact of aneuploidy on CRISPR gene manipulation has not been previously considered.

For this project, we used short read sequencing to characterise the genome of PK15, a classic pig cell line established in 1955 or 1956 from the kidney of an adult pig (Harris, 1959; Ruddle, 1961). We called variants and used read coverage in combination with within-sample allele frequency to detect potential aneuploidy in two samples of the PK15 cell line.

Materials & methods

Samples and sequencing. We extracted DNA from two *in vitro* cultured PK15 cell lines. One PK15 cell line was ordered fresh from ATCC (product name: PK(15) CCL-33) and the other were PK15 cells that have been cultured at the Roslin Institute for >10 years. Whole-genome resequencing was performed on the Illumina platform by Genewiz (Essex) for the ATCC sample and Edinburgh Genomics (Edinburgh) for the Roslin sample.

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_532, © M. Johnsson *et al.* 2022

Bioinformatics. We mapped reads to the Sscrofa11.1 reference genome (Warr *et al.,* 2020) with BWA MEM version 0.7.12 (Li, 2013) after trimming adapter and low-quality ends with Trimmomatic version 0.36. We removed duplicate alignments with Picard version 2.9.0 and called variants with GATK 3.5 (McKenna *et al.*, 2010) using the HaplotypeCaller and the GVCF workflow. We performed base quality score recalibration by first calling a preliminary set of variants, then using it as input to the GATK base quality score recalibration tool, and then calling a final set of variants using the recalibrated alignments. We filtered single nucleotide variants using the standard GATK hard filters. In order to estimate the allele frequencies within each sample, we extracted the allelic depth for each biallelic variant with bcftools version 1.10.2. In order to detect large-scale copy number variation and aneuploidy, we divided the pig genome chromosomes 1-18, X and Y into 10 kbp windows, and counted the number of reads mapping in each window with BEDTools version 2.26.0.

Results

The sequencing resulted in approximately 30X (ATCC sample) and 40X (Roslin sample) coverage of the pig genome. Variant calling detected 9,028,359 single nucleotide variants. Out of those, 3,565,539 were called as homozygous alternate in both samples, that is, they represent positions where the PK15 cell lines is likely to be fixed for a different allele than the Sscrofa11.1 reference genome.

Read coverage per chromosome suggest that several chromosomes are aneuploid in PK15. Figure 1 shows the average read coverage for a 10 kbp window on chromosomes 1-18, X and Y. In both samples, chromosomes 4, 9 and 17 have high read coverage. There are also differences between samples, such as high read coverage on chromosome 12 in the Roslin sample, not seen in the ATCC sample. When it comes to the sex chromosomes, the ATCC sample appears to be XX, consistent with the literature (Ruddle, 1961), while the Roslin sample appears to have lost one X chromosome and be X0.

Read coverage within chromosomes also suggest that there are large-scale structural variants within chromosomes. Figure 2 shows the read coverage in 10 kbp windows within four selected chromosomes. Chromosome 1, which shows little evidence of gross aneuploidy at the whole chromosome level (Figure 1), still potentially has a duplication within the Roslin sample, at around 200 Mbp. Chromosome 4 appears mostly triploid, but with several regions of higher and lower coverage especially in the Roslin sample. Chromosome 17 appears to be tetraploid in the ATCC sample and triploid in the Roslin sample.

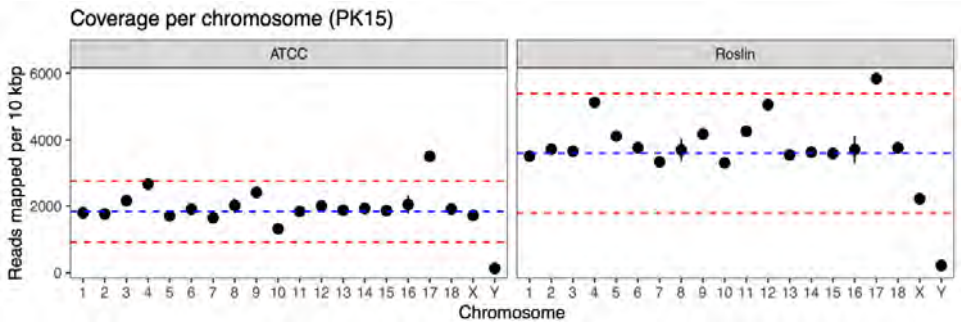


Figure 1. Average read coverage for 10 kbp windows on chromosomes 1-18, X and Y in the PK15 genome. Panels show the two samples of the cell line. Horizontal lines indicate the median coverage for a 10 kbp window (blue) as well as 0.5 × and 1.5 × the median (in red).

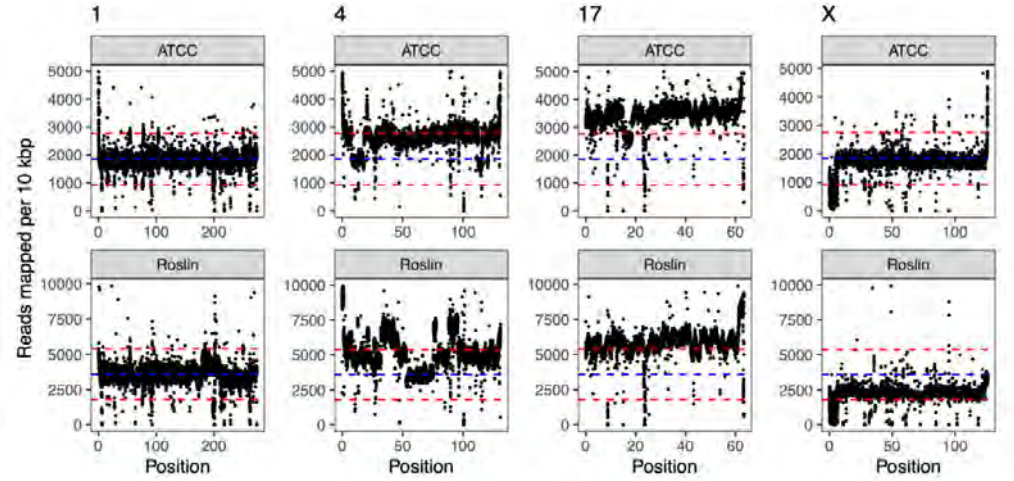


Figure 2. Read coverage in 10 kbp windows on chromosomes 1, 4, 17 and X in the PK15 genome. Panels show the two sample of the cell line. Horizontal lines indicate the median coverage for a 10 kbp window (blue) as well as 0.5 × and 1.5 × the median (in red).

Within-sample allele frequencies (Figure 3) for heterozygous sites also suggest aneuploidy. We find examples of chromosome that are likely diploid, triploid, tetraploid and haploid:

- For a diploid chromosome, one would expect a frequency of 1/2. Considering chromosome 1, the ATCC sample appears mostly diploid. In the Roslin sample, however, there are two distinct modes of allele frequency, suggesting that some part of the cell line has an aneuploid karyotype.
- For a triploid chromosome, allele frequencies should instead be 1/3 and 2/3. In both samples, chromosome 4 shows two allele frequency modes, suggesting triploidy.
- For a tetraploid chromosome resulting from a duplication of both homologous chromosomes, the expected allele frequency is 1/2. On chromosome 17, the ATCC has allele frequency centred on 0.5, while the Roslin sample shows two modes. As they both have high read coverage (Figure 2), we hypothesise that chromosome 17 is tetraploid in the ATCC sample and mostly triploid in the Roslin sample.
- For a haploid chromosome, only frequencies 0 and 1 are possible. In the Roslin sample, the X chromosome has very few heterozygous sites, suggesting that it is haploid.

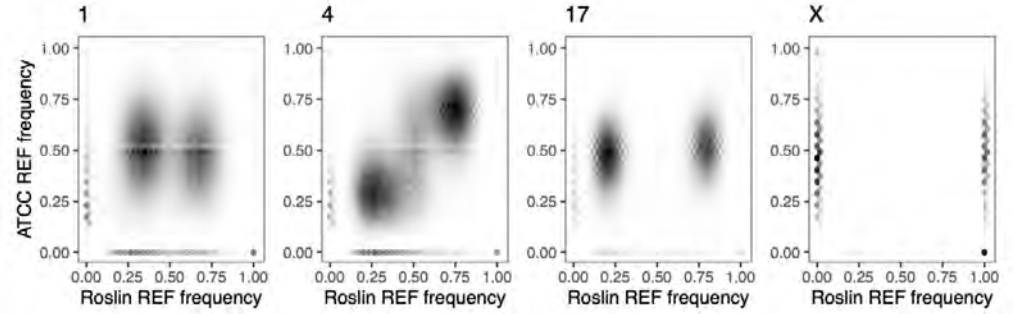


Figure 3. Within-sample allele frequency for the reference allele, calculated from allelic read counts, for single nucleotide variants that are heterozygous in at least one of the samples on chromosomes 1, 4, 17 and X.

Discussion

Our results suggest that the PK15 cell line has aneuploidies and complex structural variants in its genome. While such genome anomalies have not impacted over-expression studies that have been popular for the last 10-15 years (see review by (Prelich, 2012)), this cell line aneuploidy and clonal variation will directly impact CRISPR genome editing/genome manipulation studies; a genome editing assay with 20% homology-directed repair efficiency would generate repair rates of 4% in diploid and 0.8% in triploid genome sites respectively.

We are currently generating a more thorough bioinformatic to detect structural variants from short reads and segment the genome by ploidy. However, to fully describe the genome structure of a cell line, both long read sequencing and genome assembly is required. While currently too time consuming and expensive for routine cell line characterization, a method to examine cell line aneuploidy in more detail should be considered for key, routinely genome-manipulated and edited cell lines.

In summary, we suggest that genome aneuploidy should be considered when planning genome editing assays in *in vitro* cell lines.

References

Harris, M. (1959) Cancer Res 19, 1020–1024.

Li, H. (2013) arXiv:1303.3997. https://arxiv.org/abs/1303.3997

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, *et al.* (2010) Genome Res 20, 1297–1303. https://doi.org/10.1101/gr.107524.110

Prelich, G. (2012) Genetics 190, 841–854. https://doi.org/10.1534/genetics.111.136911

Ruddle, F.H. (1961) Cancer Res 21, 885.

Sheltzer, J.M., Torres, E.M., Dunham, M.J., Amon, A. (2012) Proc Natl Acad Sci USA 109, 12644–12649. https://doi.org/10.1073/pnas.1209227109

Walker, L.R., Engle, T.B., Vu, H., Tosky, E.R., Nonneman, *et al.* (2018) PLoS Genet 14, e1007750. https://doi.org/10.1371/journal.pgen.1007750

Warr, A., Affara, N., Aken, B., Beiki, H., Bickhart, D.M., *et al.* (2020) GigaScience 9, giaa051. https://doi.org/10.1093/gigascience/giaa051

533. Conserved non-coding elements of the bovine genome

N.K. Kadri* and H. Pausch

ETH Zürich, Universitätstrasse 2, 8092, Switzerland; naveen.kadri@usys.ethz.ch

Abstract

Conserved non-coding elements (CNEs) are common in the eukaryotic lineage. These sequences with unknown function are conserved over millions of years of evolutionary divergence. We compared bovine and human non-coding sequences and identified 23,401 unique CNEs with a median size of 226 bp (up to 1,884 bp) spanning 6.37 mb of the bovine genome. Most CNEs were found in the intergenic region and were spread non randomly in clusters across bovine autosomes. Most intergenic CNEs and their spatially associated genes were conserved in syntenic blocks across the two genomes. The associated genes were significantly enriched for 'developmental process' and 'anatomical structure development'. Further, these elements showed strong evolutionary constraint compared to their surrounding sequences and overall harboured 32% less variation than the protein coding sequences. These observations suggest that the CNEs detected are enriched for functional elements that may be essential for processes that are evolutionarily conserved across species.

Introduction

Non-coding genomic sequences that exhibit conservation across species are common in the eukaryotic lineage (Vavouri *et al.*, 2007). These sequences are commonly referred to as conserved non-coding elements (CNEs). Comparative genomics approaches have revealed CNEs even in distantly related species that may have diverged more than 400 million years ago (Polychronopoulos *et al.*, 2017). Although a regulatory role, particularly for genes involved in development processes is speculated and experimentally confirmed for some CNEs (Visel *et al.*, 2008), a definite functionality triggering the extreme conservation of non-coding sequence has not been fully elucidated (Harmston *et al.*, 2013; Polychronopoulos *et al.*, 2017). We herein survey the cattle genome for CNEs and examine their distribution, associated genes, and the functional constraint on them.

Materials & methods

Identification of CNEs. Conserved non-coding elements between cattle and human genomes were identified using CNEFinder (Ayad *et al.*, 2018). Soft masked reference sequences and annotation files of the human (GRCh38) and bovine (ARS-UCD1.2) genomes were retrieved from ensembl (v104). Following masking of the coding regions in the two genomes, non-coding sequences longer than 150 bp with at least 95% similarity between cattle and human were considered as CNE.

The evolutionary constraint on the identified CNEs and the surrounding sequence was quantified by counting the number of variants detected per 100 bp of the sequence. A variant catalogue containing 29.4 million variants previously established from whole genome sequencing data of 266 animals from 11 cattle breeds was used for this purpose (Kadri *et al.*, 2021).

Synteny and gene ontology enrichment analysis of CNE-associated genes. CNEs located outside annotated genes (n=14,025) were examined if their spatially associated genes are conserved in syntenic blocks across the two genomes. Specifically, we quantified the number of genes within 1 mb of the CNE on the bovine genome (CNE-associated genes) that had an orthologous gene (retrieved using Ensembl rest API; https://rest.ensembl.org/) within 1 mb of the corresponding conserved sequence on the human genome. CNE-associated genes conserved in synteny were then analysed for Gene Ontology (GO) enrichment. The

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_533, © N.K. Kadri and H. Pausch 2022

enrichment analysis was carried out on http://geneontology.org/. Statistical significance of enrichment was estimated using a binomial test, and Bonferroni correction was applied to account for multiple testing.

Simulation. A simulation study was carried out to investigate if the location of CNEs along the bovine genome matches with expectations assuming random placement. New locations were randomly assigned to CNEs on each chromosome restricting their placement to non-coding regions. This process was repeated 400,000 times and the mean distance between randomly placed CNEs was calculated for each replicate.

Results

CNEs appear in clusters and mostly reside in intergenic sequences. We searched for non-coding sequences longer than 150bp in the bovine genome that had more than 95% similarity with human non-coding sequences. This screen revealed a total of 23,401 unique bovine conserved non-coding elements (72 CNEs mapped to multiple regions of the human genome), with a median size of 226bp (up to 1,884 bp) covering 6.37 mb of the bovine genome. The density of CNEs varied considerably across chromosomes. While the mean density of CNEs was 2.5 kb/mb along the bovine genome, it ranged from 0.28 kb/mb on BTA3 to 6.9 kb/mb on BTA18 (Figure 1A). Most CNEs, as expectedwere in the intergenic part of the bovine genome (n=14,025, 59.93%). Of the remainder (n=9,376), 9,324 CNEs mapped to 2,156 unique genes, that included 1,967 protein coding, 151 lncRNA, 14 pseudogenes, 14 miRNA, 1 SnRNA and 7 miscRNA genes. CNEs were non-uniformly spread across the non-coding bovine genome. On average a CNE was identified within 100.7 kb of a neighbouring CNE. Such a clustering is highly unlikely assuming a random placement of CNEs along the genome (P<2.5e-06). In none of the simulated replicates the mean distance between consecutive CNEs was less than 100.7 kb.

CNEs are associated with genes coding for developmental process. Among the 9,376 CNEs in the non-coding part of the annotated genes, 8,851 mapped to 1,967 protein coding genes. The majority of these CNEs (n=8,150; 92.08%; affecting 1,735 genes) had similarity with sequences within their human orthologous genes. A GO enrichment analysis revealed that these 1,735 genes were significantly (>1.7-fold enrichment; P<2.49e-50) enriched for GO terms 'regulation of gene expression and 'regulation of macromolecule metabolic process'.

We examined if CNEs located outside annotated genes (n=14,025) and their spatially associated genes (within 1 mb) are conserved in syntenic blocks in the two genomes. Interestingly for >90% of CNEs (n=12,760) at least one human orthologue was found for CNE-associated genes and for more than half of the CNEs (n=7,496; 53.4%) human orthologues were found for more than 50% of CNE-associated genes. The resulting list of 9,612 CNE-associated genes conserved in syntenic blocks was also enriched for GO terms 'developmental process' and 'anatomical structure development' (>1.2-fold enrichment; P<8.74e-19).

When the enrichment analysis was repeated considering only the genes closest to CNEs, the resulting 712 CNE-associated genes were also enriched for GO terms 'developmental process' and 'anatomical structure development' (>2.3-fold enrichment; P<3.96e-45). In addition, GO terms 'nervous system development', 'multicellular organism development', 'system development', 'cell differentiation', 'cellular developmental process', and 'anatomical structure morphogenesis' were also significantly enriched (>2.3-fold enrichment; P<8.03e-42) among the genes closest to CNEs.

CNEs are depleted for variants. Next, we compared the evolutionary constraint on CNEs and their surrounding sequences by counting the number of overlapping variants. Variability was assessed separately for CNEs mapping inside and outside of the annotated genes. Nucleotides overlapping CNEs were less variable than their surrounding sequence (Figure 1B). Overall, CNEs were 32% less variable than the

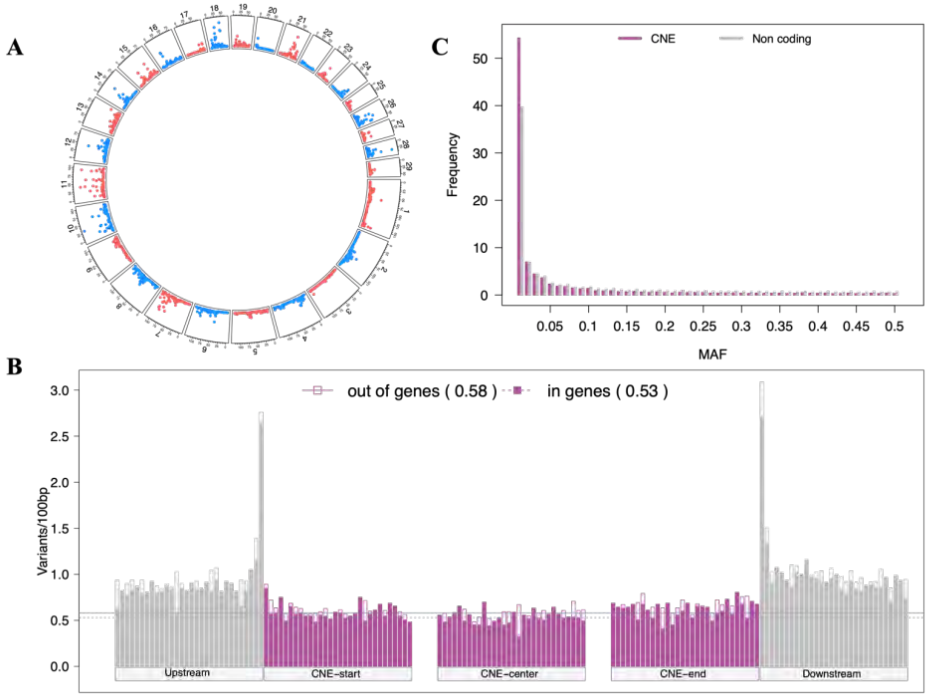


Figure 1. Distribution and functional constraint on the conserved non-coding elements (CNEs) in the bovine genome. The distribution (percentage of nucleotides overlapping CNEs in bins of 1 mb) along the bovine autosomes (A). The variability of nucleotides in and around the CNEs, quantified as number of variants per 100bp (B). CNEs mapping to inside and outside of the annotated genes are respectively represented with coloured and un-coloured bars. Minor allele frequency distribution of variants overlapping the CNEs (C).

coding sequence (0.56 per 100bp vs 0.81 per 100 bases; $P<2.2e-16$, Fisher's exact test). The constraint on CNEs mapping to annotated genes was significantly higher than for those in intergenic regions (0.53 vs 0.58 variants/100bp; $P=3.1e-17$, Fisher's exact test). Variants overlapping CNEs were enriched in lower Minor allele frequency (MAF) bins compared to variants overlapping non-conserved non-coding sequences in the bovine genome (Figure 1C).

Discussion

We identified CNEs spanning 6.3 mb of the bovine genome that have survived ~100 million years of evolution since the last cattle-human common ancestor (Liu *et al.*, 2006). Consistent with previous studies in other vertebrates, the identified CNEs appeared in clusters in the bovine genome (Woolfe *et al.*, 2005), mostly in so-called 'genomic regulatory blocks' with their associated genes (Polychronopoulos *et al.*, 2017). Genes in these syntenic blocks, as described previously (Harmston *et al.*, 2013), were enriched for those involved in developmental processes. Our study further revealed that nucleotides overlapping CNEs were less variable than the surrounding sequences and harboured variants enriched in low MAF bins suggesting purifying selection. Taken together, these observations suggest that the CNEs detected in this study are enriched for functional elements that may be essential for processes that are evolutionarily conserved across species.

We hypothesize that a larger fraction of the non-protein coding bovine genome is under purifying selection than detected in this study. The number of CNEs identified depends on the evolutionary divergence between the compared species and the thresholds used for similarity and sequence length (Harmston *et al.*, 2013). Comparison of the bovine genome with more closely related species, for example, other members of the Bovidae, is likely to yield a much larger set of lineage-specific CNEs (Chen *et al.*, 2019). An analysis of distantly related species, on the other hand, is expected to yield less CNEs but might increase the signal-to-noise ratio by filtering out slow-evolving neutral sequences (Woolfe *et al.*, 2005).

As more and more trait-associated variants are detected in non-coding regions of the genome (Georges *et al.*, 2018), systematic prediction of generally ignored non-coding functional features (e.g. Kadri *et al.*,2021) and detection of elements conserved without known functions using comparative genomics approaches seem necessary towards deciphering the molecular mechanism underlying statistical association.

References

Ayad L.A.K., Pissis S.P., Polychronopoulos D. (2018) Bioinformatics 34, i743–i747. https://doi.org/10.1093/bioinformatics/bty601

Chen L., Qiu Q., Jiang Y., Wang K., Lin Z., *et al.* (2019) Science (80-.). 364. https://doi.org/10.1126/science.aav6202

Georges M., Charlier C., Hayes B. (2018). Nat. Rev. Genet. 20. https://doi.org/10.1038/s41576-018-0082-2

Harmston N., Barešić A., Lenhard B. (2013). Philos. Trans. R. Soc. B Biol. Sci. 368. https://doi.org/10.1098/rstb.2013.0021

Kadri N.K., Mapel X.M., Pausch H. (2021). Commun. Biol. 4, 1–13. https://doi.org/10.1038/s42003-021-02725-7

Liu G.E., Matukumalli L.K., Sonstegard T.S., Shade L.L., Van Tassell C.P. (2006) BMC Genomics 7, 1–13. https://doi.org/10.1186/1471-2164-7-140

Polychronopoulos D., King J.W.D., Nash A.J., Tan G., Lenhard B. (2017). Nucleic Acids Res. 45, 12611–12624. https://doi.org/10.1093/nar/gkx1074

Vavouri T., Walter K., Gilks W.R., Lehner B., Elgar G. (2007). Genome Biol. 8. https://doi.org/10.1186/gb-2007-8-2-r15

Visel A., Prabhakar S., Akiyama J.A., Shoukry M., Lewis K.D. (2008) Nat. Genet. 40, 158–60. https://doi.org/10.1038/ng.2007.55.Ultraconservation

Woolfe A., Goodson M., Goode D.K., Snell P., McEwen G.K., *et al.* (2005). PLoS Biol. 3. https://doi.org/10.1371/journal.pbio.0030007

534. A massively parallel reporter assay to screen bovine regulatory variants

M.D. Littlejohn[1,2*], T.J. Lopdell[1], A. Trevarton[3], J. Moody[3], K.M. Tiplady[1], K.A. Burborough[3], C.P. Prowse-Wilkins[4], A.J. Chamberlain[4], M. Goddard[4] and R.G. Snell[3]

[1]Livestock Improvement Corporation, Hamilton, New Zealand; [2]Massey University, New Zealand; [3]The University of Auckland, New Zealand; [4]Agriculture Victoria, Australia; mathew.littlejohn@lic.co.nz

Abstract

Differentiating causative variants from linked variants is a major challenge of modern genetics. Here, we describe development of a massively parallel reporter assay (MPRA), an *in vitro* method that enables simultaneous functional testing of thousands of candidate causative regulatory variants. Candidates for the MPRA were identified through analysis of two large molecular quantitative trait loci (QTL) datasets derived from sequencing of bovine mammary biopsies, comprising an expression QTL resource (n=371 cows), and a histone QTL dataset (n=99 cows). Candidates that overlapped with lactation QTL discovered from genome-wide association studies were further prioritised to yield a final library of 242,340 oligos representing 24,234 regulatory sequences for expression testing. This paper gives a brief description of the MPRA method, our candidate selection criteria, cloning approach, and the composition of the final oligo pool for expression analysis.

Introduction

Genome-wide association studies (GWAS) have identified thousands of highly associated variants for myriad traits. However for a given locus, identifying which variants are responsible for phenotypic variation is challenging, given that variants in strong linkage disequilibrium (LD) may appear equivalently associated. An emerging, lab-based method to differentiate between functional regulatory variants and those that are merely statistically correlated is the massively parallel reported assay (MPRA). While single gene reporter assays have been applied for decades, Tewhey *et al.* (2016) described a highly parallelised version of this approach that drastically changes the throughput potential of reporter-based tests. Here, pooled cloning approaches are used to quantify the expression of tens of thousands of alleles simultaneously. This method involves synthesis of oligonucleotide sequences (typically 100-200bp long; limited by microarray-based synthesis constraints) to test the ability of SNPs or short indels to drive expression of a reporter sequence. For a given biallelic variant, a typical design would therefore aim to compare short sequences that differ only by the two alleles of interest. These sequences are then cloned into an expression plasmid upstream of a reporter transcript, with constructs then transfected into the cell line of interest for expression testing. Unlike traditional reporter assays that use fluorescence or luminescence of the translated reporter to quantify expression, the MPRA uses RNA-sequencing of the expressed transcript, where barcodes engineered into the 3'UTR allow assignment of sequence reads to oligos (and by extension, alleles).

We have previously identified hundreds of QTL for bovine lactation traits (Lopdell *et al.* 2017; Tiplady *et al.* 2021). Using a large gene expression resource derived from sequencing of mammary biopsies from 371 lactating cows, we observe that expression QTL (eQTL) likely drive many of these effects, with regulatory mechanisms suggested for most loci. We have also generated chromatin immunoprecipitation-sequence data (ChIP-seq) for histone modifications using mammary biopsies from a subset of 99 animals from this eQTL dataset. These data have allowed identification of thousands of histone modification QTL (hQTL), suggesting further candidate causative variants that may underpin variation in lactation traits (Prowse-Wilkins *et al.*, *in preparation*). While these data highlight many candidate variants, individual loci are near-universally represented by multiple variants. To further fine map these signals, we herein report

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_534, © M.D. Littlejohn *et al.* 2022

development of a bovine MPRA, where we attempt to resolve hundreds of eQTL, hQTL, and lactation QTL to single causative nucleotides.

Methods

We targeted two major sources of QTL for variant selection and MPRA oligo design, described under the 'eQTL with lactation effects', and 'Mammary hQTL peak variants' subheadings below. A total of 3,144 haplotypes generated as part of a separate study were also included in the design, though this content was unrelated to the majority of the mammary/lactation focused content and thus is not described in detail in this paper.

eQTL with lactation effects. We intersected the list of lactation and milk FT-MIR QTL recently reported by Tiplady *et al.* (2021) with eQTL highlighted from the same study to target loci likely underpinned by regulatory mechanisms. Expression QTL were prioritised based on assessment of correlations with lactation/FT-MIR trait QTL, computed using $-\log_{10}$ *P*-values for the lists of variants between traits. Expression QTL selection differed from that described in Tiplady *et al.* (2021) in that the interval of variants used for correlation calculations was narrower (±10 kbp of the transcript in the current study), and the nominal correlation threshold applied was higher (minimum Pearson coefficient of 0.8 R^2). These criteria identified 139 lactation/ FT-MIR QTL:eQTL pairs, where we further categorised these signals based on three levels of priority. The categories included whether they presented hallmark genes that had been previously published in multiple lactation GWAS studies (first priority candidates), with second and third tier candidates showing more limited prior literature-based implication and less compelling correlations. The number of variants interrogated per locus was constrained on this basis, with a maximum attribution of 130, 80, and 30 variants for the first, second, and third tiers respectively. For selecting the subsets of variants from each locus, variants with an LD R^2>0.8 with the top-associated variant were targeted, with number constraints then applied to these lists. For oligos that contained multiple candidate variants, we also represented all haplotypic strings of these alleles, though additional constraints were applied to cap the total number of haplotypes for the most variable loci. We also included six negative control variants per gene, targeting sites from the 139 loci of interest that were unassociated with expression (*P*-value>0.75).

Mammary hQTL peak variants. We targeted candidate variants discovered through QTL-mapping of ChIP-seq data, based on quantitative histone modification assays for the H3K4Me1, H3K4Me3 and H3K27ac histone marks. These assays were applied to biopsies from 99 lactating cows that overlapped with the RNAseq population referenced in the preceding section. A manuscript describing this dataset is currently being drafted (Prowse-Wilkins *et al. in preparation*) and is further described in a separate WCGALP paper (Prowse-Wilkins *et al.* 2022). Candidates were selected if they overlapped a ChIP-seq peak, were significantly associated with histone modification for that peak, and were also associated with gene expression changes in *cis* (using the RNAseq dataset described above). However, unlike the '*eQTL with lactation effects*' candidates, we did not investigate haplotypes for the hQTL variant targets. We did, however, duplicate this content on the array for synthesis, a rule applied to all variants that weren't represented more than once via haplotypes (i.e. to provide additional redundancy in the oligo pool).

Cloning and barcoding strategy, oligo design and synthesis. We used the vectors and cloning strategy outlined by Tewhey *et al.* (2016). Exhaustive methods are reported in that paper, though we made modifications to their approach to enable greater flexibility regarding cloning of insert sequences, and the barcoding approach. For the former, we included additional restriction sites (SfiI, FseI, AsiSI, and NotI) to allow sticky end cloning options for integration of the oligo library and reporter transcript. Regarding barcoding, we adopted a hybrid approach that included 15bp barcodes generated as part of oligo synthesis (similar in approach to Choi *et al.* 2020) while maintaining capability to expand barcode diversity by PCR,

post-synthesis (as per Tewhey *et al.* 2016). For the 15bp 'hard-coded' barcodes, each oligo was represented by 10 sequences. For oligo design, custom scripts were implemented to extract 135bp strings from the ARS-UCD1.2 genome, centred on the variant of interest. Following *in silico* addition of barcodes, Gibson homology sequences, and restriction and primer binding sites, 208bp oligos were submitted to Agilent Technologies (Australia) for synthesis using the 244k format SurePrint oligo platform.

Results

Our selection criteria yielded 24,234 candidate regulatory oligos targeting 4,814 biallelic variants. Among the two major categories of targets, there were 15,600 haplotypes representing 2,787 variants for the 'eQTL with lactation effects' pool. Figure 1 shows the eQTL-implicated genes from this selection, highlighting 13, 18, and 55 genes for the first, second, and third priority tiers of candidates. For the 'mammary hQTL peak variant' selections, 3,976 regulatory oligos representing 994 variants were included. A total of 1,514 regulatory oligos were also included as negative controls in the design. When paired with ten 15bp barcodes per oligo, a final library 242,340 oligos was submitted for synthesis. At time of writing, the oligo library had been synthesised with characterisation, cloning, and expression testing pending.

Discussion

We herein report design and synthesis of a high density MPRA for the functional testing of bovine candidate regulatory variants. This experiment is motivated by our observation that many lactation QTL appear to be underpinned by regulatory mechanisms(Tiplady *et al.* 2021; Lopdell *et al.* 2017), a finding that is well-described for human complex traits and diseases(Maurano *et al.* 2012). Further, most loci present multiple non-coding candidates. While this is true in most species (including humans), GWAS in livestock species such as cattle present additional challenges, where selection has led to population structures that increase the range of LD and present a greater number of correlated variants as candidates(Goddard and Hayes 2009). An alternative approach to fine mapping is to test alleles for their functional consequences – using *in vitro* methods that remove variants from their native genomic and haplotypic contexts. In humans, a number of studies have demonstrated use of the MPRA to this end (e.g. investigation of melanoma (Choi *et al.* 2020) and hominid evolution variants (Jagoda *et al.* 2021)), though we are so far unaware of such methods being applied in domestic species.

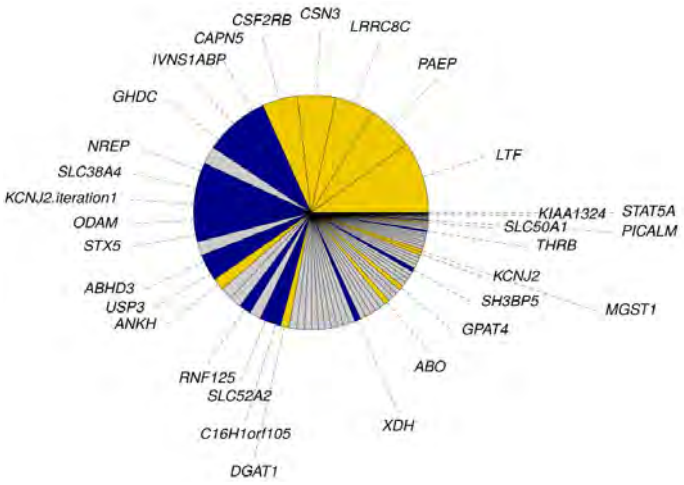


Figure 1. Genes with eQTL/lactation QTL targeted by the MPRA. Gene symbols shown for first (yellow) and second (blue) tiers only, slice area represents the number of haplotypes designed per gene.

We have taken a pragmatic design approach to the MPRA, representing hQTL and eQTL that overlap with loci associated with lactation and FT-MIR milk phenotypes. While cloning and transfection were pending at time of writing, our intention is to conduct expression testing in bovine mammary epithelial cells. Given the expression of major milk proteins in transformed mammary lines is limited at best (Huynh, Robitaille, and Turner 1991), we will target primary lactating mammary cells for transfection and have optimised methods to this end.

References

Choi J, Tongwu Z, Vu A, Ablain J, Makowski M *et al.* 2020. 'Massively Parallel Reporter Assays of Melanoma Risk Variants Identify MX2 as a Gene Promoting Melanoma.' Nature Communications 2020 11:1 11 (1): 1–16.

Goddard, ME., and Hayes BJ. 2009. 'Mapping Genes for Complex Traits in Domestic Animals and Their Use in Breeding Programmes.' Nature Reviews Genetics 2009 10:6 10 (6): 381–91.

Huynh HT, Robitaille G, and Turner JD. 1991. 'Establishment of Bovine Mammary Epithelial Cells (MAC-T): An in Vitro Model for Bovine Lactation.' Experimental Cell Research 197 (2): 191–99.

Jagoda, E, Xue JR, Reilly SK, Dannemann M, Fernando Racimo *et al.* 2021. 'Detection of Neanderthal Adaptively Introgressed Genetic Variants That Modulate Reporter Gene Expression in Human Immune Cells.' Molecular Biology and Evolution, October.

Lopdell TJ., Tiplady K, MStruchalin M, Johnson TJ, Keehan M *et al.* 2017. 'DNA and RNA-Sequence Based GWAS Highlights Membrane-Transport Genes as Key Modulators of Milk Lactose Content.' BMC Genomics 18 (1): 968.

Maurano M, Humbert R, Rynes E, Thurman RE, Haugen E *et al.* 2012. 'Systematic Localization of Common Disease-Associated Variation in Regulatory DNA.' Science (New York, N.Y.) 337 (6099): 1190–95.

Prowse-Wilkins CP, Lopdell TJ, Xiang R, Vander Jagt CJ, Littlejohn MD *et al.* 2022. 'Regulatory QTL and Exon Expression QTL in the Mammary Gland of Dairy Cows.' In World Congress on Genetics Applied to Livestock Production.

Tewhey R, Kotliar D, Park DS, Liu B, Winnicki S *et al.* 2016. 'Direct Identification of Hundreds of Expression-Modulating Variants Using a Multiplexed Reporter Assay.' Cell 165 (6): 1519–29.

Tiplady KM., Lopdell TJ, Reynolds E, Sherlock RG, Keehan M *et al.* 2021. 'Sequence-Based Genome-Wide Association Study of Individual Milk Mid-Infrared Wavenumbers in Mixed-Breed Dairy Cattle.' Genetics Selection Evolution 2021 53:1 53 (1): 1–24.

535. Detection of causative genes for resistance against cardiomyopathy syndrome in Atlantic salmon using omics data

M.L. Aslam[1*], S. Gonen[2], S.A. Boison[2], E. Leder[1], M. Baranski[2], A. Norris[2], N. Robinson[1], S. Koga[1], T.K.K. Østbye[1] and B. Dagnachew[1]

[1]Nofima, P.O. Box 210, 1431 Ås, Norway; [2]Mowi Genetics AS, P.O. Box 4102 Sandviken, 5835 Bergen, Norway; *luqman.aslam@nofima.no*

Summary

The objective of this study was to fine map the quantitative trait loci region of chromosome 27 which was previously reported as a major contributor of total genetic variation for resistance against cardiomyopathy syndrome in Atlantic salmon. The fine mapping using whole genome resequencing of individuals increased resolution of the targeted genomic region (Ssa27:10-12 Mbps), resulting in uncovering the two immune related genes (*HA1F* and *MR1*). Transcriptome analysis revealed only two differentially expressed genes, *MR1* and *HA1F* across resistant vs susceptible individuals. The gene *MR1* was highly expressed in the susceptible group with $\log_2$ fold change of 6.405 compared to resistant individuals. Contrary to *MR1* expression, *HA1F* gene was lowly expressed ($\log_2$ fold change of 2.099) in susceptible group. Moreover, the proteomics results partially validated the results of transcriptomics with the product of *MR1* being differentially produced across resistant vs susceptible individuals.

Introduction

Cardiomyopathy syndrome (CMS) is a severe cardiac disease of Atlantic salmon caused by piscine myocarditis virus (PMCV). The outbreak of CMS typically occurs several months after the sea transfer (during second year of production cycle) when the fish is a relatively large size (~3.6 kg). In most cases, no clinical signs are observed until the onset of mortality leading to compromised fish health/welfare and significant economic losses for salmon producers. The disease pathology develops slowly, and minor stressors such as treatment for sea lice infestation in the presence of CMS may trigger mass mortality (Garseth Å *et al.* 2018).

Cardiac histopathological examination of affected fish presents severe inflammation and necrosis mostly limited to the endocardium and spongy myocardium in the atrium and ventricle (Rodger *et al.* 2014), with a higher level of infiltration through mononuclear cells i.e. lymphocytes and macrophages (Ferguson *et al.* 1990; Timmerhaus *et al.* 2012). The heritability for resistance against CMS falls in the range of low to high depending on different factors e.g. type of data (survival, virus load or histology), challenge environment (field vs controlled) and/or population. Additionally, genomic regions on chromosomes 12 and 27 had previously been detected with effects on genetic variation for resistance against CMS (Boison *et al.* 2019; Hillestad *et al.* 2020). The genomic region of chromosome 27 had the largest effect, explaining ~30% of the genetic variance. In the current study, we aimed to fine map the genomic region(s) focusing mainly the quantitative trait loci (QTL) detected on chromosomes 27 to identify causative gene/mutation underlying the host resistance against CMS.

Materials & methods

The test population of this study originated from the breeding nucleus of MOWI ASA, with sampling performed from four year classes (YC2016, YC2017, YC2019 and YC2021). Individuals were selected from these four year classes were selected for two purposes: (1) whole genome resequencing to fine map the detected QTL region(s), where fin clips were sampled; and (2) functional level validation of causative genes through expression profiling (tissue from atrium and blood plasma for transcriptomics and proteomics,

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_535, © M.L. Aslam *et al.* 2022

respectively). Whole genome resequencing was performed on samples selected from YC2016, YC2017, and YC2019 while samples for transcriptomics and proteomics were obtained from YC2021.

All selected individuals had pedigree information available, were phenotyped for CMS either through company's routine challenge test or field outbreak and genotyped using NOFSAL03 Affymetrix axiom array (~55K SNPs). The selection of individuals was systematically performed limiting the number of sibs per family (offspring per sire ≤4 and dam ≤2), considering genomic breeding value and presence/state of the previously detected significantly associated haplotype. Individuals were categorized into three groups, resistant FUL (gEBVs>1.25 SD and 2 copies of the favorable haplotype), intermediate HET, (gEBVs>0.625 SD and one copy of the favorable haplotype), and susceptible DEL (gEBVs<1.25 SD and zero copy of the favorable haplotype). One hundred and fifty-seven samples were selected for whole genome resequencing from each of the three year classes distributed to all three groups (FUL, HET, and DEL). Additionally, 30 samples (10 from each FUL, HET, and DEL category) were selected from YC2021 for transcriptomics and mass spectrometry-based label-free proteomics, with each individual sampled for both atrium and blood plasma samples.

The whole genome resequencing was performed using the BGISEQ platform with 150 bp PE reads. The raw sequence reads were trimmed/filtered using Trimmomatic, and subsequently quality sequence data was aligned to the Atlantic salmon reference genome sequence (assembly ICSASG_v2) using BWA-MEM (version 0.7.13-r1126). The *mpileup* function of SAMtools (version 1.10) was then used to call variants and the call option of *bcftools* was used to call the genotype at each variant site for each individual. Informative SNP variants were annotated using the ensembl variant effect predictor to categorize variants based on their predicted effects (McLaren *et al.* 2016).

The transcriptome sequencing was performed using Illumina NovaSeq 6000 platform with 150bp paired end sequencing. After quality trimming/filtering of reads, higher quality reads were processed using STAR program (version 2.7.9a) to obtain count values of genes for each individual. The count values were then analyzed for differential expression of genes using edgeR (Robinson *et al.* 2009).

The plasma samples from each individual were digested by trypsin. The tryptic peptides were analyzed using a Q-Exactive hybrid quadrupole-orbitrap mass spectrometer, and the mass spectral data were analyzed by MaxQuant (version 2.0.3.0) for protein identification and quantification (Cox and Mann 2008). The statistical analyses were then carried out by Perseus, version 1.6.15.0 (Tyanova *et al.* 2016) and Unscrambler.

The selected individuals (n=157) used for fine mapping had phenotype information available, were genotyped using axiom array along with the new SNPs discovered through sequencing. A genome-wide association analysis was performed using the following linear mixed animal model implemented in GCTA program with '--mlma-loco' function (Yang *et al.* 2011).

$$y = \mu + Xb + Zu + e$$

where y is a vector of 'n' phenotypic records, μ is an overall mean, b is a vector of fixed effects of year class, u is a vector of additive genetic effects distributed as $u \sim N(0, G\sigma_u^2)$, where σ_u^2 is the additive genetic variance, G is a genomic relationship matrix; X and Z are the corresponding incidence matrices for fixed and additive effects, respectively, and e is the vector of random residual effects with $e \sim N(0, I\sigma_e^2)$.

Results & discussion

As the study was mainly focused on the major QTL of chromosome 27 specifically the region from 10 to 12 Mbp, hence 8,813 variants were discovered through sequencing within the region of interest. Out of these identified variants, 5 SNPs were annotated as high impact, broadly covering ~13 protein coding genes. Out of these ~13 genes, two genes (*MR1*, major histocompatibility complex class I-related gene protein-like and *HA1F*, class I histocompatibility antigen, F10 alpha chain-like) were directly related with immune function. Genome-wide association analysis including added SNPs discovered through resequencing produced a strong signal at the region of interest with increased resolution (Figure 1). However, the added SNPs did not show higher significance level than what was obtained from already available SNPs in the medium density array (Figure 1).

Pairwise comparison for differential expression of genes revealed only two differentially expressed genes, *MR1* and *HA1F* across FUL vs DEL individuals that crossed Benjamini and Hochberg adjusted *P*-values (Figure 2). The gene *MR1* was highly expressed in susceptible (DEL) group with $\log_2$ fold change 6.405 compared to resistant (FUL) individuals. Contrary to *MR1* expression, *HA1F* gene was lowly expressed in susceptible group with 2.099 $\log_2$ fold differences (Figure 2). The proteomics results validated the findings of transcriptomics showing that '*A0A1S3PYN4*' protein, which is a product of the *MR1* gene being significantly differentially produced across FUL vs DEL individuals. Both '*A0A1S3PYN4*' production and *MR1* gene expression was the highest in the susceptible group. The concordance of results at three levels across susceptible vs resistant groups, identified candidate genes through fine-mapping, transcriptomics, and proteomics certainly points to an underlying role of detected causative genes for resistance against CMS. The findings from the study indicates that the role of both genes is crucial for host to provide resistance against virus possibly through protein interactions at optimal level of expressions. As *HA1F* gene is expressed at a very low level (close to zero) in DEL and HET groups compared to FUL group which might be the causative factor for differential immune responses.

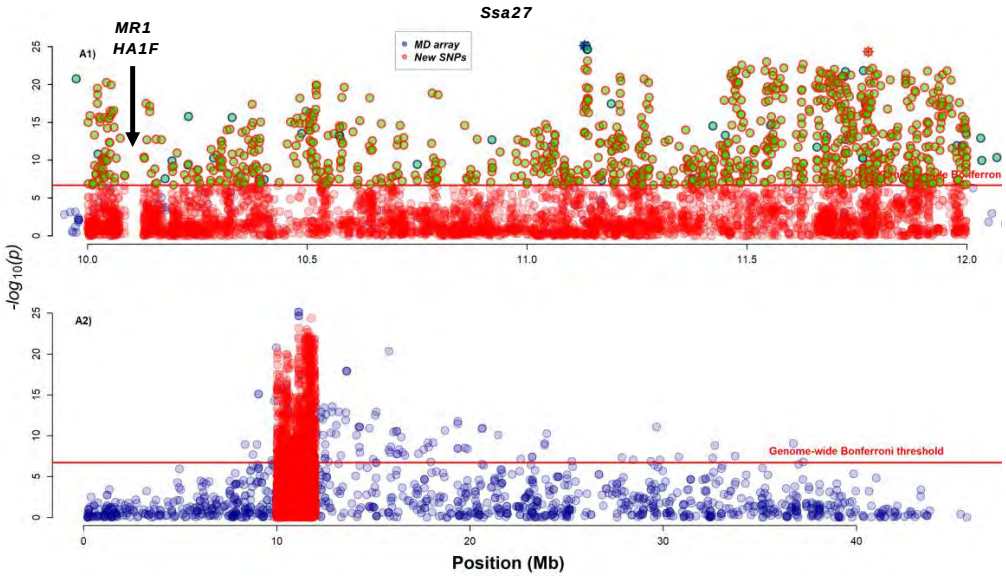


Figure 1. Genome-wide association of SNPs at chromosome 27 (Ssa27) depicting increased resolution using added SNPs from the sequence data. The solid red line is a genome-wide Bonferroni corrected threshold line. The asterisk blue and red SNPs are the highest significantly associated SNPs from the sequence and the array data, respectively. The black arow illustrates approximate position of candidate immune related genes.

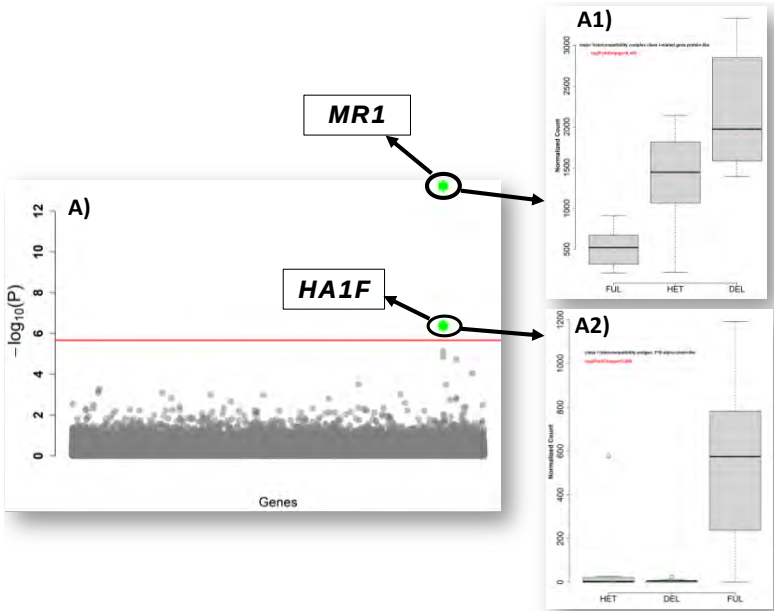


Figure 2. Significant differentially expressed genes across groups. (A); Two green asterisk genes crossed Hochberg adjusted threshold (solid red line), (A1); depicts distribution of expression levels of MR1 gene across individuals from all three groups, (A2); represents distribution of expression levels of HA1F gene across individuals from all three groups.

Acknowledgements

The results produced in current study are a part of 'CMS-Edit' and the 'Genomics4Robust' projects. CMS-Edit project was funded by The Research Council of Norway under grant agreement number 294504 while Genomics4Robust project is an internally funded project from Nofima AS under project number 13281.

References

Boison S., Ding J., Leder E., Gjerde B., Bergtun P.H. *et al.* (2019) Journal of Heredity 110(6): 727-737. https://doi.org/10.1093/jhered/esz042

Cox J., and Mann M. (2008) Nature Biotechnology 26(12): 1367-1372. https://doi.org/10.1038/nbt.1511

Ferguson H.W., Poppe T. and Speare D.J. (1990) Dis Aquat Organ 8(225-231.

Garseth Å H., Fritsvold C., Svendsen J.C., Bang Jensen B. and Mikalsen A.B. (2018) J Fish Dis 41(1): 11-26. https://doi.org/10.1111/jfd.12735

Hillestad B., H Kristjánsson Ó., Makvandi-Nejad S. and Moghadam H.K. (2020) Genes (Basel) 11(6): 608. https://doi.org/10.3390/genes11060608

Mclaren W., Gil L., Hunt S.E., Riat H.S., Ritchie G.R.S. *et al.* (2016) Genome Biology 17(1): 122. https://doi.org/10.1186/s13059-016-0974-4

Robinson M.D., Mccarthy D.J. and Smyth G.K. (2009) Bioinformatics 26(1): 139-140. https://doi.org/10.1093/bioinformatics/btp616

Rodger H.D., Mccleary S.J. and Ruane N.M. (2014) J Fish Dis 37(10): 935-939. https://doi.org/10.1111/jfd.12186

Timmerhaus G., Krasnov A., Takle H., Afanasyev S., Nilsen P. *et al.* (2012) BMC Genomics 13(1): 205. https://doi.org/10.1186/1471-2164-13-205

Tyanova S., Temu T., Sinitcyn P., Carlson A., Hein M.Y. *et al.* (2016) Nat Methods 13(9): 731-740. https://doi.org/10.1038/nmeth.3901

Yang J., Lee S.H., Goddard M.E. and Visscher P.M. (2011) Am J Hum Genet 88(1): 76-82. https://doi.org/10.1016/j.ajhg.2010.11.011

536. Genetic regulation of ear size in Slovenian and Greek sheep breeds

J. Klawatsch[1], D. Papachristou[2], P. Koutsouli[2], D. Seichter[3], I. Russ[3], M. Simčič[4], I. Bizelis[2] and I. Medugorac[1*]

[1]Population Genomics Group, Department of Veterinary Sciences, Ludwig Maximillians University Munich, Munich, Germany; [2]Agricultural University of Athens, Iera Odos, Athens, Greece; [3]Tierzuchtforschung e.V. Munich, Grub, Germany; [4]Biotechnical faculty, University of Ljubljana, Ljubljana, Slovenia; i.medugorac@lmu.de

Abstract

Natural variation in ear length is present in sheep (*Ovis aries*). *Anotia* describes the condition of the absence of the outer ear and *microtia* the condition of small ears. So far, the genetic basis of ear length regulation was studied in the Awassi, Duolang, Valle del Belice and Altay sheep breeds. The present study should add new insights to the genetics of ear length regulation in Slovenian and Greek sheep breeds. Here, a combined linkage disequilibrium and linkage analysis (cLDLA) approach for genetic mapping was applied. The conducted cLDLA on 166 individuals yielded two genome-wide and three suggestive significantly associated loci to the studied phenotype. In proximity to the strongest genetic signal the previously described *HMX1* locus is present. Our findings suggest that, while breed specific genetic regulation of ear length is possible, *HMX1* may be a main regulator of ear length in sheep breeds of a broad geographic region.

Introduction

In sheep (*Ovis aries*), natural variation in presence and length of head appendages could already be described more than 100 years ago in the case of horn (Arkell 1912) and the outer ear (Ritzman 1920; Lush 1930). If no outer ear structure is present, the condition is dubbed *anotia* and in the case of small ears *microtia*. Both are distinct to regular ear lengths of about 13 cm in our study cohort. As has already been shown in cattle (Allais-Bonnet *et al.* 2013), the genetic regulation of one head appendage (e.g. horn) can overlap with the gene regulatory network of another head appendage (e.g. eyelash). The same could apply to the presence and length of ears and horns in sheep, making *anotia* and *microtia* a very interesting breeding and developmental study target. To date, the genetics of the condition has been studied in the Awassi (Jawasreh *et al.* 2016; Jawasreh and Al-Omari 2020), Duolang (Gao *et al.* 2018), Valle del Belice (Mastrangelo *et al.* 2018) and Altay (He *et al.* 2020) sheep breeds. In any of those breeds, GWAS, local sequence alignment techniques or locus-specific SNP genotyping by PCR, yield different genes showing significant contributions to the described phenotype. The diversity of the results, pointing to the loci, *GATA6* (GATA binding protein 6), *DCC* (deleted in colorectal carcinoma), *CLRN1* (Clarin 1) and *HMX1* (H6 family homeobox 1), indicate a breed specificity in the genetic regulation of ear length. Here, we study the genetic basis of ear length variation in one Slovenian sheep breed, the Bovec sheep, and the seven Greek sheep breeds Argos, Chios Fat Tail, a local Peloponesus breed, Ikaria Fat Tail, Karpathos, Kasos and Lesbos Fat Tail and compare here investigated loci with previous described ones to examine commonalities and differences in the genetic regulation of ear length variation across sheep breeds.

Materials & methods

Animals, breeds, populations and phenotype recording. We measured the ear length (cm) of 166 (122 female and 44 male) sheep, aged two to 16 years. Twenty-two animals were from the Slovenian Bovec sheep breed and the remaining 144 from seven Greek breeds: Argos (20), Chios Fat Tail (33), a local Peloponesus breed (4), Ikaria Fat Tail (4), Karpathos (10), Kasos (6) and Lesbos Fat Tail (67). The ear lengths measured

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_536, © J. Klawatsch *et al.* 2022

ranged from about two to 15 cm and formed three categories: 1 'short-eared' with ear lengths of two to three cm, 2 'middle-eared' with six to nine cm and 3 'normal-eared' with ear lengths of ten to 15 cm. These three categories were used as phenotype for the genetic mapping study. According to our field observation, we consider category 1 as *anotia* and category 2 as *microtia*. The geographical origin and the distribution of ear length phenotypes of the studied individuals are shown in Figure 1.

Sampling, collection of biological material, DNA isolation and SNP genotyping. Blood samples were taken from studied animals in the course of routine examinations under good veterinary practice. Sampling took place in the period from 2018 to 2021 and samples were stored at -20 °C until lab analysis. DNA was isolated using the QIAmp DNA Blood Mini Kit (Qiagen, Valencia, USA) according to the manufacturer's instructions. SNP genotyping was conducted on the OvineSNP50 BeadChip (Illumina, San Diego, CA, United States) using standard procedures (http://www.illumina.com). The chip covers 54,241 SNPs.

SNP quality control, haplotyping, unified additive relationships (UAR) and genetic mapping. After genotyping and quality control, 166 animals with high genotype and phenotype quality remained in the data set. In order to impute missing genotypes and infer haplotype we used the Beagle software package (v 5.0) (Browning and Browning 2007; Browning *et al.* 2018) and included the genotype and pedigree information of all available ovine data from the in-house database, which also contained animals that are not phenotyped for ear length. To consider the genome-wide relationships between all animal pairs, a unified additive relationship (UAR) matrix was estimated according to Yang *et al.* (2010) by using the implementation in the R package snpReady (Granato *et al.* 2018). For performing the actual genetic mapping, a combined linkage disequilibrium and linkage analysis (cLDLA) approach was conducted as described in Meuwissen and Goddard (2001); and Gehrke *et al.* (2020). In addition to the genome-wide additive relationship matrix, we estimated a locus-specific IBD matrix and used ASReml (Gilmour AR 2009) to estimate the maximum likelihood, variance components, and fixed and random effects simultaneously. The following mixed linear model was applied:

$$y = X\beta + Z_1u + Z_2q + e \quad (1)$$

where y is the vector of the ear length phenotype; β is the vector of fixed effects including the overall mean (μ) and age at phenotyping; u is the vector of *n* random polygenic effects for each animal; q is the vector of

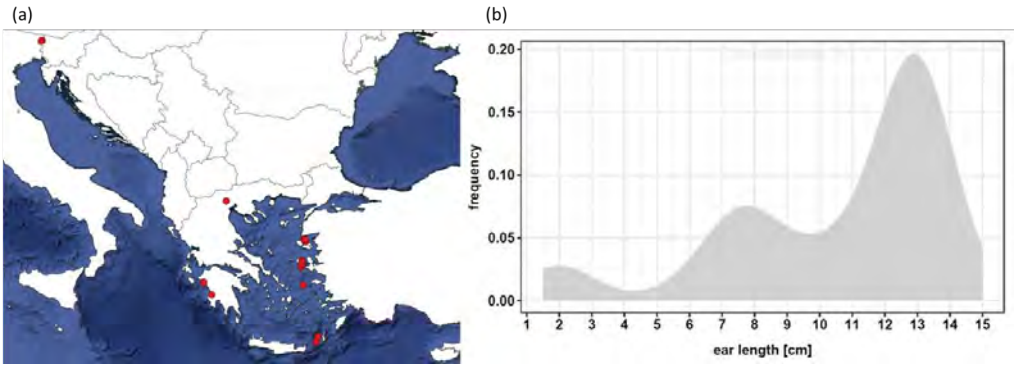


Figure 1. Map of the sampling locations of the studied breeds constructed with QGIS version 3.16.4 (QGIS.org 2021) (a) and distribution of the ear length phenotype in the study cohort (b).

random additive genetic effects of the putative QTL. Random residual effects were included in the vector e. The matrices X, Z_1 and Z_2 are the incidence matrices for the fixed and random effects. Finally, a likelihood ratio test (LRT) for the goodness-of-fit between the null hypothesis (H_0: model without a QTL effect) and the alternative hypothesis (H_1: model including a QTL effect) at each SNP interval midpoint was calculated.

Results

From the 166 studied individuals, 13 were short-eared, 50 medium-eared and 103 regular-eared. *LRT* values of cLDLA indicated two genome-wide and three chromosome-wide significant mapping results (Figure 2a). The strongest signal with an *LRT* value of 32.60 (P=1.13×10^{-8}) is on ovine chromosome 6 (OAR6) at position 6:114002547. In proximity to that position the *HMX1* locus can be found. Since a duplication of 76 bp in an evolutionarily conserved region near *HMX1* was previously described in Altay sheep (He *et al.* 2020) and later characterised in more detail in Awassi sheep (Jawasreh and Al-Omari 2020), we genotyped our mapping population for this candidate mutation. A second cLDLA which considered the *HMX1* genotype in our mapping population as a fixed effect was conducted and is shown in Figure 2b.

The cLDLA with *HMX1* as fixed effect clearly explains complete phenotypic variance on OAR6, reduces the LRT-peak on OAR4 and yield only some suggestive mapping signals. A mutation at *GATA6* exon 7 (OAR23) was discussed in Jawasreh and Al-Omari (2020) as a candidate in an epistatic interaction with *HMX1*. We also genotyped this candidate mutation in our mapping population and could not confirm any effects in our material. Also, we checked for further variations in the previously described target genes of *HMX1*, that are *SLC6A9*, *EPHA6*, *SGCG*, *TSHZ2*, *PTPRO* and *SEMA3F* (Boulling *et al.,* 2013). No noticeable difference in allelic frequencies of those genes per ear length group were found. Also, none of these loci is close to the suggestive significant threshold for association to ear length. In order to allow an insight for the size of the effect of *HMX1* and how it manifests across ear size categories, allele frequencies are presented in Table 1.

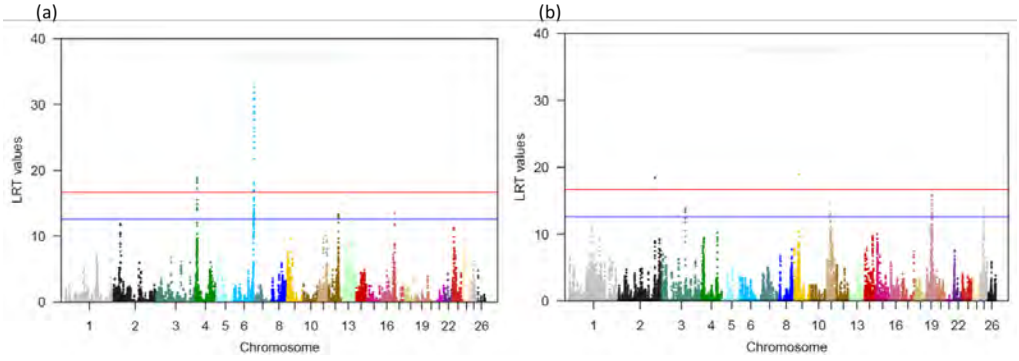


Figure 2. Likelihood ratio test (LRT) values for each genotyped position of a cLDLA for outer ear length without any genotype fixed effects (a) and with *HMX1* genotype data as a fixed effect (b). Genome-wide significance (red) and suggestive (chromosome-wide) significance (blue) thresholds are indicated as straight lines.

Table 1. Allele identities and frequencies of *HMX1* in the studied animals respective to the ear size categories. Frequencies for the allele of the reference sequence (d) and the allele with the 76 bp duplication (D) is shown.

	Anotia		Microtia		Regular ear length	
Allele identity	**d**	**D**	**d**	**D**	**d**	**D**
Allele frequency	0.038	0.96	0.53	0.47	0.97	0.034

Discussion

Anotia and *microtia* have been described in numerous Asian and European sheep breeds, and various previous mapping studies have yielded somewhat contradictory results. The study conducted here confirmed only *HMX1* as a plausible candidate cause and could not confirm any interaction with *GATA6*. We could also show that *anotia* and *microtia* in sheep from Central Asia to the Alps are caused by the same, presumably very old mutation and that both trait expressions only reflect different genotypes at the same locus. A large proportion of the sheep examined here were also phenotyped for horn status. This material will be supplemented by further phenotyped and genotyped sheep and used for further horn status mapping with focus on breeds with variable horn status and southern European origin which can complement the results discussed in Luhken *et al.* (2016).

References

Allais-Bonnet A., Grohs C., Medugorac I., Krebs S., Djari A. *et al.* (2013) PLoS One 8(5)**:** e63512. https://doi.org/10.1371/journal.pone.0063512

Arkell T.R. (1912) Journal of Heredity os-8(1)**:** 561-568. https://doi.org/10.1093/jhered/os-8.1.561

Boulling A., Wicht L. and Schorderet D.F. (2013) Mol Vis 19(1779-1794.

Browning B.L., Zhou Y. and Browning S.R. (2018) Am J Hum Genet 103(3)**:** 338-348. https://doi.org/10.1016/j.ajhg.2018.07.015

Browning S.R., and Browning B.L. (2007) Am J Hum Genet 81(5)**:** 1084-1097. https://doi.org/10.1086/521987

Gao L., Xu S.S., Yang J.Q., Shen M. and Li M.H. (2018) Anim Genet 49(4)**:** 345-348. https://doi.org/10.1111/age.12670

Gehrke L.J., Capitan A., Scheper C., Konig S., Upadhyay M. *et al.* (2020) Genet Sel Evol 52(1)**:** 6. https://doi.org/10.1186/s12711-020-0525-z

Gilmour Ar G.B., Cullis Br, Thompson R (2009), pp.

Granato I.S.C., Galli G., De Oliveira Couto E.G. E Souza M.B., Mendonça L.F. *et al.* (2018) Molecular Breeding 38(8). https://doi.org/10.1007/s11032-018-0844-8

He S., Zhang Z., Sun Y., Ren T., Li W. *et al.* (2020) Anim Genet 51(1)**:** 132-136. https://doi.org/10.1111/age.12876

Jawasreh K., Boettcher P.J. and Stella A. (2016) Anim Genet 47(4)**:** 504-506. https://doi.org/10.1111/age.12431

Jawasreh K.I.Z., and Al-Omari H.D. (2020) Genes (Basel) 11(6). https://doi.org/10.3390/genes11060597

Luhken G., Krebs S., Rothammer S., Kupper J., Mioc B. *et al.* (2016) Genet Sel Evol 48(1)**:** 78. https://doi.org/10.1186/s12711-016-0256-3

Lush J.L. (1930) Journal of Heredity 21(3)**:** 107-112. https://doi.org/10.1093/oxfordjournals.jhered.a103296

Mastrangelo S., Sottile G., Sutera A.M., Di Gerlando R., Tolone M. *et al.* (2018) Anim Genet 49(6)**:** 636-640. https://doi.org/10.1111/age.12719

Meuwissen T.H., and Goddard M.E. (2001) Genet Sel Evol 33(6)**:** 605-634. https://doi.org/10.1186/1297-9686-33-6-605

Qgis.Org (2021) QGIS Geographic Information System, pp. QGIS Association.

Ritzman E.G. (1920) Journal of Heredity 11(5)**:** 238-240. https://doi.org/10.1093/oxfordjournals.jhered.a102008

Yang J., Benyamin B., Mcevoy B.P., Gordon S., Henders A.K. *et al.* (2010) Nat Genet 42(7)**:** 565-569. https://doi.org/10.1038/ng.608

537. Large-scale 'omics fine-mapping identifies expression quantitative trait loci significantly affecting cattle phenotypes

R. Xiang[1,2*], M.E. Goddard[1,2], C.P. Prowse-Wilkins[1,2], C.M. Reich[2], B.A. Mason[2], J.B. Garner[2], L.C. Marett[2], I.M. MacLeod[2] and A.J. Chamberlain[3]

[1]Faculty of Veterinary & Agricultural Science, University of Melbourne, Parkville, VIC, 3010, Australia; [2]Agriculture Victoria, AgriBio, Centre for AgriBiosciences, Bundoora, VIC, 3083, Australia; ruidong.xiang@unimelb.edu.au

Abstract

Variants that change the expression of genes, i.e. expression quantitative trait loci or eQTL, provide an alternative to identify causal mutations. However, previous studies found little overlap between eQTL and conventional QTL. Here, we mapped and compared eQTL from blood and milk cells in ~400 dairy cattle with complex trait genome-wide association studies (GWAS) and Bayesian mixture models (BayesR) in ~103k cows. Increasing the sample size of eQTL mapping significantly increases the overlap between eQTL and QTL. As an example, a cis eQTL at Chr13:36387662(rs20846618) or the gene mohawk homeobox (*MKX*, p_{eQTL}=4.7e-08) significantly affects rump angle ('pin set': p_{GWAS}=4.5e-10): a trait related to fertility and longevity. This eQTL is the top variant in BayesR mapping for rump angle. *MKX* is a transcription factor gene and is associated with tendon development and tail morphology in mice. Therefore, large-scale multi-omics fine-mapping can prioritise biologically sensible causal mutations for cattle complex traits.

Introduction

Most cattle traits of economic importance are complex where many DNA variants affect the trait, and each contributes a very small effect. This makes the identification of causal variants very difficult. Using sequence variants in genome-wide association studies (GWAS) increases the chance of identifying causal variants (Daetwyler *et al.* 2014). However, due to the very small effective population size (N_e) in livestock species, the linkage disequilibrium (LD) between variants is very high. This results in many significant variants being found in sequence GWAS yet little is known which ones are the causal variants. Therefore, strategies to refine GWAS results are needed. One strategy is for GWAS to have a very large sample size, potentially with a multi-breed population (Bouwman *et al.* 2018; van den Berg *et al.* 2020). However, QTL do not always overlap across breeds and it is not always easy to obtain large samples for small breeds, such as Jersey or Red cattle. An alternative approach is to use functional annotation (Clark *et al.* 2020), including how sequence variants change the expression of genes (Xiang *et al.* 2019; Liu *et al.* 2021; Xiang *et al.* 2021b). However, some primary efforts showed that there is very little overlap between eQTL and conventional QTL, such as QTL for protein% and fat% (Van Den Berg *et al.* 2019; van den Berg *et al.* 2020). However, these studies used a relatively small sample size for mapping of the eQTL (N~130).

In the current study, we significantly increased the sample size for the mapping of the eQTL (N~400). We show that with a large sample size in eQTL mapping, the current study detects many overlaps between eQTL and conventional QTL. We highlight examples and provide important updates on how eQTL contributes to QTL via biological mechanisms.

Materials & methods

RNA-seq data. The RNA-seq data was generated by Agriculture Victoria Research (AVR) in Victoria, Australia. Blood and milk samples were taken from 400 Holstein cows and 29 Jersey cows (milk sample only). The processing of milk samples followed what was described in (Chamberlain *et al.* 2018; Xiang *et al.* 2018). RNA sequence data of 356 Holstein and 26 Jersey with >50 million reads for milk cells or

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_537, © R. Xiang *et al.* 2022

>25 million reads for white blood cells and had concordant alignment rate >80% were used. After quality control, raw reads of these data were aligned to the ARS-UCD1.2 reference bovine genome and gene counts were extracted.

Sequence genotypes. The sequence variants used were based on what was described in (Xiang *et al.* 2021a). Briefly, we used more than 16 million sequence variants imputed using Run7 of the 1000 Bull Genome with imputation accuracy R^2>0.4 and the minor allele frequency (MAF)>0.005 in both bulls and cows.

Conventional phenotypes. Data was collected by farmers and processed by DataGene Australia (http://www.datagene.com.au/) for the official May 2020 release of National breeding values. Phenotype data had a total of 103,350 cows, including Holstein (87,003), Jersey (13,353♀), cross-breed (5,037♀) and Australian Red dairy (3,379♀) breeds. We examined traits of protein%, fat% and pin set [detailed in (Xiang *et al.* 2021a)].

eQTL and QTL mapping. For expression of each gene after normalisation, a linear mixed model was fitted: $y = X\beta + Zg + Wv + e$ (1); Where y is an n × 1 vector gene expression or conventional phenotype (n is the number of individuals with RNA-seq data or conventional phenotype), β is a vector of fixed effects like breeds, different experiments; g is an n × 1 vector of the total genetic effects of the individuals with $g \sim N(0, G\sigma^2)$ where G is the genomic relationship matrix (GRM) built by all the variants; W is the design matrix of variant genotypes (0, 1, 2) and v is the variant additive effect; e is the error term. cis QTL is defined as variants within ±1 Mb of the gene. Significant eQTL or QTL were called if GWAS P-value <5e-8. In the milk cell eQTL mapping, the 1st 2 PCs were extracted from the expression count matrix and fitted as fixed effects along with other fixed effects. This aimed at adjusting for unrecorded fixed effects which may distort the eQTL mapping.

BayesR. We fitted the following model in BayesR (Erbe *et al.* 2012): $y = Wv + X\beta + e$ (2); where y was a vector of conventional phenotype, W was the design matrix of genotypes centred and standardised to have a unit variance; v was the vector of variant effects; X was the design matrix allocating phenotypes to fixed effects; β was the vector of fixed effects of breeds. As a result of the Monte Carlo Markov Chain (50,000 iterations with 25,000 burn-in), the effect v for each variant was jointly estimated with other variants. The prior of variant effects v followed a mixture distribution of four normal distributions including a null distribution, zero-effect $[N(0,0.0\sigma^2_g)]$, and three others: small-effect $[N(0,0.0001\sigma^2_g)]$, medium-effect $[N(0,0.001\sigma^2_g)]$ and large-effect $[N(0,0.01\sigma^2_g)]$, where σ^2_g was the additive genetic variance for the trait. to reduce the computational burden for bayesR, we used up to 2 million sequence variants which are from HD SNPs plus sequence variants under more than 1 ChIP-seq peaks (Kern *et al.* 2021; Prowse-Wilkins *et al.* 2021). The posterior inclusion probability (PIP, ranged from 0-1) of each variant in the model was obtained and was used to rank variants.

Results & discussion

We detected significant overlaps between eQTL (sample size ~400) and QTL (sample size ~104,000 of three breeds) for analysed traits (Table 1). For protein% and fat%, ~16.11% (1,828 variants) of QTL were eQTL from blood and ~8.05% of QTL were eQTL from milk cells. This represents a substantial increase in the concentration of eQTL in QTL, compared to the previous study by Van den Berg *et al.* (2020), where ~4.6% (1,496 variants) of milk percentage QTL were eQTL from blood and ~0.04% (13 variants) of QTL were eQTL from milk cells. In van den Berg *et al.*, the sample size of eQTL mapping was ~130 and GWAS used 94,321 cattle from 8 breeds. Thus, the sample size for the QTL mapping between the two studies is similar, but the sample size for eQTL mapping is about 2-3 times larger in the current study than in the previous

Table 1. Overlap between eQTL and conventional QTL. The total number (#) of variants, significant eQTL from blood and milk cells, and significant QTL for protein %, fat %, and pin set are given. The *P*-value ($p_{overlap}$) of the overlap was based on the hypergeometric test.

Tissue	# total variants	# sig eQTL	# sig QTL	# sig eQTL&QTL	% QTL as eQTL	$p_{overlap}$
Blood	11,669,532	486,311	16,855 (Prot%)	1,754	10.41%	1.2E-258
			11,345 (Fat%)	1,828	16.11%	0
			174 (Pin set)	34	19.54%	6.0E-14
Milk_cell	11,157,718	218,461	16,687 (Prot%)	825	4.94%	2.0E-121
			11,164 (Fat%)	899	8.05%	6.2E-269
			174 (Pin set)	25	14.37%	1.3E-14

study. Therefore, we demonstrate that increasing sample size is a key factor to map eQTL and to identify overlap with QTL.

There is a particular increase in the detection of eQTL in milk cells and its overlap with QTL in the current study compared to Van den Berg *et al.* (899 eQTL representing 8.05% of QTL VS 13 eQTL representing 0.04% of QTL). Along with a large sample size, adjusting PCs from the expression count matrix of milk cells in the linear model may also contribute to the increased power of the current eQTL mapping. When not correcting for PCs, there were a lot fewer significant eQTLs were called.

We also found ~20% of QTL for pin set being eQTL. Pin set is a type trait also known as rump angle. In Australian and US linear assessment systems, pin set is related to calving difficulty, fertility and longevity. To our knowledge, no QTL has been reported for this trait. We report a candidate causal mutation at Chr13:36387662 (rs208466183, ARS-UCD1.2) for pin set involving cis eQTL effects (Figure 1). This mutation has a p_{GWAS}=4.5e-10 and BayesR PIP=0.74 which is the highest PIP for BayesR mapping. Mutation at Chr13:36387662 affected the expression of mohawk homeobox (*MKX*) (p_{eQTL}=4.7e-08). In addition, after adjustment of fixed effects, the expression of *MKX* in milk cells is significantly associated with pin set (*P*=4.1e-11). *MKX* is a transcription factor and is associated with tendon development and tail morphology in mice (Ito *et al.* 2010). We, therefore, suggest the strong candidacy of Chr13:36387662 for pin set via cis-regulatory effects on *MKX*.

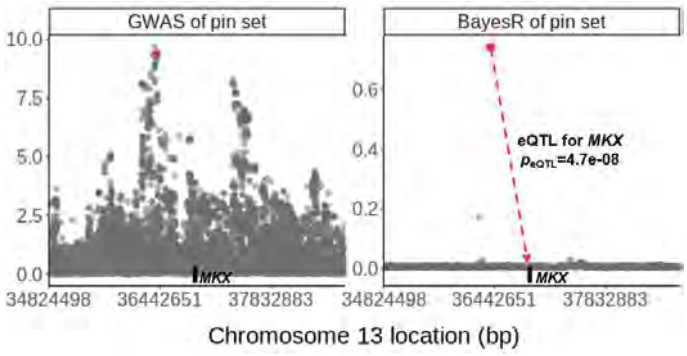


Figure 1. Candidate causal mutation Chr13:36387662 for pin set as a cis eQTL for mohawk homeobox (*MKX*). The y-axis for the left and right panel indicates -log10(p) of GWAS and posterior inclusion probability (PIP) of BayesR, respectively. Red dots indicate the mutation Chr13:36387662 and the dashed line indicates that Chr13:36387662 is a cis eQTL for *MKX*.

Conclusions

We found that increasing the sample size of this eQTL study significantly increased the identification of eQTL and its overlap with QTL. We promote the causal candidacy of Chr13:36387662 (rs208466183) for pin set via cis-regulatory effects on *MKX* expression, illuminating biological mechanisms with which cattle genes shape complex traits.

References

Bouwman A.C., Daetwyler H.D., Chamberlain A.J., Ponce C.H., Sargolzaei M., *et al* (2018) Nat. Genet.50, 362. https://doi.org/10.1038/s41588-018-0056-5

Chamberlain A., Hayes B., Xiang R., Vander Jagt C., Reich C., *et al* (2018) Proc. of the 11th WCGALP, Auckland, New Zealand: Volume Molecular Genetics, p. 254.

Clark E.L., Archibald A.L., Daetwyler H.D., Groenen M.A., Harrison P.W., *et al* (2020) Genome Biol. 21, 1-9. https://doi.org/10.1186/s13059-020-02197-8

Daetwyler H.D., Capitan A., Pausch H., Stothard P., Van Binsbergen R., *et al* (2014) Nat. Genet. 46, 858. https://doi.org/10.1038/ng.3034

Erbe M., Hayes B., Matukumalli L., Goswami S., Bowman P., *et al* (2012) J. Dairy Sci. 95, 4114-29. https://doi.org/10.3168/jds.2011-5019

Ito Y., Toriuchi N., Yoshitaka T., Ueno-Kudoh H., *et al* (2010) PNAS 107, 10538-42. https://doi.org/10.1073/pnas.1000525107

Kern C., Wang Y., Xu X., Pan Z., Halstead M., *et al* (2021) Nat. Commun. 12, 1821. https://doi.org/10.1038/s41467-021-22100-8

Liu S., Gao Y., Canela-Xandri O., Wang S., Yu Y., *et al* (2021) bioRxiv, 2020.12.01.406280. https://doi.org/10.1101/2020.12.01.406280

Prowse-Wilkins C.P., Wang J., Xiang R., Garner J.B., Goddard M.E., *et al* (2021) Front. genet 12. https://doi.org/10.3389/fgene.2021.664379

Van Den Berg I., Hayes B., Chamberlain A. & Goddard M. (2019) BMC Genomics 20, 1-18.

van den Berg I., Xiang R., Jenko J., Pausch H., Boussaha M., *et al* (2020) Genet. Sel. Evol. 52, 1-16. https://doi.org/10.1186/s12711-020-00556-4

Xiang R., Berg I.v.d., MacLeod I.M., Hayes B.J., Prowse-Wilkins C.P., *et al* (2019) PNAS 116, 19398-408. https://doi.org/10.1073/pnas.1904159116

Xiang R., Breen E.J., Bolormaa S., Jagt C.J.V., Chamberlain A.J., *et al* (2021a) Commun. Biol. 4, 1-10. https://doi.org/10.1038/s42003-021-02874-9

Xiang R., Hayes B.J., Vander Jagt C.J., MacLeod I.M., Khansefid M., *et al* (2018) BMC Genom. 19, 521. https://doi.org/10.1186/s12864-018-4902-8

Xiang R., MacLeod I.M., Daetwyler H.D., de Jong G., O'Connor E., *et al* (2021b) Nat. Commun. 12, 1-13. https://doi.org/10.1038/s41467-021-21001-0

Part 32

Functional annotation of genomes: structural variation and variants

538. A functional copy number gain at indicine cattle footprint modulates actin-myosin and immune response genes in muscle

V.H. da Silva[1*], J. Afonso[2], N.S. Morosini[1], F.C. da Veiga[1], A.S.M. Cesar[1], B. Silva-Vignato[1], G.B. Mourão[1], A.Z. Neto[3], L.C.A. Regitano[2] and L.L. Coutinho[1]

[1]University of São Paulo (ESALQ-USP), Av. Pádua Dias 11, CEP 13418-900, Piracicaba, SP, Brazil; [2]Embrapa Pecuária Sudeste, Rodovia Washington Luiz Km 234, CEP 13560-970, São Carlos, SP, Brazil; [3]Embrapa Informática Agropecuária, Av. André Tosello 209, CEP 13083-886, Campinas, SP, Brazil; viniciushs@usp.br

Abstract

There are remarkable phenotypic and physiological differences between taurine (*Bos taurus*) and indicine (*Bos indicus*) breeds such as their distinguishable meat quality and immune response. Here, we used muscle RNA-seq and genotyping data from 187 Nelore steers to explore the effect of copy number variations, which are located at taurus-indicus dissimilar genomic regions and ATAC-seq peaks, on the expression of local and distal genes. Therefore, we inferred the co-expression network for the target transcripts and performed a gene enrichment with the correlated transcripts. From seven polymorphic copy number variations identified at these regions, one gain at chromosome 15 was associated with a trans-upregulation of the transcription factor *PITX1* (Bonferroni adjusted *P*-value<0.05). This gain can be valuable for animal breeding as the *PITX1* network genes are enriched for meat quality related proteins such as actin and myosin as well as immune system proteins such as cytokine and toll-like receptors.

Introduction

The domestication of aurochs (*Bos primigenius*) gave rise to different taurine and indicine (*Bos indicus*) breeds. Beef production can be based on taurine, indicine or mixed breeds. However, the choice of the best breed for a specific production system depends on different factors such as climate and market positioning. Indicine are more resistant to tropical parasites whereas taurine breeds are adapted to temperate environments (Porto-Neto *et al.* 2014). On the other hand, taurine and indicine breeds have distinguishable meat quality properties such as fiber size and marbling (Rodrigues *et al.* 2017). Therefore, the differences between their genomes are expected to uncover candidate regions for the immune system and meat quality regulation. Genomic regions underlying taurine-indicine subspeciation can be uncovered by the sequence blocks that are dissimilar between their genomes, which can be detected by a synteny analysis. Although the term synteny classically refers to gene order, more recent definitions extended it to refer to any similar genome block between sequences, considering that important regulators may be located at intergenic intervals (i.e. microsynteny, Engström *et al.* 2007). Sequences showing synteny breaks among different species are usually enriched with copy number variations (CNVs), which are crucial for speciation events (Pös *et al.* 2021). Interestingly, frequent CNVs are common at regulatory regions with open chromatin and enriched for CpG islands. It has been shown that the effect of CNVs on the cattle muscle gene regulation is mostly distal and involves the interplay of different transcription factors (Geistlinger *et al.* 2018). Here, we evaluated the effect of CNVs on muscle gene expression (i.e. CNV-eQTLs) in our Nelore cattle population, focusing on regions potentially relevant to taurine-indicine subspeciation that are functional in muscle. Thus, we analysed exclusively those concomitantly located at open chromatin and microsynteny breaks (i.e. functional CNV-eQTLs located at subspeciation footprints). Lastly, we constructed a co-expression network for the target genes to account for the indirect effects of targeted transcription factors.

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_538, © V.H. da Silva *et al.* 2022

Materials & methods

Detection of chromatin accessibility. We performed ATAC-seq in four meat samples to obtain the FASTA files (Morosini *et al.* unpublished data). Therefore, we produced the ATAC-seq pileup files from FASTA files using the nfcore/atacseq pipeline implemented in Nextflow (https://nf-co.re/atacseq). A consensus peak mapping was obtained from the four samples by considering only regions with counts higher than zero in all. In these regions, the average among samples was calculated and values higher than 0.5 were selected as genomic regions in an ATAC-seq peak.

Screening of synteny breaks between taurus-indicus genomes. The reference genomes ARS-UCD1.2 and UOA_Brahman were used to represent taurine and indicine sequences, respectively. We searched for sequences that chained into collinear blocks of synteny with the *FindSynteny* and mapped the results onto ARS-UCD1.2 genome with *AlignSynteny* function, both implemented in the R/Bioconductor package DECIPHER (Wright 2016).

Quantification of RNA abundances. The FASTQ files of 187 Nelore steers generated by the HiSeq 2500 sequencer were mapped and quantified by the Salmon software (Patro *et al*. 2017) with full decoy. Transcripts with abundance equal to zero in more than 5% of the animals were filtered out.

Genotyping and CNV detection. The *log2 R ratio* and *B allele frequency* values from SNP probes from 777 Nelore Steers genotyped with the Illumina 700k HD, mapped to ARS-UCD1.2 reference genome, were used to detect CNVs by a hidden Markov model based approach that is implemented in the PennCNV (Wang *et al.* 2007).

Identification of CNV effect on gene expression. The effect of CNVs on gene expression was analysed by an edgeR's ANOVA-like test implemented in the CNVRanger package (da Silva *et al.* 2020). The raw *P*-values were corrected for the number of transcripts to account for multiple tests using the Bonferroni correction. The effect on a target transcript (i.e. CNV-eQTLs) was classified as a cis-effect when the distance between CNV and gene was less than 1 Mb and trans-effect otherwise.

Construction of co-expression networks. We constructed co-expression networks of all filtered transcripts using the results from the PCIT algorithm (Reverter *et al.* 2008). The transcripts belonging to the co-expression network of the transcripts targeted by CNVs was subset and used for the target-specific gene enrichment.

Enrichment of genes within co-expression networks. All transcripts belonging to the network of affected transcripts were enriched for gene ontology terms and KEGG pathways with the *enrichKEGG* and *enrichGO* functions from the clusterProfiler package (Yu *et al*. 2012) using a q-value<0.05 as significant.

Results

We detected 7,756 consensus ATAC-seq peaks and 5,081 intervals with break of sequence microsynteny between indicus-taurus genomes with mean sizes of 2.3 kb and 3.2 kb, respectively. The microsynteny break regions encompassed 16.53 Mb of the autosomes (0.66%). We identified 5,956 CNVs with a mean size of 54.36 kb calls that are distributed among 187 Nelore steers with available RNA quantification. The RNA abundance of 14,471 transcripts remained after filtering.

A total of seven polymorphic CNV regions (>1% in frequency) overlapped with both ATAC-seq peaks and regions of microsynteny break. Only the CNV region at chr15-45962744:46353257 had at least ten samples with CNVs (i.e. gain samples) to attend to the minimum default for testing (Figure 1A). This region has two

possible loci, which are reflected by sample frequency (Figure 1A, 'gain samples' track). When this CNV region was refined to comprise only the downstream locus ATAC-seq (i.e. chr15-46210000-46360000, red rectangle), a copy gain was significantly associated with the abundance of the *PITX1-201* transcript (Bonferroni adjusted *P*-value<0.05, Figure 1B).

We found 96 transcripts in the *PITX1-201* co-expression network, which were enriched for four gene ontology terms and two KEGG pathways. The gene groups enriched were associated with muscular structure and protein binding related to myosin and actin as well as with immune system dynamics through cytokine interaction and toll-like signaling (Table 1).

Discussion

It is crucial to explore the functional effects of regions lacking synteny to understand how certain molecular mechanisms differ among species. We found evidence that a copy number gain, located at a taurus-indicus microsynteny break block with open chromatin in muscle, upregulates the expression of the transcription factor *PITX1*. The trans-effect of this gain at chromosome 15 on *PITX1* (chromosome 5) expression might be caused by an enhancer reflected by the ATAC-seq peak. This gene determines the morphology of muscle and tendons (DeLaurier *et al.* 2006) and is also associated with immunity-longevity tradeoff (Otarigho and Aballay 2021). Interestingly, taurine and indicine breeds have known differences regarding meat quality and immune response. Actin and myosin related pathways, which are known to affect meat quality (Asghar and Yeates 1978), were enriched in the *PITX1-201* co-expression network. Moreover, immune response pathways related to cytokine and toll-like receptors were also enriched. This copy gain at chromosome

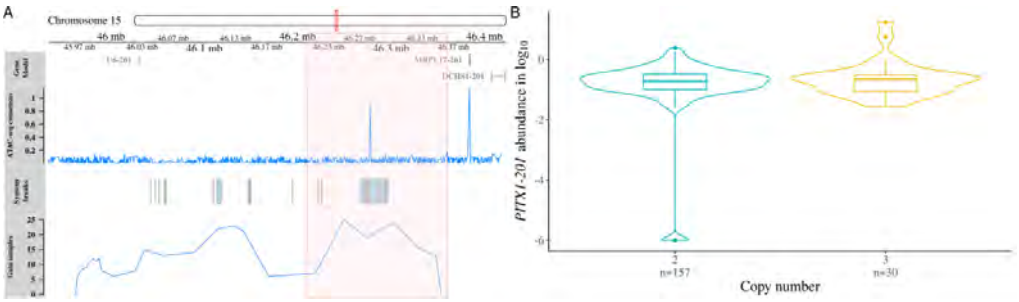


Figure 1. The trans-effect of a copy gain at the expression of the transcription factor *PITX1*. (A) Refined CNV region associated with *PITX1-201* transcript abundance. (B) *PITX1-201* abundance distribution in samples with two and three copies (*P*-value=1.23492e-06).

Table 1. Gene enrichment of the transcripts within the *PITX1-201* co-expression network.

Id	Type	Pathway name	*q*-value	Gene ratio[1]
GO:0003779	molecular function	actin binding	0.027	5/27
GO:0009888	biological process	tissue development	0.028	9/25
GO:0015629	cellular component	actin cytoskeleton	0.021	5/30
GO:0032982	cellular component	myosin filament	0.031	2/30
bta04060	KEGG pathway	cytokine-cytokine receptor interaction	0.022	6/28
bta04620	KEGG pathway	Toll-like receptor signaling pathway	0.022	4/28

[1] Total number of genes in the term/pathway by the ones that are present in the enrichment list.

15 may be relevant to animal breeding as it could contribute to the phenotypic differences among cattle subspecies. Future analysis may incorporate more animals and breeds to build a better understanding on the genomic variability within each subspecies and how it relates to this regulatory region. The distal effect of CNVs on the expression of transcription factors may help explain the regulatory landscape of genes affecting phenotypes of interest in cattle such as meat quality and immune response.

References

Asghar A., and Yeates N.T. (1978) Crit. Rev. Food Sci. Nutr. 10(2):115-45. https://doi.org/10.1080/10408397809527247

da Silva V., Ramos M., Groenen M., Crooijmans R., Johansson A. *et al.* (2020) Bioinformatics 36(3):972-973. https://doi.org/10.1093/bioinformatics/btz632

DeLaurier A., Schweitzer R., and Logan M. (2006) Dev. Biol. 299(1):22-34. https://doi.org/10.1016/j.ydbio.2006.06.055

Engström P.G, Ho Sui S.J., Drivenes Ø., Becker T.S., and Lenhard B. (2007) Genome Res. 17(12): 1898–1908. https://doi.org/10.1101/gr.6669607

Geistlinger L., da Silva V.H., Cesar A.S.M., Tizioto P.C., Waldron L. *et al.* (2018) Sci. Rep. 8:1399. https://doi.org/10.1038/s41598-018-19782-4

Otarigho B., and Aballay A. (2021) Cell Rep. 35(8):109187. https://doi.org/10.1016/j.celrep.2021.109187

Patro R., Duggal G., Love M.I., Irizarry R.A., and Kingsford C. (2017) Nat. Methods 14:417-419. https://doi.org/10.1038/nmeth.4197

Porto-Neto L.R., Reverter A., Prayaga K.C., Chan E.K.F., Johnston D.J. *et al.* (2014) PLOS ONE 9(11): e113284. https://doi.org/10.1371/journal.pone.0113284

Pös O., Radvanszky J., Buglyó G., Pös Z., Rusnakova D. *et al.* (2021) Biomed. J. 44(5)548-559. https://doi.org/10.1016/j.bj.2021.02.003

Reverter A., and Chan E.K.F. (2008) Bioinformatics 24(21):2491-7. https://doi.org/10.1093/bioinformatics/btn482

Rodrigues R.T.S., Chizzotti M. L., Vital C.E., Baracat-Pereira M.C., Barros E. *et al.* (2017) PLOS ONE 12(1): e0170294. https://doi.org/10.1371/journal.pone.0170294

Wang K., Li M., Hadley D., Liu R., Glessner J. *et al.* (2007) Genome Res. 17:1665-1674. https://doi.org/10.1101/gr.6861907

Wright E.S. (2016) R J. 8(1):352-359. https://doi.org/10.32614/RJ-2016-025

Yu G., Wang L., Han Y., and He Q. (2012) OMICS 16(5):284-287. https://doi.org/10.1089/omi.2011.0118

539. Functional SNPs and INDELs within regulatory elements associated with mastitis in Holstein cow using -OMICs technologies

V. Asselstine[1*], J.F. Medrano[2], P. Stothard[3], F. Miglior[1], N.A. Karrow[1], C.F. Baes[1,4], F.S. Schenkel[1] and A. Cánovas[1]

[1]Centre for Genetic Improvement of Livestock, Department of Animal Biosciences, University of Guelph, Guelph, ON, N1G 2W1, Canada; [2]Department of Animal Science, University of California-Davis, Davis, CA 95616, USA; [3]Department of Agricultural, Food and Nutritional Science/Livestock Gentec, University of Alberta, Edmonton, Alberta, T6G 2R3, Canada; [4]Institute of Genetics, Vetsuisse Faculty, University of Bern, 3012 Bern, Switzerland; vasselst@uoguelph.ca

Abstract

Integrating data from high-throughput transcriptomic and genomic technologies offers a new opportunity to characterize the biological basis underlying complex traits, such as mastitis. Previous research using RNA-Sequencing technology from milk somatic cell samples from Holstein dairy cows identified 7 differentially expressed (DE; FDR<0.05, |FC|>2) functional candidate genes associated with mastitis resistance. Whole genome sequencing (WGS) was then used to deeply study the 7 DE functional candidate genes to identify single nucleotide polymorphisms (SNPs) and insertions / deletions (INDELs) that were fixed and uniquely present within an independent Holstein population. In total, WGS analysis identified 14 SNPs and 7 INDELs located within the 7 DE functional candidate genes associated with mastitis resistance. The identified functional variants could impact the host's response to the mastitis causing agent, aiding in breeding healthier dairy cows that are better able to cope with or resist mastitis.

Introduction

Mastitis is an inflammatory disease of the udder caused by different species of bacteria and it is extremely prevalent in the dairy industry. As mastitis is a very complex trait, it has yet to be effectively controlled at a low incidence rate. Whole genome sequencing (WGS) is a high-throughput technology to identify key differences in the genome, such as single nucleotide polymorphisms (SNP) and insertions / deletions (INDELs) that may impact the host's response to mastitis. Previous research using RNA- Sequencing technology (RNA-Seq) identified 7 functional candidate genes differentially expressed (DE) between healthy and mastitic samples (Asselstine *et al.*, 2019). Therefore, the objective of this study was to use WGS to identify potential SNPs and INDELs that were fixed and unique between a group of healthy and a group of mastitic Holstein dairy cows. Key variants which may create non-functional proteins could then be implemented into breeding strategies to select for and breed dairy cattle that are more resistant to mastitis disease, which would greatly benefit the dairy industry.

Materials & methods

Animal material and sample collection. Health records for a random subset of animals housed at the Ontario Dairy Research Center, Elora, ON were searched manually to identify healthy and mastitic animals. Animals were classified as 'mastitic' if they had at least 1 case of either clinical mastitis, diagnosed visually at the time of milking or subclinical mastitis, diagnosed via a high somatic cell count on a dairy herd improvement test, which was cultured to confirm the presence of an intramammary infection (n=7). In total, 7 animals had at least 1 case of mastitis in their lives. Alternatively, animals were classified as healthy if at the time of analysis, they had no reported cases of either clinical or subclinical mastitis (n=7). In total, 14 animals (7 healthy and 7 mastitic) were used to perform WGS analysis. Hair samples (approximately 80-100 follicles) from the tail end were individually collected from each of the 14 animals, sealed into a pre-labelled Zoetis hair sample collector (https://www.ordergeneticstestkits.com/storefrontB2CWEB/) and stored at room temperature in a sealed bag until WGS analysis could be completed.

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_539, © V. Asselstine *et al.* 2022

Whole genome sequencing analysis. Whole genome sequencing analysis was completed at Neogen Canada (Edmonton, AB) using the Illumina HiSeq X (https://www.illumina.com/systems/sequencing-platforms/hiseq-x.html) with an average of 23.4±5.8 × coverage. All programs and parameters for SNP and INDEL detection were selected in accordance with the 1000 Bull Genome Project guidelines (http://www.1000bullgenomes.com). Briefly, single nucleotide polymorphisms (SNP), insertion / deletion (INDEL) detection was performed using the Read Mapping and Variant Calling workflow. The Raw fastq files were downloaded from the Nanuq interface on Génome Québec's website (https://cesgq.com/en-nanuq). Using FastQC, read quality control was performed on the high-throughput sequencing data. Trimmomatic, was then used to trim and crop the Illumina (FASTQ) data (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/; Bolger *et al.*, 2014). The trimmed reads were then mapped to the bovine reference genome ARS – UCD1.2 assembly using Burrow – Wheeler Aligner (BWA; https://github.com/lh3/bwa). Lastly, SNPs and INDELs were identified using the Genome Analysis Toolkit (GATK) Haplotype Caller (https://gatk.broadinstitute.org/hc/en-us). For each SNP and INDEL category, one VCF file was generated with the data for all animals included (7 mastitic and 7 healthy Holstein dairy cows).

Variant identification within DE candidate genes and variant effect predictor. After the VCF files for SNP and INDEL from the WGS analysis were obtained, CLC Genomics Workbench software (Aarhus, Denmark) was used to upload the VCF files for SNP and INDEL independently. Track tools was then used to inspect the 7 key functional candidate regions associated with mastitis resistance (Table 1) to identify SNPs and INDELs. Once these variants were identified, Variant Effect Prediction (VEP; McLaren *et al.*, 2016) from Ensembl was used to determine the potential causal effect of the variant within the functional candidate genes (Cánovas *et al.*, 2012; Cardoso *et al.*, 2018).

Results

SNPs and INDELs within functional candidate genes associated with mastitis resistance. The genomic coordinates and the related information of the 7 DE functional candidate genes used in this analysis can be found in Table 2. Looking deeply into the WGS results, in the healthy group, there are 6 unique and fixed variants (4 insertions, 2 deletions) within the functional candidate genes *FCER1G*, *GLYCAM1*, *CD74*, *NFKBIA*, and *BoLA – DRA*, which were previously identified using RNA-Seq (Asselstine *et al.*, 2019). Only two variants, one deletion in *FCER1G* and one deletion in *CD74* (both located in introns) had predicted effects of MODIFIER, meaning there is little to no evidence that these variants will have an impact on the functionality of the protein formed. However, as the remaining 4 variants (insertions) were also in intronic regions, although they did not have a predicted effect it is hypothesized they would also have a limited impact on the functionality of the protein formed. Within the 7 animals in the mastitic group,

Table 1. Candidate genes differentially expressed (FDR<0.05, |FC|>2) between 12 healthy and mastitic milk somatic cells samples from Holstein dairy cows using RNA-Sequencing.

Genes (DE)[1]	*P*-value	Fold change	FDR	Position
B2M	1.92E-04	2.62	1.64E-02	10:103116503-103123174
CD74	5.72E-04	4.11	3.70 E-02	7:61747923-61755685
GLYCAM1	1.87E-08	-55.59	1.15E-04	5:25478787-25482857
BoLA-DRA	1.51E-04	6.40	1.42E-02	23:25837353-25843155
FCER1G	1.80E-05	3.43	3.57E-03	3:8286937-8290169
NFKBIA	1.42E-04	3.38	1.36E-02	21:45639133-45643131
SDS	1.40E-04	6.27	1.36E-02	17:61030240-61040569

[1] Genes were identified by Asselstine *et al.*, 2019.

Table 2. Variants (SNPs and INDELs) found within the 7 differentially expressed (FDR<0.05, |FC|>2) functional candidate genes fixed and uniquely present in healthy or mastitic group of 14 Holstein dairy cows using whole genome sequencing.

Health group	Gene	Position	Variant type	Location
Mastitis	*GLYCAM1*	5:25479440	SNP	Intron
	CD74	7:61749473	SNP	Intron
	B2M	10:103116650	SNP	Intron
		10:103116656	SNP	Intron
		10:103116719	SNP	Intron
		10:103116846	SNP	Intron
		10:103116928	SNP	Intron
		10:103116980	SNP	Intron
		10:103116988	SNP	Intron
	SDS	17:61030286	SNP	5′UTR
		17:61030372	SNP	5′UTR
		17:61031090^61031091	Insertion	Unannotated[1]
	BoLA-DRA	23:25839069	SNP	Intron
		23:25837511	SNP	3′UTR
		23:25837540	SNP	3′UTR
Healthy	*FCER1G*	3:8289327..8289351	Deletion	Intron
	GLYCAM1	5:25480359^25480360	Insertion	Intron
	CD74	7:61754710..61754711	Deletion	Intron
	NFKBIA	21:45641090^45641091	Insertion	Intron
		21:45642227^45642228	Insertion	Intron
	BoLA-DRA	23:25841128^25841129	Insertion	Intron

[1] Unannotated: this refers to variants that are not yet annotated and thus the location in the gene is unknown.

there were 14 unique SNPs and 1 unique insertion fixed among all animals in the genes *GLYCAM1, CD74, B2M, SDS* and *BoLA – DRA*. Within the genes *GLYCAM1* and *B2M*, there were 2 unique variants (both located in introns) with predicted effects of MODIFIER and it is not expected that these specific variants would impact the protein functionality. There is also a SNP within *CD74,* but it is located in an intron. The gene with the most unique and fixed variants (7 SNPs) is *B2M*, however all of these variants are located in introns and aren't expected to impact the functionality of the protein. Alternatively, the genes *SDS* and *BoLA-DRA* have variants in the 5' UTR region (*SDS;* 2 SNPs) and 3' UTR region (*BoLA-DRA*; 2 SNPs) and variants located in these regions might modify regulatory elements, in turn affecting the interaction of UTRs and proteins and microRNAs.

Discussion

Although key variants were identified in the previous identified functional candidate genes using transcriptomics, future research should be focused on those located in exons, 3' UTR or 5' UTR regions, or those that are not yet annotated, as these could positively impact the functionality of these genes and could aid in the host responding to the mastitis causing agents. Similarly, emphasis should be placed on the candidate genes with the most functional and structural relevance, such as those which impact *B2M*. This gene is directly related with the major histocompatibility complex molecules and has multiple roles in immune regulatory mechanisms of defense toward internal or invasive threats (Liu *et al.*, 2018; Asselstine *et al.*, 2019). This gene also possesses antibacterial activity against *S. aureus*, which is a Gram-positive bacteria

and is considered one of the most contagious mastitis pathogens worldwide (Harmon, 1994; Botaro *et al.*, 2015).

In summary, our research used two different -OMICs technologies and approaches (RNA-Seq and WGS) to study the transcriptome and the genome across two independent populations of Holstein dairy cows. This allowed us to not only identify key DE functional candidate genes using RNA-Seq, but also identify potential SNPs and INDELs within these functional candidate regions that might impact mastitis resistance or the host's immune response using WGS. This research could benefit and positively impact the dairy industry by integrating functional SNPs and INDELs into genomic breeding programs, thus, helping producers decide the best animals to breed to decrease the prevalence of mastitis in their herds.

Acknowledgements

The authors acknowledge partners and collaborators of the efficient dairy genome project (https://genomedairy.ualberta.ca/) and the resilient dairy genome project (http://www.resilientdairy.ca).

References

Asselstine, V. *et al.* (2019). J. Dairy Sci. 102(10):9043–9059. https://doi.org/10.3168/jds.2019-16504

Bolger, A. M., Lohse, M. & Usadel, B. (2014). Bioinformatics. 30(15):2114–2120. https://doi.org/10.1093/bioinformatics/btu170

Botaro, B. G. *et al.* (2015). Trop Anim Health Prod 47:61–66. https://doi.org/10.1007/s11250-014-0683-5

Cánovas, A. *et al.* (2012). PLoS ONE 7(4):e35583. https://doi.org/10.1371/journal.pone.0035583

Cardoso, T. F. *et al.* (2018). BMC Genomics 19:145. https://doi.org/10.1186/s12864-018-4515-2

Harmon, R. J. (1994). J. Dairy Sci. 77: 2103–2112. https://doi.org/10.3168/jds.S0022-0302(94)77153-8

Liu, Y., Liao, J., Ku, T., Li, X. & Sheppard, A. M. (2018). CyTA – J. Food 16(1):281–286. https://doi.org/10.1080/19476337.2017.1394367

McLaren, W. *et al.* (2016). Genome Biol 17:122. https://doi.org/10.1186/s13059-016-0974-4

540. Allele specific expression as an indication of ploidy in pig IPECJ2 and chicken SL-29 cell lines

J. de Vos[1*], R.P.M.A. Crooijmans[1], M.F.L. Derks[1], S.L. Kloet[2], M.A.M. Groenen[1] and O. Madsen[1]

[1]Wageningen University & Research, P.O. Box 338, 6700 AH Wageningen, the Netherlands; [2]Leids Universitair Medisch Centrum, the Netherlands; jani.devos@wur.nl

Abstract

Cell lines are useful for investigating traits of interest e.g. intestinal absorption, feed efficiency and immunity in farm animals. We earlier investigated an intestinal cell line in pig and a fibroblast cell line in chicken and found chromosomal abnormalities by whole genome sequence (WGS) data analysis. Results from RNA-seq allele-specific expression analysis (ASE) in 4 cell lines showed aneuploidy in some chromosomes. In this paper we show that RNA-seq can be used to detect whole/partial chromosomal abnormalities based on ASE analysis.

Introduction

Pig and chicken are the primary sources for meat production worldwide and are also important for use as biomedical models and to study embryonic development. Cell lines are a valuable tool for gaining insight into genomic architecture and regulatory regions of genomes. Cell lines are derived from specific tissues of a species, and can either continue to divide endlessly or perish after a specific number of divisions (passages) (Verma *et al.*, 2020). Immortalized cell lines (i.e. that can be grown indefinitely) often show aneuploidy (presence of an abnormal number of chromosomes in a cell) or heteroploidy which is most pronounced in cancer cell lines (Verma *et al.*, 2020; Molina *et al.*, 2021). Animal cell lines from pig or chicken have proven useful to obtain insight in e.g. intestinal transport and immune function (Nossol *et al.*, 2015).

The phenotype of an individual is determined by gene expression which is a process where information from genes, encoded within DNA, is translated into proteins through RNA (Hartwell *et al.*, 2015). The phenomenon of unequal expression between alleles in diploid cells caused by either genetic variation or epigenetic regulation (cis-regulated gene expression) is known as ASE. Furthermore, non-haploid chromosomes show imbalances in gene transcriptional activity, implying that chromosomal abnormalities can influence gene expression of different alleles, resulting in ASE. Our aim was to investigate the usability of RNA-seq data to identify whole chromosomal abnormalities, with the pig IPEC-J2 and chicken SL-29 cell lines as model.

Materials & methods

Cell lines. The pig IPECJ67 cell line (IPECJ2 cells grown for 67 passages) and chicken SL-29 cell line (grown for 4 passages) used were obtained from the cell repositories at DSMZ (https://www.dsmz.de/collection/catalogue/details/culture/ACC-701) and ATCC (https://www.lgcstandards-atcc.org/products/all/CRL-1590.aspx?geo_country=nl#generalinformation) respectively. The pig cell line is derived from intestinal epithelial cells while the chicken cell line is derived from embryonic fibroblast cells.

Data analysis. Pig (Sus Scrofa 11.1) and chicken (Gallus gallus GRCg6a) reference genomes, together with ENSEMBL annotations (Sus Scrofa 11.1 release 103 & Gallus gallus GRCg6a release 94) were utilized for all data analyses of our study.

Whole genome sequencing. Whole genome sequences were trimmed using Sickle v1.33 (https://github.com/najoshi/sickle) in paired end mode, followed by alignment (bwa mem v0.7.15 (Li., 2013)) of the

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_540, © J. de Vos *et al.* 2022

trimmed reads, and removal of duplicates (Samblaster v0.1.26 (Faust and Hall, 2014)). Mate coordinates were added using samtools (v1.9, Li *et al.,* 2009). Single nucleotide variants (SNVs) were called using FreeBayes (v1.3.1), thereafter the read support ratio was assessed for heterozygous variants within the VCF file.

RNA sequencing. RNA-seq data (pig IPECJ67 and chicken SL-29) were trimmed for adapters, and minimum length using TrimGalore v0.6.4 (http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/). FastQC v0.11.9 (Andrews *et al.*, 2010) was implemented to evaluate the quality of the data. RSEM (Li and Dewey, 2011) was used for alignment of the trimmed reads, as well as gene quantification (STAR v2.7.3a as aligner) (Dobin *et al.*, 2013). Average gene expression level per chromosome was calculated and plotted using a custom python script with the Seaborn package. Additional raw RNA-seq data from the PRJNA610529 project was downloaded from ENA. This data comprises a pig Jejunum organoid sample (12 weeks (SAMN14300021), a 5 week old pig Jejunum tissue sample (SAMN14300018), cell lines IPECJ87, an IPECJ2 cell line grown for 87 passages (SAMN14300016), and IPECJ91 an IPECJ2 cell line grown for 91 passages (SAMN14299997). We trimmed, aligned, and completed gene quantification of this data following the same procedure as the above for the IPECJ67 cell line. These samples were used to compare the average gene expression levels per chromosome in the tissue, organoid, and cell lines to the IPECJ67 cell line. Initial ASE analysis was completed with variant calls from WGS data (FreeBayes, Garrison and Marth, 2012) and aligned reads from RNA-seq using GATK ASEReader (McKenna *et al.,* 2010). Variant calling from RNA-seq alignments for SL-29, IPECJ67, IPECJ87, and IPECJ91 was completed using FreeBayes and thereafter allele specific expression analysis was completed using GATK ASEReader implementing the bam and VCF files. Results from the ASE analysis was plotted using Seaborn package in python.

Results

Chromosomal abnormalities within the cell-line genomes. In a previous study of the cell lines IPECJ67 and SL-29 we analysed the structure of the genome in multiple aspects using whole genome sequence data (de Vos *et al.* manuscript in preparation). In pig IPECJ67 we observed that chromosome 16 is diploid and chromosome 17 is triploid based on the frequency of the alleles (data not shown). Additionally, aneuploidy and structural variations were observed in many other chromosomes in the IPECJ67 cell line as well as chicken SL-29 (e.g. chromosome 20 is tetraploid).

Allele specific expression. RNA-seq data provides insight into gene expression levels at a chromosome-wide level. We investigated the expression of genes and transcripts in the pig IPECJ67, IPECJ87, IPECJ91 cell lines, jejunum intestinal tissue and jejunum derived organoid sample (data not shown). We found elevated gene expression on chromosome 17 in the cell lines compared to organoid and tissue whereas for the diploid chromosomes 15 and 16 similar gene expression levels across tissue, organoids and cell lines, with slightly higher expression levels in cell lines were observed. Investigating the allele specific expression of the IPECJ67 cell line (Figure 1) shows that the frequency of the allele expression displays a pattern that confirms chromosome 17 as triploid and chromosomes 15 and 16 as diploid.

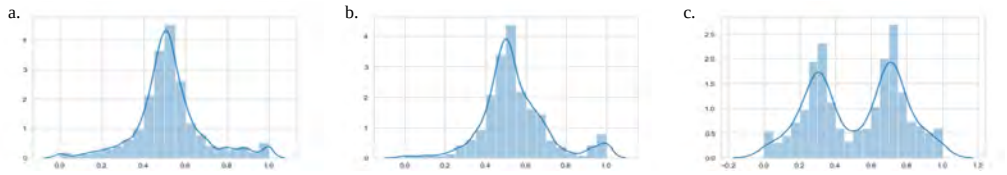


Figure 1. Allele specific expression of chromosomes 15(a), 16(b), 17(c) for the IPECJ67 cell line supports the ploidy level of the three chromosomes – allele expression of ~0.5 for diploids and allele expression of ~0.33 and ~0.7 for triploids, respectively.

Further investigation of the ASE and chromosomal abnormalities of the IPECJ2 derived cell lines cultured for a longer time also showed triploidy for chromosome 17 (Figure 2 c,f) in both cell cultures. Interestingly, chromosome 16 was observed as triploid (Figure 2 b,e) for these cell cultures, which is different from IPECJ67 which was diploid for chromosome 16 (Figure 1b).

Similar patterns of allele distribution for aneuploid chromosomes detected using ASE are observed on other chromosomes in pig IPECJ67 and in chicken SL-29 (data not shown). Lastly, we called the genome variation from the RNA-seq data and repeated the ASE analysis based on the RNA-seq called genome variation. This resulted in the same ASE results as using the variation called from the genome sequencing data (data not shown). Thus, genome sequencing is not required to detect genome aberrations based on ASE.

Discussion

Detecting chromosome abnormalities in e.g. cell lines is important as such abnormalities may influence the results obtained in cell line assays, especially if genes involved in the process studied show copy number variation as this will result in deviating expression (Figure 1c). Chromosome ploidy can be investigated by e.g. cytogenetic and whole genome sequencing methods, but these methods are generally not applied in standard cell-line research. As RNA-seq is (becoming) the general method used for gene expression detection, also in cell-lines, we investigated the potential of implementing RNA-seq as a tool for detecting chromosomal abnormalities through variant calling, ASE analysis and plotting the allele ratios. We observed similar patterns of allele support ratios from WGS variant calling and RNA-seq allelic expression variants. This suggests we do not need WGS data for detection of chromosomal abnormalities. We show that a larger number of passages of the IPECJ2 cell line results in an increased level of aneuploidy. Thus, our approach, using RNA-seq, is a sufficient and cost-effective tool to detect aneuploidy, which we suggest to be regularly applied in cell line experiments.

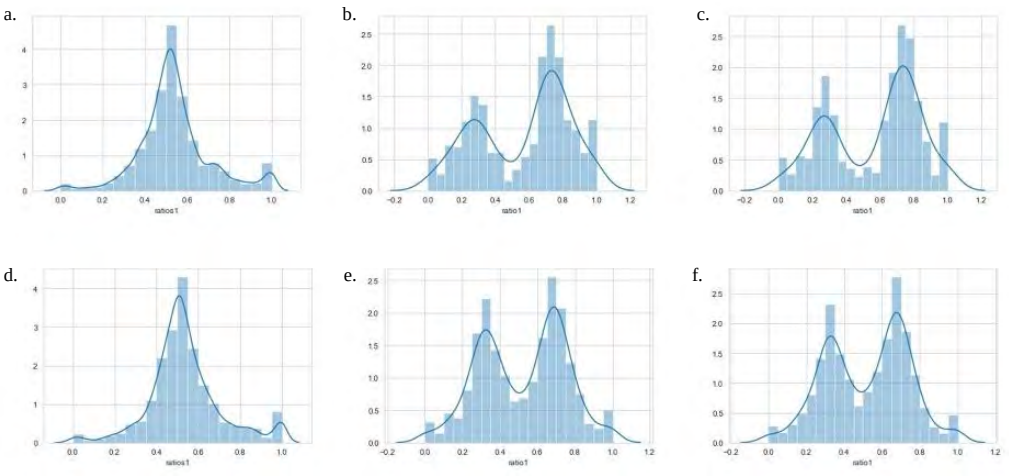


Figure 2. ASE for IPECJ2 derived cell lines cultured for a longer time (passages). IPECJ87 (a) chromosome 15, (b) chromosome 16, (c) chromosome 17. IPECJ91 (d) chromosome 15, (e) chromsome16, (f) chromosome17.

Acknowledgements

This project has received funding from the European Union's Horizon 2020 Research and Innovation Programme under the grant agreement no. 817998. Due to the restriction of the page number for this paper we were unable to include figures for all the results. The results not included can be obtained by contacting the corresponding author(s).

References

Andrews, S. (2010). FastQC: A Quality Control Tool for High Throughput Sequence Data. Berlin: ScienceOpen, Inc.

Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M. and Gingeras, T.R., 2013. STAR: ultrafast universal RNA-seq aligner. Bioinformatics, 29(1), pp.15-21.

Faust, G. G., & Hall, I. M. (2014). SAMBLASTER: fast duplicate marking and structural variant read extraction. Bioinformatics, 30(17), 2503-2505.

Garrison E, Marth G. (2012). Haplotype-based variant detection from short-read sequencing. arXiv preprint arXiv:1207.3907.

Hartwell, L., Goldberg, M.L., Fischer, J.A., Hood, L.E., Aquadro, C.F., Bejcek, B. (2015). Genetics: from genes to genomes (pp.). New York: McGraw-Hill.

McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M, DePristo MA. (2010). The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res, 20:1297-303. https://doi.org/10.1101/gr.107524.110.

Molina, O., Abad, M.A., Solé, F. and Menéndez, P. (2020). Aneuploidy in Cancer: Lessons from Acute Lymphoblastic Leukemia. Trends in Cancer.

Li, B., & Dewey, C. N. (2011). RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC bioinformatics, 12(1), 1-16.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G. and Durbin, R., 2009. The sequence alignment/map format and SAMtools. Bioinformatics, 25(16), pp.2078-2079.

Li H. (2013) Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv:1303.3997v2 [q-bio.GN].

Nossol, C., Barta-Böszörményi, A., Kahlert, S., Zuschratter, W., Faber-Zuschratter, H., Reinhardt, N., Ponsuksili, S., Wimmers, K., Diesing, A.K. and Rothkötter, H.J., (2015). Comparing two intestinal porcine epithelial cell lines (IPECs): morphological differentiation, function and metabolism. PloS one, 10(7), p.e0132323.

Verma, A., Verma, M. and Singh, A., (2020). Animal tissue culture principles and applications. In Animal Biotechnology (pp. 269-293). Academic Press.

541. Segregation of the structural variant responds for mastitis resistance in re-sequencing Nordic Holstein animals

Z. Cai* and G. Sahana

Center for Quantitative Genetics and Genomics, Aarhus University, Blichers Allé 20, 8830 Tjele, Denmark; zexi.cai@qgg.au.dk

Abstract

An important source of genome sequence variations, structural variants (SVs), are largely ignored due to the difficulty to precisely calling these variants. In this study, we used whole genome sequence (WGS) data to capture the SVs in various breeds. We have identified 84,561 SVs in 567 WGS animals. When compared the called SVs with the major quantitative trait locus for mastitis resistance on chromosome 6 near to *GC* gene, we found a duplication (DUP) is highly similar to previous reported copy number variant. Even though the linkage disequilibrium between the lead SNP for mastitis resistance and the called DUP is not complete, the SVs profile we obtained still showed the potential for identifying functional variants.

Introduction

Structural variants (SVs) are referred to the variants which are large in size (>50bp). The type of SVs include insertion (INS), deletion (DEL), duplication (DUP), copy number variants (CNV), inversion (INV), and translocations. Generally, such variants are difficult to capture using short-read sequencing due to the length of the sequence reads ranging 30-250bp. Therefore, SVs are not a typically included when using whole-genome sequence (WGS) data to obtain high density variants map. Recently, more and more reports have proven SVs are important variants that usually have high impact on the phenotypes (Mahmoud *et al.*, 2019) and have the potential to pinpoint the candidate genes in the GWAS hits. Therefore, calling for SVs for reference panel will be informative. Even though using short-read sequencing with moderate coverage to identify SVs is less powerful than long read sequencing and deep sequencing, the choice of the method could partially overcome the disadvantage of the short reads and achieve high accuracy of SVs calling.

In this study, we collected previously generated WGS dataset with moderate sequencing depth to obtain the SVs in various cattle breeds. Smoove (https://github.com/brentp/smoove) pipeline was used to call SVs, which perform the calling, filtering and annotation of SVs. The potential present of SVs around the known mastitis resistance quantitative trait locus (QTL) on chromosome 6 in Holstein cattle was investigated.

Materials & methods

Animals. In total, WGS from 567 animals were used in this study, which include 123 Holstein (HOL), 60 Jersey (JER), 175 Nordic Red Dairy cattle (RDC) and 209 from other breeds. DNA was extracted from semen samples using standard procedures at Aarhus University, Denmark. Sequencing was done using Illumina sequencing platform using shot-gun pair-end strategy with approximated 10 fold coverage.

Mapping and Structural variants calling. The raw reads were subjected to Trimmomatic 0.38 (Bolger *et al.*, 2014) to remove the adapter sequence and trim the low quality base. Then the clean reads were mapped to the cattle reference genome ARS-UCD1.2 (Rosen *et al.*, 2020) using *bwa mem* (Ma *et al.*, 2020) with parameter 'M' to mark shorter split hits as secondary. BaseRecalibrator was applied to the raw alignment bam files using GATK 3.8 (McKenna *et al.*, 2010). We applied the smoove (https://github.com/brentp/smoove) pipeline to call SVs using the base recalibrated bam files with default parameter. The statistics of the SVs in Holstein, Jersey and Nordic Red was calculated using PLINK v1.9 (Purcell *et al.*, 2007).

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_541, © Z. Cai and G. Sahana 2022

Locuszoom plot. We reused previous GWAS result (Cai *et al.*, 2018) of mastitis resistance in Danish Holstein as an example to show the potential function of structural variants. We applied the locuszoom (Pruim *et al.*, 2010) to visualize the surrounding region of the lead SNP. The regulatory elements were retrieved from previous study (Kern *et al.*, 2021) and severed as additional track.

Results & discussion

The statistics of the structural variants identified. We have identified 84,561 SVs, including 9,416 DUP, 59,709 DEL, 3,982 INV, 7,471 pairs of generic breakends (BND) across different chromosomes, and 3,983 BNDs within the same chromosome which could indicated large deletions. Furthermore, there are 66,031, 51,695 and 79,221 segregating SVs in HOL, JER and RDC respectively (Table 1). In addition, 45,894 SVs are shared by these three breeds, and RDC have largest number of breed-specific SVs (Figure 1). Besides, HOL and RDC also share a large number of SVs that are not presented in JER. The number of the SVs is affected by the number of samples and also the population, which is consistent with previous study (Kommadath *et al.*, 2019).

The QTL of mastitis resistance of chromosome 6. Previous studies have shown that there is a major QTL located on chromosome 6 near to *GC* gene for mastitis resistance. In our GWAS, we identified a SNP 6:86997761 with $-\log_{10}$ (*P*-value) equal to 10.51 as the lead SNP (Figure 2). In the SVs call, we identify 105 SVs located within the surrounding 1 Mb region (1 Mb up and 1 Mb down) of the lead SNP. We extracted

Table 1. The statistics of structural variants in Holstein, Jersey and Nordic Red cattle.

Breed	Number				Total
	DUP	DEL	INV	BND	
Holstein	6,449	39,619	3,192	16,771	66,031
Jersey	4,747	31,486	2,574	12,888	51,695
Nordic Red	7,412	49,238	3,595	18,976	79,221

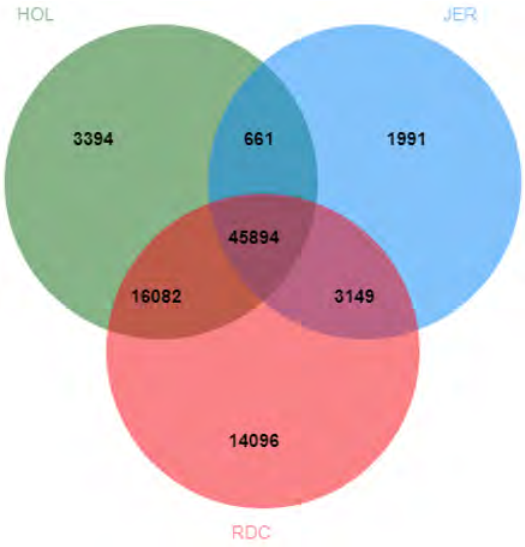


Figure 1. The venn plot shows the overlap of structural variants in Holstein (HOL), Jersey (JER) and Nordic Red Dairy cattle (RDC).

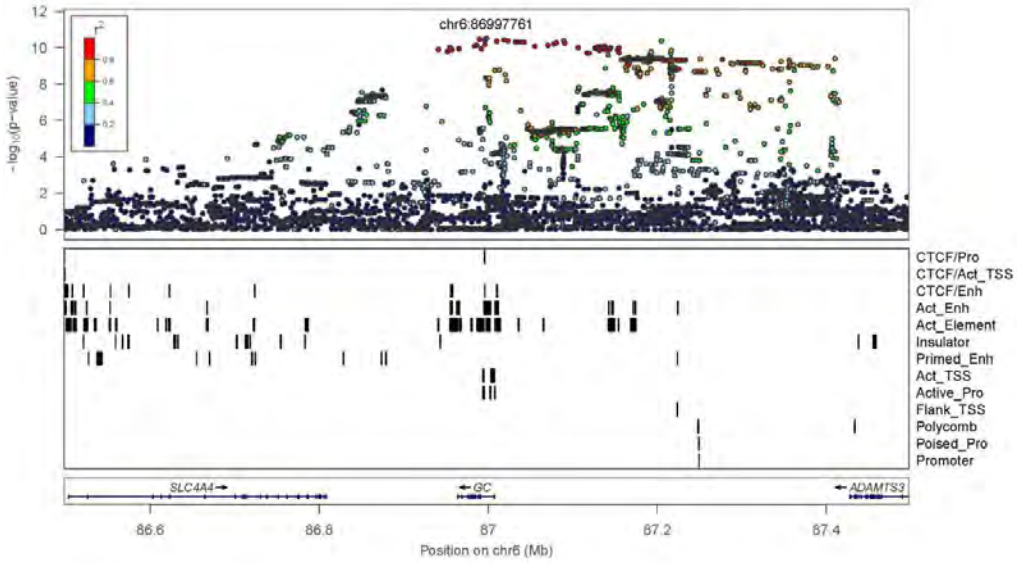


Figure 2. Locuszoom plot showed the lead SNP, LD pattern of the flanking SNPs, the flanking genes (with exon) and the predicted regulatory element from liver tissue.

the genotype of the lead SNP from the WGS animals and checking the LD pattern with the SVs. We only find 5 SVs are in LD with the lead SNP but none of them was in complete LD. These are a 863 bp DEL (r^2=0.22) start from 87,034,646, a 2,013 bp DEL (r^2=0.53) start from 87,065,781, a 272 bp DEL (r^2=0.54) start from 86,997,761, an 11,781 DUP (r^2=0.52) start from 86,949,652 and an 11,236 bp DUP (r^2=0.46) start from 86,950,196.

A previous study has identified a CNV ranging from 86,949,653 ~ 86,961,428 as a potential candidate mutation for the mastitis QTL (Lee *et al.*, 2021). In our case, we have a similar DUP ranging from 86,949,652 ~ 86,961,433. The annotation from smoove showed the start and end point could be imprecise, so it is highly possible these two mutations are the same or the region undergo similar evolutionary event. The flanking area of the lead SNPs and these five SVs are overlapped with active regulatory elements (Figure 2). Even though, the available evidence did not confirm the causality, the SVs profile show the potential for wider scanning for causal variants for QTL whole genome wide and their functional validation.

Acknowledgements

We are grateful to the Nordic Cattle Genetic Evaluation (NAV, Aarhus, Denmark) for providing the phenotypic data used in this study and Viking Genetics (Randers, Denmark) for providing samples for genotyping.

References

Bolger, A. M., M. Lohse, and B. Usadel. 2014. Bioinformatics 30(15):2114-2120. https://doi.org/10.1093/bioinformatics/btu170.

Cai, Z., B. Guldbrandtsen, M. S. Lund, and G. Sahana. 2018. BMC Genomics 19(1):656. https://doi.org/10.1186/s12864-018-5050-x.

Kern, C., Y. Wang, X. Xu, Z. Pan, M. Halstead, *et al.* 2021. Nature Communications 12(1):1821. https://doi.org/10.1038/s41467-021-22100-8.

Kommadath, A., J. R. Grant, K. Krivushin, A. M. Butty, C. F. Baes, *et al.* 2019. GigaScience 8(6). https://doi.org/10.1093/gigascience/giz073.

Lee, Y. L., H. Takeda, G. Costa Monteiro Moreira, L. Karim, E. Mullaart, *et al.* 2021. PLoS Genet 17(7):e1009331. https://doi.org/10.1371/journal.pgen.1009331.

Ma, K. C., T. D. Mortimer, M. A. Duckett, A. L. Hicks, N. E. Wheeler, *et al.* 2020. arXiv preprint arXiv:1303.3997. https://doi.org/10.1101/2020.03.24.006650.

Mahmoud, M., N. Gobet, D. I. Cruz-Dávalos, N. Mounier, C. Dessimoz, *et al.* 2019. Genome Biology 20(1):246. https://doi.org/10.1186/s13059-019-1828-7.

McKenna, A., M. Hanna, E. Banks, A. Sivachenko, K. Cibulskis, *et al.* 2010. Genome research 20(9):1297-1303.

Pruim, R. J., R. P. Welch, S. Sanna, T. M. Teslovich, P. S. Chines, *et al.* 2010. Bioinformatics 26(18):2336-2337. https://doi.org/10.1093/bioinformatics/btq419.

Purcell, S., B. Neale, K. Todd-Brown, L. Thomas, M. A. Ferreira, *et al.* 2007. The American journal of human genetics 81(3):559-575.

Rosen, B. D., D. M. Bickhart, R. D. Schnabel, S. Koren, C. G. Elsik, *et al.* 2020. Gigascience 9(3):giaa021. https://doi.org/10.1093/gigascience/giaa021.

542. Rare CNVs in the bovine genome are not captured well by 50K density genotyping array SNPs

Y.L. Lee[1*], M. Bosse[1], W. Coppieters[2,3], R.F. Veerkamp[1], L. Karim[3], C. Oget-Ebrad[2], T. Druet[2], M.A.M. Groenen[1], M. Georges[2], A.C. Bouwman[1] and C. Charlier[2]

[1]Wageningen University & Research, Animal Breeding and Genomics, P.O. Box 338, 6700 AH Wageningen, the Netherlands; [2]Unit of Animal Genomics, GIGA-R & Faculty of Veterinary Medicine, University of Liège, 4000 Liège, Belgium; [3]GIGA Genomics Platform, GIGA Institute, University of Liège, 4000 Liège, Belgium; younglim.lee@wur.nl

Abstract

Understanding how genomes, particularly genetic variants, instruct animals to develop and function is crucial for animal breeders. Copy number variants (CNVs), the gain or loss of DNA segments, can affect gene expression and alter phenotypes. However, most large-scale genetic analyses in animal breeding use single nucleotide polymorphism (SNP) genotypes from genotyping arrays, and efforts to incorporate CNVs are limited. In theory, variation from CNVs can be captured by SNPs, if they are in high linkage disequilibrium (LD). We evaluated the LD of CNV-SNP pairs in whole genome sequencing and genotyping array data. Our whole genome sequencing data showed that most CNVs have tagging SNPs. However, CNV-SNP pairs from genotyping array data were mostly in low LD, because most CNV-SNP pairs had unmatching allele frequencies. We conclude that most rare CNVs may not be fully captured by genotyping arrays.

Introduction

Bovine genomes harbour diverse types of genetic variants. Identifying and utilizing impactful genetic variants is of prime interest for animal breeders, given the economic importance of cattle populations. Copy numbre variants (CNVs) are a subset of structural variants, which include deletions (DELs) and duplications (DUPs) of DNA segments larger than 50 bp. Although CNVs are not as abundant as small genetic variants (e.g. SNPs), CNVs contribute disproportionally more to gene expression than SNPs (Chiang *et al.,* 2017; Scott *et al.,* 2021), hinting that they could be associated with phenotypes. Recent reports in farm animals, which unravelled high impact CNVs as causative variants underlying QTL of complex traits, also support the importance of CNVs (Lee *et al.* 2021; Derks *et al.* 2018; Kadri *et al.* 2014).

Despite the functional evidence of CNVs, their utilization in animal breeding is limited. To date, large-scale routine genetic analyses in breeding programmes, such as genomic prediction, have mostly focused on SNPs obtained from genotyping arrays. In theory, if those SNPs are in high LD with CNVs, the variation from CNVs can be captured by SNPs. To this end, we aimed at: (1) detecting CNVs in deeply sequenced genomes of Holstein Friesian (HF) cattle; (2) validating a subset of CNVs in an independent cohort using a direct genotyping approach; and (3) evaluating pairwise LD in CNV-SNP pairs in whole genome sequencing (WGS) and genotyping data.

Materials & methods

WGS data. A family cohort (DAMONA) consisting of 266 HF animals (including 127 trios) was sequenced using the Illumina HiSeq 2000 instrument. The data was aligned using BWA mem (version 0.7.9a-r786) (Li 2013) to the bovine reference genome ARS-UCD1.2 (Rosen *et al.* 2020). All samples had a minimum mean sequencing coverage of 15X, and the mean coverage was 26X.

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_542, © Y.L. Lee *et al.* 2022

SNP and CNV discovery in the WGS data. SNP variant calling was done using the GATK workflow (v4.1.7), and recalibrated using the following algorithms: BaseRecalibrator, HaplotypeCaller, GenomicsDBImport, GenotypeGVCF, GatherVcfs, Variant Recalibrator (DePristo *et al.* 2011; McKenna *et al.* 2010; Auwera *et al.* 2013). Variant Quality Score Recalibration (VQSR) at a truth sensitivity filter level of 97.5 was used to remove spurious variants. CNVs were detected using the Smoove pipeline, which utilized split and discordant reads evidence to discover CNVs in each sample, followed by population-wide genotyping (https://github.com/brentp/smoove). Subsequently, we used the fold-coverage change of read depth in CNVs to retain accurately genotyped biallelic CNVs, using Duphold (Pedersen and Quinlan 2019).

Genotyping data. To validate CNV discovery results, we genotyped a subset of CNVs by adding them in the custom part of the EuroGenomics SNP genotyping array, which harbours ~50K SNPs (Boichard *et al.* 2018). Probes targeting breakpoint sequences of 372 CNVs (342 deletions and 30 duplications) that appeared in non-repetitive regions, were added. Genotyping was done for 815 independent HF animals using their ear punch or blood samples.

LD analyses. Pairwise LD (r^2) for CNV-SNP pairs was calculated, to evaluate whether CNVs had tagging SNPs. This analysis was done separately for the WGS and EuroGenomics array data sets. SNPs located within 100-kb distance from CNV breakpoints (WGS) or CNV probes (array) were considered for LD calculation. Additionally, LD decay was compared between CNV-SNP and SNP-SNP pairs limited to the Eurogenomics array data. This analysis was done for common (allele frequency≥0.05) and rare (allele frequency<0.05) DELs separately. Due to a limited number, DUPs was not considered (n=19). The LD calculation was done using the PLINK software (Purcell *et al.* 2007).

Results

Overall variant discovery in the WGS data set. The SNP variant calling pipeline resulted in 11,030,905 SNPs from the 266 sequenced samples. Using the same WGS data set, we discovered 13,731 CNVs (12,200 deletions and 1,531 duplications). Genotyping accuracy of CNVs is lower than that of SNPs (Chiang *et al.* 2017). Hence, we applied read-depth based filters, resulting in 4,011 accurately genotyped biallelic CNVs (3,827 deletions and 184 duplications).

Validation of CNVs using genotyping arrays. We validated a subset of WGS CNVs, using a direct genotyping approach in 815 HF animals. All genotyped samples had a sample call rate >0.99). Of the 284 CNVs that passed the variant level filter, 229 were segregating (210 deletions and 19 duplications), showing a validation rate of 80% (229/284). The SNP genotyping results showed that 50,342 SNPs were segregating in the population. The minor frequency spectrum of CNVs was skewed towards rare variants, whereas that of SNPs were skewed towards common variants (Figure 1).

LD analyses in the WGS and the genotyping data. To evaluate whether WGS CNVs have tagging SNPs in WGS data, we calculated the LD of CNV-SNP pairs. A majority of biallelic WGS CNVs (97% of DELs and 93% of DUPs) had at least one WGS SNP in LD (r^2>0.8) within a 100-kb distance from the CNV. However, the same analyses based on the EuroGenomics array data revealed that only 15.4% of DELs and 4.5% of DUPs included on the array had a tagging array SNP (r^2>0.8) within a 100-kb distance. Lastly, the LD in the EuroGenomics array data confirmed that the CNV-SNP pairs have low LD than SNP-SNP pairs (Figure 2; analyses was done limited to DELs, due to low numbers of DUPs).

Discussion

A large percentage of the CNV in our WGS data set were well tagged by at least one WGS SNP. In a human study, only ~80% of CNVs had at least one tagging SNPs, based on low-pass WGS data (4-7X coverage;

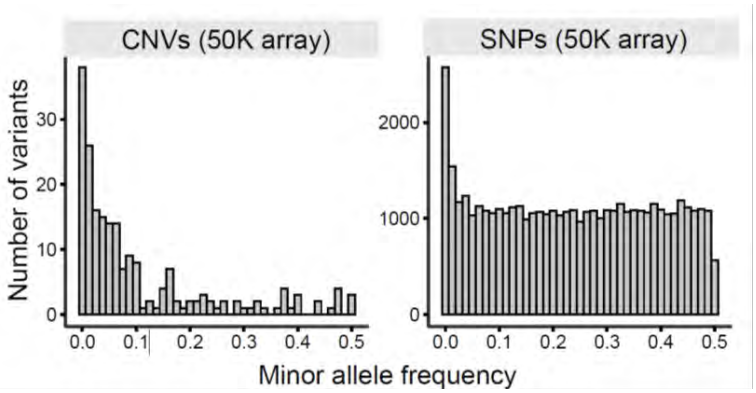


Figure 1. Minor allele frequency spectra of CNVs and SNPs obtained from the EuroGenomics array.

Sudmant *et al.* 2015). The higher LD in our data set is likely due to: (1) high relatedness among the animal samples, as opposed to the non-related human samples in Sudmant *et al.* (2015); and (2) deeper WGS data in our study. The deeper WGS contributed to more accurate genotyping of CNVs, which may have resulted in higher LD between CNV and SNP genotypes.

The LD between CNV and SNP in our array genotyped data set was low. The frequency spectra of CNVs and SNPs were discordant: CNVs were skewed towards rare variants, whereas SNPs were enriched for common variants. Thus, only a small fraction of CNVs, with high allele frequencies, was tagged by SNPs in the EuroGenomics array. Furthermore, the generally low LD in CNV-SNP pairs compared to SNP-SNP pairs confirmed previous findings based on the Illumina BovineHD Genotyping BeadChip data (Figure 2; Lee *et al.* 2020).

Taking these results together, we conclude that most CNVs have tagging SNPs, which can be identified in deep WGS data. However, such tagging SNPs may not be present in genotyping arrays, as arrays tend to harbour common SNPs. This, in turn, suggests that a small fraction of CNVs, segregating at high frequency, may be tagged well by SNPs in the genotyping arrays. However, most CNVs are rare, and hence not properly captured with the set of SNPs in genotyping arrays. To mitigate this gap, we propose to utilize the WGS CNV catalogue to prioritize likely functional CNVs, particularly rare ones, and directly genotype them.

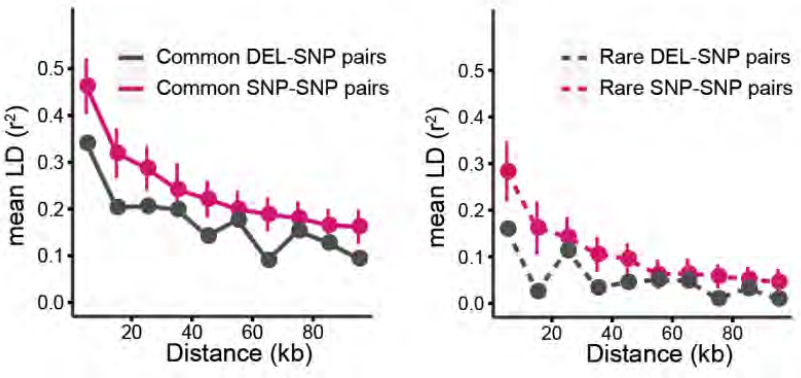


Figure 2. LD decay patterns in DEL-SNP and SNP-SNP pairs.

Funding

This study was financially supported by the Dutch Ministry of Economic Affairs (TKI Agri and Food project 16022) and the Breed4Food partners Cobb Europe, CRV, Hendrix Genetics and Topigs Norsvin. The HF WGS data set was funded by the DAMONA ERC advanced grant to MG.

References

Auwera G.A. Van der, Carneiro M.O., Hartl C., Poplin R., Angel G. del, *et al.* 2013 Curr Protoc Bioinformatics 11(1110): 11.10.1–11.10.33. https://doi.org/10.1002/0471250953.bi1110s43

Boichard D., Boussaha M., Capitan A., Rocha D., Sanchez M.P. *et al.* 2018. In 11th World Congress on Genetics Applied to Livestock Production, 11(675) pp. 1–6.

Chiang C., Scott A.J., Davis J.R., Tsang E.K., Li X. *et al.* 2017. Nat Genet 49: 692–699. https://doi.org/10.1038/ng.3834

DePristo M.A., Banks E., Poplin R., Garimella K.V., Maguire J.R. *et al.* 2011. Nat Genet 43: 491–8. http://dx.doi.org/10.1038/ng.806.

Derks M.F.L., Lopes M.S., Bosse M., Madsen O., Dibbits B. *et al.* 2018. PLoS Genet 14: 1–20. http://dx.doi.org/10.1371/journal.pgen.1007661.

Kadri N.K., Sahana G., Charlier C., Iso-Touru T., Guldbrandtsen B. *et al.* 2014. PLoS Genet 10: 1-11 https://doi.org/10.1371/journal.pgen.1004049

Lee Y.-L., Bosse M., Mullaart E., Groenen M.A.M., Veerkamp R.F. *et al.* 2020. BMC Genomics 21: 1–15. https://doi.org/10.1186/s12864-020-6496-1

Lee Y.-L., Takeda H., Moreira G.C.M., Karim L., Mullaart E. *et al.* 2021. PLoS Genet 17: 1–27. http://dx.doi.org/10.1371/journal.pgen.1009331.

Li H. 2013. arXiv Prepr arXiv 00: 3. http://arxiv.org/abs/1303.3997.

McKenna A., Hanna M., Banks E., Sivachenko A., Cibulskis K. *et al.* 2010. Genome Res 20: 1297–1303. https://doi.org/10.1101/gr.107524.110

Pedersen B.S. andQuinlan A.R.. 2019. Gigascience 1–5. https://doi.org/10.1093/gigascience/giz040

Purcell S., Neale B., Todd-brown K., Thomas L., Ferreira M.A.R. *et al.* 2007. Am J Hum Genet 81: 559–575. https://doi.org/10.1086/519795

Rosen B.D., Bickhart D.M., Schnabel R.D., Koren S., Elsik C.G. *et al.* 2020. Gigascience 9: 1–9. https://doi.org/10.1093/gigascience/giaa021

Scott A.J., Chiang C., Hall I.M.. 2021. Genome Res. 2021. 31: 2249-2257. https://doi.org/10.1101/gr.275488.121

Sudmant P.H., Rausch T., Gardner E.J., Handsaker R.E., Abyzov A. *et al.* 2015. Nature 526: 75–81. https://doi.org/10.1038/nature15394

543. Genome-wide DNA methylation of Gonad in sex-biased zebrafish families

S. Hosseini[1,2,3*], N. Trakooljul[4], M. Hirschfeld[1,2], K. Wimmers[4], J. Tetens[1,3], A.R. Sharifi[1,3] and B. Brenig[1,2,3]

[1]University of Goettingen, Department of Animal Sciences, Goettingen, Germany; [2]Institute of Veterinary Medicine, Goettingen, Germany; [3]University of Goettingen, Center for Integrated Breeding Research (CiBreed), Goettingen, Germany; [4]Leibniz Institute for Farm Animal Biology (FBN), Institute for Genome Biology, Genomics Unit, Dummerstorf, Germany; shahrbanou.hosseini@uni-goettingen.de

Abstract

To investigate the effect of DNA methylation on sex ratio variation in response to environmental temperature, in this study we used two sex-biased zebrafish families for genome-wide DNA methylation analysis using reduced representation bisulfite sequencing (RRBS). The results showed a higher number of differentially methylated CpG sites (DMCs) in the male-biased family (218,850) compared to the female-biased family (211,943). Likewise, a greater number of methylated positions was identified in the male gonads (4,151) than the female gonads (1,993). The effect of temperature showed more methylated positions in high incubation temperature than the control temperature. We also found differential methylation of many reproduction-related genes (e.g. *sox9a, gata4, lhx8a*) and genes involved in epigenetic mechanisms (e.g. *dnmt3bb.1, dimt1l, h1m*) in both families. We conclude that epigenetic modifications would be more critical in male-skewed sex ratio, which may have negative consequences for population fitness in thermosensitive species under rapid environmental changes.

Introduction

Phenotypic sexual plasticity in response to environmental conditions leads to sex-reversed individuals and variation in sex ratio, which consequently influence population dynamics and loss of biodiversity. This phenomenon is found in animal species with genotype-environment interaction (G×E) sex-determining system such as reptiles, amphibians, and fish (Piferrer *et al.*, 2019). The plasticity of sexual phenotype can be influenced by epigenetic mechanisms, which integrate genomic and environmental factors to generate a particular phenotype through changes in the transcriptional activity of genes in response to environmental stimuli (Piferrer *et al.*, 2019). Understanding the mechanisms underlying G×E sex-determining system is a fundamental research interest for conservation of natural populations and livestock species (Shao *et al.*, 2014) in order to develop counterstrategies for dealing with unforeseeable ecological consequences in light of global climate change. Zebrafish (*Danio rerio*), is an excellent research model to study the mechanism of sex determination and the effect of environmental changes on gonadal differentiation in vertebrates (Ribas and Piferrer, 2014). Hence, the main goal of this study was to investigate the epigenetic landscape of DNA methylation in the context of sexual plasticity in thermosensitive animal species using the zebrafish. Therefore, we performed genome-wide DNA methylation analysis in zebrafish gonad to address the effect of epigenetic modifications on sex ratio variation of sex-biased families in response to environmental temperature.

Materials & methods

In this trial, the fertilized eggs from 17 zebrafish families were divided in equal proportions and subjected to two different incubation temperatures: (1) control group was kept at a constant temperature of 28 °C throughout the experiment; (2) high incubation temperature group (treatment group) was exposed to high water temperature of 35 °C from 5 to 24 hours post fertilization. The phenotypic sex of all individuals was determined in the adult stage (90-120 days post fertilization). To select the male-biased (more than 60%

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_543, © S. Hosseini *et al.* 2022

males in the control group) and the female-biased (less than 40% males in the control group) families (Liew *et al.*, 2012) for RRBS analysis, two decision criteria using a linear logistic model were applied: (1) a high difference in the proportion of males between the families under control condition; and (2) a high difference in the proportion of males between control and treatment groups within the families. After selecting the two sex-biased families with significant differences in sex ratio, a total number of 72 gonad tissue samples were collected from adult males and females of the control and treatment groups for laboratory analysis. The genomic DNA was isolated using the AllPrep DNA/RNA Mini Kit (Qiagen), which were then subjected to bisulfite conversion using the EZ DNA Methylation-Gold Kit (Zymo Research) according to the manufacturer's protocols. The RRBS libraries were sequenced for 121bp single reads on the HiSeq 2500 system according to the manufacturer's recommendations using the HiSeq SR Cluster and HiSeq SBS Kit v4 (Illumina). For bioinformatics analysis of raw data, the base call (BCL) files were converted into fastq files. The fastq files were further processed for quality control of sequencing reads. The cleaned reads were then mapped to the *Danio rerio* reference genome (GRCz11) and the methylation calls were performed using Bismark software. Differential methylation analysis was conducted using 'edgeR' package of R-program to identify the DMCs between (male-biased versus female-biased family in each gonad type) and within (male versus female gonads in each family type) families. The standard false discovery rate (FDR) approach was applied for methylation differences between the comparison groups (FDR<0.05). To assess the effect of DNA methylation on genes involved in sex determination of zebrafish, we generated a list of reproduction- and epigenetic-related genes (n=174 candidate genes) to align the results of DNA methylation with the candidate genes.

Results

Differentially methylated CpG sites between male-biased and female-biased families. To assess the differences in DNA methylation of sex-biased zebrafish families, we compared the methylation profiles of male-biased (MB) versus female-biased (FB) family within female and male gonads in control (FC, MC) and treatment (FT, MT) conditions in different experimental groups: FCMB vs FCFB, MCMB vs MCFB, FTMB vs FTFB, MTMB vs MTFB (Figure 1A and 1B). The gonadal tissue samples of 6 females and 12 males in each experimental group were analysed in this study. Analysis of the methylation profiles between the two sex-biased families in each gonad type revealed that a total of 805 CpGs in the control group and 1,188 CpGs in the treatment group were differentially methylated in the female gonads. The same comparison in the male gonads showed the differentially methylation of a total of 1,969 CpGs and 2,182 CpGs in the control and treatment groups, respectively. This observation indicates that the male gonads harbored more hypermethylated CpG sites than the female gonads and the treatment group more than the control group when comparing the two sex-biased families with different genetic background of sex determination. The analysis of the association between the DNA methylation changes of the nearest annotated DMCs genes (±10 kb distances away from transcription start site, TSS) comparing male-biased versus female-biased family (Figure 1A) and the candidate genes revealed the identification of 23 genes, in which two methyltransferase genes (*trmt10c* and *shmt2*) and one chromatin-related gene (*hmgn3*) and a subset of reproduction-related genes (e.g. *gdf9, star, nr5a2, piwil1, fmr1, sox3, lef1*) were detected in different experimental groups.

Differentially methylated CpG sites in male versus female gonads within sex-biased families. To understand the epigenetic mechanisms underlying plasticity of sexual development in zebrafish, we compared male versus female gonads within the two sex-biased families in different groups: MCMB vs FCMB, MCFB vs FCFB, MTMB vs FTMB, MTFB vs FTFB (Figure 2A and 2B). The DMCs within the male-biased family revealed a total of 201,139 hypermethylated sites in the male gonads, whereas only 17,711 hypermethylated sites were found in the female gonads. Similarly, in the female-biased family, a high number of hypermethylated CpGs were identified in the male gonads (193,121) compared to a

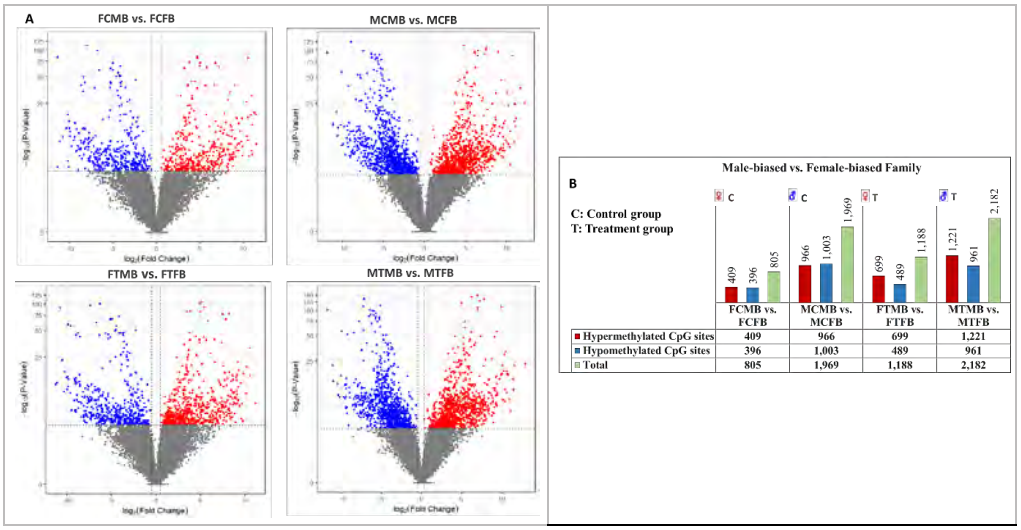


Figure 1. Differentially methylated CpG sites (DMCs) between sex-biased zebrafish families. (A) DMCs in male-biased versus female-biased family in each gonad type. (B) Overview of DMCs comparison between the two sex-biased families.

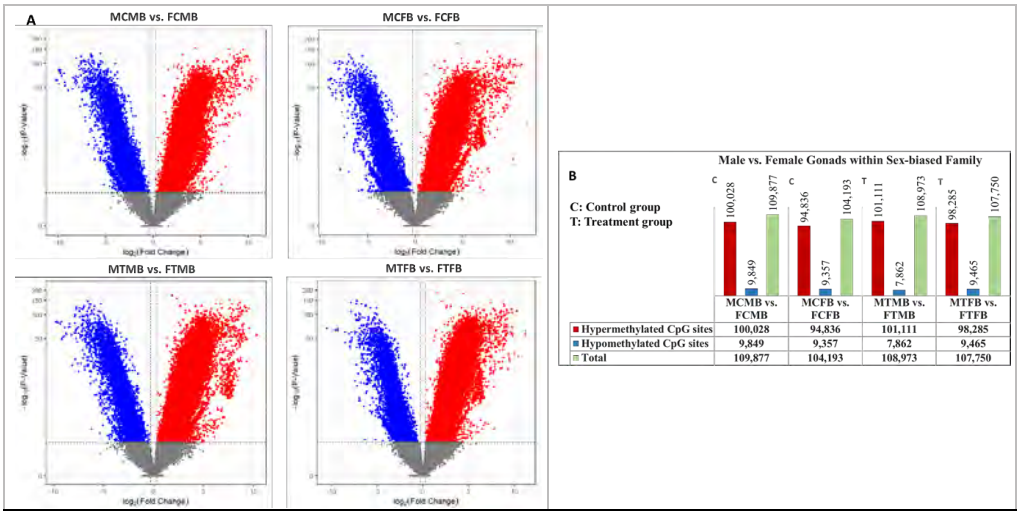


Figure 2. Differentially methylated CpG sites (DMCs) within each sex-biased zebrafish family. (A) DMCs in male versus female gonads within male-biased and female-biased families. (B) Overview of DMCs comparison within each sex-biased family.

lower number of hypermethylated CpGs in the female gonads (18,822). Our results demonstrate a higher number of hypermethylated sites in the male gonads versus a lower number of hypermethylated sites in the female gonads. Considering the effect of temperature on DMCs within both sex-biased families showed more hypermethylated CpGs in the treatment group than the control group. To find the effect of DNA methylation changes on genes within families (Figure 2A), we first annotated the DMCs with nearest genes (±10 kb distances away from TSS) comparing male versus female gonads, resulting in the detection of significant unique genes in each family (488 genes in MB and 389 genes in FB families in control group, 477 genes in MB and 426 genes in FB in treatment group) and significant common genes in overlap between the two sex-biased families (8,954 genes in control group and 8,988 genes in treatment group). These results were then used for the association between unique genes and candidate genes defined above, which revealed differential methylation of a few genes in all compared groups. However, in contrast to the unique genes, in the common genes of the two sex-biased families we identified a series of candidate reproduction-related (e.g. *sox9a, sox3, ctnnb1, lhx8a, gata4*) and epigenetic-related (e.g. *dnmt3bb.1, hdac4, hdac11, h1m*) genes in zebrafish gonad (100 candidate genes in control group and 104 candidate genes in treatment group).

Discussion

Epigenetically regulated sex determination in response to the environment plays a key role in phenotypic plasticity. Our study in sex-biased zebrafish families revealed pronounced differences in DMCs between male-biased and female-biased families, where the male gonads in the male-biased family showed more hypermethylated CpGs than the female gonads.

Identification of sex ratio variation in sex-biased zebrafish families with different genetic background of sex determination and different methylation in gonadal tissue in these families are evidence for epigenetic modifications underlying phenotypic sexual plasticity. The phenotypic divergence between the two sexes requires sex-specific gene regulation and differential sex-biased gene expression, which can be influenced by epigenetic regulatory mechanisms (Grath and Parsch, 2016). In this study, we identified the methylation of unique genes and a subset of the candidate genes studied here in each sex-biased family, especially in the group exposed to high temperature. These genes may directly or indirectly influence the gonadal development in downstream pathways, resulting in skewed sex ratios in a population. Among the differentially methylated candidate genes, the genes involved in testicular differentiation and spermatogenesis such as *sox9a*, *sox3*, and *gata4* were identified in our study. Likewise, the methylation of sex-biased genes involved in ovarian differentiation and folliculogenesis, such as *ctnnb1* and *lhx8a* was also observed in our analysis. We conclude that the effect of epigenetic modifications could be more critical in a population with male-skewed sex ratio with respect to rapid environmental changes under the predicted global warming scenario, which could not only lead to the extinction of natural populations but also have an impact on farm fish species.

References

Grath S., and Parsch J. (2016) Annu Rev Genet. 50:29-44. https://doi.org/10.1146/annurev-genet-120215-035429

Liew W.C., Bartfai R., Lim Z., Sreenivasan R., and Siegfried K.R., *et al.* (2012) PLoS One. 7(4):e34397. https://doi.org/10.1371/journal.pone.0034397

Piferrer F., Anastasiadi D., Valdivieso A., Sánchez-Baizán N., and Moraleda-Prados J., *et al.* (2019) Front Genet. 10:857. https://doi.org/10.3389/fgene.2019.00857

Ribas L., and Piferrer F. (2014) Rev Aquacult. 2014; 6(4): 209-240. https://doi.org/10.1111/raq.12041

Shao C., Li Q., Chen S., Zhang P., and Lian J., *et al.* (2014) Genome Res. 24(4):604-15. https://doi.org/10.1101/gr.162172.113

544. An approach to study the association between the blood cell methylome with feed efficiency traits

A. López-Catalina[1,2*], A. Bach[3], M. Gutiérrez-Rivas[1], R. Peiró-Pastor[1] and Ó. González-Recio[1,2]

[1]Departamento de Mejora Genética Animal, Instituto Nacional de Investigación y Tecnología Agraria y Alimentaria (INIA-CSIC), Crta. de la Coruña km 7.5, 28040 Madrid, Spain; [2]Departamento de Producción Agraria, Escuela Técnica Superior de Ingeniería Agronómica, Alimentaria Y de Biosistemas, Universidad Politécnica de Madrid, Ciudad Universitaria s/n, 28040 Madrid, Spain; [3]Marlex Research and Education and Institució Catalana de Recerca i Estudis Avançats (ICREA), Barcelona, Spain; alopez.catalina@inia.es

Abstract

Genotyping strategies enable a large volume of information that can be used in livestock breeding programmes. Here, we analysed whether nanopore sequencing shows equivalent results to whole-genome bisulfite sequencing and whether the methylation patterns of a cow affect or modify its phenotype. A depth of 10× with Nanopore sequencing showed similar results to bisulfite sequencing (Pearson correlation=0.753). More efficient cows had lower methylated sites along their genomes. Animals on their first lactation showed a larger number of methylated sites than those in subsequent lactations. Understanding the mechanisms that control the imprinting of methylation marks and their effect on productive traits of the animals can lead to an improvement of breeding programmes.

Introduction

The study of the epigenome can improve our understanding of gene expression, regulation, and changes on the functional genome during the prenatal reprogramming or during long stress periods. The most common and well known epigenetics mark are cytosine methylations which consist in the addition of a methyl group to the fifth carbon of a cytosine. These changes usually occur in the context of cytosine-guanine dinucleotides (CpG).

Epigenetic studies are shedding light on different modifications and new approaches to improve animal adaptation to the environment and resistance to stress situations. The association between the methylome and productive traits can affect the health, welfare, and performance of animals. Studies in pigs have found that differentially methylated regions are partially responsible for meat quality (Li *et al.*, 2012). An actual association of heat stress and methylation pattern has yet to be found in dairy cattle but has already been suggested in some studies (Hao *et al.*, 2016). The genetic merit of an animal is estimated using its own information and that from its relatives. Genetic evaluation procedures include environmental effects to isolate genetic information from other effects that are not inherited. Epigenetic marks can interfere with this information because marks such as DNA methylations can modify the expression of certain genes and alter the expected phenotype. Methylation marks can be understood as the way the environment and the genome interact and add useful information to the actual genetic evaluations.

The challenge with epigenetic information lies in the difficulty of obtaining it from commercial populations and the costs associated to the process. Nanopore sequencing offers a different approach than whole-genome bisulfite sequencing (WGBS), e.g. the genome can be sequenced without PCR amplification or chemical labelling of the samples (Clarke *et al.*, 2009). Investigating how this technology can complement or substitute bisulfite sequencing would open new possibilities to the study of the epigenetic marks of livestock animals at farm scale, reducing costs, and accelerating the analyses.

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_544, © A. López-Catalina *et al.* 2022

The objective of this study was to evaluate the performance of Oxford Nanopore Technologies (ONT) sequencing platforms to evaluate methylation in livestock and perform a preliminary association analysis between methyl marks and feed efficiency and number of lactation.

Materials & methods

Animal sampling. Blood samples were extracted from 22 cows via venipuncture and stored in a Vacutainer™ with EDTA. Data from the official milk recording scheme were available, in addition to dry matter intake and feed efficiency (FE). Animals with a FE equal or greater than 1.5 were classified in the high efficiency group, where as those with a FE<1.5 were grouped as low efficiency.

DNA extraction. We used the Monarch® HMW DNA Extraction Kit (T3050L) for Cells & Blood (New England Biolabs®) to extract DNA for nanopore sequencing. The DNA for WGBS was extracted from a blood sample using the MasterPure™ DNA Purification Kit (Epicentre). The concentration samples were estimated using a NanoDrop UV/Vis spectrometer (NanoDrop Technologies Inc.) and a Qubit fluorometer (ThermoFisher Scientific, 150 Waltham, MA, USA).

DNA sequencing. DNA was sequenced using the MinION sequencer (Oxford Nanopore Technologies. ONT, Oxford, UK). Sample for the comparison with Illumina sequencer was run in a single flow cell to achieve a greater coverage meanwhile samples for methylation analyses were multiplexed and sequenced at low depth. The library was prepared using the Ligation Sequencing Kit (SQK-LSK109) and following the protocol provided by its manufacturer. The samples were multiplexed and the resulting library was sequenced using two R9.4 flow cell until exhaustion. Additionally, the sample from a cow, Montross Martha, was sequenced individually, after fragmenting DNA to an average length of 8 kb, at larger depth using a single R9.4 flow cell until exhaustion. This sample was also sequenced by WGBS by an external service (CNAG-CRG. Centro Nacional de Análisis Genómico), using the Kit Enzyme Methyl (NEB). Basecalling was performed using Guppy version 5.0.14 (nanoporetech.com/community). Quality control was performed using NanoPlot (v1.36.2) (de Coster *et al.* 2018), and a minimum threshold for Q-score was set to 9. The quality control of the WGBS was performed using FastQC (v0.11.7) (Andrews, 2010)

Methylation calling. The pipelines were implemented in the CESGA super-computing centre (Galicia, Spain). First, Nanopolish (v0.13.2) (Loman *et al.*, 2015) was used to extract the methylated sites from the samples. Sequences from WGBS were analysed using Bismark (v0.23.0) (Krueger and Andrews, 2011)

Comparison between ONT and WGBS. We compared the sites detected by each software and the correlation between their methylation likelihoods was estimated. Several coverage thresholds (1×, 5×, 10×, 15×, and 20×) were evaluated to determine the effect of sequencing depth and the potential implications in the reliability of the results.

Differential methylation. PycoMeth (Leger, 2020) was used to detect the methylation patterns based on likelihood ratios. A filter for sites with a sequencing depth larger than 10x was established. Sites with a *P*-value higher than 0.05 were discarded. We compared the methylation pattern of the high FE animals to the one of low FE animals and animals from first lactation with animals from second or further lactations. GCF_002263795.1_ARS-UCD1.2_genomic.gff was used to annotate the differentially methylated genes.

Results

Comparison of WGBS and nanopore sequencing. A total of 88.2 Gb was obtained by WGBS (coverage ~30×). The mean Q-score was of 38, which is equivalent to an error rate <0.01%. Sequences obtained using the MinION resulted in 28.4 Gb, which is equivalent to a coverage of ~7×. Their mean Q-score achieved

was 9.9, corresponding to an average error rate of 10%. We detected 21,694,649 possible methylated sites with the MinION and Nanopolish and 52,403,895 using bisulfite sequencing and Bismark.

Similarity between WGBS and ONT. After filtering on different coverages (Figure 1) we found that selecting sites with a coverage >10× resulted in a correlation of 0.75 and 4,652 sites detected using both sequencing platforms.

Differential methylation. Samples were sequenced with a mean coverage of 0.5×. After filtering for coverage>10× and *P*-value<0.05 we obtained 25 genome intervals showing differential methylated status between the FE groups and 29 of them for lactation groups. We observed more methylated sites in animals in their first lactation and more methylated sites in low FE animals. Genes like *gnas*, *nelfcd*, *insl3* and *b3gnt3* were the most differentially methylated between FE groups and *amot*, *ago2*, *chrac1*, *sept9* in the lactation groups.

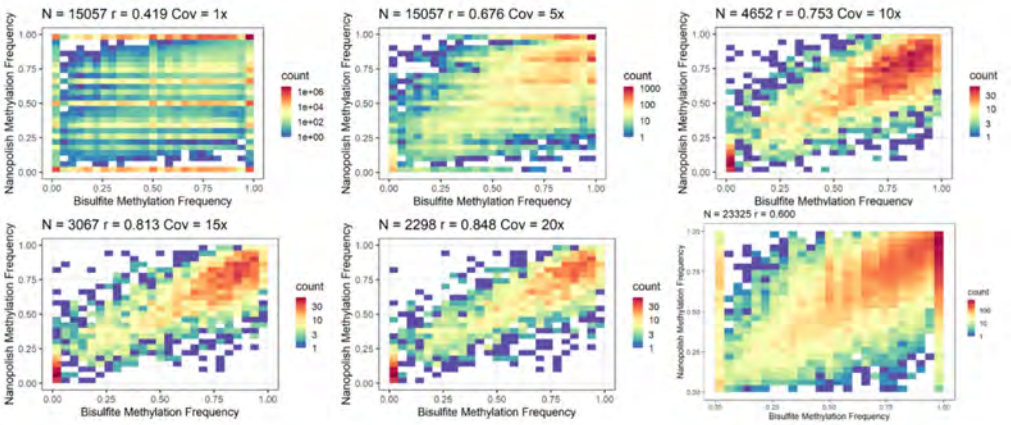


Figure 1. Different combinations of coverage and their respective correlations (r) and sites detected (N). Correlation filtered for 1×, 5×, 10×, 15×, 20× and last image illustrates filtering ONT sites by 15× and Illumina for 1×.

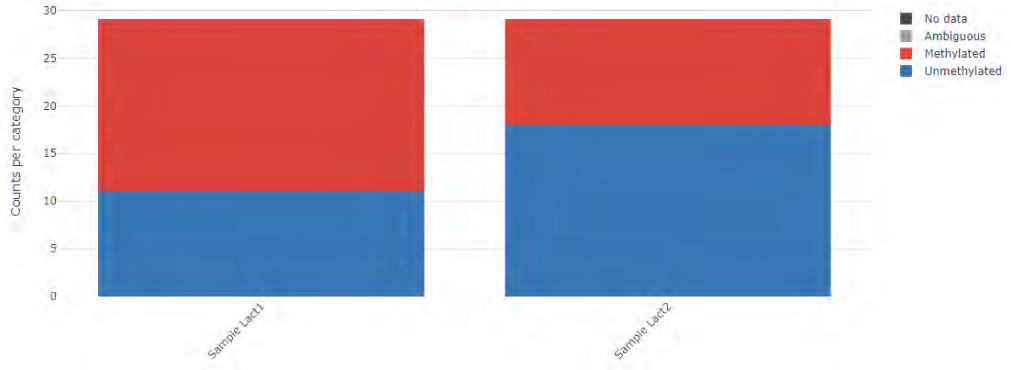


Figure 2. Comparison of methylation patterns alongside the genome of animals in their first lactation (LACT1) and animals in second or further lactations (LACT2). Red blocks represent methylated regions and blue ones represent unmethylated ones.

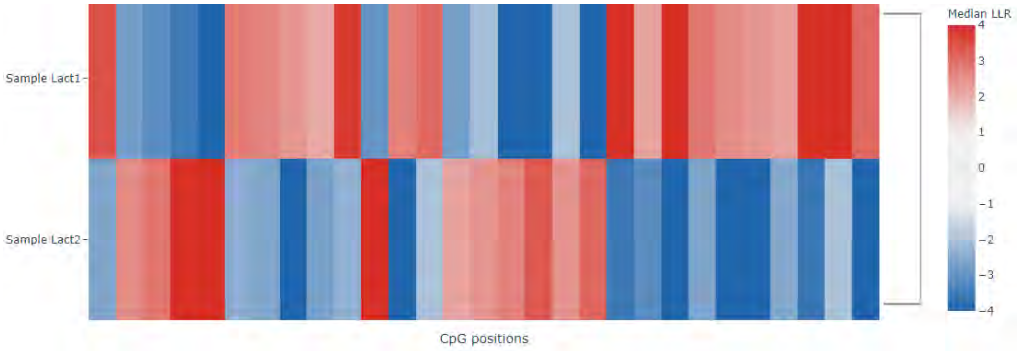


Figure 3. Comparison of methylation patterns alongside the genome of animals in their first lactation (LACT1) and animals in second or further lactations (LACT2). Red blocks represent methylated regions and blue ones represent unmethylated ones.

Discussion

Third-generation sequencing can help us to understand how the genome interacts with the environment and envision new opportunities to modify management practices and include epigenetic information in the breeding programmes. Nanopore sequencing enables new ways of studying epigenetics marks at farm level. Low cost and multiplexing capacity may allow implementing new technologies to improve animal wellbeing and adaptation to its environment. ONT offers the possibility to obtain genomic variants and genome methylation status simultaneously. We observed a good correlation between ONT and WGBS after filtering sites with a sequencing depth>10×. For future analysis, we suggest filtering the samples and keep sites with a coverage>10×.

Methylation patterns seem to clearly vary between different groups of animals. High feed efficient ones have less methylated sites than low FE animals. Preliminary selected genes like *gnas*, associated with performance traits (Sikora *et al.*, 2011), suggest that methylation marks could have an impact on the phenotype of the cow, by modulating gene expression and having and impact in the productive rates of the animal. Initially, we have observed that methylation levels are different between animals with different phenotypes (Lactation number and FE). It must be pointed out that our results were obtained from a reduced number of individuals and with a reduced genome coverage. Further studies may clarify how the methylation marks affect other traits of interest, and whether this information can increase the accuracy of genetic evaluations.

References

Clarke, J., Wu, H.C., Jayasinghe, L., Patel, A., Reid, S., *et al.* (2009). Nature Nanotechnology 4:4 4, 265–270. https://doi.org/10.1038/nnano.2009.12

Hao, Y., Cui, Y., Gu, X. (2016). Scientific Reports 2016 6:1 6, 1–13. https://doi.org/10.1038/srep27507.

Leger, A., 2020. a-slide/pycoMeth: v0.4.25. https://doi.org/10.5281/ZENODO.4110144

Loman, N.J., Quick, J., Simpson, J.T. (2015). Nature Methods 2015 12:8 12, 733–735. https://doi.org/10.1038/nmeth.3444

Sikora, K.M., Magee, D.A., Berkowicz, E.W., Berry, D.P., Howard, D.J., *et al.* (2011). BMC Genetics 12, 1–16. https://doi.org/10.1186/1471-2156-12-4/TABLES/3

545. Multi-dimensional functional annotation of bovine genome for the BovReg project

G.C.M. Moreira[1*], S. Dupont[1], D. Becker[2], M. Salavati[3], R. Clark[4], E.L. Clark[3], G. Plastow[5], C. Kühn[2,6] and C. Charlier[1] on behalf of the BovReg consortium

[1]Unit of Animal Genomics, GIGA Institute, University of Liège, 4000 Liège, Belgium; [2]Institute of Genome Biology, Research Institute for Farm Animal Biology (FBN), 18196 Dummerstorf, Germany; [3]The Roslin Institute, University of Edinburgh, EH25 9RG, Edinburgh, United Kingdom; [4]Genetics Core, Edinburgh Clinical Research Facility, The University of Edinburgh, EH4 2XU, Edinburgh, United Kingdom; [5]Livestock Gentec, Department of Agricultural, Food and Nutritional Science, University of Alberta, T6G 2R3, Edmonton, Canada; [6]Faculty of Agricultural and Environmental Sciences, University Rostock, 18059 Rostock, Germany; gcosta@uliege.be

Abstract

A multi-dimensional transcriptome map was generated using a diverse catalogue of tissue samples collected from six individuals of both sexes, different ages, kept in different environments and from three divergent breeds/crosses. More than 15k genes exhibit at least one potentially novel transcript in comparison with existing annotations. Tissue-specific miRNAs were detected contributing to the understanding of vital body functions in bovine. PIWI-interacting RNAs (piRNAs) were detected in ovary and testis only and, the repertoire of piRNA clusters in the male germline was expanded. A regulatory landscape was also generated including open chromatin regions, putative active promoters and transcription start sites, active enhancers, repressive states and insulators. This constitutes a key repository dataset for biology-driven genomic prediction that will help the scientific community and industry stakeholders to apply genomics to address the challenges facing cattle production.

Introduction

In humans, a catalogue of functional elements in the genome was generated by the ENCODE consortium (ENCODE Project Consortium, 2012; Davis *et al.,* 2018) to link genotype to phenotype. In bovine, a comprehensive catalogue of functional elements still does not exist despite the huge contribution of bovids to global food production (Adesogan *et al.,* 2020). Recent efforts from the FAANG consortium have contributed to the functional annotation of the bovine genome but, with a limited number of tissues, individuals and populations analyzed (Foissac *et al.*, 2019; Kern *et al.,* 2021). Multi-dimensional functional annotation of the bovine genome built from a diverse catalogue of tissue samples collected from individuals of both sexes, different ages, kept in different environments and from divergent breeds/crosses is critical in linking genotype to phenotype in cattle, to further improve biology-driven genomic prediction models.

Materials & methods

Samples and tissues. A diverse catalogue of 129 tissue samples were collected from six individuals of both sexes, different ages, kept in different environments and from three divergent breeds/crosses: Belgian dairy (Holstein), Canadian beef composite (Kinsella) and German beef/dairy cross (Charolais × Holstein).

RNA isolation and library preparation. For transcriptome map generation, Total RNA was extracted using miRNeasy kit (QIAGEN) from the 129 snap-frozen tissues samples. After quality control (quantity, purity and integrity assessment), Total RNA libraries were generated using TruSeq Stranded Total RNA Library Prep Gold (Illumina); mRNA libraries using TruSeq Stranded mRNA Library Prep (Illumina); and small RNA libraries using QIAseq miRNA Library (QIAGEN).

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_545, © G.C.M. Moreira *et al.* 2022

For regulatory landscape generation, ATACseq libraries were generated using the ATACseq kit (Active Motif). The chromatin immunoprecipitations (ChIP) were performed for five antibodies (H3K4me3, H3K4me1, H3K27me3, H3K27ac and CTCF – all from Diagenode) using the iDeal ChIP-seq kit for TF (Diagenode); and ChIPseq libraries were prepared using the NEBNext Ultra DNA library prep kit for Illumina libraries (New England Biolabs).

Sequencing and analyses. Libraries were sequenced on the Illumina NovaSeq 6000 (150nt 25M paired-end – mRNA; 150nt 40M paired-end – Total RNA; 100nt/50nt 25M single-end – small-RNA; 150nt 50M paired-end – ATACseq; 100nt paired-end – ChIPseq 30M reads for narrow marks and CTCF; 35M reads for broad marks).

Data analyses for mRNA and Total RNA assays were performed using the BovReg pipeline (https://github.com/BovReg/rnaseq) and for small RNA, ATACseq and ChIPseq assays, using Nextflow/nf-core pipelines (smrnaseq-1.0.0, atacseq-1.2.1, chipseq-1.2.2). The ARS-UCD1.2_Btau5Y reference file from 1000bulls project was utilized. The GTF file from Ensembl v.102 with the coordinates lifted over to the 1000bulls reference genome was used.

For piRNA analyses, small-RNA were mapped to the genome and after the selection of reads ranging between 18-35nt, reads that mapped to rRNAs, miRNAs, snoRNAs, snRNAs, and tRNAs were removed. PiRNA clusters were detected using proTRAC software (Rosenkranz and Zischler, 2012) with default parameters.

Results

Transcriptome map. A *de novo* transcriptome assembly from mRNA and Total RNA assays was created. 43,117 genes were assembled and from those, more than 15K genes exhibit at least one potentially novel transcript in comparison with existing annotations (Table 1).

Based on mRNA assays, tissues originating from the central nervous system exhibit a greater number of genes with elevated expression compared to the other tissues (Figure 1A). Charolais × Holstein/adult testis exhibited four-fold more genes with elevated expression than Holstein/neonate testis and most of those are tissue enriched (elevated expression in testis compared to any other tissue), with GO terms enriched for spermatogenesis related processes known to occur in post-puberal testis (Figure 1A). When considering only tissues shared between the six animals, Holstein/neonate animals exhibited a greater number of genes expressed/assembled and also a greater number of genes with at least one potentially novel transcript compared to Kinsella crossbred/juvenile and Charolais × Holstein.

Table 1. Results from the comparison between BovReg transcriptome assembly and existing annotations.

BovReg annotation features	Compared with	
	Ensembl v.104[1]	NCBI v.106
Genes with at least one potentially novel transcript assembled	15,954	18,636
Complete transcripts with exact match of intron chain	37,257	34,430
Transcripts with partial match, encompassing or within reference transcripts	211,616	223,926
Novel transcripts – no match with reference (unknown, intergenic)	28,925	19,442

[1] To compare the *de novo* transcriptome assembly from BovReg with existing annotations, Ensembl v.104 annotation was utilized instead of v.102.

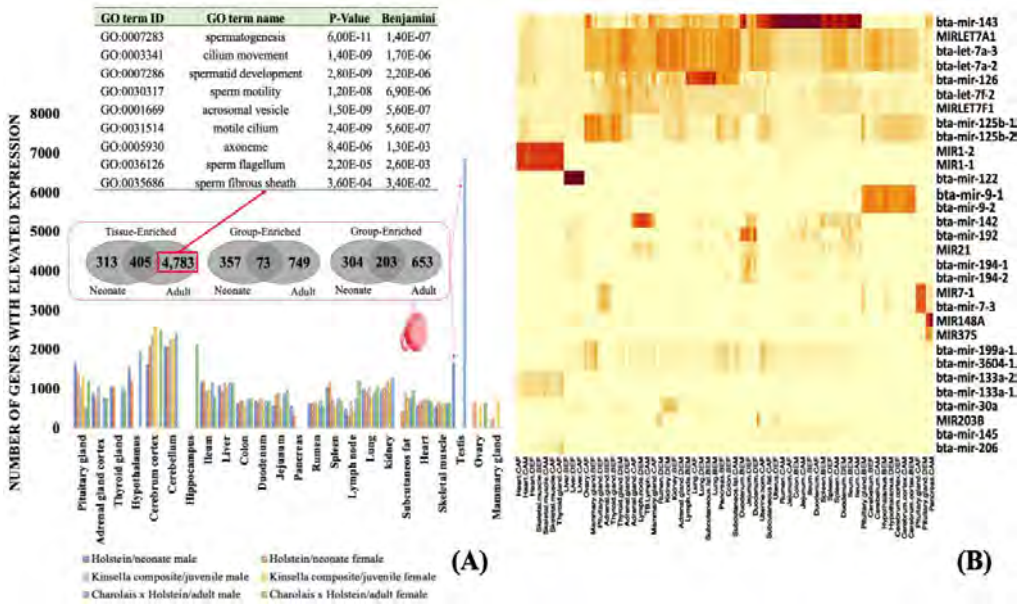


Figure 1. Overview of mRNA and miRNA specificity profiles in the bovine transcriptome (A) Genes with elevated expression (normalized by TPM) in each tissue; classification and GO enrichment of genes with elevated expression in testis. Elevated expression encompass three subcategory types (Jain and Tuteja, 2019): tissue enriched (at least 4× higher mRNA level compared to any other tissue); tissue enhanced (at least 4× higher average mRNA level in a group of 2-5 tissues compared to any other tissue) and; group enriched (at least 4x higher mRNA level compared to the average level in all other tissues). (B) Heatmap of the top 5 known miRNAs with the highest expression (normalized by CPM) in each tissue. MiRNAs in y-axis and tissues on x-axis; BEM: Belgian animal male; BEF: Belgian animal female; CAM: Canadian animal male; CAF: Canadian animal female; DEM: German animal male; DEF: German animal female.

In the small RNA dataset, miRNAs were the predominant class detected (with 81.1% of the reads corresponding to miRNAs, 0.4% to rRNA, 0.3% to tRNA, 3.02% to snoRNA, 0.93% to snRNA and 14.25% to either mitosRNA, unknown – including piRNAs – or artifacts). Taking the top five known miRNAs with the highest expression in each tissue, tissue-specific miRNAs were confirmed (Figure 1B) and most of these miRNAs are already known to play important roles in key biological processes. As an example, *Bta-mir-206* is highly expressed in skeletal muscle and is involved in muscular hypertrophy in Texel sheep (Clop *et al.*, 2006).

PIWI-interacting RNAs (piRNAs) were only detected in ovary and testis. PiRNAs are known to silence transposons and regulate gene expression (Czech and Hannon, 2016). In testis they represent up to 95% of small-RNAs in adult and around 30% in neonate. A huge proportion of piRNAs, especially those expressed in late spermatogenesis, are produced from large genomic regions known as piRNA clusters (Aravin *et al.*, 2007). A greater number of piRNA clusters were detected on neonate testis compared to adult testis and also, from what is known from the database (https://www.smallrnagroup.uni-mainz.de/piRNAclusterDB/), expanding the repertoire of piRNA clusters in bovine.

The data presented here will be combined with a new map of transcription start sites (TSS) (CAGE), lncRNA and circRNA, which will be generated in collaboration with other partners in the BovReg consortium, to create a multi-dimensional transcriptome map.

Regulatory regions. Preliminary ATACseq data revealed approximately 666K open chromatin regions in the bovine genome. From peak annotation relative to gene features, using the *de novo* transcriptome assembly ~4-fold more peaks were annotated in exons, ~3-fold more peaks annotated in promoter-TSS and in transcription termination sites (TTS) and ~2-fold less annotated in intergenic regions, compared with the annotation obtained using Ensembl v.102 assembly.

From preliminary ChIPseq data, 230,923 putative active promoters and transcription start sites (TSS) states (H3K4me3), 627,454 active enhancer states (H3K27ac and H3K4me1), 659,150 repressive states (H3K27me3) and 409,062 insulators bound by CTCF were detected.

Discussion

Despite a greater number of genes harboring potentially novel transcripts that are expressed in purebred/neonate compared to composite breeds/juvenile and adult animals, it is difficult to claim if the greater number is driven by age and/or breed. However, our findings demonstrate the importance of having different ages, breeds/crosses and environments to build a comprehensive functional annotation of bovine genome. Our investigation of gene expression regulation by miRNAs, especially the novel ones, provides important biological knowledge to better understand bovine physiology.

The repertoire of miRNAs, piRNAs and piRNA clusters in the bovine germline was expanded. There is a lack of information about piRNAs in the bovine germline and their potential impact on gene expression regulation, which emphasises the importance of the new annotation information provided by BovReg.

The data presented here already represents a substantial improvement on the functional annotation of the bovine genome. An integrative approach with the *de novo* transcriptome map, chromatin states as well as TSS catalogue will provide additional evidence to support the novel genes, miRNAs and piRNA clusters detected.

The BovReg project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 815668.

References

Adesogan A.T., Havelaar A.H., McKune S.L., Eilittä M. and Dahl G.E. (2020) Glob Food Secur 25:100325. https://doi.org/10.1016/j.gfs.2019.100325

Aravin A.A., Hannon G.J. and Brennecke J. (2007) Science 318(5851):761-764. https://doi.org/10.1126/science.1146484

Clop A., Marcq F., Takeda H., Pirottin, D., Tordoir, X. *et al.* (2006). Nat Genet 38(7):813-818. https://doi.org/10.1038/ng1810

Czech B. and Hannon G.J. (2016) Trends Biochem Sci 41(4):324-337. https://doi.org/10.1016/j.tibs.2015.12.008

Davis C.A., Hitz B.C., Sloan C.A., Chan E.T., Davidson J.M. *et al.* (2018) Nucleic Acids Res 46(D1):D794-D801. https://doi.org/10.1093/nar/gkx1081

ENCODE Project Consortium (2012) Nature 489(7414):57-74. https://doi.org/10.1038/nature11247

Foissac S., Djebali S., Munyard K., Vialaneix N., Rau A. *et al.* (2019) BMC Biol 17(1):108. https://doi.org/10.1186/s12915-019-0726-5

Jain A. and Tuteja G. (2019) Bioinformatics 35(11):1966-1967. https://doi.org/10.1093/bioinformatics/bty890

Kern C., Wang Y., Xu X., Pan Z., Halstead M. *et al.* (2021) Nat Commun 1821. https://doi.org/10.1038/s41467-021-22100-8

Rosenkranz D. and Zischler H. (2012) BMC Bioinform 13 (5). https://doi.org/10.1186/1471-2105-13-5

546. European project AQUA-FAANG: the epigenetic landscape of the Atlantic Salmon; focus on liver tissue

M.-O. Baudement*, G.B. Gillard, T. Podgorniak, L. Grønvold, S. Lien and M.P. Kent

Centre for Integrative Genetics (CIGENE), Faculty of Biosciences, Norwegian University of Life Sciences, 1430 Ås, Norway; marie-odile.baudement@nmbu.no

Abstract

As part of the European project AQUA-FAANG, we produced a body map of the epigenetic and regulatory landscape of Atlantic salmon, spanning different tissues in both sexes and at two stages of maturity. For this purpose, we developed and applied refined protocols for Chromatin ImmunoPrecipitation assays (ChIP-seq) and Transposase-Accessible Chromatin assays (ATAC-seq), complemented with RNA sequencing (RNA-seq). Our design of combining epigenetic and gene expression data sourced from the same individuals provides detailed insights into genome regulatory mechanisms. We showcase results here by comparing regulatory differences between liver of both mature and immature male and female salmon. We identify that differential expression of genes is associated with chromatin histone modifications and/or accessibility dichotomy between sex or maturity.

Introduction

The European project AQUA-FAANG aims to describe genome regulation of the most commercially important fish species in European aquaculture with refined protocols, which in turn will promote a better understanding of the biology of these organisms to improve fish production, welfare and commercial product quality (description of the project available at: https://www.aqua-faang.eu).

As part of this project, we will perform an epigenetic and regulatory analysis across different tissues, sexes and maturity states of the Atlantic salmon. We sampled five tissues (liver, brain, gonad, muscle, gill) from three mature and three immature male and female individuals (12 fish total, 60 samples) to analyse histone modifications by ChIP-seq (H3K4me1, H3K4me3, H3K27ac, and H3K27me3), chromatin accessibility by ATAC-seq and RNA-seq for gene expression. We describe here our results from liver tissue as a proof of concept. We find examples of highly expressed genes (*GYS2*), gender specific expression (*lipocalin*) as well as differentially expressed genes related to maturity (*EVLA*, *nAChRα4*). Together, these results corroborate the high quality of the genomic information obtained through our protocols and analysis, and future ability to describe the epigenetic landscape.

Materials & methods

Dissection of Atlantic salmon. Fish were euthanised by an overdose of tricaine methanesulfonate and sexed. Tissues were dissected, snap-frozen and conserved at -80 °C.

Chromatin ImmunoPrecipitation assays. Tissues were dissociated by dounce homogenization and crosslinked with 1% formaldehyde. Glycine was added and material was washed and recovered by centrifugation. The nuclei were resuspended with sonication buffer containing 1% Sodium Dodecyl Sulfate. Sonicated material was purified by proteinase K treatment and columns, quantified by Qubit and fragmentation was assessed using a Bioanalyzer. Tubes containing fragments with expected size (200-700 base pairs; 350 median) was diluted and used for immunoprecipitation. Detailed protocol available at: NMBU_SOP_ChIP_protocol

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_546, © M.-O. Baudement *et al.* 2022

Transposase-accessible chromatin assays. Tissues were dissociated by dounce homogeneization. Nuclei were recovered, exempt from mitochondria contamination, after slow centrifugation in iodixanol gradient and washing step. 50,000 purified nuclei were resuspended with reaction mix containing transposase (100 nM final). Transposed DNA was then cleaned-up, PCR-amplified and assessed using a Bioanalyzer. Detailed protocol available at: NMBU_SOP_ATAC_protocol

Gene expression assays. PolyA RNA was isolated and purified with RNAeasy kit from pieces of tissues conserved in RNAlater buffer. RNA quality was assessed using a Bioanalyzer and Nanodrop. Detailed protocol available at: NMBU_SOP_RNAextraction_protocol

Library preparation and library sequencing. ChIP libraries were prepared with the Microplex v3 kit from Diagenode. ATAC libraries were prepared with NEBNext Ultra II Q5 Mater mix of Illumina. PolyA RNA libraries were prepared with NEBNext Ultra directional RNA Library Preparation kit for Illumina. All libraries were sequenced with an Illumina NovaSeq 6000 (150 base pair, paired end reads).

Library alignment, quality assessment, peak-calling, differential analysis, and repository. Sequencing data was processed through relevant nf-core pipelines (available at: https://nf-co.re); chipseq, atacseq or rnaseq, to produce results for data quality, read alignments, peak annotations, and gene expression analysis. Libraries of satisfactory quality were submitted to ENA (project PRJEB47410). The latest Ensembl rapid release genome of the Atlantic Salmon (v3) was used for alignment and gene annotations (available at: Salmo_salar_GCA_905237065.2)

Results

Assessment of epigenetic data generated. In this project we have generated and sequenced more than 300 libraries for Atlantic salmon (ChIP, ATAC and RNA). We showcase here a subset of results obtained for liver. To assess if our adaptation of ChIP and ATAC methods for this fish species was reliable (Table 1). We evaluated the similarity of the different ChIP libraries (H3K4me3 and H3K27ac peaks) generated by calculating the Jaccard index (Figure 1A). We saw as expected a progressive reduction of correlation from replicates with the highest correlation to histone modification showing independency of signals, with intermediate correlation by sex or maturity. We saw a large proportion of ChIP peaks overlapping with genomic features as expected (Figure 1B), with a concentration surrounding gene promoters and the transcription start sites (TSS). We saw relatively less concentration of ATAC peaks within promoters.

Table 1. Metrics about ChIP-seq data generated for liver tissue.

Type of library	Number of libraries[1]	Average number of read pairs (millions)		Number of peaks detected (thousand)[2]
		sequenced	aligned	
ChIP-H3K4me3	8	57 (8 SD)	31 (5 SD)	29 (3 SD)
ChIP-H3K27ac	8	56 (7 SD)	30 (5 SD)	27 (10 SD)
ATAC	12	49 (8 SD)	24 (4 SD)	146 (25 SD)
RNA	12	32 (2 SD)	30 (2 SD)	-

[1] Two sexes × two maturity stages × two replicates for ChIP, or three replicates for ATAC and RNA.

[2] MACS2 narrow peak detection for H3K4me3 and H3K27ac, broad peak for ATAC.

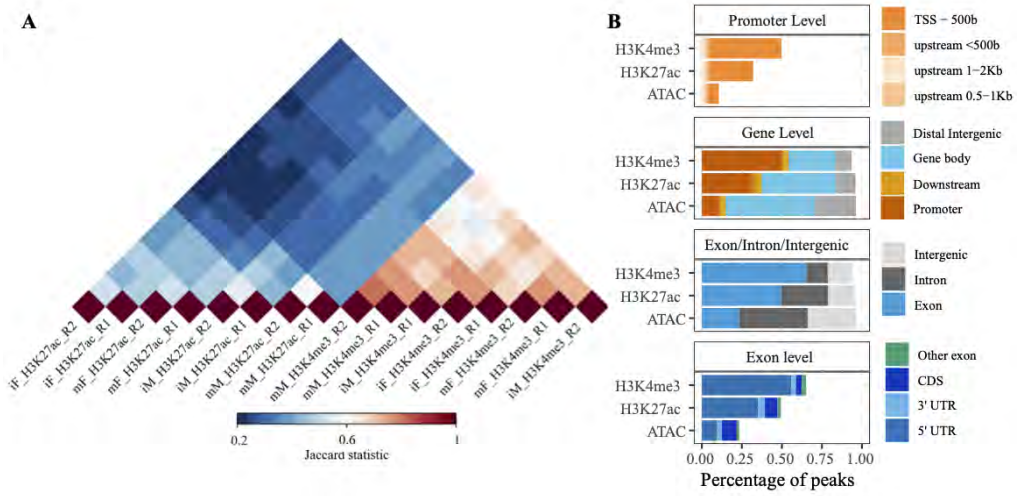


Figure 1. Assessment of peaks detected by ChIP and ATAC in liver. Similarity of peak datasets based on Jaccard statistics for two ChIP marks, H3K27ac and H3K4me3, across different liver libraries: immature female (iF), mature female (mF), immature male (iM), mature male (mM), and replicates (R1 and R2). B. Percentages of ChIP and ATAC peaks overlapping different genomic features.

Sex and maturity specific gene expression relate with chromatin environment. Differentially expressed genes (DEGs) between sex and/or maturity were identified (Figure 2). The chromatin state surrounding these DEGs was assessed by viewing H3K4me3, H3K27ac and ATAC peaks as tracks in a genome browser, alongside RNA expression levels. The gene *GYS2*, coding for a glycogen synthase, was found highly expressed in the liver of all individuals, in agreement with earlier studies in Masu salmon [Furukawa *et al.*, 2018] A gene coding for a lipocalin-like protein, suspected to be involved in sex(male) specificity in Tilapia [Shirak *et al.*, 2008] was found to be specifically expressed in males in our data. The gene *EVLA*, coding for Ena/Vasp protein is reported here has being expressed only in mature liver (independent of gender). EVLA has already been described to be involved in cytoskeleton remodelling (Ena) [Sechi and Wehland, 2004] and gluconeogenesis (vasodilator-stimulated phosphoprotein Vasp) (Tateya *et al.*, 2019), but a possible maturity-specific expression of this gene has not been reported previously. Lastly, a gene coding for a cholinergic receptor, nAChRα4, is presented here as being expressed specifically in immature individuals. Together, this first data exploration shows the high potential of better characterisation and new discoveries of genes potentially important for aquaculture.

Conclusions

We showcased here a first look at linking gene expression with the epigenetic landscape. Deeper exploration of this dataset will help to establish a better understanding of gene regulation of the salmonid genome as well as the discovery of new traits crucial for the sustainability of the fish industry for the coming years.

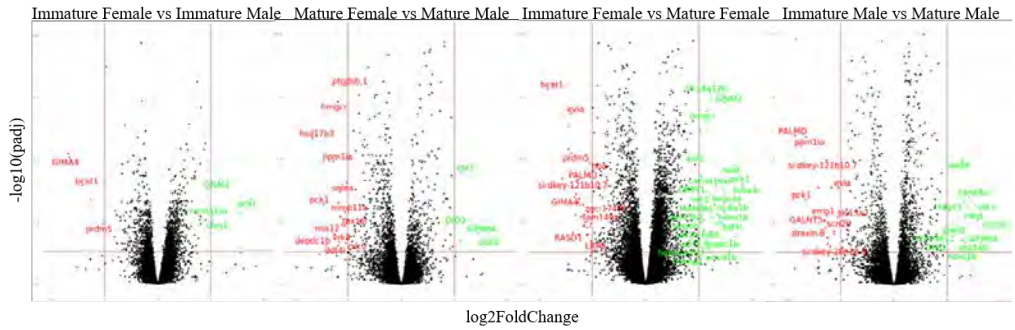


Figure 2. Differentially expressed genes in liver. Volcano plot highlighting genes highly differentially expressed between maturities or sexes in liver. Down-regulated (in red) and up-regulated genes (in green) between conditions are highlighted (FDR adjusted *P*-value<0.5 and absolute log2 fold change >5). For an overview, we display results within x and y limits up to an absolute value of 10.

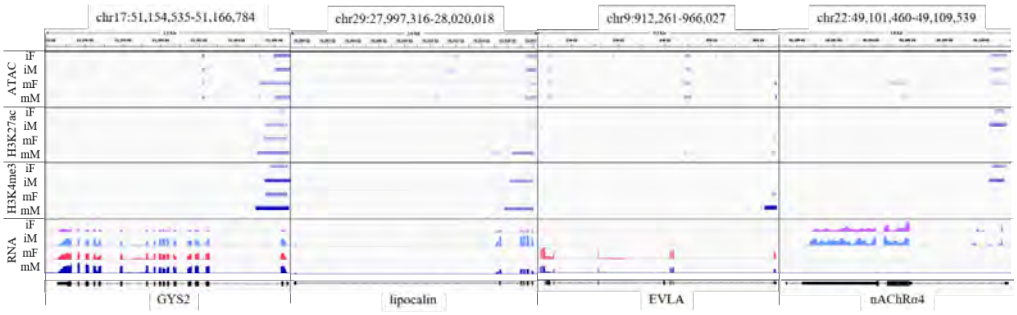


Figure 3. Epigenetic patterns of differentially expressed genes. Peaks from ChIP (H3K4me3, H3K27ac) and ATAC data are observed alongside RNA expression levels for select genes expression specificity: *GYS2* (no specificity), *lipocalin* (male-specific), *EVLA* (mature-specific) and *nAChRα4* (immature-specific).

References

Buenrostro J., Wu B., Chang H., and Greenleaf W. (2016) Curr Protoc Mol Biol 109:21.29.1–21.29.9. https://doi.org/10.1002/0471142727.mb2129s109

Chawla A., Nagy C., and Turecki G. (2021) IJMS 22(14) 7612. https://doi.org/10.3390/ijms22147612

Furukawa F., Irachi S., Koyama M., Baba O., *et al.* (2018) CBPA 225:74–82. https://doi.org/10.1016/j.cbpa.2018.07.003

Ma S. and Zhang Y. (2020) Mol Biomedicine 1(1) 9. https://doi.org/10.1186/s43556-020-00009-w

Sechi A. S. and Wehland J. (2004) Frontiers in Bioscience-Landmark 9(2) 1294-1310. https://doi.org/10.2741/1324

Shirak A., Golik M., Lee B-Y., Howe A. E., *et al.* (2008) Heredity 101:405–415. https://doi.org/10.1038/hdy.2008.68

Tateya S., Rizzo-De Leon N., Cheng A. M., Dick B. P., *et al.* (2019) PloS ONE 14(4):e0215601. https://doi.org/10.1371/journal.pone.0215601

547. Integration of methylome and transcriptome profiles reveals candidate biomarkers for *S. uberis*-subclinical mastitis

M.Q. Wang[1,2], M. Laterrière[3], P.L. Dudemaine[1], N. Bissonnette[1], D. Gagné[3], X. Zhao[4], J.P. Roy[5], M.A. Sirard[2] and E.M. Ibeagha-Awemu[1*]

[1]Sherbrooke Research and Development Centre, Agriculture and Agri-Food Canada (AAFC), Sherbrooke, Quebec, Canada; [2]Department of Animal Science, Laval University, Quebec, Quebec, Canada; [3]Quebec Research and Development Centre, AAFC, Quebec, Quebec, Canada; [4]Department of Animal Science, McGill University, Ste-Anne-De-Bellevue, Quebec, Canada; [5]Department of Clinical Sciences, University of Montreal, Ste Hyacinthe, Quebec, Canada; eveline.ibeagha-awemu@agr.gc.ca

Abstract

To investigate the roles of DNA methylation and gene expression during *Streptococcus uberis*-induced subclinical mastitis, the methylome and transcriptome profiles of milk somatic cells (MSC) from *S. uberis*-positive and control cows (n=3/group) were obtained by next-generation sequencing. Integration analysis between methylome and transcriptome profiles by MethGET revealed clear changes in the methylation levels of 585 promoters (pMeDEGs) and gene bodies of 424 genes (gbMeDEGs) in conjunction with gene expression changes (*P*-value<0.0005). The genes were significantly enriched in immune-related functions (antigen processing and cytokine activities). Applying the DIABLO method to pMeDEGs, the methylation status of 25 promoters and the expression of 10 genes were identified as candidate biomarkers as they were highly correlated and able to discriminate sick from healthy cows. Therefore, the integration analysis between the methylome and transcriptome profiles revealed important roles of pMeDEGs and gbMeDEGs in the host immune response to *S. uberis* infection, and the identified discriminant biomarkers could contribute to strategies for controlling *S. uberis-induced* mastitis.

Introduction

Subclinical mastitis is considered as one of the main challenges of the dairy industry. It is caused by different pathogens including *Streptococcus uberis*. Several investigations in the past decades have reported on the molecular mechanisms, and the genetic and epigenetic aspects of sub-clinical mastitis with aim to identify reliable biomarkers. DNA methylation, the most studied epigenetic mechanism, has been reported to play important regulatory roles in mammary gland health, including the development of mastitis due to *Staphyloccocus aureus* and *Escherichia coli* (Wang and Ibeagha-Awemu, 2020). However, the regulatory roles of DNA methylation during subclinical mastitis caused by *S. uberis* are not known. Therefore, this study profiled the global methylome and transcriptome of milk somatic cells from *S. uberis*-positive and control cows. Integration analysis of the methylome and transcriptome profiles was used to identify the methylome alterations that possibly caused changes in gene expression (transcriptome), their possible regulatory roles and candidate biomarkers that discriminate *S. uberis*-positive cows from healthy controls.

Materials & methods

Animals and samples. Two commercial farms (Quebec, Canada) were selected for the study. Cows with consistently very high (>350,000 cells/ml) or very low (<100,000 cell/ml) milk somatic cell counts (SCC) for three consecutive months were tested for the presence of mastitis pathogens. About 5 ml milk was collected per quarter of all selected cows and subjected to bacteriology (https://www.biovet-inc.com/). Cows positive for *S. uberis* only or negative for all mastitis pathogens were further selected. 200 ml milk was sampled from one quarter positive for *S. uberis* from each of three cows to constitute the test group. Similarly, 200 ml milk (composite of four quarters) was sampled from each of three cows to constitute the control group.

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_547, © M.Q. Wang *et al.* 2022

MSC were isolated immediately after milk collection by low speed centrifuge (1,500×*g* for 15 min at 4 °C), divided into two portions (TRIzol reagent added to the portion for RNA isolation) and stored at -80 °C.

DNA and RNA isolation, sequencing and bioinformatics analysis. Genomic DNA and total RNA were isolated from MSC by DNeasy Blood and Tissue Kit (Qiagen Inc., Toronto, ON, Canada) and RNeasy Mini Kit (Qiagen Inc., Toronto, ON, Canada), respectively. Purified DNA and RNA were subjected to whole-genome bisulfite sequencing (WGBS) and RNA-sequencing (RNA-Seq), respectively. Standard bioinformatics pipelines (methylseq and rnaseq) from nf-core (https://nf-co.re/methylseq/) were used to process the raw WGBS and RNA-Seq data. CpG sites or genes with at least 7 reads coverage were further analysed. MethGET software was used to correlate the methylome and transcriptome profiles (Teng *et al.*, 2020). Briefly, the DNA methylation status of the promoter (2 kb region upstream of the transcript start site) and gene body (exons and introns) of each gene were calculated per sample by averaging the methylation levels of all the CpG sites located in the corresponding regions. Then, the Gaussian Mixture Model (GMM) was incorporated to identify differential genes with DNA methylation alterations in their promoter regions (pMeDEG) or gene bodies (gbMeDEG) (P-value<0.0005 in GMM). Functional enrichment of methylated DEGs was done with ClusterProfiler (Wu *et al.*, 2021) and FDR<0.05 was considered as significant. Promoter methylation and gene expression values of pMeDEGs were then submitted to core DIABLO method from MixOmics R package to detect candidate biomarkers able to discriminate *S. uberis*-positive and control groups (Rohart *et al.*, 2017).

Results

Methylated differentially expressed genes and functional enrichment. A total of 585 pMeDEGs and 424 gbMeDEGs were identified (P<0.0005) by correlating gene expression and methylation status at the promoter (Figure 1A) or gene body, respectively. Merging pMeDEGs and gbMeDEGs revealed a total of 804 differentially expressed genes (MeDEGs), including 205 genes identified as both pMeDEGs and gbMeDEGs. Functional enrichment results were similar between pMeDEGs (2 GO and 14 KEGG pathways) and gbMeDEGs (1 GO and 20 KEGG pathways), with 11 commonly enriched KEGG pathways. However, functional annotations related to antigen processing (GO:0048002, GO:0019882 and bta04612) were only enriched by pMeDEGs. Meanwhile, more pathways with immune-related and disease functions were enriched by gbMeDEGs. Functional enrichment of MeDEGs indicated four biological processes GO terms (Figure 1B) and 28 KEGG pathways (Figure 1C), mostly having immune and disease related functions (Figure 2A,B).

Candidate biomarkers of S. uberis-induced mastitis. The application of DIABLO method to the promoter methylation and gene expression levels of pMeDEGs identified the methylation status of 25 promoters and the expression level of 10 genes as candidate biomarkers (Figure 2A). The *S. uberis*-positive and control groups were placed into two separate clusters, each with a corresponding set of candidate biomarkers (Figure 2B). Interestingly, both the promoter methylation and gene expression of *CCDC68* and *LOC112449619* were identified as candidate biomarkers. *CCDC68* was the second and third ranked promoter and gene biomarker, respectively. It should be noted that *FABP4* was enriched in inflammatory response (GO:0006954) GO term. Besides, *KLRC1,* a gene biomarker was enriched in three immune-related pathways, including antigen processing and presentation (bta04650), graft-versus-host disease(bta05332) and natural killer cell mediated cytotoxicity (bta04650).

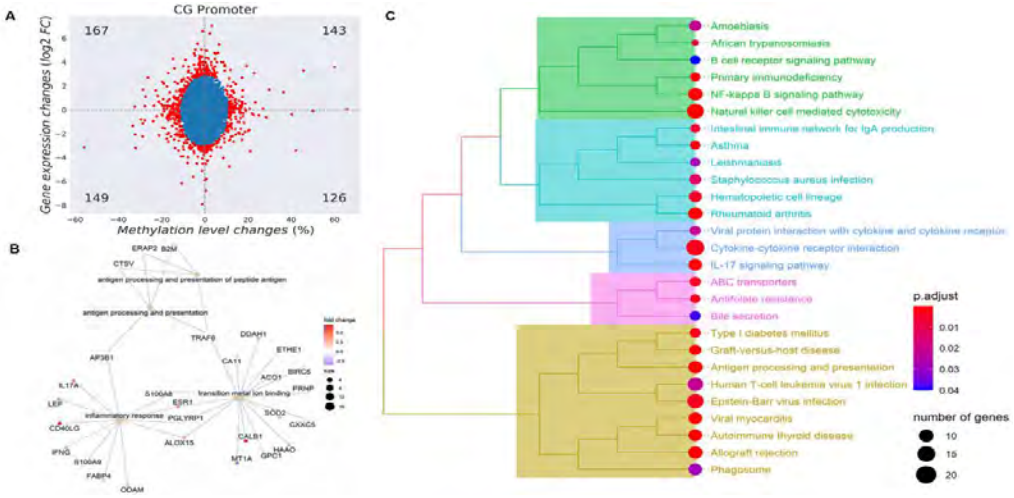


Figure 1. pMeDEGs and functional enrichment of MeDEGs. (A) Scatter plot showing changes in promoter DNA methylation and gene expression of pMeDEGs. (B) Enriched GO terms for MeDEGs showing interactions between genes. (C) Tree plot showing clusters of KEGG pathways enriched by MeDEGs.

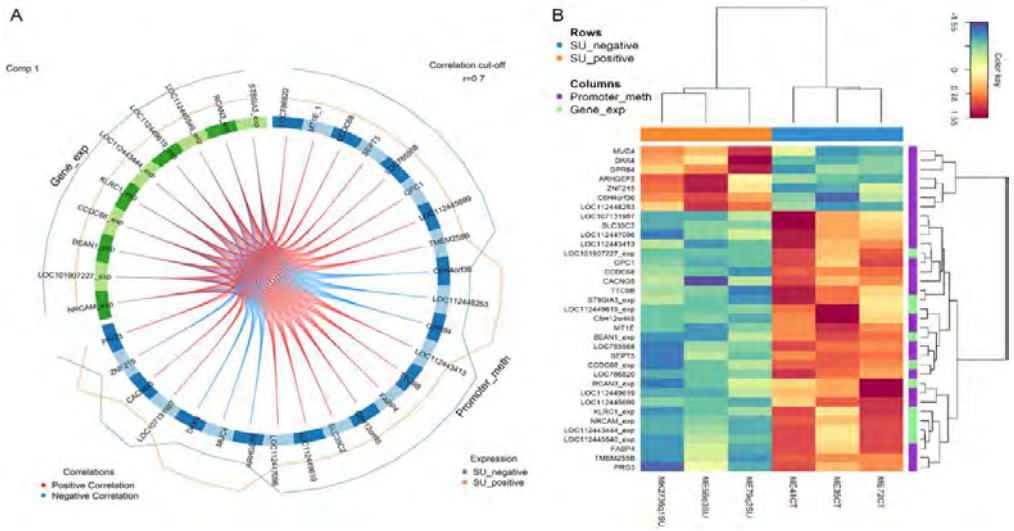


Figure 2. Candidate biomarkers distinguishing cows with *S. uberis*-induced subclinical mastitis from healthy controls. (A) The correlation between candidate biomarkers including the methylation status of 25 promoters (Blue) and the expression level of 10 genes (Green). (B) Plot showing two candidate biomarker clusters. Gene names with the extension '_exp' are gene biomarkers showing expression levels in corresponding rows, while others are promoter biomarkers showing promoter methylation status.

Discussion

Integration analysis between the global methylome and transcriptome profiles of milk somatic cells from *S. uberis*-positive and control cows revealed a total of 804 MeDEGs between the two groups. These MeDEGs with clear DNA methylation alterations in either promoters or gene bodies with associated gene expression changes gives clarity on the effects of DNA methylation changes on corresponding genes compared with consideration of DNA methylation alterations only. The functional enrichment of MeDEGs revealed their potential involvement in immune processes (Figure 1B,C), such as antigen processes, cytokine activities, and singling pathways of immune cells (B cell, T cell and natural killer cells), thereby highlighting the potential regulatory roles of DNA methylation alterations in the host immune response to *S. uberis*-induced mastitis. It also revealed the enrichment of key MeDEGs such as *IFNG*, *IL17A*, *CXCL2, etc.,* in multiple immune related pathways.

DNA methylation biomarkers intermediate environmental factors (exposome) and host genetics, and brings additional layer of information for exploring the unexplained portion of phenotypic variance in livestock traits (Ibeagha-Awemu and Yu, 2021). Compared to most studies which identify biomarkers independently from various types of data, this study incorporated the DIABLO discriminant analysis method to integrate the methylome and transcriptome datasets of the same samples and to identify highly correlated candidate biomarkers, including the methylation status of 25 promoters and the expression levels of 10 genes. The promoter methylation status and gene expression level of *CCDC68* and *LOC112449619* genes were both identified as candidate biomarkers. Furthermore, the promoter methylation status of CCDC68 was highly and significantly correlated with its gene expression level. *CCDC68* proteins are known to participate in a variety of regulatory functions associated with the development and evolution of various malignant tumors (Li *et al.*, 2021), suggesting potential roles in immune processes during *S. uberis* mastitis. To the best of our knowledge, this is the first association of methylated *CCDC68* gene with bovine mastitis, and deserves further investigation. Further candidate biomarkers (*FABP4*, *GPC1*, *KLRC1* and *NRCAM*) enriched in immune activities, emphasize their potential involvement in the host response to *S. uberis* infection. Therefore, the identified candidate discriminant biomarkers could support the development of management programs for mastitis control or breeding for mastitis resistance. However, confirmation of our results in a bigger population is necessary before application of the reported biomarkers for mastitis management due to the limited sample size of the current study and the possible effects of genetic variation on DNA methylation alterations not accounted for.

References

Ibeagha-Awemu E.M. and Yu Y.. (2021) Animal Frontiers 11:7-18. https://doi.org/10.1093/af/vfab057

Li X., Li H., Pei X., Zhou Y. and Wei Z.. (2021) Journal of Interferon Cytokine Research 41:12-19. https://doi.org/10.1089/jir.2020.0193

Rohart F., Gaurier B., Singh A., Lê Cao K.M.. (2017) PLoS Computational Biology 13(11):e1005752. https://doi.org/10.1371/journal.pcbi.1005752

Teng C.S., Wu B.H., Yen M.R. and Chen P.Y.. (2020) BMC Genomics 21:1-10. https://doi.org/10.1186/s12864-020-6722-x

Wang, M.Q. and Ibeagha-Awemu E.M.. (2021) Frontiers in Genetics 11: 613636. https://doi.org/10.3389/fgene.2020.613636

Wu T.Z., Hu E., Xu S.B., Chen M.J., Guo P.F., *et al.* (2021) The Innovation 2:100141. https://doi.org/10.1016/j.xinn.2021.100141

548. eQTL mapping in Brown Swiss bulls to identify variants associated with male fertility

X.M. Mapel*, M. Hiltpold, A. Lloret-Villas and H. Pausch

Animal Genomics, ETH Zürich, Universitätstrasse 2, 8092 Zürich, Switzerland; xenamarie.mapel@usys.ethz.ch

Abstract

Fertility is an essential component of the livestock industry. In cattle, numerous QTL for male reproductive success fall within regulatory regions. However, the effects of these loci have not been investigated in detail or on a large scale. Here, we assemble a sizeable cohort of mature bulls to detect expression quantitative trait loci (eQTL) and assess their effects on fertility-related genes. To do this, we sequenced genomes and total RNA from the testes of 72 bulls. We recovered 13,185,795 DNA sequence variants with minor allele frequency >5%, an average of 283,587,831 clean RNA reads per sample and 18,528 testis-expressed genes (TPM>0.2 in 75% of samples). In total, 2,178 genes had significant *cis*-eQTL at a false discovery rate of 5% (11.76% of expressed genes). Several genes associated with fertility, including *SPATA4* and *SPATA22* (which are responsible for reproductive processes), had significant *cis*-eQTL and variation in transcript abundance across genotypes.

Introduction

Efficient breeding operations require animals with consistent and reliable reproductive success. Artificial insemination centres often collect hundreds of ejaculates from a single bull which are then sold to a multitude of producers. Thus, bulls with impaired or complete loss of reproductive ability can be costly to both parties. The genetic basis of such subfertility is complex. Indeed, quantitative trait loci (QTL) associated with measures of male reproductive success have been identified in several breeds at varying frequencies (Hiltpold *et al.* 2020; Sweett *et al.* 2020). Further insight into these QTL and detection of additional expression QTL (eQTL) that impact the abundance of fertility-related genes can help producers select breeding bulls with high reproductive ability.

Complex traits – such as those associated with semen quality and male reproductive success – are influenced by a multitude of polymorphic sites, with many of those residing in regulatory regions (Xiang *et al.* 2019). For example, a QTL associated with sire non-return rate has been identified in a region that possibly regulates the expression of *SPATA16* (Hiltpold *et al.* 2021), a gene involved with spermatogenesis. Previously, testis eQTL were identified by the cGTEx consortium (https://cgtex.roslin.ed.ac.uk/). However, this dataset relied on samples that were collected at different points of time from various bovine sub-species and breeds and did not investigate the specific effects of testis eQTL. A uniform cohort with an abundance of sequence data would identify additional eQTL and enable the exploration of their effect on reproductive ability.

Here, we harness a cohort of Braunvieh (BV) bulls to identify *cis*-acting eQTL in male reproductive tissue and assess their effect on genes vital for reproduction. To do this, we relied on whole-genome sequence data and total RNA from the testis tissue of 72 bulls.

Materials & methods

Sampling, extraction, sequencing. We sampled testis tissue from 72 bulls (between 8 and 34 months of age) with various degrees of BV ancestry (i.e. Brown Swiss, Original BV, or a cross between one of these and a different breed). After regular slaughter, tissue was flash frozen with liquid nitrogen and kept at -80 °C until DNA and RNA extraction. Frozen testis tissue was homogenized with a Roche MagNA Lyser in 2M DTT

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_548, © X.M. Mapel *et al.* 2022

and RTL Buffer (QIAGEN). DNA and total RNA were extracted simultaneously with the Qiagen AllPrep Mini Kit (QIAGEN) following manufacture protocols. DNA concentration and quality was assessed with a Qubit 2.0 fluorometer (ThermoFisher) and RNA integrity was quantified with a RNA integrity number (RIN) inferred from a Bioanalyzer RNA 600 Nano assay (Agilent Tech.). Genomic libraries (paired-end, 150 bp) were sequenced on the Illumina NovaSeq6000. Total RNA libraries (paired-end, 150 bp) were prepared with the Illumina Stranded Total RNA kit and Illumina Ribo-Zero Plus, then sequenced with the Illumina NovaSeq6000. Library preparation and sequencing was conducted by the Function Genomic Centre Zurich.

DNA sequence data. We used fastp (v0.19.4; Chen *et al.* 2018) with default parameters to remove adapter sequences and low-quality bases, and trim poly-G tails from raw DNA sequence data. After quality control, between 70,493,763 and 307,416,205 reads per sample were aligned to the ARS-UCD1.2 reference genome (Rosen *et al.* 2020) with the mem-algorithm from BWA (v0.717; Li 2013) and the -M flag. We used Sambamba (v0.2.2; Tarasov *et al.* 2015) to sort the aligned reads by coordinates, then combined the read-group specific BAM files to create sample specific sorted BAM files. The MarkDuplicates module from Picard tools (https://broadinstitute.github.io/picard/) was used to mark duplicated reads in the merged and coordinate sorted BAM files. We used mosdepth (v0.2.2; Pedersen and Quinlan 2018) to infer the sequencing coverage at a given genomic position, which was then used to estimate the average coverage per sample and chromosome. Only high-quality reads (mapping quality >10 or without SAM flag 1796) were considered when calculating the average coverage. We called sequence variants for the 72 samples with 24 additional samples from multiple breeds (Brown Swiss=6, Fleckvieh=1, Holstein=7, Original BV=5 Simmental=5), totalling to 96 samples. SNPs and indels were genotyped with the multi-sample variant discovery approach from the Haplotype Caller, Genomics DBImport and GenotypeGVCFs modules in GATK (McKenna *et al.*, 2010). We filtered variants according to GATK's best-practice guidelines. PLINK (v1.9; Purcell 2007) was used to produce a genomic relationship matrix (GRM) for 72 bulls and was constructed with variant genotypes that had minor allele frequency (MAF) >0.5%.

RNA sequence data. We used fastp to remove and trim low-quality bases, adapter sequences, poly-A tails and poly-G tails from the raw RNA sequence data. Filtered reads were aligned to the Ensembl gene annotation (release 104) for cattle with STAR (v2.7.9a; Dobin *et al.* 2013) using the recommended default parameters. This resulted in between 191,160,837 and 386,773,085 clean reads per sample (mean: 283,587,831±43,284,185). Gene abundance, quantified as transcripts per million (TPM), was calculated with the QTLtools 'quan' function with the --filter-mapping-quality and --filter-failed-qc flags to ignore low-quality reads (v1.3.1; Delaneau *et al.* 2017). Reads were mapped to 27,270 coding and non-coding genes from the most-recent Ensembl gene annotation for cattle.

eQTL analysis. We used QTLtools to identify *cis*-acting eQTL for testis expressed genes. Considering our small sample size, we filtered genomic variants to only include genotypes with MAF >5% in the 72 samples. We filtered the TPM matrix to only contain genes that had expression >0.2 TPM in 75% of samples. Age of the bull, RIN value, mass of the sampled testis, season of sampling and the first ten PCs of the GRM (to account for relatedness) were included as covariates for eQTL mapping. We used the *cis* function to identify eQTL within 1 Mb of the transcription start site (TSS) and implemented the --normal flag to rank-normal transform the gene counts, then performed 1000 permutations to infer beta-approximated *P*-values. To identify independent-acting loci and account for multiple testing, we calculated 5% false discover rate (FDR) and used the conditional mode in QTLtools. Allelic fold change (aFC) was calculated for variants of interest with methods described in Mohammadi *et al.* (2017).

Results & discussion

The 72 genomes had moderate sequence depth, which ranged from 6.66 to 27.94 (mean=12.84±4.23). Variant calling identified 20,401,236 variants with MAF>0.5% and 13,185,795 variants with MAF>5.0%. We detected 18,528 genes expressed with >0.2 TPM in 75% of samples.

Of the 18,528 expressed genes considered, 2,178 had at least one significant eVariant (11.76% of testis expressed genes; Figure 1). Each gene's top eVariant was often within 100 kb of the TSS (1,438 variants; 66.0%) and the average distance from the most-significant variant to the TSS was 149,656±229,236 bp (median: 40,824 bp). The number of genes with significant eQTL varied chromosome-by-chromosome (Figure 1B), with BTA18 having the most (165 genes with at least one eQTL). We detected 466,573 significant variants total (using 5% FDR), which is similar to the number of splice-site associated variants in the same cohort of bulls (over 495,000 sQTL; Kadri *et al.* 2021). The conditional analysis revealed that 86 genes had two independently acting eQTL (4.0% of genes with eQTL, 0.5% of all testis expressed genes). We identified more genes with significant *cis*-eQTL than the cGTEx consortium (cGTEx: 809 expressed genes with *cis*-eQTL in bovine testis), which is likely due to our larger sample size, deeper RNA sequencing coverage and more-uniform dataset.

Several genes that are associated with crucial male reproductive processes had *cis*-eQTL. For example, we detected *cis*-eQTL for two genes from the *SPATA* gene family – a gene-set responsible for spermatogenesis and fertilization (Figure 1a; Table 1). Expression of *SPATA4* was associated with a QTL on BTA27 (lead variant: G>A; position 7,704,939; adjusted P=1.51×10^{-9}) while expression of *SPATA22* was associated with a QTL on BTA19 (lead variant: C>T position 24,111,667; adjusted P=7.61×10^{-7}). Lead variants for both eQTL demonstrated additive effects: the abundance of each gene decreased in animals that carried the alternate allele, with an aFC of -0.24 and -0.36 for *SPATA22* and *SPATA4*, respectively (Figure 2). Abnormalities in *SPATA22* and *SPATA4* have been linked to infertility in multiple species (Ishishita 2013; Sujit 2019). Both genes were present in cGTEx, however they did not identify proximal eQTL. Future analyses will further inform if, and how, these and similar eQTL impact male reproductive ability in cattle.

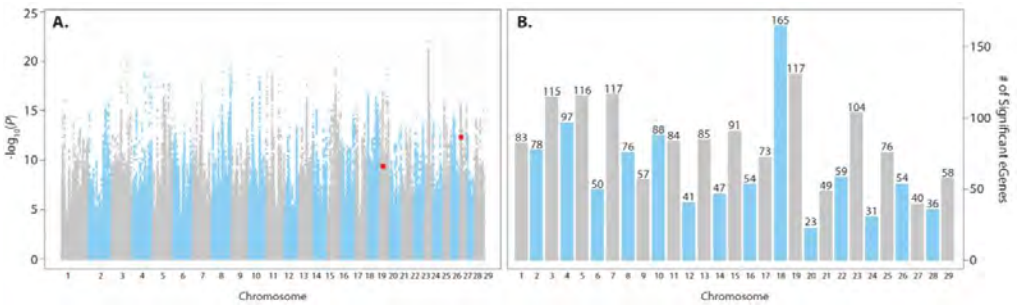


Figure 1. (A) Manhattan plot of genome-wide *cis*-eQTL with lead eVariants for *SPATA22* (chromosome 19) and *SPATA4* (chromosome 27) in red. (B) Number of testis-expressed genes with significant *cis*-eQTL per chromosome.

Table 1. Lead significant eVariants for *SPATA4* and *SPATA22*.

Gene	Position	# Variants in *Cis* window	aFC	Nominal *P*-value	Adjusted β *P*-value
SPATA4	27_7704939	16,450	-0.36	4.74×10^{-13}	1.51×10^{-9}
SPATA22	19_24111667	15,736	-0.24	3.99×10^{-10}	7.61×10^{-7}

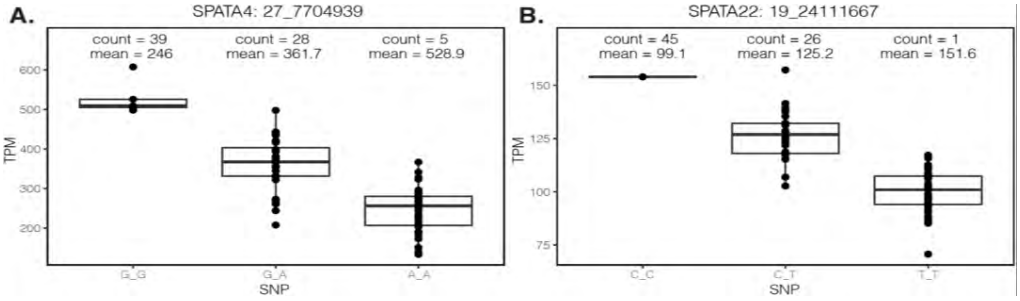


Figure 2. Abundance of *SPATA4* (A) and *SPATA22* (B) for different eVariant genotypes.

Conclusions

Here, we report the first analysis of a testis eQTL study in bulls and identify a large number of eQTL that warrant further investigation. Testing for *trans*-eQTL can uncover additional variants of interest. Furthermore, transcriptome-wide association analyses may help prioritize variants that are associated with variation in male reproduction.

Acknowledgements

Funding for this research was provided by an ETH Research Grant and Swissgenetics.

References

Chen S., Zhou Y., Chen Y., and Gu J. (2018) Bioinformatics 34:i884–i890. https://doi.org/10.1093/bioinformatics/bty560

Delaneau O, Ongen H, Brown A, Fort A, Panousis NI, *et al.* (2017) Nat. Commun. 8:15452.10.1038/ncomms15452

Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, *et al.* (2013) Bioinformatics. 29:15–21. https://doi.org/10.1093/bioinformatics/bts635

Hiltpold M, Niu G, Kadri NK, Crysnanto D, Fang ZH, *et al.* (2020) PLoS Genet. 16(5):e1008804. https://doi.org/10.1371/journal.pgen.1008804

Hiltpold, M, Kadri NK, Janett F, Witschi U, Schmitz-Hsu F, *et al.* (2021) BMC Genom. 22(1):1–19. https://doi.org/10.1186/s12864-021-07523-3

Ishishita S, Inui T, Matsuda Y, Serikawa T, and Kitada K (2013) Exp. Anim. 62(3):219-227. https://doi.org/10.1538/expanim.62.219

Kadri NK, Mapel XM, and Pausch H (2021) Commun Biol 4, 1206. https://doi.org/10.1038/s42003-021-02725-7

Li H (2013) http://arxiv.org/abs/1303.3997.

McKenna A, Hanna M. Banks E, Sivachenko A, Cibulskis K, *et al.* (2010) Genome Res. 20(9):1297-1303. https://doi.org/10.1101/gr.107524.110

Mohammadi P, Castel SE, Brown AA, and Lappalainen T (2017) Genome res., 27(11):1872. 1884. https://doi.org/10.1101/gr.216747.116

Pedersen BS and Quinlan AR (2018) Bioinformatics 34:867–868. https://doi.org/10.1093/bioinformatics/btx699

Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, *et al.* (2007) Am. J. of Hum. Genet.,81. https://doi.org/10.1086/519795

Rosen BD, Bickhart DM, Schnabel RD, Koren S, Elsik, CG *et al.* (2020) Gigascience 9: giaa021. https://doi.org/10.1093/gigascience/giaa021

Sujit KM, Singh V, Trivedi S, Singh K, and Gupta G (2020) Andrology, 8(3):602-609. https://doi.org/10.1111/andr.12742

Sweett H, Fonseca PAS, Suárez-Vega A, Livernois A, Miglior F, *et al.* (2020) Sci Rep 10:20102. https://doi.org/10.1111/andr.12742

Tarasov A, Vilella AJ, Cuppen E, Nijman IJ, and Prins P (2015) Bioinformatics 31:20322034. https://doi.org/10.1093/bioinformatics/btv098

Xiang R., Van Den Berg I, MacLeod IM, Hayes BJ, Prowse-Wilkins CP, *et al.* (2019) PNAS. 116(39):19398–19408. https://doi.org/10.1093/bioinformatics/btv098.

549. Bioinformatics workflow for the detection of eQTL in the cattle genome using Nextflow DSL2

P.K. Chitneedi[1*], F. Hadlich[1] and C. Kuehn[1,2]

[1]*Research Institute for Farm Animal Biology (FBN), Wilhelm-Stahl-Allee 2, 18196 Dummerstorf, Germany;* [2]*Agricultural and Environmental Faculty, University Rostock, Justus-von-Liebig-Weg 6, 18059 Rostock, Germany;* *chitneedi@fbn-dummerstorf.de*

Abstract

The *in silico* detection of expression quantitative trait loci (eQTL) demands high throughput processing from hundreds of samples, which is often a challenge to handle and run such large datasets. In order to focus on the core analysis, it is convenient to have simple coding and hassle-free installation of different software tools required for the bioinformatics workflow. In this context, the newly available technologies like workflow managers and software containers enabled to develop workflows with less complexity. In this study, we developed an eQTL bioinformatics pipeline with the workflow manager Nextflow and docker container software, for coding and installing the required software tools. This workflow can be portable to a different computer environment, and the results are reproducible. We tested the functionality of our workflow with a sample dataset and the runtime estimates from this demo run will provide important information in planning future analyses with much larger datasets.

Introduction

In recent years, large, cost-effectively sequenced transcriptome (RNA-Seq) and whole genome (WGS) data sets enabled association studies with gene expression as phenotype. These eQTL association studies identify the transcriptomic regulation at a genome wide scale by mapping the expression phenotype to a region on the genome (Gilad *et al.* 2008; Majewski and Pastinen 2011). The eQTL phenotype can be gene, transcript, exon expression or alternative splicing events. After high throughput sequencing, the main challenge in eQTL detection is to handle and process the large amount of generated data. When the data size is small (few gigabytes), it is convenient to implement the bioinformatics workflow with a simple computational script. However, things get complicated when dealing with large data especially in studies like eQTL, where generally thousands of RNA-Seq and WGS samples were analysed. In addition, it is not always possible to install multiple bioinformatics tools with different system requirements on a server maintained by an organization with centralized information technology support. Thus, to run such complex bioinformatics pipelines it is convenient to have a simple syntax declaration for calling the input data, executing the process and standard guidelines for installing all the required software tools. In this context, the concept of workflow managers and containers comes into play. Compared to traditional computational scripts used for bioinformatics workflows, the scripts developed using workflow managers are portable and reproducible on multiple server platforms. The most popular workflow managers for implementing bioinformatics pipelines are Snakemake (Molder *et al.* 2021) and Nextflow (Di Tommaso *et al.* 2017) among others. They use domain specific languages (DSLs), which have simple commands to run complex workflows, and the installation of all required software tools was standardised by using an open source package and environment systems like Conda (Anonymous 2020) or by container software with complete runtime environment like docker (Merkel 2014) or singularity (Kurtzer *et al.* 2017).

Materials & methods

In this current study, we used Nextflow_v21.04.1.5556 DSL2 for developing the workflow for eQTL detection in the bovine genome. Each step in the workflow was defined as an independent entity called 'Process' and they communicate with other process via input and output components called 'Channel'. In

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_549, © P.K. Chitneedi *et al.* 2022

the Nextflow DSL2, different processes can be written on separate files called 'Modules' independent of the main workflow script, which offers more flexibility to write complex workflows. All the required software to run the workflow were installed using the container software docker_v20.10.8, build 3967b7d. The docker environment was enabled for Nextflow by declaring in a configuration file 'nextflow.config'. In order to make the workflow easy to understand and troubleshoot potential errors, we divided the eQTL workflow into four NextFlow scripts as shown in Figure 1. Each script was further divided into individual modules, which perform a specific task within the workflow. In NextFlow DSL2, these modules were defined as individual entities, and they were invoked in the main script.

Script 0: Bovine genome STAR index. As a pre requisite to perform alignment using STAR (Dobin *et al.* 2013), it is required to perform the indexing of a reference genome, which was used repeatedly for aligning multiple samples. We built the STAR index for the latest bovine genome ARS-UCD1.2 and Ensembl annotation v102 using the STAR command '--runMode genomeGenerate'.

Script 1: RNA-Seq mapping and quantification of expression counts. Quality control and trimming of reads from samples with paired-end reads was carried out with FASTQC (http://www.bioinformaticsbabraham.ac.uk/projects/fastqc/) and trimmomatic v_0.36 (Bolger *et al.* 2014). The cleaned paired reads of each sample were forwarded to alignment using STAR_v2.7.0d. For the gene and transcript level counts, the reads were sorted by coordinates and for detecting splice junctions the reads were unsorted and we used '--outSAMstrandField intronMotif ' to determine the read origin either from intron or from splice junction. Using the StringTie_v1.3.3 (Pertea *et al.* 2015) assembler, we carried out the quantification and normalization at gene and transcript level, and where appropriate the strandedness of the RNA-Seq library was addressed with the StringTie options --rf for the first and --fr for the second stranded libraries. For detecting splice junctions the unsorted reads were indexed and sorted using Samtools_v1.9-4 and then exon-exon junctions were extracted from these bam files using regtools_v0.6.0 (Cotto *et al.* 2021) 'junctions extract', which was followed by clustering of introns found in the junction files to generate an across sample expression count matrix per each chromosome. Using leafcutter_v0.2.9 (Li *et al.* 2018) we estimated the top ten principal components (PCs). The count matrices and PCs were used in the model to detect splice eQTL in script 3.

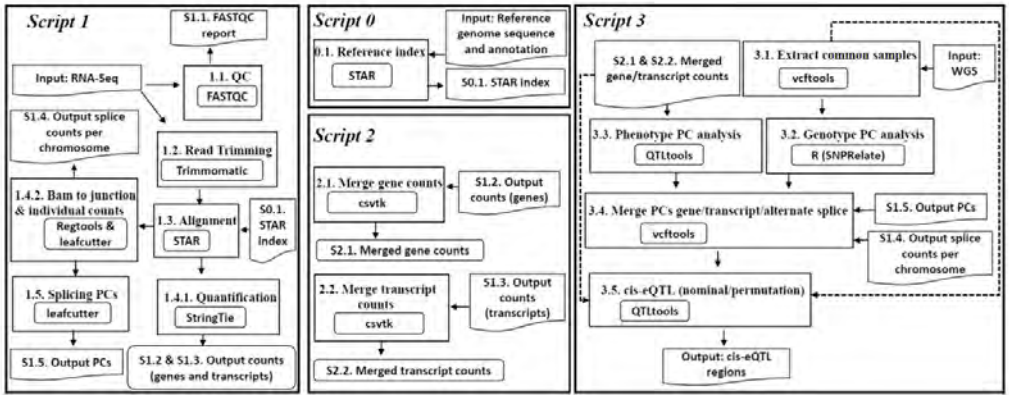


Figure 1. Diagrammatic representation of Bioinformatics workflow for eQTL detection designed using Nextflow DSL2. Each block represents individual scripts with different modules, input and output data flow.

Script 2: Merging count matrices at gene level and transcript level. After StringTie quantification, the output read count files are stored as .tsv (gene level) and .gtf (transcript level) output files. The TPM normalized expression counts of each sample at gene and transcript level were merged into matrix tables using csvtk --merge (https://bioinf.shenwei.me/csvtk/usage/). These tables were provided as input files to script 3 for the detection of eQTL.

Script 3: Estimating genotype and phenotype covariates and eQTL detection. The eQTL mapping was carried out with QTLtools_v1.3 (Delaneau *et al.* 2017). It expects three input files for mapping, genotype data in VCF format, phenotype data in .bed format and the principal components (PCs) as covariates. The input genotype data was filtered for the samples common to the phenotype input. Additionally, we filtered out the SNPs with minor allele frequency less 0.05. The population stratification present due to systematic ancestry differences in the data set was addressed by performing a PC analysis with genotype input using the R package SNPRelate (Zheng *et al.* 2012). Similarly, to address the outlier samples and batch effects in the RNA-Seq data at gene and transcript level, a phenotype PC analysis was performed with the QTLtools 'pca' command. For the eQTL mapping, the covariates include the top ten genotype PCs and top ten phenotype PCs. Finally, the nominal cis-eQTL mapping was carried out by QTLtools with the parameter --nominal 0.01, which provides as output only phenotype-variant pairs with a nominal *P*-value below 0.01, and the parameter --normal enforced the phenotypes to match a normal distribution N(0,1). In order to obtain the adjusted *P*-value for the cis associated phenotype and top variants, we performed permutation based QTL mapping also using QTLtools with the parameters --normal and --permute 1000.

Results & discussion

We developed this eQTL workflow with the primary goal of eQTL detection in the BovReg (https://www.bovreg.eu/) project with RNA-Seq phenotype and WGS data. To check the functionality of this eQTL workflow, we tested the workflow with 88 bovine liver RNA-Seq and the corresponding WGS data. The analysis was ran on Intel(R) Xeon(R) @ 2.10GHz server with 144 CPUs, 18 cores and 1TB RAM memory. Tests with a simple dataset will give an idea of how much runtime is required and potential pitfalls to consider, when running this workflow with much larger sample size. For the current test run, the WGS include 19,590,389 bi-allelic, polymorphic (MAF>0.005) variants across 29 autosomal chromosome, which were imputed using a stepwise strategy from 6k to 50k to HD to WGS (Pausch *et al.* 2016) with a mean imputation accuracy of 0.97 across different steps. The RNA-Seq samples consist of paired end libraries sequenced on the Illumina HiSeq 2500 system (Nolte *et al.* 2019; Heimes *et al.* 2020). The average number of input reads across 88 RNA samples were 56 million read pairs varying between 43 to 74 million read pairs. Although for this demo we used only paired-end reads, this pipeline will also work for single-read RNA-samples. The eQTL mapping will be carried out per chromosome by taking the WGS data for each chromosome for each iteration. By providing the configuration file and the scripts, this workflow can be portable, the analysis can reproducible on any Nextflow and docker installed computer. When executing the scripts the Nextflow installs all the required tools and users just have to take care of declaring the correct paths of input data. The Nextflow stores all log files and output data in a directory named as 'work'. This workflow was designed to copy all the important results from each step in a separate directory path. To run a workflow with large samples requires a rough estimation of runtime on a given computer infrastructure, as it takes few days to months to finish the analyses for large sample sizes. Thus, with the current demo of our eQTL workflow, we also checked the average runtime for 88 RNA-samples in the different processes on our local server. The alignment is the most time consuming process and took 108 hours for aligning all the 88 samples, while the runtime of the other processes in Script 1 varied from 1 to 10 hours. The runtime for Script 2 was just 2 minutes 41seconds, this was expected as this script only merges the available count information of different samples into a table. The final Script 3 with the PC analysis, QTL mapping and permutation across different levels (gene, transcript and splicing events) completed in 2 hours 28 min. This

runtime estimates provided us with valuable information to plan future analyses with higher sample size. Although this workflow was analysed using bovine datasets it can used for detecting eQTL in other species. In the future, the iteration of this workflow also includes allele specific expression and detection of trans-eQTL across different bovine tissues.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 815668.

References

Anonymous (2020) Available online at https://docs.anaconda.com/

Bolger A.M., Lohse M. and Usadel B. (2014) Bioinformatics 30(15)**:** 2114-2120. https://doi.org/10.1093/bioinformatics/btu170

Cotto K.C., Feng Y.-Y., Ramu A., Skidmore Z.L., Kunisaki J. *et al.* (2021) BioRxiv https://doi.org/10.1101/436634

Delaneau O., Ongen H., Brown A.A., Fort A., Panousis N.I. *et al.* (2017) Nat Commun 8:15452. https://doi.org/10.1038/ncomms15452

Di Tommaso P., Chatzou M., Floden E.W., Barja P.P., Palumbo E. *et al.* (2017) Nat Biotechnol 35(4)**:** 316-319. https://doi.org/10.1038/nbt.3820

Dobin A., Davis C.A., Schlesinger F., Drenkow J., Zaleski C. *et al.* (2013) Bioinformatics 29(1)**:** 15-21. https://doi.org/10.1093/bioinformatics/bts635

Gilad Y., Rifkin S.A. and Pritchard J.K. (2008) Trends Genet 24(8)**:** 408-415. https://doi.org/10.1016/j.tig.2008.06.001

Heimes A., Brodhagen J., Weikard R., Seyfert H.M., Becker D. *et al.* (2020) Front Immunol 11:715. https://doi.org/10.3389/fimmu.2020.00715

Kurtzer G.M., Sochat V. and Bauer M.W. (2017) Plos One 12(5). https://doi.org/10.1371/journal.pone.0177459

Li Y.I., Knowles D.A., Humphrey J., Barbeira A.N., Dickinson S.P. *et al.* (2018) Nature Genetics 50(1)**:** 151-158. https://doi.org/10.1038/s41588-017-0004-9

Majewski J., and Pastinen T. (2011) Trends Genet 27(2)**:** 72-79. https://doi.org/10.1016/j.tig.2010.10.006

Merkel D. (2014) Linux journal 2014(239)**:** 2.

Molder F., Jablonski K.P., Letcher B., Hall M.B., Tomkins-Tinch C.H. *et al.* (2021) F1000Res 10:33. https://doi.org/10.12688/f1000research.29032.2

Nolte W., Weikard R., Brunner R.M., Albrecht E., Hammon H.M. *et al.* (2019) Front Genet 10:1130. https://doi.org/10.3389/fgene.2019.01130

Pausch H., Emmerling R., Schwarzenbacher H. and Fries R. (2016) Genet Sel Evol 48:14. https://doi.org/10.1186/s12711-016-0190-4

Pertea M., Pertea G.M., Antonescu C.M., Chang T.C., Mendell J.T. *et al.* (2015) Nature Biotechnology 33(3)**:** 290-295. https://doi.org/10.1038/nbt.3122

Zheng X., Levine D., Shen J., Gogarten S.M., Laurie C. *et al.* (2012) Bioinformatics 28(24)**:** 3326-3328. https://doi.org/10.1093/bioinformatics/bts606

550. Extensive functional genomics information from early developmental time points for pig and chicken

H. Acloque[1], P.W. Harrison[2], W. Lakhal[3], F. Martin[2], A.L. Archibald[4], M. Beinat[1], M. Davey[4], S. Djebali[5], S. Foissac[6], S. Guizard[4], C. Guyomar[6], R. Kuo[4], C. Kurylo[6], O. Madsen[7], K. Miedzinska[4], M. Mongellaz[1], J. Smith[4], J. Smith[4], A. Sokolov[2], J. de Vos[7], E. Giuffra[1] and M. Watson[4*]

[1]Paris-Saclay University, INRAE, AgroParisTech, GABI, 78350 Jouy-en-Josas, France; [2]EMBL – European Bioinformatics Institute, Wellcome Genome Campus, Hinxton, CB10 1SD, United Kingdom; [3]Epigenetics R&D, Diagenode S.A., Liège Science Park, Ru du Bois Saint-Jean 3, 4102 Liège, Belgium; [4]The Roslin Institute and R(D)SVS, University of Edinburgh, Easter Bush, EH25 9RG, United Kingdom; [5]IRSD, Université de Toulouse, INSERM, INRA, ENVT, UPS, 31024 Toulouse, France; France; [6]GenPhySE, Université de Toulouse, INRAE, ENVT, 31326 Castanet Tolosan, France; [7]Animal Breeding and Genomics, Wageningen University and Research, 6700 AH Wageningen, the Netherlands; mick.watson@roslin.ed.ac.uk

Abstract

The global Functional Annotation of Farm Animal Genomes initiative (FAANG) aims to improve animal breeding by improved genomic prediction via integration of functional genomics information. The GENE-SWitCH project has produced extensive functional genomics information for a variety of important tissues at early embryonic timepoints for both chickens and pigs. These datasets will be integrated to produce both tissue and time-point specific transcript, gene, and regulatory annotation for both species. In this paper, we describe the aims of the project, and the initial release of both raw and processed data.

Introduction

Genome annotation is often presented in a static, linear, fashion, as a set of genes, transcripts and regulatory regions which exist on the genome. However, we know that gene, transcript, and regulatory regions differ in both a time- and tissue- specific manner, hence current genome annotation data sets may miss important temporal and tissue-specific differences in the expression and use of functional genomic regions, and these regions may underlie key traits in livestock species. Addition of functional genomics information is therefore expected to improve animal breeding through improved genomic prediction (Andersson *et al.* 2015, Clark *et al.* 2020).

'The regulatory GENomE of SWine and CHicken: functional annotation during development (GENE-SWitCH)' project contributes to the global FAANG initiative and to EuroFAANG (https://eurofaang.eu), a synergy of five Horizon 2020 projects that share the common goal to discover links between genotype to phenotype in farmed animals and meet global FAANG objectives. A core activity of GENE-SWitCH is to generate extensive functional genomics information for key tissues during early development of both chickens and pigs, with the overall aim of improving genomic selection via more accurate genomic prediction using models that incorporate knowledge of functional regions of the genome.

Tissue- and time-point specific genome annotations are generated by standardized approaches (sampling, functional assays and analyses) to identify genes (coding and non-coding), transcripts, and regulatory sequences.

This paper constitutes a 'marker' paper of GENE-SWitCH in the context of the Toronto Statement on Pre-Publication Sharing (Toronto International Data Release Workshop Authors *et al.* 2009) in which we describe our plans for the first use of these data for genome-wide analyses and publication.

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_550, © H. Acloque *et al.* 2022

Materials & methods

Animals, tissues, time-points. Post-mortem samples were harvested from embryos (chicken: E8, E15; pig: 30, 70 days post-fertilisation) to capture early and late organogenesis; and from new born/hatched piglets and chicks. Large White pigs, that are one of the most widely used commercial pigs, were used. The chickens were from the Roslin broiler line.

Seven different tissues (liver, small intestine, kidney, lung, hindlimb muscle, brain, skin) were chosen to represent the three layers of embryonic and fetal development (endoderm, mesoderm, ectoderm) and the physiological functions of most relevance for the further integration in genomic prediction models.

Tissues and developmental stages were chosen to complement those being characterised in complementary USDA-funded chicken and pig projects, minimising duplication, whilst including some element of replication. Links to the data and metadata for all samples are available at https://data.faang.org/projects/GENE-SWitCH

Functional genomic assays. The samples have been processed to yield chromatin, DNA and RNA for the following assays: Promoter Capture Hi-C, ATAC-seq, methylation profiling by whole genome and reduced-representation bisulphite sequencing, long read RNA sequencing using Pacific BioSciences (PacBio) Iso-Seq, RNA-seq (poly-A+), small RNA-seq, and chromatin immunoprecipitation sequencing (ChIP-seq) for sequences bound by modified histones (H3K4me1, H3K4me3, H3K27Ac and H3K27me3) and CTCF. Analyses of the sequence data from these assays will allow us to identify open chromatin indicative of active elements of the genome and key regulatory elements including promoters and enhancers as well as a more comprehensive view of the chicken and porcine transcriptomes and their dynamics during development.

Pipelines of analysis. As a basis for bioinformatics analysis of the GENE-SWitCH data, the nf-core pipelines have been adopted (Ewels *et al.* 2020). In the case of RNA-Seq, a GENE-SWitCH specific pipeline has been developed, TAGADA, which has additional features, in particular annotation of long non-coding RNAs. For DNA methylation analysis an extension of the nf-core/methylseq pipeline (GSM), including downstream analysis, was developed. In the case of IsoSeq data, an nf-core pipeline did not exist, therefore an IsoSeq pipeline has been developed and will be submitted to nf-core. All GENE-SWitCH bioinformatics pipelines are available on the FAANG github page: https://github.com/orgs/FAANG/.

Results

The data production phase of the project is almost complete with the generation of ChIP-seq data due for completion in Spring 2022. The sequence data have been deposited in the public domain sequence databases in accordance with the FAANG Data Sharing guidelines (https://www.faang.org/data-share-principle). These data are indexed on the GENE-SWitCH project pages in the FAANG Data Portal (https://data.faang.org/projects/GENE-SWitCH). Only those assays which have been completed are listed below.

RNA-Seq data. RNA-Seq reads have been used to identify and quantify expression of both known and novel transcripts and genes. The vast majority of the known reference genes (73.1% for pig and 81.5% for chicken) and transcripts (72.0% for pig and 78.8% for chicken) are expressed (TPM≥0.1) in at least one sample. In terms of novel genes, we predict an additional 2,103 for pig and 1,642 for chicken. Meanwhile, there are two to three times as many nc transcripts in the novel gene annotation as in the reference gene annotation (+92% for pig and +192% for chicken). From the novel annotation spliced transcripts, 8.9 and 22.1% extend a reference transcript by adding at least one exon in 5' or 3' for pig and chicken respectively, and 34.2 and 41.8% represent completely novel spliced transcripts. Around 15,000 transcripts are predicted as lncRNAs in the two species, of which the majority (around 70%) are intergenic.

The hierarchical clustering based on reference gene expression for pig can be seen in Figure 1. Samples cluster first by tissue for cerebellum, kidney and liver while for lung, ileum, muscle and skin, this is the developmental stage which is the main criterion for clustering, especially for the earliest stage.

IsoSeq data. IsoSeq is a method which uses PacBio sequencing to produce high quality, full length transcript sequences. In the GENE-SWitCH project, samples were pooled across individuals to produce one IsoSeq dataset per tissue, time-point and species (42 datasets). At the point of submission, IsoSeq data is only available for chicken, and IsoSeq has previously been applied to chicken tissues by Kuo *et al.* (2017) providing a useful comparator. An initial annotation for chicken using the IsoSeq data has been produced using standard IsoSeq filtering and clustering techniques, chimera removal, and removal of low confidence annotations. The number of detected transcripts is four times higher than in the Ensembl v105 annotation, and 2.5 times higher than in Kuo *et al.* (2017). The number of detected genes is similar to Ensembl v105 (Ensembl 24,356, GENE-SWitCH 24,152), however, the GENE-SWitCH IsoSeq annotation predicts 6,865 potential new genes.

Small RNA-Seq data. Between 30 and 120 million reads were sequenced per sample. Out of those, a median of 57.6% for pig and 45.0% for chicken could be mapped to mirBase mature miRNAs. From the remaining reads, a median of 39.0% reads for pig and 53.5% reads for chicken could be mapped to mirBase hairpin miRNAs. Similar to RNA-Seq, hierarchical clustering based on miRBase mature miRNA normalized expression led to a clustering by tissue (results not shown). As for miRNA discovery, 6,012 and 9,905 miRNAs were detected for pig and chicken, of which 4,873 (81.1%) and 8,124 (82.0%) were predicted as novel.

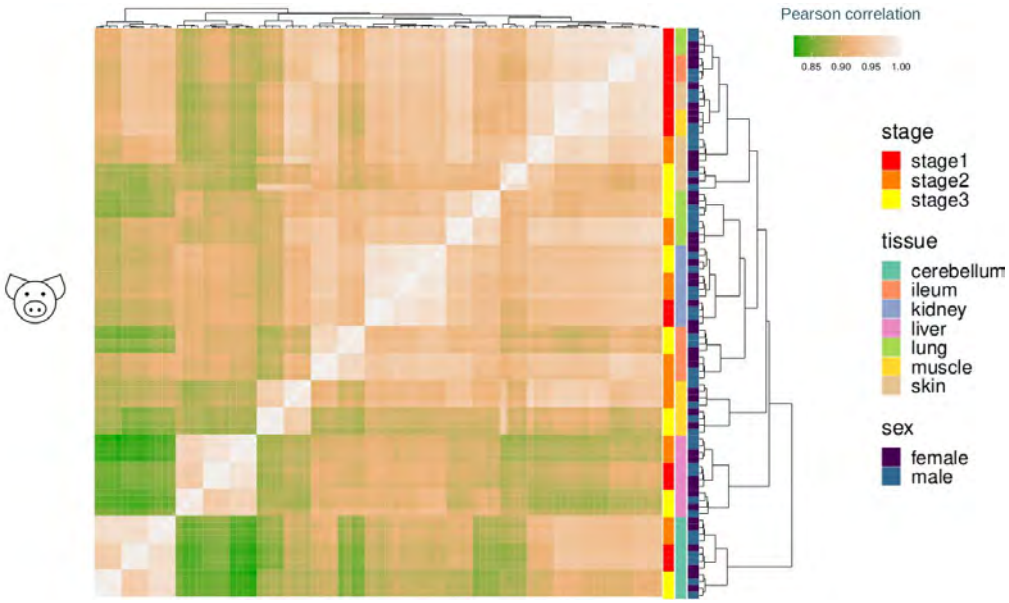


Figure 1. Hierarchical clustering of RNA-Seq data for pig. Hierarchical clustering of pig RNA-Seq data demonstrates that, as expected, samples largely cluster first by tissue, and then by developmental stage.

Bisulfite sequencing. Reduced-representation bisulfite sequencing was performed on 63 samples, and the remaining 21 samples were submitted for whole-genome bisulfite sequencing. Quality of all samples as well as mapping rates for both pig and chicken were very high. Clustering analysis of the samples provides insight into the development of the tissues, where we observe both tissue- and time-point- specific clustering in both species (results not shown).

Discussion

Functional genomics data is hypothesized to improve animal breeding by enabling more accurate genomic prediction, and a core aim of the GENE-SWitCH project is to produce extensive functional and regulatory genomics information for early developmental time-points in both pigs and chickens. Here, we have described an initial data release of both raw and processed data, as well as describing early results of how these data are likely to impact the genome annotation of these species. All data are available through the FAANG data portal.

We are using these multiple datasets: (1) to characterise the complexity of the coding and non-coding transcriptome and its dynamics during development in pigs and chickens; (2) to characterise the regulatory and epi- genomes of pigs and chickens; (3) to compare the regulatory landscape and its influence on gene expression and development in species separated by millions of years of evolutionary time (mammals, e.g. pigs and birds, e.g. chickens). In addition, to publishing the results of these analyses, enhanced annotation of the chicken and pig genomes will be released via the Ensembl Genome Browser (https://www.ensembl.org/index.html) initially as additional tracks in the Rapid Release site (https://rapid.ensembl.org/index.html). This enhanced annotation will facilitate the development of improved models for genomic prediction that incorporate knowledge of the likely functions of genome sequences as well as accounting for genetic variation.

Acknowledgements

This work is part of the GENE-SWitCH project that has received funding from the European Union's Horizon 2020 Research and Innovation Programme under the grant agreement no. 817998.

References

Clark, E.L., Archibald, A.L., Daetwyler, H.D. *et al.* From FAANG to fork: application of highly annotated genomes to improve farmed animal production. (2020) Genome Biology 21: 285. https://doi.org/10.1186/s13059-020-02197-8

Ewels, P.A., Peltzer, A., Fillinger, S. *et al.* (2020) The nf-core framework for community-curated bioinformatics pipelines. Nature Biotechnology 38:276–278. https://doi.org/10.1038/s41587-020-0439-x

Kuo, R.I., Tseng, E., Eory, L. *et al.* (2017) Normalized long read RNA sequencing in chicken reveals transcriptome complexity similar to human. BMC Genomics 18:323.

Andersson, L., Archibald, A.L. *et al.* (2015) Coordinated international action to accelerate genome-to-phenome with FAANG, the Functional Annotation of Animal Genomes project. Genome Biology 16:57.

Toronto International Data Release Workshop Authors, Birney E., Hudson T.J., Green E.D., Gunter C., *et al.* (2009) Nature 461(7261):168-170. https://doi.org/10.1038/461168a

551. Demonstrating the use of a mammalian methylation microarray for epigenome wide association studies in livestock

A. Caulton[1,2*], K.G. Dodds[1], K.M. McRae[1], C. Couldrey[3], S. Horvath[4] and S.M. Clarke[1]

[1]AgResearch Limited, Invermay Agricultural Centre, Puddle Alley, Mosgiel 9092, New Zealand; [2]Department of Biochemistry, University of Otago, Dunedin 9016, New Zealand; [3]Livestock Improvement Corporation, Hamilton 3286, New Zealand; [4]Department of Human Genetics, David Geffen School of Medicine, University of California Los Angeles, Los Angeles, CA 90095, USA; alex.caulton@agresearch.co.nz

Abstract

DNA methylation is an important epigenetic mark that is essential for genomic stability and maintenance throughout development. Advances in 'omics' technologies has fuelled investigation into the epigenome as a tool to enhance livestock selection and breeding practices; research into DNA methylation will deepen our understanding of gene-by-environment interactions and their influence on phenotypic outcomes in livestock. The recent release of a mammalian methylation array has enabled array-based 'epityping' of methylation across the Mammalia class. This array facilitates greater accessibility for investigation in the methylome within the livestock sector. Here we demonstrate the novel use of the mammalian methylation array in an epigenome-wide association study (EWAS) for age across four key ruminant species; deer, cattle, goat and sheep and identify many associated sites including two regions of note *TNRC6A and LHFPL4,* which have previously been correlated with the ageing process in independent studies.

Introduction

A wide range of techniques are available to explore DNA methylation from the single-gene level through to genome-wide profiling. Bisulphite treatment of DNA, the most widely adopted tool to profile methylation, converts unmethylated cytosines to uracil while leaving methylated cytosines unchanged. Bisulphite converted DNA can be analysed by sequenced based approaches, including whole genome bisulphite sequencing, the gold standard approach for comprehensive methylation profiling, and reduced representation bisulphite sequencing, a more cost-effective derivative. Alternatively, micro array-based analysis can provide single base resolution 'epi-typing' of targeted CpGs if a suitable BeadChip array is available for the species of interest. Until recently methylation arrays have only been available for three model species: human, rat and mouse. Recently, the mammalian methylation array was developed in an attempt to fill this gap (Arneson *et al.*, 2022). Built using probes conserved across the Mammalia class, it was designed to apply to all mammals, in order to drive down the cost of the chip through economies of scale. The BeadChip contains approximately 36,000 CpG sites built from sequence data from over 50 different mammalian species, facilitating the uptake of this technology in the animal genetics sector. Perhaps the greatest advantage of microarray technologies over sequence-based methods is the simple interpretation of the data without the requirement for computationally demanding bioinformatic analyses. Furthermore, no PCR amplification step is required, therefore, there is no bias towards shorter, less CpG rich fragments as is the case with bisulphite-sequencing. Here we demonstrate the application of the mammalian methylation array across four ruminant species and evaluate its potential as a tool to investigate the global methylation landscape and discover methylated sites associated with phenotypic outcomes in livestock, using age as an exemplar.

Materials & methods

Samples. The EWAS was performed on a previously described dataset (Caulton *et al.*, 2022). Briefly, this dataset consisted of 384 ear tissue samples from four different species: deer (Wapiti and Red pure and cross bred; ages ranged from 1 month to 13 years and 8 months old; n=96), sheep (composite breed; ages ranged

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_551, © A. Caulton *et al.* 2022

from 16 days to 7 years and 1 month old; n=90), cattle (Jersey/Friesian cross breed; ages ranged from 1 month to 11 years and 1 month old; n=96) and goat (Saanen breed; ages ranged from 2 weeks to 8 years and 7 months of age; n=96).

DNA preparation and Infinium microarray. DNA extraction, bisulphite conversion of DNA, typing of the samples on the custom Illumina methylation array 'HorvathMammalMethyl40' (Illumina, Inc., San Diego,CA, USA; Arneson *et al.*, 2021) was performed as described in Caulton *et al.*, 2022.

Characterising the global methylation landscape. The distribution of cytosines assayed via the mammalian methylation array was determined using the sheep dataset set as an example. Probe sequences were aligned to the Oar Rambouillet v1.0 genome, GenBank accession number: GCA_002742125.1. Global methylation levels were determined by averaging the methylation levels at shared cytosines (minimum of 5× coverage) between all sheep samples. A cytosine was defined as being highly methylated if on average the proportion of methylated reads to unmethylated reads was ≥0.5, moderately methylated if its methylation proportion was between 0.1 and 0.5 and lowly methylated if its proportion of methylated reads to unmethylated reads was ≤0.1. To determine how methylation was distributed within various features of the genome, the location of lowly, highly, and moderately methylated cytosines in relation to CDS, intronic regions, exonUTRs, 1 kb flanks, intergenic regions, transposable elements and lncRNAs was determined using intersectBed (Quinlan and Hall, 2010).

Age transformation. To test for associations with age across species with differing maximum lifespans, chronological age was transformed to relative age, in line with previous work for development of multispecies epigenetic clocks (Caulton *et al.*, 2022; Mammalian Methylation Consortium *et al.*, 2021).

$$RelativeAge = \frac{Age+2\times GT}{MaxAge+2\times GT} \quad (2)$$

$$LoglogAge = -log\big(-log(RelativeAge)\big) \quad (3)$$

Where *GT* is the species' gestational time and *MaxAge* its maximum lifespan. *RealtiveAge* is between 0 and 1 (for all species) and the double log transformation accounts for aging effects occurring more rapidly at younger (relative) ages.

Data quality control and epigenome-wide association study (EWAS). A Bartlett's test for homogeneity of variance of *LoglogAge* across species was non-significant and therefore a single association analysis was run including all species. Technical bias and batch effects were corrected for with the SeSAMe analysis suite (Zhou *et al.*, 2018) as described in Caulton *et al.,* 2022. Any residual effects were handled by including supplied (sex and species) and surrogate variables obtained by independent surrogate variable analysis (ISVA; Teschendorff *et al.*, 2011) in the EWAS models run in R (v4.1.0) using the Bioconductor package Meffil (Min *et al.*, 2018).

Results & discussion

The probes included on the mammalian methylation array were principally designed to target CpGs that are highly conserved across the Mammalia class. In addition, sites were specifically chosen for being distributed across a wide range of fractional methylation levels and preferentially selected for the array if they resided in gene or promoter regions (Arneson *et al.*, 2021). In human DNA, around 80% of the mammalian methylation array probe sequences are either in a gene body or promoter. In the sheep dataset (Figure 1), the proportion was much lower at 66.4%, likely a consequence of the sheep genome having poorer annotation compared to the human genome. The distribution of high, moderate, and lowly methylated sites within the sheep array dataset was 64.0, 20.7 and 15.3% respectively.

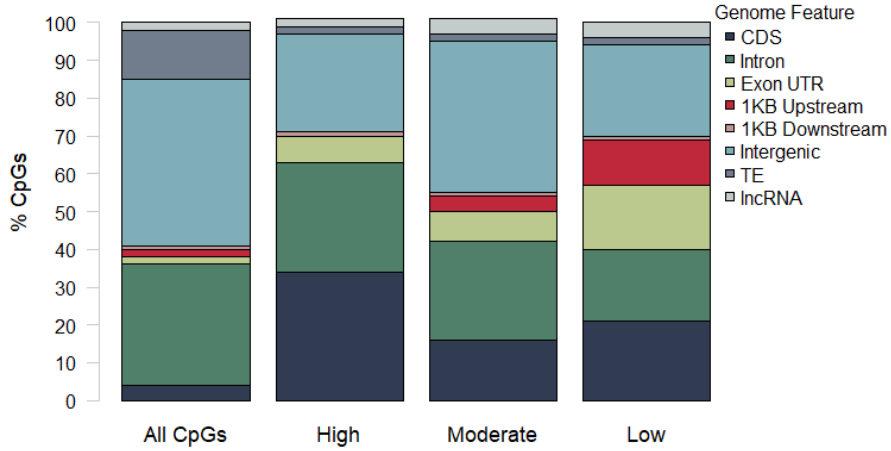


Figure 1: Distribution of CpG sites within different functional categories of the sheep genome (Oar Rambouillet v1.0). 'All CpGs' refers to the global distribution of CpGs in the complete genome. 'High', 'Moderate' and 'Low' refer to highly methylated, moderately methylated, and lowly methylated CpGs that were sampled via the array, respectively.

Genome wide profiling of DNA methylation using the Mammalian methylation array to determine methylation levels at ~36,000 CpG sites revealed a large number of positions that were strongly associated with age. The most highly associated cytosine cg09461098 (P=2.35e-50) resides on chromosome 16 in the gene *TNRC6A*. This cytosine was found to be highly predictive of chronological age in previous work to develop a biomarker for ageing in livestock (Caulton *et al.*, 2022) and has been linked to ageing in similar biomarker studies in prairie voles (Sailer *et al.*, 2020) and primates (Horvath *et al.*, 2020). In a recent study to develop a universal epigenetic biomarker of age, two cytosines cg12841266 and cg11084334 were highlighted as being the most predictive of age across all species (Mammalian Methylation Consortium *et al.*, 2021). In the EWAS performed here the same two positions on chromosome 3 ranked the 9th (P=4.54e-36) and the 25th (P=8.38e-24) most significant, respectively. Both sites are located in exon 2 of the *LHFPL4* gene. The association of this gene with ageing has been emphasised in multiple independent studies although its role is still unclear.

Here we describe the first use of the mammalian methylation array for EWAS studies in livestock species and demonstrate its potential as a tool to identify differentially methylated sites associated with phenotypic outcomes. The distribution of cytosines captured via the array are biased towards genic regions (gene body and promoter) which will aid in the discovery of genes under epigenetic regulation for traits of interest.

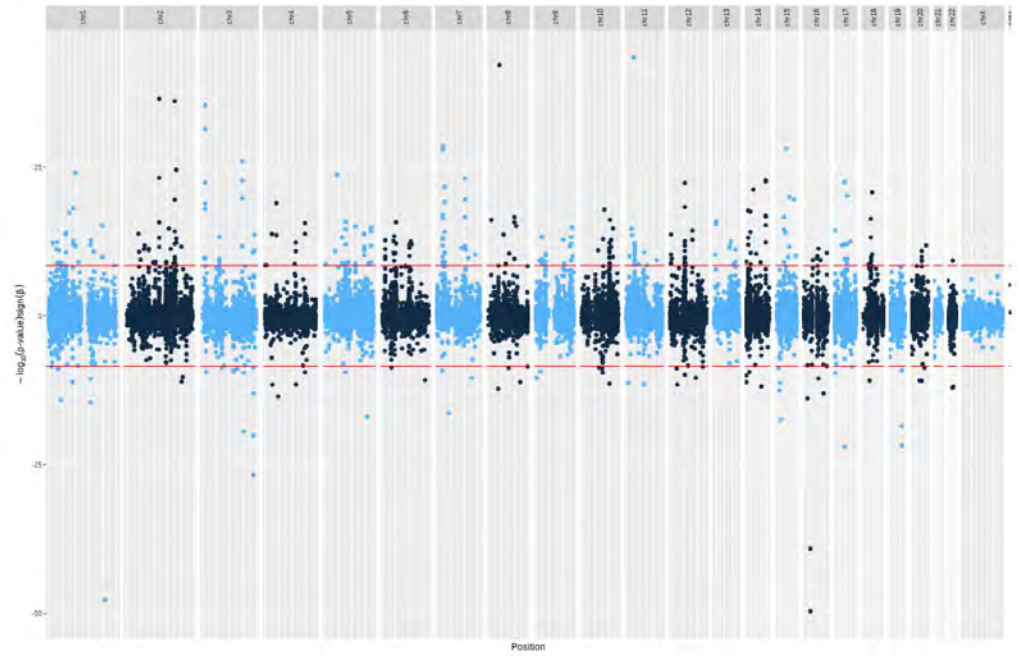


Figure 2. Manhattan plot of log(*P*-values) × effect sign from an epigenome-wide association analysis of relative age in New Zealand livestock (deer, cattle, sheep and goat), red lines indicate the Bonferroni significance threshold *P*=3.472e-9.

References

Arneson A., Haghani, A., Thompson, M. J., Pellegrini, M., Kwon, S. B., *et al.* (2022) Nature Commun. 13(1), 783. https://doi.org/10.1038/s41467-022-28355-z

Caulton A., Dodds, K. G., McRae, K. M., Couldrey, C., Horvath, S., and Clarke, S. M. (2022) Genes, 13(1), 96. https://doi.org/10.3390/genes13010096

Horvath S., Haghani, A., Zoller, J. A., Ernst, J., Pellegrini, M., *et al.* (2020). BioRxiv 2020.11.29.402891. https://doi.org/10.1101/2020.11.29.402891

Mammalian Methylation Consortium, Lu, A. T., Fei, Z., Haghani, A., Robeck, T. R., Zoller, J. *et al.* (2021). BioRxiv 2021.01.18.426733. https://doi.org/10.1101/2021.01.18.426733

Min J. L., Hemani, G., Davey Smith, G., Relton, C., and Suderman, M. (2018) J. Bioinform. 34(23), 3983–3989. https://doi.org/10.1093/bioinformatics/bty476

Pidsley R., Zotenko, E., Peters, T. J., Lawrence, M. G., Risbridger, G. P., *et al.* (2016). Genome Biol. 17. https://doi.org/10.1186/s13059-016-1066-1

Quinlan A. R., and Hall, I. M. (2010) J. Bioinform. 26(6), 841–842. https://doi.org/10.1093/bioinformatics/btq033

Teschendorff A. E., Zhuang, J., and Widschwendter, M. (2011) J. Bioinform. 27(11), 1496–1505. https://doi.org/10.1093/bioinformatics/btr171

Zhou W., Triche, T. J., Jr, Laird, P. W., and Shen, H. (2018). Nucleic Acids Res. 46(20), e123–e123. https://doi.org/10.1093/nar/gky691

552. Regulatory QTL and exon expression QTL in the mammary gland of dairy cows

C.P. Prowse-Wilkins[1,2*], T.J. Lopdell[3], R. Xiang[1,2], C.J. Vander Jagt[1], M.D. Littlejohn[3], A.J. Chamberlain[1] and M.E. Goddard[1,2]

[1]Agriculture Victoria, AgriBio, Centre of AgriBioscience, 5 Ring Road, Bundoora, Victoria, 3082, Australia; [2]Faculty of Veterinary & Agricultural Science, University of Melbourne, 3010 Parkville, Victoria, Australia; [3]Research and Development, Livestock Improvement Corporation, Private Bag 3016, Hamilton, 3240, New Zealand; c.prowsewilkins@gmail.com

Abstract

Causal variants are enriched in functional non-coding regions of the genome, where it is hypothesised that they influence regulation of gene expression. Histone modifications are thought to mark many non-coding functional regions. Additionally, the height of some histone modification peaks has been found to be correlated with nearby gene expression, demonstrating a potential causal relationship between the two. Therefore, SNP which associate with differences in gene expression and with differences in peak height could be strong candidate causal variants. We found that QTL for peak height for three histone modifications (H3K4Me1, H3K4Me3 and H3K27ac) in ~100 mammary samples in dairy cows overlapped substantially with QTL for gene expression in ~400 mammary samples, thus identifying putative regulatory variants in the bovine genome.

Introduction

Gene expression QTL studies identify putative causal variants for differences in gene expression, which is thought to be a major contributor to phenotypic variation in other complex traits (Schadt *et al*., 2005). Most putative causal variants are found in non-coding regions of the genome (Cano-Gamez *et al*., 2020). It has been suggested that this is because they are in regulatory regions, which are controlling gene expression.

Regulatory regions can be identified by assaying for histone modifications using a technique called Chromatin Immunoprecipitation followed by sequencing (ChIP-seq) (Prowse-Wilkins *et al*., 2021). This method identifies the location of functional regions such as enhancers (marked by H3K4Me1), promoters (marked by H3K4Me3) and active regions of the genome (marked by H3K27ac). However, this assay also quantitively measures histone modification peak height. Peak height has been found to be correlated with gene expression (Cheng *et al*., 2012, Prowse-Wilkins *et al*., 2021), so it is thought that differences in peak heights may be altering gene expression. Therefore, a SNP which is altering peak height and gene expression may be altering gene expression by altering binding of histone modifications.

To identify SNP which are associating with differences in peak height and gene expression we undertook allele specific QTL analysis for these two molecular phenotypes. Allele specific QTL analysis measures the phenotype (in this case peak height or gene expression) on the maternal chromosome of an animal in comparison to the phenotype on the paternal chromosome (Chen *et al*., 2016). If a QTL is acting in cis and is heterozygous in the animal, the allele count at each homologous chromosome would be different. This analysis has the benefit of being measured within one homogeneous environment (one animal), which should limit background noise.

Allele specific QTL analysis was performed in ~400 RNA-seq samples and ~100 samples for 3 different histone modifications in the mammary gland of dairy cows. Our results revealed that, as expected, there was substantial overlap between SNP which were regulatory QTL (allele specific binding QTL or asbQTL) and SNP affecting gene expression (allele specific expression QTL or aseQTL).

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_552, © C.P. Prowse-Wilkins *et al.* 2022

Methods

ChIP-seq and RNA-seq data. RNA-seq and data analysis for 411 mammary biopsies was described in Littlejohn *et al.* (2016) and Lopdell *et al.* (2017). A subset of 101 mammary biopsies were assayed with ChIP-seq for H3K4Me3 and H3K4Me1 and a further subset of 37 of these were assayed for H3K27ac. The ChIP-seq assay and subsequent data analysis was as described in Prowse-Wilkins *et al.* (2021).

Allele counts. To prevent mapping bias to reference alleles, a masked genome was created from ARS-UCD1.2. with a neutral allele at variants which were heterozygous in the sequenced animals (MAF>1% n=14,536,882). Imputation to whole genome sequence and variant calling was described in Reynolds *et al.* (2021). Consensus peaks were produced by randomly sampling all bam files for each mark with SAMtools (Li *et al*., 2009) and calling peaks from combined reads. Exons were from the RefSeq protein-coding gene database (annotation release 106).

For each heterozygous SNP, alleles were defined as maternal or paternal based on the sire's genotype at that SNP. If the sire was homozygous, this allele was designated the paternal allele, if the sire was heterozygous, the maternal/paternal classification was based on the phasing of the previous SNP. Maternal and paternal read counts were calculated by counting alleles at heterozygous SNP under a consensus peak (peak SNP or pSNP) or within an exon (transcript SNP or tSNP) using GATK HaplotypeCaller and GenotypeGVCF using the option 'depth per allele by sample'. SNPs which were monoallelic were excluded from analysis as this was assumed to be an imputation error (the individual was not heterozygous at this SNP). If there were multiple heterozygous SNP in a peak/exon, the maternal counts for the SNP were summed and the paternal counts for the SNP were summed, to create one maternal count and one paternal count per feature.

QTL analysis. All SNPs within 1 mb of the midpoint of the peak/exon were tested for association with the phenotype (maternal and paternal allele counts at the exon or peak). We called these SNPs the driver SNPs (dSNPs). Two statistical tests were used to assess the significance of the relationship. The first test was used to filter for small sample sizes and/or where there were only a small number of animals heterozygous at the dSNP:

- For the first test, where: 0= reference allele and 1=alternate allele
- For each individual i that has dSNP genotype 1|0, let M_i = number of counts of the maternal allele at the phenotype
- For each individual i that has dSNP genotype 1|0, let P_i = number of counts of the paternal allele at the phenotype
- For each individual j that has dSNP genotype 0|1, let N_j = number of counts of the maternal allele at the phenotype
- For each individual j that has dSNP genotype 0|1, let Q_j = number of counts of the paternal allele at the phenotype

And $T = \sum M_i + \sum P_i + \sum N_j + \sum Q_j$ and $A = \sum M_i + \sum Q_j$

Then: $Z = \left(\frac{A}{T} - 0.5\right) \times \sqrt{4T}$ is approximately normally distributed.

P-values for the normal distribution were calculated in R. dSNPs which were significant at $P<0.0001$ were taken through to the second test.

For the second test: We calculated a linear model: $Y_i = a + bX_i$ where, for each individual i,

$$Y_i = \ln\left(\frac{maternal\ allele\ count + 10}{paternal\ allele\ count + 10}\right)$$

When individual i has dSNP genotype 0|0 or 1|1, *let* $X_i = 0$
When individual i has dSNP genotype 1|0, *let* $X_i = 1$
When individual i has dSNP genotype 0|1, *let* $X_i = -1$
P-values were calculated in R.

Results

ChIP-seq and RNA-seq data. After filtering for ChIP-seq quality (JSD>0.25) and for animals with a known sire, there were 337 RNA-seq samples and 99, 97 and 37 samples assayed for H3K4Me3, H3K4Me1 and H3K27ac respectively. Merging ChIP-seq bam files resulted in 97-199 million mapped reads for each mark and 400-700 thousand consensus peaks.

Allele Specific QTL analysis. There were more than 48,000 exons and more than 200,000 peaks tested for allele specific expression QTL and asbQTL (Table 1). Of these, 15,308 exons representing ~6,500 genes had significant aseQTL and between ~7,000 and 20,000 peaks had significant asbQTL.

Overlap between gene expression and histone modification QTL. There was more overlap between variants which were asbQTL and variants which were aseQTL than expected by chance (OR>3.1 Table 2). In total, 31% of aseQTL variants were also asbQTL variants for one of the three marks. SNPs which were both aseQTL and asbQTL variants had the same direction of effect on the peak height or gene expression more than 57% of the time (Table 2).

If the same dSNP was associated with differences in peak height and exon expression, we propose that it is possible to link a peak to the exon it is potentially regulating. We found between 5,350 and 24,423 peak-exon pairs which shared a significant dSNP in the same direction (Table 3), representing ~1,500-3,000 genes potentially regulated by ~2,000-6,000 functional regions.

Table 1. Summary of allele specific analysis results.

	Samples	Number of phenotypes tested	Number of phenotypes with at least one significant dSNP (*P*<0.0001)	Number of significant (*P*<0.0001) dSNPs
RNA	337	48,456	15,308	1,999,869
H3K4Me3	96	224,183	15,918	1,055,069
H3K4Me1	95	220,588	20,555	981,026
H3K27ac	36	283,235	7,963	154,739

Table 2. The number of SNPs identified as aseQTL or asbQTL variants, the overlap between them (odds ratio in brackets) and the total percentage where the direction of the SNP effect was the same.

	Total dSNPs tested	aseQTL	asbQTL	Overlap (odds ratio)	Same direction of SNP effect (%)
H3K4Me3			1,055,069	410,022 (4.3)	61.5
H3K4Me1	13,397,022	1,999,869	981,026	339,821 (3.4)	57.4
H3K27ac			154,739	53,829 (3.1)	66

Table 3. The number of peaks and exon with shared significant dSNP and the distance between the peak-exon pairs.

	Number of peak-exon pairs	Median distance (bp) between peak and exon	Peaks	Exons	Average exons per peak	Average peaks per exon
H3K27ac	5,350	344,811	2,017	2,949	3.1	2.2
H3K4Me3	24,423	340,445	6,052	7,466	4.8	4.1
H3K4Me1	19,541	375,841	6,145	6,603	3.7	3.7

Discussion

This study presents the largest survey of histone modifications in the bovine genome to date. Utilising the benefits of allele specific analysis to identify cis-QTL, we present millions of SNPs which were potentially influencing regulatory regions in the mammary gland of dairy cows. Further, we show that many of these SNPs were also affecting gene expression in the same tissue. By overlapping these two analyses we hypothesise that these SNPs are causal variants for differences in gene expression by affecting regulatory regions which are controlling gene expression. If that's the case, peaks which share a significant QTL variant with an exon are potentially regulating that exon. Therefore, we identified putative regulatory regions for between 1,500 and 3,000 genes in the bovine genome.

There are some major limitations to this study. While multiple studies have found correlations between peak height and gene expression (Cheng *et al*., 2012, Prowse-Wilkins *et al*., 2021), there are few that have proven this relationship is causal. Therefore, it may be a coincidence that QTL for both traits overlap, although there is 3-4 more times overlap than expected by chance. Secondly, allele specific QTL studies are subject to linkage disequilibrium (LD), meaning the significant SNP may be tagging the causal variant rather than be the causal variant themselves. Lastly, we provide no explanation for how the variants identified in this study are altering histone modification peak height. Future work will ascertain whether the variants fall within regulatory regions such as transcription factor binding sites and, ideally, would validate the potential causal variants identified.

References

Cano-Gamez E and Trynka G. (2020) Front. Genet. 11:424. https://doi.org/10.3389/fgene.2020.00424

Chen J., Rozowsky J., Galeev T.R., Harmanci A., Kitchen R. *et al* (2016) Nat Comm 7,11101 https://doi.org/10.1038/ncomms11101

Cheng C. and Gerstein M. (2012) Nuc Acids Res 40(2),553-568. https://doi.org/10.1093/nar/gkr752

Li H., Handsaker B., Wysoker A., Fennell T., Ruan J. *et al* (2009) Bioinf 16,2078-2079. https://doi.org/10.1093/bioinformatics/btp352

Littlejohn M.D., Tiplady K., Fink T.A., Lehnert K., Lopdell T.J. *et al.* (2016) Sci Rep 6, 25376. https://doi.org/10.1038/srep25376

Lopdell T.J., Tiplady K., Struchalin M., Johnson J.J., Keehan M. *et al.* (2017) BMC Genomics 18, 968. https://doi.org/10.1186/s12864-017-4320-3

Prowse-Wilkins C.P., Wang J., Xiang R., Garner J.B., Goddard M.E. *et al.* (2021) Front. genet. 12:664379. https://doi.org/10.3389/fgene.2021.664379

Reynolds E.G.M., Neeley C., Lopdell T.J., Keehan M., Dittmer K. *et al.* (2021) Nat Genet 53, 949–954. https://doi.org/10.1038/s41588-021-00872-5

Schadt E.E., Lamb J., Yang X., Zhu J., Edwards S. *et al.* (2005) Nat Genet 37, 710-717 https://doi.org/10.1038/ng1589

553. H3K27me3 predicted regulation of gene expression linked to calcium quantification in Nelore muscle

J. Afonso[1], W.J. Shim[2,3], M. Boden[2], N. Palpant[3], M.R.S. Fortes[2], W.J.S. Diniz[4], A.O. Lima[5], M.I.P. Rocha[6], T.F. Cardoso[1], J.J. Bruscadin[6], C.F. Gromboni[7], A.R.A. Nogueira[1], G.B. Mourão[8], A. Zerlotini[9], L.L. Coutinho[10] and L.C.A. Regitano[1*]

[1]Embrapa Pecuária Sudeste, São Carlos, São Paulo, Brazil; [2]School of Chemistry and Molecular Biosciences, The University of Queensland, Brisbane, Australia; [3]Institute for Molecular Bioscience, The University of Queensland, Brisbane, Australia; [4]Department of Animal Sciences, Auburn University, Auburn, AL, USA; [5]Division of Medical Genetics, Department of Genome Sciences, Department of Medicine, University of Washington, Seattle, WA, USA; [6]Post-graduation Program of Evolutionary Genetics and Molecular Biology, Federal University of São Carlos, São Carlos, São Paulo, Brazil; [7]Bahia Federal Institute of Education, Science and Technology, Ilhéus, Brazil; [8]Department of Agroindustry, Food and Nutrition, University of São Paulo/ESALQ, Piracicaba, Brazil; [9]Bioinformatic Multi-user Laboratory, Embrapa Informática Agropecuária, Campinas, São Paulo, Brazil; [10]Department of Animal Science, University of São Paulo/ESALQ, Piracicaba, São Paulo, Brazil; luciana.regitano@embrapa.br

Abstract

The histone modification H3K27me3 is linked to the regulation of different cell states. However, no study has focused on the influence of H3K27me3 over gene expression in Nelore cattle phenotypes. We used the TRIAGE model (a predictive method based on a score calculated using the inverse relationship of H3K27me3 and gene expression), modified here to analyze Nelore muscle samples contrasting for calcium (Ca) content. The aim was to identify putatively regulated genes, called Discordantly Ranked Genes (DRGs), related to this phenotype. We identified 209 DRGs, from which 10 (e.g. *COMP, MAFB* and *ITAG11*) were also previously reported as differentially expressed genes between animals contrasting for Ca content and 44 (e.g. *KCNJ5*, *IGF2* and *HOXA10*) underlie enriched pathways related to signaling, synaptic, neural, regulatory and addiction events, all pathways related to Ca content. Our approach can identify candidate genes to be regulated in bovine muscle according to Ca content.

Introduction

Despite the previous identification of genes and genomic regions involved with the calcium (Ca) concentration in Nelore cattle muscle (Tizioto *et al.*, 2014; Diniz *et al.*, 2019; Afonso *et al.*, 2019; Afonso *et al.*, 2020), the mechanisms regulating these genes are not yet fully understood. Amongst gene regulators are the histone modifications (HM), which can silence or activate gene expression, depending on the type of the modification (Kimura *et al.*, 2013). One of these HM is the histone-3-lysine-27-trimethylation (H3K27me3), which usually causes gene inactivation through heterochromatin formation (Cai *et al.*, 2021; Young *et al.*, 2011). To our knowledge, no study focused on the influence of H3K27me3 over gene expression in Nelore muscle tissue, the most common beef cattle in Brazil, or related it to phenotypes in bovine. One way of studying H3K27me3 is through a prediction methodology called TRIAGE (Shim *et al.*, 2020). TRIAGE is an *in silico* model to analyze the impact of this HM over gene expression of cell identity genes, based on the general repressive tendency mediated by H3K27me3 previously observed in human genes. As H3K27me3 still allows genes to be activated in different cell states (Cai *et al.*, 2021), e.g. differences in the mineral amount, we hypothesize that there is a relationship between H3K27me3 distribution in the genome and the variation in Ca concentration in muscle. Here we explore evidence of this possible relationship.

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_553, © J. Afonso *et al.* 2022

Materials & methods

Samples, phenotypes and expression data. From a subset of 113 *Longissimus thoracis* samples out of a Nelore steer population described in detail by Tizioto *et al.* (2013), we used the Genetic Estimated Breeding Value (GEBV) calculated using the spectrometry quantification of Ca raw values (Tizioto *et al.*, 2015) to select two contrasting sample groups of six samples each regarding Ca concentration. The RNA-seq data from the contrasting samples was previously obtained as described by Oliveira *et al.* (2018). The expression data were mapped to the bovine reference genome ARS-UCD 1.2 using the STAR software v.2.5.4 (Dobin *et al.*, 2013) as previously described (Diniz *et al.*, 2016). We normalized the expression data with the VST function from DESeq2 software (Love *et al.*, 2014) and adjusted for batch effects with the ARSyNseq function of the NOISeq software v.2.16.0 (Tarazona *et al.*, 2015).

TRIAGE methodology. Based on the TRIAGE methodology (Shim *et al.*, 2020), we calculated the discordant score (DS) for each gene in each sample extrapolating the repressive tendency score from human data (Shim *et al.*, 2020) to orthologous genes in *Bos taurus*. After that, the genes from each sample were ranked by the DS to identify genes putatively being regulated in Nelore muscle in regard to Ca content, called DRGs.

Discordantly ranked genes (DRGs). To identify the DRGs, we compared the ranks between the contrasting groups using the RankProd R package (pfp<0.01) (Hong *et al.*, 2006).

Functional enrichment analysis and integration data. A functional enrichment analysis with the DRGs was performed using the STRING app plugin (Doncheva *et al.*, 2016) to identify enriched metabolic pathways and biological functions. To incorporate a new layer of information in the gene regulation in muscle-related to Ca content, the DRGs were compared with previously published differentially expressed genes (DEGs) between the same contrasting sample groups for Ca (Afonso *et al.*, 2019).

Results

DRGs. We identified 209 DRGs for Ca content. Biologically, DRGs for Ca content are genes presenting their expression varying more between contrasting groups for this mineral in relation to the repressive tendency prediction made by the TRIAGE method. This means that these genes have a high tendency to be repressed by H3K27me3 in several tissues and present different transcription abundance between sample groups, suggesting that they are predicted to be regulated in Nelore muscle in relation to Ca mass fraction.

Functional enrichment analysis and integration. From the 209 DRGs for Ca content, 44 genes take part in enriched pathways (Figure 1). The enriched pathways are related to signaling, synaptic, neural, regulatory and addiction events. We identified known interactions among the coded proteins of most DRGs taking part in enriched pathways. Considering all DRGs, 10 genes were previously published DEGs for Ca content (i.e. *TMEM233, SCN3B, COMP, MAFB, C1QB, ITAG11, MKX, PDGFRA, KY* and *IGF2*) (Afonso *et al.*, 2019).

Discussion

The TRIAGE methodology can be applied to any cell type or species where H3K27me3 is a conserved epigenetic mechanism for the regulation of cell identity (Shim *et al.*, 2020). The results from our Nelore samples pointed to genes already related to muscle Ca mass fraction as presenting different rankings by the DS. These genes have predicted divergences in the regulation by this HM, having a high chance to be regulated by it in some tissues, such as muscle. Our results suggest that differences in DS between samples from the same tissues can be related to Ca quantification. There is a lack of studies relating H3K27me3 with phenotypes other than developmental stages or diseases, less so to mineral quantification. However, in

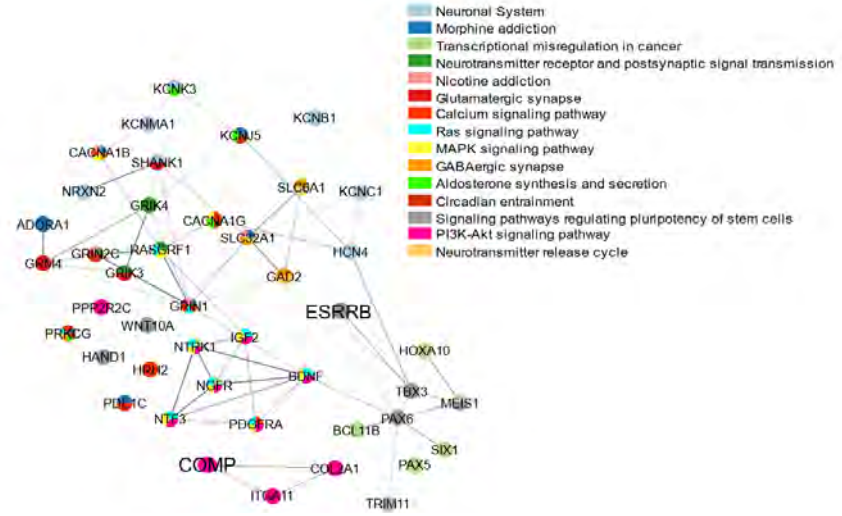


Figure 1. Protein-protein interaction network of the discordantly regulated genes (DRGs) taking part in enriched pathways.

Arabidopsis thaliana, it is known that this HM can contribute to the transcriptional regulation of TFs that are key regulators in iron uptake in a dependent Fe deficiency response. Hypothetically, H3K27me3 regulates the induction of these key regulator TFs target genes under Fe deficiency to facilitate iron acquisition while preventing iron overload (Park *et al.*, 2019), thus indicating that a relationship between Ca quantification and H3K27me3 could also be hypothesized.

Our results indicate genes putatively regulated in bovine muscle according to Ca concentration. Regarding the DRGs here identified that were described as DEGs (Afonso *et al.*, 2019), *MAFB* and *MKX* are also known bovine transcription factors (TFs) linked to Ca content. In humans, *MAFB* is an important TF for parathyroid embryogenesis and its expression correlates to extracellular Ca concentrations (Fabbri *et al.*, 2020). *MKX* is a TF linked to tendon repair in mice (Milet and Duprez, 2015), a tissue that increases its Ca content in response to mechanical deformation due to tensile strain, fluid flow, compression and vibration (Wall and Banes, 2005). Additionally, the enriched pathways mostly related to signaling, synaptic, neural, regulatory and addiction events, have known links to Ca content, showing the relationship among the DRGs and Ca content. Ca initiate synapses by acting in the release of neurotransmitters (Sudhof, 2012), thus linking it to addiction pathways also enriched here. More studies are needed to further understand the possible relationship between the H3K27me3 gene regulation and Ca content in Nelore muscle; however, this study proposes a new perspective to incorporate a layer of information regarding gene regulation.

References

Afonso J., Coutinho L.L., Tizioto P.C., Diniz W.J.S., Lima A.O. *et al.* (2019) Scientific Reports 9:12715. https://doi.org/10.1038/s41598-019-49089-x

Afonso J., Fortes M.R.S., Reverter A., Diniz W.J.S., Cesar A.S.M *et al.* (2020) Scientific Reports 10:8436. https://doi.org/10.1038/s41598-020-65454-7

Cai Y., Zhang Y., Loh Y.P., Tng J.Q., Lim M.C. *et al.* (2021) Nature Communications 12:719 https://doi.org/10.1038/s41467-021-20940-y

Da Silva Diniz W.J., Coutinho L.L., Tizioto P.C., Cesar A.S.M., Gromboni C.F. *et al.* (2016) Plos One 11(8): e0161160 https://doi.org/10.1371/journal.pone.0161160

Diniz W.J.S., Mazzoni G., Coutinho L.L., Banerjee P., Geistlinger L. *et al.* (2019) Frontiers in Genetics 10: 210 https://doi.org/10.3389/fgene.2019.00210

Dobin A., Davis C.A., Schlesinger F., Drenkow J., Zaleski C. *et al.* (2013) Bioinformatics 29(1): 15-21 http://doi.org/10.1093/bioinformatics/bts635

Doncheva N.T., Horris J.H., Gorodkin J., Jensen L.J. (2019) J Proteome Res 18(2): 623-632 http://doi.org/10.1021/acs.jproteome.8b00702

Fabbri S., Zonefrati R., Galli G., Gronchi G., Perigli G. *et al.* (2020) J Endocr Soc 4(7): bvaa058 http://doi.org/10.1210/jendso/bvaa058

Hong F., Breitling R., McEntee C.W., Wittner B.S., Nemhauser J.L. *et al.* (2006) Bioinformatics 22(22): 2825-2827 https://doi.org/10.1093/bioinformatics/btl476

Kimura H. *et al.* (2013) J Hum Gene 58(7):439-45 http://doi.org/10.1038/jhg.2013.66

Love M., Huber W., Anders S. *et al.* (2014) Genome Biology 15:550 https://doi.org/10.1186/s13059-014-0550-8

Milet C., Duprez D. (2015) Ann Transl Med 3(suppl 1): S33 http://doi.org/10.3978/j.issn.2305-5839.2015.03.64

Oliveira G.B., Regitano L.C.A., Cesar A.S.M., Reecy J.M., Degaki K.Y. BMC Genomics 19:126 http://doi.org/10.1186/s12864-018-4514-3

Park E., Tsuyuki K.M., Hu F., Lee J., Jeong J. (2019) Front Plant Sci 10:627 http://doi.org/10.3389/fpls.2019.00627

Shim W.J., Sinniah E., Xu J., Vitrinel B., Alexanian M. *et al.* (2020) Cell Syst 11(6): 625-639.e13 http://doi.org/10.1016/j.cels.2020.11.001

Sudhof T. (2012) Cold Spring Harb Perspect Biol 4(1): a011353 http://doi.org/10.1101/cshperspect.a011353

Tarazona S., Furió-Tarí P., Turrá D., Di Pietro A., Nueda M.J. *et al.* (2015) Nucleic Acid Res 43(21): e140 http://doi.org/10.1093/nar/gkv711

Tizioto P.C., Decker J.E., Taylor J.F., Schnabel R.D., Mudadu M.A. *et al.* (2013) Physiological Genomics 45: 1012-1020 https://doi.org/10.1152/physiolgenomics.00066.2013

Tizioto P.C., Gromboni C.F., Nogueira A.R.A., Souza M.M., Mudadu M.A. *et al.* (2014) Meat Science 96(1): 436-440 https://doi.org/10.1016/j.meatsci.2013.08.001

Tizioto P.C., Taylor J.F., Decker J.E., Gromboni C.F., Mudadu M.A. *et al.* (2015) Genetics Selection Evolution 47(15) https://doi.org/10.1186/s12711-014-0083-3

Wall M.E., Banes A.J. (2005) J Musculoskelet Neuronal Interact 5(1):70-84 https://pubmed.ncbi.nlm.nih.gov/15788873/

Young M.D., Wilson T.A., Wakefield M.J., Trounson E., Hilton D.J. *et al.* (2011) Nucleic Acids Res 39(17): 7415-27 http://doi.org/0.1093/nar/gkr416

554. Comparative analysis of CAGE-Seq across tissues reveals transcription start sites unique to cattle

M. Salavati[1*], R. Clark[2], D. Becker[3], C. Kühn[3,4], G. Plastow[5], G.C.M. Moreira[6], C. Charlier[6,7] and E.L. Clark[1] on behalf of the BovReg consortium

[1]The Roslin Institute, University of Edinburgh, EH25 9RG, Edinburgh, United Kingdom; [2]Genetics Core, Edinburgh Clinical Research Facility, The University of Edinburgh, EH4 2XU, Edinburgh, United Kingdom; [3]Institute of Genome Biology, Research Institute for Farm Animal Biology (FBN), 18196 Dummerstorf, Germany; [4]Faculty of Agricultural and Environmental Sciences, University Rostock, 18059 Rostock, Germany; [5]Livestock Gentec, Department of Agricultural, Food and Nutritional Science, University of Alberta, T6G 2R3, Edmonton, Canada; [6]Unit of Animal Genomics, GIGA Institute, University of Liège, 4000 Liège, Belgium; [7]Faculty of Veterinary Medicine, University of Liège, 4000 Liège, Belgium; mazdak.salavati@roslin.ed.ac.uk

Abstract

Recently, expressed genomic information e.g. expression quantitative loci (eQTL), have been utilised in both dairy and beef breeding to link genotype to phenotype. The fine mapping of promoters (transcription start sites; TSS) and enhancers (divergent amplifying segments local to TSS) in different populations of cattle across multiple tissues can help to unravel breed- and tissue-specific phenotypes. This study provides a high-resolution analysis of TSS complexity including drivers of transcription using Cap Analysis Gene Expression (CAGE) sequencing in 3 populations of cattle (Holstein, Charolais × Holstein and Kinsella beef composite; KC). We identified 55,139 TSS and 2,543 TSS-Enhancer regions shared by 3 populations along with population-specific sets. Comparative analysis revealed a cattle-specific set of TSS. Improving our understanding of the genomic architecture of gene expression and regulation in this way can help to inform the application of genomic technologies in breeding programmes for cattle.

Introduction

Mapping of the actively transcribed regions of the genome helps identify the drivers of gene expression, regulation and phenotypic plasticity. TSS within promoter regions provide information about how genes controlling traits of interest are expressed and regulated. These TSS can be specific to a tissue or pervasive across the breed/species. To improve TSS and enhancer annotation of the current cattle genome (ARS-UCD1.2), we used CAGE sequencing to create base-pair resolution expression profiles of TSS and their co-expressed short-range enhancers (<1 kb) in the ARS-UCD1.2_Btau5.0.1Y reference genome (1000bulls run7). We utilised publicly available CAGE datasets for human, chicken, mouse, rat, macaque monkey, dog and sheep for a cross-species comparative analysis of TSS and TSS-Enhancers and their co-expressive interaction. Using comparative analysis this study aims to provide a cattle-specific set of TSS and TSS-Enhancers in multiple tissues from dairy (Belgian Holstein), and cross-bred composite beef (KC) and beef/dairy cross (Charolais × Holstein) cattle.

Materials & methods

CAGE sequencing and analysis pipeline. A total of 105 samples from 24 different tissues were collected from 6 animals (3 populations, different ages and 2 sexes). Tissue representation for each breed was as follows: dairy (Holstein, n=43), composite beef (KC-composite, n=31) and beef/dairy cross (Charolais × Holstein, n=31). CAGE libraries were prepared as for (Salavati *et al.*2020) with some modifications of the barcodes to facilitate sequencing on an Illumina NextSeq 550 (50nt SE) and mapped against the ARS-UCD1.2_Btau5.0.1Y assembly using the nf-cage pipeline (BovReg 2021).

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_554, © M. Salavati *et al.* 2022

Transcription start site and enhancer prediction analysis. The base pair (bp) resolution output files were processed by CAGEfightR v.15.1 for cattle. For comparative analysis a CAGE dataset from sheep (Salavati *et al.* 2020) was mapped to ARS-UI_Rambv2.0 and also processed by CAGEfightR v.15.1. The uni- and bi-directional clustering algorithm of the CAGEfightR was used to predict putative TSS and TSS-Enhancers within each tissue individually (i.e. tissue-specific) and all tissues combined (species and population). Co-expression of predicted TSS and TSS-Enhancer regions was tested using a Kendall correlation test (P<0.05 sig.). Finally a hierarchical clustering of the super-enhancers was performed to identify longer (>1 kb) regulatory stretches of the genome.

Re-analysis of the Fantom5 CAGE datasets. Already mapped CAGE datasets from human (152), rat (13), mouse (17), chicken (32), dog (13) and Macaque monkey (15) were obtained from (Fantom5 2021). After conversion to bp resolution tracks similar analysis to the cattle and sheep was performed to identify putative TSS and TSS-Enhancer regions.

Results

An average (± SE) of 15.5±0.53 million reads per CAGE library were produced and a 94% mapping rate was achieved for the cattle dataset. Initially more than 4.5 million putative TSS and 57,412 TSS-Enhancer clusters were identified in total (min 10 read counts/cluster – Figure 1).

We considered a putative region, real/reproducible only when present across at least 2/3rd of the tissues. After such filtering 55,139 TSS and 2,543 TSS-Enhancers were detected for cattle (Table 1). Tissue-specific analysis captured, on average per tissue (± SE), 253,852±24,713 TSS clusters, 41.6% of which were novel. On average per tissue 12,138±889 TSS-Enhancer clusters were captured (27.6% novel). Population-specific analysis showed differences in TSS coordinates and expression levels between the populations. The highest number of population-specific TSS were in the KC-composite (3,102) followed by 1,152 in Holstein and

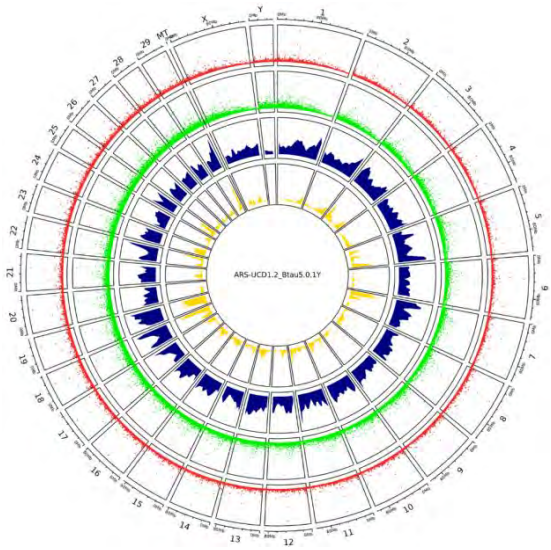


Figure 1. Distribution of uni-directional (TSS) and bi-directional (TSS-Enhancer) CAGE clusters within the cattle genome (ARS-UCD1.2_Btau5.0.1Y). The TSS clusters (red), TSS-Enhancer (green), significant positive (blue) and negative (yellow) correlation between co-expressed Enhancer and TSS(s) are shown in genomic tracks. The height of the tracks shows scaled expression or correlation coefficients (0-1).

1,092 in Charolais × Holstein. The same pattern was observed in the TSS-Enhancer clusters (419 in KC-composite, 286 in Charolais × Holstein and 202 in Holstein). A comparison of the sheep dataset and the cattle dataset is shown in Table 1. The higher number of identified regions in cattle is partly due to inclusion of multiple tissue samples from 3 populations (compared to a single animal for the sheep data).

The correlation (Kendall) analysis between a bi-directional cluster and multiple uni-directional clusters (significantly co-expressed TSS and TSS-Enhancers) showed on average 3.73±0.05 (median 3) and 22.8±0.02 (median 18) links in the sheep (3,694 regions) and cattle (641,028 regions) datasets respectively. The Kendall estimates of these significant co-expressed links were 0.547 and 0.363 for the sheep and cattle tissues respectively. The analysis of the super enhancers (stretches of bi-directional CAGE clusters) encompassing multiple enhancers within each stretch in the sheep dataset resulted in 2 super enhancer predictions. These stretches were formed of 6 TSS-Enhancer clusters (total 1,493) with the longest stretch of 5,172bp [cluster of 3 enhancers]. The similar analysis of the cattle CAGE dataset from 3 populations resulted in 20 super enhancer stretches from 65 TSS-Enhancer clusters (total 2,543). The longest stretch was 25,678bp which contained 6 TSS-Enhancers. The number of discovered super enhancer regions overall was higher in the cattle dataset (3 populations) compared to the sheep (single breed). A multi species metrics for available CAGE datasets is shown in Table 2.

Discussion

Recently there has been an influx of breed-specific reference assemblies in livestock species (Crysnanto *et al.*2021, Li *et al.*2019). To harness the full potential of these assemblies, identifying breed or population-specific promoter complexity is valuable within cattle genomic breeding programmes (Clark *et al.* 2020). To this aim, we used CAGE sequencing and publicly available datasets in a comparative analysis, to create further annotation underpinning the phenotypic diversity seen in 3 diverse cattle breeds. We identified more than 55k unique putative TSS coordinates (38% un-annotated regions of the cattle genome) compared to 27k TSS in sheep (7% unannotated regions). The promoter plasticity captured by the usage

Table 1. Mapped and annotated CAGE uni-directional clusters (TSS) in sheep and cattle using reference assembly gene models (min 2/3rd tissue representation).

Genomic region	ARS-UI_Ramb_v2.0				ARS-UCD1.2_Btau5.0.1Y			
	Novel	Anno[1]	Total	%[2]	Novel	Anno	Total	%
promoter	0	13,372	13,372	39.4	694	8,740	9,434	17.1
proximal	0	944	944	6	252	2,807	3,059	5.5
fiveUTR	0	873	873	6.7	141	2,901	3,042	5.5
threeUTR	0	2,197	2,197	7.9	38	2,257	2,295	4.2
CDS	0	4,513	4,513	16.1	363	8,977	9,340	16.9
exon	0	295	295	1.9	0	0	0	0.
intron	0	2,386	2,386	10	485	8,100	8,585	15.6
antisense	1,034	0	1,034	4.2	5,991	17	6,008	10.9
intergenic	1,397	0	1,397	7.9	13,315	61	13,376	24.3
Total TSS	2,431	24,580	27,011	100	21,279	33,860	55,139	100
Total TSS-En[3]	34	1,459	1,493		575	1,968	2,543	
Annotated genes/transcripts			13,771 / 45,298				13,251 / 24,756	

[1] Annotated using the reference assembly gff3 track.

[2] Percentage calculated based on total per genomic region category / total TSS clusters.

[3] TSS-Enhancers (Bi-directional clustering).

Table 2. Comparison of the mapped TSS and annotated genes identified in other CAGE datasets (Fantom5, OvineFAANG and BovReg). Column 'Genes' corresponds to only the genes that were annotated using the CAGE data (using the 2/3rd representation rule). The table is sorted (descending) by the number of unique TSS identified in each genome.

Species	Genome	TSS↓	Genes
Human	hg38	209,911	31,184
Mouse	mm10	164,672	30,501
Cow	ARS-UCD1.2_Btau5.0.1Y[1]	55,139	13,251
Chicken	galGal5	32,015	7,759
Rat	rn6	28,497	13,719
Sheep	Oar rambouillet v1.0[2]	28,148	13,912
Sheep	ARS-UI_Ramb_v2.0[3]	27,011	13,771
Rhesus monkey	rheMac8	25,869	8,047
Dog	canFam3	23,147	5,288

[1] Ensembl gff3 annotation v103 track lifted over to 1000bulls reference genome.

[2] NCBI RefSeq gff3 annotation v100.

[3] NCBI RefSeq gff3 annotation v104.

of 3 divergent populations showed population-specific dominant TSS for the same genes across the breeds e.g. a *DGAT1* Holstein-specific TSS (overlapping milk yield QTLs) without expression in other breeds or *PLAG1* Charolais-specific TSS-Enhancer region (overlapping cow weight QTLs) that was not present in the other 2 populations. We will also compare the sheep and cattle datasets with other publicly available TSS and Enhancer genomic tracks to further identify regulatory regions specific to the cattle genome. Such information could be used to understand how the genome controls traits in different species, and to identify regions that are important to conserve in breeding programmes. The CAGE data produced for this study when combined with transcriptomic datasets (mRNA, miRNA and total RNA-Seq) produced by BovReg partners will provide a new comprehensive transcriptome annotation for the cattle genome. This high resolution annotation of the cattle genome will help to inform the application of genomic technologies in breeding programmes. The BovReg project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 815668.

References

BovReg's nf-cage pipeline 2021. Available at: https://github.com/BovReg/nf-cage.git

Clark E.L., Archibald A. L., Daetwyler H.D., Groenen M. A.M., Harrison P.W. *et al.* (2020). Genome Biol. 21, 1–9. https://doi.org/10.1186/s13059-020-02197-8.

Crysnanto, D., Leonard, A. S., Fang, Z. H., and Pausch, H. (2021). Proc. Natl. Acad. Sci. U. S. A. 118, e2101056118. https://doi.org/10.1073/pnas.2101056118.

Fantom5 data repository 2021. Available at: https://fantom.gsc.riken.jp/5/datafiles/basic

Li R., Fu W., Su R., Tian X., Du D. *et al.* (2019). Front. Genet. 10, 1169. https://doi.org/10.3389/fgene.2019.01169.

Salavati M., Caulton A., Clark R., Gazova I., Smith T.P.L. *et al.* (2020). Front. Genet. 11, 1184. https://doi.org/10.3389/fgene.2020.580580.

555. Transcriptomic characterisation of claw horn disruption lesions in the peripheral blood leucocytes of dairy cattle

X. Dai[1*], D. Xia[1], B.E. Griffiths[2], A. Anagnostopoulos[2], M. Barden[2], B. Li[3], M. Coffey[3], D. Werling[1], G. Oikonomou[2], G. Banos[3], and A. Psifidi[1*]

[1]The Royal Veterinary College, Hawkshead Lane, AL9 7TA, Hatfield, Hertfordshire, United Kingdom; [2]Department of Livestock and One Health, Institute of Infection, Veterinary & Ecological Science, University of Liverpool, Leahurst Campus, CH64 7TE, Neston, Liverpool, United Kingdom; [3]Scotland's Rural College, Edinburgh, Easter Bush, EH25 9RG, Midlothian, United Kingdom; xdai@rvc.ac.uk; apsifidi@rvc.ac.uk

Abstract

This study aimed to investigate the transcriptomic signature of claw horn disruption lesions (CHDL) in peripheral blood leucocytes (PBL) in dairy cattle. The PBL transcriptome of 24 cows, twelve cows susceptible (cases) and 12 resistant (controls) to CHDL were compared. Total RNA was extracted for RNA sequencing using Illumina technology. Raw reads were mapped to the cattle genome (ARS-UCD1.2) using the STAR aligner and differentially expressed gene (DEG) analysis was performed using DESeq2. Eighty-eight DEG genes were found between CHDL cases and controls; 82 were downregulated in CHDL cases. We found enrichment for Ca^{2+} transport and immune system related signalling among the DEG. In addition, the mineralisation associated genes *SPARC and CLEC3B* were downregulated and upregulated in CHDL cases, respectively. Moreover, three genes (*YAP1, DSP* and *PALLD*) associated with keratinocyte development were upregulated in the CHDL cases, which may suggest the presence of a reparative mechanism in the CHDL cases.

Introduction

Lameness is a debilitating and painful condition, which considerably compromises animal welfare and is associated with significant economic losses. Claw horn disruption lesions (CHDL), such as sole ulcers (SU), sole haemorrhage (SH), and white line disease (WLD), are responsible for most of the lameness cases in dairy cattle. CHDL have adverse impact on milk production (Amory *et al.*, 2008), reproductive performance (Charfeddine and Pérez-Cabal, 2017) and premature culling (Booth *et al.*, 2004). Considering its high prevalence within dairy herds (Murray *et al.*, 1996), CHDL is one of the dairy industry's most significant challenges. However, the aetiopathogenesis of CHDL is not entirely understood.

Systemic inflammation around calving could potentially compromise the suspensory apparatus via activating matrix metalloproteinases, which play a central role in the degradation of suspensory apparatus connective fibres. In addition, the activation of proinflammatory mediators associated with direct stimulation of lipolysis could also lead to fat mobilisation from the digital cushion and a reduction of the latter's protective properties (Kushibiki *et al.*, 2003). However, the complex role of inflammation in the development of CHDL has not yet been thoroughly addressed. Therefore, the objectives of this study were to: (1) shed new light on the role of systemic inflammation, and other potential underlying molecular mechanisms in CHDL pathogenesis; and (2) identify specific genes and pathways involved in CHDL susceptibility using RNA-Seq in peripheral blood leucocytes (PBL).

Materials & methods

Sample collection and RNA isolation. Blood samples were collected from 24 (12 cases (days in milk (DIM: 198±102)) and 12 controls (DIM: 195±100)) Holstein dairy cows (selected from more than 200 cows raised in two herds in northwest England, UK. The animals were selected based on the CHDL score, which was a combined SH, SU and WLD score. The SH, SU and WLD of individual cows were recorded for all

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_555, © X. Dai *et al.* 2022

four feet as described before (Egger-Danner *et al.*, 2015) at four time-points: pre-calving stage (mean=55.3 days before calving), calving stage (mean=5.4 days in milk), early lactation stage (mean=83.9 days in milk), and late lactation stage (mean=199.6 days in milk). Each lesion severity was calculated as the average score across all assessed feet. The lesion was calculated for each time point, so each animal had four repeated records. Therefore, the final CHDL score was the sum of the mean score of SH, SU and WLD of four feet divided by number of measurements. Foot lesion identification and recording were performed by a qualified veterinarian. The selected animals were on the extreme of the CHDL phenotype distribution (CHDL average (SD)) scores for controls and cases were 0.16±0.05 and 0.67±0.17, respectively). Controls and cases were sampled in pairs within the same herd, ensuring they were of similar age, parity, and lactation stage to minimise any systematic noise affecting the subsequent RNA-Seq analysis. Peripheral blood leukocytes (PBL) were isolated by lysis of erythrocytes as described before (Werling *et al.* 1998). Total RNA from PBL was extracted using the RNeasy Mini kit (Qiagen, Hilden, Germany), according to the manufacturer's instructions.

RNA-seq and bioinformatic analysis. RNA-Seq library preparation was performed using Illumina TruSeq ribo-zero gold kit (Illumina, San Diego, USA) following the manufacturer's instruction. Paired-end, 2×150 bp, reads were obtained using the NovaSeq 6000 instrument with NovaSeq 6000 S4 reagent kit v1.5. The quality of sequence reads was assessed using FastQC (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/) before and after quality control included quality based read trimming and adapter removal using Trimmomatic v0.39. Reads that passed quality control were mapped to the ARS-UCD1.2 bovine genome assembly using STAR v2.7.0 with default parameters. Read counts per gene were obtained using the gene annotation file corresponding to the ARS-UCD1.2 bovine genome assembly (Ensemble release 104) and the featureCounts v2.0.2. Differentially expressed genes (DEG) between cases and controls were identified using DESeq2 v1.35.0. Genes with a very low expression (gene expression £2 crossing all the samples) were excluded from further analysis. The gene counts were normalised using the default parameters ('*sizeFactors*') in DESeq2. The model used for the DEG analysis included age, parity, and herd. We applied two thresholds for differential expression a) False Discovery Rate (FDR)£5% and Log Fold Change (LFC) ³2 and a milder threshold (b) absolute LFC³2, baseMean ³10 and unadjusted *P*-value£0.05. Gene Set Enrichment Analysis (GSEA) of the selected genes sets was performed using the clusterProfiler v3.14 and the genome-wide annotation for Bovine org.Bt.eg.db v3.14. Both Gene Ontology (GO) terms and the Kyoto Encyclopedia of Gene and Genomes (KEGG) pathway enrichment were applied to identify functional clusters of genes that differ in CHDL cases and controls.

Results & discussion

The sequencing generated, on average, 60,041,254 paired reads per sample. On average, 92.6% of the read pairs uniquely mapped to the bovine reference genome except for two CHDL cases, which were highly contaminated (with *Homo sapiens* reads) and removed from the downstream data analysis. After quantifying the expression of the 27,607 genes annotated from the *Bos taurus* reference genome, we excluded 2,714 genes due to low expression. We found 88 DEG between CHDL cases and controls, with 82 downregulated in CHDL cases. These downregulated genes were enriched for 12 molecular functions (MF) mainly associated with ion channel activity, and *CSN2* and *FUT9* were primarily involved in these GO pathways. *CSN2* encodes a member of the beta casein protein family, and associated diseases include milk allergy in humans and mastitis in dairy cow. GO annotations related to this gene include calcium ion binding, enzyme inhibitor activity and interleukin 2 (IL-2) pathway activity. The protein encoded by *FUT9* belongs to the glycosyltransferase family. It is localised to the golgi apparatus, and it is responsible for the expression of the glycan Lewis X factor (CD15). CD15 is a carbohydrate antigen associated with the cell surface of glycoproteins or glycolipids, including CD11b/CD18 (Mac-1) and CR3 β-chain. It is highly expressed by mature neutrophils, mediates neutrophil adhesion to dendritic cells, and has been shown to

affect a number of neutrophil functions, including adhesion to endothelium, phagocytosis, degranulation, and the respiratory burst. Downregulation of these two genes may suggest a reparative mechanism in the CHDL case.

Using the milder threshold, we found another 33 genes that could be potentially associated with CHDL (Table 1). Downstream analysis using this gene list identified enrichment for both Ca^{2+} binding and regulation of cytosolic Ca concentration and inflammatory signalling. These results may suggest a different Ca^{2+} transportation activity within the cytosol of the CHDL cow. Ca^{2+} binding proteins participate in numerous cellular functions by acting as Ca^{2+} transporters across the cell membrane or as Ca^{2+}-modulated sensors (Elíes *et al.*, 2019). Among these genes, *CLEC3B (*also called tetranectin, C-type lectin domain family 3 member B) binds plasminogen and is highly expressed in neutrophils. It shows high homology to cartilage proteoglycan core protein and has been shown to be involved with mineralisation (probably through the Ca^{2+} dependency), during osteogenesis. Downregulation of *CLEC3B* has been reported in various diseases, and *CLEC3B*-deficient mice impeded the mineralisation process in osteogenesis (Iba *et al.*, 2013). Moreover, it has been found that adipocytes are the primary source of circulating *SPARC (*secreted protein acidic and rich in cysteine, also known as Osteonectin (ON)) and that *SPARC* inhibits adipogenesis (Kos *et al.*, 2009). *SPARC* binds to a number of different extracellular matrix components, including thrombospondin 1, vitronectin, entactin/nidogen, fibrillar collagens (types I, II, III, and V), and collagen type IV, the prevalent collagen in basement membranes. Therefore, *SPARC* has the potential to contribute to the organisation of matrix in connective tissue as well as basement membranes. More importantly, similarly to *CLEC3B*, *SPARC* is a glycoprotein in the bone that binds calcium. It is secreted by osteoblasts during bone formation, initiating mineralisation and promoting mineral crystal formation. Osteonectin also shows affinity for collagen in addition to bone mineral calcium. Thus, the observed downregulated of *SPARC* and upregulated of *CLEC3B* in CHDL cases may suggests a substantial impairment of CHDL-affected cows to form new connective tissue/cartilages, and/or may be an underlying factor for the development of exostoses on the pedal bone that cause further damage to the digital cushion and the dermis seen in cows with chronic CHDL. Furthermore, three genes (*YAP1, DSP* and *PALLD*) associated with keratinocyte development were upregulated in CHDL cases, which might suggest a reparative mechanism in the CHDL case since keratinocytes may be forming the base of the claw horn in these cows.

Table 1. The higher abundant expressed protein-coding genes between CHDL cases and control cows[1].

Ensembl gene ID	Gene	baseMean[2]	LFC[3]
ENSBTAG00000014835	*SPARC*	98.4	-4.27
ENSBTAG00000021326	*CCL20*	11.1	-3.97
ENSBTAG00000040103	*SYT8*	350	-2.77
ENSBTAG00000010366	*HCRTR1*	22.1	2.03
ENSBTAG00000038112	-	160	2.05
ENSBTAG00000019453	*PTGES*	10.9	2.62
ENSBTAG00000018331	*CLEC3B*	19.2	2.74
ENSBTAG00000039307	*YAP1*	14.6	3.60
ENSBTAG00000013281	*SPATA21*	150	4.15
ENSBTAG00000001081	*PALLD*	30.4	4.68
ENSBTAG00000015106	*DSP*	24.4	6.80
ENSBTAG00000036061	-	177.	7.14
ENSBTAG00000046595	-	14.8	10.18

[1] Gene were selected based on: LFC [3]2, unadjusted *P*-value £0.05 and baseMean[3]10.

[2] baseMean = average of normalized count values, divided by size factor calculated from DESeq2.

[3] LFC = log fold change.

Moreover, we found the gene that encodes *CCL20* was downregulated in CHDL cows. *CCL20* or macrophage inflammatory protein 3 Alpha acts as a ligand for C-C chemokine receptor *CCR6*. Signals through binding and activation of *CCR6* induces a strong chemotactic response and mobilisation of intracellular calcium ions. It is responsible for the chemotaxis of dendritic cells, effector/memory T-cells and B-cells and plays an important role at skin and mucosal surfaces under homeostatic and inflammatory conditions, as well as in pathology, including cancer and various autoimmune diseases. Moreover, three MHC 1 related protein-coding genes (*ENSBTAG00000038112*, *ENSBTAG00000036061*, *ENSBTAG00000046595*) were upregulated in CHDL cases. This suggests that a systematic inflammatory response may be present in CHDL cases. According to the current RNA-Seq analysis several DEG were uncharacterised protein-coding genes, or non-coding genes which encode structural or regulatory RNAs, especially miRNA (e.g. bta-mir-758 bta-mir-2284f, etc.). The next steps of this project would be to analyse small RNA and lncRNA data to identify other potential pathways and mechanisms underlying CHDL susceptibility. Moreover, we are currently generating RNA-Seq data from foot biopsy samples collected from the same CHDL cases and controls. The integration of the results of PBL and foot tissue RNA-Seq analysis will improve our understanding of the molecular architecture of a complex trait associated with bovine foot health. Identified and characterised variants can then be duly incorporated in the genomic evaluation process of individual animals and increase the accuracy of genomic prediction and, consequently, of genomic selection.

References

Amory J.R., Barker Z.E.,Wright J.L., Mason S.A., Blowey R.W. *et al.* (2008) Prev. Vet. Med (83):381–391.

Charfeddine N., and Pérez-Cabal M.A. (2017) J. Dairy Sci (100):653–665.

Egger-Danner C., Nielsen P., Fiedler A., Müller A., Fjeldaas T. *et al.* (2015) ICAR Atlas of Claw Health. Rome, Italy

Elíes J., Yáñez M., Pereira T.M.C., Gil-Longo J., MacDougall D.A., *et al.* (2019) Calciium Singnalling. Springer. New York, USA.

Iba K., Abe Y., Chikenji T., Kanaya K., Chiba H. *et al.* (2013) J. Bone Miner.Metab (31):399–408.

Murray, R., Downham, D.Y., Clarkson, M.J., Faull, W.B., Hughes, J.W. *et al.* (1996) Vet Rec (138): 586-591.

Werling D., Howard C., Niederer E., Straub O., Saalmüller, A. *et al.* (1998) Vet. Immunol Immunopathol. (62):185–195.

556. Evidences of INFt and ISG gene expression in leukocites of pregnant hair sheep ewes from a thermal stress experiment

J.A. Gómez-Guzmán[1], J.F. Vázquez-Armijo[2], A.L. Lara-Rivera[3], A.M. Sifuentes-Rincón[1], R. Ramírez-Valverde[4], R. Nuñez-Domínguez[4], J.C. Martínez-González[5], W. Arellano-Vera[1], G.M. Parra-Bracamonte[1*]

[1]Instituto Politécnico Nacional, Centro de Biotecnología Genómica, 88710, Reynosa, Tamaulipas, México; [2]Universidad Autónoma del Estado de México, Centro Universitario Temascaltepec, 51300, Temascaltepec, México; [3]Universidad Autónoma de Nuevo León, Facultad de Ciencias Biológicas, 66450, San Nicolás de los Garza, Nuevo León, México; [4]Universidad Autónoma Chapingo. Departamento de Zootecnia, 56230, Texcoco, México; [5]Universidad Autónoma de Tamaulipas, Facultad de Ingeniería y Ciencias. 81149, Victoria, Tamaulipas, México; gparra@ipn.mx

Abstract

Reproductive efficiency is one of the most important factors to achieve profitable ovine production systems. Reproduction during recognition of gestation might be compromised by external influences such like the thermal stress. The present study was aimed to assess the effect of this kind of stress on the gene expression of Interferon Tau (INFt) and Interferon Tau stimulated (ISG) genes. An experiment was designed considering a stress and a control group of Katahdin ewes under tropical conditions. RNA was obtained from leukocytes and a gene expression analysis performed using a Targeted RNAseq analysis. Non-clear evidence on differential gene expression between thermal stress animals and control group was found. A reclassification of groups by pregnancy revealed evidence of Interferon regulatory factor 1 (IRF-1) and INFt gene expression during the experiment independently of stress. Further validation of these outcomes will confirm expression of IRF-1 and INFt in peripherical blood nuclear cells during the maternal recognition stage.

Introduction

Reproductive efficiency is one of the most important factors to achieve profitable ovine production systems. However, acceptable reproductive levels in some regions depend on technification and environmental influences related to thermic stress. Some evidence indicate that the embryo maternal recognition viability might be compromised by unfavorable thermic conditions.

Embryo development during fertilization and implantation involves dynamic growth, cell proliferation, differentiation and many gene expression changes. Henceforth, if maternal body is exposed to thermal influences during this stage, the development of preimplantation is likely to be compromised directly by thermal stress and indirectly for influenced changes in reproductive tract (Takahashi, 2012).

The maternal recognition of the embryo in ruminants needs the embryo conceptus elongation and expression of Interferon tau (IFNt) gene (Hansen *et al.*, 2017). The INFt induces the increasing of several gene expression in endometrium following the successful implantation of the conceptus. The genes called Interferon tau stimulated genes (ISG) increase in specific stages during the first third of gestation and are considered as fundamental for uterine receptivity and implantation. ISG increasing and regulation has been identified not only in uterus, but also in corpus luteum, and circulating immune cells (Mauffré *et al.*, 2016). Other studies has demonstrated increasing of ISG also in liver, and ovine bone marrow. Furthermore, some evidence of ISG15, MX1 and MX2 ISG genes has been reported in cervix and vaginal mucosal in pregnant bovines (Kunii *et al.*, 2018). Studies suggest that some portion of uterine tissue secreted INFt goes

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_556, © J.A. Gómez-Guzmán *et al.* 2022

through the uterine vein and directly regulates ISG expression in peripheric blood cells and corpus luteum (Matsuyama *et al.*, 2012).

The objective of the present study was to study the ISG and INFt gene expression in blood nuclear cells of hair sheep ewes submitted to thermal stress in tropical conditions.

Materials & methods

An experiment was designed considering ten Katahdin ewes of same age and condition. One group (n=5) was submitted to thermal stress in a pen without shadow access during the day, and control group (n=5) with shadow access as main difference in management between groups. Feeding, water and space were the same. Experiment was carried on during late autumn season in 'El Salitre ranch' from the Centro Universitario Temascaltepec of Universidad del Estado de México, in Temascaltepec, Mexico. Both groups were managed with estrous synchronization protocols and natural mating. Assuming conception, different physiological and environmental indicators were recorded to assure standardized conditions as described by Gómez-Guzman *et al.* (2021). Peripherical blood samples were obtained at day 10, 13, 16, 18 and 23, considering the mating day as 0. Blood samples preserved in EDTA were maintained immediately in -70 °C liquid nitrogen freezing until their processing. Pregnancy test was performed at 52 days to corroborate ewe's condition. RNA extraction was performed using the RiboPure™ ThermoFisher commercial kit, following a verification and quantification using the Qubit 4.0 kit (Qubit™ dsDNA HS Assay Kit. Invitrogen, Carlsbad, CA, USA), and a final concentration by a GeneJET™ RNA Cleanup and Concentration Micro Kit (ThermoFisher, Waltham, Massachussets, EUA). Gene expression analysis was performed using the TruSeq Targeted RNA Expression in a Miniseq of Illumina following the degraded RNA protocol. All cDNA libraries were sequenced by single-end strategy (read longitude=150 bp). The expression panel included ISG15 (NM_001009735.1), MX1 (NM_001009753.1), MX2 (NM_001078652.1) and RTP4 (NM_001104930.1) genes (Mauffré *et al.*, 2016) and three new genes of importance for maternal recognition of gestation INFt (X56343.1), IRF-1 (NM_001009751.1) y IRF-2 (NM_001009740.1) (Choi *et al.,* 2001) and constitutive genes as controls.

Gene expression analysis consisted in a differential expression pipeline. From Illumina platform outcomes FASTQ files were generated for each sample, and a quality control analysis was performed by FastQC 0.11.2. Libraries adaptors were removed by trimmomatic. Alignments were performed by KALLISTO, HISAT, STAR using Ovis Aries reference genome (Nov. 2015, ISGC Oar_v4.0/oviAri4) from The UCSC Genome Browser. Quantification of reads from reference genes and expression normalization by transcripts per million (thousand) RSEM was used. Differential expression analysis was performed by Deseq2, as described by Gómez (2021).

Results

Non-clear evidence on differential gene expression was found between thermal stress and control group; however, all constitutive genes showed important amount of expression in both groups. Following this outcomes, a reclassification of groups was made independently of treatment group, considering the pregnancy test performed after the experiment as a sorting criterion. As shows in Figure 1, the transcripts abundance analysis (thousands of transcripts per million) indicated important differences in two assessed genes, IRF-1 and INFt. For IRF-1, during day 13 was significant expression differences ($P<0.0001$) between pregnant (n=4) and non-pregnant ewes (n=6). For INFt was found significant transcript abundance during days 10, 16, 18 and 23 ($P<0.0001$).

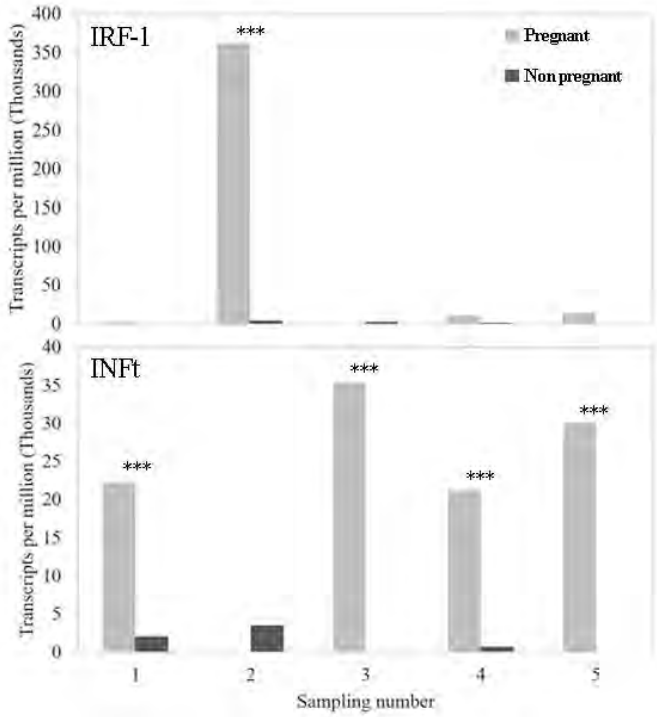


Figure 1. IRF-1 and INFt abundance of transcripts found in peripherical nuclear blood cells in pregnant and non-pregnant ewes.

Discussion

By targeted RNAseq was possible to find evidence of two maternal recognition genes expression in nuclear blood cells during the INFt expression window. Although not evidence on differential gene expression was possible to find under the original experiment arrangement (thermal stress vs control group), abundance analysis as a reliable biological transformation based on the transcript number, was able to show transcription evidences for IRF-1 and INFt genes. IRF-1 has been detected in stroma and glandular epithelia induced by INFt during early gestation (Choi *et al.*, 2001). INFt, has a key role in maternal recognition of gestation and is expected to express since day 8 to 10 in ewes, with secretion found in mononuclear trophectoderm cells from the concept (Hansen *et al.*, 2017). Transcripts from these two genes were found for the first time in leukocytes. Previously, ISG15, MX1, MX2 and RTP4 gene expression was found in these cells (Gifford *et al.*, 2008; Kose *et al.*, 2016). The potential use of these findings in gestation diagnostics has been proposed in ewes (Mauffré *et al.*, 2016) and heifers (Green *et al.,* 2010). Further validation of these outcomes will confirm the novelty of findings on INFt and IRF-1 gene expression in peripherical nuclear blood cells and their potential possible applications on pregnant ewes. Increasing of sample size and revision of technical limitations on RNA quality and availability from samples would improve future experiments for thermal stress differential expression assessments.

References

Choi Y., Johnson G.A., Burghardt R.C., Berghman L.R., Joyce, M.M. *et al.* (2001) Biol. Reprod., 65(4):1038-1049. https://doi.org/10.1095/biolreprod65.4.1038

Gomez-Guzman J.A., Vazquez-Armijo J.F., Hernandez-Melendez J., Lara-Rivera A.L and Parra-Bracamonte, G.M. (2021) Rev. Cient. LUZ. 31(1):25-30 https://doi.org/10.52973/rcfcv-luz311.01

Gómez G.J.A (2021). Expresión de Genes de Reconocimiento Materno de la Gestacionen Ovejas Bajo Estrés Térmico. Ms.C. Dissertation. Instituto Politécnico Nacional. Reynosa, México.112p.

Gifford C.A., Assiri A.M., Satterfield M.C., Spencer T.E. and Ott T.L. (2008). Biol. Reprod. 79(3): 518-524. https://doi.org/10.1095/biolreprod.108.069468

Green J.C., Okamura C.S., Poock S.E. and Lucy M.C. (2010). Anim. Reprod. Sci. 121(1-2): 24-33. https://doi.org/10.1016/j.anireprosci.2010.05.010

Hansen T.R., Sinedino L.D. and Spencer T.E. (2017). Reproduction, 154(5):F45-F59.https://doi.org/10.1530/REP-17-0315

Kose M., Kaya M.S., Aydilek N., Kucukaslan I., Bayril T. *et al.* (2016). Theriogenology, 85(6), 1161-1166. https://doi.org/10.1016/j.theriogenology.2015.11.032

Kunii H., Koyama K., Ito T., Suzuki T., Balboula A.Z., *et al.* (2018). J. Dairy Sci. 101(9): 8396-8400. https://doi.org/10.3168/jds.2017-14251

Matsuyama S., Kojima T., Kato S. and Kimura K. (2012). Reprod. Biol. Endocrinol. 10(1): 1-10. https://doi.org/10.1186/1477-7827-10-21

Mauffre V., Grimard B., Eozenou C., Inghels S., Silva L. *et al.* (2016). Animal 10(11): 1856-1863. https://doi.org/10.1017/S175173111600077X

Takahashi M. (2012). Heat stress on reproductive function and fertility in mammals. Reproductive medicine and biology, 11(1), 37-47. https://doi.org/10.1007/s12522-011-0105-6

557. Identification of differentially expressed transcription factors in Brangus skin infested with *Rhipicephalus australis*

E. Mantilla Valdivieso[1*], L. Nguyen[1], E. Ross[1], A. Raza[1], P. James[1], B. Hayes[1], N. Jonsson[2] and A. Tabor[1]

[1]The University of Queensland, Queensland Alliance for Agriculture and Food Innovation, St. Lucia, 4072, Brisbane, Australia; [2]The University of Glasgow, Glasgow, Institute of Biodiversity Animal Health and Comparative Medicine, Glasgow, United Kingdom; e.mantillavaldivieso@uq.edu.au

Abstract

Tick resistance is a desirable trait in beef cattle to minimise the incidence of cattle tick outbreaks and production loss in tropical and subtropical regions. Identification of genetic determinants of this trait will be beneficial for improved breeding programs. Using RNA-Sequencing, this study investigated gene expression data from skin of tick-infested Brangus cattle with high (n=5) and low (n=6) resistance. There were 229 significant differentially expressed genes (DEGs; FDR<0.05), of which 212 were target genes for 1,866 transcription factors expressed in skin. Regulatory impact factor analysis showed 158 significant TFs (P<0.05) of which *GRHL3*, and *DTX1* were also DEGs in the experiment. Biological pathway enrichment showed the significant TFs and DEGs feature in pathways such as response to infection, transcriptional regulation, and Notch signalling. These results highlight the application of bioinformatic approaches in the identification of key transcriptional factor that are potential drivers of differential expression of host resistance trait in cattle. Thus, these candidate genes could be further explored in gene selection programs to breed for tick resistance in crossbred cattle.

Introduction

Cattle tick infestations negatively affect animal health and cattle production in tropical and subtropical regions worldwide. Tick resistance in cattle is an attractive trait to include in breeding programs as a sustainable solution to control tick outbreaks. However, the selection of tick resistant cattle is still difficult in the field due to reliance on phenotypic assessment by tick scoring methods which are not widely standardised. The application of omics technologies, such as RNA-seq, offer an opportunity to investigate transcripts and regulatory elements associated to this trait. In this context, this paper presents data on skin gene expression from Brangus cattle of divergent host resistance phenotypes (high and low) following 12 weeks of artificial tick challenge with the cattle tick *Rhipicephalus australis*. We hypothesize it is possible to identify the key regulators among skin expressed transcription factors through bioinformatic analysis.

Materials & methods

Animals. 30 Brangus steers (~9 months old) without previous exposure to ticks were recruited for this study conducted under animal ethics approval (QAAFI/469/18). The animals were exposed to artificial infestation with approximately 10,000 tick larvae (*R. australis*) over 12 weeks, during which animals were ranked for their resistance to infestation and skin biopsies were collected on day 0 and week 12 of the trial. The number of developing adult ticks after an infestation cycle (21 days) was estimated with a tick scoring scale from 1 (<50 ticks = High resistance, HR) to 5 (>300 ticks = Low resistance, LR). Animals were ranked according to their mean tick score collected from week 8 to 15 of the trial and the top 6 and bottom 6 animals were selected for skin RNA-Sequencing experiments.

RNA isolation from skin biopsies and sequencing. RNA was extracted from frozen skin preserved in RNAlater reagent (Thermofisher, USA) using a protocol adapted from *Piper et al.* (2008) using the miRNeasy mini kit (QIAGEN, USA) as per manufacturer's instructions. RNA samples were treated with

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_557, © E. Mantilla Valdivieso *et al.* 2022

DNAse and RNA was quantified using the Nanodrop 2000 (Thermofisher, USA). The RNA RIN quality analysis was evaluated with the 2100 Bioanalyzer Instrument (Agilent Technologies, USA). The cDNA libraries were prepared with the TruSeq Stranded Total RNA kit with Ribo-depletion (Illumina, USA) and sequenced as 100 bp paired-end reads on one full S1 flowcell on the NovaSeq 6000 sequencer (Illumina, USA) through the Australian Genome Research Facility (Melbourne, AUS).

Bioinformatic pipeline. Quality control, mapping and quantification of RNA reads was performed with methods described previously (Mantilla Valdivieso *et al.* 2021) using the ARS_UCD1.2 genome and HISAT2 as the mapper tool. Gene expression was modelled with the edgeR package in RStudio (Robinson, McCarthy and Smyth 2009) using the trimmed mean of M-values (TMM) method. The linear model ~MTS+TPS+BIC+RIN was implemented to test for differentially expressed genes in the comparison of animals of low versus high mean tick score (MTS) accounting for the effects of timepoint tick score on week 12 (TPS), the proportion of individual *Bos indicus* content (BIC) (Hayes, Fordyce and Landmark 2019), and the RIN quality of the sample (RIN). Differentially expressed genes (DEGs) were considered significant based on a false discovery rate (FDR)<0.05. RIF analysis (Reverter *et al.* 2010) was implemented through the CeTF package in RStudio (Oliveira de Biagi *et al.* 2020) using filtered and normalised gene expression matrices (logCPMs) as input. The list of TFs and cofactors were obtained from the AnimalTFDBB bovine database (http://bioinfo.life.hust.edu.cn/AnimalTFDB/#!/download) to identify expressed TFs in our data. Functional enrichment of KEGG pathways was performed with the ClusterProfiler R package (Wu *et al.* 2021).

Results

A total of 17512 genes were expressed in cattle skin of low versus high host resistance Brangus cattle. EdgeR identified 229 significantly differentially expressed genes (DEGs; FDR<0.05) of which 128 were upregulated and 101 downregulated. Of these, 212 were target genes for 1866 transcription factors (TFs) expressed in skin (Figure 1A). Regulatory Impact Factor (RIF) analysis to predict key regulatory elements showed 158 significant TFs (*P*-value<0.05) based on metrics RIF1 and RIF2 metrics. Briefly, RIF1 captures TFs showing differential connectivity to DEGs found between high and low resistance phenotypes, whereas RIF2 focuses on TFs showing evidence as predictors of change in abundance of these DEGs. Furthermore, 2 transcription

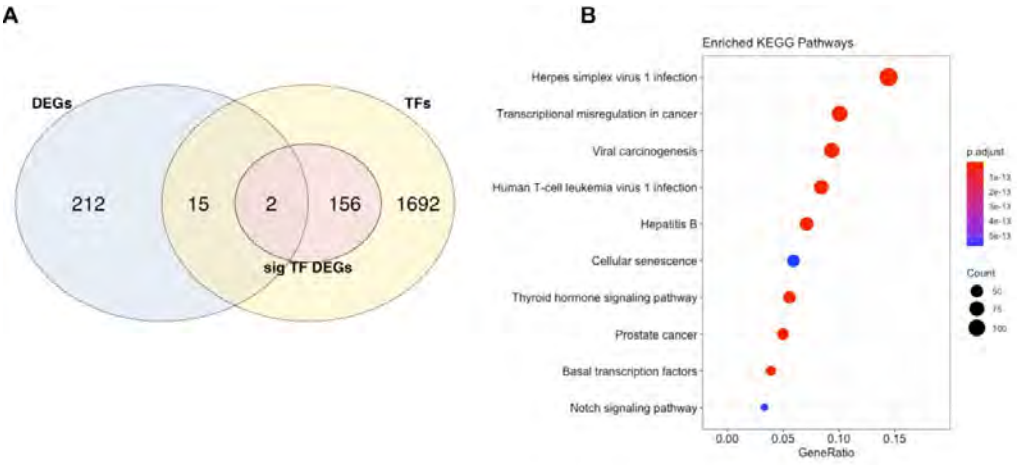


Figure 1. (A) Unique and common genes between sets of skin DEGs, expressed TFs, and significant TFs (by RIF and DEG). (B) Enriched KEGG biological pathways in the set of skin DEGs and significant TFs.

Table 1. Transcription factors identified as significant by regulatory impact factor metrics (RIF1 and RIF2, $P<0.05$) and differential gene expression (FDR<0.05).

TF symbol	Name	RIF1[1]	RIF2[1]	logFC[2]	FDR[3]
GRHL3	grainyhead like transcription factor 3	0.50	-2.68	1.24	0.048
DTX1	deltex E3 ubiquitin ligase 1	-0.48	-1.96	-2.26	0.017

[1] RIF significance value based on z-scores above 1.96 standard deviation from the mean.
[2] FC = log2 fold change between low and high tick resistance animals.
[3] FDR = false discovery rate corrected (Benjamini-Hochberg) *P*-values.

factor genes *GRHL3*, and *DTX1* were found to be significant by both RIF metrics and differential genes expression in this experiment (Table 1). Functional biological enrichment of the significant TFs and DEGs (2077 genes) showed that the most represented pathways were related to immune response to infection and transcriptional misregulation in cancer (Figure 1B). Enriched pathways with fewer genes included basal transcription factors and the Notch signalling pathway.

Discussion

Artificial challenge with *R. australis* ticks promoted divergent phenotypes of host resistance as determined by their average tick burden during the infestation experiment. Skin transcriptomic data revealed 229 significant DEGs and 158 significant TFs in the comparison of low versus high host resistance Brangus steers. Of these, *GRHL3* (grainy head like transcription factor 3) and *DTX1* (deltex E3 ubiquitin ligase 1) were identified as key regulators by both differential expression and RIF analysis, with high RIF2 metric indicating these genes are potential useful predictors of change in abundance of the skin DEGs. *GRHL3* is an epidermal transcription factor which regulates the differentiation and migration transitions of the progenitor epidermal cell (Klein *et al.* 2017). The regulation of these steps in the epidermal differential process appears to be relevant during wound healing and tissue homeostasis. The upregulation of *GRHL3* in the skin of low host resistance cattle suggests participation in the modulation of gene expression in response to tick infestation. On the other hand, *DTX1* is a transcriptional factor target of the Notch signalling cascade with possible immune effects (Brandstadter and Maillard 2019). Likewise, functional enrichment analysis captured both immune and transcriptional regulation pathways overrepresented in the significant DEGs and TFs including the Notch signalling pathway, which plays multiple functions in the innate and adaptive immune cell development and is consequently relevant in host defence. Overall, these results contribute towards uncovering some of the biology behind this complex trait putting forward a list of candidate regulator genes that may be drivers of differential expression of tick resistance in cattle. Further cross-validation of the significant DEGs and TFs with other skin or blood gene expression studies will facilitate the identification of biomarkers of tick resistance. In this way, improved phenotyping approaches with predictive markers will provide an opportunity rapidly build reference populations for genetic evaluation in breeding programs for tick-endemic regions.

References

Brandstadter J.D., and Maillard I. (2019) Open Biology 9(11). http://doi.org/10.1098/rsob.190187

Hayes B., Fordyce G., Landmark S. (2019) Proc. of the 23rd Conference of the Association for the Advancement of Animal Breeding and Genetics (AAABG), Armidale, NSW, Australia.

Klein R.H., Lin Z., Hopkin A.S., Gordon W., Tsoi L.C., Liang Y., *et al.* (2017) PLoS Genetics 13(4): e1006745. https://doi.org/10.1371/journal.pgen.1006745

Mantilla Valdivieso E., Ross E., Raza A., Hayes B., Jonsson N., *et al.* (2021) Proc. of the 24th Conference of the Association for the Advancement of Animal Breeding and Genetics (AAABG), Online, Australia.

Oliveira de Biagi C., Nociti R.P., Funicheli B.O., Ruy P.dC., Ximenez J.P.B., *et al.* (2021) *BMC genomics* 22(624). https://doi.org/10.1186/s12864-021-07918-2

Piper E.K., Jackson L.A., Bagnall N.H., Kongsuwan K.K., Lew A.E., *et al* (2018) Veterinary Immunology and Immunopathology 126(1):110-119. https://doi.org/10.1016/j.vetimm.2008.06.011

Robinson M.D., McCarthy D.J., Smyth G.K. (2009) Bioinformatics 26(1):139-140. https://doi.org/10.1093/bioinformatics/btp616

Reverter, A, Hudson N.J, Nagaraj S.H., Pérez-Enciso M., and Dalrymple B.P. (2010) Bioinformatics 26:896–904. https://doi.org/10.1093/bioinformatics/btq051

Wu T., Hu E., Xu S., Chen M., Guo P., *et al.* (2021) The Innovation 2(3):100141. https://doi.org/10.1016/j.xinn.2021.100141.

558. Unravelling regulatory variants affecting gene expression in four porcine tissues

M. Gòdia[1*], M.F.L. Derks[1,2], B. Harlizius[2], O. Madsen[1] and M.A.M. Groenen[1]

[1]*Wageningen University & Research, Animal Breeding and Genomics, Wageningen, the Netherlands;*
[2]*Topigs Norsvin Research Center, Beuningen, the Netherlands;* marta.godia@wur.nl

Abstract

The aim of this work was to identify functional variants potentially associated to complex phenotypes, to improve genomic selection. In this study, we combined genomics, transcriptomics and epigenomics data. First, we genotyped 100 pigs with the high-density genotyping array (with 660K markers). Then, we generated RNA-seq data on these animals in 4 biologically divergent tissues: liver, spleen, lung and muscle. We performed an expression GWAS (eGWAS) resulting in the identification of over 150,000 significant associations in each of the tissues. We were able to detect over 1,300 genes that showed a significant eGWAS associations. We also found an enrichment of eGWAS hits within regulatory elements such as enhancers and promoters, which indicates their pivotal role as drivers of phenotypic variation and thus are of highly relevance towards understanding the genetics of complex traits.

Introduction

Unravelling the genetic basis of economically important traits in livestock remains a challenge for the animal breeding sector, and this knowledge is essential to be able to meet the increased demands for higher sustainability.

Genome-wide association studies (GWAS) have been widely used in the sector to detect genetic variants involved in relevant traits. However, this approach alone is insufficient to pinpoint the causal variants for complex traits. A major proportion of phenotypic variation is expected to be due to variation in gene expression. Hence, additional novel approaches such as RNA-seq, that quantify gene expression levels from a given tissue type is inevitable to link genomic variation to changes in gene expression. The integration of GWAS and RNA-seq data, known as expression GWAS (eGWAS) can be used to identify genetic variants associated to gene expression levels.

The eGWAS studies provide novel regulatory variation affecting gene expression. Results can be further supported by annotated regulatory elements provided by the Functional Annotation of Animal Genomes (FAANG) consortium (FAANG data portal: https://data.faang.org/home) using different epigenetic assays. These annotated regulatory elements (RE), include among others, promoters, enhancers and TSS in a variety of mammalian tissues.

The aim of this study was to perform an eGWAS to identify regulatory variants affecting gene expression in a commercial pig breeding line in four different tissues: liver, spleen, lung and muscle. Moreover, we integrated epigenomic data recently generated by the FAANG consortium to assess the likelihood that the significant variants are found within RE such as promoters and enhancers.

Materials & methods

RNA extraction, sequencing and analysis. Four different tissues: liver, spleen, lung and muscle were extracted from 100 crossbred animals. The crossbred animals are resulting from a F1 sow (Landrace×LargeWhite) mated to a Synthetic boar line. RNA extractions were done with the Rneasy kit (Qiagen) following manufacturer's instructions. RNA samples were subjected to quality control parameters

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_558, © M. Gòdia *et al.* 2022

and subjected to library preparation with NEBNext® Ultra™ Directional RNA Library Prep Kit (New England Biolabs). Samples were sequenced at 150 bp paired-end on an Illumina 6000 sequencing platform.

The bioinformatic analysis was followed as in Kern *et al.* (2021). Briefly, RNA-seq reads were trimmed using TrimGalore! (https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/). Filtered reads were aligned to the *Sus scrofa* genome (Sscrofa11.1) with Ensembl version 104, using STAR v.2.7.8 (Dobin *et al.* 2013). Gene counts were assessed with htseq-count v.0.11.1 (Anders *et al.* 2014) and TMM normalized with EdgeR (Robinson *et al.* 2009).

The gene regulatory elements enhancers and promoters, specific for each of the tissues, were extracted from the FAANG database.

DNA extraction and genotyping. DNA from the 100 animals was extracted from spleen using standard phenol-chloroform protocol. After passing the quality control filter parameters, DNA was genotyped using the high-density Axiom™ Porcine Genotyping Array (Thermo Fisher Scientific) that queries 660K variant markers.

eGWAS and enrichment for regulatory regions. eGWAS included 87 samples with RNA-seq and genotype data that passed quality control filters. The RNA expression levels of the detected genes (average >1 CPM) were taken as quantitative traits and tested for association with the genotypes that passed quality control using a linear model in each of the tissues using GCTA v1.25.3 software (Yang *et al.,* 2011) as:

$$Y_i = \mu + SNP_i + kinship + e_i$$

where (Y_i) is the CPM gene expression as a function of the population mean (μ), fixed effect of each SNP (SNP_i), *kinship* based on the genomic relationship matrix (GRM) (Yang *et al.,* 2011 and a random residual effect (e_i). Only associations of FDR≤0.05 were considered.

To identify enrichment of eGWAS hits within RE, genomic regions of the eGWAS hits were iteratively randomized 100 times using BEDtools shuffle (Quinlan *et al.,* 2010). Then, the enrichment of the real eGWAS hit coordinates at gene regulatory regions was assessed by comparison with the overlap of the randomized locations with the regulatory regions using a one-sample T-test (one-tailed).

Results

RNA sequencing. On average, we obtained 44 M reads per sample and tissue. Of these, a mean of 99.9% passed quality filters and on average, 95.1, 92.0, 93.8 and 94.7% of the reads were mapped to the porcine genome in liver, spleen, lung, and muscle, respectively. Considering an expression average abundance >1 CPM in the 100 samples, the number of genes detected in liver, spleen, lung, and muscle, was 12,556, 12,568, 13,211 and 12,516, respectively. Yet, the overall profile was very different between tissues.

eGWAS. The eGWAS analysis resulted in several significant associations, with spleen having the highest number of associations (262,545), and lung the highest number of different genes (2,692) (Table 1).

The numbers of significant associations annotated as *cis*-regulatory variants (≤1 Mbp distance from the associated gene) were 40,211, 67,166, 44,870 and 32,070 in liver, spleen, lung, and muscle, respectively. We also identified several hotspot variants, where the variant was associated with the expression of several genes (≥10). The tissue with the highest number of hotspot variants was lung (3,820 variants), and the tissue with the lowest number was muscle (1,118 variants).

Table 1. Expression GWAS analysis results from the association between genotypes and RNAseq data in the four tissues.

Tissue	Number of significant associations	Number of variants involved	Number of genes involved
Liver	191,622	95,030	1,655
Spleen	262,545	124,353	1,301
Lung	241,047	87,664	2,693
Muscle	149,425	76,300	1,490

Some eGWAS peaks of interest included genes such as *RDH16* (Retinol Dehydrogenase 16) in liver and *TF* (Transferrin) in lung (Figure 1). *RDH16* is involved in vitamin A metabolism, which plays an important role in the regulation of feed efficiency in pigs by affecting energy metabolism, that mediates fatty acid biosynthesis and steroid hormone metabolism (Zhao *et al.*, 2016). *TF* has been linked to play a role in acute respiratory distress syndrome (Upton *et al.,* 2003).

Enrichment at regulatory elements. The location of significant eGWAS hits identified were queried for their overlap within RE regions. We found that eGWAS hits were enriched in promoters and enhancers regions (*P*-value<2.2E-16) in all the four tissues studied. In contrast, the number of eGWAS hit variants found in other REs including insulators, polycomb repressed or low signal elements, was significantly lower than the random set, supporting the role of certain regulatory elements in gene expression levels.

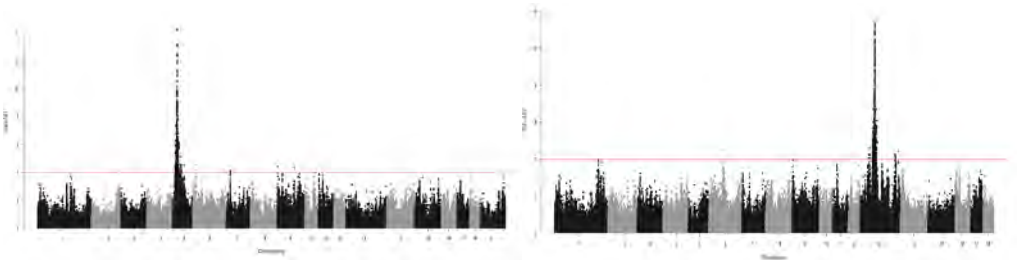


Figure 1. Manhattan plot of *RDH16* in liver (left) and *TF* in lung (right).

Discussion

In this study we report an analysis that compared the genome, transcriptome, and epigenome of biologically diverse tissues in pig. Results showed that the integrative approach of merging different techniques could help to identify variants or narrow down genomic regions involved in complex phenotypes of interest. The results can be integrated with existing QTL regions to identify potential causal variation underlying traits of interest and to understand and improve genomic selection in pig breeding programs.

Funding

This study was financially supported by the Netherlands Organization for Scientific Research (NWO) and Topigs Norsvin.

References

Anders S., Pyl P.T. and Huber W. (2015) Bioinformatics 31(2):166–169. https://doi.org/10.1093/bioinformatics/btu638

Dobin A., Davis C.A., Schlesinger F., Drenkow J., Zaleski C. *et al.* (2013) Bioinformatics 29(1): 15-21. https://doi.org/10.1093/bioinformatics/bts635

Kern C., Wang Y., Xu X., Pan Z., Halstead M. *et al.* (2021) Nat Comm 12(1):1821. https://doi.org/10.1038/s41467-021-22100-8

Quinlan A.R., and Hall I.M. (2010). Bioinformatics 26(6):841–842. https://doi.org/10.1093/bioinformatics/btq033

Robinson M.D., McCarthy D.J., and Smyth G.K. (2009) Bioinformatics 26:139–140. https://doi.org/10.1093/bioinformatics/btp616

Upton R.L., Chen Y., Mumby S., Gutteridge J.M. *et al.,* (2003) Eur Respir J 22(2):335–341. https://doi.org/10.1183/09031936.03.00075302

Yang J., Lee S.H., Goddard M.E., and Visscher P.M. (2011) Am J Hum Genet 88(1):76–82. https://doi.org/10.1016/j.ajhg.2010.11.011

Zhao Y., Hou Y., Liu F., Liu A., Jing, L. *et al.,* (2016) G3 6(11):3615–3624. https://doi.org/10.1534/g3.116.032839

559. Unravelling the transcriptomic control of growth traits in sheep during prenatal and postnatal development

S.A. Woolley*, M. Salavati and E.L. Clark

The Roslin Institute and Royal (Dick) School of Veterinary Studies, The University of Edinburgh, Easter Bush Campus, EH25 9RG, Midlothian, United Kingdom; shernae.woolley@roslin.ed.ac.uk

Abstract

Understanding gene expression in sheep can offer numerous opportunities to investigate the genomic control of health and productivity, and can help to inform future breeding programs. This study aims to identify transcriptionally active genes associated with growth through the analysis of RNA-sequencing (RNA-seq) data in Texel × Scottish Blackface sheep. Gene expression levels were estimated across tissues and developmental stages and allele-specific expression analysed. Tissue- and developmental stage-specific differences in gene expression were observed. Between day 100 of gestation and one week of age, differentially expressed genes related to muscle growth and development, were upregulated including *COL2A1* and *CDCA8* in liver and *MYH3*, *GDF5* and *COL9A2* in skeletal bicep muscle. Allele-specific expression analysis revealed genes exhibiting allelic imbalance in gene families related to growth. Integration of these results with GWAS data will assist the identification of expressed variants that are associated with growth traits in sheep.

Introduction

To meet future food production targets, and provide sufficient protein derived from red meat for a growing global population, improving the efficiency of sheep production is paramount. Growth traits in sheep are important for production and are often polygenic (Moghaddar *et al.* 2019), making the identification of trait-specific markers for genomic selection challenging. Genomic studies have accelerated the identification of key genes and variants of interest for growth traits, such as the mutation in the 3'UTR region of the myostatin gene associated with the double-muscling phenotype in Texel sheep (Clop *et al.* 2006). Investigation of the sheep transcriptome can offer insight into how gene expression influences growth traits, by measuring, for example, the up- or down- regulation of genes involved in muscle growth and differentiation through pre and postnatal development. This study aims to identify transcriptionally active genes associated with growth through the analysis of RNA-seq data across different developmental stages and tissue types in Texel × Scottish Blackface sheep.

Materials & methods

Animals. A dataset consisting of 48 samples from three prenatal and four postnatal developmental stages from Texel × Scottish Blackface sheep was generated for tissues associated with growth and development. The tissues used in this study were collected for the sheep gene expression atlas project (Clark *et al.* 2017). Prenatal time points: Day 23 whole embryo (n=3), day 23 maternal caruncle (n=2); Day 35 liver (n=3), placentome (n=2), maternal caruncle (n=2) and Day 100 liver, ovary, skeletal bicep muscle (n=4), placentome (n=2). Postnatal time points: Newborn liver (n=3), skeletal bicep muscle (n=3), ovary (n=2); One week of age liver (n=3), skeletal bicep muscle (n=3), ovary (n=2) and eight weeks of age liver (n=3), skeletal bicep muscle (n=3).

RNA extraction and sequencing. RNA was extracted from <100 mg of -80 °C frozen tissue samples using Trizol after homogenisation with ceramic beads. Samples with an RNA integrity number >6 were selected for RNA-seq and were sequenced by Novogene (Cambridge, United Kingdom) using the NEB Next® Ultra™ RNA Library Prep Kit. Libraries were sequenced on the NovaSeq 6000 platform (Illumina®, United States) at an expected depth of 60 million reads per sample with 150bp paired end reads.

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_559, © S.A. Woolley *et al.* 2022

Transcriptome analysis. After trimming and quality control, mapping of the RNA-seq data to the Oar_rambouillet_v1.0 (GCA_002742125.1; v103) assembly was performed by HISAT2 (v2.1.0) using standard parameters (Kim *et al.* 2019). To estimate RNA transcript abundance, transcript per million (TPM) counts were quantified with Kallisto (v0.44.0) (Bray *et al.* 2016). Based on the TPM estimates, gene-gene network analyses were performed in Graphia (v2.1) (Freeman *et al.* 2020). Functional gene enrichment using EnrichR (Chen *et al.* 2013) in RStudio (R v.3.6.1) was performed for the first 20 gene clusters generated from the gene-gene network analysis.

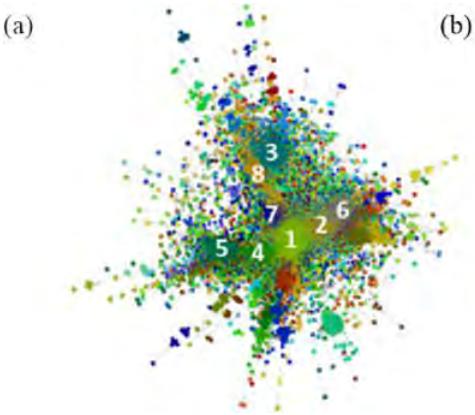
Differential gene expression. Differential gene expression between prenatal day 100 of gestation and postnatal one week of age was analysed using DESeq2 (Love *et al.* 2014), and the non-averaged transcript abundance counts per tissue type from Kallisto for liver and bicep muscle tissue. A likelihood ratio test (LRT) was used to compare a reduced model across each time point per tissue. The thresholds used for the reduced LRT model analysis included a p-adjusted value (p*adj*) <0.1 and a $\log_2$ change ($\log_2$FC) of 5. DESeq2 results were visualised in R studio using EnhancedVolcano (Blighe *et al.* 2021).

Allele-specific expression. In brief, allele-specific expression (ASE) analysis was conducted on all 48 RNA-seq samples. Reference mapping bias and ASE analysis was performed as previously described (Salavati *et al.* 2019). The Liptak score (transformed z score value) from the GeneiASE output (Edsgärd *et al.* 2016) was used as the proxy for the level of allelic imbalance.

Results

Transcriptome analysis. Gene-gene network analysis revealed large clusters of genes exhibiting tissue-specific expression in ovary, liver and skeletal bicep muscle tissues (Figure 1a and 1b). For skeletal bicep muscle tissue the genes within cluster four showed high expression levels at day 100 of gestation, and cluster five showed high expression levels in newborn to one-week-old lambs. Functional gene enrichment for cluster four showed associations of genes within this cluster with GO terms for muscle fibre organisation and development, and cluster five showed associations for muscle fibre organisation and contraction. The Myostatin (*GDF8*) gene, in cluster four, for example, which controls muscle development, showed increased expression at day 100 of gestation relative to the other developmental stages.

(a)



(b)

Cluster	Nodes	Tissue
1	6,288	All
2	1,154	Placentome and caruncle
3	767	Liver
4	591	Day 100 skeletal bicep
5	529	Skeletal bicep
6	523	Day 23 caruncle
7	426	Day 23 embryo
8	349	All

Figure 1. (a) Gene-gene network analysis showing large clusters of genes exhibiting tissue- and developmental stage-specific gene expression. Clusters are labelled according to the tissue type with the highest expression level for that cluster, with a Pearson correlation matrix (positive only) of 0.831 and a MCL granularity of 2.2, with the removal of components with a size ≤5. (b) The eight largest clusters with the number of nodes (genes) and the associated tissue type.

Differential gene expression. Rather than presenting all comparisons for differential gene expression analysis, here we focus on liver and skeletal bicep muscle tissue between one prenatal time point (day 100 of gestation) and one postnatal time point (one week of age). For liver tissue, comparing day 100 of gestation and one week of age showed 12% upregulated and 7.9% downregulated genes. In skeletal bicep muscle tissue, 23% of genes were upregulated and 23% were downregulated. Some differentially expressed genes associated with growth had a log2FC>5 (Figure 2). Genes in families associated with growth such as the myosin, TGF-beta and fibrillar collagen families were identified in skeletal bicep muscle tissue. This included the upregulation of the genes *MYH3*, *GDF5* and *COL9A2*. In liver tissue, genes in the fibrillar collagen family were also upregulated, such as the *COL2A1* gene.

Allele-specific expression. Within the top 20 highest liptak scores, for allelic imbalance, the genes *ACTB* and *ANXA2* showed a liptak score of ~946 for a day 100 of gestation sample and ~641 at a one week of age sample, respectively, in liver. For skeletal bicep muscle tissue, the genes *ACTB*, *IGF2* and *MYH2* had liptak scores of ~813 for a day 100 of gestation sample, ~408 for a day 100 of gestation sample and ~334 for a one week of age sample respectively. These genes were found to be members of gene families associated with growth, including either the actin, annexin, insulin family of polypeptide growth factors and myosin family of genes.

Discussion

Improved understanding of the sheep transcriptome provides opportunities to incorporate gene expression information in genomics enabled breeding programmes for sheep (Yuan *et al.* 2021). The initial results from this study identified large tissue-specific clusters of genes associated with growth and development. Differential gene expression analysis revealed upregulation of genes from both the collagen and myosin families, between pre and postnatal time points consistent with previous RNA-seq transcriptome studies (Clark *et al.* 2017). Differential gene expression and initial ASE analysis showed that genes within the myosin and TGF-beta families, including *MYH2*, *MYH3* and *GDF5*, exhibited significant allelic imbalance and/or upregulation in developing skeletal bicep muscle tissue, with *MYH3* associated with growth traits in cattle (Xu *et al.* 2014). Integration of our results with available eQTL datasets for tissues associated

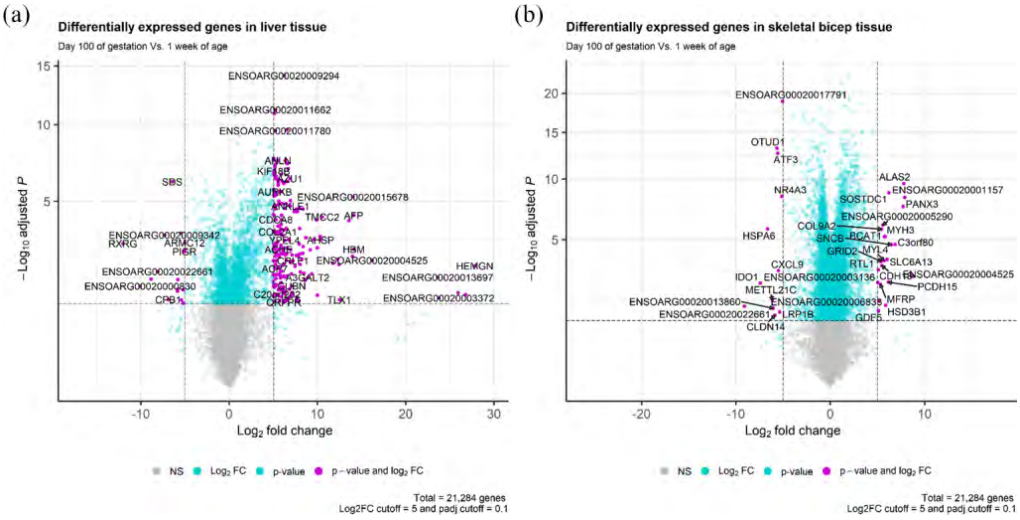


Figure 2. Volcano plots for prenatal time point day 100 of gestation and postnatal one week of age in (a) liver and (b) skeletal bicep muscle tissue, showing differential gene expression using DESeq2.

with growth (Yuan *et al.* 2021), and with information from GWAS, will identify expressed variants located in genomic regions associated with growth traits, which can be exploited in genomics-enabled breeding programmes for sheep. This work was funded by BBSRC award number BB/S01540X/1.

References

Blighe K., Rana S., Lewis M. (2021) R Packag. version 1.10.0. https://doi.org/10.18129/B9.bioc.EnhancedVolcano.

Bray N. L., Pimentel H., Melsted P. and Pachter L. (2016) Nat. Biotechnol 34:525–527. https://doi.org/10.1038/nbt.3519.

Chen E.Y., Tan C.M., Kou Y., Duan Q., Wang Z. *et al.* (2013) BMC Bioinform 14:128. https://doi.org/10.1186/1471-2105-14-128.

Clark E.L., Bush S.J., McCulloch M.E.B., Farquhar I.L., Young R. *et al.* (2017) PLOS Genet. 13:e1006997. https://doi.org/10.1371/journal.pgen.1006997.

Clop A., Marcq F., Takeda H., Pirottin D., Tordoir X. *et al.* (2006) Nat. Genet 38:813–818. https://doi.org/10.1038/ng1810.

Edsgärd D., Iglesias M.J., Reilly S.-J., Hamsten A., Tornvall P. *et al.* (2016) Sci. Rep 6: 21134. https://doi.org/10.1038/srep21134.

Freeman T.C., Horsewell S., Patir A., Harling-Lee J., Regan T. *et al.* (2020) bioRxiv. https://doi.org/2020.09.02.279349.10.1101/2020.09.02.279349.

Kim D., Paggi J.M., Park C., Bennett C., and Salzberg S. L. (2019) Nat. Biotechnol 37:907–915. https://doi.org/10.1038/s41587-019-0201-4.

Love M.I., Huber W. and Anders, S. (2014) Genome Biol 15:550. https://doi.org/10.1186/s13059-014-0550-8.

Moghaddar N., Khansefid M., Van Der Werf J.H.J., Bolormaa. S., Duijvesteijn N. *et al.* (2019) Genet. Sel. Evol. 51:72. https://doi.org/10.1186/s12711-019-0514-2.

Salavati M., Bush S.J., Palma-Vera S., McCulloch M.E.B., Hume D.A. *et al.* (2019) Front. Genet. 10:863. https://doi.org/10.3389/fgene.2019.00863

Xu Y., Shi T., Cai H., Zhou Y., Lan X. *et al.* (2014) Gene 535:106–111. https://doi.org/10.1016/j.gene.2013.11.057.

Yuan Z., Sunduimijid B., Xiang R., Behrendt R., Knight M. I. *et al.* (2021) Genet. Sel. Evol. 53:8. https://doi.org/10.1186/s12711-021-00602-9.

560. Large-scale cis-eQTL analysis of gene expression in blood of young healthy pigs using PigGTEx

L.M. Kramer[1*], J. Teng[2], K.S. Lim[1], Y. Gao[3,4], H. Yin[5], L. Bai[5], G.E. Liu[3], Z. Zhang[2], L. Fang[6], G.S. Plastow[7], C.K. Tuggle[1] and J.C.M. Dekkers[1]

[1]Iowa State University, 806 Stange Rd, Ames, IA 50011, USA; [2]South China Agricultural University, Guangzhou, 510642 Guangdong, China; [3]Animal Genomics and Improvement Laboratory Beltsville Agricultural Research Center, USDA, Beltsville, MD 20705, USA; [4]University of Maryland, College Park, MD 20742, USA; [5]Agricultural Genome Institute at Shenzhen, Chinese Academy of Agricultural Sciences, 518120 Shenzhen, China; [6]University of Edinburgh, Edinburgh, EH4 2XU, United Kingdom; [7]University of Alberta, Edmonton, AB, Canada; lmkramer@iastate.edu

Abstract

RNA sequencing and high-density genotyping or whole genome sequencing can be used in tandem to identify genomic regions associated with the control of gene expression, also known as expression quantitative trait loci (eQTL) mapping. Whole blood is a convenient tissue for many purposes due to its ease of collection and relation to health. Whole blood RNA-seq data and high-density SNP array genotypes were generated on ~1,600 healthy nursery pigs as part of a study of resilience to a polymicrobial disease challenge. This data was used for genotype imputation to whole genome sequence with the reference panel and bioinformatics pipeline developed by the PigGTEx consortium to identify cis-eQTL. Cis-eQTL were typically associated with expression of one or two genes. Chromosome 12 was enriched for cis-eQTL. The results provide insights into the control of gene expression in whole blood, allowing further study of its relationship with disease resilience.

Introduction

Selection and breeding of livestock species thrives on understanding the genetic control of economically important phenotypes. Extensive research is ongoing to better understand the genetic links between phenotypes and genetic markers through intermediate phenotypes, such as gene expression. This research utilizes various tissues to identify these intermediate phenotypes, with one of the most common tissues being blood due to its ease of collection and its relevance to health. As a tissue, blood is interesting as it interacts with other tissues across the organism, which allows it to be used for the detection of abnormalities or changes within the body, such as the onset of disease, or a change in behavior or physiological status. Blood's interconnectedness to other tissues in an organism also provides difficulties, as gene expression levels in blood can be influenced by external factors such as stress, cell composition, and temporal transcriptional differences. This makes it of utmost importance to characterize these influential factors to enhance the informativeness of blood as a tissue and its use to address a multitude of other research questions. In addition to improving blood as a source of gene expression data, it is also important to link these data to economic traits in livestock such as disease resistance/resilience, response to environmental changes, or even just production traits. The use of gene expression data and dense marker genotypes throughout the entire genome allows for eQTL mapping to better understand regions of genomic regulatory control of transcriptomic variation and response to external stressors. Identification of eQTL may indicate hotspots that are indicative of specialized regions of overarching genetic control. The purpose of this study is to capitalize on a uniquely large dataset with both genome-wide SNP genotypes and RNA-seq data to identify cis-eQTL for gene expression in whole blood of young healthy pigs for their potential use to predict future trait phenotypes for the animal.

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_560, © L.M. Kramer *et al.* 2022

Materials & methods

Animals and samples. Blood samples were collected in Tempus tubes from ~1,600 healthy Landrace × Yorkshire nursery barrows during the quarantine phase of the natural disease challenge model described by Putz *et al.* (2018) and Cheng *et al.* (2020). Piglets were brought to the nursery in batches of 60 or 75 from one of seven different breeding companies. All pigs were genotyped for approximately 650k markers across the genome.

RNA-seq. RNA-seq data were obtained as per Lim *et al.* (2021). RNA-seq libraries were generated using QuantSeq 3' mRNA Library Prep Kit FWD for Illumina and the RNA Removal Solution Globin Block (Lexogen, Austria) (Lim *et al.* 2019). QuantSeq libraries from 96 samples were multiplexed and sequenced with single end 50bp using Illumina HiSeq 3000 Sequencing Systems (Illumina, USA) for the first 887 samples from Lim *et al.* (2021) and with single-end 100 bp using NovaSeq 6000 Sequencing Systems (Illumina, USA) for an additional 707 samples. QuantSeq reads were trimmed and filtered using BBDuk. Read quality control was performed using FASTQC and mapped using STAR 2.5.3a. Mapped reads were removed if they had a zero count for more than 80% of the samples or were mapped to globin genes, *HBA* and *HBB*. The remaining reads were normalized using EdgeR (Robinson *et al.* 2010).

PigGTEx. PigGTEx is a part of the FarmGTEx project, aiming to develop a comprehensive catalog of regulatory variants in the pig transcriptome. The eQTL mapping pipeline that was used was developed by PigGTEx as a derivative of the cattle GTEx (Liu *et al.* 2021), and was designed to impute genotypes up to a whole genome sequence reference and then perform eQTL mapping using the imputed genotypes and RNA-seq data, using the following softwares: Java, Plink (Chang *et al.* 2015; Purcell and Chang), Beagle (Browning *et al.* 2018), ConformGT, Bcftools (Li, 2011), and TensorQTL (Taylor-Weiner *et al.* 2019). Imputation to whole genome sequence was performed on a chromosome basis using ~1,600 external pigs with whole genome sequence data based on the reference genome Susscrofa11.1. eQTL mapping was performed using TensorQTL by utilizing gene expression data corrected for read counts following Lim *et al.* (2021) and the imputed genotypes. Additionally, PEER estimation (Stegle *et al.* 2012) was used to obtain ten factors from the expression data and PCA was used to obtain the first ten principle components from the imputed genotypes to utilize as covariates in eQTL mapping. Cis-eQTL were defined as eQTL that were within one megabase of the gene whose expression is being evaluated. Multiple testing correction of identified cis-eQTL was done using a q-value calculation based on the approach of Nettleton *et al.* (2006).

Results & discussion

A set of cis-eQTL was obtained for each chromosome for our pig population. From this, we were able to identify the total number of eQTL after q-value control for each chromosome (Table 1), as well as the distribution of the number of genes whose expression is associated with each cis-eQTL. Chromosome six had the largest number of unique cis-eQTL after filtering (268,389), while chromosome 16 had the smallest number of cis-eQTL (51,688). The number of detected cis-eQTL per chromosome generally decreased with chromosome size, except for chromosome 12, which had many unique cis-eQTL, despite its smaller physical size. The largest number of genes whose expression was significantly associated with a given cis-eQTL by chromosome ranged from 12 for chromosome 16 to 53 for chromosome 7. Most cis-eQTL, however, had less than 10 associated genes, with the median number of genes associated with an eQTL equal to five and most eQTL only being associated with one or two genes (Figure 1). The total number of unique genes with a significant cis-eQTL per chromosome ranged from 179 for chromosome 11 to 1,053 for chromosome 6. Again, chromosome 12 had many unique genes for its size, which suggests that chromosome 12 has an abundance of cis-eQTL associated with gene expression in whole blood.

Table 1. Total number of unique cis-eQTL and associated expressed genes (# genes) by chromosome and the maximum number (Max #) of genes whose expression is associated with a single eQTL.

Chromosome	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
# cis-eQTL	247,046	229,560	194,465	150,175	118,033	268,389	154,336	119,889	140,710
# genes	864	952	763	568	569	1,053	661	342	493
Max # genes	34	50	51	44	35	48	53	20	25
Chromosome	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**
# cis-eQTL	76,294	52,045	166,380	220,995	225,801	103,632	51,688	95,557	53,619
# genes	235	179	621	685	668	386	194	300	208
Max # genes	17	15	41	29	26	24	12	25	23

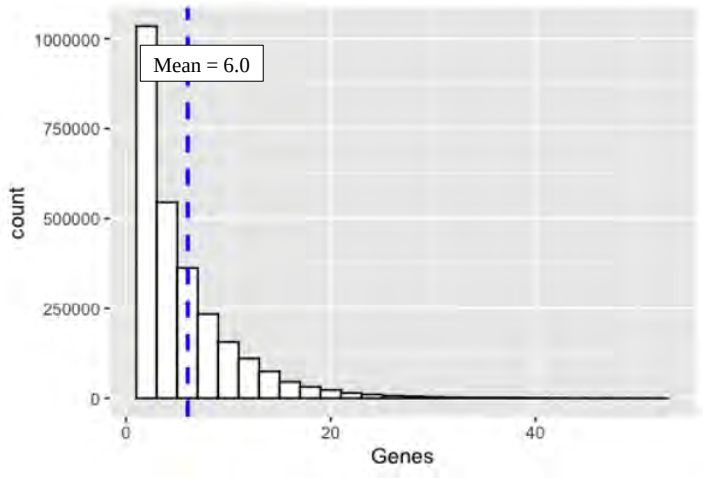


Figure 1. Frequency distribution of the number of genes whose expression is associated with each cis-eQTL.

Future analyses of eQTL for whole blood for this data set will involve an analysis of trans eQTL and of the associated genes and eQTL with disease resilience traits, following Lim *et al.* (2021). Cis- and trans-eQTL analyses after deconvolution of the gene expression data into various cell-type-specific proportions rather than gene expression of whole blood will also be conducted to identify putative regulatory regions associated with gene expression in these individual cell populations, in order to further develop and utilize blood as a tissue for understanding the biological processes behind economically important traits in swine. Through this approach it could become possible to use individual blood component gene expression profiles to identify individuals that are more or less resilient to disease.

Acknowledgements

This work was funded by USDA-NIFA grants # 2017-67007-26144 and #2021-67015-34562, along with Genome Canada and PigGen Canada. The natural disease challenge model was designed and implemented by John Harding (University of Saskatchewan), Michael Dyck (University of Alberta), Frederic Fortin (Centre de Développement du Porc du Québec, Inc.), Jack Dekkers and Graham Plastow together with PigGen Canada.

References

Andrews SFASTQC. *A quality control tool for high throughput sequence data*. 2010 Available at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc/

Browning B.L., Zhou Y., Browning S.R. (2019) Am J Hum Genet 103(3):338-348 https://doi.org/10.1016/j.ajhg.2018.07.015

Chang C.C., Chow C.C., Tellier L.CAM., Vattikuti S., Purcell S.M., *et al.* (2015) GigaScience 4(1) https://doi.org/10.1186/s13742-015-0047-8

Cheng J., Putz A.M., Harding J.C.S., Dyck M.K., Fortin F., *et al.* (2020) J Anim Sci., 98(8) https://doi.org/10.1093/jas/skaa244

Li H. (2011) Bioinformatics 237(21):3987-93 https://doi.org/10.1093/bioinformatics/btr509

Lim K.S., Dong Q., Moll P., Vitkovska J., Wiktorin G., *et al.* (2019) BMC Genomics, 20(1):1-10. https://doi.org/10.1186/s12864-019-6122-2

Lim K.S., Cheng J., Putz A.M., Dong Q., Bai X., *et al.* (2021) BMC Genomics 22:614 https://doi.org/10.1186/s12864-021-07912-8

Liu S., Gao Y., Canela-Xandri O., Wang S., Yu Y., *et al.* (2021) bioRxiv https://doi.org/10.1101/2020.12.01.406280

Nettleton D., Hwang J.T.G., Caldo R.A., Wise R.P. (2006) JABES 11, 337-356. https://doi.org/10.1198/108571106X129135

Purcell S, Change C. PLINK Available at: www.cog-genomics.org/plink/1.9/

Putz A.M., Harding J.C.S., Dyck M.K., Fortin F., Plastow G.S., *et al.* (2019) Front Genet 9:660 https://doi.org/10.3389/fgene.2018.00660

Robinson M.D., McCarthy D.J., Smyth G.K. (2010) Bioinformatics 26(1):139-140 https://doi.org/10.1093/bioinformatics/btp616

Stegle O., Parts L., Piipari M., Winn J., Durbin R. (2012) Nat Protoc 7(3): 500-7 https://doi.org/10.1038/nprot.2011.457

Taylor-Weiner A., Aguet F., Haradhvala N.J., Gosai S., Anand S., *et al.* (2019) Genome Biology 20(228) https://doi.org/10.1186/s13059-019-1836-7

561. Unravelling the structural variability in chicken genomes by long-read sequencing

J. Geibel[1,2*], J. Schauer[3], A. Weigend[3], C. Reimer[1,2,3], D.-J. de Koning[4], H. Simianer[1,2] and S. Weigend[2,3]

[1]University of Goettingen, Animal Breeding and Genetics Group, Albrecht-Thaer-Weg 3, 37075 Göttingen, Germany; [2]University of Goettingen, Center for Integrated Breeding Research, Carl-Sprengel-Weg 1, 37075 Göttingen, Germany; [3]Friedrich-Loeffler-Institut, Institute of Farm Animal Genetics, Höltystraße 10, 34535 Neustadt, Germany; [4]Swedish University of Agricultural Sciences, Department of Animal Breeding and Genetics, Box 7023, 750 07 Uppsala, Sweden; johannes.geibel@uni-goettingen.de

Abstract

Structural variants (SV) have gained increasing interest within the last years. However, calling SVs from array genotypes or short-read data is known to suffer from high false-positive rates and low specificity at the same time. In contrast, long-read technologies promise to derive accurate and sensitive SV callsets. We hereby present the first study that uses PacBio and Nanopore long-reads to derive quantitative estimates of the amount of SVs in chicken, based on four trios. We estimated the mean size of the reference genome affected by at least heterozygous SV calls to range from 0.2% for insertions to 0.7% for deletions. We further compared the results to a short-read based calling approach. This revealed more than 2/3 of short-read based initial SV calls not being backed by long-read based calls, while more than 50% of the long-read based calls would have been missed by the short-read approach.

Introduction

During the last years, some prominent phenotypes were shown to be causally affected by structural variants (SV), e.g. different comb phenotypes in chickens (Imsland *et al.*, 2012). Effects of SV may, however, be missed by analyses based on single nucleotide polymorphism (SNP) markers, as linkage disequilibrium between SVs and neighbouring marker SNPs is not necessarily high (Geibel *et al.*, 2021). This has driven the interest in identifying SVs along the genome. SV calling so far was mostly based on SNP-array data or short-read sequence data, and heavily relied on auxiliary information like probe intensity, coverage, insert size distributions, split reads, or local assembly. This led to high false positive rates and low specificity at the same time, especially when calling insertions (INS; Delage *et al.*, 2020), or SVs in repetitive regions (Belyeu *et al.*, 2021). Previous estimates of the SV content of chicken genomes therefore varied massively and reached estimates up to 12.84% of the autosomal genome size (Fernandes *et al.*, 2021).

Many problems of calling SVs from short reads can be overcome by using long reads from Oxford Nanopore Technologies (ONT) or Pacific Biosciences (PacBio) sequencing, which are nowadays considered to strongly improve SV identification (Ho *et al.*, 2019; Sedlazeck *et al.*, 2018b). However, studies identifying SVs based on long read technologies in chickens have not been published to this date.

The goal of this study was to provide a reliable quantitative estimate of the affectedness of the chicken genome by SVs using long-read sequencing of four chicken trios. We further verified how many of these structural variants would have been discovered by short-read sequencing, providing an insight into the false-positive and false-negative rates of the latter approach.

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_561, © J. Geibel *et al.* 2022

Materials & methods

For the discovery of SVs, four chicken trios were sequenced. The trios were from a commercial White Leghorn line (WLA), a commercial Rhode Island Red line (BLA), a New Hampshire line (L68) and the intercross of an Araucana × WLA backcross experiment. Sequencing was done using Illumina short read sequencing (13 – 42X, 2×151 bp), ONT sequencing (6 – 22X; 3 – 22 kb mean read length) and PacBio HiFi sequencing (3 – 25X; 16 – 22 kb mean read length). Note that for the BLA trio, a fullsib of the offspring chicken, which was sequenced by Illumina and PacBio, was sequenced with ONT due to a shortage of available DNA.

SV calling was performed separately for the different sequencing techniques. The short reads were mapped against GRGC6a by bwa-mem (Li, 2013) and SVs were called by a consensus calling approach that involved parallel SV calling by delly (Rausch *et al.*, 2012), manta (Chen *et al.*, 2016) and lumpy (Layer *et al.*, 2014), followed by feeding the union of the SV calls into svtyper (Chiang *et al.*, 2015) for genotyping. As svtyper is not able to genotype INS, only the calls of manta were considered for INS. Non-default filters for the short read data were not applied at this stage, as the main intention was to see whether short-read data would have provided support for SVs found by long reads. Long reads (PacBio and ONT) were mapped by ngmlr (Sedlazeck *et al.*, 2018a), followed by SV calling and genotyping through sniffles (Sedlazeck *et al.*, 2018a), requiring a threshold of at least five supporting reads. Long-read-based SV detection was performed for the technologies separately as well as based on merged bam alignment files (LRcombined). Finally, a merged VCF was produced for further analysis, using survivor (Jeffares *et al.*, 2017) together with custom awk scripts to account for incompatibilities between the different tools. To account for bad breakpoint resolutions, SV of the same type with breakpoint differences of 1 kb or less were further merged into one SV. Analyses were thereby limited to autosomal deletions (DEL), duplications (DUP), insertions (INS), and inversions (INV) between 50 bp and one Mb. Note that these are early stage results, which will require further validation, e.g. by assessing Mendelian errors in the trios.

Results & discussion

The short-read based consensus calling approach (SRconsensus) identified a total of 32,111 SVs (15.6k DEL, 6.9k DUP, 6.6k INS, 2.9k INV), while 18,487 SVs were identified by the combined set of PacBio and ONT reads (LRcombined; 8k DEL, 1.2k DUP, 8.8k INS, 445 INV). Figure 1 shows the overlap between the different callsets. It is remarkable that more than 50% of the LRcombined SVs (10.5k) did not show even modest support in the SRconsensus set, while 2/3 of the SRconsensus SVs (21.6k) were not supported by long-read based calls. This indicates the known limitations of relatively non-sensitive and unreliable short-read calling algorithms requiring high manual curation effort (Belyeu *et al.*, 2021). However, against expectation (Delage *et al.*, 2020), more than 50% of the INS calls by manta were supported by long reads. Further, only very few DUP (1,244) were supported by the LRcombined set with close to no overlap to the SRconsensus set. This might partly be due to the fact that a high share of raw DUP calls in short-read data sets is expected to stem from mapping errors in badly assembled repetitive regions (Geibel *et al.*, 2021), which might be better resolved by long reads and called as INS rather than DUP then.

Generally, increasing the depth by combining the long-read sets improved the confidence in variant calls and 4k SVs were only discovered in the LRcombined set. Interestingly, slightly fewer SVs (3.3k) were called when only the PacBio set was used, but no longer when combining the sets. This might be a calling artefact, as sniffles has a special HiFi mode that trusts mappings with more confidence, but which could not be used for LRcombined, as LRcombined also included less accurate ONT data. Anyway, this needs further investigation.

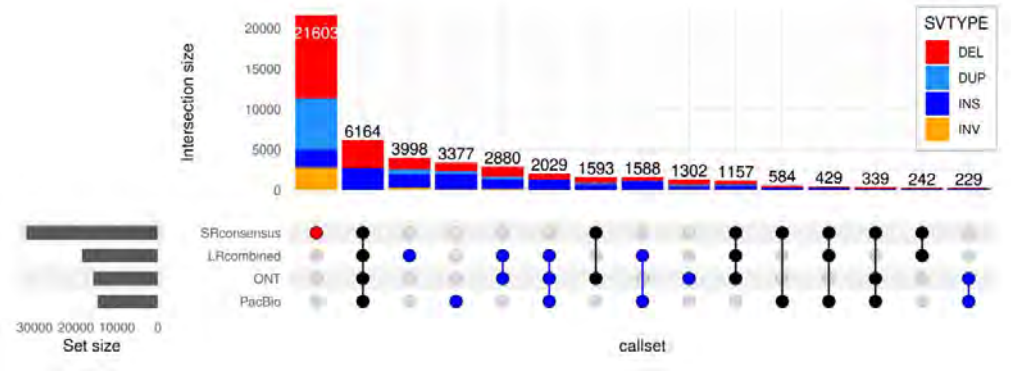


Figure 1. Upset plot showing the overlap between the callsets. Colours in the top plot indicate the proportion of SV types in the intersection. Colours in the intersection matrix instead indicate whether the SVs were found by long reads solely (blue), short reads solely (red) or by both techniques (black).

The length of the called SVs was only investigated for the most reliable set, LRcombined. Length distributions of the SVs (Figure 2A) spanned the complete investigated range between 50bp and one Mb, with lengths >100 kb (DUP and INV) respectively >10 kb (DEL and INS) being rare events. Median sizes of 913bp (DUP), 379bp (INV), 153bp (DEL), and 132bp (INS), however, showed that the vast majority of called SVs was even smaller than 1 kb.

We finally calculated the fraction of the autosomal reference genome that was affected by SV per individual (Figure 2B). This number varied on average between 0.2% for INS and 0.7% for DEL. This is more than what was discovered previously by highly curated short-read based callsets (e.g. Geibel *et al.*, 2021), but by far less than what array-based CNV studies regularly estimate (e.g. Fernandes *et al.*, 2021).

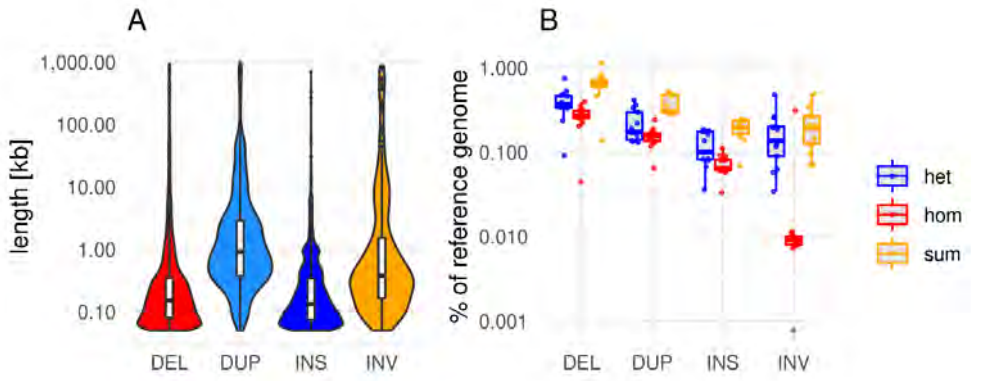


Figure 2. Distribution of SV lengths in the LRcombined callset (A). Percent of autosomal reference genome affected by heterozygous (het), homozygous (hom), or any (sum) SV genotypes by SV type in the LRcombined callset for the different chicken samples (B). Note the log-scaled y-axes.

Acknowledgements

The sequencing of the chicken was funded within the scope of the project IMAGE (www.imageh2020.eu; European Union's Horizon 2020 research and innovation programme under grant agreement No 677353). The Illumina sequencing was performed by the SNP&SEQ Technology Platform in Uppsala, Sweden. The facility is part of the National Genomics Infrastructure (NGI) Sweden and Science for Life Laboratory. The SNP&SEQ Platform is also supported by the Swedish Research Council and the Knut and Alice Wallenberg Foundation. Further, the PacBio sequencing was performed at NGI/ Uppsala Genome Center that has been funded by RFI/ VR and Science for Life Laboratory, Sweden. It was further supported by the National Genomics Infrastructure (NGI)/ Uppsala Genome Center and UPPMAX by providing assistance in massive parallel sequencing and computational infrastructure.

References

Belyeu, J.R., Brand, H., Wang, H., Zhao, X., and Pedersen, B.S., *et al.* (2021) Am J Hum Genet 108(4): 597–607. https://doi.org/10.1016/j.ajhg.2021.02.012

Chen, X., Schulz-Trieglaff, O., Shaw, R., Barnes, B., and Schlesinger, F., *et al.* (2016) Bioinformatics (Oxford, England) 32(8): 1220–1222. https://doi.org/10.1093/bioinformatics/btv710

Chiang, C., Layer, R.M., Faust, G.G., Lindberg, M.R., and Rose, D.B., *et al.* (2015) Nature methods 12(10): 966–968. https://doi.org/10.1038/nmeth.3505

Delage, W.J., Thevenon, J., and Lemaitre, C. (2020) BMC genomics 21(1): 762. https://doi.org/10.1186/s12864-020-07125-5

Fernandes, A.C., Da Silva, V.H., Goes, C.P., Moreira, G.C.M., and Godoy, T.F., *et al.* (2021) BMC genomics 22(1): 354. https://doi.org/10.1186/s12864-021-07676-1

Geibel, J., Praefke, N.P., Weigend, S., Simianer, H., and Reimer, C. (2021) Research Square (preprint). https://doi.org/10.21203/rs.3.rs-861830/v1

Ho, S.S., Urban, A.E., and Mills, R.E. (2019) Nature Reviews Genetics. https://doi.org/10.1038/s41576-019-0180-9

Imsland, F., Feng, C., Boije, H., Bed'hom, B., and Fillon, V., *et al.* (2012) PLoS genetics 8(6): e1002775. https://doi.org/10.1371/journal.pgen.1002775

Jeffares, D.C., Jolly, C., Hoti, M., Speed, D., and Shaw, L., *et al.* (2017) Nature Communications 8 14061. https://doi.org/10.1038/ncomms14061

Layer, R.M., Chiang, C., Quinlan, A.R., and Hall, I.M. (2014) Genome Biology 15(6): R84. https://doi.org/10.1186/gb-2014-15-6-r84

Li, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. https://arxiv.org/pdf/1303.3997

Rausch, T., Zichner, T., Schlattl, A., Stütz, A.M., and Benes, V., *et al.* (2012) Bioinformatics (Oxford, England) 28(18): i333-i339. https://doi.org/10.1093/bioinformatics/bts378

Sedlazeck, F.J., Rescheneder, P., Smolka, M., Fang, H., and Nattestad, M., *et al.* (2018a) Nature methods 15(6): 461–468. https://doi.org/10.1038/s41592-018-0001-7

Sedlazeck, F.J., Lee, H., Darby, C.A., and Schatz, M.C. (2018b) Nature Reviews Genetics 19(6): 329–346. https://doi.org/10.1038/s41576-018-0003-4

562. Detection and characterisation of *de novo* structural variants in pigs

M. Steensma[1*], Y.L. Lee[1], A.C. Bouwman[1], C. Pita Barros[1], M.F.L. Derks[1,2], M.C.A.M. Bink[3], B. Harlizius[2], A.E. Huisman[3], R.P.M.A. Crooijmans[1], M.A.M. Groenen[1], H.A. Mulder[1] and C.M. Rochus[4]

[1]Wageningen University & Research Animal Breeding and Genomics, P.O. Box 338, 6700 AH Wageningen, the Netherlands; [2]Topigs Norsvin Research Center, Schoenaker 6, 6641 SZ Beuningen, the Netherlands; [3]Hendrix Genetics, P.O. Box 114, 5830 AC Boxmeer, the Netherlands; [4]University of Guelph, Centre for Genetic Improvement of Livestock, 50 Stone Rd E, N1G 2W1, Guelph, ON, Canada; marije.steensma@wur.nl

Abstract

De novo mutations arising in the germline add to genetic variation. The number of *de novo* mutations occurring every generation, especially structural variants, has not been well studied in most species, including livestock. We used whole-genome sequencing from 46 pig trios from two commercial lines to identify *de novo* structural variants (dnSVs) present in the offspring. We characterised these dnSV by identifying their parent-of-origin, predicting their causal mechanisms, and identifying their functional annotations. We identified four dnSVs, including two clusters of mutations. One of these clusters contained a deletion, and three duplications, one of which was inverted. This cluster was the only dnSV that could be phased and was located in the paternal haplotype of the proband. All four identified dnSVs were located within the introns of genes. Our study is the first of its kind to identify and characterise dnSVs using whole genome shotgun sequence data in pigs.

Introduction

Mutation is the fundamental source of genetic variation. There are two types of mutations that can be found in the germline, and therefore can be transmitted to the next generation: *de novo* mutations (DNM) which arise in the germline during gametogenesis (Bishop *et al.*, 2020) and mosaic mutations which have arisen early in the zygote cell during embryogenesis (Jin *et al.*, 2018; Sasani *et al.*, 2019). The genetic variation introduced by mutations, allows for continued selection and adaption in a population, with the tradeoff that DNMs can be deleterious and impact the fitness of an individual (Belyeu *et al.*, 2021).

Structural variants (SVs) are altered DNA segments larger than 50 base pairs, that are a change in copy number, chromosomal location or orientation (Escaramís *et al.*, 2015). The detection of *de novo* SVs (dnSVs) is important because they can change gene dosage and structure, and affect gene expression and function by gains or losses of DNA segments (Bhanuprakash *et al.*, 2018; Ran *et al.*, 2018). We aimed to detect dnSV in two commercial pig lines using whole genome sequencing of trios and characterise these dnSV, including predicting their causal mechanisms, functional annotations, and where possible, identifying the parent-of-origin.

Materials & methods

Samples. DNA from 117 individuals (46 trios from two commercial pig lines) was sequenced (Illumina paired-end sequencing, mean coverage = 30X) and aligned to the Sscrofa11.1 assembly (NCBI, 2017). Nine trios were excluded because of poor DNA sequence quality.

SV detection and de novo candidate filtering. We used a Smoove (v.0.2.8) pipeline (v.0.1.0) for calling SVs (Pita Barros, 2021). Three types of SVs were identified for our study: deletions, duplications, and inversions. We filtered SVs using Bcftools (v.1.9), VCFtools (v.4.0.0) and custom R (v.3.6.2) and Python

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_562, © M. Steensma *et al.* 2022

(v.2.7.15) scripts. SV were declared a dnSV when the proband was heterozygous for the SV and parents and unrelated trios did not have the SV. The sum of genotype quality (GQ) scores for a trio had to be greater than 120. We included additional filters based on SV type, checking depth fold-change, minimum and maximum allelic balance for probands and parents respectively, and a maximum of three reads supporting the dnSV were allowed in either parent. Finally, candidate sites were manually inspected using Integrative Genome Viewer (IGV) (Robinson *et al.*, 2011), to remove spurious dnSV cases.

Characterising dnSVs. To identify the parent-of-origin, we looked at informative SNPs located within the dnSVs. For deletions, the informative SNPs we used were homozygous for the alternative allele in the proband, and homozygous for the reference allele in one of the parents and for duplications, informative SNPs were heterozygous with a 2:1 ratio and one parent homozygous (for either allele) and the other heterozygous (Brandler *et al.*, 2016). Genes located within the dnSVs were retrieved from Ensembl (release 104, May 2021) and NCBI annotation (release 106, May 2017) using NCBI Genome Data Viewer (Rangwala *et al.*, 2021). We identified the most likely causal mechanisms for dnSV formation based on split reads spanning the breakpoints of the dnSVs. Based on the degree of sequence homology at junctions we classified them as either non-homology (Belyeu *et al.*, 2021) or microhomology (Hastings *et al.*, 2009; Quinlan *et al.*, 2010; Yan *et al.*, 2007).

Results

Identified dnSVs. A total of 90,031 structural variants (SVs), including 46,478 deletions, 6,541 duplications, 5,977 inversions, and 31,035 breakend class variants (not included in further analyses) were identified. After filtering for dnSVs, we retained 187 candidate dnSVs and after visual inspection, we identified four dnSVs (Table 1). We estimated a mutation rate of 0.108 per generation (1 dnSV per 9 offspring). We identified two DNM clusters. The first involved a 187-bp duplication, a *de novo* single nucleotide variant within the duplication and a small 11-bp deletion at the distal breakpoint of the duplication. The second was a complex dnSV cluster, which consisted of one deletion and three duplications, of which one duplication was inverted.

Characterising dnSVs. The complex dnSV cluster had informative SNPs in each of the four components identifying it as a paternal haplotype. The other dnSVs had no informative SNPs located within the breakpoints. All gene annotations and predicted mutational mechanisms are reported in Table 1.

Table 1. *De novo* structural variants detected in two pig lines using whole genome sequencing of trios.

dnSV type	Position	Mechanism	Gene
Duplication	Chr1:2940197-2940384	Non-homology	[1]Paralogue of *AGP7*[2]
Complex cluster			
- Deletion	Chr1:95207720-95210792	Micro-homology	*SLC14A2*[2]
- Duplication	Chr1:95210182-95213638	Complex	
- Duplication	Chr1:95213834-95214842	Non-homology	
- Duplication	Chr1:95214800-95232460	Micro-homology	
Duplication	Chr7:37837360-37837933	Non-homology	*BICRAL*[3]
Deletion	Chr15:18965177-18965241	Non-homology	*NCKAP5*[2]

[1] *LOC110256592*.
[2] Source: NCBI.
[3] Source: Ensembl.

Discussion

This is the first study to detect and characterise dnSV in pigs. The dnSV rate, 0.108, was similar to that found in humans Belyeu *et al.* (2021) (0.122). Some dnSVs could remain undetected because it is challenging to uniquely map short-reads to repetitive DNA (Weckselblatt and Rudd, 2015). The use of more costly long-read sequencing technologies would span these repetitive regions and aid in dnSV detection there. Regardless, the DNMs identified in this could contribute to the genetic improvement programs of the evaluated commercial lines in this study, if future studies show they affect production and functional traits.

Complex mutation clusters. Two of the four high evidence dnSVs co-occurred with other DNMs and were complex mutational clusters. The observation of mutational clusters is consistent with other studies (Brandler *et al.*, 2016; Hermetz *et al.*, 2014). A possible explanation are mutational hotspots, which means we would expect an increased number of standing variants near the breakpoints of dnSVs too (Brandler *et al.*, 2016). Understanding these complex mutation clusters is necessary to fully interpret the effects on phenotypic characteristics.

Mutational mechanisms. We found that three out of four high evidence dnSVs were mediated by mechanisms that do not require sequence homology at breakpoints. This agreed with the study by Belyeu *et al.* (2021), which showed that 75% of the detected dnSVs in humans lacked sequence homology. We found that the dnSV cluster arose through several mechanisms, including non-homology, micro-homology and an unidentified complex mechanism. In other SV studies classifications are more specific (Hastings, Lupski, *et al.*, 2009), however, identifying the exact causal mechanisms is complex, and mechanisms can act together (Hastings *et al.*, 2009).

Intronic dnSVs and gene expression. We found that the four dnSVs were overlapping within the first introns of protein-coding genes, where regulatory elements are known to be preferentially located (Rigau *et al.*, 2019). The three known genes, presented in Table 1, are: (*BICRAL)* associated with the Coffin-Siris Syndrome 3 in humans which causes congenital malformation including feeding difficulties and poor growth; (*NCKAP5*) associated with Attention Deficit-Hyperactivity Disorder and Drug-Induced Lupus Erythematosus in humans; and (*SLC14A*) involved in the urea transport family (GeneCards). However, the effects of the detected dnSVs on the pigs they were identified in, remains to be discovered.

Ethics statement

All biological material used in this study was collected as part of routine commercial operations.

Funding

This study was financially supported by the Dutch Ministry of Economic Affairs (TKI Agri and Food project 16022) and the Breed4Food partners Cobb Europe, CRV, Hendrix Genetics and Topigs Norsvin. This research is supported by the NWO domain TTW.

References

Belyeu J.R., Brand H., Wang H., Zhao X., Pedersen B. S., *et al.* (2021) Am. J. Hum. Genet. 108(4):597–607. https://doi.org/10.1016/j.ajhg.2021.02.012

Bhanuprakash V., Chhotaray S., Pruthviraj D.R., Rawat C., Karthikeyan A., and Panigrahi M. (2018) Vet. World 11(4):535–541. https://doi.org/10.14202/vetworld.2018.535-541

Bishop M.R., Diaz Perez K.K., Sun M., Ho S., Chopra P., *et al.* (2020) Am. J. Hum. Genet.:107(1):124–136. https://doi.org/10.1016/j.ajhg.2020.05.018

Brandler W.M., Antaki D., Gujral M., Noor A., Rosanio G., *et al.* (2016). Am. J. Hum. Genet 98(4):667–679. https://doi.org/10.1016/j.ajhg.2016.02.018

Escaramís G., Docampo E., and Rabionet R. (2015) Brief. Funct. Genomics 14(5):305–314. https://doi.org/10.1093/bfgp/elv014

GeneCards. Available at: https://www.genecards.org/.

Hastings P.J., Ira G., and Lupski J.R. (2009) PLoS Genet 5(1):e1000327. https://doi.org/10.1371/journal.pgen.1000327

Hastings P.J., Lupski J.R., Rosenberg S.M., and Ira, G. (2009) Nat. Rev. Genet. 10(8):551–564. https://doi.org/10.1038/nrg2593

Hermetz K.E., Newman S., Conneely K.N., Martin C.L., Ballif B.C., *et al.* (2014) PLoS Genet. 10(1):e1004139. https://doi.org/10.1371/journal.pgen.1004139

Jin Z.-B., Li Z., Liu Z., Jiang Y., Cai X.-B., and Wu J. (2018) Biol. Rev. 93(2):1014–1031. https://doi.org/10.1111/brv.12383

NCBI. (2017) Available at: https://www.ncbi.nlm.nih.gov/assembly/GCF_000003025.6/.

Pita Barros, C. (2021). Available at: (https://carolinapb.github.io/population-structural-var-calling-smoove/).

Quinlan, A. R., Clark, R. A., Sokolova, S., Leibowitz, M. L., Zhang, Y., *et al.* (2010) Genome Res 20(5):623–635. https://doi.org/10.1101/gr.102970.109

R Core Team. (2021) R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna. https://www.r-project.org

Ran X.-Q., Pan H., Huang S.-H., Liu C., Niu X., *et al.* (2018) J. Anim. Physiol. Anim. Nutr. 102(5):1320–1327. https://doi.org/10.1111/jpn.12922

Rangwala S.H., Kuznetsov A., Ananiev V., Asztalos A., Borodin E., *et al.* (2021) Genome Res. 31(1):159–169. https://doi.org/10.1101/gr.266932.120

Rigau M., Juan D., Valencia A., and Rico D. (2019) PLoS Genet. 15(1):e1007902. https://doi.org/10.1371/journal.pgen.1007902

Robinson J.T., Thorvaldsdóttir H., Winckler W., Guttman M., Lander E.S., *et al.* (2011) Nat. Biotechnol. 29(1):24–26. https://doi.org/10.1038/nbt.1754

Sasani T.A., Pedersen B.S., Gao Z., Baird L., Przeworski M., *et al.* (2019) eLife 8:e46922. https://doi.org/10.7554/eLife.46922

Weckselblatt B., and Rudd M.K. (2015) Trends Genet. 31(10):587-599. https://doi.org/10.1016/j.tig.2015.05.010

563. Efficient detection of interchromosomal rearrangements in Holstein bulls using large SNP datasets

J. Jourdain[1,2*], A. Pinton[3], H. Barasc[3], A. Barbat[2], C. Hozé[1,2], M. Philippe[4], D. Boichard[2] and A. Capitan[1,2]

[1]ALLICE, 75012 Paris, France; [2]Université Paris-Saclay, INRAE, AgroParisTech, GABI, 78350 Jouy-en-Josas, France; [3]GenPhySE, Université de Toulouse, INRAE, ENVT, 31320 Castanet-Tolosan, France; [4]EVOLUTION, 35538 Noyal-sur-Vilaine, France; jeanlin.jourdain@inrae.fr

Abstract

Chromosomal structural rearrangements are the cause of reproductive problems. They can be transmitted between generations by progeny carrying a balanced translocation. Detection and elimination of carrier bulls and their carrier progeny is desirable to increase on-farm conception rates. In this paper, we propose to detect these chromosomal abnormalities based on a significant association disequilibrium between markers of different chromosomes in the progeny of a bull. This approach applied to 2,445 Holstein artificial insemination (AI) sire families resulted in the detection of 10 candidate rearrangements (0.41% of bulls). Most of the suspected bulls had low conception rates (relative loss of 29-65% compared to breed average). While historically, translocation studies in cattle focused on targeted cases, our method can be used routinely to detect chromosomal rearrangements based on SNP genotypes available for genomic evaluation.

Introduction

Chromosomal abnormalities have been reported in many livestock species (e.g. Mallepaly *et al.*, 2017) with serious negative consequences for health and reproductive performances. In contrast to other species, in cattle, there are few systematic cytogenetic controls by banding, which probably leads to an underestimation of the phenomenon (De Lorenzi *et al.*, 2012). To date, only 20 reciprocal translocations and 42 Robertsonian fusions have been reported (Iannuzzi *et al.*, 2021). In this paper, a new routine method, using genomic information from halfsibs families, is proposed to detect structural rearrangements in AI sires.

Materials & methods

Genomic and cytogenetic analyses. Phased and imputed Illumina Bovine SNP50 genotypes were obtained from the French genomic evaluation database as described in Mesbah-Uddin *et al.*, 2019. We considered a total of 2,445 Holstein paternal families comprising from 30 to 12,600 halfsibs, all genotyped in France. The average number of informative SNPs on the 29 autosomes was 14,561±607 per bull. Genomic data were processed with R 3.6.2. Association correlations between SNPs were calculated using R package gaston 1.5.6. Significance thresholds under H0 were obtained empirically as follows: 100 bulls with more than 1,500 genotyped offspring and with good reproduction scores (i.e. assumed to have normal karyotype) were selected. For each bull, 10 samples of n progeny were randomly drawn for 12 given values of n between 25 and 1000. For each halfsib family of n individuals, a two-dimensional genome scan was used to detect the maximum correlation between markers of two distinct chromosomes (Rmax). This Rmax was scored if 19 informative adjacent markers on each chromosome had R levels values greater than 0.75 Rmax. If this condition was not met, we considered the next highest correlation between markers of distinct chromosomes and so on. For a given n, 1000 (=10×100) halfsib families were analysed. Thresholds corresponding to $P<0.001$ were obtained for any n between 25 and 1000 by fitting a regression curve to these data (n, Rmax). These thresholds, ranging from 1 to 0.24 for 30 to 1000 halfsibs, enabled us to detect bulls with putative chromosomal rearrangements. Correlation graphs displayed in the results were plotted with ggplot2 3.0.0. package of the R software. Finally, karyotypes of sire C and eight daughters of sire E were obtained from blood lymphocytes as described in Ducos *et al.*, 1998.

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_563, © J. Jourdain *et al.* 2022

Reproduction data. Out of the 2,445 sires considered, the conception rate (CR, i.e. the number of calving occurring between 264 and 294 days after AI divided by the number of AIs) and non-return rates (NRRx; defined as proportion of cows not re-inseminated x days after AI) were calculated for 2,252 bulls that totalized at least 200 AIs with conventional semen between 2010-01-01 and 2020-12-31 (35 million AIs considered).

Results & discussion

In our dataset, ten bulls (0.41%) were suspected to carry a structural rearrangement with $P<0.001$ (Table 1). Surprisingly, none of them was Robertsonian, contrarily to what would be expected in a species with only acrocentric autosomes. The estimated size of the translocated segments ranged from 2 to 25 Mb (mean=10.3±8.2 Mb) and 15 out of the 29 autosomes were affected with two occurrences for BTA2 and three for BTA8 and BTA12.

We have begun collecting blood samples from these sires or their offspring for karyotyping. At time of writing, we had results for 2 bulls, confirming our findings. For bull C, we observed a reciprocal translocation between BTA 3 and 8 in both analyses (Figure 1a). For bull E, delineation of the rearrangement boundaries was made difficult by a run of homozygosity on BTA8 (b, up) and the small size of the segments involved. We observed additional material only on BTA8 (dark band) in the karyotype of some of the daughters. We hypothesize that this rearrangement is a simple insertion of chromosomal material from BTA4 into BTA8.

Table 1. Characteristics of the 10 putative rearrangements.

Bulls ID	A	B	C	D	E	F	G	H	I	J
Affected chromosomes	2,13	10,12	3,8	5,28	4,8	24,29	2,12	6,8	11,16	12,23
Genotyped progeny	36	35	512	68	519	59	49	100	43	41
CR (rank among 2,252 bulls in %)	21.8	24.1	21.2	14.3	28.7	21.5	15.7	43.8	26.1	21.6
	(99.7)	(99.6)	(99.8)	(100)	(99.2)	(99.8)	(99.9)	(29.0)	(99.6)	(99.7)

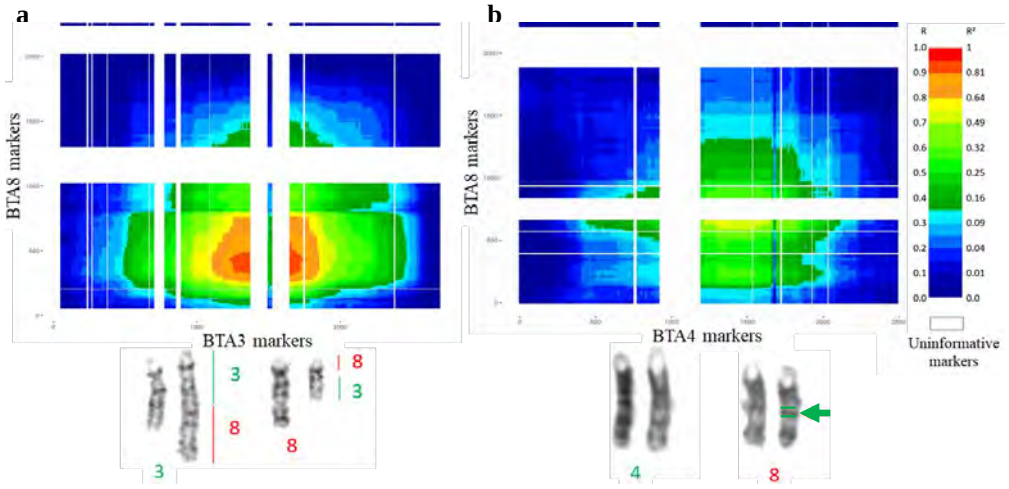


Figure 1. Linkage disequilibrium map of affected chromosomes (up) and details of karyotype (down) for bulls C (a) and E (b).

To precisely identify breakpoints and to search for motifs in their vicinity that could have favored chromosomal rearrangements, genomes of bulls A to J will be sequenced using PacBio CLR technology.

In a second step, we studied the reproductive performances of these ten bulls. Heterozygous carriers of structural chromosomal abnormalities, when they are fertile, can produce balanced (normal, rearranged) and unbalanced (with nullisomy and disomy) gametes. In most cases, unbalanced gametes cause early embryos or fetuses death (Raudsepp and Chowdhary, 2016). Nine bulls had a CR within the lowest percentile of their breed (Table 1) and represented 40% (9/2,252×100) of this worst percentile. Analyses of NRR over time supported early embryonic loss with distinct profiles (Figure 2).

For example, despite their low reproductive scores, bull E and D showed marked differences in the decrease of NNR as compared with the breed average, and thus of embryo and fetal losses (Figure 2a). Daily cumulative proportions of NRR loss also showed a diversity of patterns with bulls losing embryos rapidly (Bull C) versus bulls suspected of causing late fetal death (Bull J; Figure 2.b). This could be explained by the possible effects of rearrangements on semen production (Raudsepp and Chowdhary, 2016) and the diversity of consequences on early embryo or late fetal death (during organogenesis) depending on the genes carried and the size of the rearranged regions (Priya *et al.*, 2018). In our results, one bull (H) did not show evidence of reduced fertility. Further investigation should enable us to determine if this was a false positive or if this animal carries a rearrangement compatible with the survival of aneuploid conceptuses up to birth.

The high consequences on bull fertility of the rearrangements we have identified show that breeding companies should be immediately notified in case of high return rate (in worst 1% of the breed) and the suspect bulls should be at least karyotyped before continuing to sell their semen. Regardless of the genomic evaluation of sires, we advise companies to cull corresponding bulls to avoid a wide spread of abnormalities with long-term and costly consequences on the dairy sector. Non-carrier bulls can of course, be selected for the next generation.

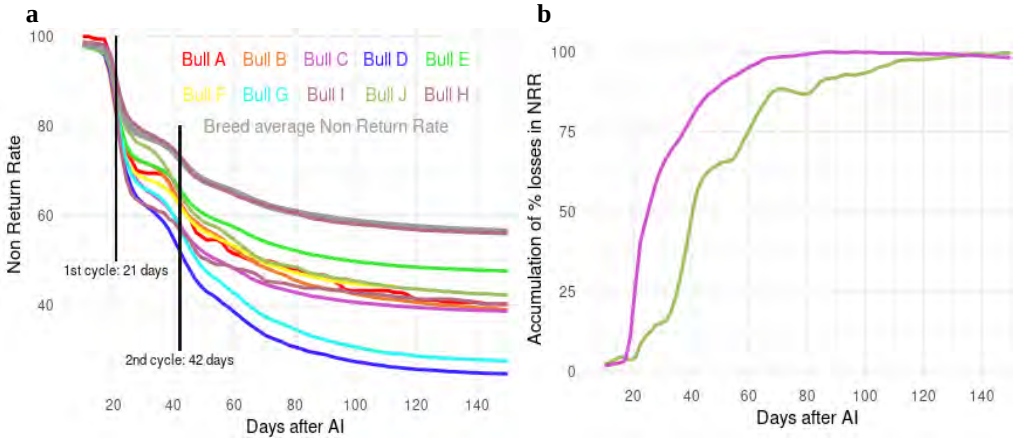


Figure 2. Non-Return Rate vs breed average (a) and daily accumulation of the proportion of losses in NRR for bulls C and J (b).

The method we propose in this communication may underestimate the occurrence of chromosomal abnormalities. Indeed, chromosomal abnormalities can cause sterility by azoospermia (Ghieh *et al.*, 2021). On the other hand, fertile bulls, because of their low fertility, might be prematurely culled because of their low NRR. They also have fewer calves and therefore fewer calves to be genomically tested by breeders, which decreases the accuracy and detection power of the method. We are working on complementary methods to detect suspect bulls based on genotypes of less than 30 daughters.

Despite this possible underestimation, our results are similar to studies in other species. Indeed, in human and swine populations, approximately 0.1 to 0.8% of individuals would carry a *de novo* chromosomal rearrangement (Barasc *et al.*, 2016; Tucker *et al.*, 1994). Furthermore, in humans, chromosomal abnormalities are estimated to cause half of embryo and fetus losses (Pylyp *et al.*, 2018).

Conclusions and perspectives

Chromosomal abnormalities have serious consequences for the dairy industry, especially with the high selection intensity of genomic breeding programs, narrowing down the number of bulls while allowing massive use of their semen before detecting problems with progeny tests. On the other hand, the large datasets generated in the framework of the genomic evaluation allow detection of rearrangements with high efficiency and higher resolution than the karyotyping banding method (5 to 10 Mb). The method proposed in this article can be used routinely to identify bulls with abnormal karyotypes and their balanced progeny to manage these defects in selection.

Acknowledgements

J. Jourdain is supported by a CIFRE PhD grant from Allice, with the financial support of the Association Nationale de la Recherche et de la Technologie and APIS-GENE (Paris, France).

References

Barasc, H., Congras, A., Mary, N., Trouilh, L., Marquet, V. *et al.* (2016) Chromosome Res. 24:511-527. https://doi.org/10.1007/s10577-016-9533-9.

De Lorenzi, L., Morando, P., Planas, J., Zannotti, M., Molteni, L. *et al.* (2012) J. Anim. Breed. Genet. 129:409–416. https://doi.org/10.1111/j.1439-0388.2011.00983.x.

Ducos, A., Berland, H. M., Pinton, A., Guillemot, E., Seguela, A. *et al.* (1998) J. Hered. 89:136-142. https://doi.org/10.1093/jhered/89.2.136.

Ghieh, F., Barbotin, A.-L., Prasivoravong, J., Ferlicot, S., Mandon-Pepin, B. (2021) Basic Clin. Androl. 31:27. https://doi.org/10.1186/s12610-021-00145-5.

Iannuzzi, A., Parma, P. and Iannuzzi, L. (2021) Animals 11:802. https://doi.org/10.3390/ani11030802.

Mallepaly, R., Butler, P. R., Herati, A. S. and Lamb, D. J. (2017) Monogr. Hum. Genet. 21: 1–16. https://doi.org/10.1159/000477275.

Mesbah-Uddin, M., Hoze, C., Michot, P., Barbat, A., Lefebvre, R. *et al.* (2019) J. Dairy. Sci. 102:6340-6356. https://doi.org/10.3168/jds.2018-16100.

Priya, P. K., Mishra, V. V., Roy, P. and Patel, H. (2018) J. Hum. Reprod. Sci. 11:337–342. https://doi.org/10.4103/jhrs.JHRS_132_17.

Pylyp, L. Y., Spynenko, L. O., Verhoglyad, N. V., Mishenko, A. O., Mykytenko, D. O. *et al.* (2018) J. Assist. Reprod. Genet. 35(2): 265–271. https://doi.org/10.1007/s10815-017-1069-1.

Raudsepp, T. and Chowdhary, B. P. (2016) Annu. Rev. Anim. Biosci. 4(1):15–43. https://doi.org/10.1146/annurev-animal-021815-111239.

Tucker, J. D., Lee, D. A., Ramsey, M. J., Briner, J., Olsen, L. *et al.* (1994) Mutat. Res., Sect. environ. mutagen. relat. subj. 313(2):193-202.

564. A pan-genome of commercial pig breeds

M.F.L. Derks[1,2*], A. Boshove[2], B. Harlizius[1], E. Sell-Kubiak[5], M.S. Lopes[1,4], E. Grindflek[3], E. Knol[1], M.A.M Groenen[2] and A.B. Gjuvsland[3]

[1]Topigs Norsvin Research Center, Schoenaker 6, 6641 SZ Beuningen, the Netherlands; [2]Wageningen University & Research Animal Breeding and Genomics, P.O. Box 338, 6700 AH Wageningen, the Netherlands; [3]Norsvin SA, Hamar, Norway; [4]Topigs Norsvin, Curitiba, Paraná, Brazil; [5]Poznań University of Life Sciences, Department of Genetics and Animal Breeding, Wołyńska Poznań, Poland; martijn.derks@wur.nl

Abstract

Genomics related research in animal breeding is usually performed by comparing genomic information to an existing reference genome. However, even if the reference genome is of high quality, the use of a single reference genome has clear drawbacks. Therefore, the breeding community is shifting towards the construction of a pan-genome for important agricultural species. In this study we produced a pig pan-genome based on four different breeds (Landrace, Large White, Synthetic, Duroc) using the nanopore long read sequencing technology. We produced chromosome arm level assemblies comparable to the current *Sus scrofa* 11.1 reference genome. We identified between breed structural variation, which gives a unique insight in the genomic structural variation that define and differentiates breeds. The pig pan-genome will facilitate the discovery of novel variation providing a unique fundamental insight into breed genomic characteristics, which can subsequently be utilized for breeding.

Introduction

In recent decades, high quality reference genomes have become available for most important livestock species. The availability of the pig reference genome (of Duroc origin) together with gene annotation have revolutionized pig genomics and genetics research over the past decade (Groenen 2016). Despite the high quality of the reference genome, working with a single reference genome also has clear drawbacks (Ballouz *et al.* 2019), meaning that sequences deviating considerably from the reference will be interpreted as low-quality, so-called reference bias. One consequence of much (structural) variation is often missed. Structural variation includes various types of variation in which longer stretches of DNA are altered. Structural variants can have a large effect on phenotypes but they are often ignored or remain unidentified (Bickhart and Liu 2014). The developments in long-read sequencing technology, now enables these shortcomings to be addressed. Hence, current focus shifts towards the generation of a pangenome sequence (i.e. all genes and genetic variation within a species that is built from the alignment of different reference genomes) in many important agricultural species, including pigs (Tian *et al.* 2020). The pig pan-genome is particularly useful to identify presence/absence gene variations, structural variation, and other miscellaneous variations. In this study we aim to construct breed specific reference genomes of four commercial pig breeds and assess their structural variation.

Materials & methods

Samples. DNA from 30 individuals was sequenced using the nanopore technology to obtain long reads. Four individuals were sequenced on three flowcells to obtain high coverage sequence data to build breed specific reference genomes (Figure 1). The breeds comprise two dam lines (Large White and Landrace) and two sire lines (Duroc and Synthetic). Additionaly 26 animals were sequenced using one flowcell each. The Circulomics kit was used for DNA extraction. All flow cells were subject to three loadings to obtain optimum output. The average number of gigabases sequenced was 120 Gb, reflecting an average coverage of 48X per flowcell. The average read N50 was 42 kb.

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_564, © M.F.L. Derks *et al.* 2022

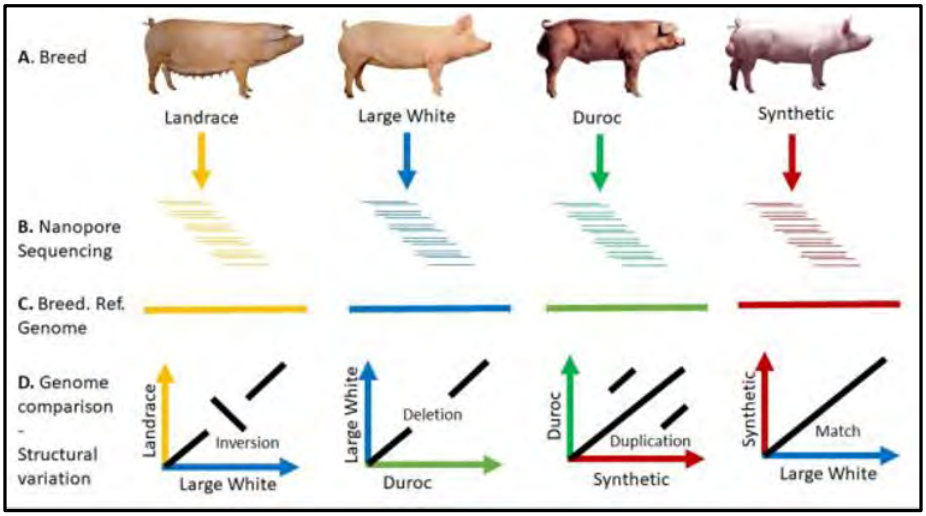


Figure 1. (A) Different pig breeds used in commercial pig breeding; (B) Long read sequencing; (C) Assembly of the long reads to produce breed specific reference genomes; (D) Genome comparison and types of structural variation between the breeds.

Genome Assembly and completeness. We used Flye 2.9 to produce a de novo assembly with the nanopore long reads to construct four breed-specific reference genomes (Kolmogorov *et al.* 2019). Further scaffolding was performed using ntLink (Coombe *et al.* 2021). Polishing was performed using Polca with Illumina short read data from the same individuals (Zimin and Salzberg 2020). The assemblies were anchored to the *Sus scrofa* 11.1 reference genome using RagTag (Alonge *et al.* 2019). The annotation from *Sus scrofa* 11.1 was lifted to the breed-specific reference genomes using Liftoff (Shumate and Salzberg 2020). We assessed the completeness using the BUSCO pipeline (Seppey *et al.* 2019) with the mammalian dataset.

Structural variation and functional prediction. We ran Syri to assess (structural) variation between the line specific genomes (including the *Sus scrofa* 11.1 reference genome) in a pairwise setting (Goel *et al.* 2019). Syri provides VCF files with all identified (structural) variation. We annotated the structural variation for their potential functional consequences using the Ensembl Variant Effect Predictor (VEP) build 104 (McLaren *et al.* 2016). For the one flowcell samples we mapped the samples to the breed reference genome using minimap2 (Li 2018). Subsequently we used Sniffles to assess structural variation between the sample and the breed specific reference genome.

Results

Breed-specific reference genomes. We constructed four breed specific reference genomes. We were able to generate chromosome arm level assemblies for each breed. The acrocentric chromosomes were covered in single scaffolds, whereas the other chromosomes were comprised of 2-4 scaffolds per chromosome. The N50 ranged from 66 to 84 Mb and the contig N50 ranged from 36 to 44 Mb, comparable to the contig N50 of the Sus scrofa 11.1 reference genome. The assembly statistics are given in Table 1.

The assembly completeness was 96.4% for all four breed specific reference genomes using the mammalian BUSCO dataset. The assembly completeness improved significantly after polishing with the Illumina short reads (from ~92 to ~96%).

Table 1. Assembly statistics.

	Sscrofa11.1	Duroc	Landrace	Large White	Synthetic
Assembly length	2,501,921,388	2,478,252,964	2,457,756,321	2,467,972,811	2,476,650,521
Longest sequence	274,330,532	159,819,763	160,007,938	131,827,189	173,409,986
# Scaffolds	613	503	503	607	519
Scaffold N50	138,966,237	84,174,150	77,125,246	66,289,209	79,784,715
Scaffold L50	7	11	13	15	10
# Contigs	1,124	689	683	1,005	694
Contig N50	48,231,277	43,213,177	41,185,454	36,125,157	44,463,933
Contig L50	14	18	20	24	17

Structural variation compared to the Sus scrofa 11.1 reference genome. We used Syri to assess structural differences between the breed specific reference genomes and the 11.1 reference genome (Table 2). The Duroc genome has the least differences with the reference genome especially for the SNPs, inversions, and deletions category. Landrace has a higher number of deletions compared to the other breed-specific reference genomes. This is because we don't have short read data for Landrace (yet) and therefore this breed-specific reference genome remains unpolished, reflected in the number of indels (insertions, deletions).

Large structural variations. We identified various large structural differences between the breed specific reference genomes. First, we identified a 2.8 Mb inversion on chromosome 6 present in all four breed specific reference genomes, but not present in *Sus scrofa* 11.1. This is likely an assembly error in the reference genome. Next we identified a Large White specific 6.2 Mb inversion on chromosome 17. In addition, the start of chromosome 10 is incomplete in the 11.1 reference genome. This first 2 Mb of this chromosome is present in an unplaced scaffold in 11.1 but it's attached to chromosome 10 in our four breed specific reference genomes.

Table 2. Structural variation between the breed specific genomes and the Sus scrofa 11.1 reference genome.

Variation	Large White	Duroc	Synthetic	Landrace
Inversions	147	105	129	139
Translocations	37	55	75	52
Duplications (ref)	31	3	15	15
Duplications (qry)	323	457	477	433
SNPs	7 M	5 M	6.8 M	6.8 M
Insertions	1.5 M	1.3 M	1.3 M	1.1 M
Deletions	785 K	667 K	812 K	4.2 M
Copygains	118	90	113	123
Copylosses	126	90	101	107

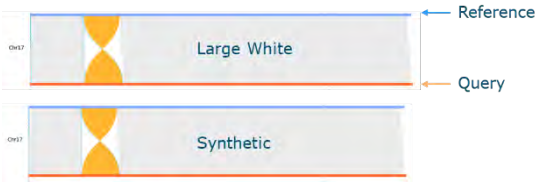


Figure 2. Large white specific (Synthetic is Large White based) inversion on chromosome 17.

Interestingly, we identified a translocation of the complete KIT locus (length 490 kb). The locus is translocated to position 55,035,942 on chromosome 8 (from 41,294,443). That translocation was not identified in the Duroc line specific genome.

Genes affected by structural variants. Table 3 shows the set of structural variants, the type of variants and the genes affected by the variants. Last four columns show the presence of the structural variants in breeds.

Discussion

The power of a commercial pig pangenome. This pig pangenome is a useful resource to identify structural variation between breeds as well as within breeds. Moreover, mapping of short read data will be improved once it is mapped to the pangenome graph instead of a single reference genome. Note that we cannot exclude that some of differences stem from assembly errors despite our high quality chromosome arm assemblies.

Genes affected by structural variation. We identified a set of genes affected by structural variation. Most of the variants disrupt part of the gene. It still remains to be investigated what the exact consequence of the variant is. It is possible that the variant only affects a single isoform of the affected gene. Nevertheless several shared variants are found between different breeds. These variants could have originated before the breed differentiation.

Within breed structural variation. Assessing within breed structural variation that segregates at moderate or low frequency will be highly valuable for breeding purposes. After identification, we plan to assess the phenotypic consequence of the structural variants and potentially add the variants to a SNPchip

Table 3. List of genes affected by structural variation (coding sequence). Notal means not aligned, and cpl means copy loss in query.

Chr	SV	Length	Gene symbol	LW	Du	Sy	LR
12	DEL	~8.8KB	ACSF2	**X**	**X**	**X**	**X**
2	DEL	~19.7KB	ARHGAP26	X		X	X
12	NOTAL	~200KB	CA10	X			X
6	DUP	~4.2KB	CCDC30		X	X	X
14	CPL	~27KB	GPAM		X		X
X	DUP	~3.7KB	GPM6B		X	X	X
8	NOTAL	~75KB	KCNIP4	**X**	**X**	**X**	**X**
16	NOTAL	~93KB	LIFR		X	X	
4	DEL	~4.6KB	MAN1A2	X		X	
8	NOTAL	~20KB	PIGG	**X**	**X**	**X**	**X**
2	DEL	~8.5KB	SAFB2	X	X		
3	DEL	~10KB	SEPTIN12	**X**	**X**	**X**	**X**
7	DEL	~6KB	TNXB		X	X	
5	DEL	~900bp	TP23		X	X	
9	DEL	~4KB	TRIM77		X	X	
14	DEL	~24KB	USP30		X	X	
X	DUP	~1.3KB	ZFX		X	X	X
6	DEL	~19KB	ZNF543	**X**	**X**	**X**	**X**

used for routine genotyping and enable selecting on identified variants. Moreover, we now map sequence reads to a reference genome constructed from an animal that belongs to the same population. This will greatly improve structural variation discovery because there are less structural differences between the sample and the reference genome.

References

Alonge M, Soyk S, Ramakrishnan S, Wang X, Goodwin S, Sedlazeck FJ, Lippman ZB, Schatz MC. 2019. RaGOO: fast and accurate reference-guided scaffolding of draft genomes. *Genome Biol* 20: 224.

Ballouz S, Dobin A, Gillis JA. 2019. Is it time to change the reference genome? *Genome Biol* 20: 159.

Bickhart DM, Liu GE. 2014. The challenges and importance of structural variation detection in livestock. *Front Genet* 5: 37.

Coombe L, Li JX, Lo T, Wong J, Nikolic V, Warren RL, Birol I. 2021. LongStitch: high-quality genome assembly correction and scaffolding using long reads. *BMC Bioinformatics* 22: 534.

Goel M, Sun H, Jiao WB, Schneeberger K. 2019. SyRI: finding genomic rearrangements and local sequence differences from whole-genome assemblies. *Genome Biol* 20: 277.

Groenen MA. 2016. A decade of pig genome sequencing: a window on pig domestication and evolution. *Genet Sel Evol* 48: 23.

Kolmogorov M, Yuan J, Lin Y, Pevzner PA. 2019. Assembly of long, error-prone reads using repeat graphs. *Nat Biotechnol* 37: 540-546.

Li H. 2018. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* 34: 3094-3100.

McLaren W, Gil L, Hunt SE, Riat HS, Ritchie GR, Thormann A, Flicek P, Cunningham F. 2016. The Ensembl Variant Effect Predictor. *Genome Biol* 17: 122.

Seppey M, Manni M, Zdobnov EM. 2019. BUSCO: Assessing Genome Assembly and Annotation Completeness. *Methods Mol Biol* 1962: 227-245.

Shumate A, Salzberg SL. 2020. Liftoff: accurate mapping of gene annotations. *Bioinformatics* https://doi.org/10.1093/bioinformatics/btaa1016.

Tian X, Li R, Fu W, Li Y, Wang X, Li M, Du D, Tang Q, Cai Y, Long Y *et al.* 2020. Building a sequence map of the pig pan-genome from multiple de novo assemblies and Hi-C data. *Sci China Life Sci* 63: 750-763.

Zimin AV, Salzberg SL. 2020. The genome polishing tool POLCA makes fast and accurate corrections in genome assemblies. *PLoS Comput Biol* 16: e1007981.

Part 33

Aquaculture

565. Conserved QTL and chromosomal inversions affect resistance to columnaris disease in two rainbow trout populations

F.C.F. Calboli[1*], H. Koskinen[2], A. Nousianen[2], C. Fraslin[3], R.D. Houston[3] and A. Kause[1]

[1]Natural Resources Institute Finland (Luke). Myllintie 1, FI-31600, Jokioinen, Finland; [2]Natural Resources Institute Finland (Luke), Neulaniementie 5, FI-70210, Kuopio. Finland; [3]The Roslin Institute and Royal (Dick) School of Veterinary Studies, University of Edinburgh, Easter Bush, EH25 9RG, United Kingdom; federico.calbolit@luke.fi

Abstract

We present a comparative genetic analysis of the loci underlying genetic variation in mortality caused by warm water columnaris disease (CD) in farmed rainbow trout (*Oncorhynchus mykiss*) populations. We provide evidence for the conservation of major QTLs, and the putative role played by chromosomal rearrangements in determining resistance to CD. We genotyped 3,962 individuals from two populations experiencing a natural *Flavobacterium columnaris* outbreak. Association analyses identified a major QTL in both populations on chromosome Omy3 and a secondary on Omy5. The QTL on Omy5 was within a naturally occurring major inversion, for which we were able to determine that the ancestral haplotype was associated with lower mortality. Genetic correlation between mortality in the two populations was estimated at 0.64, implying that the populations share much of the same genetic architecture for mortality. Our QTL validation suggests that QTLs could be used in marker-assisted selection, or specifically modelled in genomic evaluation to improve resistance by selective breeding.

Introduction

Flavobacterium columnaris is the etiological cause of columnaris disease (CD), whose mortality is especially high in younger fish, killing fry between 12 to 48 hours after infection (Declercq *et al.* 2013). A vaccine is not available for salmonids, and disease management relies on antibiotics. There is therefore a strong interest in selecting for increased resistance in host species. Rainbow trout (*Oncorhynchus mykiss*) is a prominent aquaculture species worldwide and is one of the main species susceptible to CD. This species has been the subject of extensive genetic analysis, which has generated a reference genome, a commercial SNP chip (Palti *et al.* 2015), and detailed information about naturally occurring genomic rearrangements, such as inversions (Pearse *et al.* 2019), that are present in this species.

In aquaculture species, quantitative trait loci (QTL) are unusually prominent for disease resistance (Fraslin *et al.* 2020), perhaps because of the early phase of domestication. There are two approaches to use QTL information in selective breeding. In a genomic breeding value evaluation with tens of thousands of SNP markers, SNPs can be given a weighting relative to their allele substitution effect on a phenotype. In a simpler marker assisted selection (MAS), pedigree-based evaluation can be augmented by a few markers linked to major QTLs. Both approaches are possible in aquaculture (Houston *et al.* 2020).

We present the results of a genome-wide association study (GWAS) for mortality after a CD outbreak in rainbow trout from two Finnish broodstocks. We addressed the questions of: (1) whether the same chromosomal regions are associated with mortality in both populations; (2) what the heritability of mortality due to infection is; and (3) whether we observe a positive genetic correlation between the mortality trait in both populations. We discuss our results in the light of replicability of association analysis results, and of the conservation of effects of QTL, and other genomic structural variants.

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_565, © F.C.F. Calboli *et al.* 2022

Materials & methods

Populations and genotyping. The fish used in this study come from two broodstocks, one from Savon Taimen Oy (ST), and the second from the national breeding programme maintained by Natural Resources Institute Finland (Luke). 48 sires and 33 dams, and 25 sires and 100 dams were used to make offspring in the Luke and ST populations, respectively. In ST population, the offspring were all-female due to the use of neomale sires, while the Luke population was a mixed sex population. Offspring were raised in one farm in 6 tanks (3 for each population). A natural CD outbreak affected all tanks, and tissue samples were collected for dead and surviving fish (4,505 samples). DNA was extracted and genotyped using the commercially available 57K SNP Axiom™ Trout Genotyping Array. After quality control we had data for 25,853 SNPs and 1,284 ST samples, and 2,691 LUKE samples (1,341 females, 1,337 males). All markers were mapped against the Omyk_1.0 genome.

Statistical analysis. We used PLINK to calculate F_{ST}, and pairwise additive genetic relationship coefficients. We used Genome-wide Complex Trait Analysis (GCTA) to calculate the effective genomic relationship matrix (GRM), which was used in a subsequent GWAS. To study consistency of QTLs, GWAS was performed on each population and sex independently, later combining all samples together, and combining LUKE males and ST females. We used a mixed linear model with leave-one-chromosome-out approach (MLMA-LOCO) using Bonferroni corrected 5% significance thresholds at the genome-wide and chromosome-wide levels. We tested for the presence of a large inversion on chromosome Omy 5, using tagging SNPs (Liu *et al.* 2021). We were able to unequivocally identify the genotype of each fish for the presence of the Ancestral (A) or Rearranged (R) haplotypes. We used the threshold model using BLUPF90 software for the estimation the GRM and of heritability of mortality in each population independently, and to calculate the genetic correlation between the two populations.

Results

The average F_{ST} between ST and LUKE offspring was 0.03, indicating that the two populations were closely related. Pairwise additive genetic relationship between the two populations ranged between 0.00 and 0.34, with median of 0.00. Every single sample in both populations has at least one non-zero additive genetic relationship value with the other population.

GWAS identified a QTL on Omy3 above chromosome-wide significance in all cases (Figure 1A, B, D, E), and above genome-wide significance in all cases except ST fish (Figure 1C). The SNPs associated with this QTL explained between 4 and 12% of the genetic variance. We also observed an inversion on Omy5 (red dots), which is associated with a QTL of different effect in LUKE females and males (Figure 1A, B), reaching chromosome-wide significance in all samples together (Figure 1D). The same inversion reaches genome-wide significance in ST samples and in ST and LUKE male combined samples (Figure 1C, E). In LUKE males, ST, and both combined datasets the SNPs in the QTL on Omy5 explained between 3 and 15% of the genetic variance.

We observed that heritability is of the same magnitude in both populations. The posterior estimates of h^2 on the liability scale are for LUKE h^2=0.32, and for Savon Taimen h^2=0.36. The genetic correlation estimate is 0.64.

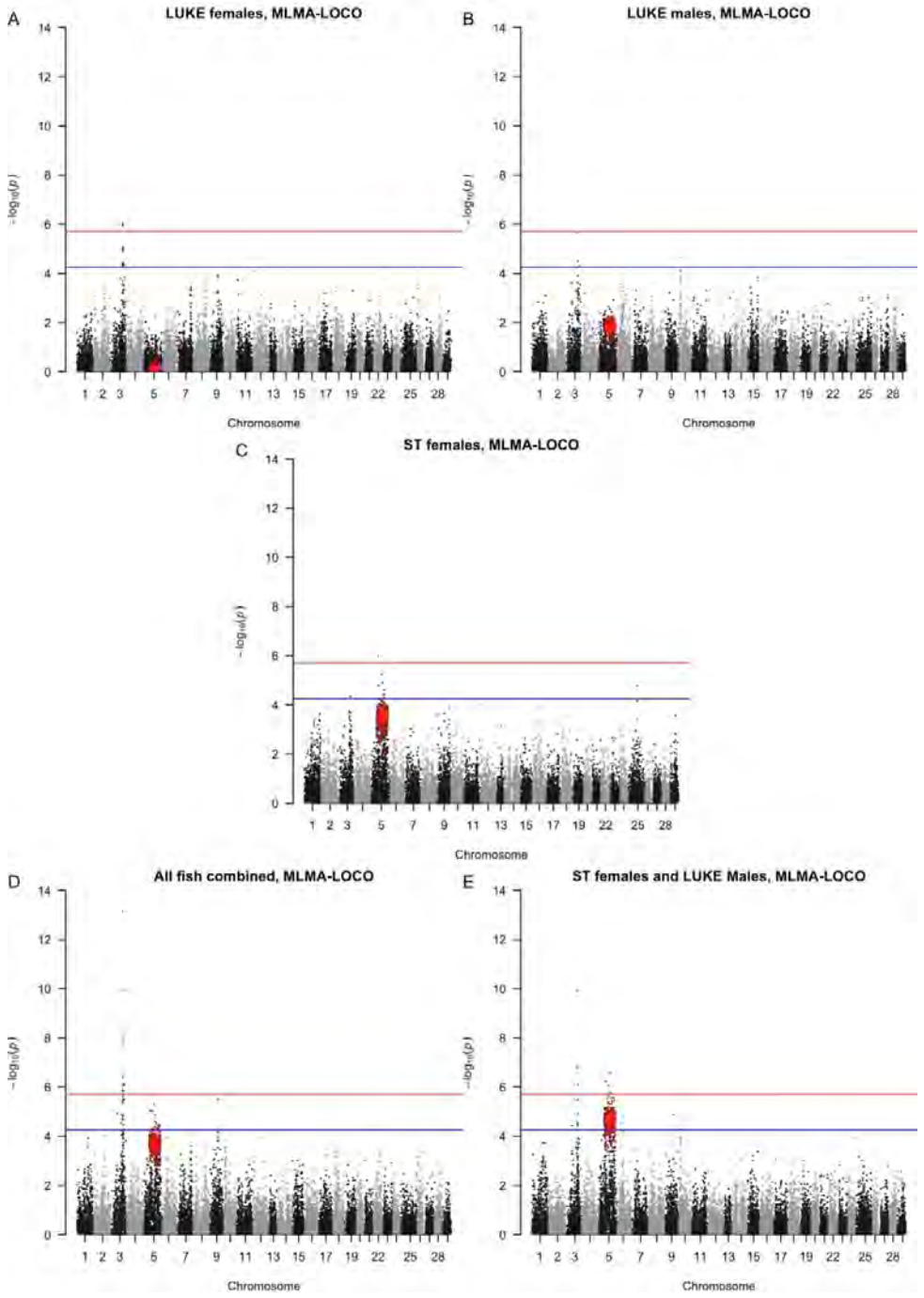


Figure 1. Manhattan plot of the genome wide association analysis in five subsets of the data. In red and in blue are the significance thresholds, respectively, genome-wide and chromosome-wide threshold. Red dots are SNPs tagging an inversion in Omy 05.

Discussion

Our results suggest that the two population we tested are closely related. Based on the F_{ST} comparisons reported by D'Ambrosio *et al.* (2017), the F_{ST} value we observe is consistent with populations separated by about five to ten generations. Our data provides an example of the conservation of QTL location and effect in closely related populations, and our approach to assess the genetic distance between two stocks used in a QTL mapping effort aids to understand the degree to which the stocks share common genetic determination.

The QTL on chromosomes Omy3 is found in both populations, the inversion on Omy 5 is also found in both populations, both with the same direction of allele substitution effect on mortality (in all cases, with the major allele associated with decreased mortality), even though the magnitude of allele substitution effects and the statistical significance shows variation across populations and sexes. The QTL on Omy 3 explains 4-12% of genetic variance, and was observed consistently in both populations. The most novel result we observed is the presence of a QTL associated with mortality on the large inversion present on chromosome Omy 5. In our data haplotype A is associated with a reduced mortality to CD especially in the ST population. In LUKE males, the same QTL effect on mortality is visible but it is weaker and non-significant. By combining LUKE males with ST data, the significance levels of this QTL were greatly elevated, implying the existence of a common QTL. In natural populations the A haplotype is strongly associated with the steelhead anadromous morph in females, and R haplotype is associated with a sedentary freshwater morph in males. The A haplotype is also associated with slightly slower growth, and exhibits a clinal distribution with the highest prevalence in southern rivers, and the frequency of individuals carrying the A haplotype, that also increases resistance against warm water CD, increases with warming water temperature (Pearse *et al.* 2019).

Heritability estimates indicate a moderate genetic effect on the phenotype, and genetic correlation was also moderate, indicating common shared genetic determinants across populations. The correlation between populations allowed to estimate heritability with higher precision than with one population alone. These results reflect both the consistency of QTL effects and of other minor genetic effects in determining the phenotype.

In our data we observe that the main QTLs explain 7 to 11% of the genetic variance, implying they would be a valid choice for MAS or weighted genomic evaluation for decreased mortality for CD. The inversion on Omy5 may be an effective target for MAS, because it can be extremely reliably tagged due to the inversion precluding recombination.

References

Declercq, A.M., Haesebrouck F., Van den Broeck W., Bossier P., and Decostere A. (2014). Vet Res 44:27. https://doi.org/10.1186/1297-9716-44-27

D'Ambrosio J., Phocas F., Haffray P., Bestin A., Brard-Fudulea S., *et al.* (2019). Genetics Selection Evolution 51(1):26. https://doi.org/10.1186/s12711-019-0468-4.

Fraslin C., Quillet E., Rochat T., Dechamp N., Bernadet J-F. (2020). Front. Gen. 11: 677. https://doi.org/10.3389/fgene.2020.00677

Houston, R.D., Bean T.P., Macqueen D.J., Gundappa M.K., Jin Y.H., *et al.* (2020). Nature Reviews Genetics 21:389–409. https://doi.org/10.1038/s41576-020-0227-y

Liu S., Gao G., Layer R.M., Thorgaard G.H., Wiens G.D., *et al.* (2021). Front. Genet. 12:639355. https://doi.org/10.3389/fgene.2021.639355

Palti, Y., Gao G., Liu S., Kent M.P., Lien S., *et al.* (2015). Molecular Ecology Resources 15:662–672. https://doi.org/10.1111/1755-0998.12337

Pearse, D.E., Barson N.J., Nome T., Gao G., Campbell M.A., *et al.* (2019). Nat Ecol Evol 3:1731–1742. https://doi.org/10.1038/s41559-019-1044-6

566. Multi-trait genome wide association study in correlated traits: fillet colour and body weight in Atlantic salmon

P. Sae-Lim*, S. Boison and M. Baranski

Mowi Genetics AS, Sandviksbodene 77A, 5035 Bergen, Norway; panya.sae-lim@mowi.com

Abstract

The 'redness' of the flesh ('fillet colour') is an economically important trait in Atlantic salmon. Pre-adjustment of fillet colour (measured as the concentration of the pigment astaxanthin) with body weight is usually performed in a genetic analysis to account for the correlation between these two traits, however multi-trait models are thought to offer superior power for genome-wide association studies (GWAS). In this study, allele substitution effects (*b*) and their significance levels from three different GWAS approaches were compared; single-trait (STGWAS), pre-adjusting fillet colour for body weight (STadjGWAS), and a multi-trait (MTGWAS). All GWAS analyses used 3,417 animals with 53,159 markers. All approaches consistently identified the most significant markers on chromosome 2 and 26, however MTGWAS increased *b* of fillet colour by 36.8 and 48.5% compared to STGWAS and STadjGWAS, and is therefore recommended as the optimal method to maximise the power of GWAS for correlated traits of this nature.

Introduction

The red colouration of Atlantic salmon flesh (hereafter referred to as 'fillet colour') is an economically important trait that is caused by the deposition of organic pigments, known as carotenoids. Specifically, it is the uptake of supplementary astaxanthin that is responsible and this process has been shown to be heritable with values ranging from 0.11 to 0.45 (Garber *et al.*, 2019; Kristjánsson *et al.*, 2020). Empirical data show that fast growing fish tend to deposit more supplementary astaxanthin. Hence, fillet colour usually correlates positively with body weight, therefore phenotypic variation in fillet colour may be partly explained by variation in body weight. However, pre-adjusting fillet colour by body weight may cause considerable changes in heritability and genetic correlation if fillet colour is not genetically independent from body weight. Kennedy *et al.* (1993) show that, when a derived trait is a linear function of two component traits, it can be optimally selected using an index of the two component traits. Nonetheless, GWAS tries to account for body weight by using it as a covariate when identifying QTL for several focal traits (fillet colour, fillet fat, lice count, etc.). We hypothesize that using body weight as a correlated trait in the MTGWAS will be a superior model that will also improve statistical power to detect candidate genes. The aim of this study was to compare allele substitution effects (*b*) of fillet colour and their significance level estimated with three different approaches in GWAS; single-trait (STGWAS), pre-adjusting fillet colour for body weight (STadjGWAS), and a multi-trait (MTGWAS) where fillet colour and body weight were simultaneously analysed.

Materials & methods

Phenotypes & genotypes. Fillet colour in this study is defined as the concentration of astaxanthin in the fillet estimated from near infrared spectrometry. The fillet colour and final body weight were obtained from 3,420 Atlantic salmon originating from 341 full-sib families of Mowi Genetics breeding population. The mean (standard deviation) of fillet colour and body weight was 5.09 (1.38) mg/kg fillet and 4.13 (1.13) kg, respectively. The regression coefficient of fillet colour on body weight indicated that fillet colour increased by 0.49 mg/kg per 1 kg body weight. Variance components of fillet colour and body weight were preliminary estimated with REML using bivariate animal mixed model in ASReml v4.1 prior to the GWAS analysis. The heritability (SE in parenthesis) for fillet colour and body weight was 0.44 (0.04) and 0.35 (0.04), respectively. The genetic correlation (SE in parenthesis) between fillet colour and body weight was 0.35 (0.07) but reduced to 0.08 (0.09) for fillet colour adjusted for body weight. Each individual was genotyped using a

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_566, © P. Sae-Lim *et al.* 2022

ThermoFisher Axiom genotyping array containing 55,725 single nucleotide polymorphism (SNP) markers distributed over the salmon genome. Quality control of SNPs was performed in PLINK v1.9, based on the following criteria: SNPs were removed if (1) locus missing rate or individual missing rate was greater than 10%, (2) SNPs deviated from Hardy-Weinberg equilibrium with a *P*-value cut-off of 1×10^{-100}, and (3) minor allele frequency was lower than 0.01. After the quality control, 53,159 SNPs remained and the phenotypes reduced to 3,417 individuals.

Genome-wide association study (GWAS). We identified SNPs associated with fillet colour using three different approaches. For STGWAS and STadjGWAS, the analysis was performed using GCTA (Yang *et al.*, 2011), where the fixed covariates in the model were cages, sex, and principle components (PCA, number of PCA used until the inflation factor (λ) approached 1). For the MTGWAS, the analysis was performed using multivariate linear mixed model in GEMMA (Zhou and Stephens, 2014). The number of PCA dimensions in MTGWAS was set to be the same as STGWAS. A subset of top ranked markers associated with QTL for fillet colour from different approaches was taken from all the markers based on ten highest $-\log_{10}(0.05/\text{number of markers})$. The magnitude of marker effect estimates and their ranking due to *P*-value across different approaches were compared. The difference in marker estimates was quantified by calculating root mean square error (RMSE).

Results

The STGWAS identified the most significant SNPs associated with fillet colour on chromosome 2 and 26 (Figure 1). On chromosome 2, marker AX-88047706 was the most significant with an allele substitution effect estimate (*b*) of 0.26 mg/kg. However, a different marker (AX-96373699) in the same region was the most significant from the STadjGWAS and MTGWAS models. The *b* (0.35 mg/kg) of AX-96373699 from MTGWAS was greater than *b* (0.30 mg/kg) from STadjGWAS.

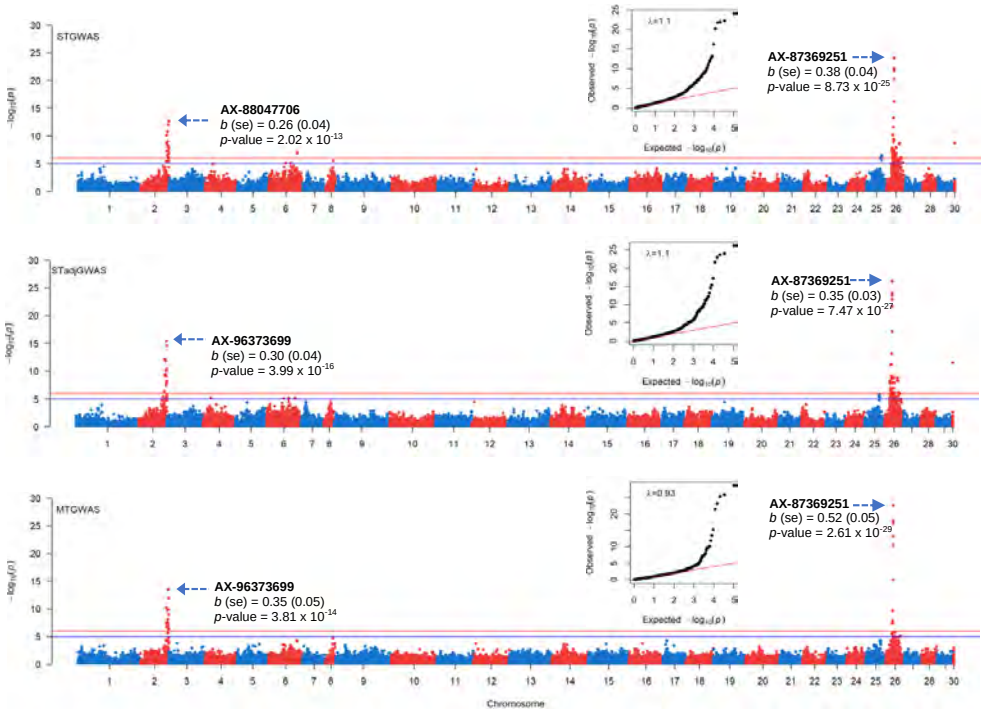


Figure 1. Single trait (STGWAS or STadjGWAS) and multi-trait genome-wide association (MTGWAS) on fillet colour and body weight in Norwegian Atlantic salmon.

On chromosome 26, marker AX-87369251 was the most significant from all the three models however, *b* from MTGWAS increased by 36.8% compared to STGWAS. Significant level based on *P*-value increased from 8.73×10^{-25} (STGWAS) to 2.61×10^{-29} (MTGWAS). The λ was equal to 1.1 in both STGWAS and STadjGWAS while λ reduced to 0.93 for MTGWAS.

The proportion of *b* to the mean of fillet colour of the top ten SNPs indicated the reduction in *b* for STadjGWAS compared to STGWAS. The MTGWAS revealed greater *b* in fillet colour for all top SNPs than STGWAS and STadjGWAS (Figure 2). The RMSE of *b* between STGWAS and MTGWAS of the ten most significant markers was 0.092 mg/kg and it slightly increased to 0.114 mg/kg for RMSE between STadjGWAS and MTGWAS.

Discussion

Two genomic regions on chromosome 2 and 26 were associated with fillet colour. The QTL peak on chromosome 26 has also been reported by Baranski *et al.* (2010) and Helgeland *et al.* (2019). Helgeland *et al.* (2019) reported *beta-carotene oxygenase 1 like* (*bco1l*) at chromosome 26 as the functional gene for fillet colour. In our study, because AX-87369251 is located not more than 500-100 bp from SNPs rs863442990 and rs863785818 reported by Helgeland *et al.* (2019), it is very likely that AX-87369251 is also associated with *bco1l*. The most significant marker on chromosome 2 was located at 70,851,651 bp and was close to the gene *ppa1b* (*Salmo salar* inorganic pyrophosphatase-like, LOC106608436). This enzyme hydrolyses inorganic pyrophosphate which could be a key process in the biosynthesis of astaxanthin (Muthuraman *et al.*, 2021).

Using a multi-trait GWAS to account for the relationship (genetic covariance) between body weight and fillet colour resulted in significant increase in allele substitution effect (*b*) and statistical significance (*P*-value). The increase in *b* could be explained by additional information from the genetically correlated trait (body weight). This phenomenon is similar to a change in breeding values when genetic evaluation is based on multiple correlated traits instead of a single trait. Furthermore, an increase in statistical power from MTGWAS models have been reported by Yoshida *et al.* (2021) for body traits in aquaculture; Broadaway *et al.* (2016) for morphological and physiological traits in cotton and Hackinger and Zeggini, (2017) in humans. Lastly, MTGWAS models may help uncover potential pleiotropic effects (Cai *et al.*, 2020; Chhetri *et al.*, 2019; Hackinger and Zeggini, 2017).

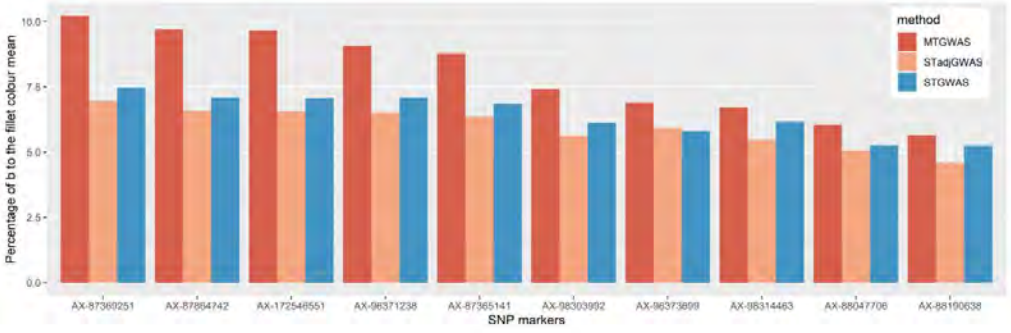


Figure 2. Percentage of allele substitution effect (*b*) to fillet colour mean of the top ten (ranked from highest to lowest) significant SNPs based on different methods of genome-wide association study.

In conclusion, two QTL regions on chromosome 2 and 26 were identified for fillet colour. The MTGWAS approach to account for relationship between fillet colour and body weight increased the magnitude of allele substitution effects of the most significant SNPs and also increased statistical power to detect the QTLs. Therefore, the MTGWAS model is recommended to identifying QTLs when a derived trait is a linear function of two component traits.

References

Baranski M., Moen T., and Våge, D. (2010) Genet. Sel. Evol. 42:17. https://doi.org/10.1186/1297-9686-42-17

Cai Z., Dusza M., Guldbrandtsen B., Sandø Lund M., and Sahana, G. (2020) Genet. Sel. Evol. 52:19. https://doi.org/10.1186/s12711-020-00538-6

Chhetri H.B., Macaya-Sanz D., Kainer D., Biswal A.K., Evans L.M., *et al.* (2019) New. Phytol. 223: 293–309. https://doi.org/10.1111/NPH.15777

Garber A.F., Amini F., Gezan S.A., Swift B.D., Hodkinson S.E., *et al.* (2019) Aquaculture 503:242–253. https://doi.org/10.1016/j.aquaculture.2019.01.001

Helgeland H., Sodeland M., Zoric N., Seilø T.J., Grammes F., *et al.* (2019) Sci. Rep. 9:20061. https://doi.org/10.1038/s41598-019-56438-3

Kennedy B.W., van der Werf J.H., and Meuwissen T.H. (1993) J. Anim. Sci. 71:3239–3250. https://doi.org/10.2527/1993.71123239x

Kristjánsson Ó.H., Gjerde B., Ødegård J., and Lillehammer M. (2020) Front. Genet. 11:573265. https://doi.org/10.3389/FGENE.2020.573265

Muthuraman A., Shaikh S.A., Ramesh M., and Sikarwar M.S. (2021) Stud. Nat. Prod. Chem. 70: 151-194. https://doi.org/10.1016/B978-0-12-819489-8.00013-2

Yang J., Lee S.H., Goddard M.E., and Visscher P.M. (2011) Am. J. Hum. Genet. 88:76–82. https://doi.org/10.1016/j.ajhg.2010.11.011

Yoshida G.M., Lhorente J.P., Carvalheiro R., and Yáñez J.M. (2017) Anim. Genet. 48:698–703. https://doi.org/10.1111/age.12621

Zhou X., and Stephens M. (2014) Nat. Methods. 11:407–409. https://doi.org/10.1038/nmeth.2848

567. Implementation of genomic selection on production and quality traits and linkage disequilibrium in *Crassostrea gigas*

A. Jourdan[1*], P. Boudry[2], F. Enez[1], P. Haffray[1], F. Chenier[3], E. Vetois[4], F. Allal[5], P. Leroy[6], F. Phocas[7], J. Bugeon[8], L. Degremont[9] and R. Morvezen[1]

[1]SYSAAF (French Poultry and Aquaculture Breeders Technical Centre), 35042 Rennes, France; [2]Ifremer, Department of Biological resources and Environment, 29280 Plouzané, France; [3]FRANCE NAISSAIN, 85230 Bouin, France; [4]SATMAR, 50760 Gatteville, France; [5]MARBEC, Université Montpellier, CNRS, Ifremer, IRD, Palavas-les-Flots, France; [6]INRAE, Physiology, Environment and Genetics for Animals and Livestock Systems, 35590 Saint Gilles, France; [7]Université Paris-Saclay, INRAE, AgroParisTech, GABI, 78350 Jouy-en-Josas, France; [8]INRAE, LPGP, Bâtiment 16A, Allée Henri Fabre,Campus de Beaulieu, Rennes, 35042, France; [9]Ifremer, RBE, ASIM, Avenue de Mus de Loup – Ronce les Bains, 17390 La Tremblade, France; ajourdan@ifremer.fr

Abstract

The recent progress and cost reduction of genotyping technologies allows for the implementation of genomic selection (GS) on more and more cultivated species. In this study, we explore the genomic determinism of commercial traits and test the effectiveness and the possible cost reduction of GS in breeding programs of the cupped oyster *Crassostrea gigas*. Two populations of more than 1000 individuals have been phenotyped and genotyped for 40,000 SNP (Guitierez *et al.*, 2017). Heritability was estimated to be moderate for commercial traits of interest (between 0.19 and 0.34). The accuracy of the genomic prediction models outperformed the classical selection on pedigree by 22 to 55%. A limited linkage disequilibrium (LD) level (less than 0.1) was observed. These results suggest that the use of GS in oyster breeding can improve the selection of breeding candidates to enhance commercial traits but need a specific account and exploration of the very low LD.

Introduction

Growth, morphology, and color are essential characteristics for the oyster farming industry and have been the focus of breeding programs. As for other aquaculture species (salmon, sea bass, sea bream), the recent development of genomic tools in the cupped oyster (57K SNP genotyping chip) opens up the possibility of integrating genetic markers into breeding programs by through genomic selection (GS) (Hollenbeck and Johnston, 2018; Boudry *et al.*,2021). According to Lillehammer *et al.* (2013) and Guitierez *et al.* (2018), a few hundred to a few thousand markers could be enough to significantly increase the quality of predictions compared to the standard method of selection on relatives but this requires a sufficiently high LD. In this study we propose to estimate the genetic parameters and to test the efficiency of GS on two commercial breeding programs, on traits of interest, by comparing the accuracy of GS with a standard, pedigree-based approach and evaluating the usefulness of the available genotyping tool, by estimating SNP quality and LD between them.

Materials & methods

Biological material. The two populations in this study were derived from the breeding Company Vendée Naissain (Bouin, France) and SATMAR (Gattevile, France). They were reared on sea shore in Normandy. The first population consists of 1,261 offspring generated from 130 parents (64 males/69 females). The second population comprises 1,136 individuals including the 104 parents (59 males/45 females).

Phenotyping. All individuals were phenotyped for the same traits in both populations: total weight before opening, length, width and height of the shell, upper and the lower valve and wet meat weight. Meat yield was estimated using residue of linear regression between meat weight and total weight. A photo of all

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_567, © A. Jourdan *et al.* 2022

the individuals was taken using a Canon EOS 2000D camera. Images were analysed with an automatic image analysis pipeline using the FIJI software to define surface measurements of the two valves and measurements of the mean external colour of the upper valve estimated in the CIE LAB space.

Genotyping. All individuals of the two populations, including parents and offspring, were genotyped on a bi-species Axiom Affimetrix 57K chip with 40,625 markers for the cupped oyster. A quality analysis carried out on the AxAS software made it possible to retain 14,469 quality markers.

Genetic parameter estimation. Genetic parameters and estimated breeding values (EBV) were estimated with BLUPF90 software suite under animal model. EBV were estimated using pedigree or genomic BLUP (Best Linear Unbiased Prediction). For each, the prediction accuracy was estimated in 40-replicated five-fold crossvalidations (training population 80%, validation population 20%). Prediction accuracy was calculated as the correlation between the predicted EBVs and the actual phenotypes of the validation population, divided by the square root of the heritability estimated. Prediction accuracy values were compared between the pedigree and genomic approaches.

Linkage disequilibrium (LD). The 14,469 good quality markers were blasted on the new reference genome (cgigas_uk_roslin_v1, GCA_902806645.1) in order to map the SNPs. Approximately 12,581 SNPs were well positioned on the 10 linkage groups. To estimate LD, we used the squared correlation based on genotypic allele counts (number of non-reference alleles at each locus) using the PLINK v1.9 software (Chang *et al.*, 2015). Pairwise LD between all SNPs in a 75-Mb long window were derived for each chromosome. The mean r^2 values was calculated for every 50 kb and covering up to 1000 kb.

Results

All of the traits measured have coefficient of variation between 13 and 53%. The two populations were phenotyped at three years post fertilization. The heritability estimated are limited to moderate ranging from 0.04 to 0.34 for color traits and between 0.19 and 0.27 for biometric traits. The gain in prediction accuracy of GBLUP compared to pedigree BLUP on classically measured and selected growth traits is presented in Figure 1. This figure shows GS is superior to selection on pedigree in most cases, with however median levels of gain in accuracy which vary according to the trait studied with values between 22.6 and 55%. The mapping of the 13,000 markers on the most recent assembly (cgigas_uk_roslin_v1, GCA_902806645.1) of the cupped oyster genome shows an important coverage heterogeneity. For example, only 159 markers belonging to chromosome 2 (73 Mb) were genotyped, while more than 2,000 were genotyped for chromosome 10 (58 Mb). The LD is represented by the r^2 as a function of the physical distance over the assembly of the oyster genome. We can observe in Figure 2 that we have a low level of DL with a slow decrease as a function of the distance. LD level is very low even over a short distance between markers (less than 0.1).

Discussion

In our study, we show moderate levels of heritability for production and quality traits. This level of heritability seems lower than when compared to other recent publications in Pacific or Portuguese oysters in Asia (Kong *et al.*, 2015). Difference in rearing methods (suspended *vs* bag) and practices (immersed vs on shore) may have a significant impact on expression of additive genetic components of oyster growth (Sheridan, 1997). Under this hypothesis, estimates of the heritability are moderate but adapted for setting up breeding programs in Europe as they are estimated in the environment of rearing. The medium density chip allows to investigate application of GS in Pacific oysters in other traits than OHsV1 genetic resistance (Guittierez *et al.,* 2017). The results obtained from cross-validation on genomic prediction are similar to

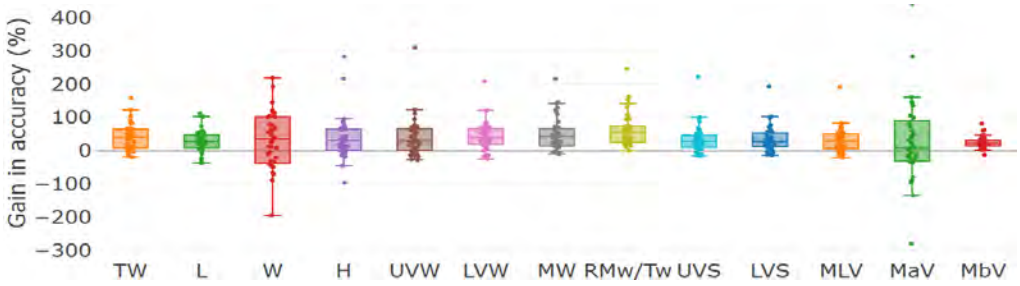


Figure 1. Boxplot of prediction accuracy gain (%) between GBLUP and BLUP for all the 40 simulations and all traits. TW = total weight; L = length; W = width; H = height; UVW = upper valve weight; LVW = lower valve weight; MW = wet meat weight; RMw/Tw = residual meat weight/total weight; UVS = upper valve surface; LVS = lower valve surface; MLV = mean upper valve L value; MaV = mean upper valve a value; MbV = mean upper valve b value.

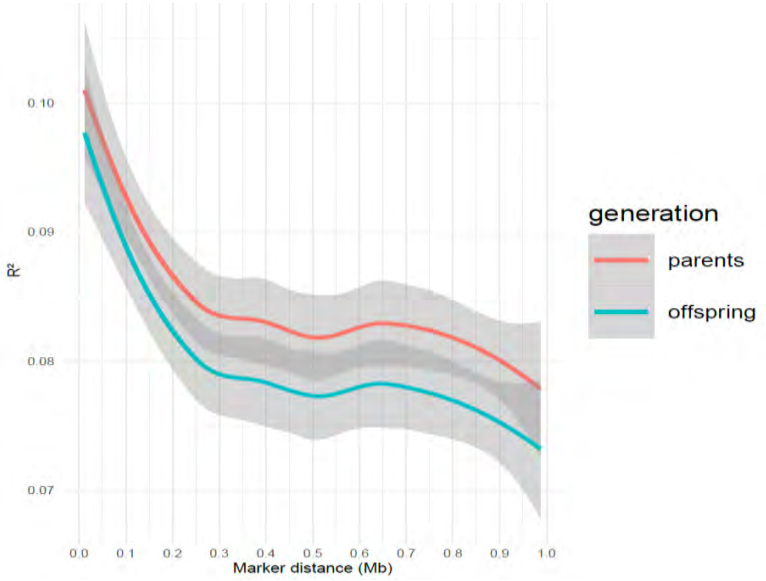


Figure 2. Decay of linkage disequilibrium (LD) with physical distance between markers for the first hatchery population.

those obtained on other aquaculture species with a clear improvement in the accuracy of the prediction. This therefore reinforces the interest of using genomic tools in the breeding programs of the cupped oyster.

One important result is very limited LD observed in two distinct populations. Indeed, to make GS more cost-effective, it is necessary to reduce the number of markers to be genotyped or to go through other methods of GS (such as GWAS) with fewer individuals to genotype. To achieve this, it is essential to have a sufficiently high level of LD in the population. Some parameters can influence the level of LD such as the number of generations for which they were raised in isolation or reduce the effective size (Ne) which would result into increase the levels of relatedness between individuals and therefore a greater extent of LD. Our LD levels are consistent with recent reports describing low levels of LD in wild populations of *C.gigas*

(Guittierez *et al.*, 2017). This result is questioning about the possibility to identify QTLs with repeated and conserved effects across generations with the actual genomic tools developed.

Acknowledgements

The data presented here were obtained in the Quality-Huitre project which received funding from the European Maritime and Fisheries Fund (EMFF) research grant number PFEA470018FA1000011. and by National Association for Research Technology (ANRT) with CIFRE CONVENTION no. 2019/1511. I would like to add a special thanks to Mr. J-B Lamy who accompanied me at the begining of this study.

References

Boudry, P., Allal, F., Aslam, M. L., Bargelloni, L., Bean, T. P., *et al.* (2021). Current status and potential of genomic selection to improve selective breeding in the main aquaculture species of International Council for the Exploration of the Sea (ICES) member countries. *Aquac Rep*, *20*, 100700. https://doi.org/10.1016/j.aqrep.2021.100700

Chang, C. C., Chow, C. C., Tellier, L. C., Vattikuti, S., Purcell, S. M., *et al.* (2015). Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience*, *4*(1), s13742-015. https://doi.org/10.1186/s13742-015-0047-8

Gutierrez, A. P., Turner, F., Gharbi, K., Talbot, R., Lowe, N. R., *et al.* (2017). Development of a medium density combined-species SNP array for Pacific and European oysters (Crassostrea gigas and Ostrea edulis). *G3-Genes Genom Genet*, *7*(7), 2209-2218. https://doi.org/10.1534/g3.117.041780

Gutierrez, A. P., Matika, O., Bean, T. P., & Houston, R. D. (2018). Genomic selection for growth traits in Pacific oyster (Crassostrea gigas): potential of low-density marker panels for breeding value prediction. *Front genet*, *9*, 391. https://doi.org/10.3389/fgene.2018.00391

Gutierrez, A. P., Bean, T. P., Hooper, C., Stenton, C. A., Sanders, M. B., *et al.* (2018). A genome-wide association study for host resistance to ostreid herpesvirus in Pacific oysters (Crassostrea gigas). *G3-Genes Genom Genet*, *8*(4), 1273-1280. https://doi.org/10.1534/g3.118.200113

Hollenbeck, C. M., & Johnston, I. A. (2018). Genomic tools and selective breeding in molluscs. *Front genet*, *9*, 253. https://doi.org/10.3389/fgene.2018.00253

Kong, N., Li, Q., Yu, H., & Kong, L. F. (2015). Heritability estimates for growth-related traits in the Pacific oyster (Crassostrea gigas) using a molecular pedigree. Aquac Res, *46*(2), 499-508. https://doi.org/10.1111/are.12205

Lillehammer, M., Meuwissen, T. H., & Sonesson, A. K. (2013). A low-marker density implementation of genomic selection in aquaculture using within-family genomic breeding values. *Genet Sel Evol*, *45*(1), 1-8. https://doi.org/10.1186/1297-9686-45-39

Sheridan, A. K. (1997). Genetic improvement of oyster production – a critique. *Aquaculture*, *153*(3-4), 165-179. https://doi.org/10.1016/S0044-8486(97)00024-0

568. Genetic correlation between IP and cohabitation challenge test model for *Piscirickettsia salmonis* in Atlantic salmon

C. Manosalva[*1], M.L. Pavez[1], P. Sae-Lim[2], A. Norris[3], M. Baranski[2] and S.A. Boison[2]

[1]Mowi Chile, Camino a Chinquihue KM 12 S/N Puerto Montt, 5500000 Chile; [2]Mowi Genetics AS, Sandviksbodene 77AB, 5035 Bergen, Norway; claudia.manosalva@mowi.com

Abstract

Salmonid Rickettsial Syndrome (SRS) is one of the infectious disease causing significant mortality of Atlantic salmon in Chile. Survival to SRS is mainly recorded from two different challenge test models; intraperitoneal injection (IP) and cohabitation but their genetic correlation has not been estimated. The aim of this study was to estimate the genetic correlation between survival to SRS measured from IP and cohabitation experimental challenge test. Variance components were estimated from 3,488 animals using multi-trait animal mixed models. Heritability for survival day and survival binary was moderate and consistent across IP and cohabitation. Genetic correlation of survival day between IP and cohabitation was close to unity (0.933) but lower for genetic correlation of binary survival between IP and cohabitation (0.825). In conclusion, survival to SRS from different challenge test models are highly correlated. Therefore, IP model can be adopted as it results in lower cost of the experimental challenge test.

Introduction

Salmonid Rickettsial Syndrome (SRS) caused by a gram negative bacteria *Piscirickettsia salmonis* is one of the largest disease causing pathogen of Atlantic salmon in Chile (Departamento de Salud Animal, 2020). In 2020, SRS was estimated to cause ~67.3% of mortalities due to infectious diseases in the Chilean aquaculture industry (Departamento de Salud Animal, 2020). Genetic variation for resistance to SRS in Atlantic salmon has been reported in several populations and heritabilities range from 0.11 to 0.43 (Lhorente *et al.*, 2014; Moraleda *et al.*, 2021; Yáñez *et al.*, 2014). Data for estimating heritability and selection of candidates for genetic improvement in a breeding programme can be obtained from field outbreaks or laboratory challenge tests. Challenge trials have been commonly used because they can be routinely undertaken to provide reliable and repeatable phenotypic information for the breeding programme. Two main challenge test models: (1) intraperitoneal injection (IP) of the test fish; and (2) cohabitation of IP injected shedder fish with test fish; are commonly practiced in aquaculture. The IP model usually results in faster disease development and shorter challenge test duration. The cohabitation challenge model is developed to be reflective of the usual routes of infection which occurs during field outbreaks. Therefore, cohabitation challenge models results in slower disease development and longer challenge test duration than IP challenge models. In cases where both challenge models can be developed for the disease causing pathogen (e.g. SRS), IP models are usually adopted as it results in lower cost of the experimental challenge test. It can be argued that to adopt IP challenge test model for routine evaluation of resistance to disease pathogens (e.g. SRS) in breeding programme, the genetic correlation with cohabitation challenge test should be moderate to high. The aim of this study was to estimate the genetic correlation between resistance to SRS measured from IP and from cohabitation experimental challenge test.

Materials & methods

Animals. The fish used in this study originated from the 2017G yearclass of Mowi breeding nucleus in Chile. The population consisted of 109 fill-sib families from 74 sires and 109 dams with a mixture of 1:1 (66%), 1:2 (20%) and 1:3 (14%) mating ratios (sire:dam). Families were reared separately until tagging at an average body weight of 15 g.

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_568, © C. Manosalva *et al.* 2022

Experimental challenge test. Two batches of fish from the above population were generated for the experimental challenge test. The first population consisted of 1,090 fish (10 fish per family) and used for the IP challenge test, while 2,398 (22 fish per family) fish were used in the cohabitation challenge test. The average weight of the fish at start of the challenge test was 145 g. Both IP and Cohabitation challenge tests were undertaken in seawater at 15 °C and 32 ppt salinity. For the IP challenge test, the fish were equally split into three tanks and each fish was injected in the intraperitoneal region with 0.1 ml dose containing 3.5×10^3 colony forming units (cfu) of *P. salmonis* EM90 genogroup. Mortalities were recorded daily until the end of the challenge test (32 days post challenge). For the cohabitation challenge test, the fish were equally split into two tanks. A group of shedder fish (17% of the fish in the cohabitation challenge test) were injected in the intraperitoneal region with an 0.1 ml dose containing 1.1×10^5 cfu of *P. salmonis* EM90 genogroup. Dead fish were collected daily until 70 days post challenge when the experimental challenge test was terminated.

Phenotypes. We used survival days (number of days fish survived during the challenge test; all alive fish at the end of the experiment were assigned the maximum day post challenge) and binary survival phenotype (dead = 0, alive = 1 at the end of the challenge) for the analysis of resistance to SRS infection.

Statistical analysis. Genetic parameters were estimated using pedigree information for survival days (SurvDays) and for the binary survival (Surv) using a bivariate animal model within the challenge test model with the statistical model below:

$$Y = XB + ZU + WC + E,$$

where Y is the matrix of phenotypic records of two traits for all animals, X in the incidence matrix corresponding to fixed effect of tanks for the two traits, B is the matrix of fixed effect solution for tanks, Z is the incidence matrix corresponding to the random genetic effects of animals, U is the matrix of random additive genetic effects of animals with variance structure $G \otimes A$, where G is the genetic variance-covariance matrix among the two traits and A is the numerator relationship matrix constructed using pedigree information, W is the incidence matrix for the random common effects to full-sibs, C is the matrix of common effects to full-sibs, E is the matrix of random residual effects with a variance structure $R \otimes I$, where R is matrix of residual variances and covariance which was assumed to be non-zero. The same bivariate animal mixed model was applied to genetic correlations between the phenotypes from the IP and cohabitation challenge test models, except that residual covariance was assumed to be zero.

Results

The percentage mortality was 64 and 43% for IP and cohabitation challenge test model, respectively. As expected, the duration of the IP challenge test model was 45% lower than the cohabitation challenge test which ended after 70 days post challenge (Figure 1). Due to human error, we had to abruptly terminate one of the tanks for the cohabitation challenge test 48 days post challenge. The percentage mortality at the time of the termination was 32%. Initial results showed that the genetic correlation between tank 1 and tank 2 for the mortality and survival days of the cohabitation test was 0.917 (0.073), therefore the complete dataset was used for the final analysis.

Heritability estimate for survival days ranged from 0.207 to 0.241 and 0.175 to 0.253 for the binary survival phenotype (Table 1). The variance ratio of the effect common to full-sib ranged was about 14 and 2% in the IP and cohabitation challenge test model, respectively.

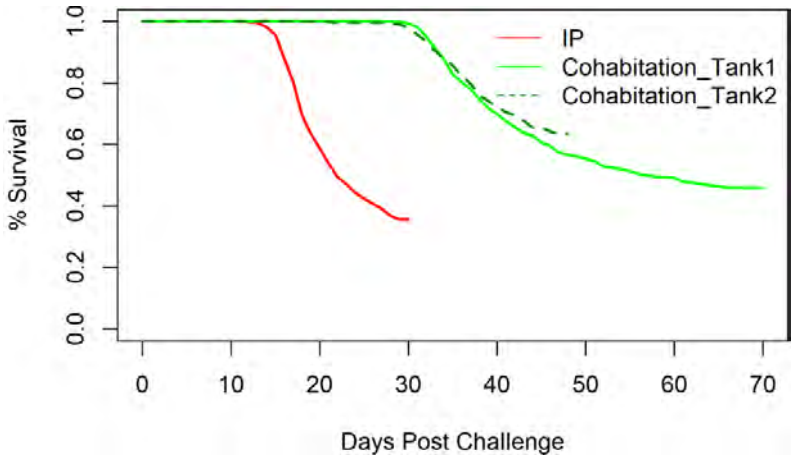


Figure 1. Survival curve for intraperitoneal (IP) and cohabitation challenge test model for SRS.

Table 1. Heritability on the diagonal and genetic correlation below the diagonal with standard errors in parentheses.

Challenge model	Trait	IP[1]		Cohabitation	
		SurvDays	Surv	SurvDays	Surv
IP	SurvDays[2]	0.241 (0.175)			
	Surv[3]	0.993 (0.043)	0.175 (0.148)		
Cohabitation	SurvDays	0.933 (0.291)	0.822 (0.353)	0.207 (0.081)	
	Surv	0.889 (0.315)	0.825 (0.350)	0.983 (0.019)	0.253 (0.085)

[1] IP = intraperitoneal challenge test model.

[2] SurvDays = survival days.

[3] Surv = binary survival.

The genetic correlation between survival days and binary survival was very high (>0.980) for both challenge test models. Finally, the genetic correlation between IP and cohabitation challenge test for survival days was 0.933 and 0.825 for binary survival.

Discussion

The aim of the study was to estimate heritability and the genetic correlation between IP and cohabitation challenge test models for resistance to SRS. As expected mortalities from the IP challenge model started 12 days post challenge, while the cohabitation test started 29 days post challenge. Similar trend in survival/mortality curve has been reported by Meza *et al.* (2019). The shorter challenge test duration from the IP model makes it a cheaper challenge test model to obtain phenotypes for selecting for resistance (arguably tolerance and/or recovery) to SRS infections.

The heritability estimate from both models were within the range of values reported (Lhorente *et al.*, 2014; Moraleda *et al.*, 2021; Yáñez *et al.*, 2014**).** Further analysis of the binary survival phenotype using a sire-dam threshold model resulted in heritability estimate on the liability scale of 0.246 (0.127) and 0.324 (0.096) for the IP and cohabitation challenge models, respectively.

Surprisingly, the estimate of the genetic correlation between IP and cohabitation challenge test models was high. The cohabitation challenge model mimics the natural route of infection, therefore multifaceted immune response (e.g. skin and gills) can be mounted by the fish against the infection, however the IP challenge test model bypasses the initial response barrier and therefore is expected trigger part of the resistance to SRS by the fish. Nonetheless, we still observe a high genetic correlation between the IP and the cohabitation challenge test. Joshi *et al.* (2020) recently reported a high (0.9) for survival of Nile tilapia to *Streptococcus agalactiae* between the IP and the cohabitation challenge test.

Conclusions

SRS is a heritable trait in breeding population of Mowi in Chile, with values ranging from 18 to 25%. Genetic correlation of IP vs cohabitation challenge test model was high, therefore we could select for SRS resistance using phenotypes generated from an IP challenge model.

References

Departamento de Salud Animal, Informe Sanitario de la Salmonicultura en Centros Marinos AÑO 2020. Available at: http://www.sernapesca.cl/sites/default/files/informe_sanitario_salmonicultura_en_centros_marinos_2020v2.pdf

Joshi R., Skaaurd A., and Tola Alvarez, A. (2021) J. Anim. Breed. Genet. 138:338–348. https://doi.org/10.1111/JBG.12516

Lhorente J.P., Gallardo J.A., Villanueva B., Carabañ O M.J., and Neira R. (2014) PLoS ONE 9(4):e95397. https://doi.org/10.1371/journal.pone.0095397

Meza K., Inami M., Dalum A.S., Bjelland A.M., Sørum H. *et al.* (2019) J. Fish Dis. 42:1001–1011. https://doi.org/10.1111/JFD.13004

Moraleda C.P., Robledo D., Gutiérrez A.P., del-Pozo J., Yáñez J.M. *et al.* (2021) BMC Genomics 22:1–13. https://doi.org/10.1186/S12864-021-07443-2/FIGURES/6

Yáñez J.M., Lhorente J.P., Bassini L.N., Oyarzún M., Neira R. *et al.* (2014) Aquaculture 433:295–298. https://doi.org/10.1016/J.AQUACULTURE.2014.06.026

569. Using 'population-specific haplotypes' to select for resistance to SRS in an admixed population of coho salmon (*Oncorhynchus kisutch*)

T. Moen[1*], J. Ødegård[1], F. Grammes[1], T. Knutsen[1], R. Deerenberg[2] and D. Chichero[2]

[1]AquaGen Norway, Postboks 1240, Torgard, Trondheim, Norway; [2]AquaGen Chile, San Francisco 340, Puerto Varas, Chile; thomas.moen@aquagen.no

Abstract

A genome-wide association study (GWAS) was used to search for QTL for resistance to Salmonid Rickettsial Salmonis (caused by *Septicaemia piscirickettsia*) in coho salmon. The coho population under study was admixed, consisting of animals from three pre-existing and distinct populations of domesticated coho. Although a major QTL was found, no single SNP was able to tag the QTL well in all three subpopulations. A system for tracking the good allele at the QTL (Q) as well as the sub-population Q was derived from was developed, facilitating more precise selection in the admixed population as well as in other populations. Later, new and better tag-SNPs were identified through fine-mapping, increasing the selection accuracy even further.

Introduction

Marker-assisted- as well as genomic selection is reliant upon DNA-markers which are sufficiently correlated to the DNA polymorphisms which affects the trait in question. This dependency is particularly pronounced when the training data, coming from relatives of the breeding candidates carrying both genotypes and phenotypes, cannot be refreshed frequently. In aquaculture, we often select for resistance to multiple pathogens, and it is not always cost-effective to measure the resistance to every pathogen in every new generation. In such cases, one may be tempted to rely on marker-trait associations identified in one or just a few underlying data sets, and even to apply marker-trait associations identified in one population on another population. This paper illustrates that this is not always a good idea and presents a generic QTL-tagging system applicable in admixed populations, based on the simultaneous detection of the beneficial allele and the population of origin of that allele. The method was applied to selection for increased resistance to the intracellular bacteria *Piscirickettia salmonis* in coho salmon (*Oncorhynchus kisutch).*

Materials & methods

Challenge tests. Challenge tests for resistance to *P. salmonis* were carried out using a cohabitant challenge model. The challenge was terminated 60 days post-infection.

Genotyping. Genotyping was carried out using Axiom Custom Arrays from ThermoFisher. Three arrays were used in the study: (1) Oki220k, the original array, containing 220k SNPs; (2) Oki72kv1, a lower-density array containing the 72k most informative SNPs from Oki220k; (3) Oki72kv2, similar to Oki72kv1 except that SNPs performing sub-optimally had been replaced by 8k SNPs located within the region of a QTL for SRS-resistance.

Genome-wide association study (GWAS). Standard GWASes were performed using the 'mixed linear model leaving-one-chromosome-out' (LOCO) option of GCTA.

Association analysis using combinations of SNPs. Within one particular chromosome, combinations of SNPs were tested for their association to the trait, as follows: (1) a data set (consisting of challenge-tested animals from a single population or a single sub-population) was phased using FImpute; (2) a region of interest was pinpointed, wherein a quantitative trait nucleotide (QTN) was thought to be located, and from

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_569, © T. Moen *et al.* 2022

this region, the genotypes of all polymorphic SNPs were extracted; (3) for each SNP-combination length L from 1 to 20, step 4 was carried out; (4) for each combination of L consecutive SNPs within the region, steps 5 to 7 were carried out; (5) the haplotypes carried by each animal were determined using the phased data set; (6) a mixed linear model was solved in DMU:

$y = 1\mu + Hh + g + e$

where y is a vector of binary phenotypes (survived = 1, dead = 0), μ is the overall mean, h is a vector of fixed haplotype effects, $g \sim N(0, A\sigma_g^2)$ is a vector of random additive genetic effects (derived from the pedigree), $e \sim N(0, I\sigma_e^2)$ is a vector of random residuals, H is a haplotype incidence matrix and A is the numerator relationship matrix; 7) a likelihood-ratio statistic for the effect of the haplotypes on the trait was calculated as $2\ln(L_{H1}) - 2\ln(L_{H0})$, where L_{H0} is the likelihood under the reduced model (no haplotype effects) and L_{H1} is the likelihood under the full model.

Identification of 'population-specific haplotypes'. A 'population-specific haplotype' was defined as a haplotype which was strongly associated to the good allele (Q) at the QTL in one of three subpopulations of an admixed population, while being absent (or having an allele frequency lower than 0.05) in the other two populations. To find such population-specific haplotypes, we first carried out association analysis using SNP-combinations, as described above. For every SNP-combination length from 1 to 20, we subsequently scanned all SNP-combinations for the presence of such population-specific haplotypes.

Genotyping of population-specific haplotypes. Once a population-specific haplotype had been found for each of the three subpopulations, the SNPs constituting all haplotypes could be genotyped in breeding candidates. The resulting genotypes were phased using the PHASE, whereupon each breeding candidate was checked for the presence of the population-specific haplotypes. If none of those were present, the animal as assigned genotype qq, if one was present it was assigned genotype Qq, and if two of three were present it was assigned genotype QQ.

Results

Initially, a challenge-test for SRS-resistance was performed on AquaGen's 2014 year-class of coho salmon. This year class was the founder population of AquaGen' 'even' year class population of coho (the species has a strict 2-year generation interval), and it was constituted by fish originating from three pre-existing and distinct populations of domesticated coho, purchased by AquaGen.

A GWAS was run on each of the three subpopulations constituting the C14 challenge test, named C14A, C14B, C14C hereafter. A large QTL for resistance to SRS was found on chromosome 21, explaining much of the genetic variation within individual subpopulations (Figure 1). However, no single SNP was able to tag the underlying QTN well in all three subpopulations. Similarly, an association analysis based on SNP-combinations was attempted, but no combination of SNPs was able to tag the QTN well.

Observing that the most significant SNPs from the GWAS performed on C14A, C14B, and C14C had: (1) a weak association to each other; and (2) low frequency of the beneficial allele (Q), we hypothesised that it might be possible to find haplotypes (multi-locus 'alleles' at SNP-combinations) which would be strongly associated to Q in one of the populations but absent in the other two. If one or more such 'population-specific' haplotypes could be found for each subpopulation, then these three or more haplotypes could jointly constitute a DNA-test for the QTN.

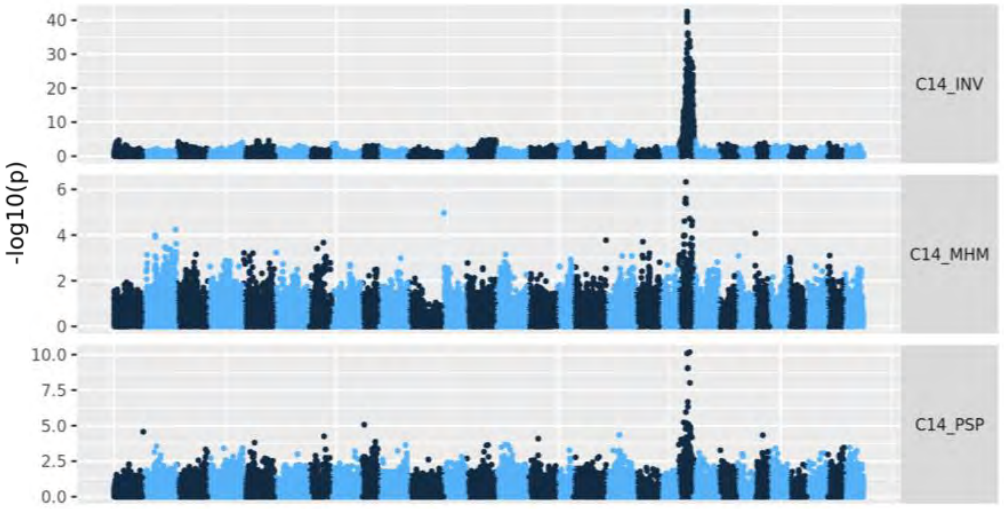


Figure 1. GWAS for resistance to *Piscirickettsia salmonis*.

The optimal combination of population-specific haplotypes was constituted by three haplotypes, each made from three distinct SNPs, i.e. nine SNPs were involved altogether. In terms of base pairs, the haplotypes covered 21 kb, 49 kb, and 56 kb, and the distance between them was 0.3 Mb and 4.6 Mb. Table 1 displays the allele substitution effect and the allele frequency of each haplotype within each of the three subpopulations.

The two-allele 'DNA-marker' derived from the three haplotypes (named 'HAP' from here on) was used for selection for a while, in situations where the breeding candidates did not have siblings with genotypes and phenotypes. In 2021, a new version of the 72k SNP-chip was made. This contained 8,342 more SNPs from within the QTL region, of which 10 were added due to their high LD to HAP (r^2>0.8). The new chip was used for genotyping a subset of the samples taken from C14A, C14B, C14C, as well C18 (grandchildren of C14), C17, and C19 (two consecutive generations of the 'odd' year class population).

Interestingly, when a GWAS was performed on the data from the new chip, two of the SNPs that had very high LD to HAP came out on top, being substantially more significant than the other SNPs.

Figure 2 shows how large a fraction of the genetic variance is explained by the original best tag-SNPs from the C14A, C14B, C14C, C17, C18, and C19 data sets, within those same data sets. HAP and the best single SNP from Oki72kv2 are also included. One can see that HAP is consistent across data sets, while none of the original tag-SNPs are. HAP explained 30% of the genetic variation across data sets. Within individual

Table 1. Properties of the best combination of population-specific haplotypes

Haplotype	Allele substitution effect[1] and frequency in:		
	C14A	C14B	C14C
C14A	0.21/0.27	NA/0.0	-0.001/0.003
C14B	0.01/0.001	0.26/0.08	0.32/0.02
C14C	0.02/0.002	NA/0.0	0.32/0.08

[1] The trait is survival (0 = affected, 1 = non-affected).

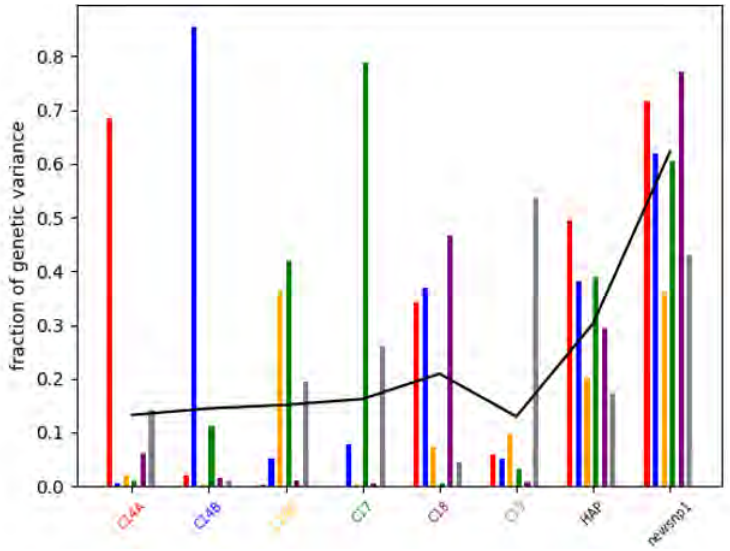


Figure 2. Fraction of genetic variance explained by various tag-SNPs.

data sets, the best tag-SNP from that population always outperforms HAP, as can be expected. The best SNPs from Oki72kv2 was both consistent across data sets, and outperformed the best original tag-SNPs in most of the data sets. Across data sets, the new best tag-SNP explained 60% of the genetic variation.

Discussion

The population-specific haplotype approach described here may be useful only under some particular circumstances. It is interesting to see that the approach was able to improve the overall accuracy of selection, thus solving the problem at hand. The case of SRS on coho salmon serves to illustrate that care should be taken before marker-trait associations identified in one population is applied on another, also when the effects of the markers is large. The results also highlight the importance of doing fine-mapping to optimise the tools of selection.

570. Potential of imputation for cost-efficient genomic selection for resistance to *Flavobacterium columnare* in Rainbow trout

C. Fraslin[1*], A. Nousiainen[2], H. Koskinen[2], A. Kause[3] and R.D. Houston[1]

[1]The Roslin Institute and Royal (Dick) School of Veterinary Studies, University of Edinburgh, Easter Bush, Midlothian, EH25 9RG, United-Kingdom; [2]Natural Resources Institute Finland (Luke), Neulaniementie 5, 70210, Kuopio, Finland; [3]Natural Resources Institute Finland (Luke), Myllytie 1, 31600, Jokioinen, Finland; clemence.fraslin@roslin.ed.ac.uk

Abstract

Columnaris disease (CD) is a major emerging disease affecting rainbow trout aquaculture. Selective breeding through genomic selection (GS) has potential to improve host resistance. However, GS is expensive partly due to the cost of genotyping high numbers of animals using high-density (HD) SNP arrays. The objective of this study was to assess the efficiency of GS for resistance to CD using ten *in silico* LD-panels combined with imputation. After a natural outbreak of CD, 2,874 challenged fish and 469 fish from the parental generation (n=81 parents) were genotyped with 27,907 SNPs. Using LD-panels (3,000 markers) alone or lower density panels (300 markers) along with imputation resulted in comparable accuracy to the 28K HD-panel, and 11% higher accuracy than pedigree-based predictions. Compared to using the commercial HD-panel, LD-panels with imputation may provide a more affordable route to genomic prediction of breeding values, supporting wider adoption of GS in aquaculture breeding programmes.

Introduction

Columnaris disease (CD) is an emerging problem for various aquaculture species worldwide. *Flavobacterium columnare*, the pathogen agent of CD, causes high levels of mortality in rainbow trout fry, causing significant annual economic losses to the industry. Improvement of disease resistance through selective breeding is a promising approach to reduce mortality, and could reduce disease incidence and reduce antibiotic use. Genomic selection (GS) uses genome-wide marker information gathered in a reference population to predict the breeding value for traits of economic interest in a target broodstock. GS is particularly applicable for traits difficult to measure in breeding candidates (such as disease resistance traits) and has been shown to be more accurate, and thus more efficient, than pedigree-based selection. However, the cost of genotyping large numbers of individuals required to perform GS is still prohibitive for many breeding programmes. Using low-density (LD) SNP panels could reduce the cost of genotyping (since LD panels are typically cheaper to genotype) while still achieving higher prediction accuracy than standard pedigree-based selection. Genotype imputation from low-density to high-density genotypes is therefore a promising approach to achieve a more affordable genomic selection. The aim of this study was to assess the efficiency of GS using low-density SNP panels along with imputation for this disease resistance trait.

Materials & methods

Fish rearing and disease outbreak management. In May 2019, 81 rainbow trout (33 females and 48 males) were selected among 567 fish from the Finnish national breeding programme maintained by Luke, based on their relationship and genetic contribution to maintain a predetermined inbreeding coefficient of less than 1% per generation. The 33 dams and 48 sires were mated to create 105 full-sib families. In June 2019, around 30,000 fry were separated into three fingerling tanks, about 100 fish per family per tank, at the farm of Hanka-Taimen Oy (Finland). From arrival at this new farm (considered as day 0 of the experiment), the fish mortality and any disease signs were monitored twice a day. On day 11 of the experiment, fish in all three tanks started to show signs of CD, some dead fish were sampled and sent to a veterinarian to confirm the CD diagnosis. From day 15 to 24, a piece of tail was taken, for later DNA extraction, from 510 fish per

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_570, © C. Fraslin *et al.* 2022

tanks, randomly chosen amongst the dead or dying fish with clear CD signs (considered as susceptible). At day 26, the three tanks were treated following the veterinarian guidelines against *F. columnare* with an approved treatment until day 32. On the last day of the experiment, day 99, a piece of tail was collected, for latter DNA extraction, on about 506 fish per tanks, randomly sampled among the fish still alive at that time (considered as resistant). In addition, the biggest fish among the sampled fish from each tank (n=164, 167, 168 for tanks 1, 2 and 3, respectively) were PIT-tagged and are considered as breeding candidates for the future broodstock

Genotyping, in silico low-density panels, and imputation. A total of 3,054 challenged fish (1,538 susceptible and 1,519 resistant) and 570 fish from the parental generation (including the 81 parents) were genotyped using the 57K SNP Axiom™ Trout Genotyping Array (Palti *et al.*, 2015). After quality controls, 2,874 challenged fish and 469 fish from the parental generation (n=78 parents of the challenged fish) genotyped with 27,907 informative SNPs were retained. Those 28K SNPs were considered as the high-density (HD) panel. The impact of reducing the SNP density on genomic prediction was tested with ten low-density (LD) SNP panels, created *in silico* (300; 500; 700; 1000; 3,000; 5,000; 7,000; 10,000; 15,000 and 20,000 SNPs). For each LD-panel, SNPs were sampled from the 28K HD-panel using the CVrepGPAcalc package (Tsairidou *et al.*, 2020). Two sampling approaches were used to create LD-panels: (1) SNPs were randomly sampled within each chromosome, with the number of SNP sampled from a given chromosome being proportional to the physical length of the chromosome in the *O. mykiss* reference genome (Omyk_0.1, Gao *et al.*, 2018); (2) SNPs were selected so that they would be equally spaced within each chromosome. Each LD-panel was replicated 10 times, with replicates allowed to overlap by chance and the final number of SNPs within each panel being allowed to slightly vary from the target density. FImpute3 (Sargolzaei *et al.*, 2014) was used to impute the *in silico* LD- genotyped offspring back to the full 28K SNPs using a combined population and pedigree based imputation method with the HD-genotyped parents (n=469) as the reference population.

Genetic parameter estimates and genomic selection. Disease resistance was analysed as a binary trait (0 = alive; 1 = dead) with the rearing tank as a fixed effect in the statistical model. Variance components and heritabilities were estimated using ASReml (v.4.1, Gilmour *et al.*, 2014) with two different approaches, a mixed linear BLUP animal model and a logistic regression model to assess the trait on the observed and underlying scale, respectively.

The (genomic) estimated breeding values {(G)EBV} of fish were computed using mixed linear BLUP animal model based on pedigree (PBLUP) or genomic (GBLUP) information using the BLUPF90 software (Misztal *et al.*, 2002). The efficiency of selection using pedigree-based or genomic information was estimated using Monte-Carlo 'leave-one-group-out' method by removing the known phenotype from 20% of the fish, and then using the phenotype and genotype data of the remaining 80% fish to predict the (G)EBVs of the 20% validation fish group. This procedure was repeated 20 times for the PBLUP, the HD-GBLUP, and for each of the 10 replicates of the LD-panels, pre and post-imputation. Accuracy of prediction was computed, for all SNP panels, as the mean over the 20 replicates of the correlation between the (G)EBV and the true phenotype of fish in the validation group, divided by the square root of the genomic-based heritability.

Results

Resistance to *F. columnare* was moderately heritable in this population with pedigree-based heritability estimated to be 0.18 (±0.038 se) and 0.35 (±0.046) on the observed and underlying logit scales, respectively. Genomic heritability was estimated to be 0.21 (±0.029 se) and 0.43 (±0.042 se) on the observed and underlying logit scales respectively. Pedigree-based prediction accuracy was estimated as 0.59 (±0.080 sd) and the use of genomic evaluation increased the prediction accuracy by 14% for the HD-GBLUP

(0.68±0.076 sd). Decreasing the number of SNPs tended to decrease the accuracy of genomic predictions (Figure 1), and no significant difference was observed between the two methods of sampling.

Prediction accuracies obtained with 300 to 500 SNPs were close to the accuracy obtained with the pedigree-based analysis (blue dotted line in Figure 1). Encouragingly, prediction accuracies obtained with LD-panels from 7,000 SNPs and above were close to the accuracy obtained with the HD-panel (-1% in accuracy compared with the HD-GBLUP). Accuracies obtained with 1000 SNPs were only 4% higher than those obtained with the pedigree, whereas the accuracy obtained with only 3,000 SNPs was 3% lower than the accuracies obtained with the HD-panel and thus 11% higher than the accuracy obtained with the pedigree only.

At the lowest densities (<3,000 SNPs, Figure 1) imputation (with FImpute3) had a positive impact on the accuracy of genomic prediction with the accuracy of genomic prediction obtained after imputation similar to the accuracy obtained with 3,000 SNPs.

Surprisingly, the accuracy of genomic prediction obtained with 3,000 SNPs was not significantly different before and after imputation and for 5,000 SNPs or higher, the accuracy of genomic prediction obtained after imputation was slightly lower than without imputation.

Discussion

The results presented here suggest that resistance to *F. columnare*, is moderately heritable in this population and could be improved through selective breeding. The results on genomic prediction are consistent with those from various aquaculture species that relatively LD SNP panels are, at least more efficient than pedigree-based selection and sometimes as efficient as full HD panels to accurately predict breeding values (see Houston *et al.*, 2020 for review). More specifically, using LD SNP panels (e.g. 3,000 to 7,000 markers) would results in comparable selection accuracy to the full 28K HD SNP panel. Imputation of very low density SNPs panels (e.g. 300-700 SNPs) resulted in an increase in genomic prediction accuracies compared to the low density panel alone, but this was not observed with imputation from higher densities (e.g. >5,000 SNPs).

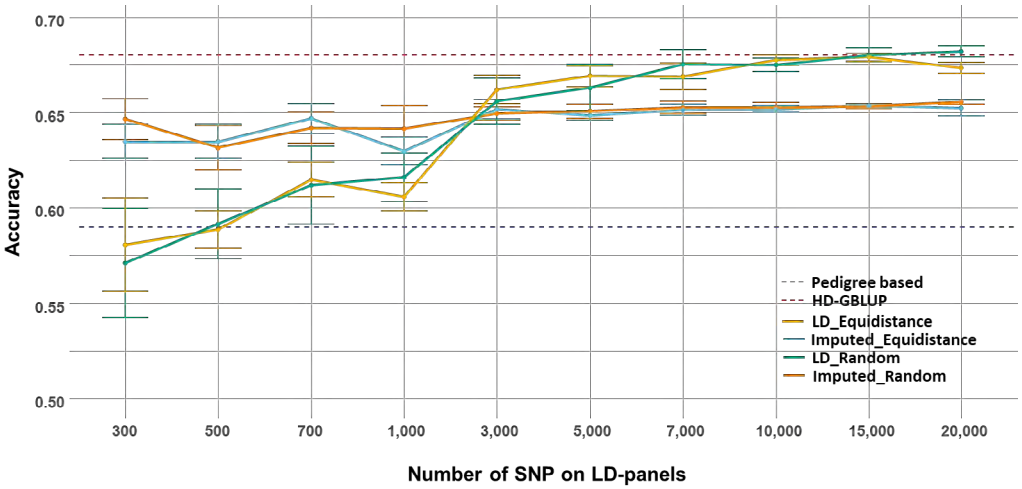


Figure 1. Accuracy of genomic prediction for resistance to *F. columnare* in rainbow trout, obtained with different SNP panels density, before and after imputation.

For rainbow trout, the high-density SNP array (57K SNPs) costs approximately 20€ per sample when genotyping a breeding population of 8,000 offspring and 200 parents. This is a total cost of about 164K €, which can be highly prohibitive for most breeding programmes. To implement genomic selection in their breeding programme, our results imply that the use of a cheaper low-density SNP array might be beneficial. For example, under the assumption a 3K SNP panel or array could be typed for all 8,200 fish at a cost of 15€ per sample, this would mean a 25% reduction in genotyping cost for a 3% decrease in accuracy compared to the full HD panel. Another possible scenario is that all offspring were typed for a very low density SNP panel (300 SNPs) at a cost of 7.5€ per sample, with parents typed for the existing 57K array at a higher per sample cost of 40€ per sample (the price is highly dependent of the number of samples genotyped), genotyping would cost 68K€. This reduces the genotyping cost by 59% compared to the price of the 57K SNP for only a 4% decrease in accuracy using the imputation approach. Moreover, *F. columnare* infects small fish, well before they can be individually tagged and identified. Therefore, even for an accurate pedigree-based evaluation, the offspring and parents needs to be genotyped in order to recover the pedigree, meaning that the 300 SNP panel could be combined purpose for both parentage assignment and imputation-based genomic selection.

In conclusion, genomic selection using low-density SNP panels may reduce costs of genomic selection in rainbow trout breeding without major reduction in breeding value prediction accuracy. Using low-density SNP panels (about 3,000 markers) or very low-density SNP panels (about 300 markers) combined with imputation may be a cost-effective genomic prediction of breeding values to improve resistance to *F. columnare*, in this rainbow trout population.

References

Gao, G., Nome, T., Pearse, D.E., Moen, *et al.*, (2018). Front. Genet. 9. https://doi.org/10.3389/fgene.2018.00147

Gilmour, A.R., Gogel, B.J., Cullis, B.R., *et al.*, (2015). ASReml User Guide Release 4.1 Functional Specification. VSN International Ltd, Hemel Hempstead, HP1 1ES, UK.

Houston R.D., Bean T.P., Macqueen D.J., Kumar Gundappa M. *et al.*, (2020). Nat. Rev. Genet. 211,389-409. https://doi.org/10.1038/s41576-020-0227-y

Misztal, I., Tsuruta, S., Strabel, T., Auvray, B., *et al.*, (2002) Proc. Of the 7th WCGALP, Montpellier, France.

Palti Y., Gao G., Liu S., Kent M.P., *et al.*, (2015) Mol. Ecol. Resour. 15,662-672. https://doi.org/10.1111/1755-0998.12337

Sargolzaei, M., Chesnais, J.P. and Schenkel, F.S. (2014) BMC Genom. 15, 478. https://doi.org/10.1186/1471-2164-15-478

Tsairidou, S., Hamilton, A., Robledo, D., Bron, J.E., Houston, R.D. (2020). G3. 10, 581–590. https://doi.org/10.1534/g3.119.400800

571. Genomic prediction of complex traits in European sea bass: random cross-validation and leave-one-family-out validation

S. Faggion[1*], D. Bertotto[1], V. Bonfatti[1], M. Freguglia[2], L. Bargelloni[1] and P. Carnier[1]

[1]Department of Comparative Biomedicine and Food Science, University of Padova, Viale dell'Università, 16, 35020 Legnaro (Padova), Italy; [2]Valle Cà Zuliani Società Agricola s.r.l., Via Gardizza, 9/B, 48017 Conselice (Ravenna), Italy; sara.faggion@unipd.it

Abstract

For European sea bass (*Dicentrarchus labrax* L.) hatcheries, viral nervous necrosis mortality, post-stress cortisol concentration, antibody titer against nervous necrosis virus and body weight are interesting traits to be included in the breeding objectives. An experimental population (n=650) generated using a commercial broodstock was phenotyped for those traits and genotyped with a genome-wide SNP panel (16,075 markers). The accuracy of the genomic model in the prediction of estimated breeding values was assessed in a 5-fold random cross-validation and in a leave-one-family-out scenario, where the prediction equation was obtained by training the model on the estimated breeding values of all animals except the offspring of the parent used as test set. The accuracies (Pearson's r) were different across traits and equal to 0.90 (random cross-validation) and 0.59 (leave-one-family-out) for mortality, 0.86 and 0.60 for cortisol concentration, 0.79 and 0.01 for antibody titre, and 0.71 and 0.18 for body weight.

Introduction

In fish species, selective breeding to enhance disease resistance is considered a feasible and sustainable approach to prevent and control mortality from disease outbreaks (Ødegård *et al.* 2011). For the European sea bass (*Dicentrarchus labrax* L.) industry, viral nervous necrosis (VNN) is a major threat. Due to a moderate heritability and a significant negative genetic correlation with VNN mortality, antibody titre against nervous necrosis virus (NNV) antigens is a reasonable indicator trait to enhance VNN resistance. Conversely, serum post-stress cortisol concentration shows no genetic relationship with mortality (Faggion *et al.* 2021). However, this trait is important in a production context because stress tolerant fish have better efficiency, productivity and resilience (Castanheira *et al.* 2015). Traditional selective breeding approaches rely on estimated breeding values (EBV) predicted using individual phenotypes of the breeding candidates or their relatives through estimated additive genetic relationships based on their pedigrees. Routine individual phenotyping for complex traits such as VNN resistance, antibody titre or cortisol concentration is either difficult or unfeasible due high costs and time requirements, and the implementation of genomic selection procedures might be greatly beneficial. In this study, we assessed, the accuracy of a genomic prediction model using both random 5-fold cross-validation and, to mimic a real genomic selection scenario, leave-one-family-out validation.

Materials & methods

Production of experimental fish and phenotyping. The experimental population was produced by artificial fertilization using a commercial NNV-free breeding stock following a mating scheme based on an incomplete factorial design (sire-to-dam mating ratio: 8 to 5).

A 29 days-NNV challenge test was carried out using 650 randomly chosen fish (age: 548 d post-hatching, average body weight, BW: 146.4 g, CV% 29.18). Before being infected through intraperitoneal injection (0.1 ml, RGNNV 283.2009, batch 7/16; $10^{8.30}$ $TCID_{50}$ per ml), fish were subjected to an acute stress-test (confinement at high density), anesthetized (MS-222), individually PIT-tagged and blood sampled. Mortality (MORT) was recorded as 1 (dead) and 0 (survived). The survival rate at the end of the challenge

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_571, © S. Faggion *et al.* 2022

test was 52%. We used indirect ELISA assay to quantify antibodies against NNV (antibody titre, AT) in the blood serum of each fish following the protocol of Nuñez-Ortiz *et al.* (2016). Cortisol was extracted from blood serum following the protocol of Bertotto *et al.* (2010), and the concentration of cortisol (HC) of each sample was analysed using a solid-phase radioimmunoassay (Simontacchi *et al.* 1995).

Genotyping. Genotypes were obtained through the 2b-RAD method (Wang *et al.* 2012). After quality controls and imputation (FImpute, Sargolzaei *et al.* 2014), the number of available SNP genotypes per animal was 16,075. Parentage assignment and pedigree reconstruction were performed combining the software CERVUS 3.0 (Kalinowski *et al.* 2007) and the R package sequoia (Huisman 2017).

Pseudo-phenotypes. Breeding values (EBV) for the traits of interest were predicted using mixed animal models implementing a Bayesian approach using the TM software package (Legarra *et al.* 2008). For MORT, the variation in the phenotypic expression of the trait (y = 0 for survivors, y = 1 for the dead fish) was assumed to be dependent upon the value of an underlying continuous liability trait λ (Falconer and Mackay 1996) and analysed using a threshold model. The liability was modelled as:

$$\lambda = 1\mu + Xb + Wu + e \quad (1)$$

where λ is a vector of underlying liabilities, μ is the model intercept, 1 is an all-ones vector, b is a vector of unknown fixed parameters (fixed effects) due to the tank, X is an incidence matrix relating λ to b, u is a vector of unknown random animal additive genetic effects and e is a vector of random residuals.

For BW, HC and AT, the prediction of EBV was based on the following linear mixed model:

$$y = 1\mu + Wu + e \quad (2)$$

where y is a vector of observed phenotypes and 1, μ, W, u, e have the same meaning as above. Variance components for random effects, used to solve the model, were those reported in Faggion *et al.* (2021). The Gibbs sampler was run using 100,000 iterations with a burn-in of 5,000 iterations and a thinning interval of 10 samples.

Bayesian regression model and prediction performance. Genotypes at the 16,075 SNPs were used as predictors of the EBV by means of a Bayesian Ridge Regression (BRR) model fitted to the data using the package BGLR/R (Pérez and de Los Campos 2014) and having the following form:

$$y = 1\mu + \sum_{j=1}^{S} x_j b_j + e \quad (3)$$

where y is a vector of EBV, μ is the model intercept, 1 is an all-ones vector, x_j is a vector of genotypes at SNP j, coded as the number of copies of the minor allele (0, 1 or 2), b_j is the allele substitution effect for the minor allele at SNP j (to be estimated), S is the total number of SNP (S=16,075) and e is a vector of random residuals which was assumed to be $N(0, I\sigma_e^2)$ where N( ) indicates a normal probability density function, I is an identity matrix of appropriate order and σ_e^2 is the residual variance. A single Gibbs chain of 800,000 samples was generated, with a burn-in of 10,000 iterations and a thinning interval of 50 samples.

Predictive accuracy of the regression model was computed in a 5-fold cross-validation (CV). The data (n=650) was randomly split into five equally-sized subsets. Four were used as a training set to obtain solutions for allele substitution effects, while the fifth served as a test set. Predictions for each test set were saved and aggregated with those of the other test sets. A leave-one-family-out (LOFO) validation procedure focused on the parents of the challenge-tested fish was designed to mimic a real genomic selection scenario. The parents of the

experimental fish were used one-at-a-time as an independent test set. A genomic prediction equation was obtained for each investigated trait by training the BRR model on the EBV of all animals except the offspring of the sire (or dam) used as test set. The EBV of the parent in the test set was then predicted from the parent genotypes using the genomic prediction equation. Prediction for each parent was saved and, at the end of the LOFO, aggregated with those of the other parents. The Pearson product-moment correlation (r) between the observed and predicted values was computed as a measure of prediction accuracy in both validations.

Results

Scatterplots of the relationships between the EBV and its genomic prediction obtained in a random 5-fold cross-validation and in the LOFO validation procedure are presented in Figure 1 and Figure 2.

Discussion

In a random 5-fold cross-validation, the genomic prediction of the EBV as pseudo-phenotypes led to very high accuracies for MORT and HC and moderately high accuracies for BW and AT. However, those accuracies are unrealistic. To obtain a more realistic insight, we developed a LOFO validation. This was done to test the consistency of prediction equations in a long-term view: in commercial breeding, genomic models are trained firstly using a reference population consisting of full- and half-sibs, but the prediction equations should perform well even in future generations, when it is likely that the animals to be predicted are distant relatives of the animals included in the reference population. The LOFO accuracies were, in fact, lower than the random 5-fold cross-validation accuracies, but to a different extent across traits. For MORT and HC, omitting the progeny information in the training of the genomic models did not affect the accuracy of the genomic predicted breeding value of the parents, suggesting that in a scenario where the reference population consists of distant relatives of the animals to be predicted, the predictive performances of such models would be still acceptable. This is of great interest in an applied context, particularly when considering traits phenotyping difficulties and complexity. Conversely, the correlations between the observed and the predicted EBV dropped remarkably for BW and AT. The degree of genomic similarity between reference and testing populations, size and composition of the reference population, and the density of the marker array (Zhang *et al.* 2019) could affect genomic prediction accuracy dependent on trait-specific features. We observed weak genomic relationships between the parents of the challenged fish

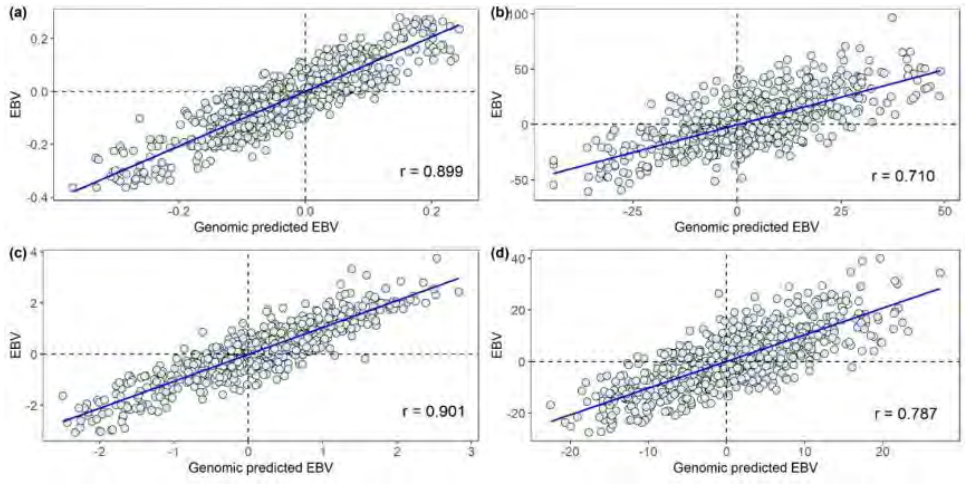


Figure 1. Relationships between EBV and their genomic predictions obtained in a random 5-fold cross-validation: (a) MORT; (b) BW; (c) HC; (d) AT. r: Pearson product-moment correlation.

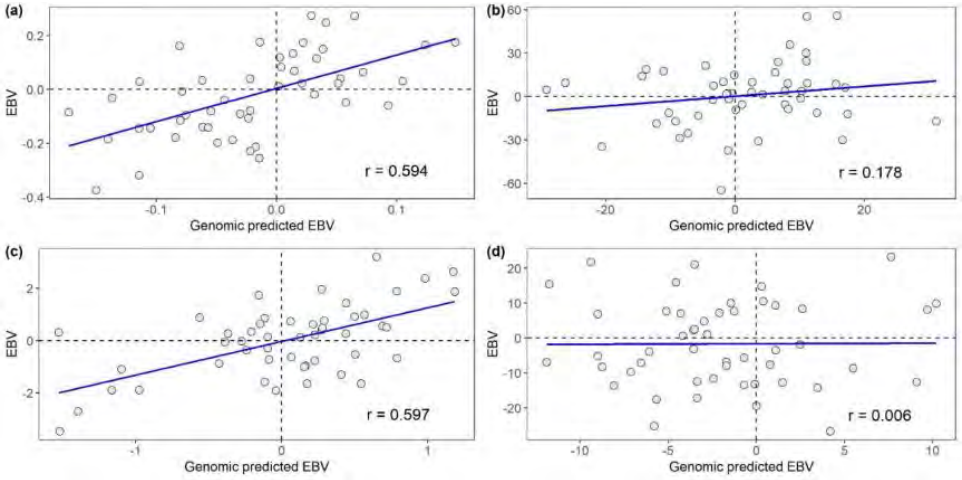


Figure 2. Relationships between EBV and their genomic predictions obtained for the parents of the challenged fish in a leave-one-family-out validation: (a) MORT; (b) BW; (c) HC; (d) AT. r: Pearson product-moment correlation.

(data not reported), implying that, in the LOFO procedure, the training of the genomic model was based on data of animals weakly related to the target individual. According to de los Campos *et al.* (2013), under perfect linkage disequilibrium between markers and causal loci, the prediction accuracy has an upper bound that depends on trait heritability and on realized genomic relationship between training and testing sets. Enlarging the size of the training population allows a high prediction accuracy (in terms of theoretical upper bound) even when genomic relationships are null or weak.

References

Bertotto, D.; Poltronieri, C.; Negrato, E.; Majolini, D.; Radaelli, G.; *et al.* (2010) Aquacult Res 41(8):1261-1267. https://doi.org/10.1111/j.1365-2109.2009.02417.x.

Castanheira, M.F.; Conceição, L.; Millot, S.; Rey, S.; Bégout, M.-L. (2015) Rev Aquacult, 9(1):23-41. https://doi.org/10.1111/raq.12100.

de los Campos, G.; Vazquez, A.I.; Fernando, R.; Klimentidis, Y.C.; Sorensen, D. (2013) PLoS Genet, 9: e1003608. https://doi.org/10.1371/journal.pgen.1003608.

Faggion, S.; Bertotto, D.; Babbucci, M.; Dalla Rovere, G.; Franch, R. (2021) Genet Sel Evol 53(1):32. https://doi.org/10.1186/s12711-021-00625-2.

Falconer, D.S. and Mackay, T.F.C. (1996) Introduction to quantitative genetics, 4th ed.; Longman: Harlow, England.

Huisman, J. (2017) Mol Ecol Res 17(5):1009-1024. https://doi.org/10.1111/1755-0998.12665.

Kalinowski, S.T.; Taper, M.L.; Marshall, T.C. (2007) Mol Ecol 16(5):1099-1106. https://doi.org/10.1111/j.1365-294X.2007.03089.x.

Legarra, A.; Varona, L.; López de Maturana, E. TM Threshold Model. Available at: http://snp.toulouse.inra.fr/~alegarra/.

Nuñez-Ortiz, N.; Stocchi, V.; Toffan, A.; Pascoli, F.; Sood, N.; *et al.* (2016) J Fish Dis 39(7):821-831. https://doi.org/10.1111/jfd.12415.

Ødegård, J.; Baranski, M.; Gjerde, B.; Gjedrem, T. (2011) Aquacult Res 42(s1):103-114. https://doi.org/10.1111/j.1365-2109.2010.02669.x.

Pérez, P.; de Los Campos, G. (2014) Genetics 198(2):483-495. https://doi.org/10.1534/genetics.114.164442.

Sargolzaei, M.; Chesnais, J.P.; Shenkel, F.S. (2014) BMC Genomics 15:478. https://doi.org/10.1186/1471-2164-15-478.

Simontacchi, C.; Bongioni, G.; Ferasin, L.; Bono, G. (1995) Atti XLIX Convegno Nazionale S.I.S.Vet., Salsomaggiore Terme (PR), Italy,

Wang, S.; Meyer, E.; McKay, J.K.; Matz, M.V. (2012) Nat Methods 9:808-810. https://doi.org/10.1038/nmeth.2023.

Zhang, H.; Yin, L.; Wang, M.; Yuan, X.; Liu, X. (2019) Front Genet 10:189. https://doi.org/10.3389/fgene.2019.00189.

572. Genetics of parasite attraction in Atlantic salmon: potential for group-level protection against sea lice

J. Ødegård[1*], S. Kjøglum[1], M. Medina[1], J.S. Torgersen[1], S.A. Korsvoll[1], R. Deerenberg[2], J.M. Yáñez[3], D. Cichero[2], P. Lopez[2] and T. Moen[1]

[1]AquaGen Norway, P.O. Box 1240, Torgard, Trondheim, Norway; [2]AquaGen Chile, San Francisco 340, Puerto Varas, Chile; [3]Facultad de Ciencias Veterinarias y Pecuarias, Universidad de Chile, Santiago 8820000, Chile; jorgen.odegard@aquagen.no

Abstract

Sea lice are important copepod ectoparasites in salmonid aquaculture. Individual lice count has moderate heritability under challenge testing and has been suggested as a selection tool for parasite control. Recent studies indicate that parasite attraction is a major underlying cause of genetic variation in lice count, which is likely to cause within-group variation in parasite load, but its potential for group-level protection is unknown. Results from an ongoing divergent selection experiment are presented, with two salmon strains divergently selected for low/high pathogen burden. Under common-garden testing, the strains were extremely different. However, when tested separately using standard protocols, the difference was not significant. An alternative testing protocol with flow-through tanks was also applied, offering less 'favourable' conditions for the parasites, indicating some group-level protection (P=0.10). Further research is needed to assess the potential of lice attraction as a tool for parasite control under large-scale field conditions.

Introduction

Sea lice (*Lepeophtheirus salmonis* and *Caligus rogercresseyi*) are important copepod ectoparasites of farmed salmonids in the northern and southern hemispheres, respectively). Spread of sea lice from salmonid farming to wild salmonid populations is considered a major threat. Selection for reduced individual lice count has been suggested to control parasite burden in salmonid fish farming, and moderate heritabilities have been estimated for the trait (Gjerde *et al.* 2011; Tsai *et al.* 2016). Recent studies have shown very high genetic correlations between early (2 days post infestation) and late (15 days post infestation) lice counts, indicating that parasite load is primarily driven by differences in initial infestation and not by immune-mediated resistance (results not shown). Furthermore, Atlantic salmon selected for reduced lice count have been found to have lower levels of kairomones compounds, used by sea lice to find suitable hosts, in their mucus than salmon selected for increased lice count (Difford *et al.* 2021). Hence, lice attraction is likely a main contributor to within-test genetic variation in lice count. AquaGen also performed a common-garden lice (caligus) challenge experiment on Atlantic and coho salmon (results not shown), where initial lice counts were much higher in coho than Atlantic salmon. However, 12 days post infestation, lice count in cohos approached zero, but were unchanged in Atlantic salmon. This may be viewed as species-specific defence strategies against sea lice, with Atlantic salmon utilizing an avoidance strategy and cohos utilizing an immune-mediated resistance strategy. AquaGen conducted a selection experiment with two divergent selection lines (LPB = low parasite burden, HPB = high parasite burden) for salmon louse count (*L. salmonis*, attached stage). To date, four consecutive generations have been tested (year-classes 2011, 2014, 2017 and 2020). Genomic selection was performed using training data from common-garden sib-testing. The first aim of this study was to quantify the differentiation of the selection lines, both genetically and phenotypically. The second aim was to confirm whether selection for reduced sea louse attraction enhances group-level protection in salmon within a tank environment.

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_572, © J. Ødegård *et al.* 2022

Materials & methods

Data. A total of 9,417 genotyped and phenotyped fish were included in the study. A genomic relationship matrix was computed using VanRaden method 1 (VanRaden 2008), utilizing data on 43,578 high-quality SNPs. Common-garden challenge tests were performed for all year-classes (Table 1). Some batches of fish were challenged twice (denoted A and B). For year class 2020, the common garden test (2020main) largely contained main nucleus families, while LPB and HPB selection strains were tested in a separate experiment (2020strain). Here, the two strains were tested in separate tanks (6 tanks per strain), and at high (25 lice/fish) and low (8 lice/fish) infestation pressure.

Under a standard challenge protocol, water level is lowered, and water flow stopped before lice copepodites are added. This may give the parasites easy access to hosts, potentially masking strain differences. A second set of tank experiments (2020flow and 2020standard) was therefore initiated. Here, fish were recruited from the main AquaGen nucleus, where two genetically differentiated groups were selected based on their genomic breeding values for lice count. These two groups were compared under both standard challenge (2020standard: 3 mixed tanks + 3 tanks per group) and flow-through challenge models (2020flow: 2 mixed tanks + 3 tanks per group). For the latter challenge model, lice copepodites were added to the tank under constant waterflow, i.e. the lice had a limited time window to attach to a host before being flushed out.

Prediction of GEBVs. The data sets were phenotypically heterogeneous (Table 1). Therefore, we fit a heterogeneous variance model for the genetic and residual variances:

$$y = X\beta + Zg + e$$

where y is a vector of phenotypes (log of lice count), β is a vector of fixed effects (year-class/challenge model/challenge group combination and regression coefficients for body size). Variance components for each group were estimated separately for the different challenge-tests using genomic animal models with the GCTA (https://yanglab.westlake.edu.cn/software/gcta/) software and used as input parameters in the final joint model. The mixed model equation system was solved using custom-made Julia scripts (https://julialang.org/).

Table 1. Number of genotyped individuals and estimated variance components for different sea lice challenge tests.

Test	Type	N	Gen. var. (SE)	Res. var. (SE)	h^2 (SE)
2011A	Sea cage	1,431	0.074 (0.017)	0.355 (0.017)	0.172 (0.036)
2011B	Sea cage	546	0.048 (0.020)	0.253 (0.022)	0.159 (0.065)
2014A	Mixed tank	1,040	0.048 (0.012)	0.195 (0.011)	0.198 (0.043)
2014B	Mixed tank	766	0.088 (0.023)	0.298 (0.020)	0.228 (0.053)
2017A	Mixed tank	1,050	0.074 (0.015)	0.209 (0.012)	0.261 (0.045)
2017B	Mixed tank	1,050	0.092 (0.018)	0.248 (0.014)	0.270 (0.045)
2020strain	Sep. tanks	587	0.065 (0.020)	0.210 (0.017)	0.235 (0.063)
2020main	Mixed tank	1,131	0.170 (0.033)	0.443 (0.026)	0.277 (0.046)
2020flow	Sep. tanks	120	-	-	-
2020flow	Mixed tanks	24	-	-	-
2020standard	Sep. tanks	179	-	-	-
2020standard	Mixed tanks	76	-	-	-

Results

Number of genotyped individuals, estimated variance components and heritabilities for the different groups are presented in Table 1. The estimated within-test genomic heritabilities were 0.16-0.17 for sea cage testing and 0.20-0.28 for tank testing. Starting from the same original sub-population of the 2011 year-class, the estimated differentiation in GEBVs between the LPB and HPB strains increased from 0 (2011), 1.2 (2014), 2.5 (2017) and finally 3.2 (2020) genetic standard deviations (Figure 1).

Under common garden testing, the estimated genetic change was accompanied by a large phenotypic differentiation (Figure 2), with the LPB having an observed lice count of -19% (P=1.0e-12) and -55% (P<2.2e-16) relative to HPB, for the year classes 2014 and 2017, respectively. For year-class 2020 (2020strain), the LPB and HPB strains were tested in separate tanks, and the phenotypic differentiation dropped considerably for both high and low infestation levels, with LPB averaging -7% relative to HPB (P=0.18). This happened despite genetic differentiation being at its most extreme (3.2 genetic standard deviations). Hence, under a standard challenge model, selection for reduced attraction did not provide significant group-level protection in this challenge model. This applied to both infestation levels.

For the two differentiated groups from the main nucleus (tests 2020standard and 2020flow), the estimated genetic difference was 1.7 genetic standard deviations (i.e. half the difference between the LPB and HPB selection strains). Under common-garden testing, the new LPB group had lower lice counts than HPB in both ordinary and flow-through challenge models (-22%, P=0.005 and -29%, P=0.11, respectively). However, under separate tank testing, this effect was absent for the ordinary challenge model (+1%, P=0.86), but a moderately favourable effect was estimated for the flow-through model (-14%, P=0.10). Experimental tank did not have any significant effect within any of the treatments.

Discussion

The results indicate that divergent selection on lice count has a large (within-tank) effect under common-garden testing. However, when tested in separate tanks under standard conditions, the strain effect was in the right direction but no longer significant. The standard challenge protocol may give lice copepodites ample opportunity to attach to available hosts, potentially masking relevant tank-level genetic effects. For this purpose, a flow-through challenge model was developed, resulting in substantial group differences under common-garden testing, and likely some group-protection effect under separate tank testing. In

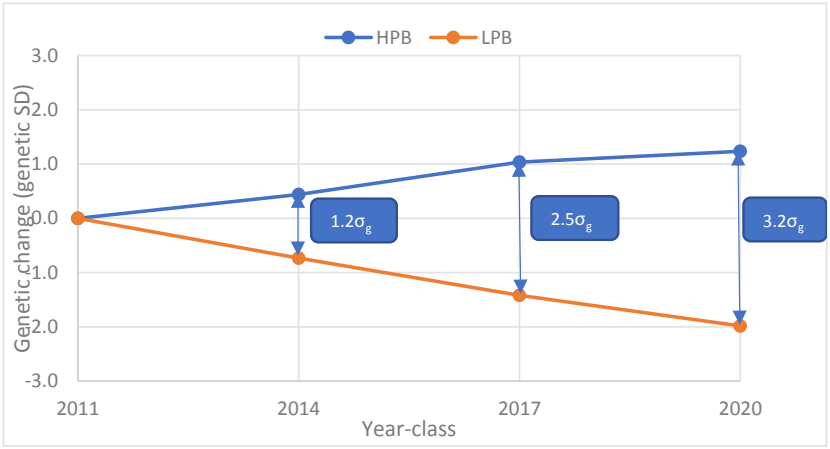


Figure 1. Estimated genetic trend for the LPB and HPB selection strains by generation (year-class).

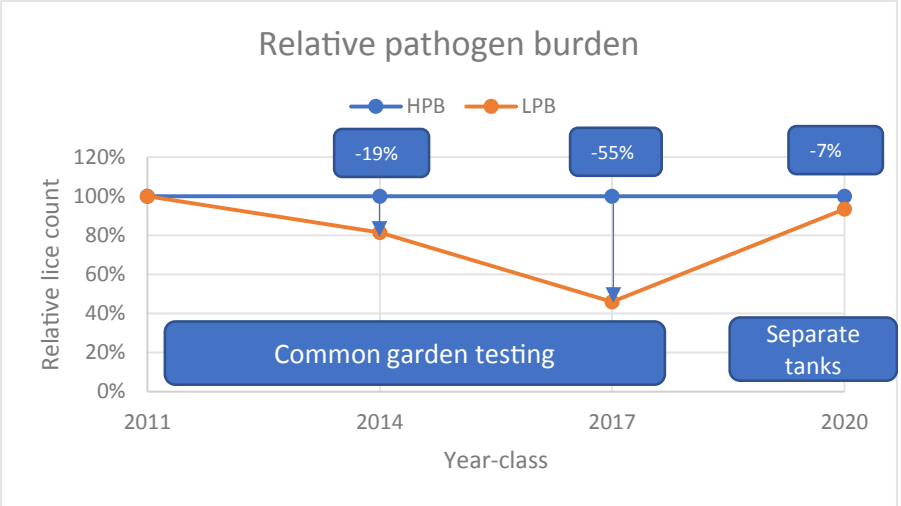


Figure 2. Relative observed phenotypic difference between the HPB and LPB selection strains by generation (year-class). Common garden testing was used for year-classes 2011-2017, while separate tanks were used for year-class 2020.

conclusion, genetic variation in salmon louse attraction had a large effect on within-tank distribution of lice, but, depending on conditions, smaller effects when strains are tested separately. Little is known about the group-level effects of parasite attraction under large-scale field conditions. More research is thus needed to what extent the effect is translated into commercial/field conditions.

References

Difford, G., Haugen, J.E., Aslam, M.L., Johansen, L.H., Breiland, M. *et al.* (2021) Res. Square. https://doi.org/10.21203/rs.3.rs-620617/v1

Gjerde, B., Ødegård, J. and Thorland, I. (2011) Aquaculture 314: 66-72. https://doi.org/10.1016/j.aquaculture.2011.01.026

Tsai, H.Y., Hamilton, A., Tinch, A.E., Guy, D.R, Bron, J.E. *et al.* (2016). Gen. Sel. Evol. 48: 47. https://doi.org/10.1186/s12711-016-0226-9

VanRaden, P. M. (2008) J. Dairy Sci. 91: 4414-4423. https://doi.org/10.3168/jds.2007-0980

573. A major QTL affects resistance to viral nervous necrosis in farmed European seabass

R. Mukiibi[1], D. Robledo[1*], C. Peñaloza[1], S. Ferraresso[2], R. Franch[2], D. Bertotto[2], M. Freguglia[3], S. Laureau[3,4], F. Pascoli[5], A. Toffan[5], C. Tsigenopolous[6], L. Bargelloni[2] and R. Houston[1*]

[1]The Roslin Institute and Royal (Dick) School of Veterinary Sciences, The University of Edinburgh, Edinburgh, United Kingdom; [2]Department of Comparative Biomedicine and Food Science, University of Padova, Viale dell'Universita' 16, Agripolis, 35020 Legnaro, Italy; [3]Valle Cà Zuliani Società Agricola s.r.l., Conselice (RA), Italy; [4]Acefas, 9 Boulevard du Roi 78000 Versailles, France; [5]OIE Reference Laboratory for Viral Encephalopathy and Retinopathy (VER), Istituto Zooprofilattico Sperimentale delle Venezie, Viale dell'Università 10, 35020 Legnaro (Padova), Italy; [6]Institute of Marine Biology, Biotechnology and Aquaculture (IMBBC), Hellenic Centre for Marine Research (H.C.M.R.) Crete, P.O. Box 2214, Gournes Pediados 715 00 Heraklion, Crete, Greece; diego.robledo@roslin.ed.ac.uk; ross.houston@roslin.ed.ac.uk

Abstract

Viral nervous necrosis (VNN) presents a significant threat to the European seabass aquaculture industry, causing extensive losses due to mortality and impaired growth of infected fish. Selective breeding presents an opportunity to develop seabass strains with increased resistance, thus reducing the impact of the disease on the industry. Knowledge of the genetic architecture of the trait is important to inform selective breeding strategies, including use of genomic tools. In the current study the genetic basis of VNN resistance in European seabass was investigated through an extensive study based on nervous necrosis virus (NNV) challenge of juvenile seabass fish. The results highlighted moderate heritability (up to 0.39±0.05) of VNN resistance in farmed seabass. A major QTL was identified on LG12 or Chromosome 13 and impacted VNN resistance when defined as binary survival or as days to death. This QTL explains up to 37% of the genetic variation in VNN resistance, and therefore has a significant value for marker-assisted and genomic selection to improve resistance in sea bass breeding programmes. Together, our results provide insight into the genetic architecture of resistance to VNN that could further be utilized for enhancing genomic selection for more resistant fish.

Introduction

European sea bass (*D. labrax*) is a fish species with major economic and cultural importance to the European aquaculture industry (Vandeputte *et al.* 2019). Viral nervous necrosis (VNN) disease causes significant losses to farmed sea bass production through mortalities and impaired growth of the infected fish (Barsøe *et al.* 2021). Despite the significant economic impact of this disease to Mediterranean aquaculture, there are limited vaccination, therapeutic, and biosecurity control measures. Genetics studies have revealed significant, moderate heritability of host resistance to VNN in different seabass populations (Palaiokostas *et al.* 2018; Faggion *et al.* 2021; Griot *et al.* 2021). Additionally, genomic regions or markers have showed significant association with VNN resistance in different European sea bass populations (Palaiokostas *et al.* 2018; Griot *et al.* 2021). However, information remains sparse on the genetic architecture of host resistance to VNN in different farmed populations. Furthermore, identification of genetic markers or regions involved in regulating VNN resistance enable marker-assisted selection or enhance the genomic selection for more resistant fish which is a sustainable route to reduce the negative economic and animal welfare impact of VNN. Therefore, in the current study, the aim was to characterise the genetic resistance to VNN in sea bass, and to identify specific genomic regions and markers associated with the trait.

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_573, © R. Mukiibi *et al.* 2022

Materials & methods

Viral nervous necrosis challenge experiment and genotyping. Approximately 1,500 European seabass from a full factorial cross of 25 dams and 25 sires were challenged with nervous necrosis virus at a temperature of 25 °C. These fish were produced by Valle Ca' Zuliani Società Agricola in April 2020 and were challenged at a weight ranging from 6 to 20 grams at the Istituto Zooprofilattico Sperimentale delle Venezie facilities under the authorization of the Italian Ministry of Health (authorisation code 975/2016-PR). Mortality was recorded twice a day for a period of 29 days as the primary phenotype for the genetic analyses, with days to death (a.k.a survival time) also used as a phenotype. The days to death phenotype was considered both in mortalities only, and in all animals where survivors were assigned a high value equivalent to the termination of the trial. A total of 1,066 fish (including the 50 parents and 1,016 NNV challenged fish) were genotyped on the Thermo Fisher Scientific's Axiom 'MedFish' SNP array which contains circa 30K seabass SNPs (Penaloza *et al.* 2021).

Pedigree reconstruction and pedigree-based heritability estimation. Since the family structure was not known at the point of challenge, genotype data were used to reconstruct the pedigree. Circa 23,000 informative SNP marker genotypes were used to uniquely assign a dam and sire to each of the offspring using the APIS package (Griot *et al.* 2020) implemented in R, allowing for a Mendelian error rate of 5%. Subsequently, survival analyses were performed for the offspring for dam and sire families using the Survminer package (Kassambara *et al.* 2017) implemented in R. The reconstructed pedigree was used to fit a linear mixed animal model to estimate the heritability of VNN resistance in this population of farmed European sea bass using the ASReml software (Gilmour *et al.* 2015).

Genotype data processing, genomic heritability estimation, and genome-wide association study (GWAS) analyses. Genotype data were filtered using Plink (Purcell *et al.* 2007) to exclude SNPs with minor allele frequencies <5%, SNPs with missing genotype rate >5%, and SNPs that strongly (P<1E-4) deviated from Hardy Weinberg equilibrium. Additionally, individuals with high genotype missing rate (>10%) were also excluded from the analysis. All of the animals that passed the QC filters were retained, and the final dataset contained genotype data for 24,740 informative SNP markers. GWAS was performed on 990 fish which had high quality phenotype and genotype data. Phenotype records (i.e. binary survival and days to death) were firstly adjusted to account for population substructure using the first two principal components that explained 77.5% of the variation between the animals. The heritability of resistance to VNN was estimated using the genomic relationships, and an equivalent mixed linear model to that described above. GWA analyses were used to assess the association between SNP marker genotypes and the corrected VNN resistance phenotype data using the GCTA software (Yang *et al.* 2011). GWAS results were then visualized using Manhattan plots using the qqman package implemented in R (Turner 2014).

Results

Parentage assignment analysis revealed that every parent (sires and dams) in the study population contributed a significant number of progeny with moderate degree of variability. Survival analyses of the families revealed marked differences in mortality levels within and between families for both dam and sire parents over the challenge study time. In general, moderate and significant heritability was found for the three VNN resistance traits (Table 1).

A total of 528, 578 and 42 markers were identified as significantly (FDR<0.05) associated with binary survival, days to death (all animals), and days to death (mortalities only), respectively. Additionally, 92, 104 and 23 markers, respectively, showed suggestive (FDR<0.1) association with the same traits. For all the three traits, most (78 to 95%) of the significant SNP markers were located on LG12 (also denoted chromosome 13, Figure 1a, b and c), indicating presence of a major QTL on this chromosome. For all the three traits,

Table 1. Heritability estimates of VNN resistance in farmed seabass.

VNN resistance definition	Pedigree based $h^2 \pm SE$	Genomic $h^2 \pm SE$
Binary survival	0.38±0.10	0.37±0.05
Days to death (mortalities and survivors)	0.23±0.08	0.39±0.05
Days to death (mortalities only)	0.06±0.10	0.21±0.08

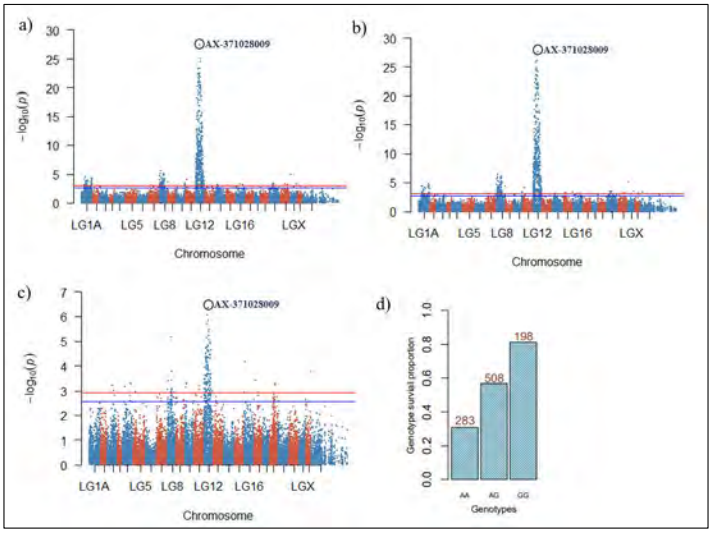


Figure 1. Manhattan plot summarizing genome-wide association study results for VNN resistance in European seabass defined as; (a) binary survival, (b) days to death for both mortalities and surviving fish, and (c) days to death with mortalities; and (d) bar plot showing survival proportion of individuals from the three genotype classes (AA, AG and GG) for the most significant SNP marker (AX-371028009), the numbers individuals with each of the three genotypes are provided above each bar in the plot.

AX-371028009 was the most significant SNP marker associated with each of the traits and explained 12.7, 14.0 and 6.3% of the phenotypic variation in VNN resistance when defined as binary survival, days to death (all animals), or days to death (mortalities only), respectively. This equates to 37.0, 36.0 and 29.8% of the additive genetic variance of the same traits, respectively. Indeed this SNP demonstrated additive genetic effect on VNN resistance in seabass with the homozygous genotype GG associated with the highest (81% survival rate) level of resistance, while the AA genotype associated with the lowest (30.7% survival rate) resistance (Figure 1d). This SNP is located within the 26th intron of the EML5 (EMAP Like 5) gene. EML5 encodes for the Echinoderm microtubule associated protein like 5, which is involved in regulating cytoskeleton rearrangement (O'Connor *et al.* 2004).

Discussion

The moderate heritability estimates of VNN resistance observed in the current study concur with the moderate to high heritability estimates of VNN resistance that have been reported in other farmed European seabass populations (Palaiokostas *et al.* 2018; Faggion *et al.* 2021; Griot *et al.* 2021). Together, these results highlight the potential for selective breeding as a viable as a sustainable approach to improve resistance and therefore reduce economic losses due to disease outbreaks on seabass farms. The GWAS

results presented herein are also in agreement with the recent discovery of a major QTL for VNN resistance on LG12 in different European seabass subpopulations (Griot *et al.* 2021) where the QTL explained up to 9.2% of the total genetic variability. Palaiokostas *et al.* (2018) did not identify such a major QTL on LG12, however they identified a SNP marker on LG12 that was significantly associated to VNN resistance and accounted for 4% of the total additive genetic variance (Palaiokostas *et al.* 2018). Therefore, the current study confirms the presence of a major VNN resistance locus on LG12 that can be utilized for breeding of more resistant seabass populations. Further analyses are underway using whole genome resequencing data for all animals, combined with gene expression comparison of resistant and susceptible animals, with a view to identifying the causative gene and variant for this major QTL.

Funding

The study was conducted as part of the AQUA-FAANG project, which has received funding from the European Union's Horizon 2020 research and innovation program under Grant Agreement No 817923.

References

Barsøe S., Allal F., Vergnet A., Vandeputte M., Olesen N.J., Schmidt J.G., Larsen C.A., Cuenca A. & Vendramin N. (2021) Different survival of three populations of European sea bass (Dicentrarchus labrax) following challenge with two variants of nervous necrosis virus (NNV). *Aquaculture Reports* 19, 100621.

Faggion S., Bertotto D., Babbucci M., Dalla Rovere G., Franch R., Bovolenta M., Laureau S., Pascoli F., Toffan A. & Bargelloni L. (2021) Resistance to viral nervous necrosis in European sea bass (Dicentrarchus labrax L.): heritability and relationships with body weight, cortisol concentration, and antibody titer. *Genetics Selection Evolution* 53, 1-12.

Gilmour A., Gogel B., Cullis B., Welham S. & Thompson R. (2015) ASReml user guide release 4.1 structural specification. *Hemel hempstead: VSN international ltd*.

Griot R., Allal F., Brard-Fudulea S., Morvezen R., Haffray P., Phocas F. & Vandeputte M. (2020) APIS: An auto-adaptive parentage inference software that tolerates missing parents. *Molecular ecology resources* 20, 579-90.

Griot R., Allal F., Phocas F., Brard-Fudulea S., Morvezen R., Bestin A., Haffray P., François Y., Morin T. & Poncet C. (2021) Genome-wide association studies for resistance to viral nervous necrosis in three populations of European sea bass (Dicentrarchus labrax) using a novel 57k SNP array DlabChip. *Aquaculture* 530, 735930.

Kassambara A., Kosinski M., Biecek P. & Fabian S. (2017) Package 'survminer'. *Drawing Survival Curves using 'ggplot2'(R package version 03 1)*.

Palaiokostas C., Cariou S., Bestin A., Bruant J.-S., Haffray P., Morin T., Cabon J., Allal F., Vandeputte M. & Houston R.D. (2018) Genome-wide association and genomic prediction of resistance to viral nervous necrosis in European sea bass (Dicentrarchus labrax) using RAD sequencing. *Genetics Selection Evolution* 50, 1-11.

Penaloza C., Manousaki T., Franch R., Tsakogiannis A., Sonesson A.K., Aslam M., Allal F., Bargelloni L., Houston R. & Tsigenopoulos C. (2021) Development and testing of a combined species SNP array for the European seabass (Dicentrarchus labrax) and gilthead seabream (Sparus aurata). *Genomics* 113, 2096-107.

Purcell S., Neale B., Todd-Brown K., Thomas L., Ferreira M.A., Bender D., Maller J., Sklar P., De Bakker P.I. & Daly M.J.J.T.A.j.o.h.g. (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. 81, 559-75.

Turner S.D. (2014) qqman: an R package for visualizing GWAS results using QQ and manhattan plots. *Biorxiv*, 005165.

Vandeputte M., Gagnaire P.A. & Allal F. (2019) The European sea bass: a key marine fish model in the wild and in aquaculture. *Animal genetics* 50, 195-206.

Yang J., Lee S.H., Goddard M.E. & Visscher P.M. (2011) GCTA: a tool for genome-wide complex trait analysis. *The American Journal of Human Genetics* 88, 76-82.

574. The potential indicator for resilience of Sukamandi strain tilapia grown in brackish water ponds

M.H.F. Aththar[1,2*], P. Setyawan[1,3], I. Imron[3], B. Gunadi[3], M. Camara[1], H. Komen[1] and J.W.M. Bastiaansen[1]

[1]Animal Breeding and Genomics, Wageningen University and Research, Droevendaalsesteeg 1, 6708 PB Wageningen, the Netherlands; [2]Research Institute for Freshwater Aquaculture and Fisheries Extension, Ministry of Marine Affairs and Fisheries Indonesia; [3]Research Institute for Fish Breeding, Ministry of Marine Affairs and Fisheries Indonesia; farid.aththar@wur.nl

Abstract

Resilience is the ability of an animal to adapt with environmental changes while maintaining consistent performance over time. A selective breeding program for growth in brackish water of Sukamandi strain tilapia is conducted in Indonesia. The resilience of fish is also important for producing tilapia in variable saline conditions. The objectives of this study were to estimate the genetic parameters of resilience for Sukamandi strain, expressed as consistency in growth. We used the log-transformed variance of deviations (LnVar) from weight, recorded five times during grow-out. We also estimate the correlation of LnVar with weight (W), daily growth coefficient (DGC), condition factor (K) and survival (S). The heritability of LnVar in time point 4 and 5 were 0.02 and 0.07, respectively. The genetic correlations were positive (0.35-0.76) for LnVar with W, DGC, S and negative (-0.31) with K. These estimates showed that selection for growth and survival will moderately increase LnVar and may lead to the decreasing resilience. However, selection for K could contribute to increased resiliency of tilapia.

Introduction

Resilience is the ability of the animal to cope with environmental changes, such as fluctuations in salinity (Berghof *et al.* 2019a, Mulder *et al.* 2017). Resilience in fish production refers to the consistency of production performance throughout time. A selective breeding program for growth rate of tilapia in brackish water is conducted in Indonesia. Selection is applied to the Sukamandi strain population, an inter-species hybrid created by mixing blue tilapia (*Oreochromis aureus*) and Nile tilapia (*Oreochromis niloticus*) (Yu *et al.* 2021). While focusing on growth, the resilience of fish is also considered important. To produce tilapia in conditions of varying salinity, a high resilience of tilapia is needed. In this paper, we estimate the genetic parameters for resilience of Sukamandi strain. The genetic parameter for resilience can be estimated based on longitudinal measurement by calculating deviations of observed from expected production (Friggens *et al.* 2017). Several potential resilience indicators have been investigated, including natural-log-transformed variance of deviations from mean (LnVar), skewness of deviations, and autocorrelation of deviations (Berghof *et al.* 2019a). LnVar was the most promising resilience trait based on its moderate heritability (Poppe *et al.* 2020; Berghof *et al.* 2019a). Furthermore, Berghof *et al.* (2019a) showed that including resilience trait in index selection would add economic benefit. Hence, correlated response of potential indicator for resilience with growth, factor condition and survival is important. In this study, we also estimate the correlation of LnVar with weight (W), daily growth coefficient (DGC), condition factor (K) and survival (S).

Materials & methods

Family production and testing environments. All families were produced using combination of one male mated to three females. After 7 days of mating, we collected the fertilized eggs or larva from a female mouth, and then incubated eggs in cone hatching jars until they became functional larva. In total 102 families were produced over a period of 28 days. We transferred 200 swim-up fry from each family for

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_574, © M.H.F. Aththar *et al.* 2022

nursery period. After 3-4 weeks, we randomly tagged and weighed 40 individuals in each family using PIT (Passive Integrated Transponder) tags. From each family, 20 fingerlings were stocked for grow-out in monoculture and 20 in co-culture with shrimp. The test location for grow out of tilapia was at Technical Implementation Unit for Brackish water Culture Karawang, at salinity around 20 ppt. We stocked all families in the happa within two ponds with a stocking density of 5.1 fish per m^2. During stocking we reduced the salinity level from around 20 ppt to 10 ppt to reduce stress and mortality.

Trait measurements. We measured weight and length at 5 time points: stocking time (t_1), 3 time points after stocking time ($t_{2\text{-}4}$) and harvest time (t_5). We calculated the standardized deviations of body weight in each time point: (Wt_n - $\overline{W}t_n$) / SD_n. For each fish, we calculated the variance of these standardized deviations and log-transformed this variance using the natural logarithm to obtain LnVar. DGC, W, and S were evaluated at time point 4 and 5. DGC was calculated as:

$$DGC = \left(\frac{W_t^{\frac{1}{3}} - SW^{\frac{1}{3}}}{growing\ days}\right) x\ 100,$$

where SW is body weight at t_1, W_t is weight at t_n and growing days is the growing time between t_1 and t_n. We calculated K at t_4 with W in grams and L in mm:

$$K = \left(\frac{W_4}{{L_4}^3}\right) \times 10^5.$$

Data analysis – phenotypic and genetic parameters. Estimation of genetic parameters was carried out for W, LnVar, DGC and S at t_4 and t_5. Estimation of genetic parameters for K was conducted only at t_4 due to lacking L information at t_5. Fixed effects of happa and sex were fitted in the animal models for all traits. Genetic parameter were estimated by fitting a univariate animal model using ASReml version 4.1 (Gilmour *et al.* 2015):

$$y_{ijklm} = \mu + Happa_i + Sex_j + Init_k + Family_l + a_m + e_{ijklm}$$

where: y_{ijklm} is vector of trait; μ is overall mean; $Happa_i$ is fixed effect of happa; Sex_j is fixed effect of sex (l =m, f); $Init_k$ is initial weight for DGC and initial weight and length for K; $Family_l$ is full sib family as random common environmental effect; a_m is random additive genetic effect of the m-th individual; e_{ijklm} is random residual effect associated with an individual. The model was used to analyse the traits DGC and K. The same model was used for direct measurement trait of W, LnVar and S but excluded the initial value. We calculated the heritability as the ratio between additive genetic variance (σ_A^2) and phenotypic variance (σ_P^2), $\frac{\sigma_A^2}{\sigma_P^2}$.

Genetic and phenotypic correlations between different traits were obtained from bivariate analyses. The animal effects were distributed as N(0,A⊗G) with the additive genetic variance covariance matrix:

$$(G) = \begin{bmatrix} \sigma_{A,1}^2 & r_{A,12}\sigma_{A,1}\,\sigma_{A,2} \\ r_{A,12}\sigma_{A,1}\,\sigma_{A,2} & \sigma_{A,2}^2 \end{bmatrix},$$

where $\sigma_{A,1}^2$ ($\sigma_{A,2}^2$) is the additive genetic variance of trait 1 (trait 2), and $r_{A,12}\sigma_{A,1}\,\sigma_{A,2}$ is the additive genetic covariance between trait 1 and trait 2. The residuals were distributed as N(0, I⊗R) with residual variance-covariance matrix (R) is:

$$\begin{bmatrix} \sigma_{e,1}^2 & r_{e,12}\sigma_{e,1}\,\sigma_{e,2} \\ r_{A,12}\sigma_{A,1}\,\sigma_{A,2} & \sigma_{e,2}^2 \end{bmatrix},$$

where $\sigma^2_{e,1}$ ($\sigma^2_{e,2}$) is th e residual variance of trait 1 (trait 2), and $r_{e,12}\sigma_{e,1}\,\sigma_{e,2}$ is the residual covariance between trait 1 and trait 2. Genetic and phenotypic correlations among traits were calculated as the covariance divided by the product of the standard deviations of traits.

Results

The average stocking weight was similar between male (11.8±3.2 g) and female (10.1±2.8 g). Overall, the mean for all measured parameter were higher in male compared with the female. The survival rate at t_4 is 75.5% and 27.3% at t_5(Table 1).

Phenotypic and genetic parameters. We found low heritability for LnVar of 0.02±0.04 in t_4 and 0.07±0.07 in t_5. The heritability were low to moderate for W, DGC, K and S in t_4 (0.04-0.13) and t_5 (0.06-0.29) (Table 2). Common environmental effects due to family were decreasing and heritability were increasing for all estimated trait between t_4 and t_5.

The genetic correlations of LnVar with W, DGC and S were positive and moderate (0.35-0.76), while the correlations were negative between LnVar and K (-0.31±0.23). The genetic correlations between W, DGC and K were high (0.87-0.99). Furthermore, the genetic correlation of survival with W, DGC and K were negative and moderate between -0.26 and -0.76 (Table 3).

Table 1. Descriptive statistics of weight (W), variance of deviation (Var-dev), log transformed variance (LnVar), daily growth coefficient (DGC), condition factor (K) and survival (S) of Sukamandi strain tilapia for data point 4 (t_4) and 5 (t_5) in brackish water ponds.

Traits	t_4			t_5		
	N	Male	Female	n	Male	Female
W	3,056	148.3±29.7	120.2±27.6	1,105	219.4±44.9	180±42.1
Var-dev	2,926	0.5±0.6	0.3±0.4	1,030	0.6±0.6	0.4±0.4
LnVar	2,926	-1.1±1.1	-1.7±1.2	1,030	-0.9±1.0	-1.5±1.0
DGC	3,015	3.1±0.4	2.8±0.4	1,091	2.6±0.3	2.4±0.3
K	3,015	3.9±0.3	3.9±0.3	-	-	-
S (%)	3,056	(75.5%)		(27.3%)		

Table 2. Phenotypic variances (σ^2_P), heritability (h^2) and common environmental effect (c^2) of weight (W), log transformed variance (LnVar), daily growth coefficient (DGC), condition factor (K) and survival (S) of Sukamandi strain tilapia for data point 4 (t_4) and 5 ($t_{4)}$) in brackish water ponds.

Traits	t_4			t_5		
	σ^2_P	h^2 *(se)*	c^2	σ^2_P	h^2 *(se)*	c^2
W	706.3	0.12 (0.06)	0.07 (0.03)	1,738.7	0.29 (0.07)	0.00 (0.00)
LnVar	1.2	0.02 (0.04)	0.06 (0.02)	1.0	0.07 (0.07)	0.05 (0.03)
DGC	0.2	0.13 (0.07)	0.08 (0.03)	0.8×10^{-1}	0.28 (0.10)	0.01 (0.03)
K	0.9×10^{-1}	0.28 (0.09)	0.08 (0.03)	-	-	-
S	0.2	0.04 (0.03)	0.05 (0.02)	0.1	0.06 (0.02)	0.00 (0.01)

Table 3. Estimated genetic correlations (above diagonal) and phenotypic correlations (below diagonal) between resilience trait (LnVar) with weight (W), daily growth coefficient (DGC), condition factor (K) and survival (S) of Sukamandi strain tilapia for data point 4 (t_4) in brackish water ponds. Standard error are in brackets.

	LnVar	W	DGC	K	S
LnVar		0.76 (0.18)	0.48 (0.37)	-0.31 (0.23)	0.35 (0.50)
W	0.10 (0.02)		0.99 (0.00)	0.91 (0.06)	-0.67 (0.38)
DGC	0.03 (0.02)	0.99 (0.00)		0.87 (0.07)	-0.76 (0.33)
K	-0.02 (0.02)	0.40 (0.02)	0.40 (0.02)		-0.26 (0.28)
S	0.01 (0.02)	-0.01 (0.02)	-0.03 (0.03)	-0.04 (0.03)	

Discussion

We reported low heritabilities for LnVar. This estimates was lower than previous estimates for heritability of LnVar such as LnVar heritability for chicken body weight of 0.10 (Berghof *et al.* 2019b). Common environmental effects decreased while heritability increased for all estimated trait between t_4 and t_5. We only estimated the genetic correlations in t_4 because the dataset became too small at t_5 due to mortality between t_4 and t_5 (Table 1). Genetic correlations between LnVar and W, DGC and S were positive and moderate (0.35-0.76). These positive correlation would mean that selection for growth and survival will moderately increase LnVar or decrease the resilience. In other words, a positive correlation is unfavorable for simultaneous improvement between resilience, growth and survival, whereas a negative correlation is beneficial. However, the negative and moderate genetic correlation estimated between LnVar and K (-0.31) showed that selection for K could contribute to greater resiliency to the salinity. Common environmental effects were strong up to t_4 and may have affected the estimates for heritability. Increased heritability of resilience was estimated at t_5, but based on a reduced dataset due to mortality.

Acknowledgements

The authors acknowledge the financial support of the Koepon Foundation and Indonesia Endowment Fund for Education, Ministry of Finance Indonesia (LPDP – RI).

References

Berghof T.V.L., Poppe M., and Mulder H.A., (2019a). Front. Genet. 9:692. https://doi.org/10.3389/fgene.2018.00692

Berghof T.V.L, Bovenhuis H., and Mulder H.A., (2019b). Front. Genet. 10:1216. https://doi.org/10.3389/fgene.2019.01216

Friggens, N. C., Blanc, F., Berry, D. P., Puillet, L., (2017). Animal 11:12. https://doi.org/10.1017/S175173111700088X

Gilmour A.R., Gogel G.J., Cullis B.R., Welham S.J., and Thompson, R., (2015). ASReml User Guide Release 4.1 Functional Specification.

Mulder H. A., and Rashidi H. (2017). J. Anim. Sci. 95: 3346–3358. https://doi.org/10.2527/jas2017.1479

Poppe M., Veerkamp R. F., van Pelt M. L., and Mulder H. A. (2020). J. Dairy Sci. 103(2): 1667-1684. doi.org/10.3168/jds.2019-17290

Yu X., Setyawan P., Bastiaansen J.W.M., Liu L., Imron I., Groenen M.A.M., Komen H., Megens, H.J. (2021). Aquaculture (*under review).*

575. Exploiting genomics to infer population structure and genetic variability in wild and farmed seabream and seabass

B. Villanueva[1*], A. Fernández[1], R. Peiro-Pastor[1], C. Peñaloza[2], R.D. Houston[2], A.K. Sonesson[3], C.S. Tsigenopoulos[4], L. Bargelloni[5], K. Gamsiz[6], B. Karahan[6], E.Ö. Gökçek[6], J. Fernández[1] and M. Saura[1]

[1]INIA-CSIC, Crta. La Coruña km 7.5, 28040 Madrid, Spain; [2]The Roslin Institute and Royal (Dick) School of Veterinary Studies, University of Edinburgh, EH25 9RG Midlothian, United Kingdom; [3]NOFIMA, Osloveien 1, 1430 Ås, Norway; [4]IMBBC-HCMR, Thalassocosmos Gournes Pediados, 71500 Heraklion, Greece; [5]Universitá degli Studi di Padova, Via 8 Febbraio 1848, 2, 35122 Padova PD, Italy; [6]Faculty of Fisheries, University of Ege, 35100 Izmir, Turkey; villanueva.beatriz@inia.es

Abstract

Knowledge of population structure and genetic variability within and between wild and farmed populations is important to achieve sustainable aquaculture production. In this study, these features were assessed for gilthead seabream and European seabass using a combined species SNP array newly developed. Population structure analyses revealed a clear differentiation between wild and farmed populations in both species, indicating that care must be taken to avoid escapes that could have undesirable genetic effects on native populations. Wild populations showed a low degree of differentiation, particularly in seabream. However, farmed populations were quite heterogeneous and showed a high degree of differentiation. In general, the effective population size was large ($>10^3$) for wild and small ($<10^2$) for farmed populations of both species. The low effective population size estimated for some farmed populations highlights the need of applying measures to control the rate at which genetic variability is lost.

Introduction

Knowledge of population structure and genetic diversity within and between wild and farmed fish populations is of paramount importance to develop optimal strategies for the conservation of native fish and to achieve sustainable aquaculture production. The success of any breeding programme is critically dependent on the way in which the base population of breeders is built. Given that base populations can be established from wild and farmed populations, it is necessary to determine the relatedness between them in order to optimise their creation (Fernández *et al.* 2014). Also, once genetic improvement programmes are operational, the control of inbreeding (or equivalently, the effective population size, N_e, a parameter inversely related to the loss of genetic variability) is needed to ensure their long-term sustainability.

Gilthead seabream (*Sparus aurata*) and European seabass (*Dicentrarchus labrax*) are two of the main cultured finfish species in Europe. Although still less developed than in salmonids, selective breeding is expected to play an increasingly important role in the commercial production of both species and different companies have initiated breeding programmes in recent years (Janssen *et al.* 2017).

Most studies investigating population structure in gilthead seabream and European seabass have been based on a limited number of genetic markers (mostly microsatellites). However, new genomic resources are now available, including a recently developed combined ~60K SNP array for both species (the MedFish array; Peñaloza *et al.* 2020).

The objective of this study was to make use of the MedFish SNP array to: (1) identify the extent of wild and farmed population stratification within species, considering the geographic origin of individuals; and (2) estimate N_e in populations of both species.

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_575, © B. Villanueva *et al.* 2022

Materials & methods

Samples available came from 11 wild (W-) and 12 farmed (F-) seabream populations and from 9 wild and 15 farmed seabass populations, distributed from the Western to the Eastern Mediterranean Sea. Additionally, three wild seabream populations sampled in the Atlantic Ocean (W-SPA-1, W-SPA-2, W-FRA) were available for analyses. Populations were ordered according to their geographic location (Figure 1).

The total number of individuals genotyped with the SNP array was 462 and 516 for seabream and seabass, respectively. After quality control, the total number of SNPs available for analysis was 25,319 for seabream and 22,507 for seabass.

To assess population structure, a model-based clustering method was implemented using genotypic data and the software Admixture version 1.3.0 (Alexander *et al.* 2009). The analysis is based on maximum likelihood estimation of individual ancestries from SNP genotypes, providing the best proportions of admixing components (clusters) for any hypothetical number (K) of genetic groups (Liu *et al.* 2013). In order to identify the value of K for which the model has the best predictive accuracy, the Admixture software uses a cross-validation procedure, and the K value chosen is that with the lowest standard error for the cross-validation. Wild and farmed populations within each species were analysed jointly.

Estimates of current effective population size (N_e) were obtained for all populations from linkage disequilibrium (measured as r^2) between independent SNPs mapped in different chromosomes or linkage groups, following Waples (2006).

Results

The admixture model results revealed that the number of ancestral populations (K) with the lowest cross-validation error was 7 for seabream (Figures 1a, 1b) and 15 for seabass (Figures 1c, 1d). The results showed a clear differentiation between wild and farmed populations.

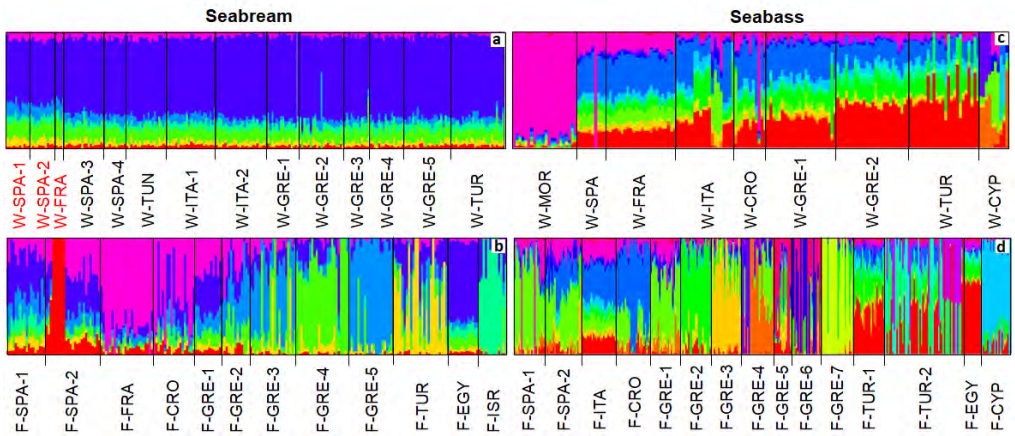


Figure 1. Admixture plots for wild (a, c) and farmed (b, d) seabream (a, b) and seabass (c, d) populations. Ancestry proportions are represented on the y-axis. Population codes (ordered according to their geographical location) are represented on the x-axis (MORocco, SPAin, FRAnce, TUNisia, ITAly, CROacia, GREece, TURkey, CYPrus, EGYpt, ISRael). Codes for populations sampled in the Atlantic Ocean are in red (a).

In general, there was very little differentiation among wild seabream populations although a slight differentiation was observed between Atlantic and Mediterranean populations (Figure 1a). Contrarily, farmed seabream populations were heterogeneous and showed different admixture proportions. It is worth to note that the F-EGY population showed admixture proportions similar to those found in wild populations (Figure 1b). It is also relevant the observation that some farmed populations (e.g. F-SPA-2) presented two clearly different groups, probably reflecting a very recent mixture of populations (Figure 1b).

Within wild seabass populations, the Moroccan population (W-MOR) was very different from the rest (Figure 1c). Also, the Cyprus population (W-CYP) had a different genetic makeup. For the rest of the wild seabass populations, a gradual change was observed in the proportions of different origins and this change was associated with the geographic locations as we move from West to East. It was also observed the presence of individuals within a particular population that clearly differed from others in the same population (e.g. in populations W-SPA and W-CRO, Figure 1c). As with seabream, farmed populations were very heterogeneous and there was no direct relationship between the genetic structure and the location of the populations. Also for seabass, some farmed populations (F-ITA, F-TUR-1 and F-EGY) showed admixture proportions similar to those found in wild populations (Figure 1d).

In general, estimates of N_e were very high for wild populations ($>10^3$) and relatively low ($<10^2$) for farmed populations of both species, but there were some exceptions (Table 1). Five farmed seabream populations and 12 farmed seabass populations had a N_e lower than 50 fish.

Discussion

This study provides new high-resolution population genetic data to inform future breeding and stock management of two key European aquaculture species. Given the clear differentiation between wild and farmed populations for both species, precautions must be taken to avoid escapes that could have undesirable genetic effects in native populations. In fact, the admixture results showed that, within a particular population, some wild individuals have a genetic composition very different from the rest of individuals in the population and they may represent escapees from local farms. These results were generally supported by Principal Component Analysis and hierarchical distance-based clustering methods (results not shown).

In general, the degree of differentiation among wild populations was low, particularly in seabream, indicating that considerable gene flow exists. Our results are in line with those obtained in previous studies that used a much more limited number of markers (e.g. Alarcón *et al.* 2004), although some studies have described differentiated clusters for Mediterranean wild seabass (e.g. Souche *et al.* 2015) rather than the gradual change in the genetic composition from West to East observed here.

Table 1. Estimates of N_e for wild (W) and farmed (F) populations of seabream and seabass.

N_e		Seabream		Seabass
$>10^3$	W:	SPA-1,3, FRA, ITA-1,2, GRE-1,3-5, TUR	W:	MOR, FRA, GRE-1,2
			F:	ITA, EGY
$<10^2$	W:	SPA-2, TUN*, GRE-2	W:	SPA, ITA*, CRO, TUR, CYP*
	F:	SPA-1,2*, GRE-1,2,3,(4-5)*, TUR*, ISR*	F:	SPA-1*,2*, CRO*, GRE-(1-7)*, TUR-1,2*, CYP*

*Populations with $N_e<50$.

On the contrary, farmed populations of both species were quite heterogeneous, which may indicate that they were generated from several origins. They also showed a high degree of differentiation, pointing to different fish sources for different companies. A few farmed populations showed genetic compositions similar to those found in wild populations, which may indicate that their selective breeding programmes are still in their beginnings or that broodstock has been recently renewed using wild fish as a means of mitigating inbreeding.

Given that there was no clear differentiation among wild populations, it is likely that the low N_e found for some wild populations can be due to escapees included in the samples or to a reduced number of families included in the samples. On the other hand, the unexpected high N_e estimated for some farmed populations agrees with the results of the Admixture analysis in which these populations showed admixture proportions very similar to those found in wild populations that may indicate that their selective breeding programmes are recent.

As in many other fish populations of different species (e.g. Saura *et al.* 2021), estimates of N_e below 50 fish were found in many farmed populations probably due to sampling a limited number of families when building them. These estimates are lower than the critical size of 50 needed to avoid inbreeding depression and retain fitness in the short-term (FAO, 2013) and highlights the need of applying selection and mating approaches designed to control inbreeding and loss of genetic variability.

Acknowledgements

This work was funded by the European Commission Horizon 2020 (H2020) Framework Programme through grant agreement no 727315 MedAID project and by MCIN/ AEI /10.13039/501100011033 (Project PID2020-114426GB-C2).

References

Alarcón J.A., Magoulas A., Georgakopoulos T., Zouros E., and Alvarez M.C. (2004). Aquaculture 230(1):65–80. https://doi.org/10.1016/S0044-8486(03)00434-4

Alexander D.H., Novembre J., and Lange K. (2009). Genome Res. 19(9):1655–1664. https://doi.org/10.1101/gr.094052.109

FAO. (2013). *In vivo* conservation of animal genetic resources. FAO Animal Production and Health Guidelines. No. 14. Rome. https://www.fao.org/3/i3327e/i3327e.pdf

Fernández J., Toro M.A., Sonesson A.K., and Villanueva B. (2014). Front. Genet. 5:414. https://doi.org/10.3389/fgene.2014.00414

Janssen K., Chavanne H., Berentsen P., and Komen H. (2017). Aquaculture 472(1):8–16. https://doi.org/10.1016/j.aquaculture.2016.03.012

Liu Y., Nyunoya T., Leng, S., Belinsky S.A., Tesfaigzi Y. *et al.* (2013). Hum. Genomics 7(1):1. https://doi.org/10.1186/1479-7364-7-1

Peñaloza C., Manousaki T., Franch R., Tsakogiannis A., Sonesson A. *et al.* (2021). Genomics 113(4):2096–2107. https://doi.org/10.1016/j.ygeno.2021.04.038

Saura M., Caballero A., Santiago E., Fernández A., Morales-González E. *et al.* (2021). Genet. Sel. Evol. 53:85. https://doi.org/10.1186/s12711-021-00680-9

Souche E., Hellemans B., Babbucci M., Mac Aoidh E., Guinand B. *et al.* (2015). Biol. J. Linn. Soc. 116(1):86–105. https://doi.org/10.1111/bij.12572

Waples R. (2006). Conserv. Genet. 7(2):167–184. https://doi.org/10.1007/s10592-005-9100-y

576. Comparison of tank and commercial sea-pen family evaluation of Chinook salmon (*Oncorhynchus tshawytscha*) in New Zealand

M.R. Scholtens[1*], K.G. Dodds[2], S.M. Clarke[2], S.P. Walker[1], M.L. Tate[3], R. Miller[4], T. Slattery[4] and J.E. Symonds[1]

[1]Aquaculture Group, Cawthron Institute, Private Bag 2, Nelson 7042, New Zealand; [2]AgResearch, Invermay Agricultural Centre, Puddle Alley, Mosgiel 9053, New Zealand; [3]Natural Food Innovation, 607 Upper Sefton Road, P.O. Box 21, Sefton 7445, New Zealand; [4]Sanford Ltd, 19 Riverside Road Kaitangata 9281, New Zealand; megan.scholtens@cawthron.org.nz

Abstract

The aim of this study was to estimate between environment genetic correlations of thermal growth coefficient (TGC), condition factor (CF) and fillet fat (FF) of Chinook (king) salmon (*Oncorhynchus tshawytscha*) reared in tanks and commercial sea-pens (SP). 1,508 individuals from 44 pedigree families were tagged and reared for 12 months in seawater tanks up to ~1.50 kg. From the same families, 1,288 individuals were reared in SPs and measured at 2yrs old. All 2,796 individuals were genotyped using genotype-by-sequencing to generate a genomic relationship matrix. Estimates of heritabilities and genetic correlations were obtained using a bivariate animal model fitting contemporary group as fixed effect. Estimates of heritabilities were slightly lower in the SP traits indicating more environmental variation. The genetic correlations between tanks and SPs for TGC, CF and FF ranged from 0.46-0.78, suggesting that the Cawthron Finfish Research Centre tank evaluation can be used to predict family performance in this population.

Introduction

Chinook (king) salmon (*Oncorhynchus tshawytscha*) is an important aquaculture species in New Zealand (NZ) and selective breeding since the 1990s has resulted in significant gains in growth traits, flesh quality traits and fat content (Symonds *et al.*, 2019). With adequate genetic improvement in growth traits, breeders are considering incorporation of other economically important traits into the breeding programs such as environmental tolerance and feed conversion efficiency.

The Cawthron's Finfish Research Centre (FRC) facilities provide the opportunity for commercial salmon breeding programs to assess genotypes in trials that cannot be conducted in sea-pens. Phenotypes recorded in these facilities can be applied to pedigree-based or genomic evaluation of broodstock candidates. This provides the opportunity to evaluate traits that are extremely difficult to measure in the sea-pens such as feed efficiency and heat and hypoxia tolerance (Lulijwa *et al.*, 2021).

This study aimed to provide information about the genetic correlations among three commercially valuable traits, evaluated in related families reared in a commercial sea pen and a land-based tank system.

Materials & methods

Sanford family evaluation. Animals used in this study originated from the 2018-year class of the Sanford Chinook Salmon breeding program where selection is operated using non-overlapping generations of 3 years each. Assessment records of 1,288 animals from the 2018 cohort, were provided by Sanford. These animals were raised in SPs at Steward Island and sampled and measured during harvest on the 24 and 25th of June 2020. These animals originated from 126 crosses between 47 sires and 70 dams. Fillet fat (FF) data were measured using AgResearch's Hyperspectral method (Reis *et al.*, 2021).

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_576, © M.R. Scholtens *et al.* 2022

Cawthron FRC evaluation and trial design. The 1,508 animals in the FRC tank evaluation were part of a larger trial evaluating family growth, feed efficiency and thermal tolerance (Lulijwa *et al.*, 2021). These salmon were sourced from the same 2018-year class families as the commercial breeding program families in the sea-pens. The 1,508 juveniles originated from 44 crosses between 33 sires and 36 dams. Fillet fat (FF) was measured using Fourier transform-near infrared reflectance (NIR) spectroscopy (Miller *et al.*, 2018).

Genotype data. All fish in this study were genotyped using Genotyping-by-Sequencing (GBS) following the methods described in Dodds *et al.* (2015), with the following modifications: Samples were digested with MspI as well as PstI and the fragments were size selected to the range 193-318bp (Pippin Prep) and sequenced with 384 samples (including controls) per lane of a HiSeq2500. Prior to this study, adapter-trimmed sequence reads from a set of 3,279 Chinook salmon from various captive stocks in NZ were used to detect 42,839 single nucleotide polymorphisms (SNPs) using UNEAK (Lu *et al.* 2013) and without using a reference genome. These SNPs were placed into a catalogue to allow counts of reference and alternate alleles for each variant using TagDigger (Clark and Sacks, 2016) on any subsequently sequenced sample.

Initially, 473 SNPs that were monomorphic in the study fish and 9 fish with mean read depth <0.3 were removed. The SNPs were further filtered to remove any with minor allele frequency <0.01, Hardy-Weinberg (HW) disequilibrium <-0.05 or with a depth-adjusted HW test (Dodds *et al.* 2018) *P*-value<10^{-100}. After filtering, 37,867 SNPs remained.

Statistical analysis. Thermal growth coefficient (TGC) was calculated using the modified formula of Cho (1992): TGC = $[(W_2^{0.333} - W_1^{0.333}) / (T \times t)] \times 1000$, where T is temperature (°C), W_1 and W_2 are the body weights (grams) at tagging and harvest, respectively and t is the number of days between those measurements. Condition factor (CF) was calculated as $(W/l^3) \times 100{,}000$, where l is length (mm). The length-weight relationship was evaluated using the log transformation formula of Le Cren (1951): $W = aL^b$ and the linear regression of the log-transformed equation: log(W) = log(a) +blog(L), where the parameters a and b were calculated with 'a' representing the intercept and 'b' the slope of the relationship.

Descriptive statistics for TGC, CF and FF were obtained using R. Genetic parameters were obtained using the Restricted Maximum Likelihood procedure in ASReml-R version 4 (Butler *et al.*, 2017) fitting a bivariate animal model represented as;

$$y_i = X_i b_i + Z_i u_i + e_i \quad (1)$$

where subscript *i* refers to trait environment, y_i is the vector of recorded phenotypes, X_i and Z_i are incidence matrices relating phenotypes with the fixed and random effects, b_i is the vector of the fixed effects of the contemporary group (defined as tank history for tank traits and harvest date for sea-pen traits), u_i is a vector of random additive genetic effects $\sim N(0, G_i\sigma^2_{gi})$ where G_i is the submatrix of G for animals in environment i and σ^2_{gi} is the additive genetic variance, e_i is a vector of random residual effects $\sim N(0, I_i\sigma^2_{ei})$ where I_i is an identity matrix and σ^2_{ei} is the residual variance. Additionally, $cov(u_1, u_2) = r_G \sigma_{g1} \sigma_{g2}$. The bivariate models were fitted for TGC, CF and FF for both environments.

Results

Descriptive statistics are summarised in Table 1.

The length-weight relationship of fish stocked in different tanks is presented in Table 2. For a 1% increase in length, we would expect a 3.39% increase in fish weight in the FRC tanks and a 2.62% increase in the sea-pens.

Table 1. Descriptive statistics of traits assessed in commercial sea-pens (average age 730 days) and the FRC tanks (average age 463 days) in New Zealand in 2021.

Location	Trait	n	Mean	SD[1]	Min	Max	CV[2] (%)
Tanks							
	Condition factor	1,397	2.33	0.21	1.25	2.97	9
	Fillet fat (g/100 g)	390	23.60	3.39	15.34	36.29	14
	TGC[3]	1,398	1.95	0.18	1.25	2.45	9
Sea-pens							
	Condition factor	1,288	1.95	0.18	1.32	3.34	9
	Fillet fat (g/100 g)	1,287	34.42	1.34	29.64	38.33	4
	TGC	1,288	2.69	0.17	2.08	3.21	6

[1] SD = Standard deviation across location.
[2] CV = coefficient of variation.
[3] TGC = thermal growth coefficient.

Table 2. The length weight relationship of Chinook Salmon reared in FRC tanks and commercial sea-pens.

Environment	Mean weight (g)	Correlation coefficient (r)	b[1]	SE[2]
Tanks	1,500	0.91	3.39	0.04
Sea-pens	6,390	0.83	2.62	0.05

[1] b = the slope of the relationship.
[2] SE = standard error of b.

Estimates of heritabilities and genetic correlations are presented in Table 3. All heritability estimates were greater in the FRC tanks compared to the commercial sea-pens. The genetic correlations between the traits were lowest for TGC and greatest for FF.

Discussion

Using a bivariate animal model, this study produced genetic correlations between TGC, CF and FF of Chinook salmon reared in two different environments. Growth performance was represented by the TGC and varied between the two environments (1.95 from tanks vs 2.69 from sea-pens). When length was regressed with weight, the slope value was significantly lower than critical isometric value (3) for the sea-pens indicating negative algometric growth; thus fish become more slender as they increase in length (Pauly 1984). In comparison the slope was greater than 3 for the animals in the tanks, indicating the fish becomes more round for its weight, as it grows longer (Thakur and Das 1974). The negative allometric growth in sea-pens could be due to different environment and/or feeding regimes in sea compared to in the tanks.

Table 3. Estimated heritabilities (±SE) and genetic correlations (rG) of weight, condition factor and fillet fat of Chinook Salmon reared in FRC tanks and commercial sea-pens.

Trait	Tank h^2	Sea-pen h^2	Genetic correlation
TGC[1]	0.52 (0.04)	0.39 (0.05)	0.46 (0.09)
Condition factor	0.42 (0.04)	0.37 (0.05)	0.60 (0.10)
Fillet fat	0.42 (0.10)	0.35 (0.04)	0.78 (0.13)

[1] TGC = thermal growth coefficient.

All heritability estimates were greater for traits observed in the tanks vs sea-pen. This could be attributed to the age differences and greater environmental variation in the sea-pen as the rearing conditions were tightly controlled throughout the FRC trial. Moderate to high heritabilities were observed for all traits. The heritabilities for TGC were similar to those obtained for body weight from Sanford of 0.30-0.48 (Symonds *et al.*, 2019) and within the range of heritabilities reported for TGC in Atlantic salmon (0.35-0.56; Thorland *et al.*, 2020).

This study provides the first heritability estimates of CF in NZ Chinook salmon. These estimates are greater than those reported in trout (0.19; Gjerde and Schaeffer 1989), similar to Atlantic Salmon (0.42; Garber *et al.*, 2019) and within the range of heritabilities (0-0.6) reported by Carlson and Seamons (2008). In the past, heritability estimates of flesh quality traits in this population ranged from 0.04-0.18 (Symonds *et al.*, 2019), which are similar estimates to those reported in Coho salmon (0.17; Neira *et al.*, 2004) and Atlantic salmon (0.28; Powell *et al.*, 2008) but are lower than the 0.35-0.42 obtained for FF in this study.

Results show substantial additive genetic variance for traits of interest that can be exploited through selective breeding. The genetic correlation for FF could be influenced by the difference in measuring FF. The low to moderate genetic correlations between the traits suggest the traits in the FRC tank evaluation are similar to those in the commercial sea-pen evaluation. These results support our hypothesis that tank evaluation can be used to predict commercial family performance, but the low correlations highlight the need to use information from sibs in both environments to optimise the response to selection. Future FRC studies will evaluate more difficult to measure traits that otherwise cannot be evaluated in the sea-pens, to enable selection for these desirable traits in broodstock.

References

Butler D.G., Cullis B.R., Gilmour A.R., Gogel B.G., *et al.* 2017. ASReml-R

Reference Manual Version 4. VSN International Ltd, Hemel Hempstead, HP1 1ES, UK.

Carlson S.M., and Seamons T.R. (2008). Evol. Appl. 1(2):222-238.

Cho C.Y. 1992. Aquac. 100:107-123.

Clark L.V. and Sacks E.J. (2016) *Source Code Biol. Med.* 11:1.

Dodds K.G., McEwan J.C., Brauning R., Anderson R.M., *et al.* 2015. BMC Genom. 16:1047. https://doi.org/10.1186/s12864-015-2252-3

Dodds K., McEwan J., Bilton T., Brauning R., *et al.* (2018) Australasian Applied Stat. Conf.:38 (abstract).

Garber A.F., Amini F., Gezan S.A., Swift B.D., *et al.* (2019). Aquac. 503:242-253.

Gjerde B.A., and Schaeffer L.R. (1989). Aquac. 80(1-2):25-44.

Le Cren E.D. (1951). J. Ani. Ecol. 20:201-219.

Lu F., Lipka A.E., Glaubitz J., Elshire R., *et al.* (2013) PLoS Genet. 9:e1003215.

Lulijwa R., Young,T., Symonds J.E., Walker S.P., *et al.* (2021). Metabolites, 11(8):547.

Miller M.R., Puddick J., Symonds J.E., Walker S.P., *et al.* (2018). NIR news, 29(5), 12-14.

Neira R., Lhorente J.P., Araneda C., Díaz N., *et al.* (2004). Aquac. 241(1-4):117-131.

Pauly D. (1984). WorldFish. 8:325.

Powell J., White I., Guy D., Brotherstone S. (2008). Aquac. 274(2-4):225-231.

Reis M.G., Agnew M., Jacobs N., Reis M.M. (2021). NZ Instit. Food. Sci. Tech. Conf.:P2 (abstract).

Symonds J.E., Clarke S.M., King N., Walker S.P., *et al.* (2019). Front. Genet. 10:27.

Thakur N.K., Das M.K., (1974). J Inland Fish Soc India. 6:95–99.

Thorland I., Thodesen J., Refstie T., Folkedal O., *et al.* (2020). Aquac. 518:734735.

577. Testing fat content or feeding behaviour as indirect selection criteria for feed efficiency using trout isogenic lines

S. Pouil[1*], T. Kernéis[2], E. Quillet[1], D. Lallias[1], L. Labbé[2], F. Phocas[1] and M. Dupont-Nivet[1]

[1]Université Paris-Saclay, INRAE, AgroParisTech, GABI, 38350 Jouy-en-Josas, France; [2]INRAE, UE 0937 PEIMA (Pisciculture Expérimentale INRAE des Monts d'Arrée), 29450 Sizun, France; simon.pouil@inrae.fr

Abstract

Phenotyping individual feed intake through direct measurements remains complex in fish, and one way to overcome this challenge is to find indirect criteria for estimating feed efficiency. This study investigated the correlations between fish residual feed intake (RFI) and fat content or the number of feed demands in ten isogenic lines of rainbow trout. We found positive correlations between RFI and fat content indicating that fatter fish tend to be less feed efficient. However, the association between RFI and fat content was not constant over time and was driven mostly by one replicate from one line. We also found positive correlations between RFI and the total number of feed demands. Our findings highlight the interest of using fat content in selective breeding programs to improve feed efficiency and highlighted the interest in exploring the relationship between feed efficiency and feeding behaviour in fish.

Introduction

Improving feed efficiency (FE) in fish can be achieved through various approaches, including husbandry, nutrition, and genetics. FE is one of the main traits of interest for fish breeding programs (Chavanne *et al.*, 2016). However, unless infrastructure is available for separately rear of families as previously done for Atlantic salmon in Norway, selection for FE remains challenging, as it requires precise recording of individual feed intake (FI). Several approaches can be used for phenotyping individual FI in fish, including video-recording during feeding, the use of radio-opaque feed, and the counting of uneaten pellets in individual rearing. Nevertheless, these methods have major drawbacks as reviewed by de Verdal *et al.* (2018).

On an experimental scale, one tool to overcome the drawbacks of typical FI assessment methods is using isogenic lines (Quillet *et al.*, 2007), within which all individuals are genetically identical. Because the genetic variability within an isogenic line is zero, fish belonging to the same line may be reared in the same tank allowing a precise assessment of FI of their shared genotype under conditions close to production systems (Knap and Kause, 2018). In addition, combining several isogenic lines gives access to the between-line genetic variability of FE (Grima *et al.*, 2008).

Many indirect criteria for estimating FE have been studied. Body weight loss and gain during feed deprivation (FD) and subsequent compensatory growth (CG) periods and fat content have been suggested as indirect estimators of FE (de Verdal *et al.*, 2018). Feed efficiency can also be linked to feeding behaviour. For example, Martins *et al.* (2006) have shown that African catfish with better FE spend less time eating. In Nile tilapia, Martins *et al.* (2011) found that FE was negatively correlated with feeding latency and positively correlated with feeding time and the number of feeding acts.

The objectives of the present study were to investigate the links between residual feed intake (RFI), fat content and feeding behaviour.

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_577, © S. Pouil *et al.* 2022

Materials & methods

Heterozygous isogenic lines. Parents were derived from the INRAE homozygous isogenic lines previously established after two generations of gynogenesis (Quillet *et al.*, 2007). Ten heterozygous isogenic lines were produced by mating females from a single maternal homozygous isogenic line (named B57) with sex-reversed XX males from ten other homozygous isogenic lines (A02, A03, A22, A36, AB1, AP2, B45, B61, N38, R25) to produce genetically identical heterozygous (h) individuals exhibiting performances across the range of the population from which the isogenic lines are derived. Only one maternal line was used to avoid maternal effects which could bias the results.

Successive experimental phases. The experiment reported here started at 137 days post fertilization (dpf). Feed intake was measured three times: from 265 dpf to 297 dpf (fi1) then from 439 to 469 dpf (fi2) and from 522 to 547 dpf (fi3). The first period (fi1) of FI assessment was performed under normal growth conditions, while the two others (fi2, fi3) followed FD periods. From 137 to 549 dpf, each line was reared separately into 1.8 m^3 outdoor tanks, with three triplicates per line, supplied with river water (11.3-16.0 °C). In order to keep a density below 50 kg m^{-3}, regular random eliminations were performed, and 100 fish per replicate remained at the end of the experiment. Commercial pellets were distributed by self-feeders (Imetronic®, France) throughout the experiment.

Weight and residual feed intake. Body weight (BW) was measured regularly throughout the experiment (n≈100 fish for each of three replicates). A way to measure FE is to estimate the residual feed intake (RFI) using linear regression analysis:

$$RFI = FI - (\beta_0 + \beta_1 xBW^{0.8} + \beta_2 xBWG) \quad (1)$$

With FI being the observed feed intake and $(\beta_0 + \beta_1 xBW^{0.8} + \beta_2 xBWG)$ being the predicted feed intake based on a linear regression of the observed FI on metabolic body weight ($BW^{0.8}$) and body weight gain (BWG).

Fat content. The total fat content in muscle was recorded three times during the experiment (at 471, 520, and 549 dpf, i.e. measurements made at the end of fi2, right before fi3 and at the end of fi3, respectively; n≈100 fish per replicate) using a Fish Torry Fat-meter® (Douirin *et al.*, 1998).

Feed demand. Self-feeders were used continuously throughout the experiment to characterize feeding behaviour for the ten isogenic lines by recording feed demands (i.e. rewarded, unrewarded, and both of them). This feeding technique is based on the learning ability of fish, and feed is delivered depending on their demand. In a self-feeding system, fish are assumed to precisely control their feeding by activating a 'trigger' sensor. All the self-feeders were connected to a computer system recording the feed demands.

Results

The correlations between RFI derived from the first measurement under normal feeding regime (fi1) and the two others measures following a FD period (fi2 and fi3) were low ($r \leq 0.25$, $P \geq 0.18$). However, the correlation between the last two RFI measured after FD, i.e. fi2 and fi3, was significant ($r=0.58$, $P<0.01$).

Muscle fat content was comparable across the lines at 471 dpf (5.9±0.8%), but a significant line effect was observed at 520 and 549 dpf. On average, fish lost between 20% (B61h) and 58% (N38h) of muscle fat during the FD period while they gained from 31% (AB1h) to 150% (N38h) of their muscle fat during the re-feeding period.

We found a positive correlation between RFI calculated from the two FI measurements following FD periods (fi2 and fi3) and the fat content measured at the end of the FI measurements (at 471 and 549 dpf). This correlation was only significant for the fi3 measurement period but mostly driven by one replicate belonging to AP2h (Figure 1). Overall, we found that RFI increased when fat content increased.

The daily profile of feeding behaviour differed significantly between lines only during the two last FI measurement periods (fi2 and fi3). Interestingly, we found significant correlations between RFI calculated from the first FI measurement period (fi1) and feeding behaviour, assessed as the total number of feed demands over the same period (r=0.42-0.49, $P \leq 0.02$; Figure 2). Such correlations were not found for the other FI measurement periods.

Discussion

Previous findings highlighted the relationship between fat content and feed efficiency in rainbow trout and other salmonid species. In rainbow trout, divergent selection for muscle fat content has been carried out (Quillet *et al.*, 2005). After four generations of selection, fish from the lean-muscle selected line had lower body and muscle lipid contents, and their feed conversion ratio (FCR) was improved by about 6-9% compared to the fat-muscle selected line (Kamalam *et al.*, 2012). Kause *et al.* (2016) have shown that rainbow trout with a low body fat content and low muscle fat content were more efficient in feed and protein utilization, highlighting that controlling fat deposition in fish breeding programs could improve FE and reduce nutrient waste. Our results, from isogenic lines, suggest that such conclusions need to be qualified.

In brown trout, Mambrini *et al.* (2004) found that selection for growth led to changes in feeding activity as assessed using self-feeders, but differences between control and selected lines were significant for only three out of the five periods investigated. We obtained similar results for rainbow trout isogenic lines. RFI was positively correlated with the number of feed demands at a routine feeding regime (i.e. without any FD),

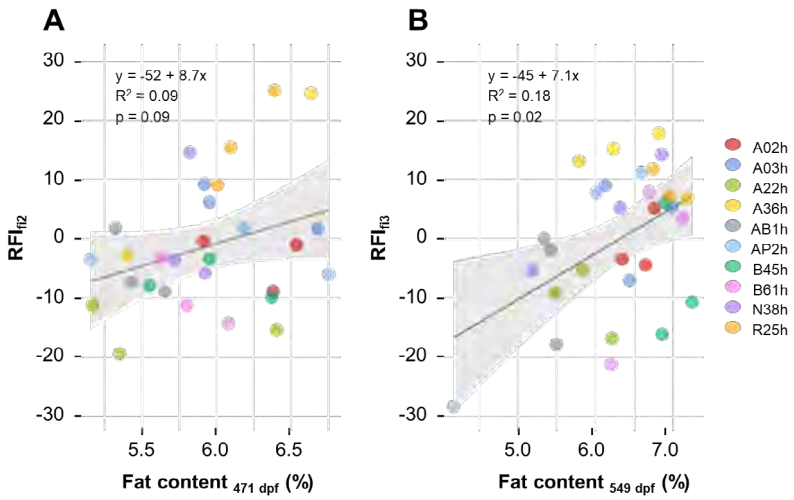


Figure 1. Relationship between fat content in muscle measured at 471 and 549 dpf, i.e. right after the two last feed intake periods: from 439 to 469 dpf (fi2) and from 522 to 547 dpf and the RFI calculated for the same periods: (A) fi2 and (B) fi3. Dots sharing the same colour are replicates of the same isogenic line. Grey shading areas indicate the 95% confidence intervals.

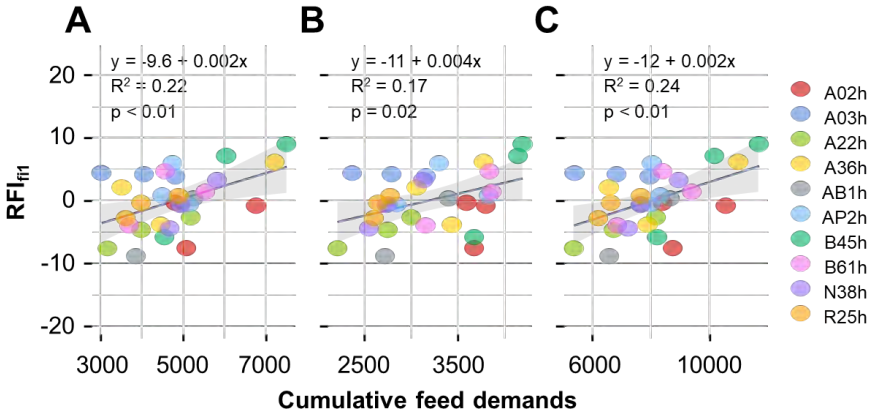


Figure 2. Relationship between the daily feed demands: (A) unrewarded, (B) rewarded and (C) total and the RFI calculated during the first feed intake period (from d 265 to 297 dpf; fi1). Dots sharing the same colour are replicates of a same isogenic line. Grey shading areas indicate the 95% confidence intervals.

but not after FD periods. In the regular feeding regime, the most feed-efficient isogenic lines (lower RFI) were the ones that requested less feed via the self-feeders. Such findings are in accordance with Martins *et al.* (2006, 2011), who found that, in African catfish or in Nile tilapia, the most feed efficient fish were the less active for eating.

To conclude, while we did not find clear evidence about the relevance of using fat content as an indirect selection criterion for FE in breeding programs, our results showed the interest in exploring other indirect selection criteria, particularly linked to feeding behaviour.

References

Chavanne H., Janssen K., Hofherr J., Contini F., *et al.* (2016). Aquacult Int 24:1287-1307. https://doi.org/10.1007/s10499-016-9985-0

de Verdal H., Komenc H., Quillet E., Chatain B., *et al.* (2018). Rev Aquac 10(4):833-851. https://doi.org/10.1111/raq.12202

Douirin C., Haffray P., Vallet J.-L., and Fauconneau B. (1998). Sci Aliments 18:527-535.

Grima L., Quillet E., Boujard T., Robert-Granié C., Chatain B., *et al.* (2008). Genet Sel Evol 40(6):607-624. https://doi.org/10.1186/1297-9686-40-6-607

Kamalam B.S., Medale F., Kaushik S., Polakof S., *et al.* (2012). J Exp Biol 215(15):2567-2578. https://doi.org/10.1242/jeb.070581

Kause A., Kiessling A., Martin S.A., Houlihan D., and Ruohonen K. (2016). Br J Nutr 116(9):1656-1665. https://doi.org/1017/S0007114516003603

Knap P.W., and Kause A. (2018). Front Genet 9:184. https://doi.org/10.3389/fgene.2018.00184

Mambrini M., Sanchez M.P., Chevassus B., Labbé L., *et al.* (2004). Livest Prod Sci 88(1-2):85-98. https://doi.org/10.1016/j.livprodsci.2003.10.005

Martins C.I.M., Schrama J.W., and Verreth J.A.J. (2006). Aquaculture 256(1-4):588-595. https://doi.org/10.1016/j.aquaculture.2006.02.051

Martins C.I.M., Conceição L.E., and Schrama J.W. (2011). Appl Anim Behav Sci 133(1-2):109-116. https://doi.org/10.1016/j.applanim.2011.05.001

Quillet E., Le Guillou S., Aubin J., and Fauconneau B. (2005). Aquaculture 245:49-61. https://doi.org/10.1016/j.aquaculture.2004.12.014

Quillet E., Dorson M., Le Guillou S., Benmansour A., and Boudinot P. (2007). Fish Shellfish Immunol 22(5):510-519. https://doi.org/10.1016/j.fsi.2006.07.002

578. The possibilities of using electronic sensors in aquaculture breeding

W. Abbink[1*], A. Palstra[1], W. Agbeti[1], G. Lembo[2] and J. Komen[1]

[1]*Wageningen University & Research, P.O. Box 338, 6700 AH Wageningen, the Netherlands;*
[2]*COISPA Tecnologia & Ricerca, via Dei Trulli 18, 70126 Bari, Italy;* wout.abbink@wur.nl

Abstract

The use of electronic sensors can greatly enhance our understanding of fish behaviour and physiology, and sensor data are emerging as key sources of information for improving aquaculture management practices. An integrated and interdisciplinary approach to monitor the activity pattern and physiological status of farmed fish, as well as environmental parameters, will enhance the sustainability and profitability of European aquaculture. For tailor made breeding programs, sensor technology offers the possibility to monitor traits in individual fish or families related to key factors that influence animal health and behaviour, such as hypoxia or sub-optimal water current. The better performing fish under these challenging environmental conditions can be selected for breeding.

Electronic sensors in aquaculture

Studying fish behaviour, physiology, health and welfare in aquaculture settings, as well as the farming environment, presents challenges because of the physical characteristics of water and the dimensions of the water bodies where the fish are cultured in. However, in recent years, the use of electronic sensors in aquaculture research is increasing, and is significantly improving our understanding of fish performance in aquaculture, as sensor data emerge as important sources of information for improving aquaculture management practices.

Collecting information of individual fish in aquaculture was difficult before sensor technology became available and often included catching, sedating and handling of the fish. Fish in aquaculture systems are grown in large water bodies in a high density, where the fish can't be distinguished individually for observation by the human eye or camera systems. Group observations during resting and feeding, monitoring fish growth and mortality, and monitoring the water quality, were, and often still are, management strategies to safeguard the fish in aquaculture systems.

Although sensors are widely used in terrestrial applications (Ellen *et al.,* 2019), their use in the marine environment and aquaculture was limited up to recent years because of the special attributes of water, the high cost of underwater sensors and malfunctions due to the harsh environment. Underwater sensors should be cost- and power-efficient, but also be robust and waterproof, non-metallic, and should have no effects on surrounding organisms (Parra *et al.,* 2018).

The various types of sensor technology that are nowadays available can be used to study the farming environment and to assess biomass and fish performance. Hence, depending on the purpose of the sensor and its physical appearance, sensors can be operated either inside the fish, attached to the fish, or placed in the water.

Biomass, environment, and fish tracking

Biomass estimators that can be installed on the feeding stations or on the cages or tanks, can provide information to monitor or expand the production. In addition, sensor- and camera systems may provide information for estimating the biomass in the cages and developing reliable fish feeding strategies.

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_578, © W. Abbink *et al.* 2022

Sensors that monitor the environment in aquaculture settings record parameters such as oxygen, temperature, salinity, sea current, pH, wind and CO_2. Whereas in open sea cages this information can be used for management practices, such as feeding strategy, in Recirculation Aquaculture Systems this information can give full adaptive control over water quality parameters.

When monitoring individual fish, electronic sensor tags can track and identify individual fish (Lembo *et al.,* 2002). This provides information on the swimming behaviour of the fish in the aquaculture setting. Depending on the type of sensor, these tags have to be recovered from the fish to obtain the data, or the sensor sends the data to an external receiver. Placed in or on fish, more advanced tags incorporate sensors that measure and record multiple environmental and/or biological parameters of the fish during production (Cooke *et al.,* 2016).

A combination of sensor technologies can form a wireless communication system to monitor large scale aquaculture activities, and increases the accuracy of data, processing and interpretation (Kessel *et al.,* 2014). These data provide the industry with information needed to facilitate performance, health and welfare and to optimise management practices.

Fish breeding

For breeding purposes, the availability of sensor technologies allows to track and monitor individual fish. A large amount of data can be obtained from individual fish for specific traits related to environmental factors that influence animal health and behaviour, such as hypoxia or water current, and this information can be linked to the genotype of the individual fish. Better performing fish on desired traits can then be selected for tailor made breeding programs.

Sensor types and possibilities

Accelerometer tags can transmit a 3D acceleration of the tagged fish when they are within the receiver array, or logged on the sensor. The fish acceleration is the result of measuring acceleration on three axes. Accelerometer tags can be used as measure of animal activity, such as swimming speed via tail beat acceleration, detecting mortality, feeding, spawning activity, nocturnal/diurnal activity, and activity responses to changing oxygen, salinity and temperature (https://www.vemco.com/products/v9ap-v13ap-accelerometer/; https://www.thelmabiotel.com/).

Other sensor types measure heart rate, activity and body temperature simultaneously. These are ideal for monitoring behaviour and stress response, animal welfare and fish physiology (https://www.cefastechnology.co.uk/products/data-storage-tags/g7/).

Sensors that can record physiological data from free-swimming fish can be placed either externally or internally on the fish. The sensors measure muscle activity using probes inserted into the musculature of the fish and can provide an estimate of the energetic costs associated with physical activity via electromyograms (EMGs). These are records of bioelectric potentials that are strongly correlated with the strength and duration of muscle contractions, and as indicators of the intensity of fish activity (http://www.lotek.com/cemg2-series-emg-transmitters.htm).

Limitations

Sensor technology that is used to obtain information on individual fish poses constraints in the form of time and money, and physical limitations. When the sensors are be placed on, and particularly in the fish, this requires surgery by qualified staff, and recovery time is needed. In addition, the sensors are expensive, with thousands or tens of thousands of euro's needed for the sensors alone. These two factors can form a

limitation to the number of fish equipped with a sensor. In addition, the environmental conditions for the sensors are challenging, especially in salt water. In addition, getting data from the sensors requires either retrieving the sensor physically from the fish, or having the fish swim in reach of receivers, which can also be challenging in (salt) water.

Case study

In a case study performed by researchers from Wageningen Livestock Research, electronic sensors were implanted in yellowtail kingfish (*Seriola lalandi*), a powerful and pelagic fish that lives in warmer waters worldwide. The STAR ODDI logger DST milli-HRT (Figure 1; https://www.star-oddi.com/products/data-loggers?sensors=heart-rate) monitors heart rate, body temperature and acceleration in the fish simultaneously and is ideal for monitoring behaviour and stress response of the fish. The logger takes a burst measurement of ECG at a programmed time interval and calculates the mean heart rate for each recording.

The fish were exposed to a series of increasing water flow speeds ranging from 0.2 to 1.0 m/s in a swim tunnel (van den Thillart *et al.,* 2004; Palstra *et al.,* 2015), forcing the fish to swim against the current. Preliminary results for heart rate and external acceleration at increasing swim speeds are shown in Figure 2. The heart rate remains at a (high) basal level at lower swim speeds, only to increase towards the highest speeds of 0.8 and 1.0 m/s. The external acceleration shows a similar pattern and appears, in this fish, to be correlated to the swim speed. However, preliminary data from other fish equipped with a sensor show that individual variation in heart rate and external acceleration at different swim speeds proves that electronic sensors indeed are a powerful tool to acquire in-dept physiological and performance information on individual fish that can be used for aquaculture management practices.

Concluding remarks

The use of electronic sensors offers great opportunities to study specific traits for breeding programs in individual fish in aquaculture settings. Enhanced environmental and biological sensor data collected by a network of electronic sensors can provide accurate measurements and predictive modelling of environmental conditions and desired traits on fish health, welfare and physiology, that can be linked to the genotype of the fish (Cooke *et al.,* 2016).

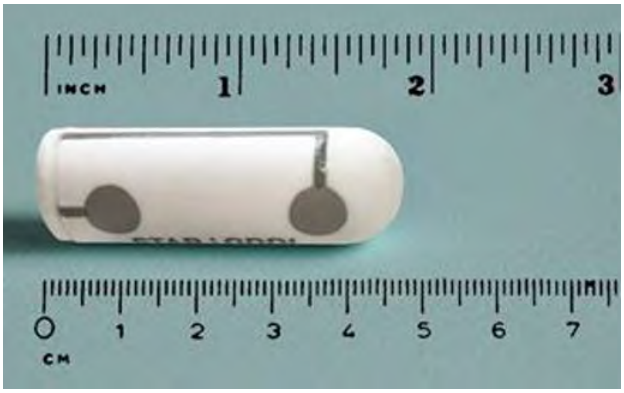


Figure 1. The STAR ODDI logger DST milli-HRT.

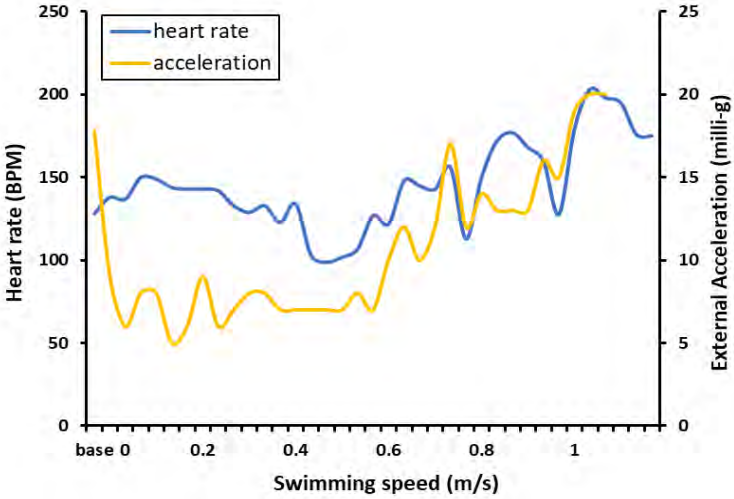


Figure 2. Heart rate and external acceleration of a yellowtail kingfish (*Seriola lalandi*) at increasing swimming speeds.

References

Cefas Technology. Available at: https://www.cefastechnology.co.uk/products/data-storage-tags/g7/

Cooke S.J., Brownscombe J.W., Raby G.D., Broell F., Hinch S.G. *et al.* (2016) Comp Biochem Physiol A 202: 23–37. http://doi.org/10.1016/j.cbpa.2016.03.022

Ellen E.D., van der Sluis M., Siegford J., Guzhva O., Toscano, M.J. *et al.* (2019) Animals 9(3), 108; https://doi.org/10.3390/ani9030108

Kessel S.T., Cooke S.J., Heupel M.R., Hussey N.E., Simpfendorfer C.A. *et al.* (2014). Rev Fish Biol Fisher 24: 199–218. https://doi.org/10.1007/s00227-018-3384-1

Lembo G., Spedicato M.T., Økland F. Carbonara P., Fleming, I.A. *et al.* (2002). Hydrobiol. 483 249-257. http://doi.org/10.1023/A:1021360419150

Lotek. Available at: http://www.lotek.com/cemg2-series-emg-transmitters.htm

Palstra A.P., Mes D., Kusters K., Roques J.A.C., Flik G., *et al.* (2015) Front. Physiol. 5:506. https://doi.org/10.3389/fphys.2014.00506

Parra L., Lloret G., Lloret J. and Rodilla, M. (2018). IEEE Sens J 18(10), 3915-3923. http://doi.org/10.1109/JSEN.2018.2817158

STAR ODDI. Available at: https://www.star-oddi.com/products/data-loggers?sensors=heart-rate

Thelmabiotel. Available at: https://www.thelmabiotel.com/

Van den Thillart G., Van Ginneken V., Körner F., Heijmans R., Van Der Linden R., *et al.* (2004) J. Fish Biol. 65(2): 312-318. https://doi.org/10.1111/j.0022-1112.2004.00447.x

Vemco. Available at: https://www.vemco.com/products/v9ap-v13ap-accelerometer/

579. Genetic parameters and genome-wide association study of resistance to acute hyperthermia in rainbow trout

H. Lagarde[1*], D. Lallias[1], P. Patrice[2], M. Prchal[3], Y. François[2], J. D'Ambrosio[2], A. Acin-Perez[4], F. Cachelou[4], P. Haffray[2], M. Dupont-Nivet[1] and F. Phocas[1]

[1]Université Paris-Saclay, INRAE, AgroParisTech, GABI, 78350 Jouy-en-Josas, France; [2]SYSAAF, French Poultry and Aquaculture Breeders Association, 35042 Rennes, France; [3]University of South Bohemia in České Budějovice, Faculty of Fisheries and Protection of Waters, South Bohemian Research Center of Aquaculture and Biodiversity of Hydrocenoses, Zátiší 728/II, 389 25 Vodňany, Czech Republic; [4]Viviers de Sarrance, Pisciculture Labedan, 64490 Sarrance, France; henri.lagarde@inrae.fr

Abstract

Selective breeding is a promising solution to reduce fish farms vulnerability to heat peaks. Objectives of this study were to give a new insight on the genetic architecture of resistance to acute hyperthermia stress in rainbow trout. At 275 days post-fertilization, 1,384 fish were phenotyped for acute hyperthermia resistance and body weight. Challenged fish were genotyped for 57K SNP and their genotypes were imputed at high-density thanks to their parents being genotyped on a 665K SNP array. Heritability estimate of resistance to acute hyperthermia was 0.32±0.04. This trait was genetically negatively correlated with body weight (-0.58±0.17). The genome-wide association study revealed that resistance to acute hyperthermia is highly polygenic as altogether the 5 detected QTLs explained less than 5% of the genetic variance. The main QTL region explained 3% of the genetic variance and contained two candidate genes previously described to be associated with temperature resistance.

Introduction

Rainbow trout (*Oncorhynchus mykiss*), a cultured salmonid of major economic importance, thrives in a narrow cold thermal range (Bear *et al.*, 2007). Exposition to high temperatures leads to the impairing of production efficiency with growth reduction, increase of disease and mortality. Extreme temperature events such as heat waves are expected to increase in intensity and frequency with global warming. The reduction of their impact represents a major challenge for the aquaculture industry. Among existing solutions, genetic improvement by selective breeding is an interesting approach: progress is cumulative from a generation to another and genetically improved animals could be exported to any farm without any prior equipment required. In rainbow trout, genetic determinism of resistance to acute temperature stress has already been studied in a Canadian commercial population (Perry *et al.* 2005) and QTLs were detected (Jackson *et al.*, 1998; Perry *et al.*, 2001). However, these studies concern a single population, distant from our study population, and recent advances in genomics may provide more accurate results for QTL detection.

Objectives of the present study were to complement existing knowledge for introducing acute hyperthermia resistance trait in rainbow trout breeding programs by: (1) assessing genetic parameters for acute hyperthermia resistance in a new commercial population; (2) detecting with a powerful tool (new high-density SNP array) QTLs associated with acute hyperthermia resistance; and (3) identifying candidate genes.

Materials & methods

The experiment was carried out according to the European guidelines; the protocols were evaluated and approved by the ethic committee ANSES/ENVA/UPC No 16 and authorized by the French ministry of higher education and research (APAFIS#24441-2020022417122193).

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_579, © H. Lagarde *et al.* 2022

Study population. A total of 1,384 fish from 573 full-sib families was produced in a partial factorial mating design from 74 dams and 99 sires derived from the Viviers de Sarrance (Nouvelle-Aquitaine, France) breeding program. At 265-days post-fertilization (dpf), fish were tagged with ISO RFID tags and fin-clipped for later DNA extraction and genotyping. At 272 dpf, fish were transported to ANSES-SYSAAF Fortior Genetics platform (Brittany, France). The challenge to hyperthermia was spread over 7 days (between 275 dpf and 285 dpf) with 198±12 fish per group. Around 9 a.m., fish from one of the 7 batches were transferred to the challenge tank, supplied with identical river water. Once the transfer completed, temperature was gradually increased from initial river water temperature (17.3±0.7 °C) at a rate of 1.5±0.2 °C/hour. Above 27 °C, fish gradually began to lose equilibrium. Each fish losing equilibrium was removed from the tank and identified. The time was recorded and the fish was weighed and euthanized. The acute hyperthermia resistance phenotype was considered as time to loss of equilibrium (TLE). Body weight (BW) was measured to estimate phenotypic and genetic correlation between BW (mean BW 87±12 grams) and TLE. Fish were genotyped with the 57K Axiom™ Trout Genotyping array (Palti *et al.*, 2015) at the INRAE genotyping Platform Gentyane. Their parents were genotyped for 664,531 SNPs with a new high-density Axiom™ Trout Genotyping array. After quality control and parental assignation with APIS software (Griot *et al.*, 2020), the final dataset contained 1,332 offspring genotyped for 30,379 SNPs (LD genotypes). The parent's final dataset contained 172 individuals genotyped for 420,079 SNPs (HD genotypes). The mean number of progenies per sire, per dam and per full-sibs family were respectively 14.0±6.4, 18.7±7.5 and 2.4±2.5. Offspring's LD genotypes were imputed to HD thanks to the parental reference HD genotypes using FIMPUTE3 software (Sargolzaei *et al.*, 2014).

Statistical analysis. Genetic (co)variance components of acute hyperthermia resistance and body weight at 275 dpf were estimated with AIREML algorithm in BLUPF90 software (Misztal *et al.*, 2002) using the following animal model:

$$y_{ij} = \mu + day_i + a_j + \varepsilon_{ij} \quad (1)$$

where y_{ij} is the performance (TLE or BW at 275 days) of animal j, μ is the overall mean of the population, day_i is the fixed effect of day of challenge i, a_j is the additive effect of animal j and ε_{ij} is the random residual error. Both pedigree and genomic relationship matrices were built. The pedigree was constituted of 20,372 animals over ten generations. The genomic matrix was built with the HD genotypes.

GWAS was performed on temperature resistance trait using the Bayesian approach BayesCπ implemented in the BESSiE software with the following model:

$$TLE_{ij} = \mu + day_i + \sum_{k=1}^{n} \delta_{kl} z_{ijk} a_k + \varepsilon_{ijl} \quad (2)$$

With TLE_{ij}, the time before loss of equilibrium, the temperature resistance phenotype of animal j, μ the overall mean of the population, day_i the fixed effect of day of challenge i and n the total number of SNPs (420,079). δ_{kl} is an indicator variable: within a cycle l, δ_{kl}=1 if the effect of SNP k is estimated in this cycle or δ_{kl}=0 if not. In each cycle, δ_{kl} is sampled from a binomial distribution with a probability π that δ_{kl} was equal to 1. The proportion π was sampled from a beta distribution B(α,β) with α=400 and β=420,079; which represents approximatively 0.1% of SNPs selected at each cycle. z_{ijk} was the genotype on locus k for individual j of day i (coded as 0, 1, or 2), a_k the effect of the reference allele of SNP k, and ε_{ijl} the residual effect. The Markov chain Monte Carlo was run with 400,000 cycles and a burn-in period of 10,000 cycles. Results were saved every 40 cycles. The Bayes factor (BF) was calculated to quantify the degree of association between a SNP and the phenotype. Evidence for a QTL was provided by a value 2×ln(BF)≥6 at a peak SNP. A credibility interval for the QTL was computed around the peak SNP including all SNPs for

which 2×ln(BF)≥3 in a sliding window of 0.2 Mb. Genes in the QTL region were annotated with the NCBI *O. mykiss* genome assembly (GCA_013265735.3 USDA_OmykA_1.1) (Gao *et al.*, 2021).

Results & discussion

Genetic parameters of TLE and BW are presented in Table 1. Pedigree heritability and genomic heritability of acute hyperthermia resistance were estimated at similar values of 0.32±0.06 and 0.28±0.04. Those values are lower than the previous heritability estimate 0.41± 0.06 (Perry *et al.* 2005), but still suggest a good potential for genetic improvement. Heritability of BW was estimated to 0.14±0.04 with the pedigree relationship matrix and 0.20±0.04 with the genomic relationship one, which is consistent with previous estimates (e.g. Leeds *et al.*, 2016). Phenotypic correlation between acute hyperthermia resistance and BW was not statistically different from zero (-0.067±0.039), but the genetic correlation was estimated significantly negative (-0.58±0.17) while Perry *et al.* (2005) found it not significant (-0.032±0.18). If our result is confirmed, it will be necessary to consider the trade-off between temperature resistance and fish growth in selective breeding programs.

With an evidence threshold of 2×ln(BF) ≥6, five QTLs were found, located on three chromosomes (Figure 1). One QTL, located on chromosome 13, had a very high evidence (2×ln(BF) >10) and explained almost 3% of the genetic variance of acute hyperthermia resistance (Table 2). This result suggests that resistance to acute hyperthermia is a highly polygenic trait. In the credibility interval of this QTL, 81 genes were identified. In a window of 100Kb around the peak SNP of this QTL, we identified two genes, *HSP70* and *FKB10*, which are directly involved in proteins folding in generic response to stress exposure. Both genes have already been

Table 1. Genetic parameters for resistance to acute hyperthermia in rainbow trout.[1]

Trait	Relationship matrix	σ^2_a	σ^2_e	σ^2_p	h^2(±se)
TLE	Pedigree	369	787	1,156	0.32 (±0.06)
TLE	Genomic	324	819	1,143	0.28 (±0.04)
BW	Pedigree	19	110	129	0.14 (±0.04)
BW	Genomic	26	103	129	0.20 (±0.04)

[1] σ^2_a = additive genetic variance; σ^2_e = environmental variance; σ^2_p = phenotypic variance; h^2(±se) = heritability ± standard error.

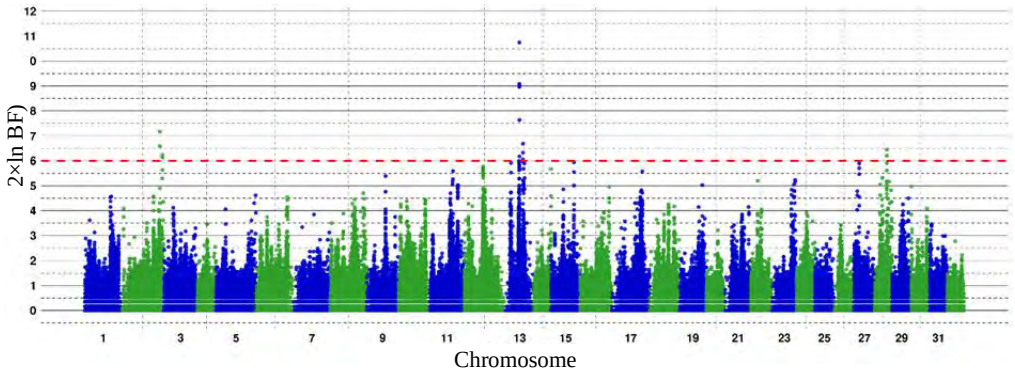


Figure 1. Manhattan plot of QTL detected for acute hyperthermia resistance trait. The red line corresponds to the QTL evidence threshold (2×ln(BF) ≥6).

Table 2. Effect and location of acute hyperthermia resistance QTL.

Chr.	Peak SNP	% of variance explained by peak SNP	2×ln(BF)	Credibility interval (Mb)	% of variance explained by QTL	# genes in QTL range
2	Affx-1237560225	0.2%	7.2	[94.49-94.95]	0.2%	15
2	Affx-277325684	0.1%	6.2	[101.37-101.43]	0.1%	3
13	Affx-1237409861	1.7%	10.7	[34.33-36.93]	2.9%	81
13	Affx-1237414100	0.1%	6.7	[45.60-45.93]	0.2%	11
28	Affx-1248614594	0.1%	6.4	[30.21-30.40]	0.2%	8

reported to be differentially expressed during temperature stress in salmonids (Akbarzadeh *et al.*, 2018) and thus may be suggested as functional candidates' genes.

References

Akbarzadeh, A., Günther, O.P., Houde, A.L., Li, S., Ming, *et al.* (2018) BMC Genomics 19. https://doi.org/10.1186/s12864-018-5108-9

Bear, E.A., McMahon, T.E., and Zale, A. V. (2007) Trans. Am. Fish. Soc. 136: 1113–1121. https://doi.org/10.1577/t06-072.1

Boerner V., and Tier B. (2016) Genet. Sel. Evol. 48:1–5. https://doi.org/10.1186/s12711-016-0241-x

Gao G., Magadan S., Waldbieser G.C., Youngblood, R.C., Wheeler, P.A. *et al.* (2021) G3 Genes, Genomes, Genet. 11. https://doi.org/10.1093/g3journal/jkab052

Griot R., Allal F., Brard-Fudulea S., Morvezen R., Haffray P., *et al.* (2020) Mol. Ecol. Resour. 20: 579–590. https://doi.org/10.1111/1755-0998.13103

Jackson T.R., Ferguson M.M., Danzmann R.G., Fishback A.G., Ihssen P.E *et al.* (1998) Heredity 80:143–151. https://doi.org/10.1038/sj.hdy.6882890

Leeds T.D., Vallejo R.L., Weber G.M., Gonzalez-Pena D., and Silverstein J.T. (2016) Aquaculture 465: 341–351. https://doi.org/10.1016/j.aquaculture.2016.08.036

Misztal I., Tsuruta S., Strabel T., Auvray B., Druet T., *et al.* (2002) Proc. of the 7th World Congress on Genetics Applied to Livestock Production. Montpellier, France.

Palti Y., Gao G., Liu S., Kent M.P., Lien S., *et al.* (2015) Mol. Ecol. Resour. 15: 662–672. https://doi.org/10.1111/1755-0998.12337

Perry G.M.L., Danzmann, R.G., Ferguson M.M., and Gibson J.P. (2001) Heredity (Edinb). 86, 333–341. https://doi.org/10.1046/j.1365-2540.2001.00838.x

Perry G.M.L., Martyniuk C.M., Ferguson M.M., and Danzmann R.G. (2005) Aquaculture 250: 120–128. https://doi.org/10.1016/j.aquaculture.2005.04.042

Sargolzaei M, Chesnais JP, and Schenkel FS. (2014) BMC Genomics 15: https://doi.org/10.1186/1471-2164-15-478

580. QTLs controlling swimming performance and their effect on growth in Nile tilapia (*Oreochromis niloticus*)

X. Yu[1*], S. Bekele Mengistu[1,2], H.A. Mulder[1], A.P. Palstra[1], J.A.H. Benzie[3,4], T. Quoc Trinh[3], M.A.M. Groenen[1], H. Komen[1] and H.J. Megens[1]

[1]Wageningen University & Research, P.O. Box 338, 6700 AH Wageningen, the Netherlands; [2]College of Agriculture, P.O. Box 5, Hawassa University, Hawassa Ethiopia; [3]WorldFish Centre, Jalan Batu Maung, Bayan Lepas, Penang, Malaysia; [4]School of Biological Earth and Environmental Sciences, University College Cork, Cork, Ireland; xiaofei.yu@wur.nl

Abstract

Critical swimming speed (U_{crit}) is an important measurement of swimming performance as well as a good indicator for cardio-respiratory health. It offers a new opportunity to select fish with better fitness. Here we present results from a GWAS study on critical swimming speed and growth in Nile tilapia under conditions of recurrent hypoxia. Heritability for U_{crit} was moderate (0.31±0.04). Large fish generally swim faster than smaller fish, whereas fish with better swimming performance tend to show slower growth later in life. Nine suggestive SNPs associated with swimming performance were identified under a hypoxic environment. A clear pleiotropic effect on growth traits was found for some of the SNPs associated with swimming performance, while other SNPs had only an effect on swimming performance, but not on growth. Our analyses sheds a first light on the genomic mechanisms of swimming performance and growth in Nile tilapia.

Introduction

Swimming performance is an important feature that correlates with fitness, survival, and metabolism in many aquacultural fish species (Palstra and Planas, 2012). Swimming performance in fish can be measured in a critical swimming challenge test, by which U_{crit} can be assessed. U_{crit} is the speed at which maximum oxygen uptake occurs. In tilapia, U_{crit} has been shown to be highly correlated with maximum metabolic rate (McKenzie *et al.*, 2003). However, it is unknown which QTLs are associated with U_{crit} in tilapia.

Nile tilapia (*Oreochromis niloticus*) is one of the most important farmed fish species in the world. Farming is mostly done in ponds and cages. However, for most smallholder farmers aeration of fishponds is not available or too expensive. In non-aerated ponds extreme hypoxia can occur, especially during the night when algae become net oxygen consumers (Mengistu *et al.*, 2020). Therefore, it is crucial to select tilapia that grow better under conditions where dissolved oxygen is limited.

The main objectives in this study were: (1) to estimate the genetic correlations between U_{crit} with body size traits at the swimming test, and with growth traits using a genomic relationship matrix, (2) to identify SNPs and genes associated with swimming performance and (3) to estimate the effect of candidate SNPs for U_{crit} on the growth traits.

Materials & methods

Experimental design and traits collection. The experimental design and data collection is described in full detail in Mengistu *et al.* (2021). Nile tilapia were collected as part of the GIFT selective breeding program. U_{crit} and body weight at test (BWtest) during the swimming test were recorded. After the swimming test, fish were stocked in non-aerated ponds for growing-out and harvested after 145 or 146 days. The stocking and harvest weight (Harw) were recorded, by and daily growth coefficient (DGC) was calculated.

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_580, © X. Yu *et al.* 2022

SNP genotyping and quality control. The DNA was extracted and genotyped by Identigen (Dublin, Ireland) using an Axiom® SNP array, which contains 65K SNP markers dispersed over the Nile tilapia reference genome (Peñaloza *et al*., 2020). Data was filtered to meet a dish quality control >0.82 and call rate for samples >0.93, respectively. Next, a second quality control step was applied based on per SNP call rate and minor allele frequency statistics.

Genetic parameters estimation. Variance components and heritability for swimming performance traits including U_{crit}, BWtest, Harw and DGC were estimated using univariate models based on residual maximum likelihood method with the following model was applied:

$$y = Xb + Z_1a + Z_2c + e \quad (1)$$

Phenotypic and genetic correlations between U_{crit} and traits BWtest, Harw and DGC were estimated based on a bivariate linear model. The fixed effects were the same as in the univariate models for U_{crit}, BWtest, Harw and DGC traits.

Association analysis and candidate genes in QTLs. Animal phenotypes and genotypes were jointly analysed to identify genomic regions associated with U_{crit}. The following model was applied:

$$y = W\alpha + Z\mu + x\beta + \varepsilon \quad (2)$$

LD and haplotype blocks were analysed with LDBlockShow (Dong *et al*., 2020), while r^2>0.8 as cut-off for blocks. Putative candidate genes associated with swimming performances were characterized within a 200 kb window size (100 kb upstream and downstream) flanking the candidate SNPs.

The effect of candidate SNPs in swimming performance on growth. We estimated the genetic association between SNPs and DGC based on generalized linear model. The best fitting genetic model was evaluated based on Akaike Information Criterion score. The significant threshold was defined as adjusted *P*-value<0.05.

Results

Genetic parameters estimation with different traits at swimming test. The heritability for U_{crit} was 0.31 with standard error 0.043, while the heritability for body weight at swimming test was 0.27 (Table 1). The common environmental effects were not significant for both traits. The phenotypic and genetic correlations between U_{crit} and BWtest were 0.43 and 0.36 (Table 2), indicating that fish with better swimming capacity usually show big size. The estimated r_g between U_{crit} and harvest weight (-0.13±0.13) and U_{crit} and DGC (-0.26±0.13) were slightly negative but not different from zero, suggesting that U_{crit} might have a negative genetic correlation to body weight at harvest and growth until harvest.

Table 1. Variance components for U_{crit} and body weight traits.

Traits	σ^2_a	σ^2_e	σ^2_p	h^2
U_{crit}	5.15	11.72	16.88	0.31 (0.043)
BWtest	1.23	3.38	4.61	0.27 (0.043)

Table 2. Genetic and phenotypic correlations between U_{crit} and body weight traits.

Traits	r_g	r_p
BWtest	0.43±0.10	0.36±0.03
Harw	-0.13±0.13	0.06±0.04
DGC	-0.26±0.13	-0.01±0.04

Genome-wide association study for swimming performance. Nine SNPs located on LG12, LG13, LG14 and LG19 exceeded the genome-wide suggestive threshold for U_{crit} (Figure 1). All genes in the up- and down-stream 100 kb regions of the 9 suggestive SNPs in a high LD (r^2>0.8) were investigated, which were *limk1a*, *elna*, *lsamp, aipl1* and *dner* on LG14, *pprc1* and *rrp12* on LG13, *hectd1 and smyd1* on LG19 and LG12.

Pleiotropic SNPs with swimming performance and growth. In this study, we investigated the impact of candidate nine SNPs associated with U_{crit} on growth traits. Three of these nine SNPs were significantly associated with both harvest weight and DGC, and SNP genotypes that presented a low mean U_{crit} had a high mean harvest weight and DGC.

Discussion

The observed genetic correlation between U_{crit} and Harw, and U_{crit} and DGC, suggests an negative interaction between growth and critical swimming speed under hypoxic environment.

Our results show several QTLs involved in the swimming performance in Nile tilapia. Overall, 16 genes in high LD with the identified SNPs, could be interesting candidate genes in regulation of swimming performance. For instance, the *rrp12* gene has been associated with adaptation to low oxygen (Foll *et al.*, 2014) and *pprc1* is potentially involved in cardiac and exercise-related muscle activity (Uguccioni *et al.*, 2010).

Our results show only part of the SNPs to be significantly associated with harvest weight and DGC, which located on several genes (*hip1*, *hectd1*, *dner*). They potentially can be applied as biomarkers to select fish with good cardio-respiratory health and good growth.

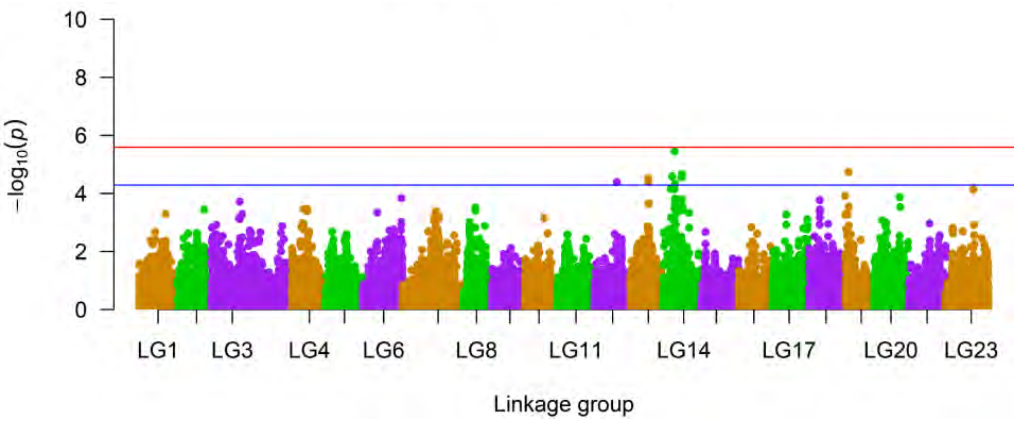


Figure 1. Manhattan plot for U_{crit}. The orange and blue horizontal line represent the genome-wide significance (2.57E-06) and suggestive significance threshold value (5.15E-05), respectively.

Funding

This work was funded by Koepon Foundation, International Fund for Agricultural Development (IFAD) Grant number: 2000001001, the CGIAR Research Program on Fish Agri-Food Systems (FISH) led by WorldFish. The FISH program is supported by contributors to the CGIAR Trust Fund. The authors also thank to the financial support from China Scholarship Council for Xiaofei Yu's PhD study.

References

Dong, S.-S., He, W.-M., Ji, J.-J., Zhang, C., Guo, Y., *et al.* 2020. LDBlockShow: a fast and convenient tool for visualizing linkage disequilibrium and haplotype blocks based on variant call format files. Briefings in Bioinformatics 00, 1–6. https://doi.org/10.1093/bib/bbaa227

Foll, M., Gaggiotti, O.E., Daub, J.T., Vatsiou, A., Excoffier, L., 2014. Widespread signals of convergent adaptation to high altitude in Asia and America. American Journal of Human Genetics 95, 394–407. https://doi.org/10.1016/j.ajhg.2014.09.002

McKenzie, D.J., Martinez, R., Morales, A., Acosta, J., Morales, R., *et al.* 2003. Effects of growth hormone transgenesis on metabolic rate, exercise performance and hypoxia tolerance in tilapia hybrids. Journal of Fish Biology 63, 398–409. https://doi.org/10.1046/j.1095-8649.2003.00162.x

Mengistu, S.B., Mulder, H.A., Benzie, J.A.H., Khaw, H.L., *et al.* 2020. Genotype by environment interaction between aerated and non-aerated ponds and the impact of aeration on genetic parameters in Nile tilapia (*Oreochromis niloticus*). Aquaculture 529, 735704. https://doi.org/10.1016/j.aquaculture.2020.735704

Mengistu, S.B., Palstra, A.P., Mulder, H.A., Benzie, J.A.H., Trinh, T.Q., Roozeboom, C., Komen, H., 2021. Heritable variation in swimming performance in Nile tilapia (Oreochromis niloticus) and negative genetic correlations with growth and harvest weight. Scientific Reports 11, 1–11. https://doi.org/10.1038/s41598-021-90418-w

Palstra, A.P., Planas, J. V, 2012. Swimming Physiology of Fish: Towards using exercise to farm a fit fish in sustainable aquaculture. Springer Science & Business Media.

Peñaloza, C., Robledo, D., Barría, A., Trinh, T.Q., Mahmuddin, *et al.* 2020. Development and validation of an open access SNP array for Nile Tilapia (oreochromis niloticus). G3: Genes, Genomes, Genetics 10, 2777–2785. https://doi.org/10.1534/g3.120.401343

Uguccioni, G., D'souza, D., Hood, D.A., 2010. Regulation of PPAR coactivator-1 function and expression in muscle: effect of exercise. PPAR research 2010.

581. Optimizing genotyping effort in aquaculture breeding programs by pre-selection of candidates

C. Roozeboom*, B. Gulzari, H. Komen, M. Camara and J.W.M. Bastiaansen

Wageningen University & Research, P.O. Box 338, 6708 PB Wageningen, the Netherlands; chantal.roozeboom@wur.nl

Abstract

Highly fecund aquaculture species provide more opportunities to optimize genotyping effort in genomic selection programs through selective genotyping and pre-selection than other livestock species. The aim of this study is to optimize pre-selection in a simulated fish breeding program to achieve the highest genetic gain per genotyped fish under restricted inbreeding. We simulated breeding programs in R with pre-selection among all selection candidates, between full sib families and within full sib families and several pre-selection intensities within each of the strategies. Pre-selection always reduces genetic gain. However, when genotyping costs are taken into account, genetic gain per genotyped fish can be increased with pre-selection. Pre-selecting 40% of the candidates from all available selection candidates produced the highest additional genetic gain per genotyped fish.

Introduction

Aquaculture species are highly fecund and can produce large numbers of candidate offspring, and most fish breeding programs can produce sufficient numbers of sibs for sib testing if desired. Depending on the size of the candidate population, genotyping all available candidates can be prohibitively costly. However, genotyping costs can be reduced, for example by selectively genotyping the reference population (Gowane *et al.*, 2019), genotyping using cheaper, lower density SNP panels (Kriaridou *et al.*, 2020) or pre-selection of candidates. Pre-selection based on phenotypes measured at an early age can increase genetic gain (Martinez *et al.*, 2006), but Sae-Lim *et al.* (2013) showed that postponing pre-selection can be more effective because the genetic correlations between phenotypes measured closer to harvest and breeding goal traits are higher than for early phenotypes.

In this study, we focus on reducing genotyping costs by adding a pedigree-based pre-selection step immediately before genotyping and genomic selection. We aim to optimize this pre-selection to achieve the highest genetic gain per genotyped fish under restricted inbreeding, by simulating and comparing the predicted outcomes from a range of scenarios.

Materials & methods

To perform pedigree based preselection we simulated a family-based rainbow trout breeding program with controlled spawning, using stochastic simulations in R (Figure 1). We simulated the traits tagging weight, weight in nucleus measured on selection candidates, and weight at sea measured on sibs of the candidates in the production environment. Heritabilities and genetic correlations used (Table 1) are estimates from a real breeding program.

Table 1. Heritabilities (diagonal) and genetic correlations.

	Tagging weight	Weight nucleus	Weight sea
Tagging weight	0.26	0.38	0.29
Weight nucleus		0.33	0.64
Weight sea			0.32

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_581, © C. Roozeboom *et al.* 2022

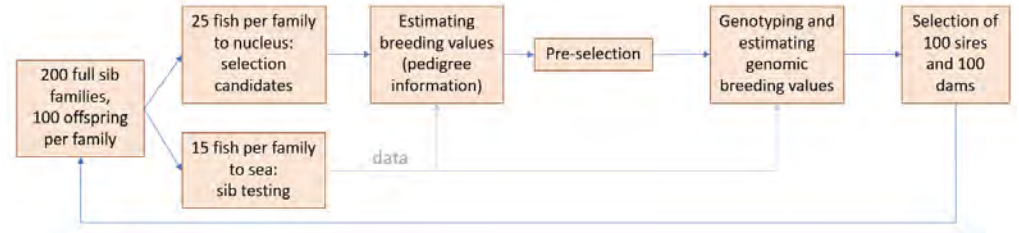


Figure 1. Design of the breeding program.

We simulated pre-selection based on a multi-trait selection index of traditional, pedigree-based estimated breeding values (EBVs) to identify fish for genotyping and final selection based on an index of genomically estimated EBVs (GEBVs) and optimal contribution selection to identify 100 replacement sires and 100 replacement dams each generation.

We simulated three pre-selection strategies: (1) the top 10, 20, 40, 60 or 80% of all selection candidates regardless of family membership (ALL); (2) the top 10, 20, 40, 60 or 80% selection candidates within each full sib family (WITHIN); or (3) the top 10, 20, 40, 60 or 80% of the full sib families (BETWEEN). The pre-selected and genotyped proportion varied between 400 and 3,200 selection candidates. The accuracy of the GEBV of weight at sea was simulated as a constant based on a reference population of 2,000 fish per generation. The accuracy of the GEBV for weight in the nucleus depended on the number of pre-selected and genotyped fish.

We simulated ten generations of selection for each scenario and calculated the average genetic gain and rate of inbreeding from 20 replicates. The additional genetic gain per genotyped fish was calculated as the difference in genetic gain with and without pre-selection divided by the number of genotyped fish.

Results

Without pre-selection, the genetic gain was 136.4 and 217.9 grams per generation for weight in the nucleus and weight at sea with genomic selection, and 106.7 and 147.2 grams per generation with pedigree selection. The rate of inbreeding was 0.73 and 0.44% without pre-selection with final selection based on EBV and GEBV, respectively.

Pre-selection always reduced genetic gain compared to genotyping all candidates (Figure 2). Genetic gain for both weight in the nucleus and weight at sea declines more rapidly with decreasing pre-selected proportion for the WITHIN strategy than for the ALL and BETWEEN strategies, except at the lowest pre-selected proportion. Regardless of the pre-selected proportion, the highest genetic gains were obtained for the ALL strategy. Pre-selected proportions smaller than 0.20 always resulted in genetic gains that were comparable or lower than gains from selection based on pedigree-based EBVs.

In the ALL and BETWEEN strategies, the rate of inbreeding increased when the pre-selected proportion decreased, but for the WITHIN strategy, the rate of inbreeding decreased. The rate of inbreeding increased most for the BETWEEN strategy to a value slightly exceeding the acceptable threshold of 1% with a 10% pre-selected proportion.

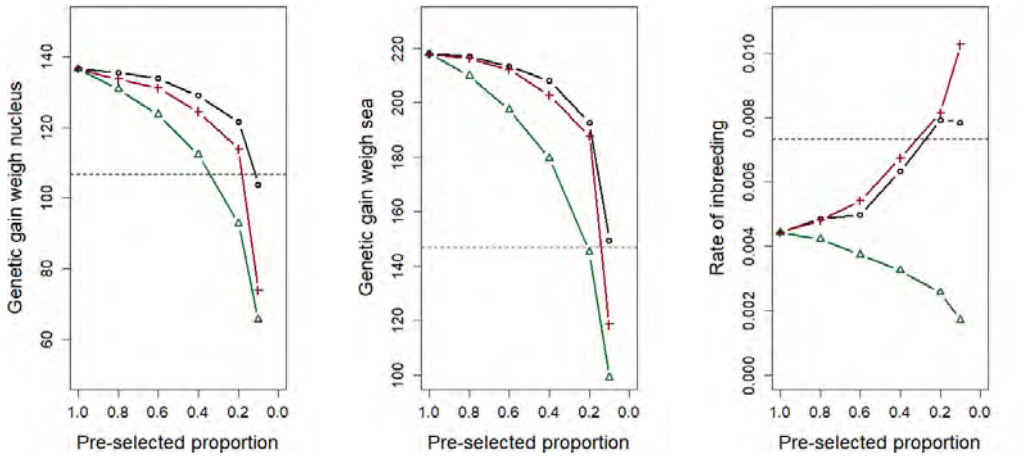


Figure 2. Genetic gain and rate of inbreeding. Genetic gain and rate of inbreeding by pre-selected proportion for the ALL (– O –), BETWEEN (– + –) and WITHIN (– Δ –) strategies. The --- reference line is the genetic gain and rate of inbreeding without genotyping.

The ALL and BETWEEN pre-selection strategies increased additional genetic gain per genotyped fish (Figure 3). For these two strategies, additional genetic gain per genotyped fish was highest at a pre-selected proportion of approximately 40%. Additional genetic gain per genotyped fish for the WITHIN strategy was maximized when nearly all candidates were genotyped (Figure 3). Additional genetic gain per genotyped fish for weight in nucleus followed a similar pattern to weight at sea (results not shown).

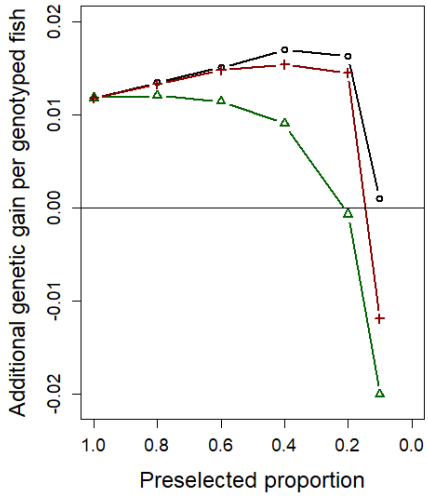


Figure 3 Additional genetic gain per genotyped fish for weight at sea. Lines are for the strategies ALL (– O –), BETWEEN (– + –) and WITHIN (– Δ –).

Discussion

Our simulations predict that pre-selection always reduces genetic gain compared to genotyping all candidates, but the relationship is highly non-linear. This non-linearity leads to an intermediate optimum in terms of genetic gain per genotyped fish. If a fixed budget is available for the breeding program, then pre-selection intensity can be adjusted to efficiently use the available budget. Applying the optimal pre-selection strategy and intensity can reduce genotyping cost while limiting the effects on genetic gain. Pre-selection at an earlier age can also limit cost for the breeding program (Sea-Lim *et al.*, 2013, Martinez *et al.*, 2006). However, this reduction is cost is due to lower rearing cost. The pre-selection in this study will not reduce rearing cost, only genotyping cost.

The choice of pre-selection strategy is important because it influences both the genetic gain and rate of inbreeding. To limit the effect of pre-selection on rate of inbreeding, optimal contributions in the pre-selection could be implemented in the ALL strategy. The rate of inbreeding was reduced when pre-selection took place within a full sib family. However, the WITHIN strategy is not recommended for pre-selection, given that the rate of genetic gain reduces more rapidly with this strategy. This rapid reduction may be because the number of families was much larger than the number of candidates per family.

Pre-selection among all selection candidates resulted in the highest genetic gain, because more fish from the best families could be selected compared to the scenarios in which pre-selection took place within a family. In the BETWEEN strategy, the number of selected families was lower than in the ALL strategy. Therefore, it prevents selecting good fish from families with lower average breeding values, especially at lower selected proportion.

Postponing pre-selection can result in higher genetic gains due to higher genetic correlations between the traits (Sae-Lim *et al.*, 2013). In this study, the pre-selection was immediately before genomic selection. However, pre-selection decreased the genetic gain, because the pre-selection was based on EBVs instead of GEBVs, which resulted in lower accuracies in the pre-selection step compared to genotyping all candidates and therefore lower genetic gains.

For a low genotyping budget, the BETWEEN strategy could be further optimized because the pre-selected proportion will be small, and final selection takes place mostly within the small number of pre-selected full sib families. This provides an opportunity to reduce the genotyping costs by using low density SNP panels and genotyping only the phenotypically most extreme fish from the training population, because within-family haplotype blocks can be accurately traced using low-density SNP panels and most of the within-family variance can be captured by genotyping only the phenotypically most extreme sibs for the training population (Ødegård and Meuwissen, 2014).

Conclusions

Our simulation results indicate that the reduction in genetic gain due to pre-selection is smallest when pre-selection takes place among all selection candidates. When genotyping cost are taken into account, higher genetic gains per genotyped fish are achieved with pre-selection. Pre-selecting approximately 40% from all selection candidates resulted in the highest additional genetic gain per genotyped fish.

Acknowledgements

This study is part of the AquaIMPACT project and was funded by the European Union's Horizon 2020 research and innovation programme under grant agreement No 818367.

References

Gowane, G. R., Lee, S. H., Clark, S., Moghaddar, N., Al-Mamun, H. A., *et al.* (2019). Journal of Animal Breeding and Genetics, 136(5), 390-407. https://doi.org/10.1111/jbg.12420

Kriaridou, C., Tsairidou, S., Houston, R. D., & Robledo, D. (2020). Frontiers in genetics, 11, 124. https://doi.org/10.3389/fgene.2020.00124

Martinez, V., Kause, A., Mäntysaari, E., & Mäki-Tanila, A. (2006). Aquaculture, 254(1-4), 195-202. https://doi.org/10.1016/j.aquaculture.2005.11.011

Ødegård, J., & Meuwissen, T. H. (2014) Genetics Selection Evolution, 46(1), 1-8. https://doi.org/10.1186/1297-9686-46-3

Sae-Lim, P., Komen, H., Kause, A., Martin, K. E., Crooijmans, R., *et al.* (2013). *Aquaculture*, *372*, 89-96. https://doi.org/10.1016/j.aquaculture.2012.10.031.

582. Prediction of production traits by using body features of gilthead seabream (*Sparus aurata*) obtained from digital images

B. Gulzari[1*], A. Mencarelli[2], C. Roozeboom[1], H. Komen[1] and J.W.M. Bastiaansen[1]

[1]*Animal Breeding and Genomics Group, Wageningen University & Research, the Netherlands;*
[2]*Wageningen Plant Research, Wageningen University & Research, the Netherlands; benan.gulzari@wur.nl*

Abstract

Gilthead seabream (*Sparus aurata*) is a key aquaculture species in the Mediterranean and surrounding regions. While many traits are of interest in seabream breeding programs, phenotypes of them cannot always be easily or accurately measured. Therefore, the objectives are to predict phenotypes of production traits by using automated measurements of body features from digital images and to obtain genetic correlations between measured and predicted phenotypes. The production traits analysed were harvest weight (HW), fillet weight (FW), and fillet percentage (F%). Each image feature was tested for prediction of phenotypes by using 10-fold cross-validation. The phenotypic and genetic correlations of measured and predicted phenotypes were 0.98 and 0.996 for HW, 0.93 and 0.99 for FW, 0.27 and 0.70 for F%, respectively. The relative efficiency of mass selection was higher for predicted phenotypes, except for HW.

Introduction

Gilthead seabream (*Sparus aurata*) is a key aquaculture species in the Mediterranean and surrounding regions. The aquacultural production of gilthead seabream increased by 90% between 2009 ande 2019 to 260,000 tons across 20 countries (FAO, 2021). However, only 60% of farmed seabream originated from commercial breeding programs in 2015 (Janssen *et al.*, 2015). In fish breeding programs, various traits are selected for to increase the production level, efficiency, and welfare of the fish. These traits are important for farmers as well as for other stakeholders. However, phenotypes of them cannot always be easily or accurately measured. Some traits, such as product quality traits, cannot be measured on live fish. Sib information can be used to select for such traits but using phenotypes of selection candidates is preferable and has three main advantages. First, both within family and between family variation can be utilized. Second, the accuracy of the estimated breeding values (EBV) will be higher. Third, using phenotypes of the selection candidates simplifies the breeding program and reduces the costs because fewer animals are needed to reach a certain accuracy of the EBVs.

Instead of actual phenotypes of difficult to measure traits, predictions of those phenotypes can be used for selection candidates. Predictions of the phenotypes from body measurements has been studied in various fish species. Manual measurements work well to predict the phenotypes of difficult to measure traits; however, manually measuring body features is labour intensive. Computer vision allows digital image acquisition and image processing for measurements of the fish by using dedicated software (Dowlati *et al.*, 2012). The measurements obtained from the images of fish can be used to establish predictions of the phenotypes. Furthermore, obtaining measurements from digital images can be automated (Navarro *et al.*, 2016).

Therefore, the objectives are to predict phenotypes of production traits by using automated measurements of body features from digital images and to obtain genetic correlations between measured and predicted phenotypes.

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_582, © B. Gulzari *et al.* 2022

Materials & methods

Phenotypic data collection. For this experiment, a population of juveniles was produced by mass spawning of 33 males and 20 females on a single day. When fish reached an average weight of 3 grams, a batch of juveniles was stocked in a sea cage in southeast of Spain and grown under commercial conditions. Data on production traits were collected from the commercially produced fish that were harvested after a grow-out period of 500 days. After measuring harvest weight (HW), the fish were filleted, and fillet weight was recorded. Fillet percentage (F%) was calculated as (FW / HW) × 100.

Each fish was photographed after being taken out of the water. Measurements of body features from digital images were automated by using the Halcon 18.11 framework (MVTec Software GmbH, Munich, Germany). The body features measured from images included total area, fillet area, trunk area, head area, tail area, total length, standard length, height, head height, head length, total volume, and head excluded volume. Head proportion was calculated as head area divided by total area.

Prediction of traits. All variables measured on images were considered individually and combined as candidate predictors for each phenotype. Prediction accuracy was defined as the correlation between measured and predicted phenotypes and tested by using 10-fold cross-validation. Cross-validation was repeated 10 times to treat all the parts of the data as validation and test sets. Predictors were added to linear regression models with forward selection. For genetic analyses, each phenotype was predicted by using a linear model with all data and the combination of predictors that yielded the highest accuracy in cross-validation.

Estimation of heritabilities and correlations. Genetic correlations (r_g) between measured and predicted phenotypes were estimated by using a bivariate animal model in ASReml version 4.1 (Gilmour *et al.*, 2015). The bivariate animal model is:

$$y = X\beta + Zu + e \quad (1)$$

where y is a vector of measured and predicted phenotypes, β is the vector of fixed effect 'sampling day', u is the vector of random animal additive genetic effects:

$$\sim \left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, G \begin{bmatrix} \sigma^2_{a,T_M} & r_{a,T_{M,P}} \sigma_{a,T_M} \sigma_{a,T_P} \\ r_{a,T_{M,P}} \sigma_{a,T_M} \sigma_{a,T_P} & \sigma^2_{a,T_P} \end{bmatrix} \right),$$

where G is the genomic relationship matrix and σ^2_{a,T_M} is the additive genetic variance of measured trait, σ^2_{a,T_P} is the additive genetic variance of predicted trait, $r_{a,T_{M,P}}$ is the additive genetic correlation between the measured and predicted traits, and e is the vector of random residual effects

In case of mass selection, the relative efficiency (RE) was calculated as:

$$RE = \left(\sqrt{h^2_P / h^2_M} \right) \; x \; r_A \quad (2)$$

where h^2_P is the heritability of predicted phenotypes, h^2_M is the heritability of measured phenotypes, and r_A is the genetic correlation between measured and predicted phenotypes.

Results

Prediction of harvest weight. The accuracy of predicting phenotypes was 0.98 for HW, and the genetic correlation between measured and predicted phenotypes was 0.996±0.001. The model that yielded the highest accuracy of prediction contained head excluded volume and total area of the fish. The heritability

of the predicted HW (0.52±0.05) was slightly lower than measured HW (0.54±0.05). The RE of indirect selection for HW was 98%.

Prediction of fillet weight and fillet percentage. The accuracy of predicting phenotypes was 0.92 and 0.27 for FW and F%, and the genetic correlation between measured and predicted phenotypes was 0.99±0.005 and 0.74±0.13. The model that yielded the highest correlation for FW contained head excluded volume and total area of the fish. The heritability of the predicted FW (0.52±0.05) was slightly higher than measured FW (0.47±0.05). The relative efficiency of indirect selection for FW was 104%. The model that yielded the highest correlation for F% contained head proportion and head length of the fish. The heritability of the predicted F% (0.31±0.05) was higher than measured F% (0.12±0.04). The RE of indirect selection for F% was 119%.

Discussion

Our findings show that predicting phenotypes of gilthead seabream from digital images has strong potential to replace manual phenotyping in breeding programs. In case of mass selection, changing to predicted phenotypes would yield 4 and 19% more response for FW and F%, respectively. For HW, using predicted phenotypes would yield only 2% less response than using measured phenotypes.

Automation of phenotyping for fillet traits with digital images is advantageous due to two interrelated reasons. First reason is that the labour required for manual phenotyping increases if the breeding goal includes more than HW, especially when traits cannot be measured on live fish and siblings need to be phenotyped. In such cases, using image features to predict the phenotypes simplifies the breeding program because separate measurements for different traits will no longer be necessary and the fish does not have to be slaughtered. With a single image of the fish, it is possible to predict phenotypes for many traits. The second reason is that phenotypes can be obtained for selection candidates. Our findings imply that using digital images to phenotype for fillet traits is not only possible but also preferred in fish breeding programs. If predicted phenotypes are used, more response would be obtained in fillet traits and the response in HW would only slightly reduce in comparison to using measured phenotypes. Prediction of phenotypes provides benefits in genomic selection schemes as well. Predicted phenotypes become available on the selection candidates and the reference population contributes images and measured phenotypes to the training dataset for the genomic selection as well as the phenotype prediction models. Furthermore, prediction of phenotypes from images could enable application of genomic selection without genotyping the reference population. Genomic EBVs can be calculated for the selection candidates with the phenotypes predicted from image features. Not genotyping the reference population will affect the accuracy of genomic selection. This trade-off between the accuracy of genomic selection and accuracy of predicting phenotypes needs to be further investigated and optimized for each genomic selection scheme.

Certain strengths and limitations were recognized in this study. The main strengths of this study are that: (1) the measurements from digital images were automated, so the measurement error is minimal; (2) a genomic relationship matrix was used to estimate the genetic parameters. The advantage of using a genomic relationship matrix is that it improves the accuracy of the parameter estimates in comparison to using pedigree relationships (Veerkamp *et al.*, 2011). This study was performed with images from a single experimental seabream population that was produced as one batch with a full sib and half sib family structure. Accuracy of predicted phenotypes may reduce across different production batches or in unrelated populations of seabream. To avoid effects of overfitting, the predictions of phenotypes were evaluated with cross-validation, which is expected to result in a model with the strongest prediction ability to other populations. Before applying these models to predict the phenotypes of other seabream populations, it will be needed to test the model accuracies obtained in those populations. Extending the current dataset with

other unrelated seabream populations may change the accuracy of the predictions. Similar to reference datasets in genomic selection, the level of prediction accuracy will need to be assessed for situations where reference phenotype and image data are from more diverse background or when they are more distant from selection candidates.

Funding

This study is part of the AquaIMPACT project and was funded by the European Union's Horizon 2020 research and innovation programme under grant agreement No 818367.

References

Dowlati, M., de la Guardia, M., Dowlati, M., Mohtasebi, S.S., 2012. Application of machine-vision techniques to fish-quality assessment. TrAC Trends Anal. Chem. 40, 168–179. https://doi.org/10.1016/J.TRAC.2012.07.011

FAO, 2021. Fishery and Aquaculture Statistics. Global aquaculture production 1950-2019 (FishstatJ).

Gilmour, A.R., Gogel, B.J., Cullis, B.R., Welham, Sj., Thompson, R., 2015. ASReml user guide release 4.1 structural specification. Hemel hempstead VSN Int. ltd.

Janssen, K., Chavanne, H., Berentsen, P., Komen, H., 2015. Gilthead Seabream (*Sparus aurata*)–Current status of selective breeding in Europe, in: Abstract Book of the International Symposium on Genetics in Aquaculture XII, Santiago de Compostela, Spain.

Navarro, A., Lee-Montero, I., Santana, D., Henríquez, P., Ferrer, M.A., Morales, A., Soula, M., Badilla, R., Negrín-Báez, D., Zamorano, M.J., Afonso, J.M., 2016. IMAFISH_ML: A fully-automated image analysis software for assessing fish morphometric traits on gilthead seabream (*Sparus aurata* L.), meagre (Argyrosomus regius) and red porgy (*Pagrus pagrus*). Comput. Electron. Agric. 121, 66–73. https://doi.org/10.1016/j.compag.2015.11.015

Veerkamp, R.F., Mulder, H.A., Thompson, R., Calus, M.P.L., 2011. Genomic and pedigree-based genetic parameters for scarcely recorded traits when some animals are genotyped. J. Dairy Sci. 94, 4189–4197. https://doi.org/10.3168/jds.2011-4223

583. Comparison of genomic selection barramundi breeding schemes for multiple environments with strong G×E

M. Lillehammer[1*], N.A. Robinson[1,2], D.B. Jones[3], C. Massault[3], S. Loughnan[4], H.S. Cate[4], P.J. Harrison[4], J.M. Strugnell[3], K.R. Zenger[3] and D.R. Jerry[3]

[1]Breeding and Genetics, Nofima, 1430 Ås, Norway; [2]Sustainable Aquaculture Laboratory-Temperate and Tropical (SALTT), School of BioSciences, the University of Melbourne, Parkville 3010, Australia; [3]Centre for Sustainable Tropical Fisheries & Aquaculture, The College of Science and Engineering, James Cook University, Townsville, Queensland 4811, Australia; [4]Mainstream Aquaculture Group Pty Ltd, Werribee, Victoria 3030, Australia; marie.lillehammer@nofima.no

Abstract

Barramundi is a popular fish for aquaculture due to strong consumer demand and ability to thrive in a range of salinities and culture systems. Genotype-by-environment interactions have been detected for growth rate among barramundi cultured in different production environments and genomic selection has been modelled to be efficient under such situations. The aim of this study was to explore the application of genomic selection for achieving genetic gain for growth rate in two diverse growing environments, recirculating brackish water tank and freshwater pond culture. Growth rate and 70K SNP genotype data was collected from a commercial scale cohort of families grown in the two environments. Estimated genomic breeding values, genetic parameters and economic data were used to compare predicted genetic gain and cost-benefit under three alternative breeding program designs. Results suggest that a single breeding program using genomic selection (general effects from an interaction model) would achieve high benefit-cost ratios.

Introduction

Barramundi is grown under a range of salinities (fresh water to fully marine) and culture systems (RAS to sea cages, Jerry, 2013). Significant genotype-by-environment interaction (G×E) has been detected for growth rate in these different environments (r_g from 0.59 to 0.94 for weight, depending on production environments compared (Domingos, *et al.*, 2013; Domingos, *et al.*, 2021; Khang, *et al.*, 2018). Genomic selection has been predicted to be effective in complex situations when strong G×E occurs (Mulder, 2016). Sukhavachana, *et al.* (2021) reported that harvest and fillet trait breeding value prediction accuracy was improved by 25% (10% for fillet yield) using genomic-informed (GBLUP), rather than traditional pedigree-based (PBLUP) prediction. The potential genetic gain from genomic selection, and the design of breeding programs servicing farms with diverse grow-out environments, needs further investigation.

The aim of this study was to compare genetic gain and benefit-cost ratios for alternative barramundi breeding program designs that service multiple environments with strong G×E.

Materials & methods

A cohort of barramundi was produced following commercial procedures at Mainstream Aquaculture Group Pty Ltd, Australia. Briefly, eight female and 16 male commercial broodstock were placed in a 30,000 l spawning tank (31 ppt salinity, 30.5 °C), cannulated to check for reproductive readiness and induced to mass-spawn using intramuscular injection of luteinizing hormone release hormone analogue (LHRHa) at a dosage of 50 µg/kg. Larviculture was conducted closely according to semi-automated commercial intensive protocols described by Schipp, *et al.* (2007). All broodstock were genotyped using a custom 70k Axiom™ myDesign™ SNP array (ThermoFisher Scientific™) (Jerry, *et al.*, 2022) and tagged with passive implant transponders (PIT). Larvae were reared to 30 days post hatch, split into groups and transferred

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_583, © M. Lillehammer *et al.* 2022

into nurseries at the grow-out sites; ponds located in Innisfail (QLD) and recirculating aquaculture tanks located in Werribee (VIC). Size grading was undertaken within both farm sites, but no culling of the smaller fish was conducted. Fish at both sites were fed a commercial pelleted diet under industry feed rates and practices until harvest (336 dph for VIC and 440 dph for QLD) at which time they were weighed.

A genomic relationship matrix (GRM), was compiled as described by VanRaden (2008), using the snpReady package in *R* (R Core Team, 2019). Estimated breeding values (EBVs) were calculated using restricted maximum likelihood (REML) in ASReml 4.0.1 (Gilmour, *et al.*, 2015) using the animal model: $y = \mu + Zu + e$, where y is the whole fish weight (WFW), μ is the overall mean for the trait, Z is the incidence matrix relating phenotypes to breeding values, u is the vector of breeding values, and e is the vector of random residual effects. The breeding values (u) were assumed to be normally distributed with variance equal to the genetic variance (V_A) and a covariance structure given by the pedigree (PBLUP) or the GRM (GBLUP). Three possible breeding program selection strategies were considered (as detailed in, Jerry, *et al.*, 2022);

1. two separate breeding programs for each farm site environment, each based on the performance of breeding candidates in that respective farm environment (*MULTIPLE*);
2. a single breeding program for both farm environments based on the general/main effect from the interaction model (*GENERAL*); and
3. a single breeding program based on the performance of breeding candidates in a single farm environment (*SINGLE*).

The expected rate of genetic gain was estimated as: $\Delta G = I * Ac * \sqrt{V_A}$, where I was the intensity of selection, Ac was the accuracy of GEBV prediction for each trait and V_A was the additive genetic variance estimated from the single trait GBLUP model for the trait within this environment, which gave the most conservative estimates of V_A.

The following parameters were assumed for the calculation of benefit-cost ratio: Phenotyping: 10,000 AU$ per cohort sampling. One cohort sampling per year for SINGLE, and two for GENERAL and MULTIPLE due to sampling in both environments. Genotyping: 40 AU$ per genotyped fish. All selection candidates were genotyped. Operational costs of 64 AU$ per fish for keeping the selection candidates and broodstock fish until phenotyping and 5,000 AU$ per broodstock fish from selection until mating. Generation interval of 3 years for males and 5 years for females. One additional kg of fish produced gives 3 AU$ extra income with 8 million fish produced per year, equally distributed between both environments. Each year 1000 selection candidates were produced and 300 broodstock selected. Profit was estimated as (income – cost), where income was calculated as the predicted additional kg fish produced and cost is the sum of phenotyping, genotyping and operational costs of each breeding program strategy.

Results

Genetic correlation of WFW between the two environments was found to be 0.54 (PBLUP) or 0.59 (GBLUP), which indicates strong G×E. Figure 1 shows the genetic gain per generation for VIC (blue bars) and QLD (orange bars). Genetic gain was higher in QLD than even in VIC, with GBLUP than with PBLUP and when the reference population was from the same environment as where gain was measured.

Figure 2 shows the accumulated profit if each of these breeding programs were run for 10 years (ignoring inflation). The most profitable strategy was GENERAL, closely followed by SINGLE QLD. The least profitable strategy was MULTIPLE. GBLUP was more profitable than PBLUP, reaching break half a year earlier, and after 10 years accumulated profit was 80% higher with GBLUP than with PBLUP.

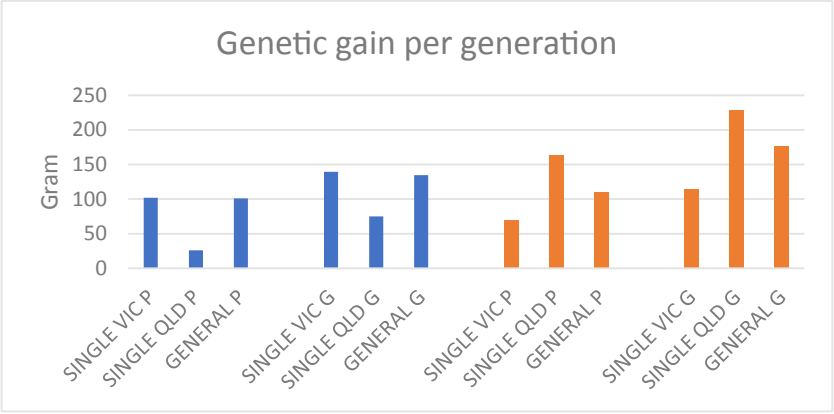


Figure 1. Genetic gain after one generation of selection in VIC (blue) and QLD (orange).

Figure 2. Accumulated profit (AU$) over 10 years for singleVIC (blue), single QLD (red), MULTIPLE (grey) and GENERAL (yellow).

Discussion

Despite the strong G×E documented in Jerry *et al*., (2022) we show that under the assumptions included here a single breeding program would be the most profitable option for barramundi,. The single breeding program could aim to improve general growth, taking phenotypes from both environments into account or to improve growth in the QLD environment, where genetic gain was highest. These two strategies would give similar profit over 10 years assuming that half the fish are produced in a VIC environment and half in a QLD environment. Although running two separate breeding programs was shown to be the least profitable strategy over 10 years, MULTIPLE would be the most profitable strategy after 14 years both for PBLUP and for GBLUP (results not shown).

This project was supported by CRC project funding from the Department of Industry, Innovation and Science, Commonwealth of Australia.

References

Domingos, J.A., Smith-Keune, C., Robinson, N., Loughnan, S., Harrison, P., Jerry, D.R., 2013. Heritability of harvest growth traits and genotype-environment interactions in barramundi, Lates calcarifer (Bloch). Aquaculture. 402, 66-75.

Domingos, J.A., Goldsbury, J.A., Gomes, G.B., Smith, B.G., Tomlinson, C., Bade, T., Sander, C., Forrester, J., Jerry, D.R., 2021. Genotype by environment interactions of harvest growth traits for barramundi (Lates calcarifer) commercially farmed in marine vs. freshwater conditions. Aquaculture. 532, 8.

Gilmour, A.R., Gogel, B.J., Cullis, B.R., Welham, S.J., Thompson, R., 2015. ASReml User Guide Release 4.1 www.vsni.co.uk

Jerry, D.R., 2013. Biology and culture of Asian seabass Lates calcarifer. CRC Press.

Jerry, D.R., Jones, D.B., Lillehammer, M., Massault, C., Loughnan, S., Cate, H.S., Harrison, P.J., Strugnell, J.M., Zenger, K.R., Robinson, N.A., 2022. Predicted strong genetic gains from the application of genomic selection to improve growth related traits in barramundi (Lates calcarifer). Aquaculture. 549, 737761.

Khang, P.V., Phuong, T.H., Dat, N.K., Knibb, W., Nguyen, N.H., 2018. An 8-Year Breeding Program for Asian Seabass Lates calcarifer: Genetic Evaluation, Experiences, and Challenges. Frontiers in Genetics. 9, 12.

Mulder, H.A., 2016. Genomic Selection Improves Response to Selection in Resilience by Exploiting Genotype by Environment Interactions. Frontiers in Genetics. 7, 11.

R Core Team, 2019. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.

Sukhavachana, S., Senanan, W., Pattarapanyawong, N., Srithong, C., Joerakate, W., Tunkijjanukij, S., Poompuang, S., 2021. Multiple-trait genomic prediction of harvest and fillet traits in Asian seabass (Lates calcarifer, Bloch 1790). Aquaculture. 544, 737069.

VanRaden, P., 2008. Efficient methods to compute genomic predictions. J Dairy Sci. 91, 4414-4423.

Part 34

Avian

584. Genome-wide association studies of chicken caecal microbiota

X. Dai[1], A.T. Hinsu[2], C. Dadousis[3], R.J. Pandit[2], M. Crotta[1], G. Limon[1], B. Fosso[4], J. Guitian[1], F.M. Tomley[1], P.G. Koringa[2], C.G. Joshi[2], D.P. Blake[1] and A. Psifidi[1*]

[1]The Royal Veterinary College, Hawkshead Lane, AL9 7TA, Hatfield, Hertfordshire, United Kingdom; [2]Department of Animal Biotechnology, College of Veterinary Science and Animal Husbandry, Anand Agricultural University, 388110, Anand, Gujarat, India; [3]Department of Veterinary Science, University of Parma, 43126 Parma, Italy; [4]Institute of Biomembranes, Bioenergetics and Molecular Biotechnologies, CNR, Bari, Italy; apsifidi@rvc.ac.uk

Abstract

The enteric microbiota has a vital role in animal health and performance. Gut microbial community structure varies widely between individuals, suggesting that in addition to environmental factors there is also a host genetic component. The aim of this study was to identify heritable host (chicken) genetic variation associated with the microbial community structure in two broiler (meat-type chicken) breeds, the indigenous Indian Kadaknath and the commercial Cobb400, raised in western parts of India under the local prevailing husbandry conditions. We collected blood and caecal lumen contents from 300 birds from each breed. All birds were genotyped with the 600K Affymetrix® Axiom® HD array and the caecal lumen microbiota was defined using DNA-metabarcoding. Variant component estimates and genome-wide association studies were conducted. Using these datasets, we identified significant heritability estimates and genomic associations at genome-wide level ($P<0.05$ Bonferroni corrected) between host genetic variation and gut microbiome composition in both chicken breeds under investigation.

Introduction

Enteric microbial communities are dynamic and complex, influencing host health and disease profoundly. It is common that microorganisms residing in the gut have established symbiotic relationships with their host (Krajmalnik-Brown *et al.*, 2012). Such microbial systems have coevolved with their host and contribute to fundamental functions such as regulation of host metabolic and immune pathways, and prevention of pathogen colonisation (Pickard *et al.*, 2017). Gut microbial composition is influenced by multiple extrinsic factors, including diet, environmental conditions (Rothschild *et al.*, 2018), and medication (Vangay *et al.*, 2015). These factors can modulate gut microbial communities leading to beneficial or detrimental changes in host physiological status. Variation within gut microbiota is also driven by intrinsic factors, such as host genetics (Zhao *et al.*, 2013). So far, there has been a limited number of studies looking at the contribution of host genetics to chicken microbiome, while the degree to which host genetics can modulate microbial composition remains largely unknown. According to our previous study chicken breed exert significant impact on caecal microbiome composition (Pandit *et al.*, 2018), suggesting that there is scope for host genetic studies to investigate the potential of selective breeding to promote protective gut microbiota. Selecting for genotypes that concourage protective enteric microbial populations can reduce dependence on antibiotic-mediated pathogen control. This study sought to investigate this possibility by studying the genomic architecture of gut microbiome composition in two chicken breeds, namely Kadaknath (an indigenous Indian breed) and Cobb400 (a commercial broiler line).

Materials & methods

Chicken populations, sample collection, and processing. Whole blood preserved on Flinders Technology Associates (FTA) cards and caecal lumen contents were collected from 600 chickens (300 Kadaknath and 300 Cobb400 of 36-40 days old) raised under the local prevailing husbandry conditions on 60 small to middle size farms located in Gujarat, India (as described by Hinsu *et al.*, 2018). Briefly, birds

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_584, © X. Dai *et al.* 2022

were reared together on 30 farms (n=5 per breed, for a total of 300 samples), or sampled directly from specialised in either Kadaknath or Cobb400 farms to increase the sample size (15 farms each breed, n=10 birds per farm, a total of 300 samples). Metadata related to bird and environmental variables were collected using a detailed questionnaire. Collection of caecal samples, total genomic bacterial DNA extraction, 16S rRNA gene amplification and 16S rDNA sequencing using the MiSeq platform were performed as described previously (Pandit *et al.*, 2018). Chicken genomic DNA was extracted from FTA cards using the Gensolve DNA recovery kit (GenTegra, Pleasnton, USA) and all chickens were genotyped using the 600K SNP Affymetrix high-density array (Kranis *et al.*, 2013).

Data and statistical analysis. 16S rDNA-sequencing data were analysed with QIIME 2. In particular, raw paired-end reads were denoised (i.e. amplification and sequencing noise removal), merged and chimera-checked by inferring the amplicon sequence variants (ASVs) with DADA2. The obtained ASVs table was normalised by rarefaction to an equal sequencing depth of 10,000 sequences per sample. All amplicon sequence variants (ASVs) were classified into the bacterial taxonomy using QIIME 2 and the SILVA database v.132. Low abundance bacterial genera (present in <5 samples and relative abundance <0.01%) were filtered out. Alpha (Inverse Simpson index) and beta diversity (weighted UniFrac and Bray-Curtis dissimilarity metrics) indices were calculated using the R *Phyloseq* package.

Alpha diversity, the first ten beta diversity Principal Coordinates Analysis (PCoA) axes and individual genera relative abundance were used as phenotypes for genome-wide association studies (GWAS). Data were analysed separately for each breed. Rare genera detected in <30% of samples were excluded from subsequent analyses. Individual genera present in <60%, but ³30% of samples were analysed as binary traits. Individual genera detected in ³60% of samples were analysed as continuous traits after normalisation using the rank-based inverse normal transformation of *GenABEL* package in R. The same normalisation function was used for alpha diversity. Multiple correspondence analyses (MCA) of bird and environmental farm variables were performed to identify key components explaining the phenotypic variance in recorded metadata using *FactoMineR* package in R.

Genotypic quality control was performed separately for each breed using PLINK v1.9 (Purcell *et al.*, 2007) with the parameters: sample call rate >90%, single nucleotide polymorphism (SNP) call rate >90%, minor allele frequency >5% and Hardy-Weinberg equilibrium P-value<10^{-6}. After these steps, 465,002 and 400,048 SNPs, 267 and 292 birds, respectively for Kadaknath and Cobb400 remained for subsequent analyses. Variance components to estimate the heritability for each trait of interest (alpha- and beta-diversity, individual genera abundance) per breed was estimated using GCTA v1.94 (Yang *et al.*, 2011). A univariate linear-mixed model was used which included the first four MCA components of farm practice categorical variables as fixed effects and the genomic relationship matrix among-individuals as a random effect. GWAS analyses was performed using the same model as the one used for the heritability estimation on GEMMA v0.98 (Zhou and Stephens, 2012). After Bonferroni correction, the P-value<0.05 genome-wide significance threshold was set at 1.1×10^{-6} and 1.2×10^{-6} for Kadaknath and Cobb400, respectively. Those SNPs that were significantly associated with microbial traits were remapped to the latest reference chicken genome assembly (GalGal6) using the NCBI remapping tool.

Results & discussion

This study was set out to evaluate the role of host genetics in shaping gut microbiota in two distinct chicken breeds. The alpha and beta diversity indices did not reveal any clear separation of the caecal microbiota between the two breeds (Figures 1A and 1B).These results could be partly attributed to the fact that the two breeds were raised in the same geographical region under similar management conditions and are consistent with previous reports showing that environmental factors have dominant roles in shaping gut

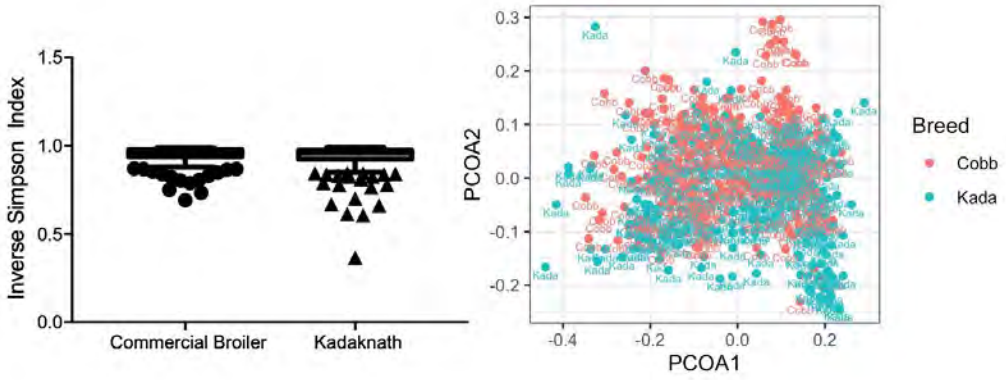


Figure 1. Characterisation of caecal microbiota sampled from Kadaknath and Cobb400 birds based upon 16S rDNA amplicon sequencing. A. Inverse Simpson index. B. Principal Coordinates Analysis (PCOA) plot based on Bray-Curtis distance.

microbiota (Rothschild *et al.*, 2018). The first four MCA components of the recorded metadata accounted for 71.2 and 75.3% of all the farm variability for Kadaknath and Cobb400, respectively, and therefore were included in the GWAS as fixed effects to reduce environmental noise.

We identified considerable heritable genetic variation for beta diversity as well as for individual microbial genera, some potentially pathogenic, in both breeds. Heritability estimates for beta diversity ranged from 0.01 to 0.76 in Kadaknath, and from 0 to 0.72 in Cobb400. In total, 47 and 25 SNPs were significantly associated at the genome wide level with gut microbiota structure in Kadaknath and Cobb400 birds, respectively. The relevant abundance of 39 and 35 individual microbial genera were analysed as binary and continuous traits, respectively. We identified several significant SNP at genome wide levels for the abundance of these genera in the caecal microbiota. Examples of genera of interest (i.e. those with pathogenic potential) included *Campylobacter, Ruminococcus, Alistipes, Helicobacter, Faecalibacterium*, and *Bacteroides* in Kadaknath (Figure 2A-F), and *Campylobacter, Enterococcus* and *Alistipes* in Cobb400 (Figure 2G-H). These potentially pathogenic bacteria could, directly and/or indirectly, influence the host and its enteric microbiota. Previous studies have shown that inflammatory immune responses induced by pathogens could influence the intestinal environment and its bacterial communities (Vangay *et al.*, 2015; Pickard *et al.*, 2017). Therefore, being able to influence their abundance through a selective breeding program would be of interest for productive performance as well as public health.

In conclusion, in this study we found that the microbial community structure and relative abundance of individual genera within chicken caecal microbiota are heritable polygenic traits. Our results corroborate the results of previous studies and support the possibility of breeding to establish enteric microbial populations in chickens that promote resistance against pathogen colonisation or improve performance. Further studies to unravel what constitutes a protective or productive microbiota are needed to facilitate future breeding programmes.

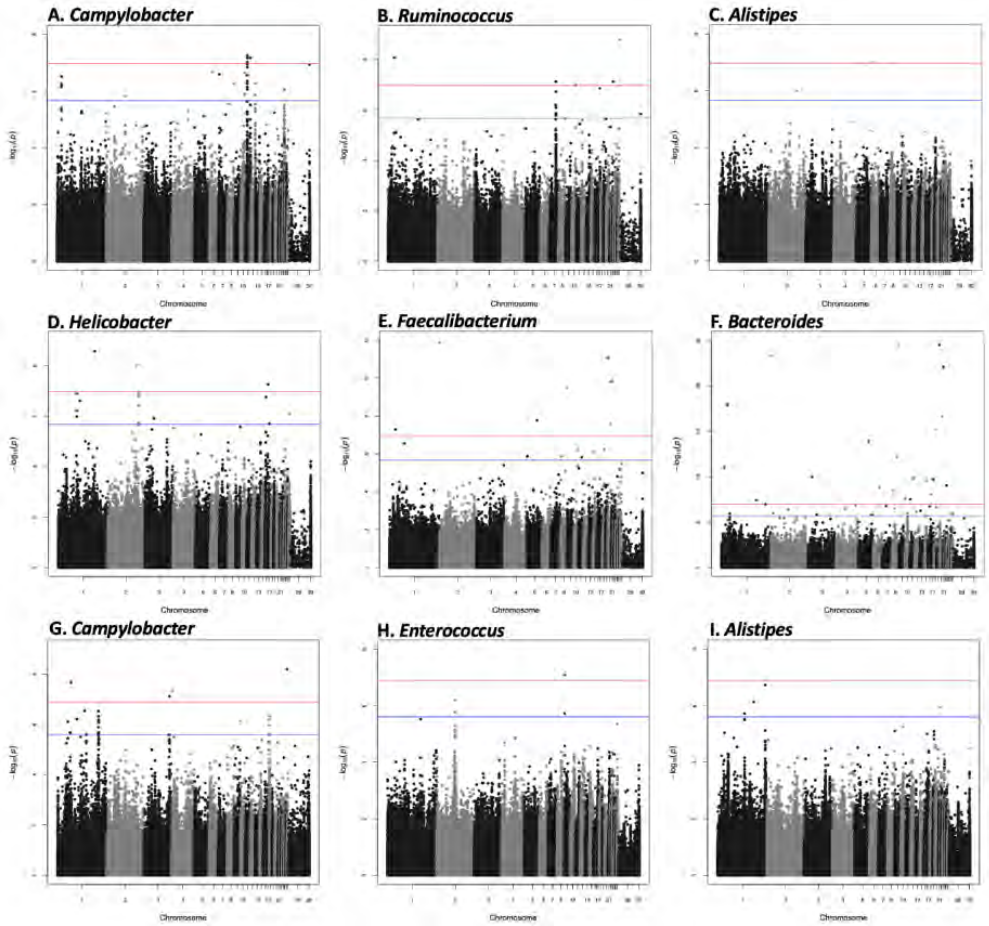


Figure 2. Manhattan plots of genome-wide association studies of relative abundance of potentially pathogenic genera in the caeca microbiota of Kadaknath (A-F) and Cobb400 (G-I) chicken breeds; in the x-axis is chromosome number; y-axis is -$\log_{10}$(*P*-value); red and blue horizontal lines show the genome-wide significant and suggestive significance (1 false discovery per genome scan) levels, respectively.

References

Hinsu A.T., Thakkar J.R., Koringa P.G., Vrba V., Jakhesara S.J. *et al.* (2018) Front. Vet. Sci (5).
Krajmalnik-Brown R.,Ilhan Z.E.,Kang D.W.,DiBaise. J.K.(2012). Nutr.Clin.Pract (27):201-214.
Kranis A., Gheyas A.A., Boschiero C., Turner F., Yu L. *et al.* (2013) BMC Genom (14):59
Pandit R.J., Hinsu A.T., Patel N. V., Koringa P.G., Jakhesara S.J. *et al.* (2018) Microbiome (6):115.
Pickard J.M., Zeng M.Y., Caruso R., and Núñez G. (2017) Immunol. Rev (279):70–89.
Psifidi A., Crotta M., Pandit. RJ., Fosso B., Koringa P.G. *et al.* (2018) Proc. of the 11th WCGALP, Auckland, New Zealand
Purcell S., Neale B., Todd-Brown K., Thomas L., Ferreira M.A.R. *et al.* (2007). Am. J. Hum. Genet.81:559–575.
Rothschild D., Weissbrod O., Barkan E., Kurilshikov A., Korem T. *et al.*(2018) Nature (555):210-215.
Vangay P., Ward T., Gerber J.S., and Knights D. (2015) Cell Host Microbe (17):553–564.
Yang J., Lee S.H., Goddard M.E., and Visscher P.M. (2011). Am J Hum Genet 88:76–82.
Zhao L., Wang G., Siegel P., He C., Wang H. *et al.* (2013) Sci. Rep (3):1163.
Zhou X. and Stephens M. (2012) Nat Genet. 44:821–4.

585. Breeding for an optimised gut microbiota composition for the improvement of phosphorus utilisation in Japanese quail

V. Haas*, S. Vollmar, A. Camarinha-Silva and J. Bennewitz

Institute of Animal Science, University of Hohenheim, Garbenstraße 17, 70599 Stuttgart, Germany;
valentin.haas@uni-hohenheim.de

Abstract

Phosphorus utilisation (PU) and correlated performance traits are typical quantitative traits with many genes with small effects. Through adaption of the microbes in the gastro intestinal tract (GIT) a better digestion of phytic acids bound P in the feed could be achieved. The classical breeding value g_i can be decomposed into a heritable contribution p_i that results from altered metabolic pathways and a heritable contribution m_i which results from altered digestion in the animal from the effect of the microbiota composition in the GIT, so $g_i = m_i + p_i$. Instead of selecting for the breeding value $\hat{g}_i$, it would be beneficial to split the breeding value for one trait into $\hat{m}_i$ and $\hat{p}_i$, by using an index. This enables a stronger selection pressure on the microbiota-mediated path, which is seen as beneficial from an animal welfare point of view. This study describes the index and its application to PU in Japanese quail.

Introduction

Efficiency traits are a major factor influencing the environmental balance in livestock production. They consist of two general components, the digestive efficiency and metabolic efficiency (Martin *et al.* 2021, Puillet *et al.* 2016). Digestive efficiency reflects the ability to absorb nutrients from the GIT into the bloodstream, whereas metabolic efficiency refers to the allocation and reallocation processes of the absorbed nutrients to the animal products. It was shown that efficiency traits are mainly conditioned by reallocation effects due to breeding. But, these reallocation effects can in turn negatively influence animal health (Huber 2018) and thus also animal welfare.

PU and its correlated traits are polygenetic traits that are influenced by many genes with small effects and have low to medium heritability (Beck *et al.* 2016; Vollmar *et al.* 2021). Increased native digestion of phytic acids P (primary form of storage in plant seeds) is targeted to reduce supplementation of mineral P in the feed and thereby reduce environmental impact. Native phytic acids degradation in the GIT can occur by phytases and phosphatases derived from the endogenous mucosa of the GIT, from some vegetable feed components, or from the gut microbiota (Rodehutscord, 2017).

The acquisition (i.e. digestion) pathway can be achieved through a proper community of the microbiota in the GIT. This can be confirmed by microbiabilites found for PU and other performance and efficiency traits (Vollmar *et al.* 2020). These microbiota influencing genes affect the phenotype not only directly by changing the metabolic pathways, but also indirectly by altering the microbiota composition of the animals GIT. The aim of this study was to review the hologenomic selection index proposed in an earlier study (Weishaar *et al.* 2019) and investigate the feasibility for the important trait PU in the model organism Japanese quail for commercial poultry.

Materials & methods

Experimental design and data collection. We used an F2 design of Japanese quail (Coturnix japonica) from two divergent lines from a previous study (Beck *et al.* 2016). The traits of this experiment were PU, body weight gain (BWG), feed intake (FI) and feed per gain (F:G), collected from the 10th to 15th day of life. The quail were slaughtered on day 15 to collect the ileum samples. All birds were incubated and slaughtered

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_585, © V. Haas *et al.* 2022

on twelve different days treated as test days in the statistical analysis (Beck *et al.* 2016). The quail were 4k SNP genotyped and the ileum microbiota composition were performed by target amplicon sequencing (Borda-Molina *et al.* 2020; Vollmar *et al.* 2021). Microbiota sequences were clustered into operational taxonomic units (OTUs) at >97% similarity. The final data set contained 758 quail with SNP genotypes, microbiota composition of 1188 OTUs, and trait records (PU, BWG, FI, and F:G).

Statistical analyses. The genomic breeding values $\hat{g}_i$ for the trait were obtained with the genomic mixed model:

$$\mathbf{y} = \mu\,\mathbf{1} + \mathbf{Z_{td}td} + \mathbf{g} + \mathbf{e} \quad (1)$$

where y is the vector of observations (PU, FI, BWG or F:G) for *n* animals, μ the trait mean and 1 the vector of ones. The vector $\mathbf{td} \sim N(0, \mathbf{I}\sigma^2_{td})$ represents the random test day effects with variance σ^2_{td} and design matrix Z_{td}. Vector $\mathbf{g} \sim N(0, \mathbf{G}\sigma^2_a)$ contains the random animal effects with the genomic relationship matrix G, using SNP genotypes following Method 1 of VanRaden (2008), with additive genetic variance σ^2_a. The random residuals are stored in the vector $\mathbf{e} \sim N(0, \mathbf{I}\sigma^2_e)$.

The estimates $\hat{k}_i$ of the animal microbiota effects k_i on the phenotype were obtained using a microbial mixed model, which is similar to model (1), but with vector k instead of vector g:

$$\mathbf{y} = \mu\,\mathbf{1} + \mathbf{Z_{td}td} + \mathbf{Ik} + \mathbf{e} \quad (2)$$

vector k includes the random microbiota effect of all the animals with distribution $\mathbf{k} \sim N(0, \mathbf{M}\sigma^2_k)$, with microbial relationship matrix M (Camarinha-Silva *et al.* 2017) and microbial variance σ^2_k.

Subsequently, we used the estimated microbial animal effect $\hat{\mathbf{k}}$ for the traits from model (2) as observations in the following genomic model (3) to predict the genomic breeding value $\hat{m}_i$ for the microbiota-mediated effects m_i:

$$\hat{\mathbf{k}} = \mu\,\mathbf{1} + \mathbf{Im} + \mathbf{e} \quad (3)$$

with μ the overall mean, vector $\mathbf{m} \sim N(0, \mathbf{G}\sigma^2_m)$ the random animal effect with the genomic covariance matrix G, the SNP variance σ^2_m of the microbiota effect and e the residuals with variance σ^2_e. Models (1-3) were fitted using the package ASReml, which is available in R (Butler *et al.* 2009).

Since $\hat{m}_i$ estimates the microbiota effect m_i on the breeding value g_i, the estimates can be used to calculate the index:

$$I_i = \lambda_1 \hat{m}_i + \lambda_2 (\hat{g}_i - \hat{m}_i)$$

Microbial and genomic predictions were performed with repeated fivefold cross-validation, with a randomly selected reference population of 80% of the animals at each repetition (500 in total) for estimating OTU and/or SNP effects. The remaining 20% were used as the validation population from which the animal effects were predicted. The averaged Pearson correlations between the predicted animal effect and the observed animal phenotype were used as the accuracy of prediction. Based on these correlations, the confidence intervals were calculated.

We applied three kinds of predictions. In G-BLUP, we used model (1), in M-BLUP model (2), and for G-BLUP of the microbiota-mediated part of the trait, we used model (2) to obtain the random microbiota effect of the animals for the corresponding trait, which was subsequently used as an observation in the model (3).

Results

The results of the cross-validation of microbial and genomic predictions are presented in Table 1. The correlations $r_{\hat{g},y}$ between the genomic predictions $\hat{\mathbf{g}}$ and the observations of the traits y are at a similar level as the correlations $r_{\hat{m},y}$ between the genomic predictions of the microbiota-mediated part of the traits $\hat{\mathbf{m}}$ and the observations. The correlations $r_{\hat{k},y}$ between the microbial predictions $\hat{\mathbf{k}}$ and the trait observations y are slightly higher for PU and FI and markedly larger for BWG and F:G.

Discussion

The results from the microbial and genomic predictions (Table 1) confirmed the strong impact of the host genome and the microbiota composition on the traits. Furthermore, the two-step procedure proposed by Weishaar *et al.* (2020) to estimate a breeding value for the microbiota-mediated part of a trait was also successful in this study and it was most evident for PU and FI. The G-BLUP prediction accuracy for the microbiota-mediated part of the trait was only slightly lower than the prediction accuracy of the conventional G-BLUP. An explanation might be that the host genetic effect on the trait mediated by the microbiota is much stronger than the direct host path. The prediction methods have been further developed recently by Christensen *et al.* (2021).

The results of the two-step method proposed by Weishaar *et al.* (2020) clearly show the potential for predicting genomic breeding values for microbiota effects m_i on quantitative traits in Japanese quail. It is only possible to breed for the microbiota effect m_i of a trait, if the trait is heritable and influenced by the composition of the microbiota in the GIT. The selection response for the microbial effect m_i depends on the additive variance of the trait, the position in the GIT where the microbial samples were taken, the proportion of the additive variance that can be explained by the microbial effect m_i and the accuracy with which the OTU effects and the SNP effects were estimated (Weishaar *et al.* 2020).

PU is a typical trait for a potential breeding approach based on the GIT microbiota, i.e. for selecting the hologenome to achieve breeding success. It is known that poultry has a low endogenous phytase activity, but native phytic acid degradation in the GIT can occur by phytases and phosphatases from the gut microbiota. The proposed method enables to breed for an improved PU by optimising the GIT microbiota composition. Although it is well-known that the microbial colonisation of the GIT is influenced by the environment, there is a host genetic control that manifested in significant heritabilities of several microbial traits and this can be used for breeding, as proposed here.

Table 1. Results of the genomic and microbial trait predictions.[1]

Trait[2]	M-BLUP of y		G-BLUP of y		G-BLUP of microbiota-mediated y	
	$r_{\hat{k},y}$	95% CI	$r_{\hat{g},y}$	95% CI	$r_{\hat{m},y}$	95% CI
PU	0.221	0.091:0.351	0.181	0.048:0.320	0.162	0.006:0.313
FI	0.308	0.171:0.427	0.235	0.099:0.353	0.222	0.074:0.376
BWG	0.338	0.201:0.464	0.131	-0.010:0.254	0.137	-0.030:0.294
F:G	0.307	0.100:0.470	0.097	-0.052:0.226	0.073	-0.069:0.225

[1] Estimated accuracy of the M-BLUPs ($r_{\hat{k},y}$) and G-BLUPs ($r_{\hat{g},y}$) of the trait observations and G-BLUPs of the microbiota-mediated part of the trait observations ($r_{\hat{m},y}$) with confidence intervals (CI).

[2] PU = P utilisation, FI = feed intake, BWG = body weight gain and F:G = feed per gain.

If the amount of excess P in livestock production and thus the use of mineral P is to be reduced, this can be done by looking at the microbiota composition in the GIT using the hologenomic approach.

Funding

This research was funded by the Deutsche Forschungsgemeinschaft and was part of the research unit P-FOWL (FOR 2601).

References

Beck P., Piepho H.-P., Rodehutscord M. and Bennewitz J. (2016) Poult Sci 95(4):764–773. https://doi.org/10.3382/ps/pev376

Borda-Molina D., Roth C., Hérnandez-Arriaga A., Rissi D., *et al.* (2020) Animals 10(5):885. https://doi.org/10.3390/ani10050885

Butler D., Cullis B.R., Gilmour A.R. and Gogel B.J. (2009) ASReml-R reference manual: mixed models for S language environments. QDPIF, Brisbane, AUS.

Camarinha-Silva A., Maushammer M., Wellmann R., Vital M., *et al.* (2017) Genetics 206(3):1637–1644. https://doi.org/10.1534/genetics.117.200782

Christensen O.F., Börner V., Varona L. and Legarra A (2021) Genetics 219:iyab130. https://doi.org/10.1093/genetics/iyab167

Huber K. (2018) Animal 12(3):528–536. https://doi.org/10.1017/S1751731117002051

Martin P., Ducrocq V., Faverdin P. and Friggens N.C. (2021) J Dairy Sci 104(6):6329–6342. https://doi.org/10.3168/jds.2020-19844

Puillet L., Réale D. and Friggens N.C. (2016) GSE 48(1):72. https://doi.org/10.1186/s12711-016-0251-8

Rodehutscord M. (2017) Advances in understanding the role of phytate in phosphorus and calcium nutrition of poultry. In: Applegate T. (eds) Achieving sustainable production of poultry meat, Volume 2: Breeding and nutrition. Burleigh Dodds Sci Publ, Cambridge, UK, pp. 165–80.

VanRaden P.M. (2008) J Dairy Sci 91(11):4414–4423. https://doi.org/10.3168/jds.2007-0980

Vollmar S., Haas V., Schmid M., Preuβ S., *et al.* (2021) Anim Genet 52(1):90–98. https://doi.org/10.1111/age.13018

Vollmar S., Wellmann R., Borda-Molina D., Rodehutscord M., *et al.* (2020) G3 (Bethesda) 10(7):2553–2562. https://doi.org/10.1534/g3.120.401424

Weishaar R., Wellmann R., Camarinha-Silva A., Rodehutscord M. and Bennewitz J. (2020) J Anim Breed Genet 137(1):14–22. https://doi.org/10.1111/jbg.12447

586. High-resolution genotyping and analysis of *in silico* predicted proteins for classical MHC genes in Korean native chicken

T.K. Ediriweera[1], P. Manjula[2], E. Cho[1], M. Kim[1], S. Cho[1] and J.H. Lee[1*]

[1]Chungnam National University, 99 Daehak-ro, Yuseong-gu, 34134, Daejeon, Republic of Korea; [2]Uva Wellassa University, Passara Road, 90000, Badulla, Sri Lanka; junheon@cnu.ac.kr

Abstract

The major histocompatibility complex-*B* (MHC-*B*) region is of paramount importance in chicken immunogenesis. The Grey line, one of the five lines belong to the Korean native chicken (KNC) breed, was analysed for the MHC-*B* region by comparing with the *Gallus gallus* (Red junglefowl) reference with an especial focus on classical MHC class I and II genes (*BF-BL* genes) and their resultant proteins. This study concluded that the MHC-*B*, including classical class I and II genes and resultant proteins in the KNC Grey line, vary from the *G. gallus* (Red junglefowl) MHC-*B* remarkably by analysing the Next-Generation Sequencing (NGS) assembly along with *BF-BL* genes and their predicted proteins.

Introduction

The major histocompatibility complex-*B* (MHC-*B*) region possesses classical MHC class I and class II genes that are extremely important in recognizing foreign pathogenic peptides. They initiate the antigen presentation by producing class I and class II MHC molecules which are highly important for immune response and histocompatibility in chickens (Kaufman, 2018). The capability of identifying such foreign materials depends on the structures of peptide-binding grooves of MHC molecules (Koch *et al.*, 2007), and those structures are determined by the coding sequences (CDSs) of respective genes. Therefore, having a complete understanding of the resulting protein structures corresponding to their DNA sequence is immensely important. In consequence, obtaining the DNA sequences, identifying genes and CDSs, translating them into possible amino acid sequences, and predicting their features, need to be conducted to comprehend the importance and involvement of MHC molecules in the process of immunogenicity.

Therefore, this study focused on the MHC-*B* region of the KNC-Grey (KNC-G) line, an indigenous chicken line with superior meat quality in Korea (Nawarathne *et al.*, 2020) that currently demonstrates a trend towards commercial farming. The objectives of this study are to obtain the consensus DNA sequence from the *BG1-BF2* gene, including *BF1*, *BF2*, *BLB1*, *BLB2* genes of the KNC-G line, investigate the variants and features of *in silico* predicted class I & class II MHC proteins. Data from this study serve as base information for further studies, especially in the development of vaccines for key avian diseases (e.g. Avian influenza, Marek's disease, Newcastle disease, etc.). This will be useful in disease management, which is a critical aspect of concern in commercial farming.

Materials & methods

At the previous stage of this study, the MHC-*B* region of the KNC-G line was analysed using MHC-*B* linked microsatellite markers, and MHC-*B* based single nucleotide polymorphisms (SNP) panel having 90 SNPs. Thus, a homozygous individual for them was identified, followed by sampling for its blood and DNA extraction. Later, the extracted DNA was amplified for the MHC-*B* (from *BG1* to *BF2* gene) including *BF1*, *BF2* genes of MHC class I and *BLB1*, *BLB2* genes of MHC class II, using long-range polymerase chain reaction (PCR) techniques. After the PCR purification step, the sequencing was conducted using Illumina Next-Generation Sequencing (NGS).

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_586, © T.K. Ediriweera *et al.* 2022

The NGS reads were then processed and analysed using 'Geneious Prime molecular biology and Sequence analysis tools' [Geneious Prime 2020.0.1. (https://www.geneious.com), Biomatters, Ltd., Auckland, New Zealand]. Accordingly, the pre-processed NGS reads were assembled using the 'mapped to reference' technique, using '*Gallus gallus* genes, MHC region, partial and complete cds' sequence available in NCBI (Accession number: AB268588.1) as the reference. Then the consensus sequence was obtained, followed by variant analysis. Before mapping, the reference sequence was annotated directly with NCBI using Geneious Prime's GenBank accession tools. Accordingly, the genes and CDSs of the obtained consensus for the KNC-G line were recognized when it is accurately mapped to the reference.

The CDSs of MHC class I and II genes (*BF-BL* genes) of both reference and KNC-G line were then obtained and translated into amino acid sequences followed by prediction of antigenic regions, and transmembrane regions using EMBOSS protein analysis tools and transmembrane prediction plugins of Geneious Prime.

Results

The reference mapping-based assembly of the KNC-G line for its MHC-*B* (from *BG1* to *BF2* gene) region comprised 67,740 bp with 59,676 sequences in which the pairwise identity percentage was 80.9%. The GC content was 59.8% whereas the individual A, C, G, T contents were 20.2, 28.8, 31.0, and 20.0%, respectively. Further, there were 830 SNPs (576 transition SNPs and 254 transversion SNPs), of which 230 of them were in gene regions. In addition to SNPs, it contained 38 insertions, 40 deletions, and 65 substitutions.

However, addressing the prime focus, the DNA sequences for *BF-BL* genes were extracted from the main consensus and further analysed for their CDSs, amino acid sequences, properties, and features. Each time the considered genes of the KNC-G line have been compared with the reference. The results for predicted physical properties and features of obtained amino acid sequences based on their CDSs are presented in Table 1 and Figure 1.

Subfigures a, b, c, d of Figure 1 represent the predicted protein comparisons for *BF1*, *BF2*, *BLB1* and *BLB2* genes of the KNC-G line respectively. In each subfigure, the topmost 'identity' row indicates, the protein identity between predicted proteins of the particular KNC-G gene and its analogous reference gene. In here, the green colour means identical, and the white means non-identical. The blue lines and light-yellow lines demonstrate genes and CDSs, respectively. The antigenic regions, transmembrane regions, and potential extracellular regions for each protein have been predicted and marked with dark-red, dark-yellow, and light-pink colours, respectively. The composite colour bar remains beneath the illustrations of transmembrane and extracellular regions in reference ('Ref') side of each of the subfigures represent different amino acids of predicted protein with different colours. The bar with less dense colours immediately below this

Table 1. The properties of predicted proteins for major histocompatibility complex-*B* class I and II genes.

Gene that encodes the protein	Whether KNC/ Ref	Length (aa)	Molecular weight (kDa)	Isoelectric point	Charge at pH 7
BF1	KNC-G_BF1	355	39.201	5.27	-8.68
	Ref_BF1	360	39.713	6.77	-0.68
BF2	KNC-G_BF2	356	37.987	11.44	27.73
	Ref_BF2	356	37.851	11.53	30.56
BLB1	KNC-G_BLB1	264	26.677	11.77	20.23
	Ref_BLB1	264	26.946	11.72	20.19
BLB2	KNC-G_BLB2	263	29.294	6.12	-1.94
	Ref_BLB2	263	29.088	6.5	-0.94

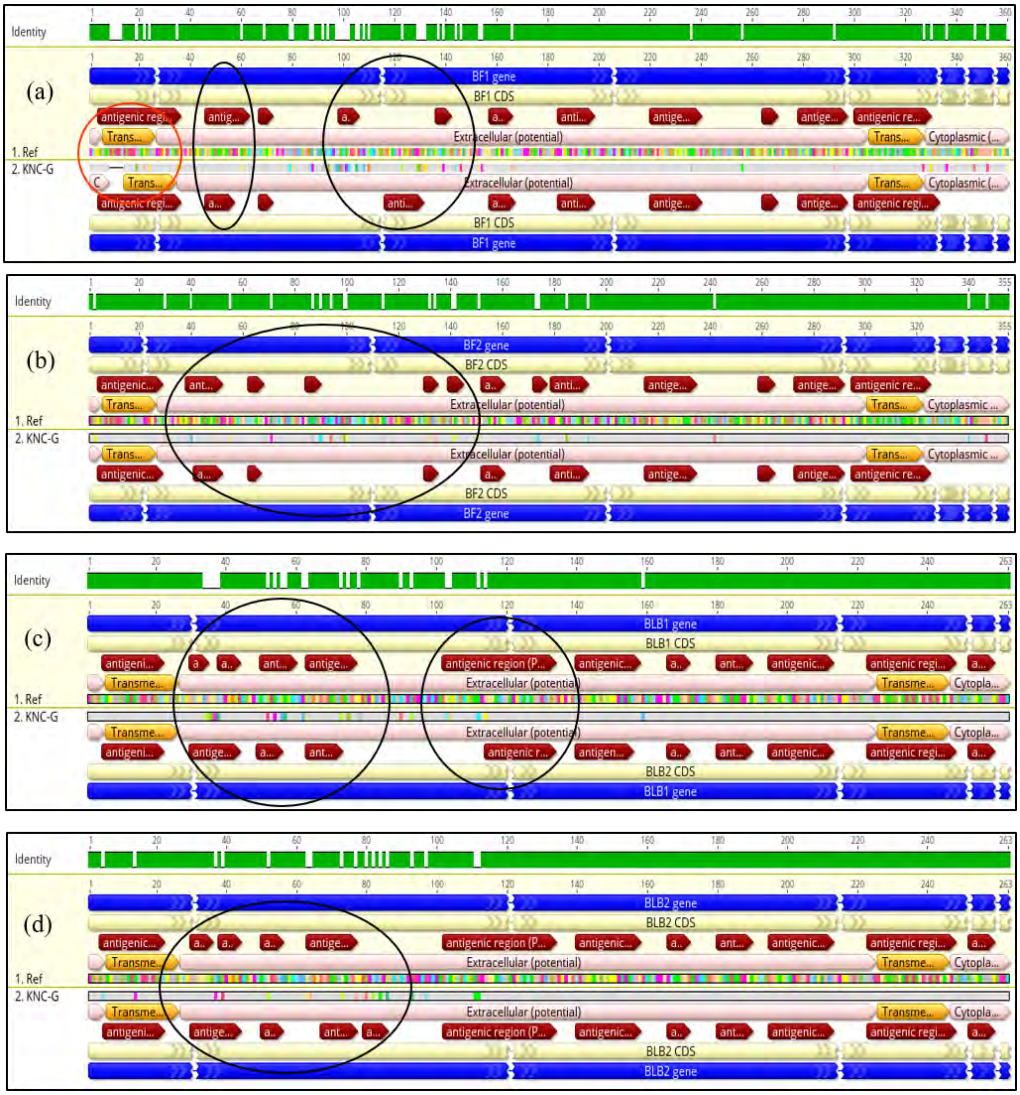


Figure 1. The features of predicted proteins for major histocompatibility complex-*B* class I and II genes.

composite colour bar indicates the amino acid disagreements of KNC-G based proteins compared to the reference. At last, the black and light-red circles highlight the prominent alterations in antigenic regions and transmembrane regions between the Reference and KNC-G based proteins respectively.

Discussion

The pairwise identity percentage of 80.9% between the KNC-G line and the reference gives a clear instinct that the MHC-*B* region has not completely conserved between the *Gallus gallus* (Red junglefowl) and the KNC-G line throughout their evolutionary process as they demonstrate a profound dissimilarity of about 19%. The high gene density in this MHC-*B* might be the reason to have a high GC content of 59.8% (Rao *et al.*, 2013).

Further, the high rate of polymorphisms suggests a high diversity within the MHC-*B* region. Especially the SNPs in gene regions reflect changes in resultant proteins. Hence, we were highly motivated to investigate the classical class I and II genes for their possible proteins.

Among all the four genes analysed for the changes in their properties and features, only *BF1* has demonstrated a deviation in amino acid length compared to the reference, and hence it can be considered a comparatively major effect. Accordingly, the other properties of *BF1*, including molecular weight, isoelectric point, and charge at pH 7, as well have noticeably diverged from the reference. In all the other three genes, despite the lengths being similar with the reference, the molecular weight, isoelectric point, and charge at pH 7 have been changed, due to amino acid variations. These changes resulting in an altered level of disease resistance/ susceptibility suggest a considerable difference in both MHC class I and II molecules.

Accordingly, they all were then analysed for antigenic regions and transmembrane regions. As demonstrated in Figure 1, black and light-red colour circles highlight the prominent alterations in antigenic regions and transmembrane regions respectively, that occurred due to the variations in their genes. These feature changes ultimately influence the competencies of both MHC class I and II molecules in pathogenic recognitions. Nevertheless, further studies need to be conducted to investigate their positive/ negative effect on infectious diseases.

Al-Shuhaib *et al.* (2019) also has employed a comparatively similar type of *in silico* approach to predict the effect of polymorphisms in resulting protein structures. Hence, this kind of *in silico* approach can be considered, fair and efficient to analyse the features of predicted proteins. However, the outcomes still need to be validated with *in vitro* experiments.

In conclusion, the MHC-*B* region of the KNC-G line considerably varies from the *Gallus gallus* (Red junglefowl) MHC-*B* region. As a result of the variations in the MHC class I and II genes, the MHC class I and II molecules are getting modified, thus possibly affecting their pathogenic identification abilities.

References

Al-Shuhaib M.B.S., Al-Kafajy F.R., and Al-Jashami G.S. (2019) Anim. Biotechnol. 32(3):1-9. https://doi.org/10.1080/10495398.2019.1683568

Kaufman J. (2018) Trends Immunol. 39(5):367-379. https://doi.org/10.1016/j.it.2018.01.001

Koch M., Camp S., Collen T., Avila D., *et al.* (2007) Immunity 27(6):885-899. https://doi.org/10.1016/j.immuni.2007.11.007

Nawarathne S.R., Lee S.K., Cho H.M., Wickramasuriya S.S., *et al.* (2020) Korean J. Agric. Sci. 47(2):255-262. https://doi.org/10.7744/kjoas.20200016

Rao Y.S., Chai X.W., Wang Z.F., Nie Q.H. and Zhang X.Q. (2013) Genet. Sel. Evol. 45(1):1-7. https://doi.org/10.1186/1297-9686-45-9

587. Genetic architecture of body weight and carcass traits in Ghanaian local chickens

R.A. Kanlisi[1], E.N. Amuzu-Aweh[1], M. Walugembe[2], A. Naazie[1], H.R. Otsyina[3], N. Chouicha[4], T.R. Kelly[5], R. Gallardo[5], J. Dekkers[2], S. Lamont[2], H. Zhou[4] and B.B. Kayang[1*]

[1]Department of Animal Science, University of Ghana, Legon, Accra, Ghana; [2]Department of Animal Science, Iowa State University, Ames, IA, USA; [3]School of Veterinary Medicine, University of Ghana, Legon, Accra, Ghana; [4]Department of Animal Science, University of California, Davis, CA, USA; [5]School of Veterinary Medicine, University of California, Davis, 95616, USA; bbkayang@ug.edu.gh

Abstract

Little information on the genetic architecture of production traits of indigenous African chicken exists. We estimated genetic parameters and performed a genome-wide association study using imputed 600K SNP genotypes on three Ghanaian chicken ecotypes (n=1,113). Variance components and heritabilities for body weight at 22 weeks (BW22), average daily gain (ADG), and carcass traits were estimated, as well as genetic and phenotypic correlations among them. Heritabilities for all traits were high (0.51-0.69), while those for carcass traits corrected for BW22 were moderate (0.21-0.38). Seven 1-Mb windows from six chromosomes each explained up to 25% of the genetic variance in these traits. These windows contained genes such as *FOXO1*, *NCARP*, *LCORL*, *LAP3*, *LDB2*, *TGFBR2*, and *TAPT1*, previously reported to have effects on growth and carcass traits. The moderate to high heritabilities and highly positive genetic correlations between them indicate that these traits can be improved through selective breeding.

Introduction

Indigenous African chickens have over the years adapted very well to harsh climatic and environmental conditions (Birteeb *et al.*, 2016). They have several desirable characteristics, including natural resistance against diseases like Newcastle disease (Mpenda *et al.*, 2018). They, however, tend to have lower growth rates and body sizes compared to exotic chicken breeds (Osei Amponsah et.al., 2011; Birteeb et.al., 2016). Therefore, in the quest to increase local chicken's body size and growth rate, many farmers introduce birds of unknown genetic composition into their flocks, potentially leading to the loss of important traits. Some studies on genetic production trait parameters of local chicken in Africa have been carried out (Osei-Amponsah *et al.*, 2013; Dekkers *et al.*, 2017), but very limited comprehensive genome-wide association studies (GWAS) to explore the genetic basis of these traits are available. Therefore, it is important to explore the genetic architecture of production traits in local chicken populations to identify their potential for genetic improvement, whilst preserving their genetic diversity and retaining their ability to withstand the harsh environmental conditions of their habitats. The objective of this study was, therefore, to estimate genetic parameters and explore the genetic architecture of body weight and carcass traits of three local chicken ecotypes in Ghana using a genome-wide association study.

Materials & methods

The chickens used in this study were offspring of those described in Walugembe *et al.* (2020). This study was conducted in four replicates, with a total of 1,113 chickens made up of the Forest, Interior Savanna and Coastal Savanna chicken ecotypes. Chickens were housed in deep litter pens and fed on a standard chick starter mash from day 1 to week 8 and on a standard chick grower mash from week 9 to week 22. Water was provided *ad libitum*. Vaccination, feeding, and general husbandry practices were the same for all the birds.

Body weights were taken bi-weekly from hatch until 22 weeks of age. Average daily gain (ADG) was calculated as the linear regression of body weight on age. At week 23, birds were humanely sacrificed, and

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_587, © R.A. Kanlisi *et al.* 2022

several carcass traits were measured, including weights of breast, drumstick, thigh, wings, overall dressed weight and dressed percentage.

Blood samples were collected from chicks at 5 weeks of age and DNA was isolated for genotyping by sequencing. The resulting genome sequences were run through a customised in-house SNP-pipeline which resulted in 5K SNP genotypes for each bird. These were then imputed to 382,240 SNPs using post-quality control high-density genotype data of relatives, as described by Walugembe *et al.* (2020). Imputation was performed using Fimpute (Sargolzaei *et al.,* 2014). The three local ecotypes studied here were derived from three ancestral populations (Walugembe *et al.*, 2020). Therefore, to infer the proportions of ancestral subpopulations for all chickens, we performed admixture analyses on the imputed genotypes using the Admixture software (Alexander *et al.,* 2009). The resulting ancestral subpopulation proportions were used as covariates in the downstream genetic analyses.

Variance components and heritabilities were estimated using a univariate model with the fixed effects of replicate, sex, and pen by replicate, ancestral subpopulation proportions as covariates, and random effects of animal genetics and a residual. The model (Model 1) is given by: $y = Xb + Z_a a + e$, where y is the vector of phenotypes; X is the incidence matrix relating fixed effects and covariates to vector y; b is the vector of fixed effects and covariates; Z_a is the incidence matrix relating the phenotypes to the vector of random bird genetic effects, *a*, with a genomic relationship matrix computed using method 1 of Van Raden (2008); and *e* is the vector of random residuals. Some carcass traits were also analyzed by including body weight at week 22 as a fixed covariate (referred to as Model 2). To estimate genetic and phenotypic correlations, we fit pairwise bivariate models with the same effects as in the univariate models. All models were implemented in ASReml 4 (Gilmour *et al.*, 2015).

We performed a GWAS using Bayes-B (Meuwissen *et al.*, 2001), as implemented in the JWAS package (Cheng *et al.*, 2018) to estimate the genetic variance explained by 1-Mb windows of SNPs across the genome. Both Models 1 and 2 were used. One-Mb windows that explained more than 1% of the genetic variance for a trait were considered significant. We used the Genome Data Viewer in NCBI with the *Gallus gallus* 6 genome build to identify candidate genes within the significant 1-Mb windows.

Results & discussion

Means and estimates of heritabilities, genetic and phenotypic correlations of the studied traits are in Table 1. The effects of ancestral subpopulation proportions were not significant for any of the traits. Replicate and sex were significant for all traits except for breast weight adjusted for BW22. When BW22 was included as a covariate (Model 2), it was significant for all traits.

Estimates of heritabilities based on Model 1 were high for all traits, ranging from 0.51 for thigh weight to 0.69 for ADG (Table 1), in line with the findings of Rance *et al.* (2002). Estimates of carcass trait heritabilities adjusted for BW22 were moderate and ranged from 0.21 for thigh weight to 0.38 for wing weight, and were similar to those reported by Gaya *et al.* (2012). All heritability estimates reduced when BW22 was included as a covariate, but the magnitude and direction of the phenotypic and genetic correlations did not change. Estimates of genetic and phenotypic correlations were all positive with moderate to high heritabilities. These findings agree with those previously reported by Zarehdaran *et al.* (2004) but not with the findings of Osei-Amponsah (2010), who reported weak to moderate negative correlations of live weight with most carcass traits at 24 weeks of age, except for carcass weight, for which the correlation was high and positive. The positive genetic correlations of body weight with carcass traits suggest that the carcass traits will improve with selection on body weight.

Table 1. Means and estimates of heritabilities (along diagonal) and of genetic (above diagonal) and phenotypic (below diagonal) correlations. In each cell, an estimate without BW22 covariate is on top and an estimate with covariate is beneath it.[1]

	BW22	ADG	Dressed wt	Dressing %	Breast	Thigh	Wing	Drumstick
Trait mean	1,179.2	7.9	798.2	70.3	194.1	128.4	97.3	116.3
BW22	**0.58 (0.07)**	0.96 (0.01)	0.95 (0.02)	0.24 (0.17)	0.86 (0.04)	0.92 (0.03)	0.88 (0.03)	0.88 (0.03)
		0.96 (0.01)	0.95 (0.02)	0.24 (0.17)	0.78 (0.06)	0.90 (0.04)	0.85 (0.04)	0.84 (0.05)
ADG	0.90 (0.006)	**0.69 (0.07)**	0.91 (0.02)	0.22 (0.16)	0.79 (0.04)	0.86 (0.04)	0.83 (0.04)	0.82 (0.04)
	0.91 (0.06)		0.81 (0.05)	0.28 (0.16)	0.60 (0.09)	0.77 (0.06)	0.70 (0.07)	0.69 (0.07)
Dressed wt	0.87 (0.0008)	0.82 (0.01)	**0.60 (0.07)**	0.53 (0.13)	0.94 (0.02)	0.95 (0.02)	0.93 (0.02)	0.91 (0.02)
	0.87 (0.008)	0.65 (0.02)	**0.30 (0.07)**	0.82 (0.09)	0.77 (0.08)	0.75 (0.1)	0.71 (0.09)	0.61 (0.11)
Dressing %	0.004 (0.03)	0.06 (0.03)	0.41 (0.03)	**0.15 (0.06)**	0.61 (0.13)	0.47 (0.16)	0.44 (0.15)	0.38 (0.16)
	0.004 (0.03)	0.09 (0.04)	0.69 (0.02)		0.75 (0.11)	0.51 (0.2)	0.45 (0.17)	0.28 (0.2)
Breast	0.72 (0.02)	0.67 (0.02)	0.87 (0.00)	0.43 (0.03)	**0.52 (0.07)**	0.84 (0.0.5)	0.86 (0.04)	0.79 (0.05)
	0.60 (0.03)	0.41 (0.03)	0.74 (0.01)	0.56 (0.02)	**0.29 (0.07)**	0.39 (0.17)	0.48 (0.13)	0.22 (0.18)
Thigh	0.78 (0.01)	0.73 (0.02)	0.82 (0.01)	0.27 (0.03)	0.65 (0.02)	**0.51 (0.07)**	0.89 (0.04)	0.94 (0.03)
	0.71 (0.02)	0.59 (0.03)	0.58 (0.02)	0.34 (0.02)	0.30 (0.03)	**0.21 (0.07)**	0.58 (0.14)	0.76 (0.12)
Wing	0.78 (0.01)	0.73 (0.02)	0.87 (0.01)	0.30 (0.03)	0.73 (0.02)	0.73 (0.02)	**0.62 (0.07)**	0.94 (0.03)
	0.72 (0.02)	0.53 (0.03)	0.67 (0.02)	0.40 (0.03)	0.45 (0.03)	0.43 (0.03)	**0.38 (0.08)**	0.75 (0.09)
Drumstick	0.82 (0.01)	0.71 (0.02)	0.85 (0.01)	0.35 (0.03)	0.68 (0.02)	0.72 (0.02)	0.80 (0.01)	**0.55 (0.07)**
	0.65 (0.02)	0.53 (0.03)	0.67 (0.02)	0.46 (0.03)	0.35 (0.03)	0.42 (0.03)	0.57 (0.02)	**0.29 (0.07)**

[1] wt: weight; all weights are given in grams.

After quality control, data on 1,113 birds and 382,240 SNPs were used for the GWAS analysis. When using Model 1, a 75 Mb region on chromosome 4 was significant for all traits except dressing %. This region explained 19.7, 11.6, 12.2, 20.6, 25.2, 19.6, and 14.0% of the genetic variance of BW22, ADG, breast weight, drumstick weight, wing weight, thigh weight and dressed weight, respectively. This region contains the *NCARP, LCORL, LAP 3, LDB2* and *TAPT1* candidate genes. The *NCAPG* gene has been reported to be associated with body size, ADG, carcass traits, and muscle development in beef cattle (Lindholm-Perry *et al.*, 2013). Polymorphisms in *LCORL* have been reported to be associated with growth traits in cattle (Lindholm-Perry *et al.*, 2013), while *LDB2* has been associated with body weight and ADG in chicken (Gu *et al.*, 2011). A 170-172 region on chromosome 1 explained 12.49, 10.6, 6.18, 3.11, 11.91, 11.12, and 14.04% of the genetic variance of BW22, ADG, dressed weight, breast weight, drumstick weight, thigh weight, and wing weight, respectively. This region includes the F*oxO1* gene, a transcription factor that plays a critical role in muscle growth, metabolism, and cell differentiation, and has been shown to be involved in regulating muscle fibre type specification (Yuan et.al., 2011). A 19 Mb window on chromosome 2 was found to explain 1.3% of the genetic variance of BW22. This region contains the *TGFBR2* gene, which, together with TGFB-2 proteins, seems to play a role in cell proliferation and differentiation, apoptosis, regulation of muscle tissues, body fat, and bone development. When BW22 was added to the model as a covariate (Model 2), there was a reduction in genetic variance for the carcass traits. Only 6.6% of the genetic variance of breast weight was now explained by three 1-Mb windows on three chromosomes. For drumstick weight, three 1-Mb windows on three chromosomes together explained only 4.4% of the genetic variance. One 1-Mb window explained 1% of the genetic variance of thigh weight, and four windows on four chromosomes together explained 10.2% of the genetic variance of wing weight.

Conclusions

The moderate to high heritabilities obtained for the studied traits and the positively high genetic correlations between them, coupled with the several 1-Mb windows that explained >1% of the genetic variance indicate that these traits can be improved in Ghanaian indigenous chicken populations through selective breeding.

Acknowledgements

This study was made possible by the generous support of the American people through the United States Agency for International Development (USAID) Feed the Future Innovation Lab for Genomics to Improve Poultry (cooperative agreement number AID-OAA-A-13-00080). The contents are the responsibility of the Feed the Future Innovation Lab for Genomics to Improve Poultry and do not necessarily reflect the views of USAID or the United States Government.

References

Alexander, D.H., Novembre, J., and Lange, K. (2009) Genome Res. 19:1655-1664. https://doi.org/10.1101/gr.094052.109

Birteeb, P.T., Essuman, A.K., and Adzitey, F. (2016). J. World's Poult. 6(3):153-160.

Cheng, H., Fernando, R.L., and Garrick, D.J. (2018). Proc. of WCGALP, 11.859. Auckland, New Zealand.

Dekkers, J., Botchway, P.K., Amuzu-Aweh, E.N., Naazie, A., Aning, K.G., *et al.* (2017) Proc. of WCGALP, 11.812

Gaya, L.D.G., Mourão, G.B., Ferraz, J.B.S., Mattos, E.C.D., Costa, A.M.M.A.D., *et al.* (2011) http://dx.doi.org/10.1590/S0103-90162011000600002

Gilmour, A.R., Gogel, B.J., Cullis, B.R., Welham, S.J., and Thompson, R. (2015) Hemel Hempstead: VSN Int. Ltd, 1–30.

Gu, X., Feng, C., Ma, L., Song, C., Wang, Y., *et al.* (2011) doi.org/10.1371/journal.pone.0021872

Lindholm-Perry, A.K., Kuehn, L.A., Oliver, W.T., Sexten, A.K., Miles, J.R., *et al.* (2013) https://doi.org/10.1371/journal.pone.0080882

Meuwissen T, Hayes B, Goddard M. (2001) Genetics 157:1819–1829.

Mpenda, F.N., Schilling, M.A., Campbell, Z., Mngumi, E.B. and Buza, J. (2018) J. Appl. Poult. Res. https://doi.org/10.3382/japr/pfy063

Osei-Amponsah, R., Kayang, B.B., and Naazie, A. (2013) https://doi.org/10.1017/S2078633613000271

Osei-Amponsah, R., Kayang, B., Naazie, A., Arthur, P., Barchia, I. (2011) https://doi.org/10.1007/s11250-011-9825-1.

Osei-Amponsah, R. (2010) PhD Thesis. Department of Animal Science, University of Ghana, Legon.

Rance, K. A., McEntee, G. M., and McDevitt, R. M. (2002) https://doi.org/10.1080/00071660220000044426

Sargolzaei, M., Chesnais, J.P. and Schenkel, F.S. (2014) BMC Genomics 15:478 https://doi.org/10.1186/1471-2164-15-478.

VanRaden, P.M. (2008) J. Dairy Sci. 91:4414–4423. https://doi.org/10.3168/jds.2007-0980

Walugembe, M., Amuzu-Aweh, E.N., Botchway P, K., Naazie, A., Aning, G., *et al.* (2020) https://doi.org/10.3389/fgene.2020.00739

Yuan, Y., Shi, X.E., Liu, Y.G., and Yang, G.S. (2011) doi.org/10.1007/s11010-010-0640-1

Zerehdaran, S.A.L.J., Vereijken, A.J., Van Arendonk, J.A.M., and Van der Waaijt, E.H. (2004) http://dx.doi.org/10.1093/ps/83.4.521

588. Estimating environmental direct-maternal covariance in real data – illustration on broilers

H. Romé[1*], T.T. Chu[1,2], D. Marois[3], C.H. Huang[3], P. Madsen[1] and J. Jensen[1]

[1]Aarhus University, Blichers Allé 20, 8830 Tjele, Denmark; [2]Faculty of Animal Science, Vietnam National University of Agriculture, Gia Lam, Hanoi, Vietnam; [3]Cobb-Vantress, P.O. Box 1030, 72761-1030 Siloam Springs, AR, USA; helene.rome@mbg.au.dk

Abstract

At early age performance of an individual might be influenced by maternal effects of both genetic and environmental origin. Because of this, a covariance structure arises between the direct and the maternal components. Estimating direct-maternal genetic covariance is easy. However, estimating direct-maternal environmental covariance is more complicated. This corresponds to estimating a covariance between residual environmental effect of the dam and her effect on her offspring. The aim of our paper was to illustrate a method to estimate easily direct-maternal environmental covariance using standard available breeding software. For this, we focused on body weight in two commercial lines of broilers, which is known to be influenced by maternal effects. Our model was implemented in DMU-software by augmenting the model with a dummy residual with a variance that was constrained to a fixed value. In both lines investigated, we found a negative direct-maternal genetic covariance and a positive direct-maternal environmental covariance.

Introduction

The phenotype of an individual (P_i) is determined by different factors such as the individual direct genetic effect (G_i) and individual environmental effect (E_i). This is represented in the famous equation $P = G + E$. At young age, the phenotype of an individual (P_i) could be also influenced by maternal effects such as maternal direct genetic effect (G_m) and maternal environmental effect (E_m). In such case, the phenotype of individual is then equivalent to the sum of individual effects and maternal effects. Thus, now we have $P_i = G_i + E_i + G_m + E_m$. This corresponds to the maternal animal model (MAM) developed by Willham (1963). In such situation, covariance structure arises between the individual and the maternal components. Therefore, one might assume that $cov(G_i,G_m)$ and $cov(E_i,E_m)$ are non-null.

Numerous studies have investigate the genetic direct-maternal covariance in different species (Cantet *et al.*, 1988; Koerhuis and Thompson, 1997; Meyer, 1992; Ogawa *et al.*, 2021; Robinson, 1996) ignoring the environmental covariance, e.g. assumed zero. Very few have focused on the environmental direct-maternal covariance (Bijma, 2006; Munilla and Cantet, 2015). This could be due to the ease of estimating genetic direct-maternal covariance using current breeding software, but also by the fact that it could be easier to understand and visualize the meaning of such covariance. Bijma (2006) explained nicely the environmental covariance structures that could arise depending on the number of batches of offspring and the number of offspring per batches per dam. He showed that when there are multiple offspring in multiple batches (and therefore raised in different environments), it is enough to estimate only two environment maternal effect; the maternal permanent environmental effect and the residual effect of the dam; and the correlation between them. Chicken is a good model for this, since it is possible to obtain multiple batches of progeny, each of them containing multiple progenies. Moreover, we know that traits as body weight are influenced by maternal effects (Chu *et al.*, 2020; Koerhuis and Thompson, 1997), especially at an early age. Munilla and Cantet (2015) proposed a method to estimate the environmental direct-maternal covariance, but because of heavy computation, its application to large dataset might be difficult. The aim of this paper is to propose a simple way to estimate the environmental direct-maternal covariance using REML and to estimate the

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_588, © H. Romé *et al.* 2022

genetic direct-maternal covariance while considering the environmental direct maternal covariance at the same time.

Materials & methods

Populations. Our analysis focused on two commercial purebred broiler lines (A and B). For both lines, body weight (BW) at similar age were recorded (in grams) for all birds tested in 40 rounds of testing and selection. For Line A, BW was recorded on 179,897 females and 178,299 males. For Line B, BW was recorded on 366,680 females and 357,241 males. BW within sexes was considered as different traits because of known sexual dimorphism in broilers affecting body weight averages and variances Except for dams in the early selection rounds, all dams had their phenotypes (BW) included in the data. In the pedigree, there were 407,474 birds for Line A, and 828,299 birds for Line B.

Model. Body weight in males and in females were considered as different traits based on observation done on pre-analysis. Thus, we built a bivariate models, where subscript F and M stand for the sex, including the fixed effects (incidence matrix X, vector b), an additive genetic effect (incidence matrix Z, vector a), a maternal genetic effect (incidence matrix W vector m), a residual (incidence matrix B, vector e), a permanent maternal environmental effect (incidence matrix W vector pe) and an extra residual ε with constant variance set equal to c^2, where c^2 is a constant. The constant could be any value superior to 0 and lower than the residual variance, in our case, the constant was set to 1. Therefore, our model was as follow:

$$\begin{bmatrix} \boldsymbol{BW_F} \\ \boldsymbol{BW_M} \end{bmatrix} = \begin{bmatrix} \boldsymbol{X_F} & 0 \\ 0 & \boldsymbol{X_M} \end{bmatrix} \begin{bmatrix} \boldsymbol{b_F} \\ \boldsymbol{b_M} \end{bmatrix} + \begin{bmatrix} \boldsymbol{Z_F} & 0 & \boldsymbol{W_F} & 0 \\ 0 & \boldsymbol{Z_M} & 0 & \boldsymbol{W_M} \end{bmatrix} \begin{bmatrix} \boldsymbol{a_F} \\ \boldsymbol{a_M} \\ \boldsymbol{m_F} \\ \boldsymbol{m_M} \end{bmatrix} + \begin{bmatrix} \boldsymbol{B_F} & 0 & \boldsymbol{W_F} & 0 \\ \boldsymbol{0} & \boldsymbol{B_M} & 0 & \boldsymbol{W_M} \end{bmatrix} \begin{bmatrix} \boldsymbol{e_F} \\ \boldsymbol{e_M} \\ \boldsymbol{pe_F} \\ \boldsymbol{pe_M} \end{bmatrix} + \begin{bmatrix} \boldsymbol{\varepsilon_F} \\ \boldsymbol{\varepsilon_M} \end{bmatrix} \quad (1)$$

$$\begin{bmatrix} \boldsymbol{a_F} \\ \boldsymbol{a_M} \\ \boldsymbol{m_F} \\ \boldsymbol{m_M} \end{bmatrix} \sim MVN \left[0, \boldsymbol{A} \otimes \begin{pmatrix} \sigma^2_{a_F} & \sigma_{a_F a_M} & \sigma_{a_F m_F} & \sigma_{a_F m_M} \\ \sigma_{a_F a_F} & \sigma^2_{a_M} & \sigma_{a_M m_F} & \sigma_{a_M m_M} \\ \sigma_{a_F m_F} & \sigma_{a_M m_F} & \sigma^2_{m_F} & \sigma_{m_F m_M} \\ \sigma_{a_F m_M} & \sigma_{a_M m_M} & \sigma_{m_F m_M} & \sigma^2_{m_M} \end{pmatrix} \right]$$

$$\begin{bmatrix} \boldsymbol{e_F} \\ \boldsymbol{e_M} \\ \boldsymbol{pe_F} \\ \boldsymbol{pe_M} \end{bmatrix} \sim MVN \left[0, \boldsymbol{I} \otimes \begin{pmatrix} \sigma^2_{e_F} & 0 & \sigma_{e_F pe_F} & \sigma_{e_F pe_M} \\ 0 & \sigma^2_{eM} & 0 & 0 \\ \sigma_{e_F pe_F} & 0 & \sigma^2_{pe_F} & \sigma_{pe_F pe_M} \\ \sigma_{e_F pe_M} & 0 & \sigma_{pe_F pe_M} & \sigma^2_{peM} \end{pmatrix} \right] \quad (2)$$

$\begin{bmatrix} \boldsymbol{\varepsilon_F} \\ \boldsymbol{\varepsilon_M} \end{bmatrix} \sim N(0, \boldsymbol{I}c^2)$, where c^2 were constrained to 1.

Where A is the pedigree relationship matrix and I is an identity matrix.

Variance components were estimated using the DMU software package (Madsen and Jensen, 2013). Setting the covariance between the additive genetic effect and the maternal genetic effect is quite simple, as we can just assign these two components to the same covariance matrix. Same goes with the covariance between the permanent maternal environmental effect and the environmental effect. However, in most of the software to estimate variance components, the residual matrix is set automatically as the last matrix. Finally, the estimate of residual variance are $\sigma^2_{e_F} + c^2$ and $\sigma^2_{e_m} + c^2$.

Results

(Co)variance components estimated in both lines are presented in Table 1. We were able to estimate both genetic and environmental direct-maternal correlation. Phenotypic variances were higher in males than in females for both lines. Heritability estimates for body weight were slightly higher in Line A than in Line B. In both lines, we observed a negative genetic direct-maternal correlation and a positive environmental

Table 1. Variance components estimated for the 2 lines.

		σ^2_a	σ_{am}	σ^2_m	σ^2_e	σ^2_{epe}	σ^2_{pe}	r_{am}	r_{epe}	h^2
Line A	f	4,952 (340)	-795 (178)	1,016 (146)	14,772 (179)	1,048 (130)	727 (68)	-0.35 (0.06)	0.32 (0.04)	0.23 (0.01)
	m	8,170 (568)	-1,436 (292)	1,608 (237)	23,523 (297)	1,232 (164)	1,254 (107)	-0.40 (0.06)	0.29 (0.04)	0.24 (0.01)
Line B	f	5,300 (271)	-1,187 (145)	1,005 (114)	19,578 (145)	1,511 (108)	1,365 (57)	-0.51 (0.04)	0.29 (0.02)	0.19 (0.01)
	m	8,309 (429)	-1,465 (211)	1,244 (163)	29,763 (228)	1,641 (135)	1,983 (88)	-0.46 (0.04)	0.26 (0.02)	0.20 (0.01)

direct-maternal correlation. Whereas environmental direct-maternal correlation were similar between the two lines (r_{epe}: 0.29~0.32), genetic direct-maternal correlation more extreme in line B (r_{am}: -0.46~-0.51) than in Line A (r_{am}: -0.35~-0.40).

Discussion

Negative direct-maternal covariance have been reported in many different cases but have often been received with scepticism (Cantet *et al.*, 1988; Koerhuis and Thompson, 1997; Meyer, 1997; Robinson, 1996). It was believed that this negative covariance was due to some negative environmental covariance that were ignored in the model (Falconer and Geerts, 1965; Robinson, 1996). Here, we were able to estimate both genetic and environmental direct-maternal covariance simultaneously. Whereas, the genetic correlation was strongly negative, the environmental one was positive. The negative genetic correlation could be easily explained by the resource allocation theory (Beilharz *et al.*, 1993). Indeed, one can imagine that a genotype selected on its ability to grow, will used most of its resources to this function and not to reproduction functions, especially if food input were limited. Therefore, egg contents maybe not optimal for embryonic development resulting in smaller chicks. The environmental direct-maternal correlation can be seen from different, partly overlapping, viewpoints. First, we can hypothesize that there is a correlation between the environment of the dam (considered as an individual) and the permanent effect on the group of offspring. In such case, a dam having a favourable environment for growth will influence positively the performances of its offspring by producing for example bigger eggs, resulting in bigger chicks at hatch (Abiola *et al.*, 2008; Pinchasov, 1991). As a second viewpoint, we can look at the environmental effect as the residual effect, which is really is, and so assume that this correlation is due to effects which are ignored in the model. We can also hypothesize that dam and offspring might share a common environment, which will be reflected in the covariance. For example, microbiome is heritable and is correlated to body weight (Meng *et al.*, 2014). As said before, the different point of view are not exclusives, and so the observed environmental direct-maternal covariance could be due to different factors. Adding extra layers of information (microbiome, non-additive genetic effects, etc.) into our model might help to better dissect the environmental direct-maternal covariance.

Munilla and Cantet (2015) proposed another approach that was described as computationally intense, and so that could be complicated to apply to large dataset. Here we presented a simple way to estimate the environmental direct-maternal covariance using REML. We applied our method to large datasets without increasing computation time or memory compared to models where the direct-maternal environmental covariance were ignored.

Furthermore, it would be interesting to investigate the effect of accounting for both genetic and environmental direct-maternal covariance on accuracy of prediction and on inflation of breeding values.

Conclusions

It is possible to estimate the environmental direct-maternal (i.e. dam-offspring) covariance using REML. In our case, the negative genetic direct-maternal covariance was associated with a positive environmental direct-maternal covariance.

References

Abiola, S. S., Meshioye, O. O., Oyerinde, B. O., & Bamgbose, M. A. (2008). *Archivos de Zootecnia*, *57*(217), 83–86.

Beilharz, R. G., Luxford, B. G., & Wilkinson, J. L. (1993). JABG *110*(1–6), 161–170. https://doi.org/10.1111/j.1439-0388.1993.tb00728.x

Bijma, P. (2006). JAS *84*(4), 800–806. https://doi.org/10.2527/2006.844800x

Cantet, R. J. C., Kress, D. D., Anderson, D. C., Doornbos, D. E., Burfening, P. J., *et al.* (1988). JAS *66*(3), 648–660. https://doi.org/10.2527/jas1988.663648x

Chu, T. T., Madsen, P., Norberg, E., Wang, L., Marois, D., *et al.* (2020). JABG 137(2), 245-259. https://doi.org/10.1111/jbg.12448

Falconer, D. S., & Geerts, S. J. (1965). *Genetics Today, Proceedings of the XIth International Congress on Genetics*. Pergamon, Oxford.

Koerhuis, A., & Thompson, R. (1997). *GSE*, *29*(2), 225–249. https://doi.org/10.1186/1297-9686-29-2-225

Madsen, P., Jensen, J., 2013. A user's guide to DMU. A package for analysing multivariate mixed models version 6 (dmuv6_guide.5.2.pdf (au.dk))

Meng, H., Zhang, Y., Zhao, L., Zhao, W., He, C. et al (2014). *PLOS ONE*, *9*(3), e89862. https://doi.org/10.1371/journal.pone.0089862

Meyer, K. (1992). *Livestock Production Science*, *31*(3), 179–204. https://doi.org/10.1016/0301-6226(92)90017-X

Munilla, S., & Cantet, R. J. C. (2015). JAS, *93*(6), 2669–2677. https://doi.org/10.2527/jas.2014-8512

Ogawa, S., Yazaki, N., Ohnishi, C., Ishii, K., Uemoto, Y.et al. (2021). JABG, *138*(2), 237–245. https://doi.org/10.1111/jbg.12505

Pinchasov, Y. (1991). *British Poultry Science*, *32*(1), 109–115. https://doi.org/10.1080/00071669108417332

Robinson, D. L. (1996). *Livestock Production Science*, *45*(2), 111–122. https://doi.org/10.1016/0301-6226(96)00002-4

Willham, R. L. (1963). *Biometrics*, *19*(1), 18–27. https://doi.org/10.2307/2527570

589. Host transcriptome response of meat-type chickens infected with *Eimeria maxima* and reared under heat stress environment

S.E. Aggrey[1*], A.F.A. Ghareeb[1], A.L. Fuller[1], M.C. Milfort[1], J.C. Foutz[1], G.H. Schneiders[1,2], W.S. Habashy[1,3], K. Adomako[4] and R. Rekaya[1]

[1]University of Georgia, Athens, GA, USA; [2]Merck Animal Health, Madison, NJ, USA; [3]Damanhour University, Damanhour, Egypt; [4]Kwame Nkrumah University of Science and Technology, Kumasi, Ghana; saggrey@uga.edu

Abstract

Molecular responses of broiler chickens to stressors would provide the essential architecture required to devise innovative strategies to mitigate them. Ileum tissues from chickens infected with *E. maxima* and reared under either thermoneutral (TN) or heat stress (HS) conditions were used to study the molecular mechanisms that underlie responses to these stressors. Chickens modulate the ileum transcriptome by catabolizing amino acids for energy when subjected to HS alone. Toll-like receptor and cytokine-cytokine receptor signaling were downregulated while PPAR and fatty acid degradation were upregulated. Fatty acid degradation, tryptophan and glutathione pathways were downregulated in the in chickens infected with *E. maxima* and reared under TN conditions. When chickens were infected with *E. maxima* and subjected to HS, they downregulated calcium signaling, tryptophan and fatty acid metabolism to reduce the nutrients available for scavenging by the parasite. The reduced nutrients available for the parasite to scavenge limit their reproductive ability.

Introduction

Eimeria spp. are apicomplexan parasites that cause coccidiosis in poultry. The economic impact of coccidiosis is estimated to be $3B globally. *Eimeria (E.) maxima* is one of the seven recognized species of coccidia that infect chickens. The disease is characterized by reduced growth and feed intake, apathy, diarrhea and in severe cases, mortality. *Eimeria maxima* specifically infects chicken at mid-gut regions proximal and distal to the Meckel's diverticulum. During infection, the parasite destroys the intestinal mucosa and impacts digestibility and nutrient absorption. The intestine undergoes enterocytic damage and epithelia sloughing predisposing the chicken to invasion of pathogens that causes other enteric diseases (Dubey and Jenkins, 2018).

Heat stress (HS) could be seasonal or an endemic stressor in the poultry industry that impacts health, feed intake, and growth of chickens leading to severe economic losses. Heat stress can lead to oxidative stress and that could have several negative physio-pathological effects on the birds. Impaired growth during HS can be attributed to reduced nutrient availability (Habashy *et al.*, 2017). Changes in nutrient metabolism and the heat load can significantly affect protein biosynthesis. Understanding the host molecular responses to *E. maxima* infection and temperature would allow for the development of novel mitigation strategies. We hypothesized that *E. maxima* infection and HS will elicit changes in the ileum transcriptome.

Materials & methods

Freshly sporulated *E. maxima* oocysts were counted suspended in PBS at a concentration of one 500 oocysts/ml Single oocysts were observed in 2µl droplets under light microscopy and collected. Twelve 14 days old-broilers kept in sterile isolators were infected with one oocyst suspended in 200µl of PBS via gavage. Fecal matter was collected from each isolator at 7 days post infection (dpi) and analyzed for the presence of oocysts by salt flotation. One volume of feces was solubilized in 9 volumes of saturated salt solution and the supernatant used to verify for the presence of oocysts in the presence of oocysts in a

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_589, © S.E. Aggrey *et al.* 2022

McMaster chamber. The oocysts recovered from one of the chickens were sporulated in three consecutive passages. The purity of the *E. maxima* clone was verified by PCR (Jenkins *et al.*, 2006).

Six hundred 14-d old commercial broiler chickens were divided into 2 groups and reared at 20 °C (thermoneutral, TN) or 35 °C (HS). Each temperature group comprised of 30 groups of 10 chicks. Within each temperature group, half of the chicks were infected via gavage with 2×10^5 *E. maxima* sporulated oocysts suspended in water and the other half were mock infected with water. All chickens were raised in batteries with wired floor, with *ad libitum* access to feed and water. In all, there were 4 experimental groups: TN control [TNc], TN infected [TNi], HS control [HSc] and [HSi]. At 6 dpi, 5 chickens per group were randomly sampled, and ileum tissues were collected for RNA extraction and then sequenced using NGS Illumina sequencing platforms. Total RNA was extracted from the ileum tissue using Trizol reagents and purified with RNeasy mini kits. Concentration of the extracted RNA was assessed with UV absorbance and the $OD_{260/280}$ ratios for all samples were >1.9. All RNA samples displayed an RNA Integrity Number (RIN) of ≥9. cDNA libraries were prepared with 4 µg total DNA using TruSeq RNA sample Preparation Kit to obtain cDNA fragments of 229 bp average size, or 355 bp including adapter sequences. All 20 cDNA libraries were sequenced with 150 bp paired-end read using the Illumina HiSeq 2000 system.

Quality reads were aligned to *Gallus gallus* 5.0 using STAR aligner v.2.5.2b (Dobin *et al.*, 2013). Hit counts were counted using featureCounts v.1.5.2 (Liao *et al.*, 2014), using only unique reads. DESeq2 (Anders and Huber, 2010) was used for differential gene expression analysis. The Wald test was used to generate *P*-values and $\log_2$ fold changes (FC). Fold change were conducted for all orthogonal comparisons. Genes were considered as differentially expressed (DEGs) at FDR ≤0.05 and $\log_2$ FC>1. We performed gene ontology (GO) and KEGG pathway analyses for DEGs using ClueGo Plugin v1.x-v2.x.x within Cytoscape v3.7.2. (Shannon *et al.*, 2003). ClueGo integrated GO terms and KEGG pathways to create functionally organized pathways.

Results

Subjecting chickens to HS lead to gene expression changes in the ileum. The aldehyde dehydrogenase 7 family member A1(ALDH7A1), histidine ammonia lyase (HAL), glycine cleavage system protein H (GCSH) cholinergic receptor nicotinic Alpha 7 Subunit (CHRNA7), cholecystokinin A receptor (CCKAR), lipase G, endothelial type (LIPG) were upregulated in the HSc and the polymeric immunoglobulin receptor (PIGR), tumor necrosis factor receptor super family 13B (TNFRSF13B), relaxin 3, interleukin 8-like 1 (IL8L1) and DnaJ heat shock protein family (Hsp40) member A1 (DNAJA1) were downregulated in the HSc compared with the TNc group. The valine, leucine and isoleucine, tyrosine, lysine degradation and histidine metabolism pathways were upregulated while protein processing in endoplasmic reticulum, RNA degradation, cytokine-cytokine receptor interaction pathways were downregulated in the HSc group compared to the TNc group.

In the current study, the Aly/REF export factor (ALYREF), IL12β, hepatocyte growth factor (HGF), IL1β and integrin subunit alpha 2 (ITGA2) were upregulated while phosphoenolpyruvate carboxykinase 1 (PCK1), glucose 6 phosphatase catalytic subunit (G6PC), kynurenine-oxoglutarate transaminase 3 (KYAT3), propionyl-CoA carboxylase (PCCA) and the catalase (CAT) were downregulated in the TNi group compared with the TNc group. The RNA transport, spliceosome, toll-like receptor, cytokine-cytokine receptor signaling and proteasome pathways were upregulated while the peroxisome, PPAR, fatty acid degradation, tryptophan and glutathione pathways were downregulated in the TNi group compared with the TNc group.

When chickens under HS are infected with *E. maxima*, the genes that affect function in the ileum respond to the infection. The carnitine palmitoyltransferase 1A (CPT1A), crystallin alpha B (CPTT2) and acyl-CoA synthetase long chain family member 5 (ACSL5), relaxin 3 (RLN3) and IL8 were upregulated while acyl-CoA oxidase 3 (ACOX3), alcohol dehydrogenase 6 (ADH6) and acyl-CoA synthetase family member 1 (ACLS1) while dopa decarboxylase (DDC), indoleamine 2,3-dioxygenase (IDO), ryanodine receptor 2 (RYR2), calcium voltage-gated channel subunit alpha1 B (CACNA1B) and voltage dependent calcium channel protein (TPCN2) were downregulated in the HSi group compared to the HSc group. Among the upregulated pathways in the HSi group were proteasome, cell cycle, steroid biosynthesis and oxidative phosphorylation, however, calcium signaling, fatty acid degradation, tryptophan, neuroactive ligand receptor interaction and phosphatidylinositol pathways were downregulated in the HSi group compared to the HSc group.

Discussion

Several studies have shown that the chicken reorganizes its transcriptome to respond to stressors including heat. The current study shows significant upregulation of glycine, serine threonine, tyrosine and histidine metabolisms and lysine degradation. Several members of the aldehyde dehydrogenase (ALDH) are upregulated in the ileum. The ALDH7A1 gene encodes the enzyme antiquitin which is involved in the breakdown of amino acids, such as lysine. Lysine degradation is necessary for energy production (Brocker *et al.*, 2010). The hal gene encodes for the enzyme histidase that breaks down the histidine to form urocanic acid which is further catalyzed to form glutamic acid (Suchi *et al.*, 1995). The GCSH gene which was upregulated is part of the system that catalyzes the degradation of the amino acid glycine. The H protein (GCSH) shuttles the methylamine group of glycine from the P protein (GLDC) to the T protein (GCST) (Leung *et al.*, 2021). Most of the genes associated with amino acid networks encode for enzymes that degrade amino acids into other molecules for other functions or for the generation of energy. Heat stress causes significant reduction in feed intake leading to significant reductions in available nutrients and energy to the chicken. The data from the current study clearly indicate that, the energy need of chickens reared under HS may be partly met by degradation of amino acids such as lysine, glycine, histidine, valine, leucine and isoleucine. Downregulation of the cytokine-cytokine signaling and toll-like receptor pathways is an indication that immunity is suppressed during HS. The PIGR gene that binds to the polymeric IgA and IgM as well as the TNFRSF13B gene which is associated with immunodeficiency were downregulated in the HS uninfected group compared to the TN counterparts.

When chickens are infected with *E. maxima*, ALYREF, a nuclear protein and molecular chaperone which is involved in the spliceosome, RNA transport and mRNA surveillance is upregulated. It appears that there is a concerted effort to increase the maturation of the eukaryotic mRNA. There is a concomitant upregulation of the toll-like receptor pathway, cytokine-cytokine receptor interaction and other immune-related pathways. Several interleukins including IL10, IL22, IL10RA, IL17RA, IL12B, IL8L1, IL6R are all upregulated. Infection with *E. maxima* may increase oxidative stress and cause a reduction in the antioxidant capacity of the chicken. The peroxisome and glutathione metabolisms which play significant roles in ROS detoxification were downregulated. The PCCA gene involved in the catabolism of odd chain fatty acids, branched-chain amino acids isoleucine, threonine, methionine, and valine and other metabolites (Kalousek *et al.* 1980) was downregulated. The AU binding methylglutaconyl-coA hydratase, an enzyme critical in the leucine degradation pathway is downregulated. KYAT3 which encodes an enzyme involved in the catabolism of tryptophan is also downregulated. It appears that chickens infected with *E. maxima* modify their metabolism to reduce the catabolism of several amino acids to limit available amino acids available for scavenging by the parasite.

Chickens when infected with *E. maxima* and reared under HS tend to modify their physiology to limit the replication of the parasite (Schneiders *et al.*, 2020). Beta oxidation which is usually prominent under HS alone is reduced as evidenced by the downregulation of FA metabolism. The upregulation of CPT1A, CPTT2 and ACSL5 and downregulation of ACOX3, ADH6 and ACLS1 potentially limits the oxidation of long-chain FA. Tryptophan metabolism is also downregulated. Tryptophan (Tryp) is an essential amino acid which is catabolized to kynurenic acid. The IDO and DDC genes that are associated with the catabolism of tryptophan are all downregulated. Induction of IDO has been shown to deplete Tryp and subsequently limit the growth of *E. falciformis* in mouse (Schmid *et al.*, 2012). Tryptophan metabolism was downregulated regardless of the rearing temperature. The calcium signaling pathways was also downregulated. The RYR2 and RYR3 genes which form channels that transport Ca^{2+} are downregulated. These channels are embedded in the endoplasmic/sarcoplasm reticulum which acts as a storage center for Ca^{2+}. Calcium release might be triggered by increased cytoplasmic Ca^{2+} level due to the activation of the L-type calcium channel CACNA1B which was also downregulated in the HSi group compared with the HSc group. The ATP driven Ca^{2+} ion pump and TPCN were all downregulated. Long-chain FA, Tryp and Ca^{2+} may be among the nutrients and ions scavenged by *E. maxima* for replication, and downregulation of pathways may explain why replication of the parasite is limited under HS (Schneiders *et al.*, 2020).

References

Anders, S., and Huber, W. (2010). Genome Biology, 11(10), R106.

Brocker C, Lassen N, Estey T, *et al.* (2010) J. Biol. Chem. 285(24):18452-18463.

Dobin, A., Davis, C. A., Schlesinger, F. *et al.* (2013) Bioinformatics 29(1):15-21.

Dubey J.P., and Jenkins, M.C. (2018) Parasitology. (2018) 145:1051-58.

Habashy WS, Milfort MC, Adomako K, *et al.* (2017) Poult. Sci. 96(7):2312-2319.

Jenkins, M. C., Miska, K. and Klopp, S (2006). Avian Dis. 50:110–114

Kalousek, F., Darigo, M.D., Rosenberg, L.E. (1980) J. Biol. Chem. 255(1):60-65.

Leung KY, De Castro SCP, Galea GL *et al.* (2021) Front. Genet. 12:625120. https://doi.org/10.3389/fgene.2021.625120

Liao, Y., Smyth, G. K., and Shi, W. (2014). Bioinformatics 30(7): 923-930.

Schmid M, Lehmann M.J., Lucius R. *et al.* (2012). J Biol Chem. 287(24):20197-20207.

Schneiders G.H., Foutz, J.C., Milfort, M.C. *et al.* (2020) Sci Rep. 10(1):10736. https://doi.org/10.1038/s41598-020-67330-w.

Shannon, P., Markiel, A., Ozier, O., *et al.* (2003) Genome research, 13(11), 2498-2504.

Suchi, M., Sano, H., Mizuno, H, *et al.* (1995) Genomics. 29(1):98-104.

590. Changes in genetic parameters for traits under genomic selection in poultry

J. Richter[1*], J. Hidalgo[1], V. Breen[2], R. Hawken[2], I. Misztal[1] and D. Lourenco[1]

[1]*Department of Animal and Dairy Science, University of Georgia, Athens, GA 30602, USA;*
[2]*Cobb-Vantress, Inc., Siloam Springs, AR 72761, USA; jennifer.richter25@uga.edu*

Abstract

We investigated changes over time in heritability of a feed efficiency trait (FE), a carcass yield trait (CY), a growth trait (GT), and a leg disorder trait (DIS) in a broiler population. The data contained 200,093 (FE), 42,895 (CY), 203,060 (GT), and 63,349 (DIS), phenotypic records obtained over 55 mating groups then divided into 13 intervals. Pedigrees included 1,250,404 animals, of which 154,318 were genotyped. Genetic parameters were estimated with single-trait models for each interval with (GEN) and without genomic information (PED) using the BLUPF90 programs. Single-step was used in the former. Heritabilities changed from 0.38 to 0.23 (0.38 to 0.26) for GEN (PED) across performance traits and from 0.55 to 0.25 (0.55 to 0.28) for the disorder trait. Heritability estimates by PED were often higher than those from GEN, indicating bias. Genetic parameters should be updated regularly for unbiased estimation of breeding values and selection response.

Introduction

Accurate genetic parameters are crucial for unbiased estimation of breeding values and selection response in breeding programs. Genetic parameters are expected to change under selection, and the magnitude of the changes depends on the intensity of selection and the initial allele frequency (Walsh and Lynch, 2018). The additive genetic variance, and therefore, the heritability, can be reduced under selection due to the increased coancestry among animals (Walsh and Lynch, 2018). Another cause of reduction is the Bulmer effect, which creates negative linkage disequilibrium (i.e. negative covariance between pairs of loci) under directional or stabilizing selection (Bulmer, 1971). Due to strong selection, genetic parameters may change, leading to inaccurate predictions of genetic gain when using the old parameters. Thus, it is important to update the genetic parameters regularly.

Estimating genetic parameters over time is computationally costly with extensive data and under genomic selection because the genomic relationship matrix (G) is dense. According to Hidalgo *et al.* (2020), time intervals are useful to estimate genetic parameters over time so that only a fraction of the data is utilized in each analysis. Intervals need to be large enough and with sufficient data to avoid biases due to the averaged changes (Misztal *et al.*, 2021). The objective of this study was to estimate changes in heritability over time for growth, efficiency, and leg disorder traits in a broiler population under genomic selection using time intervals including or not genomic information.

Materials & methods

Data. The dataset used in this study was provided by Cobb-Vantress Inc. (Siloam Springs, AR). The whole dataset contained 200,093 phenotypes for FE, 42,895 for CY, 203,060 for GT, and 63,349 for DIS. The pedigree included 1,250,404 purebred broilers hatched over 7 years, resulting from 55 mating groups (MG). A total of 154,318 birds were genotyped with a 60K SNP panel. We defined 13 time intervals, the first interval comprised seven MG, and the remaining 12 of these intervals comprised eight MG. Each subsequent interval had four MG of overlapping data. For example, the first interval included data from MG one through seven, the second interval included data from MG four through eleven; therefore, the MG four to seven overlapped in intervals one and two. Only the most recent time intervals in the dataset

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_590, © J. Richter *et al.* 2022

contained genotypes (i.e. from MG 24 to 55). Intervals including genotypes had, on average, 38,000 genotyped animals except for the first interval only having ~16,000 genotyped animals.

Model and analysis. Single-trait linear models were used for the efficiency, yield, and growth traits (FE, CY, GT). These linear models included fixed effects of contemporary group (CG) and sex while having a random animal effect. Variance components for the linear models were estimated using the average information restricted maximum likelihood algorithm implemented in the AIREMLF90 software (Misztal *et al.*, 2014). A single trait threshold model was used for the leg disorder trait (DIS). This threshold model included a fixed effect of sex with a random effect of CG and a random animal effect. Variance components for the threshold model were estimated using a Bayesian approach via the Gibbs sampling algorithm as implemented in the THRGIBBS1F90 software (Misztal *et al.*, 2014). A single Gibbs chain of a total length of 100,000 rounds was initially generated. After discarding the initial 10,000 samples as burn-in, 1 in every 10 samples were stored to compute means and standard deviations of the posterior distributions. The means were used as estimates of the variance components, and their posterior standard deviations were considered a measurement of their estimation errors. The analyses to estimate variance components for all traits were performed, including (GEN) or not the genomic information (PED) for each interval.

The full form of the linear models was:

$$y = Xb+Za+e, \quad (1)$$

where y is a vector of phenotypes; b and a are vectors of fixed and random animal effects, respectively; X and Z are incidence matrices for b and a respectively; and e is a vector of random residuals.

For the categorical trait (DIS), the threshold model assumed an underlying distribution L. The model was:

$$L = X_l b_l + Z_l a_l + W_l c_l + e_l, \quad (2)$$

where L is a vector of underlying distribution of phenotype y; b_l is a vector of systematic effects; a_l is a vector of random animal effects; c_l is a vector of random CG effects; X_l, Z_l, and W_l are incidence matrices for b_l, a_l, and c_l, respectively; and e_l is a vector of random residuals for this threshold model. During the estimation process, the residual variance for DIS was fixed to 1 so that the model is identifiable.

Results

The posterior means and standard deviations for heritabilities estimated by GEN and PED are shown in Figure 1 for broiler performance and leg disorder traits; heritabilities for DIS are presented in the liability scale. Changes in heritability estimated by GEN (PED) from the first to the last interval, that is from 1-7 to 47-54 MG were from 0.31 to 0.19 (0.31 to 0.20) for FE, 0.60 to 0.38 (0.60 to 0.45) for CY, 0.25 to 0.11 (0.25 to 0.13) for GT, and 0.55 to 0.25 (0.55 to 0.28) for DIS.

Discussion

Although there was an overall decline, heritabilities seemed to stabilize during the last four intervals for FE, CY, and DIS. The stabilization of heritabilities happens because the recombination between unlinked loci removes linkage disequilibrium, regenerating genetic variance (Bulmer, 1971). Villanueva and Kennedy (1990) proved that a recombination-linkage disequilibrium is achieved after several generations of selection; therefore, the genetic variance and heritability will stabilize, provided that other components of the heritability remain unchanged.

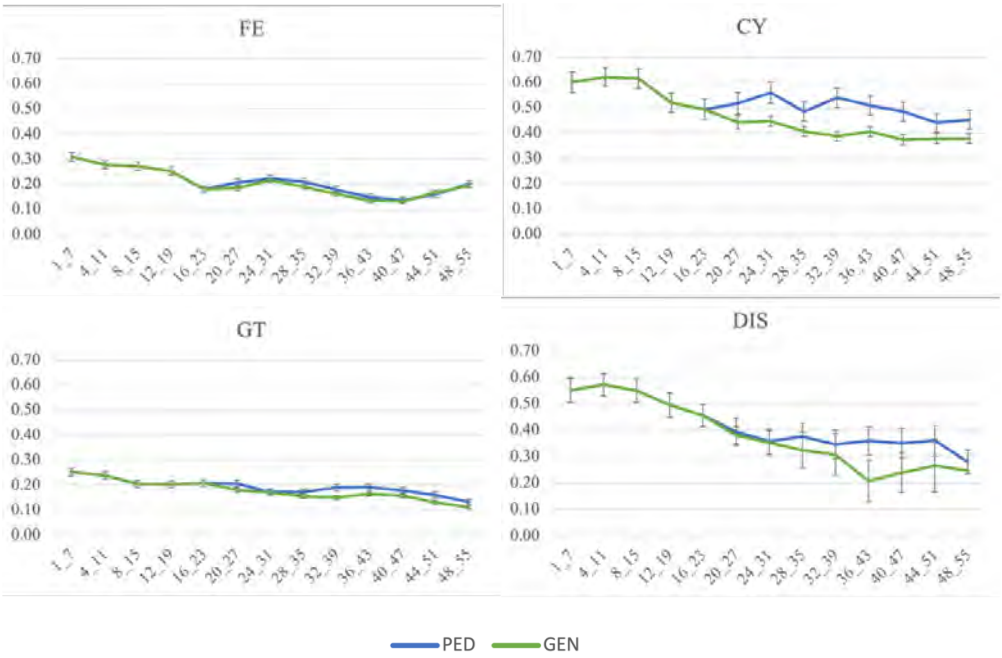


Figure 1. Posterior means and standard errors for estimated heritabilities for FE, CY, GT, and DIS estimated with (GEN) or without genotypes (PED). Heritabilities for all traits decreased over time.

Changes in heritabilities over time were similar with and without the inclusion of genomic information for FE; however, the heritability estimates by PED for CY, GT, and DIS were higher than those from GEN. Our results agreed with those found by Hidalgo *et al.* (2020). Those authors concluded that heritabilities estimated by GEN and PED were similar for traits under less intensive selection, in contrast, heritabilities differed for traits more intensively selected. Selection reduces heritabilities; therefore, it is important to update genetic parameters regularly to have unbiased estimation of breeding values and selection response. Heritability estimates may be biased if genomic information is ignored. This is a topic under further investigation.

References

Bulmer, M. G. (1976) Genet. Res. 28(2):101-117. https://doi.org/10.1017/S0016672300016797

Hidalgo, J., S. Tsuruta, D. Lourenco, Y. Masuda, Y. Huang et al. (2020) J. Anim. Sci. 99(9):1-10. https://doi.org/10.1093/jas/skab239

Misztal, I., I. Aguilar, D. Lourenco, L. Ma, J. P. Steibel *et al.* (2021) https://doi.org/10.1093/jas/skab092

Misztal, I., S. Tsuruya, D.A.L. Lourenco, I. Aguilar, A. Legarra *et al.* (2014b) Manual for BLUPF90 family of programs. Available at http://nce.ads.uga.edu/wiki/lib/exe/fetch.php?media=blupf90_all2.pdf

Villanueva, B., and B. W. Kennedy (1990) Theor. Appl. Genet. 80: 746-752. https://doi.org/10.1007/BF00224187

Walsh B., and Lynch M. (2018) Evolution and selection of quantitative traits. Oxford University Press, Oxford, UK.

591. Effects of age, line and diet on the occurrence of white striping lesions in broiler breast muscle

L.T. Fernandes[1*], J.O. Peixoto[1,2], M.S. Dal Pizzol[3], A.M.G. Ibelli[1,2], M.A.Z. Morés[1], F.C. Tavernari[1,3], A. Coldebella[1], M.E. Cantão[1] and M.C. Ledur[1,3]

[1]Embrapa Suínos e Aves, caixa postal 321, 89715-899, Concórdia, Santa Catarina, Brazil; [2]Programa de Pós-Graduação em Ciências Veterinárias, Universidade Estadual do Centro-Oeste, R. Salvatore Renna, 875, 85015-430, Guarapuava Paraná, Brazil; [3]Programa de Pós-Graduação em Zootecnia, UDESC-Oeste, Rua Beloni Trombeta Zanin, 680E, 89815-630 Chapecó, SC, Brazil; lanatf@yahoo.fr

Abstract

The White Striping (WS) is one of the most prevalent pectoral myopathies in chickens affecting meat quality. Its etiology, as well as the influence of diet, age and line, remains unclear. Therefore, effects of broiler line, nutrition and age were investigated to better understand and identify possible causative mechanisms triggering the onset and the progression of WS myopathy in the breast of broiler chickens. The most relevant macroscopic lesion observed was white striations on the surface of the breasts. For the lesion identified macroscopically, Ross had an odds ratio of 8.43 in relation to TT line, while for the microscopic lesion the odds ratio was 22.52. The high-performance diet had an odds ratio of 1.76 in relation to the regular diet for the macroscopic lesions only. Therefore, we have shown that the presence of WS increases with age even in broiler lines that have slower growth rate than commercial lines.

Introduction

In recent years, pectoral myopathies are among the main challenges of the poultry industry, since they negatively affect the sensorial and technological quality of meat (Kato *et al.*, 2019). Among the disorders that affect the breast muscles of fast-growing broilers, one of the most prevalent is the White Striping (WS), which consists macroscopically of the presence of superficial striated fat deposition that runs parallel to the muscle fibres. It affects meat quality, increasing the downgrading percentages and condemnation of breast fillets (Kuttappan *et al.*, 2013). The histological features mostly observed are myodegeneration, interstitial inflammation, infiltration of connective tissue and lipidosis (Vanhatalo *et al.*, 2021). The etiology of pectoral myopathies has been associated with several factors, from the rapid growth rate to the high-calorie diets and specific genotypes (Petracci, *et al.*, 2015). Moreover, heritability estimates for WS in different broiler lines suggest a relevant genetic component for the manifestation of WS (Lake *et al.*, 2021). Nowadays, whole transcriptome studies are making it possible to advance in the knowledge of the genetic mechanisms associated with WS (Marchesi *et al.*, 2018). However, a large amount of data from studies evaluating the physiological and histopathological changes in the *Pectoralis major* muscle across ages and lines are essential to achieve a comprehensive picture of the disorder. Hence, effects of broiler line, nutrition and age were investigated to identify possible causative mechanisms triggering the onset and the progression of WS myopathy in the breast of broiler chickens.

Materials & methods

This study was conducted with two lines of broiler chickens: a paternal broiler line developed by the Embrapa Swine and Poultry National Research Centre called TT and a commercial line of fast-growing broilers Ross AP 91. The TT pure line has been maintained under multi-trait selection, in open sided poultry houses, since 1992, in Concordia, Santa Catarina State, in the south of Brazil. The main traits under selection in TT pure line are body weight (BW), feed conversion, breast size (mass selection), carcass and cuts yield, abdominal fat (indirect selection based on non-selected sibs), fertility, hatchability of fertile eggs and chick viability (independent culling levels). For more information see Marchesi *et al.* (2018).

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_591, © L.T. Fernandes *et al.* 2022

Experimental design. In this study, a 2×2 factorial design was used, with two broiler lines (Ross and TT) and two diets (normal and superior). About 2000 male broilers were raised in 60 boxes with 34 animals each. The diets were recommended for regular (2950-3050 kcal ME/kg; 24.4-21.2% CP) and high performance (3,250-3,200 kcal ME/kg; 22-24% CP) of broilers, according to the Brazilian Tables for Poultry and Swine (Rostagno, 2017).

Histopathological analysis. Approximately 30 chickens of each line and diet were weighed and slaughtered by cervical dislocation at 28, 35 and 42 days of age, following the procedures of the Ethics Committee for Animal Use (CEUA) from the Embrapa Swine and Poultry National Research Centre, under protocol # 08/2019. At necropsy, breast of all animals were visually evaluated for presence of superficial medium-to-thick white striations in the cranial part of the fillet or absence of WS (no white striations) (Kuttappan *et al.*, 2013). Subsequently, samples of 2 cm^2 from the centre-cranial part of *pectoralis major* were collected and fixed by immersion in 10% neutral buffered formalin. Fixed tissue sections were dehydrated and embedded in paraffin wax, and transverse and longitudinal fibre segment for each sample were sectioned at 4 mm and stained with hematoxylin and eosin (HE) for morphologic evaluation.

Statistical analyses. A logistic regression was applied to analyse WS as a categorical trait (0 = control and 1 = affected) with a model considering the fixed effects of age, line, diet and their interactions, which were performed using the Logistic procedure of the Statistical Analysis System© (SAS, 2012). This analysis was complemented with the evaluation of the association between the presence of the lesions with broilers body weight at slaughter. We also performed a c^2 test for the association between the micro and macroscopic evaluation of lesions.

Results

The observed means for body weight at 42 days of age (BW42) for Ross receiving the high performance diet was 3,349.93 (260.25) and for the regular diet was 3,237.97 (246.46). For the TT line, the observed BW42 was 2,809.23 (173.36) with the high performance diet and 2,806.90 (219.26) with the regular diet.

Pathology. The most relevant macroscopic lesion observed was white striations on the surface of the breasts. The main histological features of WS affected breast muscles from both broiler lines were mild to severe degeneration of muscle fibres, moderate to severe degeneration of muscle fibres associated with proliferation of connective tissue and deposition of adipocytes, and mild to severe degeneration of muscle fibres along with mild to severe increased of fat infiltrations (lipidosis).

Statistical analyses. The results from the logistic regression showed significant effects of line, diet and age for the presence of macroscopic lesions of WS in the broilers breasts. The microscopic lesions of WS were influenced only by line and age effects. There was an increase in the occurrence of lesions with age and the commercial line Ross showed more WS lesions than the TT line. Interestingly, at 28 days of age, the TT broilers showed no histological lesions (Figure 1A and 1B). For the lesions identified macroscopically, Ross had an odds ratio of 8.43 in relation to TT line, while for the microscopic lesions the odds ratio was 22.52. The high-performance diet had an odds ratio of 1.76 in relation to the regular diet for the macroscopic lesions only.

A negative association between body weight at slaughter and the presence of WS lesions was observed. However, this result was not consistent, since this association was significant for macroscopic lesions only at 42 days of age for both lines and for microscopic lesions only at 28 days for Ross and 42 days for TT (results not shown). There was a positive association ($P \leq 0.05$) between the macro and microscopic detection of WS lesions (Table 1), with an accuracy of 243/335.

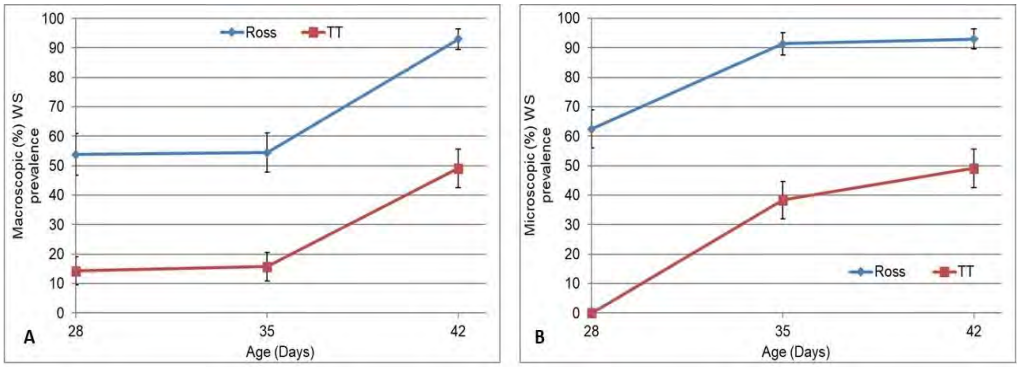


Figure 1. Prevalence and standard error of macroscopic (A) and histologic (B) presence of white striping in TT and Ross broilers breasts by age.

Table 1. Association between macroscopic and microscopic evaluation of lesions.

Macroscopic lesions	Microscopic lesions		Total
	Absent	Present	
Absent	118 (65.92%)	61 (34.08%)	179
Present	31 (19.87%)	125 (80.13%)	156
Total	149	186	335

Discussion

In the present study, histopathological changes occurring in WS indicate that the degenerative myopathy observed could be associated with increased growth rate in birds, since both gross and microscopic lesions were significantly influenced by genetic line. Swatland (1990) suggested that fast-growing broilers have muscles that outgrow their life support systems causing muscle damage. Moreover, it was observed an increase in the occurrence of lesions with age. The prevalence of WS lesions in the commercial line was much higher than in the TT in all evaluated ages, although the shape of the curves was similar. Interestingly, at 28 days of age, no microscopic lesion was detected in TT broilers. This result can be explained because the TT line was in its 27th generation of annual selection, while the commercial lines are originated from many more generations of continuous selection. The high performance diet was used especially to try to induce TT to develop white striping, as this line is a slower-growing line compared to Ross. The reason for the high-performance diet had an effect only in the macroscopic lesions of WS is still to be clarified. In our study, the association of WS lesions with body weight was significant only for specific ages, although we have not evaluated this association with breast weight and yield. We also observed a high concordance between the macro and microscopic detection of WS lesions, especially when WS was present.

Selection for breast yield traditionally focus on ages targeting hypertrophy, which occurs normally between 6 and 8 weeks of age, rather than hyperplastic growth. This results in increase in fibre size rather than fibre number, which could favour myopathies occurrence (Orlowski *et al.*, 2021). Therefore, we have shown that the presence of WS increases with age even in broiler lines that have slower growth rate than commercial lines. More physiological and histopathological data from chicken myopathies across different ages and genetic lines can provide essential information to help interpretate and better explore future integrated analyses from the omics methodologies such as RNA-seq, GWAS, exome and epigenome.

Acknowledgements

This study was financed by the Brazilian Agricultural Research Corporation (EMBRAPA) project # 10.19.03.034.00.00. LTF is supported by a visiting specialist fellowship from the National Council of Scientific and Technological Development (CNPq) grant # 380282/2021-6 and MSDP by a Fundação de Amparo à Pesquisa e Inovação do Estado de Santa Catarina (FAPESC) scholarship. MCL and FCT are CNPq productivity fellows. We thank the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), Brazil – Finance Code 001 for the free access to the journals used in the literature review.

References

Kato T, Mastelini SM, Campos GFC, *et al.* 2019. Asian-Australas J Anim Sci. 32(7):1015–1026. http://doi.org/10.5713/ajas.18.050.

Kuttappan, V. A., Shivaprasad, H. l., Shaw, D. P., Valentine, B. A., Hargis, B. M., Clark, F. D., *et al.* 2013. Poult. Sci. 92. http://doi.org/10.3382/ps.2012-02646.

Lake, J. A., Dekkers, J. C. M., and Abasht, B. 2021. Sci. Reports. 11: 6785. http://doi.org/10.1038/s41598-021-86176-4.

Marchesi, J. A. P. *et al.* 2018. Animal 12: 1126–1134. http://doi.org/10.1017/s1751731117002671

Orlowski, S.K., Dridi, S., Greene, E.S., Coy, C.S. *et al.* 2021. Front Physiol. 12:712095. http://doi/org/10.3389/fphys.2021.712095.

Petracci, M., Mudalal, S., Soglia, F., and Cavani, C. 2015. Worlds. Poult. Sci. J. 71. http://doi.org/10.1017/S0043933915000367.

Rostagno, H.S., Albino, L.F.T., Hannas, M.I., Donzele, J.L. *et al.* 2017. Tabelas Brasileiras para Aves e Suínos: Composição de Alimentos e Exigências Nutricionais. Departamento de Zootecnia-UFV, Viçosa, 488 p.

Sas Institute Inc. System for Microsoft Windows, Release 9.4, Cary, NC, USA, 2002-2012.

Swatland, H. J. 1990. Can. Inst. Food Sci. Technol. J. 23, 239–241. http://doi.org/10.1016/s0315-5463(90)70251-9.

Vanhatalo, O. E., Henderson, J. D., De La Torre, U., Garrity, C. R., Pechanec, M. Y., Mienaltowski, A., *et al.* 2021. Poult. Sci. 100. http://doi.org/10.1016/j.psj.2021.101020.

592. Variation in male autosomal recombination in turkeys (*Meleagris gallopavo)*

C.M. Rochus[1*], B.J. Wood[1,2,3] and C.F. Baes[1,4]

[1]University of Guelph, Centre for Genetic Improvement of Livestock, 50 Stone Rd E, N1G 2W1, Guelph, ON, Canada; [2]Hybrid Turkeys, Suite C, 650 Riverbend Drive, Kitchener, Ontario, N2K 3S2, Canada; [3]School of Veterinary Science, University of Queensland, Gatton, Queensland, 4343, Australia; [4]University of Bern, Vetsuisse, 3001 Bern, Switzerland; crochus@uoguelph.ca

Abstract

Recombination, a feature of sexual reproduction, creates new combinations of genetic material during meiosis. Breeding strategies could benefit from an increased knowledge of recombination in avian livestock species, especially on recombination occurring in microchromosomes. We estimated recombination rates in male turkeys using medium density SNP genotype data from three commercial selection lines. Line 1 had a significantly higher genome-wide recombination rate than the other lines, however, the three lines shared the majority of the detected recombination hotspots. Recombination rate on microchromosomes was 2.3 to 2.7 times higher than on macrochromosomes. A genome-wide regression analysis was used to estimate SNP effects on crossover number. We identified several candidate genes and three functional enrichments. We have provided an updated estimate of the recombination rate in turkeys and found variation in recombination rate between lines and chromosomes.

Introduction

Recombination occurs when two homologous chromosomes break and exchange segments during meiosis. Higher recombination rates have been associated with regions that have higher GC content, the ends of chromosomes, and during meiosis in females. Recombination frequency and hotspot location have been associated with specific genes, for example, *PRDM9* and *RNF212* in mammals (Sandor, *et al.*, 2012). Simulations have shown that increasing the recombination rate in breeding program populations could improve long-term selection, by increasing response to selection and decreasing genetic variance loss (Battagin, *et al.*, 2016).

Birds and reptiles have microchromosomes, which are small, gene-rich chromosomes with a high GC content (Waters, *et al.*, 2021). Most birds species have about 31 microchromosomes that have higher recombination rates, mutation rates, number of interchromosomal interactions, and have more open chromatin than macrochromosomes (Aslam, *et al.*, 2010; Axelsson *et al.*, 2004; Perry *et al.*, 2020; Waters, *et al.*, 2021). Despite the dynamic nature of microchromosomes, proteins from their genes are highly conserved, which means that these proteins have esstential functions and microchromosomes are under greater selection pressure than macrochromosomes (Axelsson *et al.*, 2004).

Recombination within the genome of turkeys has not been well studied. Previous recombination estimates (Aslam *et al.* 2010) used 531 informative SNPs, consequently the resolution of detected crossover locations was poor. More recently, using higher density SNP genotyping, runs of homozygosity were detected in turkeys, which could give insight into recombination rate variation (Adams *et al.*, 2021). With the availability of this affordable, higher density SNP array, we have updated the estimate of turkey recombination, and characterised the variation seen across the genome, and within and between lines. This updated estimate of recombination could be used to develop and improve avian-specific breeding strategies.

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_592, © C.M. Rochus *et al.* 2022

Materials & methods

Samples. For this study we used 60K SNP genotype data from three turkey lines which included 1,889 individuals genotyped from line 1, 2,540 from line 2, and 2,410 individuals from line 3. We removed variants within each line that had a Mendelian error rate >0.05 or a minor allele frequency less than 0.01 using PLINK 1.9 (Purcell and Chang, 2021).

Recombination. We used LINKPHASE3 to detect crossover events in genotyped individuals with at least one parent genotyped using linkage information and a hidden Markov model to account for genotyping errors (Druet and Georges, 2015). We did not account for interference. Crossover positions were indicated by their two flanking, informative SNP markers and we averaged the two positions to simplify further calculations of the recombination rate. We calculated recombination rate as the number of crossovers divided by the physical distance. We identified recombination hotspots in sliding 1 Mb windows that had a recombination rate two and a half standard deviations greater than the genome-wide average recombination rate (calculated separately for each line).

Genome-wide association analysis. We used PLINK 1.9 to estimate SNP effects in a genome-wide association analysis of crossover number (Purcell and Chang, 2021). Phenotypes were the total number of crossovers detected in haplotypes inherited by an individual's offspring divided by their total number of offspring. We used line as a covariate in the genome-wide association analysis and the Benjamini and Yekutieli (2001) step-up false discovery method to correct the inflation of *P*-values. We used USCS Genome Browser (Navarro Gonzalez, *et al.*, 2021) to identify candidate genes based on the domestic chicken RefSeq Genes located 0.5 Mb upstream and downstream of SNPs of interest from the genome-wide analysis. Finally, we used String to identify functional enrichments in the candidate genes (Szklarczyk, *et al.*, 2019).

Results

Samples. After filtering SNPs for Mendelian errors and minor allele frequency in each line, 59,749 SNPs remained for line 1, 60,778 for line 2, and 61,704 variants for line 3.

Recombination. To accurately estimate recombination rates, only genotyped males (toms) with five or more offspring genotyped were included in our calculations. There were a total of 405 toms that fit this criterion: 84 from line 1, 159 from line 2, and 162 from line 3. There were only 12 females (hens) with genotypes that had at least five offspring (two from line 1, seven from line 2, and three from line 3) and these were removed from further analyses.

The first line had an average genome-wide recombination rate (3.9×10^{-8}) that was significantly higher than the other two lines (3.3×10^{-8} and 3.5×10^{-8}). Recombination rates for each chromosome, by line are shown in Figure 1. In general, as chromosome size decreased, recombination rate increased. Recombination rates were 2.3, 2.5, and 2.7 times higher in microchromosomes compared to macrochromosomes in the three lines respectively. A total of 28 recombination hotspots were identified, the majority of which were shared across the three lines. There were more than five times more hotspots identified on microchromosomes than on the macrochromosomes.

Genome-wide association analysis. Our genome-wide association analysis identified seven SNPs of interest. The candidate genes in the nearby regions are shown in Table 1. The String analyses identified three functional enrichments for all the candidate genes: 'biological pathways', 'molecular function', and 'cellular component'.

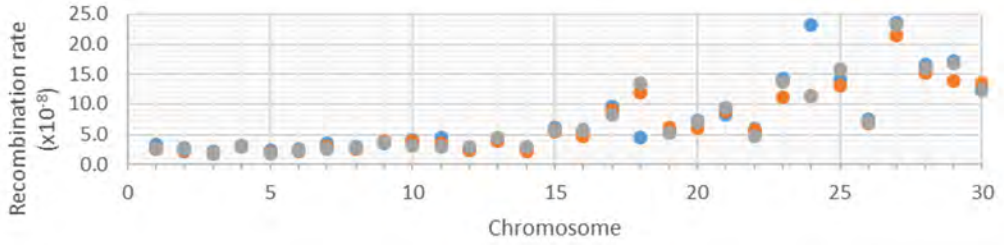


Figure 1. Male autosomal recombination rate in turkeys by chromosome and commercial selection line (line 1 (blue), line 2 (orange) and line 3 (grey)).

Table 1. SNPs associated with male recombination rate in turkeys and candidate genes based on annotation of the chicken genome.

Chr	Position (bp)	*P*-value	Candidate genes
3	72075403	5.54×10^{-8}	*PTDSS1, CPQ, RPL30, STK3*
3	72100530	2.63×10^{-7}	
5	12203971	3.97×10^{-22}	*NUCB2, TPH1, MYOD1, MYF5, FTL, SPTY2D1, UEVLD, TSG101, LDHA, GTF2H1, TPH1*
10	2561024	5.54×10^{-8}	*TGFBR3, RDH5, GLI1, SARNP, ORMDL2, FAM129A*
11	5607589	5.84×10^{-7}	*AP1S3, SERPINE2, WDFY1*
17	5075728	1.73×10^{-7}	*P2RX4, ANAPC5, HIP1R, DENR, ARPC3, ATP2A2, IFT81, ATP6V0A2, GTF2H3, TMED2, SNRNP35, SBNO1, UBB*
22	7781174	2.66×10^{-10}	*CDH4, CDH2, GATA5, GATA6, ADRM1, PSMA7,*
22	7857909	1.49×10^{-10}	

[1] Chr = chromosome.

Discussion

Our results highlighted the variation in recombination rate across the turkey genome, within and between lines. This knowledge could contribute to future, avian-specific, breeding strategies that incorporate recombination. As with other species, recombination increased as chromosome size decreased. This was especially striking in our study because of the microchromosomes. Much of avian diversification has been attributed to the microchromosomes, which has likely been aided by this increased recombination rate (Singhal *et al.*, 2016; Axelsson *et al.*, 2004).

We didn't identify candidate genes that have been associated with recombination in other species. For example, *PRDM9*, which is associated with hotspot location and is highly conserved in mammals, is not present in most avian species including chicken (Singhal *et al.*, 2016). This has resulted in recombination hotspots in avian species that are concentrated at functional elements with greater access to recombination machinery, and highly conserved hotspots shared between species separated by millions of years (Singhal *et al.*, 2016). In agreement with this, we did not see variation in the location of recombination hotspots between turkey lines.

We identified several candidate genes including, *PSMA7*, which was a paralog of a gene, *PSMA8*, that has a role in meiotic exit and fertility in mice (Gómez *et al.*, 2019). Additionally, candidate genes had functional enrichments for: 'biological pathways', 'molecular function', and 'cellular component', with genes associated with DNA and chromatin binding, which might play a role in meiosis.

Acknowledgements

This project was funded by the Government of Canada through Genome Canada and the Ontario Genomics Institute (OGI-133) as part of the project entitled 'Application of genomic selection in turkeys for health, welfare, efficiency and production traits' through the Genomic Application Partnership Program (recipients: B.J. Wood (Industry) and C.F. Baes (Academic)). The authors also acknowledge NSERC and Hybrid Turkeys for support.

References

Adams S.M., Derks M.F.L., Makanjuola B.O., Marras G., Wood B.J., Baes C.F. (2021) Poult. Sci., 100(11):101366. https://doi.org/10.1016/j.psj.2021.101366

Aslam M.L., Bastiaansen J.W., Crooijmans R.P., Vereijken A., Megens H.-J., and Groenen, M.A. (2010) BMC Genom., 11:647. https://doi.org/10.1186/1471-2164-11-647

Axelsson E., Webster M.T., Smith N.G., Burt D., and Ellegren H. (2004) Genome Res., 15:120-125. https://doi.org/10.1101/gr.3021305

Battagin M., Gorjanc G., Faux A.-M., Johnston S.E., and Hickey J.M. (2016) Genet. Sel. Evol., 48:44. https://doi.org/10.1186/s12711-016-0221-1

Benjamini Y.B., and Yekutieli D. (2001) Ann. Stat., 29(4):1165-1188. https://doi.org/10.1214/aos/1013699998.

Druet T., and Georges M. (2015) Bioinformatics, 31(10):1677–1679. https://doi.org/10.1093/bioinformatics/btu859

Gómez-H L., Felipe-Medina N., Condezo Y.B., Garcia-Valiente R., Ramos I., *et al.* (2019) PLoS Genet., 15(8):e1008316. https://doi.org/10.1371/journal.pgen.1008316

Johnsson M., Whalen A., Ros-Freixedes R., Gorjanc G., Chen C.-Y., *et al.* (2021) Genet. Sel. Evol., 53:54. https://doi.org/10.1186/s12711-021-00643-0

Navarro Gonzalez J., Zweig A.S., Speir M.L., Schmelter D., Rosenbloom K.R., *et al.* (2021) Nucleic Acids Res., 49(D1):D1046–D1057. https://doi.org/10.1093/nar/gkaa1070

Perry B.W., Schield D.R., Adams R.H., and Castoe T.A. (2020) Mol. Biol. and Evol., 38(3):904-910. https://doi.org/10.1093/molbev/msaa253

Purcell S., and Chang C. (2021). PLINK v1.9. Retrieved from www.cog-genomics.org/plink/1.9/

Sandor C., Li W., Coppieters W., Druet T., Charlier C., and Georges M. (2012). PLoS Genet., 8(7):e1002854. https://doi.org/10.1371/journal.pgen.1002854

Singhal S., Leffler E.M., Sannareddy K., Turner I., Venn O., *et al.* (2016) Science, 350(6263):928–932. https://doi.org/10.1126/science.aad0843

Szklarczyk D., Gable A.L., Lyon D., Junge A., Wyder S., *et al.* (2019) Nucleic Acids Res., 47(D1):D607–D613. https://doi.org/10.1093/nar/gky1131

Waters P.D., Hardip R.P., Ruiz-Herrera A., Álvarez-González L., Lister N.C., *et al.* (2021) PNAS, 118(45):e2112494118. https://doi.org/10.1073/pnas.2112494118

593. Functional annotations of regulatory elements in the chicken genome

Z. Pan[1], Y. Wang[1], L. An[1], Y. Bi[1], D. Guan[1], M. Delany[1], H.H. Cheng[2] and H. Zhou[1*]

[1]Department of Animal Science, University of California, Davis, CA, USA; [2]USDA, ARS, Avian Disease and Oncology Laboratory, East Lansing, MI, USA; hzhou@ucdavis.edu

Abstract

Functional annotation of the chicken genome, especially the identification of regulatory elements, will play a significant role in identifying causative genetic variants associated with economically important traits. Using a bioinformatics pipeline that incorporates and analyzes (Chromatin Immunoprecipitation Sequencing (ChIP-seq), Assay for Transposase-Accessible Chromatin using sequencing (ATAC-seq), and RNA-seq datasets, we extend our annotation efforts to seven gut-associated tissues from two male chickens used in the FAANG pilot project. ChromHMM was used to predict a total of 15 chromatin states and 1,197,186 regulatory elements (excluding Quiescent) spanning 15 tissues (include 8 core tissues done previously) with 398,050 new regulatory elements were identified. These regulatory elements included 29,823 strong promoters, 104,560 active strong enhancers, and 73,582 repressors. This result provides a great resource for the farm animal genome community to further explore causative variants of complex traits.

Introduction

Chicken is one of most important farm animal species in the world. In addition, chicken is a model organism especially for developmental biology, immunology, and epigenetic regulation of gene expression (Beacon and Davie 2021). The Encyclopedia of DNA Elements (ENCODE) projects clearly demonstrated the important role of regulatory elements and their subsequent contribution to genetic variants associated with health and diseases (Kundaje *et al.* 2005; Gorkin *et al.* 2021). Similar to the ENCODE project, the Functional Annotation of Animal Genomes (FAANG) has made significant progress on the improvement of the annotation of farm animal genomes including chickens (Kern *et al.* 2021). Our group previously reported the annotations of functional elements of eight core tissues in chicken (Kern *et al.* 2021). However, as gene regulatory elements show distinct tissue-specific patterns that relate to biological relevant functions, more tissues in more developmental stages is needed.

In the current study, seven gut-associated tissues were further characterized. In total, 30 RNA-seq libraries, 30 ATAC-seq libraries and 150 ChIP-seq libraries are presented here. Here we present an updated annotation of regulatory elements in the chicken genome.

Materials & methods

Animals and tissues. The highly inbred Line 6 and Line 7 from the USDA, ARS, Avian Disease and Oncology Laboratory (ADOL) were intermated and two male F1 birds were grown then euthanized at 20 weeks of age. Tissue samples (15 for each bird) were flash frozen in liquid nitrogen then stored at -80 °C until further assay processing. The protocol no. 18464 for these animals was approved by Institutional Animal Care and Use Committee at the University of California, Davis.

Library preparation and sequencing. RNA-seq, ATAC-seq and ChIP-seq (H3K4me3, H3K4me1, H3K27ac and H3K27me3) experiments were performed on flash-frozen tissue samples. For RNA-seq experiments, total RNA was extracted using TRIzol (Thermo Fisher, #15596026) and treated with DNase I (Thermo Fisher, #EN0521). RNA integrity number was determined using an Agilent Bioanalyzer (Agilent Technologies, Santa Clara, CA, USA) and directional RNA-seq libraries were prepared using an NEBNext

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_593, © Z. Pan *et al.* 2022

Ultra RNA Library Prep Kit for Illumina (New England BioLabs, #E7530L) and sequenced on an Illumina Hiseq 4000 (Illumina, San Diego, CA, USA) with paired-end 150–base pair (bp) reads. For ATAC-seq experiments, libraries were generated from frozen tissue samples by a modified protocol (https://figshare.com/articles/dataset/Final_ATAC_protocol_docx/13891268) according to OmniATAC and cryopreserved nuclei protocols (Corces *et al.* 2017; Halstead *et al.* 2020). For ChIP-seq experiments, samples were processed as previous described (Kern *et al.* 2021). Libraries were sequenced on an Illumina HiSeq 4000 with read length of single-end 50 bp.

Data analysis. The UC Davis FAANG Functional Annotation Pipeline (https://github.com/kernco/functional-annotation) was applied to process the ChIP-seq, ATAC-seq, and RNA-seq data, as previously described (Kern *et al.* 2021). The galGal6 genome assembly and Ensembl genome annotation (v100) were used as reference for read alignments. ChromHMM (Ernst and Kellies 2012) (v.1.20) was used to train the chromatin state prediction model by integrating ChIP-seq (H3K4me3, H3K4me1, H3K27ac, H3K27me3, and input control) and ATAC-seq data from two biological replicates of 15 tissues.

Results & discussions

A total of 210 genome-wide sequencing datasets from 15 major tissues in chicken were used including four histone modifications (H3K4me3, H3K4me1, H3K27ac and H3K27me3), inputs control, ATAC-seq, and RNA-seq. There were 5,378,700,298 new raw reads with an average remaining retention rate of 74.94% generated in the gut-associated tissues (Table 1).

After peak calling, we obtained an average of 55,419, 68,898, 76,695, 23,655, 66,965 peaks for H3K27ac, H3K27me3, H3K4me1, H3K4me3, ATAC, with average covering 4.47%, 13.17%, 4.27%, 2.53%, and 4.58% of the entire genome, respectively (Figure 1a, b).

Table 1. Summary of sequence reads.

Assay	Raw reads	Trimmed	%	Unaligned	%	Filtered	%	Duplicates	%	Remaining	%
H3K27ac	46,196,825	52,027	0.12	1,183,361	2.57	2,393,492	5.33	8,658,969	19.91	33,908,976	73.83
H3K27me3	83,672,667	63,806	0.08	1,066,446	1.29	8,370,626	10.45	8,163,354	11.00	66,008,436	78.65
H3K4me1	47,330,470	50,216	0.13	1,031,343	2.08	3,410,564	7.52	4,776,934	11.26	38,061,413	80.32
H3K4me3	45,031,286	46,316	0.12	1,043,363	2.34	3,222,928	7.32	11,487,650	27.98	29,231,031	65.15
ATAC-seq	100,870,411	89,626	0.09	5,544,252	4.73	23,949,654	25.16	11,832,366	17.90	59,454,513	59.46
RNA-seq	61,091,219	269,064	0.46	2,992,939	4.87	1,549,822	2.61	0	0.00	56,279,394	92.22

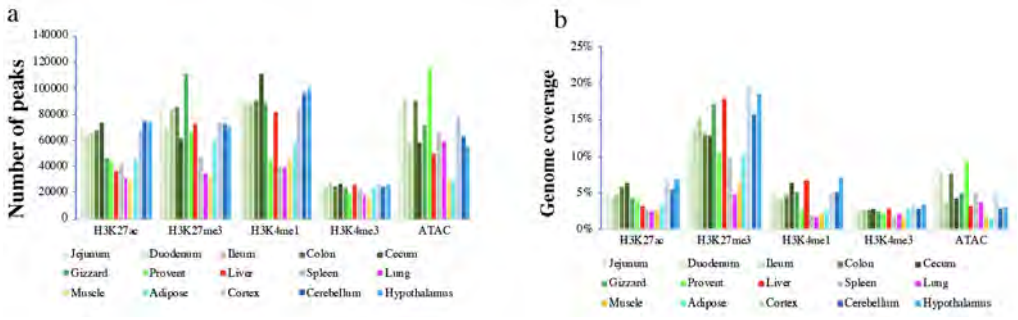


Figure 1. The summary of number of peaks and its genome coverage. a, Peak number, b peak genome coverage in each assay and each tissues.

Next, ChromHMM was used to predict 15 distinct chromatin states by combining all five epigenetic marks across 15 tissues. Each state was defined based on the combination of each marker in the previous studies (Pan *et al.* 2021) (Figure 2a). These states mainly represented promoters (TssA, TssAHet and TssBiv, covering 1.95% of the entire genome), TSS-proximal transcribed regions (TxFlnk, TxFlnkWk and TxFlnkHet, covering 1.55% of the genome), enhancers (EnhA, EnhAMe, EnhAWk, EnhAHet and EnhPois, covering 9.42% of the genome), repressed regions (Repr and ReprWk, covering 21.94% of the genome) and quiescent regions (Qui, 61.44%) (Figure 2d). In total, we identified 1,197,186 regulatory elements (excluding Qui) spanning 15 tissues, including 29,823 strong promoters (TssA), 104,560 active strong enhancers (EnhA), and 73,582 repressors (Repr). We also explored the regulator enrichment in gene element, and found that promoters had higher enrichment in TSS, 5'UTR, CpG islands than enhancers. As expected, both active promoters and enhancers were significantly enriched in the expressed gene' TSS and gene body than in the repressed genes (Figure 2e). Compared with our previous study with eight core tissues (Kern *et al.* 2021), we found 398,050 new regulatory elements (excluding Qui) in these gut-associated tissues. For the 41,802 strong enhancers (EnhA), the Gene Ontology (GO) biological process related to striated muscle cell differentiation and intestinal epithelial cell differentiation were significantly enriched, which suggest important roles of these enhancers on the regulation of genes associated with gut function.

In summary, a more comprehensive functional annotation of regulatory elements across 15 tissues in the chicken genome was generated. This resource should enhance the ability of the animal genome community to identify causative variants associated with economically important traits and further understand the function of these variants in farm animals.

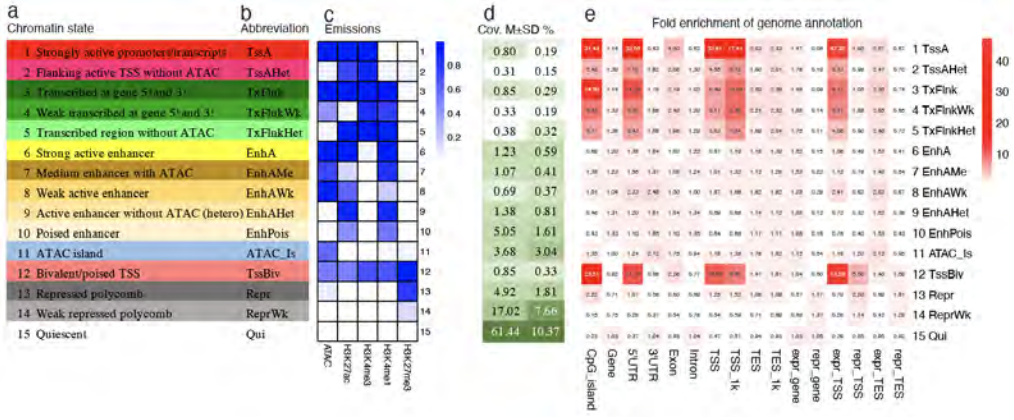


Figure 2. Prediction and characterization of chromatin states across 15 tissues. a, b, Definitions and abbreviations of 15 chromatin states. c, Emission probabilities of individual epigenetic marks for each chromatin state. The color from white to deep blue indicates emission probability (0-1). d, Genomic coverage of each chromatin state. M±SD, mean ± standard deviation. e, Average enrichment of chromatin states for genomic annotations, including CpG islands, genes, TSS/TES_1K (±1 kb around TSS and TES), expressed genes (TPM≥0.1), and repressed genes (TPM<0.1) in each tissue.

Acknowledgements

The study was supported by the United States Department of Agriculture, National Institute of Food and Agriculture, Agriculture and Food Research Initiative grant no. 2020-67015-31175 awarded to H. Z. Partial support was provided by the United States Department of Agriculture, National Institute of Food and Agriculture, Multi-state Research Project NC1170 and NRSP8 (H.Z.) and the California Agricultural Experimental Station (H.Z.).

References

Beacon, T. H., and Davie J. R. (2021) Genome 64: 476-489. https://doi.org/10.1139/gen-2020-0129

Chanthavixay, G., C. Kern, Y. Wang, P. Saelao, S. J. Lamont *et al.*, (2020) Front Genet 11: 567812. https://doi.org/10.3389/fgene.2020.567812

Corces, M. R., A. E. Trevino, E. G. Hamilton, P. G. Greenside, N. A. Sinnott-Armstrong *et al.*, (2017) Nat Methods 14: 959-962. https://doi.org/10.1038/nmeth.4396

Ernst, J., and Kellis, M. (2012) Nature Methods 9: 215-216. https://doi.org/10.1038/nmeth.1906

Gorkin, D. U., I. Barozzi, Y. Zhao, Y. Zhang, H. Huang *et al.*, (2021) Nature 589: E4. https://doi.org/10.1038/s41586-020-2841-4

Halstead, M. M., C. Kern, P. Saelao, G. Chanthavixay, Y. Wang *et al.*, (2020) Sci Rep 10: 5230. https://doi.org/10.1038/s41598-020-61678-9

Kern, C., Y. Wang, X. Xu, Z. Pan, M. Halstead *et al.*, (2021) Nat Commun 12: 1821. https://doi.org/10.1038/s41467-021-22100-8

Pan, Z., Yao, Y., Yin, H., Cai, Z., Wang, Y., *et al.* (2021) Nat commun 12:1-15. https://doi.org/10.1038/s41467-021-26153-7

Roadmap Epigenomics, C., A. Kundaje, W. Meuleman, J. Ernst, M. Bilenky *et al.*, (2015) Nature 518: 317-330. https://doi.org/10.1038/nature14248

594. Time-series transcriptome analysis of broilers under *Eimeria tenella* infection reveals host-pathogen interactions

M. Kim[1], Y. Chung[1], P. Manjula[2], D. Seo[1], S. Cho[1], E. Cho[1], T.K. Ediriweera[1], M. Yu[1], S. Nam[1], J.M. Heo[1] and J.H. Lee[1*]

[1]*Chungnam National University, 99, Daehak-ro, Yuseong-gu, 34134, Daejeon, Republic of Korea;* [2]*Uva Wellassa University, Passara road, 90000, Badulla, Sri Lanka; junheon@cnu.ac.kr*

Abstract

Coccidiosis is a catastrophic disease in the chicken industry, and *Eimeria tenella* is the most critical species. To observe the response under *E. tenella* infection, RNA sequencing was used for comparison between infected and control groups. Samples were collected at 4, 7, and 21 days post-infection (dpi) from both groups. In total, 999 significant differentially expressed genes (DEGs) were detected and grouped into three clusters based on gene expression pattern and clustered to the different three groups using co-expression value. GO and KEGG analyses were conducted for each clustered group, and GSEA was performed on 4 and 7 dpi. As a result, lists of pathways and genes involved in the host response were revealed. And we observed that the innate immune system was activated initially for defence and downregulated for anti-inflammation. The significant DEGs identified in this study will be utilized for further knowledge of coccidiosis resistance in chickens.

Introduction

Coccidiosis is a parasitic disease that causes the most remarkable economic loss in the poultry industry (Blake *et al.*, 2014). Among the causative species, *Eimeria tenella* is the most pathogenic protozoan strain, mainly parasitic on the cecum and increases severe inflammations (Georgieva *et al.*, 2006). To cope with the loss by the disease, genomic selection based on immune-associated genes can be employed to develop disease-resistant animals and reduce recurring prevention costs. In this study, genes and pathways associated with the host response of broiler chickens after *E. tenella* infection were investigated based on gene expression patterns and relationships.

Materials & methods

Animals and sample preparation. A total of 39 male broilers were used and separated into two treatments; 21 birds of positive control (PC) and 18 birds of negative control (NC). The 14 days old chickens in the PC group were orally challenged with 10 times overdosed coccidiosis vaccine (Livacox® T™; Biopharm Co., Prague, Czech Republic). At the 4, 7, and 21 days post-infection (dpi), 7 birds from PC and 6 birds from NC were humanly slaughtered respectively; also the cecum tissues of them were collected. Further, the samples were stored at -80 °C with RNA-later solution.

RNA isolation, complementary DNA (cDNA) library construction, and RNA-Seq. RNA was isolated from the tissue samples using TRIzol reagent (Invitrogen, Carlsbad, CA, USA). The cDNA library was constructed using TruSeq Stranded mRNA sample preparation kit (Illumina, San Diego, CA, USA) with RNA samples. After library construction, samples were sequenced using Illumina's NovaSeq 6000 platform.

Data processing to quantify DEGs. Generated reads from sequencing were aligned to the chicken reference genome (*GRCg6a)*, and gene annotation step was conducted using *Gallus gallus* GTF v102 (Ensembl, https://www.ensembl.org). Gene expressions were calculated by the trimmed mean of M-value (TMM) algorithm. The PC and NC groups were compared at each time point to find differentially expressed genes (DEGs). DEGs were calculated based on the significance thresholds: false discovery rate (FDR)<0.05 and absolute fold change >2.

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_594, © M. Kim *et al.* 2022

Grouping genes using expression patterns and co-expression. After defining DEGs, genes were grouped using two different ways. The first clustering method was based on the reads counts using a *k*-means clustering algorithm to categorize genes with expression patterns. For the second clustering method, the gene co-expression networks (GCNs) were constructed using correlation among gene expressions.

Functional analysis. Functional studies including gene ontology (GO) enrichment analysis and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis were conducted with each group of genes. Biological process and molecular function databases were used for GO analysis. A *P*-value less than 0.05 was set on the criteria for both analyses. Gene set enrichment analysis (GSEA) was performed using normalized reads counts of all genes at 4 and 7 dpi, and the significance was assessed with FDR less than 0.05.

Results

A total of 999 unique DEGs, containing 511, 615, and 2 DEGs at 4, 7, and 21 dpi respectively, were observed. The three *k*-means clustering groups (Type 1, 2, and 3), which comprise 436, 414, and 149 DEGs respectively, are presented in Figure 1. Three of the constructed GCNs were named GCN278, GCN53, and GCN40: 278, 53, and 40 being the number of DEGs in the respective network.

Table 1 and 2 summarise the results of GO and KEGG analysis of the entire 6 groups of DEGs, respectively. The result of GO for the Type 2 group was excluded from Table 2, as there were no significant terms.

GSEA was conducted on the GO and KEGG databases and proceeded except for 21 dpi which had a small number of DEGs observed. In the results of 4 dpi, the highest score of GO and KEGG terms for upregulation was related to the steroid biosynthesis. Additionally, the keratinization-related term was also highly significant at 4 dpi. In the case of the downregulation, only the ribosome term was identified at 4 dpi. The results of the 7 dpi KEGG database presented the functions related to the nutritional metabolism and intracellular ribosome for upregulation.

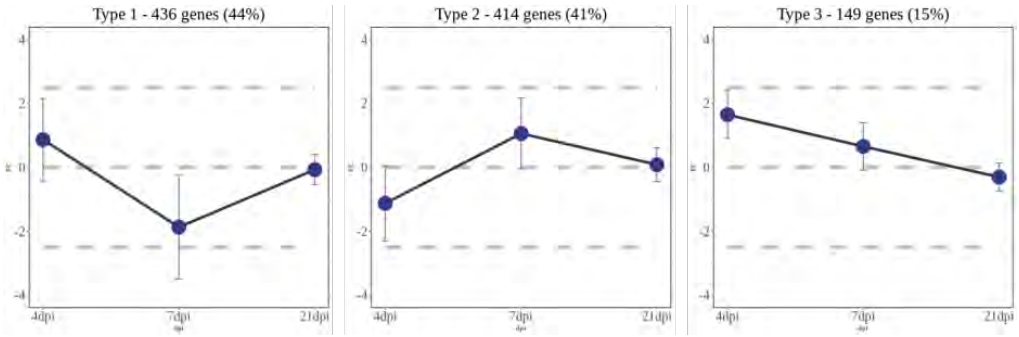


Figure 1. Three types of differentially expressed gene clusters separated based on gene expression patterns.

Table 1. GO analysis results of gene groups clustered with two different methods: expression pattern and gene co-expression. GO terms of the Type 2 were omitted.

	GO term	Count		GO term	Count
Type 1 cluster	Lipid metabolic process	12	GCN278 cluster	Signaling receptor activator activity	9
	Oxidation-reduction process	8		Hormone activity	5
	Homeostatic process	10		Cellular homeostasis	9
	Regulation of biological quality	16		Antibiotic metabolic process	4
	Regulation of hormone levels	5		Epithelium development	8
Type 3 cluster	Regulation of response to external stimulus	6	GCN53 cluster	Guanyl nucleotide binding	3
	Cytokine secretion	3		Nucleotide binding	4
	Response to organic substance	8		Defense response to virus	4
	Inflammatory response	4		Immune system process	6
	Innate immune response	3		Positive regulation of cytokine production	2
			GCN40 cluster	Protein tyrosine kinase activity	2
				Positive regulation of immune system process	3
				Antigen receptor-mediated signaling pathway	2
				Leukocyte migration	2
				Cell-cell adhesion	2

Table 2. KEGG analysis results of gene groups clustered with two different methods: expression pattern and gene co-expression.

	KEGG term	Count		KEGG term	Count
Type 1 cluster	Steroid biosynthesis	5	GCN278 cluster	Galactose metabolism	6
	Steroid hormone biosynthesis	4		Oxidative phosphorylation	8
	Cytokine-cytokine receptor interaction	10		Phagosome	8
	Linoleic acid metabolism	3		Steroid hormone biosynthesis	4
	Glycolysis / Gluconeogenesis	4		PPAR signaling pathway	4
Type 2 cluster	Drug metabolism - cytochrome P450	5	GCN53 cluster	Influenza A	7
	Metabolism of xenobiotics by cytochrome P450	5		Herpes simplex virus 1 infection	6
	ABC transporters	5		RIG-I-like receptor signaling pathway	4
	PPAR signaling pathway	6		NOD-like receptor signaling pathway	5
	Intestinal immune network for IgA production	4		Toll-like receptor signaling pathway	3
Type 3 cluster	Pentose and glucuronate interconversions	4	GCN40 cluster	Cytokine-cytokine receptor interaction	5
	Glycerolipid metabolism	3		Ether lipid metabolism	3
	Folate biosynthesis	2		alpha-Linolenic acid metabolism	2
	Phagosome	4		Glycerophospholipid metabolism	3
	ECM-receptor interaction	3		Intestinal immune network for IgA production	2

Discussion

Functional analysis results of Type 1 suggested that lipid synthesis and metabolism were activated at 4 dpi. According to Galluzzi *et al.* (2017) and Liu *et al.* (2021), endoplasmic reticulum (ER) stress is caused by *E. tenella* infection and induces the unfolded protein response (UPR) within cells. ER stress and excessive UPR disrupt cell homeostasis; therefore, lipid regulation fails. Furthermore, under the influence of ER stress, the cells tend to reduce the number of the ribosome for lessening protein synthesis and transport

(Han *et al.*, 2016, Hotamisligil, 2010). The downregulation of the ribosome term at 4 dpi in GSEA was verified for this reason. In 7 dpi, however, the ribosome term was again upregulated, possibly due to the relative relief of ER homeostasis at the particular time.

The genes belonging to the Type 3 group and GCN53 were characterized with innate immune response-based functions. It suggested that the natural immune system was initially activated following the timely expression pattern of the Type 3 group genes. According to the GCN53 KEGG result, the recognition receptor signalling pathways, as a considerable role for defence in *E. tenella* infection, were associated with the regulation of inflammation (Franchi *et al.*, 2009, Haunshi *et al.*, 2017).

Cytokine-cytokine receptor interaction term had the most significant number of genes in KEGG results of the Type 1 gene group. Further, the genes belonging to the term (*IL13RA2*, *CXCL13*, *CXCR5*, *GDF15*, *IL7R*, *IL12RB2*, *BMP7*, *TNFRSF11B*, *IL22*, and *IFNGR1*) were reported to be related with not only the inflammatory pathway but also activation of immune response (Wischhusen *et al.*, 2020, Yun *et al.*, 2000, Zou *et al.*, 1997). According to the expression patterns of the Type 1 genes, the initial innate immune responses were to defend chicken itself following downregulation of those genes was one of the mechanisms for regulating inflammation.

References

Blake D.P. and Tomley F.M. (2014) Trends Parasitol. 30(1):12-19. https://doi.org/10.1016/j.pt.2013.10.003

Franchi L., Warner N., Viani K., and Nuñez G. (2009) Immunol. Rev. 221(1):106-128 https://doi.org/10.1111/j.1600-065X.2008.00734.x

Georgieva N.V., Koinarski V., and Gadjeva V. (2006) Vet. J. 172(3):488-492. https://doi.org/10.1016/j.tvjl.2005.07.016

Galluzzi, L., Diotallevi A., and Magnani M. (2017) Future. Sci. OA 3(3):FSO198. https://doi.org/10.4155/fsoa-2017-0020

Haunshi S., Burramsetty A.K., Kannaki T.R., Ravindra K.S.R., and Chatterjee R.N. (2017) Polt. Sci. 96(9):3052-3057 https://doi.org/10.3382/ps/pex113

Han J. and Kaufman R.J. (2016) J. Lipid Res. 57(8):1329-1388 https://doi.org/10.1194/jlr.R067595

Hotamisligil G.S. (2010) Cell 140(6):900-917 https://doi.org/10.1016/j.cell.2010.02.034

Liu B., Ma X., and Cai J. (2021) Front. Vet. Sci. 8(1):756 https://doi.org/10.3389/fvets.2021.688684

Wischhusen J., Melero I., and Fridman W.H. (2020) Front Immunol. 11(1):951 https://doi.org/10.3389/fimmu.2020.00951

Yun C., Lillehoj H., and Lillehoj E. (2020) Dev. Comp. Immunol. 24(2-3):303-324 https://doi.org/10.1016/j.dci.2011.06.007

Zou J., Presky D.H., Wu C.Y., and Gubler U (1997) J. Biol. Chem. 272(9):6073-6077 https://doi.org/10.1074/jbc.272.9.6073

595. Genetic diversity of *Eimeria tenella* apical membrane antigen 1 in chickens from KwaZulu-Natal, South Africa

A.J. Fatoba[1], O.T. Zishiri[1], D.P. Blake[2], S.O. Peters[3], J. Lebepe[4], S. Mukaratirwa[4,5] and M.A. Adeleke[1*]

[1]Discipline of Genetics, School of Life Sciences, University of KwaZulu-Natal, Westville, P/Bag X54001, Durban 4000, South Africa; [2]Department of Pathobiology and Population Sciences, Royal Veterinary College, Hawkshead Lane, Hertfordshire AL9 7TA, United Kingdom; [3]Department of Animal Science, Berry College, Mount Berry, GA 30149, USA; [4]Biodiversity and Evolutionary Biology, School of Life Sciences, University of KwaZulu-Natal, Private Bag X54001, Durban 4000, South Africa; [5]One Health Center for Zoonoses and Tropical Veterinary Medicine, Ross University School Veterinary Medicine, Basseterre, St. Kitts, West Indies; adelekem@ukzn.ac.za

Abstract

Recombinant vaccine which contains immunoprotective antigens such as apical membrane antigen 1 (AMA1) has shown promising prospect in curtailing chicken coccidiosis- a global burden in poultry industry. However, understanding the naturally occurring diversity among these antigens is crucial for vaccine effectiveness. The present study aimed to determine diversity of AMA1 of *Eimeria tenella* isolates from broiler and free-range chickens in KwaZulu-Natal province of South Africa. Polymerase chain reaction (PCR) amplification of *E. tenella* AMA1 (EtAMA1) was carried out from both broilers and free-range chicken fecal samples respectively. Sequence analysis revealed low levels of diversity in broiler chickens compared to AMA1 sequences in the free-range chickens. Similarly, haplotype network analysis identified 5 haplotypes to be unique from the existing ones. This study provides baseline information on the diversity of EtAMA1 isolates from chicken farms in KwaZulu-Natal province of South Africa and is crucial to designing effective vaccines.

Introduction

The annual global productivity loss due to chicken coccidiosis has been a major concern in the poultry industry and has triggered the continuous design of effective control measures (Blake *et al.*, 2020). Control of coccidiosis through live vaccines which entails the use of live *Eimeria* parasite either in attenuated or non-attenuated form has been effective in the control of this disease. However, the cost of production of live vaccines has shifted attention to other alternatives such as recombinant vaccines. Recombinant vaccines have become popular especially with the discovery of different partially immunoprotective antigens and increase in the number of vaccine delivery systems that can be used in chickens (Clark *et al.*, 2017). The effectiveness of recombinant vaccines relies on low genetic variability among these antigens (Clark *et al.*, 2017) and as such, exploring the genetic diversity among these antigens becomes imperative. Notably among these vaccine antigens used in the design of recombinant vaccines are apical membrane antigen 1 (AMA1) and immune and mapped protein (IMP-1) (Tang *et al.*, 2019).

Apical membrane antigen 1 (AMA1) is a highly conserved microneme protein among apicomplexans. However, in *Eimeria* parasites, there is dearth of information on the diversity of AMA1 in South Africa. This study therefore aimed to determine antigenic diversity among AMA1 of *E. tenella* isolates from broiler and free-range chickens in KwaZulu-Natal province of South Africa.

Materials & methods

Sample collection and processing. Three hundred chicken fecal samples were randomly collected from January to October 2018 from 12 broiler farms together with 42 samples from indigenous chickens from different localities. Samples were collected following the procedure described by Kumar *et al.* (2014) after

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_595, © A.J. Fatoba *et al.* 2022

which they were stored at 4 °C until further use. The fecal samples were screened for *Eimeria* oocyst using fecal floatation method. In both broiler and free-range chickens, one hundred and two (102) and twenty-five (25) samples respectively had *Eimeria* oocyst. From the positive samples, oocysts per gram (OPG) of feces were determined using a McMaster counting chamber following a standard protocol. Samples with OPG greater or equal to 250 OPG in broiler (n=90) and free-range (n=19) were selected for DNA extraction.

DNA extraction and PCR amplification of Eimeria tenella. QiaAmp DNA stool Mini Kit (Germany) was used for the extraction of genomic DNA based on the manufacturer's protocol with minor modifications. Agarose gel (1.5%) and Nanodrop™ 1000 spectrophotometer (Thermo Scientific, USA) at 260 nm absorbance were used to check the quality and concentration of the extracted DNA. PCR protocol targeting internal transcribed spacer 1 region of the genome was carried out. After the detection of *E. tenella*, primers were designed with primer3 software to amplify EtAMA1 and the amplicons were sent for sequencing at Inqaba Biotech (South Africa).

Sequence analysis, polymorphism and statistical analyses. Sequences were edited, trimmed and consensus sequences were generated using BioEdit version 7.0. Sequences were further compared with published EtAMA1 sequences in the GenBank using BLAST tool and multiple sequence alignment was done using ClustalW programme in MEGA version 6.0. Nucleotide diversity was calculated using both the number of mutation and the sliding window method (100bp length and step size of 25bp). The number of haplotypes (*H*), haplotype diversity (*Hd*), average number of nucleotide difference (k), and polymorphic sites were also calculated with DnaSP 5.10.01 software. Haplotype median-joining network of the AMA1 sequences in this study and other sequences in the GenBank were constructed with Network software version 5.0 (http://www.fluxus-engineering.com). The ratio of non-synonymous to synonymous substitution (*dN/dS*) was calculated using Nei-Gojobori method (Jukes-Cantor correction) and significance was tested with Z-test using MEGA version 6.0. Departure from neutrality was evaluated using Tajima's D, Fu and Li's D* and Fu and Li's F* based on the following: (1) the number of mutations at ($P<0.05$); (2) sliding window of 100bp with a step size of 25bp.

Results

Comparison of genetic diversity between EtAMA1 sequences of broiler chicken in this study and Genbank sequences. As shown in Table 1, comparing the EtAMA1 sequences of broiler in this study with the retrieved EtAMA1 sequences from the Genbank generated 135 polymorphic sites. Nucleotide and haplotype diversities of 0.458 and 0.795 were obtained respectively from this comparison. Also, Tajima's D (1.869; $P>0.10$) showed non-significant balancing selection while Fu and Li's D** (1.561; $P<0.02$) and Fu and Li's F** (1.749; $P<0.02$) all had significant positive values which are indication of balancing selection. On the contrary, a synonymous to non-synonymous mutation (*dN/dS*) of 0.889 was obtained, indicative of a negative or purifying selection on the EtAMA1 sequences. Further analysis using sliding window technique of 100bp and 25bp step size showed all the three (3) parameters of selection with significant positive values at some nucleotide positions.

Comparison of genetic diversity between EtAMA1 sequences of free-range chicken in this study and Genbank sequences. Nucleotide and haplotype diversities of 0.367 and 0.727 respectively were obtained when the sequences of EtAMA1 from free-range chickens were compared with sequences from the Genbank. The three selection parameters: Tajima's D (-0.376; $P>0.10$); Fu and Li's D (0.754; $P>0.01$) and Fu and Li's F (0.488; $P>0.01$) all had negative and positive values but were non-significant (Table 1). Similarly, a contrasting positive or balancing selection was indicated by the synonymous to non-synonymous mutation value of 1.282. Sliding window techniques of 100bp and 25bp step size also showed the gross insignificance of values of these three parameters.

Table 1. Comparison of genetic diversity between EtAMA1 sequences from broiler and free-range chicken in this study and Genbank sequences.[1]

Chicken type	Antigen	N	S	$\pi \pm$ S.D	*dN/dS*	k	*H*	*Hd* $\pm$ S.D	Tajima's D	Fu and Li's D*	Fu and Li's F*
Broiler	EtAMA1	13	134	0.458±0.050	0.889	62.6	7	0.795±0.109	1.869	1.561**	1.749**
Free-range	EtAMA1	11	140	0.367±0.076	1.282	50.3	6	0.727±0.144	-0.376	0.754	0.488

[1] N: number of sequences tested; S: segregating or variable sites; π: nucleotide diversity based on jukes cantor; k: average number of nucleotide difference; H: number of haplotypes; Hd: haplotype diversity.

Discussion

In this study, higher diversity parameters were observed in EtAMA1 sequences from indigenous chickens compared to broilers chickens when the sequences were analyzed with GenBank sequences. This is in contrast to the report of Blake *et al.* (2015) who reported low levels of diversity of EtAMA1 sequences from different geographical regions. The scavenging life-style of indigenous chickens exposes them to myriad of diseases (Bettridge *et al.* 2014). This could contribute to the high level of diversity of EtAMA1 sequences among indigenous chickens in this study compared to broiler chicken production with intensive management system. In addition, analysis of selection signatures for EtAMA1 sequences in this study and those retrieved from Genbank were also performed and non-synonymous to synonymous substitution parameter (*dN/dS*) values of 0.889 and 1.282 were obtained for both broiler and indigenous chickens respectively which indicate both purifying/negative and positive selections. Due to this contrasting result of diversity parameters, *dN/dS* has been criticized as poor evaluator of selection signatures among *Eimeria* vaccine candidates. Similarly, considering the insignificant values of neutrality parameter (Tajima's D) for EtAMA1 in this study, EtAMA1 is under neutral selection. Although few significant values were obtained for Fu and Li's F* and Fu and Li's D* neutrality parameters in this study, we are of the opinion that this contrasting result could be due to the partial length of EtAMA1 sequenced in this study compared to the full-length of EtAMA1 reported by Blake *et al.* (2015). In conclusion, this study confirmed low diversity among EtAMA1 sequences in the study area which potentially indicated the reliability of this as a vaccine candidate. However, further studies on AMA1 in this region are still necessary possibly with larger sample size and full-length sequences of EtAMA1 from different provinces of South Africa.

References

Bettridge J.M., Lynch SE., Brena M.C., Melese K., Dessie T. *et al.* (2014) Prev. Vet. Med. 117(2):358-66. http://dx.doi.org/10.1016/j.prevetmed.2014.07.002

Blake D (2015) Vet Parasitol 212: 68–74. https://doi.org/10.1016/j.vetpar.2015.05.007

Blake DP, Knox J, Dehaeck B, Huntington B, Rathinam T *et al.* (2020). *Vet Res* 51**,** 115 https://doi.org/10.1186/s13567-020-00837-2

Clark EL, Tomley FM, Blake DP (2017) Trends in parasitology 33(3):231-41. http://dx.doi.org/10.1016/j.pt.2016.08.007

Kumar S, Garg R, Moftah A, Clark EL, Macdonald SE, *et al.* (2014) Vet Parasitol. 199: 24-31. http://dx.doi.org/10.1016/j.vetpar.2013.09.026

Tang X, Wang C, Liang L, Hu D, Zhang S, *et al.* (2019) Parasites & vectors 12(1): 347. https://doi.org/10.1186/s13071-019-3605-6

596. Chicken quantitative traits follow the omnigenic model

P. Davoodi, A. Ehsani*, R. Vaez Torshizi and A.A Masoudi

Tarbiat Modares University, 14445-336, Tehran, Iran; alireza.ehsani@modares.ac.ir

Abstract

To comprehend the association structure between genome and complex traits, we investigated the different aspects of the novel omnigenic model for 23 traits in six trait categories in F2 cross-bred chickens from the perspectives of association signals, genomic heritability, overlapped genes, and predicted gene networks. In the current study, the high SNP-based heritability was obtained when the genetic architecture of traits was predominantly additive and non-significant markers contributed greater to explaining heritability. Moreover, we distinguished overlapping genes among apparently distinct traits indicating that traits are not inherited independently. The findings showed that the GWAS signals (core genes) were few and on average with greater beta effect sizes. Conversely, peripheral genes with non-significant effects appeared highly with co-expression connectivity in predicted gene networks. All results together allowed us to experimentally infer the highlighted role for peripheral genes in chicken's complex traits which supports the establishment of the 'omnigenic' model.

Introduction

For a long time now, the main goal of animal genetics has always been genetic progress in economic phenotypes alongside discover the relationship between genotype and phenotype. The power of genome-wide association studies (GWAS) to detect an association between genomic markers and a trait is strongly limited by the experimental sample size, trait's genetic architecture, and stringent significance threshold. Therefore, the genetic association with quantitative traits seems to be undetectable by GWAS (Visscher *et al.* 2017). There is still a complex maze from sequence to biological consequence besides the environment, which can cause phenotypic variation and make this involute route even more complicated. Because the two main models of infinitesimal and finite loci, despite their advantages in genetic studies, have some inconsistencies and shortcomings (Visscher *et al.* 2012, Nelson *et al.* 2013, Turelli 2017, Visscher *et al.* 2017), the researchers attempt to answer the demand for a new model, namely the omnigenic model, for comprehensive genetic interpretation (Boyle *et al.* 2017). According to this model, there is a small number of 'core genes' having a direct biological effect on a particular trait, and also there is a large number of regulatory genes that contribute to the expression of quantitative traits. Because the 'peripheral genes' are much greater than the number of core genes affecting a trait they explain more fractions of heritability (Boyle *et al.* 2017, Liu *et al.* 2019, Mathieson 2021). In recent years, the omnigenic model has been studied in human disease (Peedicayil and Grayson 2018, Rammos *et al.* 2019) and on the behavioural traits of Drosophila Melanogaster (Zhang *et al.* 2019). However, no results of the omnigenic model have been so far reported for farm animals. Therefore, the present study aimed to investigate whether the distinct groups of quantitative traits in broiler chickens follow the omnigenic model.

Materials & methods

Up to 312 chickens for different traits in the F2 population descended from reciprocal crosses between the Arian line and Iranian native Urmia chickens were used as the experimental population. All congruent approaches of GWAS, SNP-based heritability, heritability partitioning, finding overlapped genes, and gene network prediction were performed to establish if there was an omnigenic model present across the different groups of quantitative traits including growth-related traits, feed efficiency-related traits, immune-related traits, blood parameters, carcass-related traits, and plumage colour. GWASs were performed using the mixed linear model with the leave one chromosome out (LOCO) approach. The model includes the

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_596, © P. Davoodi *et al.* 2022

overall mean, covariate variables of sex and first principal component and SNP effects (fixed effects), polygenic effects (random effects), and residual effects (random effects). For estimating the phenotypic variance proportion explained by all SNPs, the Genome-based restricted maximum likelihood (GREML) method was applied. GWASs and estimation of SNP-based heritability and heritability partitioning were performed on GCTA (Yang *et al.* 2011, Stanton-Geddes *et al.* 2013). SNP-based heritability was partitioned into four groups of significant, suggestive, non-significant, and all SNPs to determine the contribution of each group to explain phenotypic variation. Overlapped genes were discovered by the GALLO R package by using suggestive genes for each trait and accounting for pleiotropic genes (Fonseca *et al.* 2020). Integrative network prediction was performed on Cytoscape through GeneMANIA to discover different sources of gene-gene connections among significant genes in six groups of traits (Montojo *et al.* 2014).

Results

GWASs performed on 23 chicken traits, and the 44 annotated genes related to the significant markers were used for further studies. The SNP-based heritability for 23 traits and the heritability contribution explained by significant, suggestive, non-significant, and all SNPs are shown in Figure 1. The results indicated that non-significant SNPs can explain the majority of genomic heritability in each trait (orange line in Figure 1). In contrast, significant SNPs resulting from GWASs in all traits explained tiny parts of heritabilities (red line in Figure 1). Furthermore, integrated network analysis of 44 coding genes related to six groups of chicken traits (Figure 2, Left) has consisted of 51.61% of co-expression, 16.62% of genetic interactions, 16.17% of physical interactions, 7.17% of co-localization, and 6.72% of pathways. Co-expression was the dominant type of connection in this network. In the predicted gene network only black circles with the first letter of the trait categories are defined as core genes for the certain trait and other circles act as peripheral genes. Also, the overlapped genes insignificant level depicted in this network (Figure 2, Left) by including the first letter of traits on black circles (gene) so the black circles that possess more than one letter indicate, that gene affects more than one trait. Moreover, we distinguished overlapping or pleiotropic genes among apparently distinct traits with a heat map matrix of shared genes using suggestive genes (Figure 2, Right), which indicates that traits are not inherited independently and common genes or variants can affect them simultaneously.

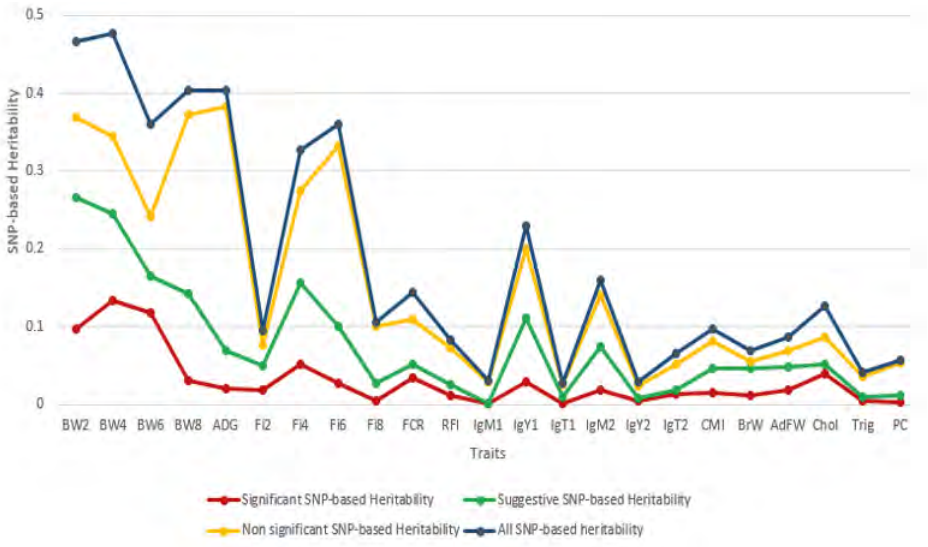


Figure 1. Heritability is explained by significant, suggestive, non-significant, and all SNPs.

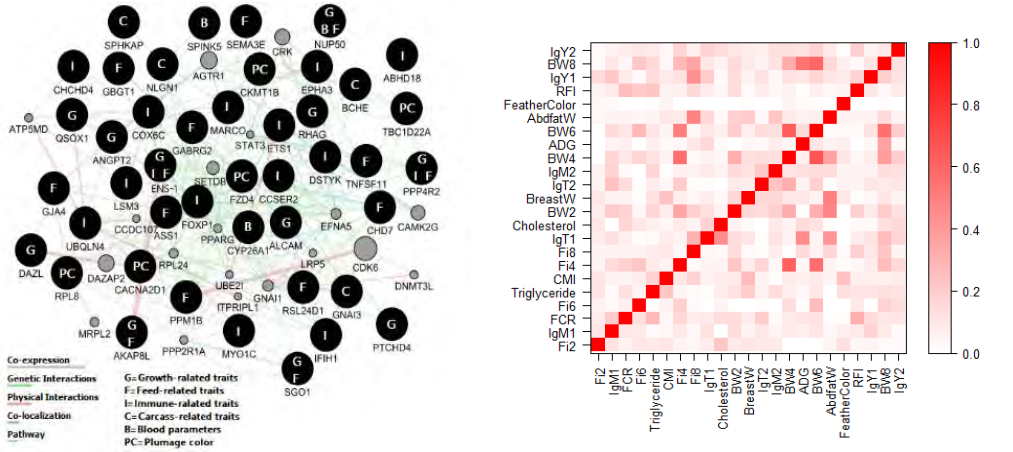


Figure 2. Left; An integrated network of 44 genes related to six groups of chicken traits. Right; the heat map matrix of overlapped genes among 23 traits in suggestive gene level (2,000 genes).

Discussion

From the results of the heritabilities partitioned into significant, suggestive, non-significant, and all SNPs, non-significant SNPs explained the highest phenotypic variation, which is in agreement with previous results (Yang *et al.* 2010, Wood *et al.* 2014). The overlapped genes in different trait groups support the evidence of pleiotropy among traits, implying that traits are not inherited independently. According to GWAS results, the core genes (related to the significant SNPs) were few with additive and traceable effects through GWASs, but they explained only a small part of the trait' heritability because they were small in number. On the other hand, peripheral genes (genes annotated for non-significant variants) were not strictly detectable individually due to their small effect sizes in genomic scans, but they were very large in number (perhaps all non-significant variants in GWAS). Therefore, their cumulative effects were much greater than the effect of significant variants (Figure 1). Furthermore, the gene network with a dominant level of co-expression in the current study and SNP-based heritability estimation revealed that the omnigenic model has better conformity to the genetic architecture in chicken traits than the polygenic model. Surprisingly by adding any gene set, even randomly selected genes, we achieved a connected gene network (not shown). It has been interesting to obtain a set of agreements with other researchers (Boyle *et al.* 2017, Shen and Shi 2017) in the light of the omnigenic theory in chickens' quantitative traits, which offers the bipartite genetic involvement in the development of complex traits, including the direct effects of core genes and the indirect effects of peripheral or helping genes. Overall, low and high contribution of significant and non-significant SNPs respectively on heritability, the high level of co-expression in a gene network, and the presence of overlapped genes experimentally support the establishment of the omnigenic model which highlights the role of peripheral genes in chicken's complex traits.

References

Boyle E. A., Li Y. I. and Pritchard J. K. (2017) Cell 169(7): 1177-1186. https://doi.org/10.1016/j.cell.2017.05.038

Fonseca, P. A. S., Suarez-Vega, A., Marras, G. and Canovas, A. (2020) GigaScience 9(12):149. https://doi.org/10.1093/gigascience/giaa149

Mathieson, I. (2021) The American Journal of Human Genetics 108(9): 1558-1563. https://doi.org/10.1016/j.ajhg.2021.07.003

Montojo, J., Zuberi, K. Rodriguez, H. Bader G. D., and Morris, Q. (2014) F1000Res 3: 153. https://doi.org/10.12688/f1000research.4572.1

Nelson, R. M., M. E. Pettersson and O. Carlborg (2013) Trends in Genetics 29(12): 669-676. https://doi.org/10.1016/j.tig.2013.09.006

Peedicayil, J. and Grayson, D. R. (2018) Journal of theoretical biology 443: 52-55. https://doi.org/10.1016/j.jtbi.2018.01.027

Rammos, A., Gonzalez, L. A. N., Weinberger, D. R., Mitchell, K. J., Nicodemus K. K and *et al.* (2019) Neuropsychopharmacology 44(9): 1562-1569. https://doi.org/10.1038/s41386-019-0410

Shen, L. and Shi, Q. (2017) Journal of Psychiatry and Brain Science 2(6): 3. https://doi.org/10.20900/jpbs.20170019

Stanton-Geddes, J., Yoder, J. B., Briskine, R., Young, N. D., and Tiffin, P. (2013) 4(12): 1151-1158. https://doi.org/10.1111/2041-210X.12129

Turelli, M. (2017) Theoretical Population Biology 118: 46-49. https://doi.org/10.1016/j.tpb.2017.09.003

Visscher, P. M., Wray, N. R., Zhang, Q. Sklar, P., McCarthy, M. I., Brown, M. A., and Yang, J. (2017). American Journal of Human Genetics 101(1): 5-22. https://doi.org/10.1016/j.ajhg.2017.06.005

Wood, A. R., Esko, T., Yang, J., Vedantam, S. Pers, T. H., et al (2014) Nature Genetics 46: 1173. https://doi.org/10.1038/ng.3097

Yang, J., B. Benyamin, B. P., McEvoy, S., Gordon, A. K., Henders, D. R., and *et al* (2010) Nature Genetics 42: 565. https://doi.org/10.1038/ng.608

Yang, J., Lee, S. H., Goddard, M. E. and Visscher, P. M. (2011). American Journal of Human Genetics 88(1): 76-82. https://doi.org/10.1016/j.ajhg.2010.11.011

Zhang, W., Reeves, R. G. and Tautz, D. (2019) bioRxiv: 639955. https://doi.org/10.1101/639955

597. Association of polymorphism of *MAP3K8* with egg production traits of White Muscovy ducks (*Cairina moschata*)

S.F. Bello[1], H.Xu[1], S. Zhang[1], K.D. Aloryi[2], R.O. Ahmed[3], A.A. Bahareldin[1], A.C. Adeola[4], M.A. Popoola[5], W. Zhu[6], D. Zhang[1,6], X. Zhang[1], C. Ji[6] and Q. Nie[1,6*]

[1]Department of Animal Genetics, Breeding and Reproduction, College of Animal Science, South China Agricultural University, Guangzhou 510642, Guangdong, China; [2]State Key Laboratory for Conservation and Utilization of Subtropical Agro-Bioresources, South China Agricultural University, Guangzhou, 510642, Guangdong, China; [3]Department of Animal and Avian Sciences, University of Maryland, College Park, MD 20742, USA; [4]State Key Laboratory of Genetic Resources and Evolution & Yunnan Laboratory of Molecular Biology of Domestic Animals, Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming 650223, China; [5]Federal College of Animal Health and Production Technology, Moor Plantation, Apata, Ibadan, Nigeria; [6]Wens Foodstuff Group Co. Ltd., Yunfu 527400 Guangdong, China; nqinghua@scau.edu.cn

Abstract

Prior studies on the transcriptomes of the hypothalamus and ovary revealed that *MAP3K8* is one of the candidate genes that might affect egg production in White Muscovy ducks. However, the functional role of this gene in the egg production traits of Muscovy ducks remains unknown. We first examined the relative expression profile and then identified variants within *MAP3K8* that may affect egg production traits (age at first egg (AFE), number of eggs at 300 days (N300D), and number of eggs at 59 weeks (N59W)) in 507 ducks. A total of seven mutations were identified and association analysis of the single nucleotide polymorphism (SNPs) with egg production traits showed variations at g.148303340 G>A, g.148303208 T>C, g.148290065 A>G, and g. 148290061 A>T of *MAP3K8* with number of eggs at 59weeks. This study sheds new knowledge on the molecular mechanism that might be regulating egg production traits in Muscovy ducks.

Introduction

Muscovy duck (*Cairina moschata*) was first domesticated in South America (Stahl 2005; Stahl 2008) and later introduced to Asia, Africa, and Europe (Rojas and Arias-Sosa 2021). In China, black and white Muscovy ducks are commonly raised. Previous studies on the transcriptomes of the hypothalamus and ovarian revealed several pathways related to egg production, including MAPK signaling, ECM-receptor interaction, and Apelin signaling pathway. These pathways were identified using the highest and lowest egg-producing White Muscovy ducks (Bello *et al.* 2021). However, the functional roles of *MAP3K8* on the tissues and egg production traits of Muscovy ducks remain unknown. *MAP3K8* gene has been assumed to be involved in the MAPK signaling pathway which in turn participates in LH and FSH-induced steroidogenesis in granulosa cells (Seger *et al.* 2017). *MAP3K8* was predicted as a target gene of miRNA-509-3p that promotes the secretion of oestradiol (Huang *et al.*, 2016), which is an important hormone in egg ovulation. Interestingly, MAP3K8 was found to be associated with the MAPK pathway which was one of the enriched pathways that affect egg production in our previous study (Bello *et al.* 2021). Although, the functional roles of *MAP3K8* in the tissues and egg production traits of white muscovy ducks remain unknown. Thus, we first examined the mRNA expression of *MAP3K8* in eight tissues and also evaluated the polymorphism of *MAP3K8* gene and its association with egg production traits to identify genetic markers for the improvement of egg production of White Muscovy ducks.

Materials & methods

Experimental animals. Same experimental birds used in our previous study (Bello *et al.* 2021).

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_597, © S.F. Bello *et al.* 2022

Blood sampling and DNA extraction. 2 ml of blood was collected from individual Muscovy ducks through their wing-web into EDTA tubes at 56 weeks of age. Genomic DNA was extracted from the 10µl whole blood of the individual Muscovy ducks using E.Z.N.A.™ NRBC Blood DNA kit (OMEGA, Bio-tek, USA). All DNA samples were diluted to a working concentration of 50 ng/µl and stored at -20 °C for further use.

Collection of tissue samples. Based on the number of the egg at 59weeks, the four lowest (LP) and four highest producing (HP) Muscovy ducks within the same egg number from each group were selected for tissues collection. The eight selected ducks were slaughtered, and eight different tissues, including the cerebrum, cerebellum, etc. were collected.

Primer design, PCR amplification, SNP detection, PCR amplification of SNP sites and genotyping by DNA sequencing. *MAP3K8* has 10 exons and is located on chromosome 2 (148288602-148307555bp) with the accession number 101791769 of the *Anas platyrhynchos* (mallard) genome. Seven pairs of primers were designed to amplify different regions of this gene. DNA mixed pool was constructed with 10µl DNA sample each of 20 randomly selected individuals with equal concentration. The PCR products were sequenced directly using the ABI-3730XL DNA analyzer (USA) by Sangon Biotech Company (Guangzhou, China). Trimming and alignment of sequences, and SNP discovery were conducted using SnapGene® 4.3.6 software.

RNA extraction and synthesis of cDNA. Total RNA was extracted from the eight different tissue selected using MagZol Reagent (Magen, Guangzhou, China). Then, total RNA samples were reversely transcribed to cDNA using HiScript® II RT SuperMix for qPCR (+gDNA wiper) (Vazyme Biotech Co, Ltd. Cat# R223-01).

Primers and quantitative real time-PCR (qRT-PCR). Primers were designed according to the NCBI database sequences of the *MAP3K8* with Accession number XM_027450562.2 using Primer premier version 5.0 software (Applied Biosystems). The MonAmp™ ChemoHS qPCR kit (Monad, Suzhou, China, Cat# MQ00401) was used to perform qRT-PCR. The qRT-PCR reaction was performed in CFX96 (Bio-Rad, United States).

Statistical analysis. Association analyses of single nucleotide polymorphisms with egg production traits of 507 Muscovy laying ducks were analyzed using the GLM procedure in SPSS 21.0. The significant differences between the average expression of the LP and HP tissues were examined with a t-test. Differences between the genotypes were analyzed using SPSS 21.0 (IBM, Chicago, IL, United States). The difference with P-value≤0.05 was considered significant in analyses. The model below was adopted: $Y_{ijkl} = \mu + K_i + L_j + M_k + e_{ijkl}$, where: Y_{ijkl} = phenotype of the egg production traits, such as age at first egg (days), number of eggs at 300 days, and number of eggs at 59 weeks; μ = population mean; K_i= fixed effect of the SNP genotypes; L_j = fixed effect of the Sire; M_k = fixed effect of the Dam; e_{ijkl}= random error.

Results

Sequences and variation at AKT3 gene. Amplification of *MAP3K8* gene was done using seven pairs of primers to determine the specific locus carrying the mutation. Further, only two of the designed primers showed seven significant SNP sites when aligned with the reference genome of *Anas platyrhynchos* (mallard). Three variation sites were identified at the exon 3 (g.148303340 G>A, g.148303241 G>C and g.148303208 T>C) while four variations were detected at the intron 9 (g.148290274 T>C, g.148290240 G>A, g.148290065 A>G and g.148290061 A>T). In addition, we did not observe any synonymous amino acid substitution at these SNPs sites. These variations were further used to test their association with egg production traits in 507 Muscovy ducks.

Association analyses of AKT3 with egg production traits

The association between the seven *MAP3K8* polymorphic sites and egg production traits (age at first egg laid, the number of eggs at 300 days of age, and the number of eggs at 59 weeks of age) were analyzed using 507 ducks samples. The least-square mean and standard error of mean (SEM) values of genotypes of each significant SNPs for egg production traits are shown in Table 1. There is a significant difference (P>0.05) in the genotypes of g.148303340 G>A, g.148303208 T>C, g.148290065 A>G, and g.148290061 A>T with the number of eggs at 59 weeks (Table 1).

Relative expression level of MAP3K8 gene in the tissues of White Muscovy ducks. The pattern of the mRNA level of *MAP3K8* across the ovary, oviduct, and hypothalamus. In the three tissues, the mRNA expression of the *MAP3K8* gene in HP ducks is significantly (P<0.01 or P<0.5) higher than in the LP with the *MAP3K8* expression highest in the oviduct among the three (Figure 1a). Cerebrum, cerebellum, and pituitary tissues had similar expression trends of the *MAP3K8* gene in the HP ducks, while the pituitary of the HP ducks showed the highest expression levels of the *MAP3K8* gene (Figure 1b). We found significantly higher expression of *MAP3K8* in HP's subcutaneous and abdominal fats compared to the LP counterpart (P<0.001). Although, the abdominal fat of the LP had a lower expression of *MAP3K8* when compared to its subcutaneous fat (Figure 1c). Together, this result suggests that there is variation in the expression of *MAP3K8* in the eight selected tissues of HP and LP.

Discussion

The study of genes that are linked to economic traits is one of the important aspects of animal breeding to aid improvement. Egg production is a major economic trait of poultry species (Bao *et al.* 2021) which declines with ovarian aging (Hao *et al.* 2021). In this study, pituitary has the highest mRNA level. It has been reported that MAP3K8 is highly expressed in the brain (Banerjee *et al.* 2006). g.148303340 G>A, g. 148303208 T>C, g. 148290065 A>G, and g. 148290061 A>T were significantly (P<0.05) associated with number of eggs at 59 weeks.

Table 1. Association analysis of the SNPs of *MAP3K8* with egg production traits of White Muscovy ducks.[1]

SNPs	Egg production traits	Least-square mean ± SEM			*P*-value
		GG (137)	**GA (252)**	**AA (118)**	
g.148303340 G>A	AFE	195.33±0.59	194.63±0.42	195.79±0.63	0.139
	E300D	97.26±0.96^{b}	98.66±0.63^{a}	97.78±0.83^{b}	0.028
	E59W	174.00±2.56^{b}	180.29±1.59^{a}	174.72±2.81^{b}	0.049
		TT (82)	**TC (232)**	**CC (192)**	
g. 148303208 T>C	AFE	194.99±0.75	194.73±0.44	195.53±0.51	0.341
	E300D	98.16±1.24	98.58±0.65	97.50±0.71	0.389
	E59W	175.65±2.93^{b}	178.61±1.74^{a}	176.41±2.19^{ab}	0.039
		AA (68)	**AG (217)**	**GG (221)**	
g. 148290065 A>G	AFE	194.94±0.90	194.77±0.45	195.47±0.46	0.441
	E300D	99.10±1.16^{a}	98.56±0.69^{ab}	97.24±0.70^{b}	0.018
	E59W	176.37±3.01^{b}	178.28±1.82^{a}	176.48±2.01^{b}	0.062
	AFE	**AA (234)**	**AT (210)**	**TT(63)**	
g. 148290061 A>T	E300D	195.41±0.45	194.70±0.46	195.21±0.91	0.462
	E39W	97.46±0.68	98.69±0.68	98.30±1.26	0.417
		177.38±1.91^{b}	178.09±1.85^{a}	174.33±3.49^{c}	0.034

[1] SNPs- Single Nucleotide Polymorphisms. AFE- age at first egg laid; E300D- number of eggs at 300days of age; E59W- number of eggs at 59weeks of age; SEM- standard error of mean. Values within the same row with different superscripts differ significantly at P<0.05. Only SNPs with significant association with egg production data are presented in Table 1.

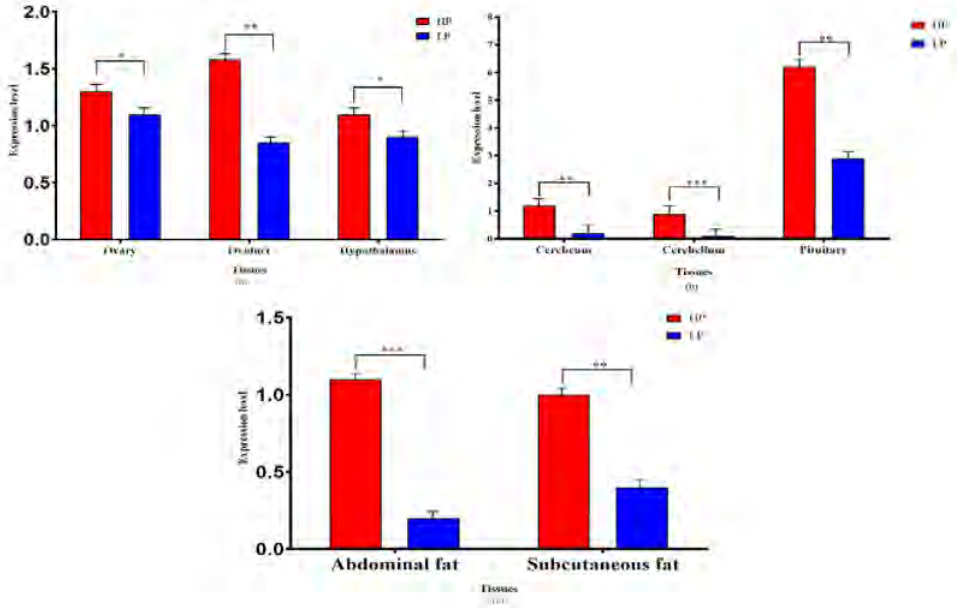


Figure 1. Expression level of the *MAP3K8* gene in (a) ovary, oviduct, and hypothalamus (b) cerebrum, cerebellum, and pituitary (c) abdominal fat and subcutaneous fat of HP and LP White Muscovy ducks. **P*<0.05 and ***P*<0.01 means there is a significant difference between the expression level of HP and LP Muscovy duck.

Moreover, g.148303340 G>A and g. 148290065 A>G were significantly (*P*<0.05) associated with number of eggs at 300 days. This is similar to a study on *FSHR* with polymorphic sites of A-227G and C-320T were associated with E33W and E59W, and AFE and E33W, respectively (Ye *et al.* 2017). *MAP3K8* might be responsible for the differences in egg production in ducks, an important finding that could be useful in duck breeding improvement programs.

References

Banerjee A., Raffi G., Martin M., Steve G. (2006). Proc Natl Acad Sci USA 103: 3274-3279. https://doi.org/10.1073/pnas.0511113103.

Bao X., Yiping S., Tao L., Shanshan Z., Lihua H. *et al.* (2021). *Genes*. https://doi.org/10.3390/genes12010057.

Bello S.F., Haiping X., Lijin G., Kan L., Ming Z. *et al.* (2021). *Poult. Sci.* 100 (9): 101310. https://doi.org/10.1016/j.psj.2021.101310.

Hao E., De-He W., Yi-Fan C., Rong-Yan Z., Hui C. *et al.* (2021). *Poult. Sci.* 100 (1): 334–47. https://doi.org/10.1016/j.psj.2020.10.005.

Huang X., Chang L., Cuifang H., Qianqing T., Riming L., *et al.* (2016). *Reproduction*. https://doi.org/10.1530/REP-16-0071.

Rojas A., Luis A. (2021). *Worlds Poult. Sci.* 77(5):565–588. https://doi.org/10.1080/00439339.2021.1921668.

Seger R., Tamar H., Revital R., Ada D., Wolfgang M. *et al.* (2017). *J Biol Chem* 292(5): 8847. https://doi.org/10.1074/jbc.A117.006852.

Stahl P. (2005). *J Archaeol.* 32(6):915–29. https://doi.org/10.1016/j.jas.2005.01.009.

Stahl P. (2008). In *The Handbook of South American Archaeology*. https://doi.org/10.1007/978-0-387-74907-5_8.

Ye Q., Jiguo X., Xinfeng G., Hongjia O., Wei L., *et al.* (2017). *PeerJ* 5(11): e4083. https://doi.org/10.7717/peerj.4083.

598. Genetic markers associated with bone quality in Rhode Island Red laying hens

M. Sallam[1*], H.A. McCormack[2], R.H. Fleming[2], P.W. Wilson[2], B. Andersson[3], M. Schmutz[3], C. Benavides[4], N. Dominguez-Gasca[4], E. Sanchez-Rodriguez[4], A.B. Rodriguez-Navarro[4], I.C. Dunn[2], D.J. De Koning[1] and M. Johnsson[1]

[1]Swedish University of Agricultural Sciences, 75651 Uppsala, Sweden; [2]Roslin Institute, University of Edinburgh, Edinburgh EH25 9RG, Scotland, United Kingdom; [3]Lohmann Breeders, 27472 Cuxhaven, Germany; [4]Departamento de Mineralogia y Petrologia, Universidad de Granada,18002 Granada, Spain; mohammed.abdallah.sallam@slu.se

Abstract

Bone weakness is a key challenge in laying hens. Detailed measurements on layers' bone combined with genotypes via GWAS analysis would identify QTLs affecting bone health. Infrared spectroscopy and thermogravimetry measurements on tibia and humerus medullary and cortical bones of 924 Rhode Island Red laying hens, were performed and combined with genotypes (57k SNP markers) for GWAS analysis. The results showed 13 SNP markers significantly ($P<10^{-5}$ to $<10^{-7}$) associated with tibia cortical lipid (chromosomes 2 and 3), humerus medullary score (chromosome 9), and tibia cortical mineral content (chromosome Z). Further analysis is needed to determine the potential correlations (LD) between these significant markers and other QTL throughout the Rhode Island Red genome.

Introduction

Most laying hens display a high tendency to suffer from bone damage (deviations or fractures), which is a major welfare challenge in the egg industry. Gregory and Wilkins (1989) reported ~30% of commercial layers with at least one bone fracture. Another recent study reported ~70% of keel bone fractures in Danish laying farms (Thøfner *et al.*, 2021), this may be due to an increase of alternative housing systems. Housing, nutrition and genetics all contribute to bone health; the genetic component is particularly relevant in White Leghorn and Rhode Island Red breeds as they both represent the main grandparents of commercial layers.

To better understand bone health in these breeds, a detailed bone phenotyping (~55 traits) study was performed, including measurements for medullary and cortical bone separately in addition to whole bone strengths, egg production, and body weights (Dunn *et al.*, 2021). The medullary measurements showed 0.18-0.41 heritability and 0.6-0.9 genetic correlations with tibia strengths traits in the Rhode Island Red breed. These moderate to strong heritabilities and strong genetic correlations suggest that genome-wide association (GWAS) could reveal genomic regions, which contribute to multiple aspects of bone health in laying hens. In this paper, we present preliminary results from a GWAS to find genetic marker associations with 47 bone measurements in a cohort of 924 Rhode Island Red laying hens.

Materials & methods

Animals and phenotyping. A group of 924 RIR hens were investigated, which are bred for Lohmann Brown commercial layers (Lohmann Breeders GmbH, Germany). Hens from four hatches were assigned into two houses equipped with cages. Each cage had a perch and was occupied by two birds: one RIR bird and a companion from another breed. Hens were euthanized at 68 weeks of age, weighed, and bone samples were collected including humerus, tibia, and keel bone for further measurements as described previously (Dunn *et al.*, 2021).

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_598, © M. Sallam *et al.* 2022

The measurements included: a bone index, (0·27×KeelDensity + 0·37×HumerusStrength + 0·61×TibiaStrength – 0·25×BodyWeight) following (Bishop *et al.*, 2000) to jointly map correlated effects on all three bone sites, the mineral content in tibia cortical bone (determined by thermogravimetric analysis, represents carbonate apatite; calcium phosphate with some CO_3 substitution of the PO_4), lipid content in tibia cortical bone (determined by infrared spectroscopy at main peak 1,710 cm^{-1}) and humerus medullary bone score (on a scale of 0-3 where 0 represents no medullary bone and 3 a diaphyseal medullary cavity filled with bone).

Genotyping. For genome-wide association studies, the hens were genotyped on 57,636 single nucleotide polymorphisms using the Illumina Infinium assay. The genotyping was performed by the SNP&SEQ Technology Platform, Uppsala University, Sweden. We aligned the sequences flanking the markers against the GRCg6a chicken reference genome to find physical marker positions. 188 SNPs were removed because of high missingness and 21,230 were monomorphic in the sample, leaving 36,218 markers for GWAS.

GWAS. For testing each SNP marker association with the trait of interest, we used the following linear mixed model implemented in GEMMA (Zhou and Stephens, 2012):

$$y = X\,b + g\,snp + Z\,hen + e \quad (1)$$

where y is the standardized trait measurement, X and Z are design matrices that relate the measurement y to the corresponding fixed and random effect, *g* is the marker effect, and hen the genomic breeding value for each hen based on genomic relationship matrix. We included hatch and house as fixed effects, and body weight as a covariate.

Results

Three traits showed associations with *P*-value<10^{-5}: tibia cortical lipid (on chromosomes 2 and 3), humerus medullary score (on chromosome 9), and tibia cortical mineral content (on chromosome Z). Figure 1 shows Manhattan plots of these three traits. Table 1 shows position, estimated marker effect and *P*-value for the most significant associations. In addition, we report the result of bone index (Figure 1) as a joint measure of bone quality, previously used in selection and genetic mapping studies.

Table 1. The most significant SNP markers with their positions, estimated marker effect and *P*-value.

	SNP position			
Trait	**Chr.**	**Base pair**	**Effect size[1]**	***P*-value[2]**
Tibia cortical lipid	2	2585350	-0.27	8.64E-06
	3	27204115	0.30	7.58E-08
	3	27351346	0.29	1.95E-07
	3	27434588	0.31	4.54E-08
	3	27548492	0.29	1.03E-06
	3	27648733	0.29	1.70E-06
Humerus medullary score	9	22641697	0.26	5.09E-06
	9	22694487	0.25	8.26E-06
	9	22716989	0.26	5.20E-06
	9	22718522	0.26	5.20E-06
	9	22749245	0.26	5.20E-06
	9	22755016	0.26	5.20E-06
Tibia cortical mineral%	Z	41701480	-0.63	1.32E-06

[1] All traits were standardized (with zero mean and one standard deviation) to facilitate effect size interpretation.

[2] Wald test *P*-value.

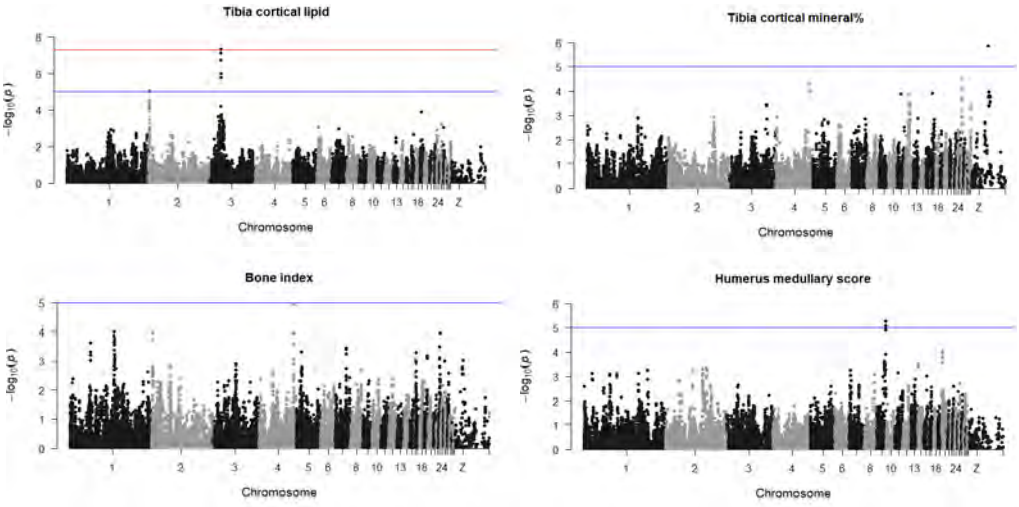


Figure 1. Manhattan plots showing $-\log_{10}$(Wald *P*-value) for each SNP marker in Tibia cortical lipid/mineral%, humerus medullary score, and bone index. The red line is to the conventional significance threshold of 5×10^{-8} and the blue a suggestive threshold of 10^{-5}.

Discussion

In this study, we detected seven loci associated with bone traits in a Rhode Island Red population. Tibia cortical lipid showed the most significant associations. Tibia cortical lipid is phenotypically correlated with the degree of mineralization (mineral% and phosphate/organic ratio) of both medulla and cortex, tibia breaking strength and density. Varying levels of lipids may be caused by variation in the cell population of the cortex. It remains to be investigated whether lipid content may be used as an intermediary physiological phenotype for bone quality.

The regions represent novel associations for bone traits. None overlap the tibial breaking strength QTL detected by Raymond *et al.* (2018) in White Leghorn layers of the Lohmann breeding program. When comparing to QTL in Chicken QTLdb, overlaps included QTL for feed intake (Mignon-Grasteau *et al.*, 2015) and comb, feather and wattle traits in a Chinese chicken breed (Sun *et al.*, 2015). The lipid association on chromosome 2 overlaps a suggestive locus for cortical carbonate content from Johnsson *et al.* (2021).

In contrast, there were no strong associations for bone breaking strength and the bone index. Tibia and humerus breaking strength show high heritabilities (~0.5) in this population (Dunn *et al.*, 2021). The genomic heritabilities estimated in the current GWAS were comparable: 0.47 for tibia breaking strength, 0.35 for humerus breaking strength and 0.48 for bone index. This suggests that the SNP chip tags most of the additive genetic variance in the trait, without detecting major QTL. Previously, a major QTL for tibial breaking strength has been detected and fine-mapped to the *cystathionine beta synthase* (*CBS*) gene in White Leghorn line (De Koning *et al.*, 2020). This locus was not detected in the current Rhode Island Red line. This finding suggests a polygenic architecture for bone-breaking strength in Rhode Island Red.

This study demonstrates the importance of developing bone measurements that enable large-scale genetic mapping studies and genomic selection of bone traits. Ideally, recording should be possible on live birds and for the keel bone, which appears the most important for layer health yet hardest to accurately phenotype (Thøfner *et al.*, 2021).

Acknowledgements

The work was partly funded by an ERANET grant to ICD (BBSRC BB/M028291/1) DJDK (Svenska Forskningsrådet Formas, 2014-01840) and ARN (Instituto Nacional de Investigación y Tecnología Agraria y Alimentaria, 291815). SM and MJ were supported by Svenska Forskningsrådet Formas (2019-02116 and 2016-01386). We acknowledge additional support through the COST Action CA15224 Keel Bone Damage.

References

Bishop, S. C., Fleming, R. H., McCormack, H. A., Flock, D. K., & Whitehead, C. C. (2000) British Poultry Science 41(1): 33–40. https://doi.org/10.1080/00071660086376

De Koning, D.-J., Dominguez-Gasca, N., Fleming, R. H., Gill, A., Kurian, D., *et al.* (2020) Genetics Selection Evolution 52(1):13. https://doi.org/10.1186/s12711-020-00532-y

Dunn, I. C., De Koning, D.-J., McCormack, H. A., Fleming, R. H., Wilson, P. W., *et al.* (2021) Genetics Selection Evolution 53(1): 11. https://doi.org/10.1186/s12711-021-00603-8

Gregory, N. G., & Wilkins, L. J. (1989) British Poultry Science 30(3):555–562. https://doi.org/10.1080/00071668908417179

Johnsson, M., Wall, H., Pinto, F. A. L., Fleming, R. H., McCormack, H. A., *et al.* (2021) BioRxiv. pre-print paper. https://doi.org/10.1101/2021.06.21.449243

Mignon-Grasteau, S., Rideau, N., Gabriel, I., Chantry-Darmon, C., Boscher, M.-Y., *et al.* (2015) Genetics Selection Evolution 47(1):74. https://doi.org/10.1186/s12711-015-0156-y

Raymond, B., Johansson, A. M., McCormack, H. A., Fleming, R. H., Schmutz, M., *et al.* (2018) Journal of Animal Science, 96(7): 2525–2535. https://doi.org/10.1093/jas/sky157

Sun, Y., Liu, R., Zhao, G., Zheng, M., Sun, Y., *et al.* (2015) G3 5(10): 2037–2041. https://doi.org/10.1534/g3.115.020883

Thøfner, I. C. N., Dahl, J., & Christensen, J. P. (2021) PLOS ONE 16(8): e0256105. https://doi.org/10.1371/journal.pone.0256105

Zhou, X., & Stephens, M. (2012) Nature Genetics 44(7): 821–824. https://doi.org/10.1038/ng.2310

599. Determining the heritability of premaxillary bone shape and size within two populations of pure line laying hens

S. Struthers[1,2*], B. Andersson[3], M. Schmutz[3], H.A. McCormack[1], P.W. Wilson[1], I.C. Dunn[1], V. Sandilands[2] and J.J. Schoenebeck[1]

[1]The Roslin Institute and Royal (Dick) School of Veterinary Studies, University of Edinburgh, EH25 9RG, Edinburgh, United Kingdom; [2]Scotland's Rural College, EH25 9RG, Edinburgh, United Kingdom; [3]Lohmann Breeders, 27472 Cuxhaven, Germany; sarah.struthers@ed.ac.uk

Abstract

Severe feather pecking (SFP) is a serious animal welfare and economic concern in commercial poultry species, particularly laying hens. The increasing demand for a ban on beak treatment presents a challenge to reducing the incidence of SFP. Considerable variation in beak shape exists within and between laying hen lines and there is the potential to genetically select hens whose beak shapes cause less damage when engaging in SFP behaviour. First, the range of phenotypes for both the external beak shapes and the underlying bones needs to be identified. The aim of this study was to estimate the heritability of the shape and size of the premaxillary bone within the top beak of two pure laying hen lines. It was found that both shape and size were heritable within the two lines, suggesting that bone shape could be incorporated into selection indices to help reduce the incidence of SFP.

Introduction

Chickens use their beaks for many important functions including feeding, drinking, maintaining plumage, and establishing social order; however, beyond these functions, birds can develop abnormal beak-related behaviours such as severe feather pecking (SFP). SFP involves the pecking at, pulling, and removal of feathers which results in large denuded areas on the body and make birds susceptible to cannibalism (Savory, 1995). Beak treatment is the shortening and blunting of beak length via hot-blade or infrared methods. Beak treatment is used to pre-emptively limit SFP damage but does not address the cause of the behaviour itself. There is increasing demand to ban beak treatment and rely on environmental enrichment and genetic selection. Selective breeding against SFP has been successful in laying hens (Rodenburg *et al.*, 2003; Grams *et al.*, 2015); however, incorporating behavioural data into a breeding programme can be challenging. A solution may be to select for traits such as plumage cover, liveability, and beak shape that are related to SFP behaviour but are easier to measure (Ellen *et al.*, 2019).

The extent to which beak shape can help reduce the incidence of SFP is not fully understood. Icken *et al.* (2017) found that beak shape was heritable and that hens with naturally shorter (blunter) top beaks had improved plumage cover and liveability. This suggests that selection for beak shape can reduce the damage that birds can inflict when engaging in SFP behaviour. Struthers *et al.* (2021) analysed the morphological variation of the premaxillary bone (within the top beak) in two pure White Leghorn laying hen lines, since the shape of the bone may influence external beak shape. The authors found that the shape and size of the premaxillary bones differed significantly between the two lines. Using the bone phenotypes from Struthers *et al.* (2021), the objective of this study was to determine the heritability of the shape and size of the premaxillary bone (within the top beak) within two pure lines of White Leghorn laying hens.

Materials & methods

This study was reviewed and approved by the Animal Welfare and Ethical Review Body at the Roslin Institute (University of Edinburgh).

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_599, © S. Struthers *et al.* 2022

Animals. Hens used for this study were from two pure line populations (Line A and B) of White Leghorn laying hens and were approximately 40 weeks of age. Both Line A (n=387) and Line B (n=432) hens contribute to the four-line breeding scheme of Lohmann Selected Leghorn (LSL) hybrid layers. Hens from both lines were housed in conventional cages (n=200) in family groups (full and/or half-siblings) with a maximum of five birds per cage.

Radiographic examination and analysis. Lateral head radiographs were collected from the birds and analysed using the methods described by Struthers *et al.* (2021). To summarise, the premaxillary bone in each radiograph was landmarked in ImageJ (v.1.53g; National Institutes of Health, Bethesda, MD) using 13 landmarks. The landmark coordinates were imported into R (v.3.6.1; R Core Team (2019)) for geometric morphometric analyses using the package Geomorph (v.4.0.0; Adams *et al.* (2021)). Within Geomorph, Procrustes superimposition was used to standardise all the landmarks in each radiograph by superimposing, re-scaling, and rotating them to a common orientation. The covariation of the transformed landmarks was analysed using principal components analysis and multivariate regression. The Procrustes superimposition also created an average bone shape by identifying the centre point (centroid) of all the landmarks in each radiograph. The centroid size was used as a proxy for relative bone size.

Statistical analysis. To estimate heritabilities and genetic and environmental correlations between principal component (PC) scores and centroid size, the following bivariate model was fitted for each line in ASReml (v4.1.1068; VSN International):

$$y = X\beta + Zu + e \quad (1)$$

where y is a vector containing the phenotypic values for each trait for all birds, β is a vector of fixed effects, and u is a vector of random genetic effects. X and Z are the known design matrices for β and u, respectively, and e is a vector of residual errors. Narrow-sense heritability was calculated as $h^2a = \sigma^2a/\sigma^2p$, with σ^2a being the additive genetic variance and σ^2p the phenotypic variance.

Results

From the geometric morphometric analyses conducted by Struthers *et al.* (2021), three PCs were found to explain 85% of the total variation in premaxillary bone shape within the two hen populations; therefore, heritability estimates were determined for the three PCs and the centroid size (Table 1 and 2).

Table 1. Estimates of heritabilities (diagonal), genetic correlations (below the diagonal) and environmental correlations (above the diagonal), with standard errors, between premaxillary bone principal component (PC) scores and centroid size for Line A hens[1].

	PC1 score	PC2 score	PC3 score	Centroid size
PC1 score	**0.41±0.14**	-0.33±0.17	0.23±0.13	**0.86±0.06**
PC2 score	-0.18±0.23	**0.54±0.15**	**0.33±0.15**	**-0.48±0.19**
PC3 score	0.01±0.42	-0.46±0.42	0.13±0.12	0.12±0.13
Centroid size	**0.62±0.15**	0.05±0.24	0.04±0.42	**0.44±0.15**

[1] Values in bold are considered significant.

Table 2. Estimates of heritabilities (diagonal), genetic correlations (below the diagonal) and environmental correlations (above the diagonal), with standard errors, between premaxillary bone principal component (PC) scores and centroid size for Line B hens[1].

	PC1 score	PC2 score	PC3 score	Centroid size
PC1 score	0.14±0.11	**-0.25±0.10**	0.07±0.16	**0.88±0.03**
PC2 score	-0.17±0.47	0.22±0.12	0.02±0.16	**-0.35±0.10**
PC3 score	-0.12±0.35	-0.14±0.30	**0.59±0.15**	0.00±0.17
Centroid size	**0.63±0.24**	-0.30±0.36	-0.01±0.27	**0.24±0.12**

[1] Values in bold are considered significant.

PC1 described the curvature of the bone tip and the width of the caudal end (towards the head) of the bone. Heritability estimates for PC1 scores were 0.41±0.14 for Line A hens and 0.14±0.11 for Line B. PC2 described bone length. Heritability estimates were 0.54±0.15 for Line A hens and 0.22±0.12 for Line B. Heritability estimates of PC scores and centroid size were higher for Line A hens than Line B apart from PC3 scores. PC3 described a slight narrowing vs widening of the caudal-most (towards the head) ventral margin (bottom edge) of the premaxillary bone. Heritability estimates were 0.13±0.12 for Line A hens and 0.59±0.15 for Line B. For centroid size, heritability estimates were 0.44±0.15 for Line A hens and 0.24±0.12 for Line B.

For both lines, there was a significant positive genetic correlation between PC1 score and centroid size (Tables 1 and 2). Although mostly non-significant, genetic correlation estimates were primarily positive for Line A hens and primarily negative for Line B. For environmental correlations, there was a significant positive correlation between PC1 score and centroid size for both lines. There was also a significant negative correlation between PC2 score and centroid size for both lines. Line A hens showed a positive correlation between PC2 and PC3 scores (Table 1). For Line B, there was also a significant negative correlation between PC1 and PC2 scores (Table 2).

Discussion

Reducing SFP within commercial laying hen flocks remains a challenge to the poultry industry and research into why hens feather peck and how to mitigate the behaviour has been ongoing since the late 1930's (Kennard, 1937). Researchers and primary breeding companies have investigated using group selection for traits such as plumage cover and liveability to reduce the incidence of SFP (Icken *et al.*, 2017; Ellen and Bijma, 2019). These traits were found to have moderate heritabilities and selection for them reduced the incidence of SFP and SFP-related injuries; however, only victims of SFP can be identified rather than the instigator (for plumage cover selection) and selection candidates are often housed individually, making group selection for liveability difficult (Brinker *et al.*, 2014; Ellen and Bijma, 2019).

Beak shape has the potential to be incorporated into selection indices to help reduce the incidence of SFP (Icken *et al.*, 2017) but in order for multivariate beak shape data to be useful, the phenotypes need to be collected in live animals. Struthers *et al.* (2021) developed the novel capability of using radiography for the application of geometric morphometrics (analysis of shape) in live, non-sedated laying hens. Radiography offers the advantage of being able to analyse both the external beak shape and the underlying bones. To our knowledge, this is the first study to estimate the heritability of the premaxillary bone within the top beak of laying hens. This experiment is part of a larger study aimed at understanding the morphology and genetics of beak shape and the underlying bones that provide its structure.

The heritability values found in the present study suggest that it would be possible to incorporate multivariate shape data into selection indices to guide the selection of hens whose beak shapes cause the least damage to other birds when engaging in SFP behaviour. The heritability estimates for bone PC scores (i.e. bone shape) are similar to beak length heritabilities reported by Icken *et al.* (2017). The authors used a device that measured the length of the top beak that overhangs the bottom. However, although assumed, it is not yet verified that external beak shape reflects the underlying shape of the bones. Thus, some caution should be used when comparing the heritability estimates of the present study to those who analyses were done using external beak shape measurements. The strong, positive correlations between PC1 scores and centroid size for each line were not surprising as Struthers *et al.* (2021) found that premaxillary bone shape was significantly associated with bone size and accounted for nearly half of the total shape variation.

The results of this study demonstrate that the heritability of premaxillary bone shape varies between the two pure line populations studied. However, there were similarities between the two lines for the strong positive genetic and environmental correlations between PC1 score and centroid size and moderate negative environmental correlations between PC2 score and centroid size. Overall, the results of the present study agree with previous studies in that the shape of the beak and its underlying bones may be advantageous to include within breeding programme selection indices.

References

Adams D.C., Collyer M.L., Kaliontzopoulou A., and Balken E.K., Geomorph: Software for geometric morphometric analyses 2021. Available at https://cran.r-project.org/web/packages/geomorph/index.html.

Brinker T., Bijma P., Visscher J., Rodenburg T.B., and Ellen E.D. (2014) Genet. Sel. Evol. 46:33. https://doi.org/10.1186/1297-9686-46-33

Ellen E.D., and Bijma P. (2019) Poult. Sci. 98(9):3431-3442. https://doi.org/10.3382/ps/pez250

Ellen E.D., van der Sluis M., Siegford J., Guzhva O., Toscano M.J., *et al.* (2019) Anim. 9(3):108. https://doi.org/10.3390/ani9030108

Grams V.S., Bögelein S., Grashorn M.A., Bessei W., and Bennewitz J. (2015) Behav. Genet. 45:228-235. https://doi.org/10.1007/s10519-014-9695-1

Icken W., Cavero D., and Schmutz M. (2017) Lohmann Inf. 51:22-27. https://lohmann-breeders.com/lohmanninfo/selection-on-beak-shape-to-reduce-feather-pecking-in-laying-hens/

Kennard D.C. (1937) Ohio Bimonthly Bulletin 22: 33-39. R Core Team. R: A lanaguage and environment for statistical computing 2019. Available at https://www.r-project.org/.

Rodenburg T.B., Buitenhuis A.J., Ask B., Uitdehaag K.A., Koene P., *et al.* (2003) Poult. Sci. 82(6):861-867. https://doi.org/10.1093/ps/82.6.861.

Savory C.J. (1995) Worlds Poult. Sci. J. 51(2):215-219. https://doi.org/10.1079/WPS19950016.

Struthers S., Andersson B., Schmutz M., McCormack H.A., Wilson P.W., *et al.* (2021) Poult. Sci. 100(12):101500. https://doi.org/10.1016/j.psj.2021.101500.

600. No G×E on egg qualities and body weight between cage and floor systems, implications for breeding programmes in laying hens

N. Bédère[1*], L. Bécot[1,2], T. Burlot[2] and P. Le Roy[1]

[1]PEGASE, INRAE, Institut Agro, 16 Le Clos 35590 Saint-Gilles, France; [2]NOVOGEN, 5 rue des Compagnons 22960 Plédran, France; nicolas.bedere@inrae.fr

Abstract

The main objective of this study was to quantify genotype-by-environment interactions for egg quality traits and body weight in laying hens between cage and cage-free systems. Records of 18,524 eggs for quality traits and 10,266 hens for body weight from a pedigreed Rhode Island Red line were used. Hens were hatched across 5 batches from 2018 to 2020, siblings are housed in two systems: a cage and a floor system. Random regressions were modelled to estimate variance components. Heritability estimates were moderate to high in both systems (0.29-0.77). Genetic correlations between the cage and the floor system were high (≥0.79), and genetic correlations between the intercept and the slope of the reaction norm ranged from -0.69 to 0.27. The results indicate that there is no need to maintain a dedicated breeding programme per system, if the breeding goals are identical, which is unlikely.

Introduction

A few breeding companies are supplying farmers with laying hens all over the world. Purebred birds of the *nuclei* of these breeding companies are housed in cages to ensure pedigree and performance monitoring. Although most of the egg production is produced in cage systems, there is a large diversity in farming systems. Some alternative systems to cage are expanding, for instance in 2020, 52% of laying hens are raised in cage-free systems in E.U. (barn, free-range, and organic; European Commission, 2021). In all these systems, breeding goals include the number of eggs produced, egg qualities (regarding the albumen, yolk, and shell), feed efficiency, behaviour, fertility and disease resistance. In the current situation, birds are selected in the *nuclei* based on performance in cage system. Some studies indicate that some major traits such as egg production traits may be under the influence of genotype-by-environment interactions (G×E), which can affect the ranking of selection candidates (Icken *et al.*, 2008; Mulder, 2017). However, it is still unclear whether there is a need to develop a breeding programme dedicated to cage-free systems because the selection candidate would be different from those in the cage system. Our objective is to quantify G×E for egg quality traits and body weight in laying hens between cage and cage-free systems.

Materials & methods

Birds and farming system. Records from the *nucleus* of a Rhode Island Red line of the breeding company Novogen were used in this study. Hens were hatched across 5 batches from 2018 to 2020. They were raised on a floor in a barn until 17 weeks of age. Then, they were either bred in collective cages or a floor system until 64 weeks of age. On average, data from 1,633 cages for the cage system and 365 hens for the floor system was retrieved per batch. The cages included a living area (with free access to food and water). The floor system included individual nests overlaid on two rows (about 1 nest for 5 hens), and a living area on floor pens (with free access to food and water). Hens were housed in cages of 5 full sisters in the cage system and raised with the entire cohort in the floor system, including the cocks (about 1 cock for 10 hens).

The pedigree for each recorded individual was traced back over at least 5 generations.

Traits. At 50 weeks of age, egg quality was recorded twice for each cage in the cage system or each hen in the floor system, and body weight once. A total of 18,524 eggs were collected: 16,331 from the cage system

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_600, © N. Bédère *et al.* 2022

and 2,193 from the floor system. A total of 10,266 hens were weighed: 9,025 from the cage system and 1,412 from the floor system. All traits were named according to the Animal Trait Ontology for Livestock (ATOL). Five egg quality traits were studied: eggshell colour (ESC), eggshell strength (ESS), eggshell shape index (ESshape), Haugh value (HV), and egg weight (EW). The body weight was recorded at 50 weeks of age (BW). ESC was determined by a chromameter, ESS was determined by a compression instrument, EW was determined by a weighing scale, ESshape is calculated from egg diameter and EW, HV was determined by a tripod methodology, and BW was determined by a weighing scale. Further details on the measurements are available on the ontology website (Picard Druet *et al.*, 2020).

Models. Because the hens in a cage are full sibs, the sire and dam of the cage were known: animal models were used for both cage and floor environments, considering each egg as an individual. Body weight was recorded at the individual level in both systems. Variance components were estimated using an average information restricted maximum likelihood method applied to univariate reaction norm models:

$$y_{ij} = \mu + CG_{ij} + \beta_1 age_{ij} + \beta_2 x_{ij} + a_{int_i} + a_{slp_i} x_{ij} + e_{ij}$$

Where y_{ij} was the observed phenotype on the j[th] egg of the i[th] 'animal' (either egg or hen, depending on the system), μ the overall mean, CG_{ij} was the fixed effect for the contemporary group (hatching date, location of the cage in the barn), β_1 was the regression coefficient on the age in weeks, β_1 was the regression coefficient on the environmental gradient x (x=0 for the cage system and x=1 for the floor system), a_{int_i} was the random additive genetic effect for the intercept of the reaction norm of the i[th] 'animal', a_{slp_i} was the random additive genetic effect for the slope of the reaction norm of the i[th] 'animal', and e was the random residual effect. For BW, the same model was used except that there is a single record per animal. The additive genetic effects were assumed to be normally distributed. Heterogeneity in residual variance was modelled with a residual variance for the cage system and a residual variance for the floor system. The residual effects were assumed to be normally distributed.

Genetic parameters were calculated from the estimated variance components (additive genetic variance for the intercept, additive genetic variance for the slope, additive genetic covariance of the intercept and the slope, and the residual variance) according to the theory (Falconer and Mackay, 1996; de Jong and Bijma, 2002).

Log-likelihood ratio test (LRT) with a reduced model without the random regression term were performed to test the significant level of the random additive genetic effect for the slope of the reaction norm.

Software. Data handling was performed with base-R (R Core Team, 2020). The variance components estimation and the calculations of genetic parameters were performed with the statistical software ASReml 4.2 (Gilmour *et al.*, 2021).

Results & discussion

We found that egg quality traits and BW were moderately to highly heritable in both environments (with heritability estimates ranging from 0.29 to 0.77; Table 1). This is consistent with other studies where heritability for these traits ranged from 0.24 to 0.60 (Beaumont *et al.*, 2010; Mulder *et al.*, 2016; Picard Druet *et al.*, 2020). Our results indicated that heritability estimates in the floor system were often higher than in the cage system. Both the additive genetic variance tended to be higher and the residual variance tended to be lower in the floor system compared to the cage system (data not shown). It is possible that the additive genetic variance was underestimated because of the loss of information about the individual that laid the egg and that estimation of a permanent environment variance was needed (repeated measures

Table 1. Genetic parameters[1] for egg production traits and mature body weight.

Traits	h^2_{cage}	h^2_{floor}	$r_{(cage,floor)}$	$r_{(int,slp)}$
ESC	0.55 (0.03)	0.74 (0.02)	0.79 (0.05)	-0.03 (0.09)[2]
ESS	0.33 (0.02)	0.31 (0.04)	0.93 (0.05)	0.15 (0.23)[2]
ESshape	0.47 (0.03)	0.58 (0.03)	0.92 (0.04)	0.19 (0.18)[2]
HV	0.29 (0.02)	0.45 (0.03)	0.96 (0.04)	0.04 (0.26)[2]
EW	0.62 (0.03)	0.77 (0.01)	0.93 (0.03)	0.27 (0.13)
BW	0.61 (0.02)	0.50 (0.01)	0.87 (0.04)	-0.69 (0.08)

[1] Estimates, with standard errors in brackets, of heritability of the traits in cage (h^2_{cage}) or floor system (h^2_{floor}), and genetic correlation between the two systems ($r_{(cage,floor)}$) and between the intercept and the slope of the reaction norm ($r_{(int,slp)}$).
[2] Not significantly different from zero (LRT).

with 2 eggs per hen), or that the unbalanced design biased the estimation of variance components (88% of the observation came from the cage system and 12% from the floor system). We tried to fit a permanent environment random effect of the cage in the cage system and of the hen in the floor system but the model did not converge.

We found that the genetic correlations between the cage and the floor system were very high ($r_{(cage,floor)} \geq 0.79$). To our knowledge, this is the first study about the same traits with the same purebred population in two environments. It is often admitted that when a genetic correlation between two traits is higher than 0.80, then these two traits are genetically similar enough to consider only one of them in the breeding programme. This genetic correlation is also used as an indicator of the importance of G×E when using a character-state model (same trait in different environments like in our study; Falconer and Mackay, 1996). Although a genetic correlation of 0.80 is high, it is not one (unity). Therefore, re-ranking of selection candidates may occur due to G×E. We used univariate random regression models to assess the stability of the genotype thanks to another genetic parameter: the genetic correlation between the intercept and the slope of the reaction norm models (de Jong and Bijma, 2002). We assumed the systems to be on a continuous scale from the reference environment (cage system) to another graduation of the environment (floor system). We found that the genetic correlations between the intercept and the slope of the reaction norms varied from -0.69 to 0.27 and that most of them were not significantly different from zero. The estimated $r_{(int,slp)}$ for EW was 0.27, meaning that the higher the breeding value for EW in the cage system (reference environment), the higher the genetic sensitivity to changing to floor system. However, this trait is particular because the breeding goal is not to increase EW but to keep it within a defined range. Giving the high $r_{(cage,floor)}$, it is possible that re-ranking only occur for extreme breeding values, which would not be a problem. The estimated $r_{(int,slp)}$ for BW is -0.69, meaning that the higher the breeding value for BW in the cage system (reference environment), the lower the genetic sensitivity to changing to floor system. The estimated $r_{(int,slp)}$ for the other traits were not significantly different from zero, indicating that selecting on the level would not change the sensitivity to environmental change.

Conclusions

In the present study, all egg quality traits and adult body weight displayed high heritability estimates, high genetic correlations between the cage and the floor systems, yet different genetic correlations between the intercept and the slope of the reaction norm models. Considering the breeding goals for all these traits, our results indicate that there is no need to maintain a breeding programme per system. However, other major traits need to be studied (egg number or disease resistance). These conclusions hold only if the breeding goals are the same between the two systems, which is unlikely (see Phocas *et al.*, 2016).

References

ATOL. Animal Trait Ontology for Livestock. Available from: http://www.atol-ontology.com

Beaumont, C., F. Calenge, H. Chapuis, J. Fablet, F. Minvielle, *et al.* 2010. Génétique de la qualité de l'œuf. INRAE Prod. Anim. 23:123–132. https://doi.org/10.20870/productions-animales.2010.23.2.3294

de Jong, G., and P. Bijma. 2002. Selection and phenotypic plasticity in evolutionary biology and animal breeding. Livest. Prod. Sc. 78:195–214. https://doi.org/10.1016/S0301-6226(02)00096-9

European Commission. 2021. Market situation for eggs. Available from: https://ec.europa.eu/info/food-farming-fisheries/animals-and-animal-products/animal-products/eggs_en

Falconer, D. S., and T. F. C. Mackay. 1996. Introduction to Quantitative Genetics. 4th ed. Longmans Green, Harlow, Essex, UK.

Gilmour, A. R., B. J. Gogel, B. R. Cullis, S. J. Welham, and R. Thompson. 2021. ASReml User Guide Release 4.2 Functional Specification. VSN International Ltd, Hemel Hempstead, UK.

Icken, W., D. Cavero, M. Schmutz, S. Thurner, G. Wendl, *et al.* 2008. Analysis of the free range behaviour of laying hens and the genetic and phenotypic relationships with laying performance. Br Poult Sci. 49:533–541. https://doi.org/10.1080/00071660802158357

Mulder, H. A. 2017. Is GxE a burden or a blessing? Opportunities for genomic selection and big data. J Anim Breed Genet. 134:435–436. https://doi.org/10.1111/jbg.12303

Mulder, H. A., J. Visscher, and J. Fablet. 2016. Estimating the purebred–crossbred genetic correlation for uniformity of eggshell color in laying hens. Genet Sel Evol. 48:39. https://doi.org/10.1186/s12711-016-0212-2

Phocas, F., C. Belloc, J. Bidanel, L. Delaby, J. Y. Dourmad, *et al.* 2016. Review: Towards the agroecological management of ruminants, pigs and poultry through the development of sustainable breeding programs: I. selection goals and criteria. Animal. 10:1749–1759. https://doi.org/10.1017/S1751731116000926

Picard Druet, D., Varenne, A., Herry, F., Hérault, F., Allais, S., *et al.* 2020. Reliability of genomic evaluation for egg quality traits in layers. BMC Genet 21, 17. https://doi.org/10.1186/s12863-020-0820-2

R Core Team. 2020. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.

601. Growth and laying performance of local chicken breeds and their crosses fed with faba beans

A.R. Sharifi[1,2*#], T. Nolte[1,2#], S. Jansen[3], S. Weigend[2,3], D. Mörlein[1], I. Halle[4], W. Link[5] and H. Simianer[1,2]

[1]University of Goettingen, Department of Animal Sciences, 37075 Goettingen Germany; [2]University of Goettingen, Center for Integrated Breeding Research, 37075 Goettingen; [3]Institute of Farm Animal Genetics, Friedrich-Loeffler-Institut, 31535 Neustadt, Germany; [4]Institute of Animal Nutrition, Friedrich-Loeffler-Institut, 38116 Braunschweig, Germany; Germany; [5]University of Goettingen, Department of Crop Sciences, 37075 Goettingen, Germany; rsharif@gwdg.de; [#]These authors contributed equally to this work

Abstract

Poultry production systems are currently dealing with some important issues like animal welfare, environmental impacts of soy imports from overseas, and the decline of genetic diversity. Therefore, an alternative approach to poultry production was tested in this study using six different chicken genotypes in terms of fattening performance of males and laying performance of females. Two traditional chicken breeds, Vorwerkhuhn (VH) and Bresse Gauloise (BG), and a commercial layer genotype, White Rock (WR) as well as crossbreds thereof were fed diets containing either 20% vicin-rich or vicin-poor faba beans. The genotype BG and the crossing genotypes with BG exhibited the best growth performance. WR and BG × WR cross showed the best performance regarding the egg production and egg quality. Taken together, the dietary effect revealed only small negative impacts of vicin. In terms of dual-purpose usage, crossing BG with WR seems to be the most promising variant studied here.

Introduction

In recent decades, poultry production has developed into a highly diversified sector with chicken lines genetically specialized either for egg production or for meat production (Krautwald-Junghanns *et al.*, 2018). In poultry production, both sexes of meat type chickens are used for meat production, whereas the male chicks of layer lines are not used and are culled on their first day of life due to their low growth potential and the associated economic disadvantages. This is problematic in respect to animal welfare and ethical aspects and is therefore no longer permitted in Germany. In this context, the fattening of the cocks of heavy layer hybrids or the use of dual-purpose-breeds have been examined (Damme *et al.*, 2003), showing that the layer cockerels are clearly inferior to commercial broilers. The use of local breeds might be an alternative solution to serve niche markets in a regional context. These breeds historically have been used for both egg and meat production and are nowadays mostly kept by hobby breeders. To supply the poultry industry with high-quality protein feed, the EU imports large quantities of soybeans from overseas. The use of genetically modified seeds is common in the main producing countries and is viewed critically by European consumers. In addition, the negative environmental impact of the soybean production, especially in South America (Barona, *et al.*, 2010), suggest a reduction in soy imports in favour of regional protein crops in Germany. A suitable regional protein crop is the faba bean (*Vicia faba*). However, antinutritional substances such as the glycosides vicin and convicin (VC) have so far restricted the use of field beans in animal feed. The main objective of this study is therefore to evaluate the fattening and laying performance of animals of two local pure breeds compared to a commercial line and their crosses, and to quantify the effect of feeding faba bean with different VC content on performance parameters.

Materials & methods

This study included three purebred and three crossbred genotypes of domestic chicken. The purebred genotypes were Vorwerkhuhn (VH), Bresse Gauloise (BG), and White Rock (WR). VH is a local chicken

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_601, © A.R. Sharifi *et al.* 2022

breed from Germany, which was originally bred for dual-purpose usage. The BG hens originated from the Bresse region in the south of the french Burgundy County. The WR is a commercial layer parent stocks from Lohmann breeders company (Cuxhaven, Germany). In order to quantify the effect of genotypes and feeding on fattening performance, in a first experiment (A) male chicks of the purebred lines (120 chicks of BG and WR each and 94 chicks VH) at 4 weeks of age were subjected to three different feeding treatments. Two diets contained faba beans as an alternative source of protein, while the third diet was a soybean-based standard feed as control (soy). The feeding treatment with faba beans comprised a diet containing 20% of a VC-rich faba bean variety Fuego (VC+) and the other 20% of a VC-poor variety Tiffany (VC-). In experiment B, 120 birds from three crosses (VH males × BG females: VBG, VH males × WR males: VWR, and BG male × WR female: BWR) were subjected to the same feeding regime from 4 weeks of age as described for the purebreds. The animals of both experiments were slaughtered at a target weight of ca. 2,100 g to assess fattening performance. Therefore, the slaughter age of the breeds varied between 10 to 17 weeks of life. Due to the difference in slaughter age and for comparability of genotypes, in this paper fattening performance was analysed for a time window of 4 to 10 weeks, considering the effects of genotypes, feeding treatment, regression effect of pre-treatment weight, regression effect of age in the linear and quadratic terms, and random effect of pen. In order to quantify the influence of the different genotypes and feeding treatments on laying performance and egg quality, two consecutive trials were conducted. As described above, trial A tested purebred hens of the VH, BG, and WR breeds. Trial B dealt with their crosses (VBG, VWR, VWR). In both experiments, the testing period lasted from the18th until the 52nd week of age. Egg weight and eggshell quality was analysed at four different time points (week 26, 34, 42 and 50). At each time point, 96 randomly selected eggs from four consecutive days were analyzed per genotype × diet combination. The statistical analysis of the laying performance was carried out using a polynomial growth function as described above for weight development. The egg quality parameters were analysed by considering the effect of genotype, diet, age, and their interactions.

Results

Weight gain and feed efficiency. The comparison of the genotypes from week five to ten showed differences between the respective genotypes, which became clearer with increasing age (Figure 1). Statistical analysis of weight at 10 weeks of age showed that the animals of BG breed had the highest weight of 1,823 g, followed by the two BG crosses VBG (1,580 g) and BWR (1,536 g), which were both significantly different from the purebred BG genotypes at this age. The VH animals weighed 1,342 g, which was significantly less than the BG roosters and its crosses, but heavier than the WR animals (1,286 g). The weight of the VWR birds (1,312 g) was in between the VH and WR roosters. There was a significant effect of feeding treatment on the weight of animals with a difference of 19 g between VC- and soy, and 21 g between VC+ and soy. Within the respective genotypes, no significant differences in the growth performance of animals were observed between the feeding treatments. In terms of feed efficiency BG (2.56 kg/kg) outperformed the other genotypes. VBG showed an FCR of 2.76, which was significantly better than VWR (2.91). BWR, VH, and WR ranked between these two genotypes with FCRs of 2.74, 2.82, and 2.84, respectively. There were no statistically significant differences in FCR between the feeding groups in general and within the respective genotypes as well as in the separate analysis of the genotypes over the entire experimental period.

Egg performance and egg quality. The egg production curves of the different genotypes are shown in Figure 2. The purebreds differed significantly from each other at all experimental times. As expected, WR achieved the highest performance in the purebred lines (peak performance 100%), followed by BG and VH. The VH hens showed the lowest laying performance, which resulted from a delayed laying maturity and a peak performance of only about 56%. The crossbreds were significantly different from each other from weeks 18 to 48, but in weeks 49 to 51, the difference between VWR and VBG became smaller and was no longer statistically significant due to the low persistency of VWR. VWR hens had similar performance

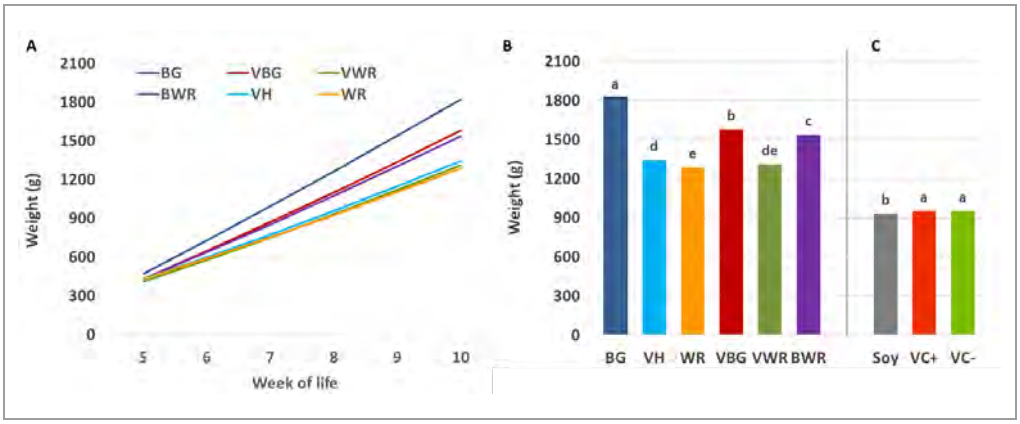


Figure 1. Weight gain of different genotypes and diets. (A) Effect of genotype on weight development from five to ten week of life. (B) LS-means for bodyweight of the respective genotypes at week ten of age. (C) LS-means for bodyweight of the different feeding groups adjusted by the mean of age.

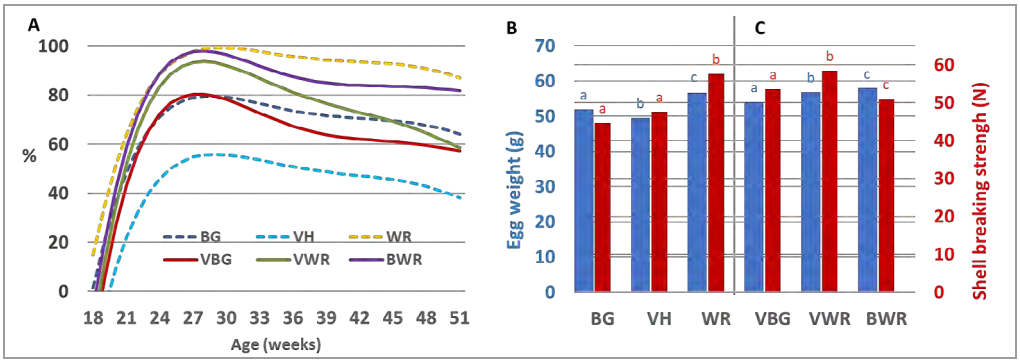


Figure 2. The effect of different genotypes on egg performance and egg quality Comparison of laying performance (A), egg weight (B) and eggshell breaking strength (C) of purebreds and crossbred genotypes.

compared to BWR hens, but were worse regarding laying persistency. The laying performance of VBG was lower compared to BWR and VWR.

There was a significant difference in egg weight between all genotypes. Among the purebred genotypes, WR hens produced the significantly highest egg weights (56.6 g), followed by BG (51.9 g) and VW hens (49.4 g). In the crossbreeds, the highest egg weights were found for the BWR genotype (58.0 g), while VBG laid the lightest eggs (54.0 g).

No significant interaction between genotypes and feeding was detected for egg weight. However, for both purebred and crossbred genotypes, the VC+ group laid significantly smaller eggs than the soybean and VC- groups, but these differences were not pronounced and amounted to 1.1 g or even less. Shell strength was significantly highest for WR (57.68 N) and for the crossbred genotype VWR (58.41 N). The purebred BG and VH hens exhibited values below 50 N. The differences between the dietary groups were only found in the crossbreds, where the VC+ group showed a significantly reduced eggshell breaking strength.

Discussion

The observed performance difference between the purebreds is caused by their genetic background, indicating the BG is a meat-type dual purpose breed, whereas the VH is a layer-type dual purpose breed and the WR is a parental layer genotype. The meat-type BG is the most suitable for fattening among the tested genotype due to its high growth rate and good feed efficiency. These findings are in line with those reported by Lambertz *et al.* (2018) for BG and their crosses with New Hampshire hens. As a cross-breeding partner, BG shortens the fattening period and improves the feed efficiency of layer genotypes VH and WR. The laying performance and egg weight of the local breeds and especially of VH breed has been considerably lower than WR breed. With regard to the shell breaking strength, there was also a clear difference between the commercial WR line and the local breeds BG and VH in the present study, which demonstrates a long-term genetic selection effect. The laying performance of the crossing genotypes in the first half of the laying period was more similar to the performance level of the maternal genotype than to the paternal. However, in the second half of the laying period, there was a decrease in direction of the paternal performance, although BWR showed a much better persistency than VWR in the laying performance. The mean egg weights of the crossing genotypes were higher than the mean of the parental pure lines for all genotypes. However, this difference cannot be construed directly as a heterosis effect, since it could be confounded with the effect of different experimental series. Concerning eggshell parameters, the values of the cross genotypes were between the parental lines. Overall, no adverse effect of VC containing feeds on laying and fattening parameters was detected, with the exception of a marginal negative effect on egg weight. Since our study deals with dual-purpose genotypes, both the fattening performance of the males and the laying performance of the females should be considered for a final conclusion. The performance test of the males showed that the BG breed improved the fattening performance of the layer-type VH and WR chickens. Therefore, considering both sexes, the BWR hybrids seem to be the most promising cross.

Funding

This research was funded by the Lower Saxony Ministry of Science and Culture, grant number: MWK 11-76251-99-30/16.

References

Barona, E.; Ramankutty, N.; Hyman, G.; Coomes, O.T. (2010). Environ. Res. Lett. 5, 1–9. https://doi.org/10.1088/1748-9326/5/2/024002

Damme, K.; Ristic, M. (2003). Worlds Poult. Sci. J. 2003, 59, 50-53.

Krautwald-Junghanns M.E., Cramer K., Fischer B., and Förster A. *et al.* (2018). Poult Sci. 97(3): 749-757. https://doi.org/10.3382/ps/pex389

Lambertz, C. Wuthijaree, K.; Gauly (2018). Poult. Sci., 97, 3564-3576. https://doi.org/10.3382/ps/pey223.

602. Heterosis of egg-laying performance in chickens

A.X. Ni[1,2*], M.P.L. Calus[2], H. Bovenhuis[2], Y.Y. Sun[1] and J.L. Chen[1]

[1]Institute of Animal Sciences, Chinese Academy of Agricultural Sciences, 100193 Beijing, China; [2]Wageningen University & Research, P.O. Box 338, 6700 AH Wageningen, the Netherlands; aixin.ni@wur.nl

Abstract

Heterosis is the phenomenon that hybrids display better performance than the parental average. In this study, we aimed to quantify egg-laying performance, and heterosis for a complete diallel cross of Beijing-You and White leghorn chickens. Egg weight, egg production and egg quality traits were measured, and heterosis was estimated. We found significant heterosis for egg weight, egg number and age at first egg in the cross lines. Favourable heterosis was also observed for egg quality traits except albumen height and Haugh unit. Moreover, for egg weight, heterosis tended to be higher if Beijing-You was the sire line and White Leghorn was the dam line which was in contrast to egg production traits.

Introduction

Heterosis is the phenomenon that performance in hybrids is superior to the average performance of their parents. At present, most eggs, chicken meat, and pork are produced from hybrid animals. Utilization of heterosis in commercial breeding schemes relies on investigating the best crossbreeding combinations performance from extensive cross-examination Therefore, in this study, a complete diallel cross of Beijing-You (Y) and White Leghorn (W) was constructed to evaluate heterosis of egg-laying performance in chickens.

Materials & methods

Population. Two breeds (Y, Beijing-You, a Chinese local breed; W, White Leghorn, a high productive breed) with a large difference in egg-laying performance were chosen. The two breeds were mated to generate four genetic groups: (1) 30 Y roosters were mated with 150 Y hens to generate YY offspring; (2) the same roosters were mated with 150 W hens to generate YW offspring; (3) 30 W roosters were mated with 150 W hens to generate WW offspring; and (4) the same roosters were mated with 150 Y hens to generate WY offspring.

Egg-laying performance. Egg laying characteristics measured were age at first egg (AFE), egg number until 43 weeks of age (EN43), average clutch length (ACL), and oviposition period (OP), weight of first egg (FEW), egg weight every 4 week from 28 weeks of age till 68 weeks of age, and egg quality traits at 32 weeks of age: shell colour (SC), shell strength (SI), shell thickness (ST), albumen height (Alh), Haugh unit (HU), egg shape index (ESI), yolk ratio (YR), and shell ratio (SR).

Statistics analysis. Heterosis and genetic parameters were estimated using the following model:

$$y_{ijk} = \mu + b_i + a_j + e_{ijk}$$

Where y_{ijk} is the phenotypic value, μ is the population average, b_i is the fixed effect of genetic group (YY, YW, WW and WY), a_j is the random animal effect with 2,485 animals included in the pedigree, e_{ijk} is the random error.

Estimated differences between YW and mid-parent value (MPV), between WY and MPV, and between YW and WY and their Wald F statistics were obtained by ASReml.

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_602, © A.X. Ni *et al.* 2022

Results

Descriptive statistics. The mean egg weight increased with aging (Table 1). The standard deviation for egg number at 43 weeks of age and age at first egg were higher than oviposition period and average clutch length (Table 1). Descriptive statistics for egg quality traits is provided in Table 2.

Genetic parameter analysis. Estimated heritability for egg weight at 28 to 64 weeks of age was high (0.56-0.77), with the highest value at 52 and 64 weeks of age (Table 3). Genetic correlations between egg weight at different ages were high (0.86-0.98), especially for consecutive time points. Age at first egg was negatively correlated with egg number until 43 weeks of age (r_g=-0.47, Table 3). The heritability for all egg quality traits at 32 weeks of age was moderate (0.31-0.50, Table 4). Shell thickness had a strong genetic correlation with both shell ratio (r_g=0.85), and shell strength (r_g=0.69, Table 4).

Table 1. Descriptive statistics for egg weight and egg production traits of all populations.

Trait[1]	EWt28	EWt40	EWt52	EWt64	EN43	AFE	OP	ACL
n	1,141	989	868	774	928	1,194	1,104	936
Mean	51.28	55.85	58.05	60.85	110.86	167.92	25.27	7.55
SD	4.86	4.99	4.85	5.01	18.40	11.83	1.11	5.95

[1] EWtx = egg weight at x weeks of age, n = number of test animals, SD = standard deviation.

Table 2. Descriptive statistics for egg quality traits of all populations.

Trait	SC	SI	ST	Alh	HU	ESI	YR	SR
n	1,075	1,068	1,074	1,074	1,071	1,074	1,071	1,065
Mean	64.46	3.79	0.34	4.53	67.29	1.31	29.00	9.98
SD	12.90	0.57	0.02	0.80	6.69	0.04	2.18	0.59

Table 3. Estimated genetic parameters for egg weight and egg production traits.

Trait category	Egg weight					Egg production			
Trait[1]	EWt28	EWt40	EWt52	EWt64	Trait	EN43	AFE(d)	OP	ACL
EWt28	0.56	0.89	0.84	0.86	EN43	0.22	-0.47	-0.30	0.36
EWt40	0.67	0.67	0.97	0.97	AFE(d)	-0.62	0.54	-0.28	0.24
EWt52	0.62	0.80	0.77	0.98	OP	-0.35	-0.01	0.38	-0.74
EWt64	0.60	0.75	0.81	0.77	ACL	0.47	0.03	-0.39	0.06

[1] Values in bold are heritability estimates, the upper triangles are genetic correlations and lower triangles are phenotypic correlations.

Table 4. Estimated genetic parameters for egg quality traits.

Trait	SC(%)	SI(kg/cm^2)	ST(mm)	Alh(mm)	HU	ESI	YR	SR
SC (%)	0.31	-0.22	-0.44	-0.44	-0.46	0.00	0.27	-0.37
SI (kg/cm^2)	-0.17	0.38	0.69	-0.14	-0.13	-0.06	0.03	0.74
ST (mm)	-0.33	0.55	0.42	-0.12	-0.10	0.00	-0.08	0.85
Alh (mm)	-0.07	-0.01	-0.03	0.32	0.97	-0.11	-0.32	-0.13
HU	-0.04	-0.02	-0.07	0.92	0.31	-0.09	-0.26	-0.09
ESI	-0.02	0.07	-0.01	0.01	-0.04	0.44	-0.03	-0.07
YR	0.10	-0.05	-0.19	-0.44	-0.35	0.00	0.44	0.07
SR	-0.20	0.61	0.76	-0.28	-0.19	-0.03	0.09	0.50

Predicted values. Both YW and WY showed significant heterosis (1.07%-6.65%) for egg weight throughout the laying period. The heterosis for egg weight increased with advancing age, and was higher for YW than for WY (Figure 1). Age at first egg and egg number until 43 weeks of age showed favourable heterosis, and the heterosis for egg number was higher for WY than for YW. While oviposition period differed between WY and MPV, and between YW and WY. Except for albumen height and Haugh unit, all egg quality traits presented favourable heterosis. For shell colour, yolk weight, shell strength, and shell thickness, hybrid performance was higher than MPV, while the reverse was observed for Haugh unit. Heterosis for yolk weight, yolk ratio, and egg shape index in YW was higher than in WY (Table 5).

Discussion

Egg-laying performance is an important economic trait. The pedigree-based heritability estimates for egg weight and egg quality traits were slightly higher than those in a Rhode Island Red (RIR) population (Liu *et al.*, 2018a, 2018b). The estimate of egg number and age at first egg were similar to RIR population (Liu *et al.*, 2019). The genetic and phenotypic correlation among all egg-laying traits were similar to previously reported estimates (Liu *et al.*, 2018a, 2018b, 2019).With the extension of the laying period egg weight increases which confirms results of other studies (Ledur *et al.*, 2002).

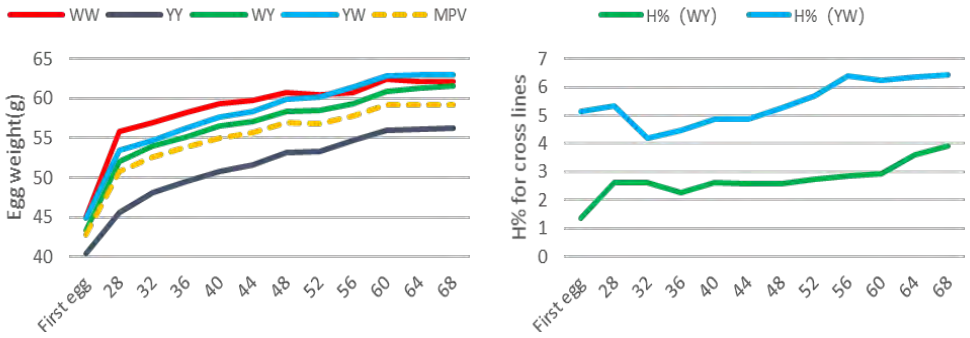


Figure 1. Predicted egg weight for the four genetic groups and heterosis for the two reciprocal crosses.

Table 5. Predicted values for egg production and egg quality traits.

Trait		MPV	H%(WY)	H%(YW)	*P*-value		
					WY vs MPV	YW vs MPV	WY vs YW
Egg production	EN43	109.97	3.01	1.87	<0.01	<0.01	<0.01
	AFE	169.21	-1.96	-0.99	0.01	<0.01	<0.01
	OP	25.26	-0.62	0.03	0.02	0.69	<0.01
	ACL	9.04	-24.29	-26.07	<0.01	<0.01	0.02
Egg quality	SC (%)	63.00	8.33	5.88	0.50	<0.01	<0.01
	SI (kg/cm^2)	3.62	5.08	10.71	<0.01	<0.01	0.32
	ST (mm)	0.33	1.08	2.65	0.01	<0.01	0.09
	Alh (mm)	4.62	-2.20	-4.28	0.04	0.22	0.50
	HU	68.49	-3.21	-3.66	<0.01	<0.01	0.75
	ESI	1.31	0.56	0.36	0.08	<0.01	0.03
	YR	28.58	1.91	2.08	<0.01	0.05	<0.01
	SR	9.86	1.32	3.00	<0.01	<0.01	0.21

In this experiment, significant heterosis was observed for egg weight, egg number, average clutch length and Haugh unit, which was consistent with other studies (Abplanalp *et al.* 1984; Ledur *et al.* 2000; Emmerson *et al.* 2002; Ledur *et al.* 2002). But heterosis for egg number and age at first egg was smaller than Williams *et al.* results in fowl (2002) and Emmerson *et al.* in turkey (2002). Our results showed that the level of heterosis is both dependent on the trait and the reciprocal cross.

Throughout the experimental period, we found higher heterosis for egg number in WY than YW, and higher for egg weight in YW than WY. There was a strong divergence in heterosis for these traits, illustrating the complexity of non-additive genetic variation and the difficulty in predicting heterosis. Sutherland *et al.* (2018) also found differences between reciprocal crosses for growth traits in chickens, and suggested that one or more Z-linked genes may underlie this difference.

In conclusion, favourable heterosis was observed for most egg-laying performance traits in a cross between Beijing-You and White Leghorn chickens. Moreover, a considerable difference in heterosis between the two reciprocal crosses was observed. The molecular mechanisms underlying heterosis will be further explored in a subsequent genome-wide association study.

Acknowledgements

This work was supported by the National Natural Science Foundation of China (32172721), and scholarships from the China Scholarship Council (202103250072).

References

Abplanalp H., Okamoto S., Napolitano D. and Len E.R. (1984) Poult Sci 63(2):234-239. https://doi.org/10.3382/ps.0630234

Emmerson D.A., Velleman S.G., and Nestor K.E. (2002) Poult Sci 81(3):316-320. https://doi.org/10.1093/ps/81.3.316

Ledur M.C., Fairfull R.W., McMillan I., and Asseltine L. (2000) Poult Sci 79(3):296-304. https://doi.org/10.1093/ps/79.3.296

Ledur M.C., Liljedahl L.E., McMillan I. and Asselstine L. *et al.* (2002) Poult Sci 81(10):1439-1447. https://doi.org/10.1093/ps/81.10.1439

Liu Z., Sun C.J., Yan Y.Y. and Li G.Q. *et al.* (2018a) Sci Rep 8(1):10832. http://doi.org/10.1038/s41598-018-29162-7

Liu Z., Sun C.J., Yan Y.Y. and Li G.Q. *et al.* (2018b) Front Genet 9:128. https://doi.org/10.3389/fgene.2018.00128

Liu Z., Yang N., Yan Y.Y. and Li G.Q. *et al.* (2019) BMC Genet 20(1):67. https://doi.org/10.1186/s12863-019-0771-7

Sutherland D.A.T., Honaker C.F., Dorshorst B. and Andersson L. *et al.* (2018) J Appl Genet 59(2):193-201. https://doi.org/10.1007/s13353-018-0435-8

Williams S.M., Price S.E., and Siegel P.B. (2002) Poult Sci 81(8):1109-1112. https://doi.org/10.1093/ps/81.8.1109

603. *In vitro* effects of probiotic *Bacillus* strains on adaptive immune cells of broiler chicken

F. Larsberg[1], M. Sprechert[1], D. Hesse[1], G. Loh[2], G.A. Brockmann[1] and S. Kreuzer-Redmer[3*]

[1]*Albrecht Daniel Thaer-Institute, Breeding Biology and Molecular Genetics, Humboldt University of Berlin, Unter den Linden 6, 10099 Berlin, Germany;* [2]*Evonik Operations GmbH, Research, Development & Innovation Nutrition & Care, 33790 Halle, Germany;* [3]*Institute of Animal Nutrition and Functional Plant Compounds, Unit of Nutrigenomics, University of Veterinary Medicine Vienna, Veterinärplatz 1, 1210 Vienna, Austria;* *susanne.kreuzer-redmer@vetmeduni.ac.at*

Abstract

Probiotic *Bacillus amyloliquefaciens* CECT 5940 *(B. amyloliquefaciens)* and *Bacillus subtilis DSM 32315 (B. subtilis)* have positive effects as feed additives on chicken health and performance. However, the underlying modes of action are still elusive, they may have a direct effect on the immune system. Here, we investigated direct effects of *B. amyloliquefaciens* and *B. subtilis* on peripheral blood mononuclear cells (PBMCs) *in vitro*. Flow cytometric analysis revealed a higher Δ relative cell count of CD4+ T-helper cells ($P<0.1$; $P<0.05$) and CD4+CD25+ activated T-helper cells ($P<0.05$; $P<0.01$) in the cultured chicken PBMCs after treatment with both *Bacillus* strains. Furthermore, we found a higher Δ relative cell count of CD8+ cytotoxic T-cells ($P<0.05$) and CD8+CD25+ activated cytotoxic T-cells ($P<0.05$) after treatment with *B. subtilis.* Hence, we suggest a direct immunomodulatory effect of *B. subtilis* and *B. amyloliquefaciens* on chicken PBMCs.

Introduction

Poultry is the cheapest source of animal protein worldwide and the second largest source of meat after pork (Wahyono and Utami, 2018). However, there is growing concern about the resistance of pathogenic bacteria against antibiotics, residual antibiotics in meat products for human nutrition (Sato *et al.*, 2019), and public health risk from zoonotic pathogens like *Salmonella* and *Campylobacter*. Since the subtherapeutic usage of antibiotics was forbidden in farming in the EU in 2006 (Council Regulation EC 70/524/ EEC), alternatives such as prebiotics, probiotics, synbiotics, and trace elements have ascended to improve the health of the animal and to reduce the usage of antibiotics. Among most widely and commonly used probiotics are *Bacillus* spp., spore forming bacteria with resistance to high temperatures and harsh storage conditions. They are generally considered as safe strains for use in poultry production (Penaloza-Vazquez *et al.*, 2019). *Bacillus amyloliquefaciens* CECT 5940 (*B. amyloliquefaciens*) and *Bacillus subtilis DSM 32315* (*B. subtilis*) were shown to have positive effects on pathogenic bacteria such as *Clostridium perfringens* (Gharib-Naseri *et al.*, 2021, Lei *et al.*, 2015) and *to* improve growth performance, increase feed efficiency, manipulate the intestinal structure and microbial composition of the gut (Ma *et al.*, 2018). However, the underlying mechanisms of action of these probiotics are still unknown. They could have a direct effect on the immune system or e.g. interact with the microbiome. This study aims to enlighten the mechanisms of probiotics. We examined the direct effects of *B. amyloliquefaciens* CECT 5940 and *B. subtilis DSM 32315 on T-helper cells and cytotoxic T-cells as two major components of the adaptive immune response. The results could* help to improve health in chicken farming and to prevent infectious diseases and the use of antibiotics.

Materials & methods

We conducted cell culture experiments using a previously described cell culture model with chicken PBMCs (Larsberg *et al.* 2021) in co-culture with *B. amyloliquefaciens* CECT 5940 and *B. subtilis DSM 32315*. The blood of Cobb500 broiler chicken was collected in polystyrene tubes pre-filled with sodium citrate (VACUETTE®, Greiner Bio-One, Kremsmünster, Austria). For lymphocyte isolation, blood was

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_603, © F. Larsberg *et al.* 2022

diluted with PBS (Gibco™, ThermoFisher Scientific, Waltham, MA, USA) containing 2 mM EDTA (Carl Roth, Karlsruhe, Germany) in the ratio 1:2 and mixed with 3% dextran (Carl Roth, Karlsruhe, Germany) in a ratio of 1:0.4. The suspension was centrifuged at 50×*g* for 20 min. The upper phase was carefully layered on an equal volume of a solution of polysucrose and sodium diatrizoate (Histopaque-1077, Sigma-Aldrich, St. Louis, MO, USA) and centrifuged at 900×*g* for 30 min without break to separate PBMCs from the residual blood. PBMCs from the interphase were transferred to a PBS-coated 50 ml tube and washed twice with PBS by centrifugation at 400×*g* for 10 min. Thereafter, PBMCs were resuspended in Roswell Park Memorial Institute (RPMI) 1640 medium (Gibco™, ThermoFisher Scientific, Waltham, MA, USA) containing glucose, 10% chicken serum, and 1% Pen/Strep before culture at 41 °C with 5% CO_2. For co-culture experiments, 1×10^6 cells (Tali® Image-Based Cytometer, Invitrogen™, ThermoFisher Scientific, Waltham, MA, USA) were treated with vital bacteria from the probiotic strains *B. amyloliquefaciens* CECT 5940 and *B. subtilis DSM 32315* in a ratio of 1:3 (PBMCs:*Bacillus*) for 24 h in RPMI 1640 supplemented with 10% chicken serum in a 24-well plate. 10 µg/ml concanavalin A (conA) was used as a positive control for proliferation. Untreated cells cultured in RPMI 1640 medium served as a negative control. Thereafter, cells were harvested, centrifuged at 400×*g* for 10 min before staining with different antibodies. Immune cells were stained with antibodies labelling T-helper cells (CD4-SPRD, CT-4, SouthernBiotech, Birmingham, AL, USA) and cytotoxic T-cells (CD8-APC, CT-8, ThermoFisher Scientific, Waltham, MA, USA). Furthermore, the T-cell activation was investigated (CD25-FITC, AbD13504, Bio-Rad Laboratories, Hercules, CA, USA). DAPI was used as a viability marker. 20,000 vital cells were acquired on a BD FACSCanto™ II (Becton Dickinson, Franklin Lakes, NJ, USA). The cell count was calculated relative to the vital PBMCs measured. Due to high variation in the immune response capacity, the results are displayed as the Δ relative cell count, representing the difference of the relative cell count between the treatment and the respective untreated negative control for every biological replicate. Statistical analysis was performed by a one sample *t*-test against a hypothetical value 0. All tests were executed using GraphPad Prism 8.0.2 (GraphPad Software, San Diego, CA, USA). Differences between groups were considered statistically significant at $P<0.05$.

Results

To investigate direct effects of *B. amyloliquefaciens* (Figure 1) and *B. subtilis* (Figure 2) on the adaptive immune cells, we assessed the Δ relative cell count of CD4+ T-helper cells (Figure 1a, Figure 2a), CD4+CD25+ activated T-helper cells (Figure 1b, Figure 2b), CD8+ cytotoxic T-cells (Figure 2c) and CD8CD25+ activated cytotoxic T-cells (Figure 2d).

As expected, conA as a positive control increased the Δ relative cell count of CD4+ T-helper cells (Figure 1a, $P<0.01$) and CD4+CD25+ activated T-helper cells (Figure 1b, $P<0.0001$). After treatment with *B. amyloliquefaciens*, we detected a higher Δ relative cell count of CD4+CD25+ activated T-helper cells (Figure 1b, $P<0.05$) and of CD4+ T-helper cells (Figure 1a, $P<0.1$). There was no difference for cytotoxic T-cells (data not shown).

As expected, conA increased the Δ relative cell count of T-helper cells (Figure 2a, $P<0.0001$; Figure 2b, $P<0.0001$) and cytotoxic T-cells (Figure 2c; Figure 2d, $P<0.0001$) also in the experiments with *B. subtilis* treatment. Furthermore, we found a higher Δ relative cell count of CD4+ T-helper cells (Figure 2a, $P<0.05$) and CD4+CD25+ activated T-helper cells (Figure 2b, $P<0.01$) after treatment with *B. subtilis*. Moreover, the Δ relative cell count of CD8+ cytotoxic T-cells (Figure 2c, $P<0.05$) and CD8+CD25+ activated cytotoxic T-cells (Figure 2d, $P<0.05$) increased in the presence of *B. subtilis*.

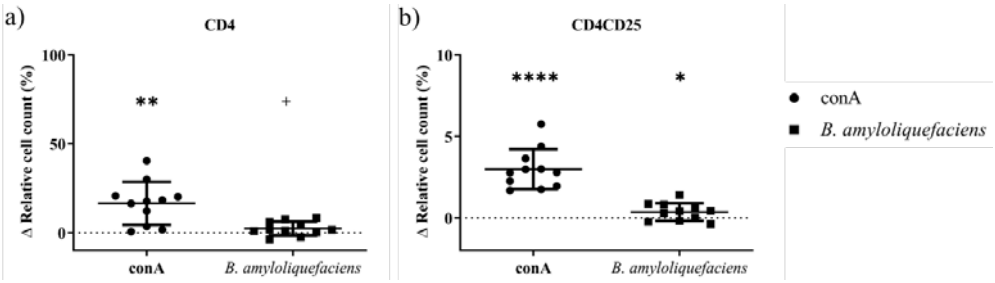


Figure 1. *In vitro* effects of *B. amyloliquefaciens* CECT 5940 on T-helper cells. CD4+ T-helper cells (a) and CD4+CD25+ activated T-helper cells (b). The Δ relative cell count was calculated as the difference of the relative cell count between the treatment and the respective untreated negative control measured on a BD FACSCanto™ II. Cells were treated with *B. amyloliquefaciens* in a ratio of 1:3 (PBMCs:*Bacillus*), with 10 µg/ml conA (positive control) or left untreated (negative control). Error bars show SD. Significance is shown as +, *P*<0.1; *, *P*<0.05; **, *P*<0.01; ****, *P*<0.0001.

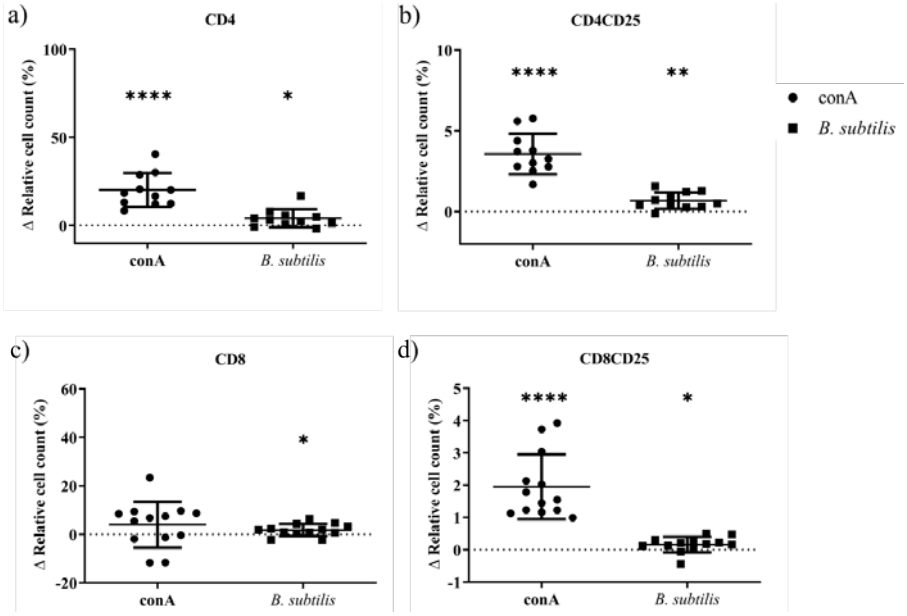


Figure 2. *In vitro* effects of *B. subtilis DSM 32315* on T-helper cells and cytotoxic T-cells. CD4+ T-helper cells (a), CD4+CD25+ activated T-helper cells (b), CD8+ cytotoxic T-cells (c), and CD8+CD25+ activated cytotoxic T-cells (d). The Δ relative cell count was calculated as the difference of the relative cell count between the treatment and the respective untreated negative control measured on a BD FACSCanto™ II. Cells were treated with *B. subtilis* in a ratio of 1:3 (PBMCs:*Bacillus*), with 10 µg/ml conA (positive control) or left untreated (negative control). Error bars show SD. Significance is shown as *, *P*<0.05; **, *P*<0.01; ****, *P*<0.0001.

Discussion

We investigated the effects of two *Bacillus* strains, *B. amyloliquefaciens* and *B. subtilis* in a cell culture system with chicken adaptive immune cells (Larsberg *et al.*, 2021). We detected stimulating effects of both strains on the activation (percentage of CD4+CD25+ activated T-helper cells), as well as on the proliferation (percentage of CD4+ T-helper cells) of T-helper cells. In the treatment with *B. amyloliquefaciens,* we found no effect on cytotoxic T-cells. That hints towards an *in vitro* stimulation of T-helper cells through *B. amyloliquefaciens* bacteria which was already observed *in vivo* (Dorigam *et al.*, 2019).

In the experiments with *Bacillus subtilis*, we found stimulating effects on T-helper cells as well as on cytotoxic T-cells. This finding is in line with observations in a previous *in vivo* study in chicken with probiotic *Lactobacillus acidophilus* LA5 (Asgari *et al.*, 2016). Moreover, *Bacillus subtilis* spores, as adjuvants, were shown to enhance CD4+ and CD8+ T-cell responses *in vivo* against avian influenza H9N2 (Lee *et al.*, 2020). In a necrotic enteritis challenge, *Bacillus subtilis* DSM 32315 was shown to reduce pathology and improve performance of broilers. Thus, *Bacillus subtilis* DSM32315 may improve the performance through modulation of T-cell responses in chicken. This ongoing study provides evidence of a direct immunomodulatory effect of *B. amyloliquefaciens* CECT 5940 and *Bacillus subtilis* DSM 32315 on chicken immune cells. Therefore, it can be suggested that the administration of these probiotics can improve the adaptive immune defence in chicken and thereby the usage of antibiotics could be reduced in chicken farming.

References

Asgari F., Madjd Z., Falak R., Bahar M.A. *et al.* (2016) Benef Microbes 7(5):669–675. https://doi.org/10.3920/BM2016.0014

Dorigam J.C.P., Yacoubi N., Doranalli K. (2019) Proc. of the 22nd ESPN 2019, Gdańsk, Poland.

Gharib-Naseri K., Dorigam J.C.P., Doranalli K., Morgan N. *et al.* (2021) Anim Nutr 7(1):185–197. https://doi.org/10.1016/j.aninu.2020.05.004

Larsberg F., Sprechert M., Hesse D., Brockmann G.A., Kreuzer-Redmer S. (2021) Animals 11(12):3600. https://doi.org/10.3390/ani11123600

Lee J.E., Kye YC., Park SM., Shim BS, *et al.* (2020) Vet Res 51:68. https://doi.org/10.1186/s13567-020-00788-8

Lei X., Piao X., Ru Y., Zhang H., *et al.* (2015) Asian-Australas. J Anim Sci 28:239–246. https://doi.org/10.5713/ajas.14.0330

Ma Y., Wang W., Zhang H., Wang J., *et al.* (2018) Sci Rep 8:15358. https://doi.org/10.1038/s41598-018-33762-8

Penaloza-Vazquez A., Ma L.M., Rayas-Duarte P. (2019) Can J Microbiol *65*(10), 762–774. https://doi.org/10.1139/cjm-2019-0019

Sato Y., Kuroki Y., Oka K., Takahashi M., *et al.* (2019) Front Vet Sci 6:26. https://doi.org/10.3389/fvets.2019.00026

Wahyono N.D. and Utami M.M.D. (2018) J Phys Conf Ser 953:012125. https://doi.org/10.1088/1742-6596/953/1/012125

604. Strain and age effects, as well as genetic parameters for adult traits in ostriches

S.W.P. Cloete[1,2*], A. Engelbrecht[3] and Z. Brand[3]

[1]Stellenbosch University, Private Bag X1, 7602 Matieland, South Africa; [2]Western Cape Department of Agriculture, 7607 Elsenburg, South Africa; [3]Western Cape Department of Agriculture, 6620 Oudtshoorn, South Africa; schalkc2@sun.ac.za

Abstract

This study investigated the effects of strain (South African Black – SAB; Zimbabwean Blue – ZB; Kenyan Red – KR) on adult traits in ostriches. SAB females outperformed ZB and KR birds for reproduction traits, but had a lower adult weight than the other strains. KR birds were heavier than SAB birds, but lighter than ZB birds. Feather weight and hair scores were poorer in KR birds than in the other strains. All traits were heritable, estimates ranging from 0.10 for chick production to 0.42 for adult weight. Corresponding values for animal permanent environment ranged from 0.17 for egg production to 0.37 for feather weight and hair score. No conclusive unfavourable genetic correlations were found among traits, thereby simplifying selection decisions for the economically important traits included in the study.

Introduction

Income from an ostrich farming enterprise is derived from skins, meat and feathers (Cloete *et al.*, 2008b). As most skins and meat are produced by surplus birds not retained for breeding in their flock of origin, the importance of reproduction is evident in commercial enterprises. Apart from reproduction, adult weight is a key input trait, which is indicative of lifetime maintenance requirements, while feather weight and hair follicle prevalence has also been identified as key traits in ostriches (Engelbrecht *et al.*, 2019). Hair follicle prevalence can only be scored on the leather obtained from slaughter birds, but the prevalence of bristle hairs on live adult birds may be used as a proxy (Engelbrecht and Cloete, 2021). Furthermore, commercial ostrich production may also benefit from genetics from other ostrich strains (Cloete *et al.*, 2008a; Davids, 2011).

Against this background, we studied strain effects of South African Black (SAB), Zimbabwean Blue (ZB) and Kenyan Red (KR) birds. We also derived genetic parameters for traits that were identified as economically important in adult ostriches.

Materials & methods

The study was undertaken on the multi-breed ostrich resource flock on the Oudtshoorn Research farm of the Western Cape Department of Agriculture (Cloete *et al.*, 2008a; Davids 2011; Brand and Cloete, 2015). Data of pure SAB, ZB and KR breeding pairs were used. Details of data recording for annual egg production (EP) and chick production (CP) of females as well as adult weight (AW) of males and females at the commencement of breeding can be obtained from Cloete *et al.* (2006). These data were recorded over a 31-year period from 1990 to 2020. Total feather weight (TFW) was defined as the total weight of six feather classes, as defined by Brand and Cloete (2009; 2015) and recorded from 2001 to 2020. The presence of bristle hairs was scored (HS) on a 5-point linear scale (negligible – excessive) at the commencement of breeding from 2016 to 2020. The number of records available for analysis amounted to 1,328 for HS, 2,818 for TFW, 3,868 for EP and CP and 8,125 for AW.

All data were analysed with ASReml (Gilmour *et al.*, 2015), fitting the fixed effects of strain (SAB, ZB or KR), sex (male or female, only for AW, TFW and HS), animal age (2- to 10+ years of age) and production year

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_604, © S.W.P. Cloete *et al.* 2022

(as defined previously). Two-factor interactions of strain with age and sex with year were also considered. Once an operational model for a specific trait were identified, the random effects of additive animal, animal permanent environment (PE) and service sire PE (only for EP and CP) were added and evaluated with the log likelihood ratio test. Three-trait analyses were then conducted to draw up a complete set of genetic, animal PE, environmental and phenotypic correlations among traits.

Results

Expressed relative to SAB performance, female EP amounted to 65% in ZB and 72% in KR females ($P<0.05$; Table 1). Corresponding figures for CP were respectively 69 and 79%. ZB and KR birds were respectively 13.3 and 5.5% heavier than SAB birds, the difference between the former strains also reaching significance ($P<0.05$), SAB and ZB birds did not differ for TFW and HS ($P>0.05$). TFW was lower in KR birds than in the other strains, while HS was higher for KR birds (both $P<0.05$).

Strain interacted with animal age for all traits ($P<0.05$). The CP of the SAB strain increased from 2- to 3-year-old females with little subsequent change (Figure 1a). In ZB females, CP increased to around 6 years of age before starting to decline. In KR females, in contrast, the initial increase in CP was followed by a further increase from 8 to 10 years. AW generally increased to maxima at 4-5 years in the SAB and ZB strains and 3-7 years in the KR strain (Figure 1b). AW tended to decline at ages subsequent to the peak weights.

Table 1. Predicted means (±SE) for the effect of strain on egg production (EP), chick production (CP), adult weight (AW), total feather weight (TFW) and hair score (HS) in adult ostriches.[1]

Strain	Trait				
	EP	CP	AW (kg)	TFW (g)	HS (n)
SA Black	46.1±0.5^{b}	21.7±0.3^{b}	120±1^{a}	748±4^{b}	2.35±0.05^{a}
Zimbabwean Blue	30.0±1.1^{a}	15.0±0.8^{a}	136±1^{b}	761±5^{b}	2.32±0.06^{a}
Kenyan Red	33.4±1.8^{a}	17.1±1.3^{a}	126±1^{b}	658±5^{a}	2.91±0.06^{b}

[1] Means with different superscripts differ in collumns ($P<0.05$).

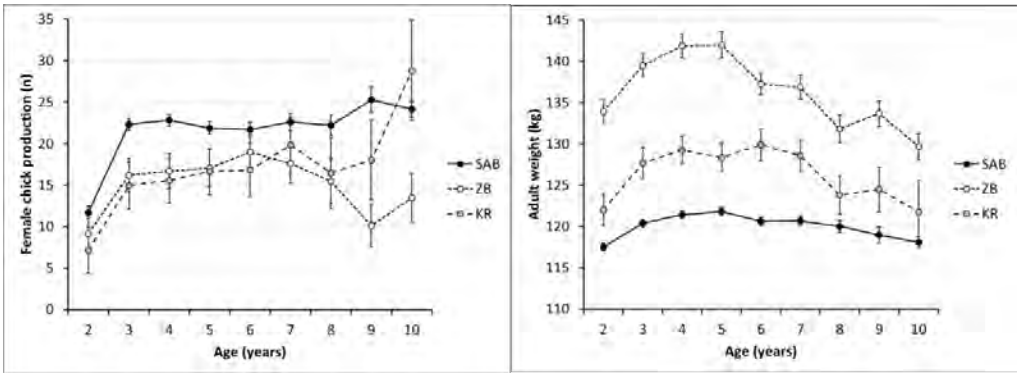


Figure 1. Predicted means (±SE) depicting the interaction of strain with animal age for chick production as a reproductive trait (a) and adult weight (b).

All traits were heritable, with estimates ranging from 0.10 for CP to 0.42 for AW (Table 2). The corresponding range for animal PE values was from 0.17 for EP to 0.37 for TFW and HS. The service sire PE effect was significant for EP at 0.03±0.01 and CP at 0.05±0.01 (not tabulated). The genetic correlation approaching unity between EP and CP was the only significant genetic correlation. The absolute direction of the negative genetic correlations of reproduction with HS was favourable, but countered by positive correlations at the animal PE level (Table 2). Other significant favourable animal PE correlations were between EP and CP, AW and TFW and TFW and HS.

Discussion

This study confirms superior reproduction but a lower mature body weight in SAB birds compared to the other strains (Table 1), which are consistent with previous studies comparing SAB and ZB birds (Cloete *et al.*, 2008a). Davids (2011) also compared all three strains and similarly reported that ZB and KR birds were heavier, but with an inferior performance for CP compared to SABs. The KR birds were intermediate for EP, not differing from either of the other strain in the latter study. The lower feather weight of KR birds compared to the SAB strain was consistent with previous results (Brand and Cloete, 2015). However, these authors, as well as Brand and Cloete (2009) also reported a lower feather weight for ZB birds, which was not found in this study. No previous results for strain effects on HS are available to support the poorer performance of the KR strain in this respect.

The age trend for SAB females for CP in Figure 1a was consistent with the trend reported by Cloete *et al.* (2006). Earlier studies by Cloete *et al.* (1998) and Bunter (2002) reported age trends more similar to those of the ZB in the present study. To our knowledge, it is the first report of the influence of age on CP in KR birds. It is possible that the increase in CP from 8 to 10 years was caused by the retention of reproductive females to later ages, as the KR was by far the smallest group and therefore difficult to maintain. The general trend in AW as animal age changed was consistent with previous reports (Bunter, 2002; Cloete *et al.*, 2006), but the trends for ZB and KR birds seem to be accentuated relative to the SAB birds. No previous studies were done on age effects in the former strains, making it impossible to relate the results obtained to the literature.

Table 2. (Co)variance ratios (±SE) for and among egg production (EP), chick production (CP), adult weight (AW), total feather weight (TFW) and hair score (HS) in adult ostriches.

Parameters and traits	Trait				
	EP	CP	AW (kg)	TFW (g)	HS (n)
Heritability in bold on the diagonal, genetic correlations below and phenotypic correlations above the diagonal					
EP	**0.15±0.04**	0.77±0.01	0.03±0.03	-0.21±0.03	0.08±0.05
CP	0.98±0.03	**0.10±0.03**	0.01±0.03	-0.22±0.03	0.07±0.05
AW	-0.07±0.14	0.02±0.15	**0.42±0.04**	0.17±0.03	-0.02±0.04
TFW	0.03±0.20	-0.20±0.21	0.06±0.15	**0.21±0.05**	-0.14±0.05
HS	-0.27±0.26	-0.34±0.29	0.20±0.19	0.28±0.32	**0.23±0.08**
Animal PE effects in bold on the diagonal, Animal PE correlations below and environmental correlations above the diagonal					
EP	**0.17±0.04**	0.76±0.01	0.16±0.02	-0.22±0.04	0.06±0.05
CP	0.76±0.05	**0.20±0.03**	0.10±0.02	-0.17±0.04	0.06±0.05
AW	-0.07±0.12	-0.13±0.10	**0.30±0.03**	0.02±0.02	-0.01±0.04
TFW	-0.44±0.14	-0.36±0.11	0.43±0.07	**0.37±0.05**	-0.01±0.08
HS	0.40±0.22	0.36±0.19	-0.23±0.13	-0.46±0.14	**0.37±0.08**

Estimates of h^2 for EP and CP were consistent, but at the lower end of the range of previous values reviewed by Cloete *et al.* (2008b). This result could be due to the inclusion of the other strains in the present analysis, as previous results were obtained on SABs only. It did correspond closely with the more recent paper by Cloete *et al.* (2019) on the SAB, though. It is thus plausible that the recent introduction of new birds from commercial strains affected the partitioning of variances between h^2 and animal PE. Service sire PE effects were also similar to the review by Cloete *et al.* (2008b) and results of Cloete *et al.* (2019). The h^2 of AW was consistent with values reviewed by Cloete *et al.* (2008b), but the h^2 of TFW was slightly lower than the previous estimate of 0.30 reported by Brand and Cloete (2009) on SAB and ZB birds only. Brand and Cloete (2015), reported a lower h^2 of 0.12, but it should be remembered that they analysed weights of six different feather classes individually instead of total feather weight. The only previous h^2 for HS in adult birds amounted to 0.23 in the same flock (Engelbrecht and Cloete, 2021). The near-to-unity genetic correlation between EP and CP accorded with previous results (Cloete *et al.*, 2008b).

To conclude, it is evident that ostrich strain affected all traits, the SAB performing better for reproduction and poorer for AW. Moreover, all traits were variable and heritable and should respond to directed selection. No conclusive unfavourable genetic correlations were found, simplifying selection decisions for adult traits in ostriches.

References

Brand, Z., and Cloete, S.W.P. (2009) Proc. of the 18th Association for the Advancement in Animal Breeding and Genetics meeting, Adelaide, Australia.

Brand, Z., and Cloete, S.W.P. (2015) South African Journal of Animal Science 45:20-29.

Bunter, K.L. (2002) PhD Rural Sciences thesis, University of New England, Australia.

Cloete, S.W.P., Van Schalkwyk, S.J., and Brand, Z. (1998) Proc. of the 2nd International Ratite Congress, Oudsthoorn, South Africa.

Cloete, S.W.P., Bunter, K.L., Lambrechts, H., Brand, Z., Swart, D. *et al.* (2006) British Poultry Science 47:147-158. https://doi.org/10.1080/00071660600610658.

Cloete, S.W.P., Brand, M.M., Hoffman, L.C., and Muller, M.M. (2008a) South African Journal of Animal Science 38:65-73.

Cloete, S.W.P., Engelbrecht, A., Olivier, J.J., and Bunter, K.L. (2008b) Animal Production Science 48:1247-1256. https://doi.org/10.1071/EA08135.

Cloete, S.W.P., Engelbrecht, A., Gilmour, A.R., Schou, M.F., Brand, Z., *et al.* (2019) Proc. of the 23rd Association for the Advancement in Animal Breeding and Genetics meeting, Armidale, Australia.

Davids, A.H. (2011) MSc Animal Sciences thesis, Stellenbosch University, South Africa.

Engelbrecht, A., Cloete, S.W.P., and Amer, P.R. (2019) Proc. of the 23rd Association for the Advancement in Animal Breeding and Genetics meeting, Armidale, Australia.

Engelbrecht, A., and Cloete, S.W.P. (2021). Bristle hairs, filoplumes and pinholes in ostriches. World Poultry Congress. https://viewr.wpcparis2022.com/2021/10/05/bristle-hairs-filoplumes-and-pinholes-in-ostriches/.

Gilmour, A. R., Gogel, B. J., and Welham, S. J. 2015. ASReml User Guide Release 4.1 Structural Specification.

Part 35

Bees and other insects

605. Economic value and production characteristics of table honey

E.A. Frost[1,2*], N.C. Chapman[3], R.G. Banks[1], S.F. Walkom[1] and S. Hermesch[1]

AGBU, a joint venture of NSW Department of Primary Industries and University of New England, 2351, Armidale, Australia; [2]New South Wales Department of Primary Industries, Paterson, NSW, 2421, Australia; [3]University of Sydney, Macleay Building A12, Science Rd, Camperdown 2006, Australia; elizabeth.frost@dpi.nsw.gov.au

Abstract

Most honey produced globally is classed as table honey, for direct consumption, as distinct from higher value medicinal grade honeys. Economic value of honey was based on domestic table honey price in AU$ derived from Australia (AU) and United States (US), from 2010-2021 and sourced from AU honey processors and the US Department of Agriculture. For 2021, average AU and US domestic honey values were AU$5.70 and AU$6.53 per kg respectively. Domestic price slopes for 2010-2021 were similar for AU and US at an increase of AU$0.28 per year. Characteristics for honey production (HP) per hive per year were collated from four manuscripts and three AU datasets, presenting a substantial range of means (8.5 to 123.3 kg), and standard deviations (2.6 to 42.8 kg). Variability in HP between studies is attributed to differences in phenotyping, genotype, environment and production scale.

Introduction

Honey accounts for 85% of Australian (AU) beekeeper income (ABARES, 2016). Surveyed AU beekeepers nominated honey production (HP) as the top trait for selection (Chapman and Frost, 2020). Due to lack of selection for the trait, a need to improve overall beekeeper management, and variable access to nectar-yielding plant species, HP is not fully realised. In regard to genetic improvement, Petersen's (2020) review of honey bee breeding program design notes selective breeding in honey bees is primarily undertaken by researchers for a single trait or in small-scale production systems, with little uptake from or benefit to beekeepers operating large-scale production systems. Such large-scale systems are common in AU. The National Honey Bee Genetic Improvement Program (Plan Bee) will develop a breeding program to improve productivity of the large-scale AU beekeeping industry. As a first step towards this aim, deriving the economic value of table honey and evaluating HP phenotype data on individual hives with fixed queens over twelve months is necessary because the beekeeping industry lacks formalised breeding objectives for large-scale production systems. Literature was reviewed for the economic value of table honey. A lack of economic value, besides subjective weighting of traits by percentage value out of 100, (Van Engelsdorp and Otis, 2000, Zakour and Bienefeld, 2014) necessitated the present study to derive it. This study derives the economic value of honey in two beekeeping industries with comparable production systems (AU and United States (US)) and summarises HP characteristics.

Materials & methods

The economic value for honey was derived and prices over time for two comparable markets (AU and US domestic and AU export) were documented. Cost of production was examined in the AU context. Published and unpublished AU data for HP per hive in large-scale systems across one full and two partial twelve month seasons were compared with other studies including HP characteristics.

Results

Production characteristics varied considerably between studies (Table 1). Variability reflects differences in phenotyping, genotype, environment and production system. Phenotyping methods included estimation of HP by number of honey combs times a factor of 2.5 (Zakour *et al.,* 2012), measuring weight difference of

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_605, © E.A. Frost *et al.* 2022

Table 1. Overview of honey production characteristics for hives with a fixed queen over single or multiple honey harvesting events in one or more twelve month periods.[1]

Study	Mean	*S*	Min/Max	PS
Banks et al. (2020) AU Pop 1 2018-19	41.5	18.2	0.0/136.0	Large
Banks et al. (2020) AU Pop 2 2019-20	123.3	42.8	25.6/195.9	Large
Bienefeld and Pirchner (1990) Germany	24.1	13.9	0.0/99.8	Mixed
Brascamp et al. (2018) Austria	30.0	15.8	0.1/202.0	Mixed
Zakour et al. (2012) Syria Pop 1	8.5	3.3	0.5/23.8	Mixed
Zakour et al. (2012) Syria Pop 2	10.2	2.6	1.0/20.0	Mixed
Plan Bee AU 2021	44.2	14.8	0.0/96.6	Large

[1] *S* = standard deviation; *Min* = minimum; *Max* = maximum; *PS* = production system; *Pop* = population.

combs before and after extraction of honey plus an estimate of honey remaining in the broodnest (Bienefeld and Pirchner, 1990), and weight difference of combs before and after honey extraction by itself (Banks *et al.*, 2020, *Brascamp et al.,* 2018, Plan Bee, 2021). Subspecies of honey bees analysed included *Apis mellifera syriaca* (Zakour *et al.,* 2012), *Apis mellifera carnica* (Bienefeld and Pirchner, 1990, Brascamp *et al.,* 2018) and *Apis mellifera* hybrids (Banks *et al.*, 2020, Plan Bee, 2021). Production systems can reflect hives fixed in one location or migrated between locations, small (<200 hives) or large scale (>200 hives), and mixed (both small and large scale included). A mean of 30 kg for HP across 18 Austrian production years (Brascamp, 2016) compared to means of 41.5 and 44.19 for partial and 123.27 for full AU production years (*Banks et al., 2020* and Plan Bee, 2021 datasets) illustrates the difference between European and AU production systems and environments. Datasets from AU (Banks *et al.*, 2020, Plan Bee, 2021) were generated from large scale, migratory production systems. Other datasets (Bienefeld and Pirchner, 1990, Brascamp *et al.,* 2018, Zakour *et al.,* 2012) included both stationary and migratory small (< 200 hives per beekeeper) and large-scale production systems (>200 hives per beekeeper).

Economic value. The economic value of honey was derived from the price in AU$ of one kilogram of honey. Average honey prices from AU and US domestic markets and AU export markets from 2010 to 2021 were sourced from the AU beekeeping industry and US Department of Agriculture. Average annual exchange rates from 2010 to 2020 (www.exchangerates.org.uk) were used to convert US$/lb to AU$/kg for US domestic and US$/kg to AU$/kg for AU export prices. Linear regression was used to quantify the change in honey price for different markets over time. The export market price for table honey was higher than domestic market prices in recent years. The high export price in 2020 is a result of depressed domestic supply and increased demand from overseas markets. The economic value of honey was AU$5.70/kg for the 2021 AU domestic market, with 2022 economic value of AU$6.42/kg expected. Linear regression of table honey price for AU and US domestic shows a steady increase in domestic price (Figure 1). Honey price (AU and US domestic) over time generated a positive, parallel slope of 0.28. Export price (AU) yielded a slope of 1.32. Price for table honey increased consistently over time with minimal variation from the overall slope which explained 81 to 85% (coefficients of determination) of the variation in price of honey over time.

Expected costs to large scale AU producers for each added kg honey per hive are hardware, labour and freight. Large variability in HP across AU hives (Table 1), indicates that an increase of 1 kg per hive can be captured within existing hardware (i.e. honey boxes, cooperative-owned bulk containers) and available labour, negating any increase in equipment or labour at this stage (Banks *et al.*, 2020, N.G. Bingley, C.C. Cooper, C.F. Horner, personal communications). Given our mean, freight is the only added cost to produce additional honey (AU$0.04/kg) at the current time so the economic value of HP is given as price per kilogram AU$5.70 on the domestic market.

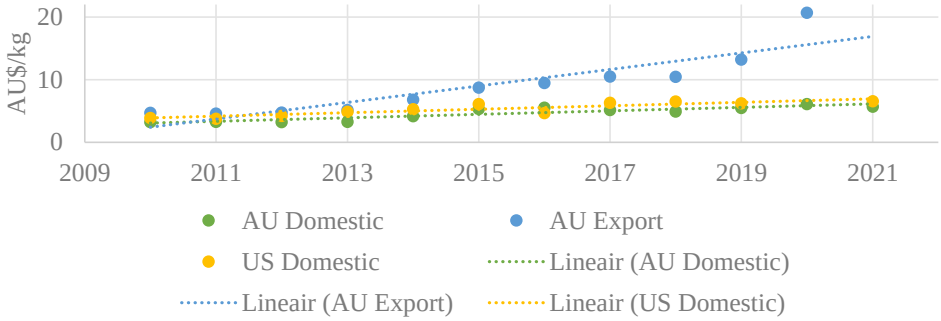


Figure 1. Honey (AU$/kg) price over time in different markets.

Discussion

Price differences illustrate variation in economic value for table honey over time and between markets which need to be considered in bee breeding programs. Although economic value was not found in the literature, the price weighting of honey as a subjective percentage (34.4±2.40% and 55.3% respectively), in addition to other traits totalling 100%, shows a range of trait emphasis in breeding objectives specific to production system, genotype and environment between Canada and Syria (Van Engelsdorp and Otis, 2000, Zakour and Bienefeld, 2014). This highlights the need for AU economic models to derive relative economic values for top AU traits. Secure growth in the AU domestic market supports increased selection for HP. Where export market targeting is possible at the producer level, an extra value upward of AU$15/kg, excluding marketing costs, is expected. Datasets from AU are testament to environmental conditions for HP which surpass those in other study environments. Additionally, the AU datasets result from large-scale production systems which migrate hives to multiple honey making events in a twelve-month period. Comparatively, other datasets include a mix of small and large-scale production systems and both migratory and stationary honey producers. The substantial range in HP characteristics between the literature and AU datasets highlights the need to establish the differences between datasets in other countries and AU genotypes, environment and production systems to ensure accurate analysis. AU datasets presented are indicative of large-scale production systems where colonies are transported to subsequent honey production environments unique in aspects of climate, plant species, nectar production and duration to those northern hemisphere environments within the literature. Phenotyping is currently underway in AU through Plan Bee. Phenotypic data for AU-preferred traits and pedigree information will enable selection tailored to the AU environment, thereby improving the viability of the AU beekeeping industry and contributing to a global model applicable to large-scale production systems.

Acknowledgements

The provision of honey data and wholesale prices by the AU beekeeping industry and critical comments by Kim Bunter are gratefully acknowledged.

References

ABARES, Australian honey bee industry: 2014–15 survey results. Available at: www.agriculture.gov.au/abares/media-releases/2016/combing-through-honey-bee-industry.

Banks R.G., Frost E., Chapman N., Boerner V. *et al.* (2020) AgriFutures Australia Publication No. 21-087. Available at: www.agrifutures.com.au/wp-content/uploads/2021/08/21-087.pdf

Bienefeld K. and Pirchner F. (1990) Apidologie 21: 175-183. https://doi.org/10.1051/apido:19900302

Brascamp E.W., Willam A., Boigenzahn C., Bijma P., and Veerkamp R.F. (2018) Apidologie 49: 462–463. https://doi.org/10.1007/s13592-018-0573-3

Chapman, N.C. and Frost, E.A. Plan Bee: Beekeeper and queen bee breeder surveys 2020. Available at: www.agrifutures.com.au/wp-content/uploads/2021/05/Plan-Bee-beekeeper-and-queen-bee-breeder-surveys-2020.pdf

Petersen G.E.L, Fennessy P.F, Amer P.R., and Dearden P.K. (2020) Journal of Apicultural Research 59(4): 638-647. https://doi.org/10.1080/00218839.2020.1715583

USDA (2021), Table 46. Available at: www.ers.usda.gov/webdocs/DataFiles/53304/TABLE46.XLS?v=0

Van Engelsdorp D. and Otis G.W. (2000) Journal of Economic Entomology 93(6): 1606–1612, https://doi.org/10.1603/0022-0493-93.6.1606

Zakour, M.K. and Bienefeld K. (2014) Journal of Apicultural Research 53(2): 314-326. https://doi.org/10.3896/IBRA.1.53.2.13

Zakour M.K., Ehrhardt K., and Bienefeld K. (2012) Apidologie 43(5): 600-607. https://doi.org/10.1007/s13592-012-0130-4

606. Evaluating the potential for mating control in honey bee breeding in three SE European countries (preliminary results)

A. Uzunov[1,2*], S. Andonov[3], B. Dahle[4], M. Kovačić[5,6], J. Prešern[7], G. Aleksovski[1,2], F. Jaman[5], M. Jovanovska[1], B. Pavlov[2,8], Z. Puškadija[5,6], J. Wegener[9] and R. Büchler[10]

[1]Faculty of Agricultural Sciences and Food, Ss. Cyril and Methodius University in Skopje, Macedonia; [2]Company for Applied Research and Permanent Education in Agriculture, Skopje, Macedonia; [3]Swedish University of Agricultural Sciences, Uppsala, Sweden; [4]Norwegian Beekeepers Association, Norway; [5]Faculty of Agrobiotechnical Sciences Osijek, Croatia; [6]Centre for applied life sciences Healthy Food chain Ltd, Osijek, Croatia; [7]Agricultural Institute of Slovenia, Slovenia; [8]Eko Pcela, Skopje, Macedonia; [9]Bee Research Institute, Hohen Neuendorf, Germany; [10]Landesbetrieb Landwirtschaft Hessen, Bieneninstitut Kirchhain, Germany;
uzunov@fznh.ukim.edu.mk

Abstract

The study was conducted to explore alternatives for mating control as a part of honey bee breeding programs for genetic improvement & conservation of local honey bee (*Apis mellifera*) populations in Croatia (HR), Macedonia (MK) and Slovenia (SI). We observed nuptial flights of 87 virgin queens (30 in HR, 35 in MK and 22 in SI) on potential sites. Mating success was related to the presence of drone producing colonies, but even at locations chosen for their isolation from known apiaries, mating success of 70% was achieved. On average, queens performed nuptial flights on 1.8 days in HR, 1.6 to 4.3 days in MK and 1.6 to 2.4 days in SI. Unsuccessful nuptial flights were two/threefold shorter than successful flights.

Introduction

Breeding programs for genetic improvement and conservation of local honey bee (*Apis mellifera*) populations have been in place for at least a decade in Croatia (HR), Macedonia (MK) and Slovenia (SI). However, the selection progress has been slower than expected due to the lack of efficient mating control as one of the crucial elements of the breeding programs (Uzunov *et al.*, 2017, Plate *et al.*, 2019). The mating behaviour of honey bees is complex: the queen performs nuptial flights and mates with 7-17 drones that can originate from many different colonies within a 10 km radius (Koeniger *et al.*, 2014). High colony density and lack of geographically isolated areas are recognised as the main factors for the unsatisfactory employment of mating control. However, by scouting possible isolated places or utilizing other means of mating control, such unfavourable situations may be counteracted. Such studies include identifying isolated locations by employing on-field observations and complementary laboratory investigations, investment in technical and human capacities in case of instrumental insemination, or intensive cooperation with surrounding beekeepers if an overflow with selected drones is the strategy. The project BeeConSel, funded by EEA & Norway Grants Fund for Regional Cooperation (2020-2023) provides an opportunity for investigation of alternatives for overcoming the deficiency of existing mating control. The preliminary results from the field investigations by observations on the queens' nuptial flights in the project's initial year are presented and discussed in detail in the next paragraphs.

Materials & methods

In June and July of 2021, the field investigations and data collection were conducted by direct observations. Virgin queens were placed into drone-free mini mating boxes (nucs) with around 700 worker bees and solid food, and kept 72 h in a dark and cool room. Forty-eight hours before the observations began, they were placed in the observation sites with an installed queen excluder on the entrance (Büchler *et al.*, 2013; Scheiner *et al.*, 2013). Nuptial flights were observed on 87 queens: 30 in Croatia (*A. m. carnica*), 35 in Macedonia (*A. m. macedonica*) and 22 in Slovenia (*A. m. carnica)*. The candidate locations were

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_606, © A. Uzunov *et al.* 2022

selected in each country based on the following main criteria: expected low density or absence of managed and/or feral colonies, geographical isolation, good logistical accessibility, and suitable weather/climate. In Croatia, observations were done on a flatland mating station (Batina, 240 m a.s.l.) flooded with drones originating from 96 drone producing colonies (DPC), headed by sister queens. No other managed colonies were expected to be present withing a diameter of 6 km. In Macedonia, the observations were done on three micro-locations in the frame of the mountainous region of Mavrovo plateau (>1,680 m a.s.l.). At the location MK-Belicki, 17 managed colonies were identified 2.5 km away at the end of observations. The locations MK-Crkva and MK-Cavkarnik were at 3 km air-distance from the next same apiary, that was lower in altitude by 400 m. Following the observations, the queens in the last two locations were allowed to freely fly for two additional days. Three locations were investigated in Slovenia in deep Alpine valleys: Vrata (1,020 m a.s.l.) with direct observation and no known managed/feral colonies present, Krma 0 (960 m a.s.l) with five DPC at the location and Krma 1 which was 1 km further in the valley at 1,015 m a.s.l. Three queens in Krma 0 were observed by video surveillance. The field study was done following a tailored protocol for direct observation (Uzunov *et al.*, *in preparation*). For at least 5 days from 11:00 to 17:00 queens' nuptial flight parameters were monitored and recorded: flight frequency, flight duration, observed mating sign and weather conditions. The direct observations were done by installing an extension on the mating box entrance with a transparent cover and queen excluder allowing observation and control of any queen flight attempt (Koeniger, 1981). At the time of the field studies, the weather conditions were suitable in all sites, although with noticeable variations (particularly wind) between the micro-locations in Macedonia.

Results

The number of observed, active and mated queens, the average number of flying days and flight frequency, as well as the queen's average flight duration with or without the mating sign per country and location are presented in the Table 1.

A mating success of over 76% was achieved in HR-Batina where 96 DPCs were placed and over 91% in MK-Cavkarnik where an apiary was at a distance of 3 km. In contrast, the mating success in SI-Krma 1, where 5 DPCs were present 1 km away from the locations with mating boxes, was only 56%. In SI-Vrata,

Table 1. Descriptive statistics of the main parameters from the observations. The minimum and maximum values are given in round brackets.

	Number of queens		Aver. per queen		Aver. flight dur./queen**	
Country	Observed [Flew]	Mated* [%]	Flying days	Flight frequency	w/o mating sign (min)	with mating sign (min)
HR-Batina[1]	30 [28]	23 [76.7]	1.8 (1-3)	3.8 (2-6)	8 (1-28)	14 (4-22)
MK-Belicki[2]	12 [12]	5 [41.7]	4.3 (2-6)	8.8 (2-16)	6 (1-21)	22 (8-32)
MK-Crkva[2]	11 [9]	6 [54.5]	1.6 (0-3)	2.6 (0-7)	7 (1-55)	19
MK-Cavkarnik[2]	12 [11]	11 [91.7]	2.2 (0-3)	3.9 (0-10)	6 (1-20)	12 (3-18)
SI-Vrata[3]	10 [7]	7 [70.0]	2.4 (1-5)	2.8 (0-5)	11 (1-32)	26 (18-34)
SI-Krma 1[1]	9 [5]	5 [55.6]	2.4 (2-4)	2.6 (0-10)	10 (2-29)	20 (17-23)
SI-Krma 0[1,4]	n.a.	n.a.	1.6 (1-2)	2.3 (0-3)	5 (3-9)	11 (3-18)

[1] With the presence of DPC at the mating station.

[2] Managed/feral colony(ies) found at the location.

[3] No managed /feral colony(ies) found at the location.

[4] Video observed, video analysis incomplete, *Values based on sampled sealed worker brood, **Values based on partial or complete observations of queens' nuptial flights.

where no managed or feral colonies were known to exist 70% of the virgin queens succeeded in mating. In the remaining two Macedonian locations MK-Belicki and MK-Crkva, the mating success was below 55%.

The average number of flying days per queen and the average flight frequencies were in the expected biological range except the high values in MK-Belicki (average of 4.3 flying days and 8.8 flights per queen). The lowest values for average number of flying days were recorded in HR-Batina (1.8) and SI-Krma 0 (1.6). Almost on all locations, the average flights' duration per queen returning to the mating box without a mating sign were two to threefold shorter than the flights with a mating sign. The highest average differences of 3.7 times, between flights with and without observed mating sign, was recorded in MK-Belicki.

Discussion

In HR-Batina, where the location was saturated with a sufficient number of drones, the results of high mating success (76.7%) show the importance of the availability of a high number of mature airborne drones. Under such conditions, the queens flew on a low number of days (1.8) and had a regular flight frequency (3.8) per queen, reducing the risk of additional flights and engendering possible losses, clearly listing the location as a good candidate for setting up a mating station, and making the approach of 'drone flooding' appear promising. With a queens' flight speed of around 20 km/h (Koeniger *et al.*, 2014) and an average flight duration of 14 minutes for those returning with a mating sign, we estimated an average distance of 1.5 km to the potential mating site(s), taking into account the time of around 5 min for the mating *rendezvous* with the drones (Koeniger *et al.*, 2014). Still, a successful flight as short as 4 minutes indicate that the queen was mated close by.

With five DPCs employed at the locations SI-Krma 1, partial mating success was obtained with 56%. According to Tiesler *et al.* (2016), the required number of DPCs ranges from 8 for 50 virgin queens to a minimum of 20 DPCs for 500 virgin queens. Nevertheless, due to the small sample size, the results from SI-Krma 0 needs to be interpreted with caution even though the flight activities indicate that the queens' mating was efficient. On the other hand, the unsatisfactory results from SI-Krma 1 might be a consequence of the ambiental conditions with brief sunny periods due to high mountain walls or rough transportation to the location. Finally, the results from the supposedly drone-free location SI-Vrata, show that there seems to be a significant drone 'noise' from the surrounding area; however, the average duration of mating flights – both with or without a mating sign, was the longest of all studied locations. This may indicate that distance alone is not always sufficient to predict the suitability of a candidate location.

In Macedonia, we recognised the possibility of using MK-Belicki as the most prosperous location for establishing a reliable mating station. All queens were flying but only 42% of queens (n=5) were producing worker brood. The above-average number of flying days and high average frequency of 8.8 flights per queen indicate a queen's excessive efforts for finding partners (Uzunov *et al.*, 2014) even under the fact that 17 managed (but not DPC) colonies were later discovered 2.5 km away. At MK-Crkva, almost 82% of the queens flew but 55% of the queens (n=6) successfully produced worker brood even with the existence of an apiary in the vicinity of 3 km. However, the queens' flight activities were lower and similar to the locations saturated with drones. The persistent wind occurrence recorded during the observations might affect queens flight performance at this location (Koeniger *et al.*, 2014). MK-Cavkarnik, with the highest mating success (91.7%) among the Macedonian locations, even it is on the same distance from the next apiary as MK-Crkva is (but in different direction), seems an unsuitable choice for setting up a mating station. Such prediction is also based on the queens' flight activities and in particular, the short duration of 3 minutes for the performed successful flight in MK-Cavkarnik. Koeniger and Koeniger (2007) and Heidinger *et al.* (2014) found shorter mating flight duration in an area saturated with drones, while in the area with a low number of colonies, mating flights were longer (Koeniger *et al.*, 1989). The obvious relationship between

the duration of mating flights and the number/vicinity of potential mating partners, also observed in the present experiment, suggest that this parameter can be used as an additional indicator for judging of the suitability of a location for isolated mating.

In order to answer the question of the achievable mating purity with greater certainty, an analysis of the drone's origin and semen composition are indispensable. This information will be available before the beginning of the next beekeeping season. Therefore, in the following two seasons, the activities will be focused on verifying the current locations using sufficient DPCs or exploring new sites by observations on the queens and drones' nuptial flights.

References

Büchler R., Uzunov A., Kiprijanovska H., and Andonov S. (2012) Proc. of SICAMM Conference, Landquart, Switzerland

Büchler R., Andonov S., Bienefeld K., Costa C., Hatjina F., *et al.* (2013). J. Api. Res. 52(1):1 -30. https://doi.org/10.3896/IBRA.1.52.1.07

Heidinger I.M.M., Meixner M.D., Berg S., and Büchler R. (2014) Insects 5:513-527. https://doi.org/10.3390/insects5030513

Koeniger G. (1981). Apidologie 12(4): 329-343. https://doi.org/10.1051/apido:19810403

Koeniger G., Koeniger N., Pechhacker H., Ruttner F., and Berg S. (1989) Insects Soc. 36:129–138. https://doi.org/10.1007/BF02225908

Koeniger N., and Koeniger, G. (2007) Apidologie 38:606-611. https://doi.org/10.1051/apido:2007060

Koeniger G., Koeniger N., and Tiesler F-K. (2014). Paarungsbiologie und Paarungskontrolle bei der Honigbiene. Buschhausen Druck und Verlaghaus – Herten, Germany

Plate M., Bernstein R., Hoppe A., and Bienefeld K. (2019) Gen Sel Evol 51(74):.1–14. https://doi.org/10.1186/s12711-019-0518-y

Scheiner R., Abramson C.I., Brodschneider R., Crailsheim K., Farina W.M. *et al.* (2013) J. Api. Res. 52(4):1–58. https://doi.org/10.3896/IBRA.1.52.4.04

Tiesler F-K., Bienefeld K., and Büchler R. (2016) Selektion bei der Konigbiene. Buschhausen Druck und Verlaghaus - Herten, Germany

Uzunov A., Büchler R., Pavlov B., Andonov S., and Kiprijanovska H. (2014) Proc. of AG Tagung Conference, Marburg, Germany

Uzunov A., Brascamp E.W., and Büchler R. (2017) Bee World 94(3):84-87. https://doi.org/10.1080/0005772x.2017.1345427

607. A model for, and early implementation of, genomic selection in the Australian honey bee population

N.C. Chapman[1], E.A. Frost[2,3] and R.G. Banks[3*]

[1]University of Sydney, Macleay Building A12, Science Rd, Camperdown 2006, Australia; [2]New South Wales Department of Primary Industries, Paterson, NSW, 2421, Australia; [3]AGBU, a joint venture of NSW Department of Primary Industries and University of New England, Armidale, NSW, 2350, Australia; rbanks@une.edu.au

Abstract

Genetic improvement has been limited in the Australian honey bee population, despite the direct and indirect economic significance of the industry. Recently, significant industry and community investment has established the foundation for a national R&D and improvement program centred on objective evaluation and genomic selection, underpinning strong industry and community engagement. This paper outlines the key elements of that program, its goals, and early steps in implementation, and highlights the transformative potential of the program.

Introduction

The Western honey bee (*Apis mellifera*) was introduced to Australia within the wider process of colonisation and development of British/European agricultural practices, and now contributes to the Australian economy in two ways:

- directly, through honey production, which is worth an estimated $AUD150m per year; and
- indirectly, through pollination 'services' provided to a range of introduced crops. In the absence of alternative means of pollination, the value of these services has been estimated at upwards of $AUD5-6bn per year.

In recent years the critical importance of the honey bee has been brought into sharper focus by awareness of threats affecting honey bee populations globally, and in particular by the risk posed by the Varroa mite (*Varroa destructor*), which is currently not present in Australia. Australia maintains strict border quarantine protocols against Varroa (and other threats to honey bees), but it is generally accepted that the mite will enter and establish in the country at some point. Further, importation of genetic material from overseas populations is very limited, again reflecting the quarantine protocols.

A further challenge is the highly variable nature of the Australian climate, with frequent droughts impacting production, and making performance recording across time challenging.

Despite these contributions and recognition of threats, there have been limited systematic wide-scale attempts at genetic improvement of the population. There have been studies outlining scope for improvement via selection and estimating potential value of such improvement (Oldroyd, 1984, Banks *et al.*, 2020). It has become apparent that this situation likely reflects, firstly, very limited performance recording, meaning that queens are marketed primarily on the reputation of the breeder, and secondly (and essentially the flipside of the same issue) no real incentive for queen breeders to invest time and effort in such recording.

Together, these facts mean that the industry has little or no awareness of objectively described differences in genetic merit and what they mean for honey production and/or pollination economics, and extremely limited datasets. While genetic improvement focussed on honey production and health traits would deliver private benefits, there is clear evidence of a market failure in their delivery, and a strong case for government and industry intervention.

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_607, © N.C. Chapman *et al.* 2022

A recent response to this risky situation has developed through two investment steps:

- establishment of a research honey bee population at Tocal College, a research and outreach centre of the NSW Department of Primary Industries, with capacity to manage upwards of 250 hives/colonies for phenotyping at any one time;
- support for a collaborative R&D program known as Plan Bee, which aims to build on the infrastructure establishment of the facility at Tocal.

This paper outlines how these responses establish a platform for industry-wide implementation of genomic selection, and the opportunities that would present.

Key elements of Plan Bee

As noted above, systematic performance recording in honey bees is very limited in Australia, with essentially three sustained programs of any scale. These three programs have provided queens to the research facility at Tocal College, augmented by sampling from a range of smaller queen breeding operations across Australia, to establish initial samples of c. 250 queens. These queens have been introduced to hives, and detailed recording of performance, including weight of honey produced, temperament, and health traits, has commenced, along with sampling for genotyping.

Alongside, and complementing this core activity, are:

- widespread industry consultation to establish a Data Definition Manual (Chapman and Frost, 2021)– describing and defining traits which could be recorded;
- development of an industry database and procedures for analysis of data collected to produce estimates of genetic merit;
- a strong engagement effort to encourage performance recording and submission of data to the industry database; and
- careful data collection focussed on pollination activities in conjunction with key providers and users of pollination services.

The target outcomes of the first, current phase of Plan Bee activity include:

- the trait definition manual (Chapman and Frost, 2021);
- extension materials for performance recording;
- extension materials on genetic information and principles of genetic improvement;
- objectively evaluated queens, as potential sources of genetic material for industry;
- the initial core of a genomic reference population, incorporating queens sampled at tocal along with queens managed by industry queen breeders;
- initial models for wider industry participation in recording and utilisation of evaluated genetic material, including industry consultation around feasibility, practicality, etc.

While there is much to be done in development of these models, the combination of a sound reference population with large-scale multiplication of elite material provides scope for relatively rapid industry change and return on industry and government co-investment. In this sense the timing of this initiative in terms of availability of genomic technologies, coupled with the inherent reproductive capacity of honey bees and use of techniques such as AI, is extremely serendipitous (and change may be less limited by the need to achieve very widespread performance recording).

Current funding has a limited term, but it is anticipated that momentum from early progress will underpin continued co-investment by industry and government to maintain a program.

Potential impacts of a successful Plan Bee

Plan Bee has the potential to transform the Australian honey bee industry and community, through a range of important outcomes.

The scope for direct utilisation of elite genetic material is essentially unlimited, and is already being taken up. Backed by strong communication around the expected differences in hive performance, and the value of performance data, it is expected that this access could rapidly lift the performance of hives on a wide scale, which in turn should strengthen already widespread community interest in bees and beekeeping, and stimulate increased data collection. The awareness of how widely people can contribute to 'helping our bees' should develop a sense of community responsibility and opportunity.

At the professional level, the scope for accessing elite material can establish the basis of a more certain business model for queen breeders, multipliers and providers of pollination services. In this context, strong extension of the concept of genetic merit and how to use objective estimates of it, will be fundamental.

Together, these point to a much more engaged and integrated industry and beekeeping community, which is in itself valuable.

Earlier studies have suggested ample scope for genetic improvement in honey bees in Australia, and there is no reason why this cannot be realised given a sufficiently large recorded population and use of sound genetic analysis. The most challenging aspect of this outcome will be to develop methods to mitigate risks around Varroa. Plan Bee data recording will include hygienic tests, but the opportunity will also exist to collaborate with research and industry in other countries to screen genomic variation in the Australian population against reference datasets that include survival and performance in the presence of Varroa. Should Varroa enter Australia, the infrastructure of a well-recorded and genomically characterised population should enable a more rapid response. In this context, it is valuable that a small importation of Varroa-resistant genetic material has been sampled for inclusion in the first-year intake for Plan Bee.

The wide sampling of the Australian honey bee population already underway, coupled with genotyping of all sampled material, will provide a basis for immediate implementation of selection using genomic information to maintain diversity. Previous work suggests the existence of diversity at specific loci and between small samples of the population, but establishing as wide a sample base of the population as possible will maximise capacity to maintain genetic progress in the absence of importation from overseas.

While the model for an ongoing program is to be developed, initial modelling suggests (Banks *et al*., 2021) that a program centred on recording and genotyping of c. 2,000 queens per year, can generate valuable progress and be attractive economically, simply based on increased honey production. The modelling assumed ongoing national costs (database and analysis, extension and communication), but there is scope to develop approaches that combine collective (government and/or industry) and private contributions. Such models will need to incorporate suitable forms of participation or membership, with a wide range of options potentially available.

Collaboration with overseas programs has been flagged earlier in the context of Varroa, but more generally, having a sound and broadly based recording and evaluation program in place in Australia will greatly enhance scope for international collaboration, both in breeding and more broadly in R&D. In particular, understanding of genomic relatedness between countries' populations may enable a more nuanced approach to import and export of genetic material, which could benefit both Australia and international partners.

Conclusions

Plan Bee offers, and is already starting to deliver, a unique opportunity for transformation in the Australian honey bee industry. This transformation can include wide involvement and participation, very valuable genetic progress, and massive reduction of production and disease risk.

While this is obviously exciting for Australia and its own honey bee industry, there may be valuable outcomes for other countries, and at the same time, a model of industry transformation focussed on genomic selection underpinned by mixed (i.e. public and private) investment and strong industry and community engagement. This may provide learnings useful to other industries in both the 'developed' and 'developing' countries (Marshall *et al*., 2019).

Acknowledgements

Support from NSW DPI, AgriFutures Australia, the Australian Department of Agriculture, Water and the Environment through the RRD4P program, breeders and industry sponsors is gratefully acknowledged.

References

Australian Bureau of Agricultural and Resource Economics and Sciences (ABARES), Australian honey bee industry: 2014–15 survey results. Available at: www.agriculture.gov.au/abares/media-releases/2016/combing-through-honey-bee-industry

Banks R.G., Frost E., Chapman N., Boerner V. *et al.* (2020), AgriFutures Australia Publication No. 21-087. Available at: www.agrifutures.com.au/wp-content/uploads/2021/08/21-087.pdf

Chapman N.C. and Frost E.A., Plan Bee: Beekeeper and queen bee breeder surveys 2020. Available at: www.agrifutures.com.au/wp-content/uploads/2021/05/Plan-Bee-beekeeper-and-queen-bee-breeder-surveys-2020.pdf

Chapman N.C. and Frost E.A. (2021), AgriFutures Australia Publication No. 21-089. Available at: www.agrifutures.com.au/wp-content/uploads/PlanBeeBreedingManual-002.pdf

Marshall K., Gibson J.P., Mwai O., Mwacharo J.M., Haile A. *et al.* (2019) Front. genet. 10:297. https://doi.org/10.3389/fgene.2019.00297

Oldroyd B. (1984) Unpublished PhD Thesis, University of Sydney

608. An over-time distribution map of the main *Apis mellifera* mitochondrial DNA lineage using honey as source of eDNA

V. Taurisano*, V.J. Utzeri, A. Ribani and L. Fontanesi

Department of Agricultural and Food Sciences, University of Bologna, Italy; valeria.taurisano2@unibo.it

Abstract

A growing interest across Europe is emerging to maintain the genetic integrity of honey bee populations and subspecies. In this study we combined different approaches to obtain an over-time distribution map of the main *Apis mellifera* mitochondrial DNA (mtDNA) lineages in the North of Italy. mtDNA was analysed from individual honey bees and from honey. Honey is a source of environmental DNA (eDNA), which also contains traces of the DNA of the honey bees that have produced it. A total of about 1,600 honey samples produced in the North of Italy over four years (2018-2021) where analysed to detect the A, C1, C2 and M mtDNA lineages. The obtained results will be useful to design conservation strategies of *A. mellifera* genetic resources in Italy. We demonstrated that eDNA from honey can be exploited to design a cost-effective population genetic study in *A. mellifera*.

Introduction

Because of the role that *Apis mellifera* plays in the agroecological systems, a growing interest across Europe is emerging to maintain the genetic integrity of honey bee populations and subspecies, which might be well adapted to local environments (De la Rúa *et al.* 2009). A combination of different factors threats the genetic integrity of *A. mellifera* populations. For example, beekeeping activities, including the transhumance of colonies and extensive trading of queen bees increase admixture between populations. The effect of harmful pesticides for the honey bees, the action of extreme climatic waves and their consequence on honey bee health can cause the loss of many colonies which could contribute to the reduction of the effective population size and incentive the use of non-autochthonous queens to replace dead families.

More than 30 *A. mellifera* subspecies and ecotypes have been described using morphometrical descriptors (Ruttner, 1988) which have been integrated with information on mitochondrial DNA (mtDNA) lineages and sub-lineages and nuclear genome polymorphisms (Cournet and Garnery,1991; Wallberg *et al.*, 2014). Due to its geographical position, Italy is a unique case in Europe as it hosts natural populations of four *A. mellifera* subspecies (Fontana *et al.*, 2018; Franck *et al.*, 2000; Ruttner, 1988; Vecchi, 1927): *A. m. ligustica* (which mainly carries the C1 mitotype but also the M7 mitotype) whose origin is from the Italian peninsula; *A. m. siciliana* (which has mitotypes of the African originated A lineage), native of the Sicily island; *A. m. mellifera* (which carries M mitotypes), originally distributed in small areas of the Western Alpine arch; and *A. m. carnica* (which mainly carries the C2 mitotype), with natural populations in the North-East part of Italy, close to the border with Austria and Slovenia (Sinacori *et al.*, 1998; Muñoz *et al.*, 2009; Franck *et al.*, 2000).

Several national and regional legislative acts have been introduced in order to protect Italian native honey bee subspecies. Starting from 2017 current legislation takes initiatives to safeguard the subspecies of *A. m. ligustica* by limiting or banning other subspecies used in beekeeping practice, including the use of hybrid (if not natural) in some Italian regions (Fontana *et al.* 2018).

We recently developed two assays to discriminate the main three mtDNA lineages (A, C and M) of *A. mellifera* and the two main mitotypes (C1 and C2) of the C lineage using honey as a source of environmental DNA (eDNA), which also contains traces of the DNA of the honey bees that have produced it (Utzeri *et*

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_608, © V. Taurisano *et al.* 2022

al. 2018, 2022). We already applied the method that discriminates the A, C and M lineages to obtain an updated distribution map of the A. mellifera mitotypes over all Italy by analysing more than 600 honey samples produced in the year 2018 (Utzeri *et al.* 2021).

In this study, using honey bees and honey samples collected from the same colonies and region, we first compared the morphometric results and the mtDNA lineage frequency obtained by analysing DNA extracted from honey bees and honey. Then using many other honey samples produced in four years in all regions of the North of Italy, we obtained a distribution map of the A, C1, C2 and M mitotypes based on eDNA recovered from this food matrix.

Materials & methods

Honey bees and honey samples. Eighty samples of honey bees (~60 workers), each from a single colony from which honey (extracted from a single honey comb) was collected, were obtained from 45 beekeepers of the Emilia-Romagna region in the year 2020. These honey bees were morphometrically characterized for the identification of the *A. mellifera* subspecies of the family. For these colonies, 3-10 workers for each colony and the corresponding honey samples were used for DNA analysis. In addition, a total of 1588 commercial honey samples, produced over four years from apiaries located in all regions of the North of Italy (from west to east and north to south: Liguria, Piedmont, Valle d'Aosta, Lombardy, Trentino-Alto Adige, Friuli-Venezia Giulia, Veneto and Emilia-Romagna) were provided directly by many beekeepers. Table 1 reports the number of honey samples per region and year of production. All these samples were used for DNA analysis.

DNA analyses. DNA extraction from honey samples was performed following the protocol previously described by Utzeri *et al.* (2018). DNA extraction from individual workers was carried out using the Wizard® Genomic DNA Purification Kit (Promega, Promega Corporation, Madison, WI, USA) following the manufacturer's instructions. DNA extracted from honey was analysed with the PCR primers and method described by Utzeri *et al.* (2018) that can distinguish the three main mtDNA lineages (A, C and M) and with the Sanger sequencing assay reported by Utzeri *et al.* (2022) that discriminate the C1 and C2 mitotypes. As honey samples, if not directly sampled from a single honey comb, can be derived from more than one colony, more mtDNA lineages can be identified with the methods described by Utzeri *et al.* (2018, 2022). DNA extracted from individual worker and larvae samples was amplified with the primer pair E2–H2 (Garnery *et al.* (1992) that target the intergenic COI-COII mtDNA region. Sanger sequencing was then used to identify the amplified mitotypes after sequence data analyses carried out with MEGA X (Kumar *et al.* 2018) and BLASTN.

Table 1. Number of commercial honey samples analysed and produced by beekeepers in different regions and over the four years (2018-2021).

Region	2018	2019	2020	2021	Total per region
Liguria	6	7	9	10	32
Piedmont	56	84	136	91	367
Valle d'Aosta	11	22	26	11	70
Lombardy	74	98	126	128	426
Trentino Alto Adige	20	21	18	14	73
Friuli-Venezia Giulia	22	13	17	26	78
Veneto	41	24	69	51	185
Emilia Romagna	100	42	153	62	357
Total per year	330	311	554	393	1588

Results

Morphometric analysis of the honey bees indicated that about 60% of colonies sampled in Emilia-Romagna could be considered of *A. m. ligustica*. All honey samples obtained from these families had only the C mitotype using the method of Utzeri *et al.* (2018). These results were confirmed by mtDNA analysis of the single honey bees belonging to the same colonies. Among the remaining 40% of these colonies, five of them had honey that had the A+C+M mitotype, confirmed by individual honey bee analyses. This could be probably due to substitutions of the queens or drift of honey bees from close colonies in the apiary.

Considering all honey samples produced in Emilia-Romagna over all four years (n. 367; the region that accounted the largest number of analysed samples), about 77-82% of them had only the C mitotype. The proportion of honey samples with only the C mitotype tended to increase in the latest years. An overview of the distribution of the honey samples with different mitotypes in the four years and divided by administrative provinces of this region is shown in Figure 1. Honey samples that had only the C mitotype were also analysed to distinguish between the C1 and C2 sub-mitotypes. About 10% of these honey samples had both C1+C2 or only C2 mitotypes.

A similar distribution of honey with different mitotypes was also obtained in all other regions of the North of Italy. Honey samples with the C2 mitotype were more frequent in the North-East regions (i.e. Trentino-Alto Adige, Friuli-Venezia Giulia and Veneto), which are close to the border with Austria and Slovenia where *A. m. carnica* can be considered the autochthonous honey bee subspecies.

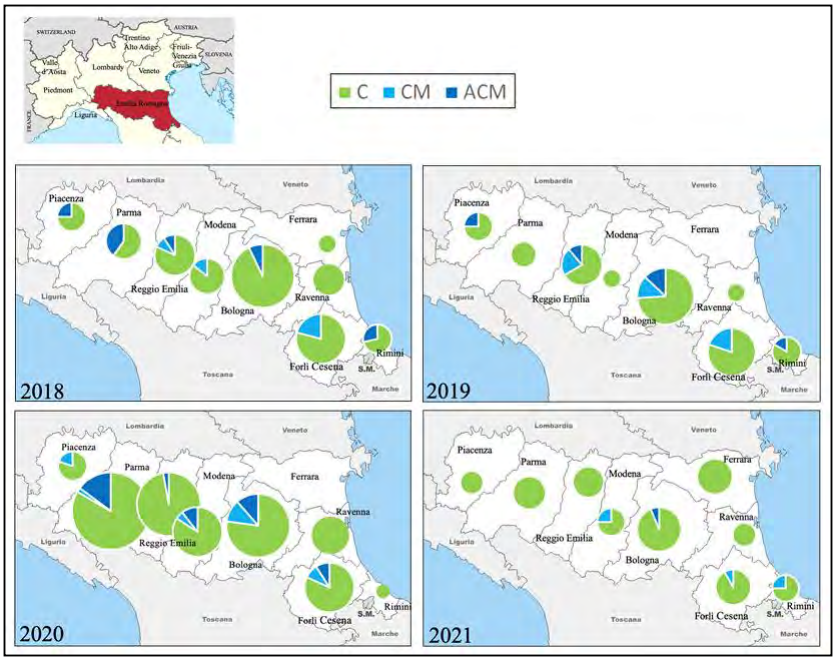


Figure 1. Distribution of honey samples produced in the Emilia-Romagna region (in red in the top set) in the years 2018-2021 that had different mitotypes, as obtained using the method of Utzeri *et al.* (2018). Information is also divided by the administrative provinces of this region. The size of the charts is proportional to the number of honey samples analysed. The different colours in the charts indicates the honey samples that had different mitotypes or their combinations.

Discussion

The combination and comparison of the different approaches and methods that we used gave a general overview of the *A. mellifera* populations in the North of Italy. As honey DNA is highly degraded (Utzeri *et al.* 2018), only short fragments can be amplified which are not fully informative. Therefore, only the main mtDNA lineages or target informative mtDNA sites can be analysed. Despite that, eDNA from this matrix could provide an interesting overview of the A. mellifera mitotypes present in the North of Italy. The current picture might be derived by the admixture of different sub-species as already reported (Utzeri *et al.* 2021). Mitochondrial DNA analysis can integrate morphometric evaluations. As similar results were obtained considering individual bee mitotype analyses and mtDNA lineage detection in honey samples, the use of eDNA from honey can be considered very useful to design a cost-effective population genetic study in *A. mellifera*, even over time. We are working to implement the analysis of nuclear genome polymorphisms using honey DNA to further strengthen the usefulness of eDNA in honey bee population genomic studies.

References

Cornuet, J. M., Garnery, L. (1991) Apidologie 22, 627-642. https://doi.org/10.1051/apido:19910606

De la Rúa, P., Jaffé, R., Dall'Olio, R., Muñoz, I.; Serrano, J. (2009) Apidologie, 40, 263–284. https://doi.org/10.1051/apido/2009027

Fontana, P., Costa, C., Di Prisco, G., Ruzzier, E., Annoscia, *et al.* (2018) Bull. Insectol., 71, 257–271. http://hdl.handle.net/10449/53488

Franck, P., Garnery, L., Celebrano, G., Solignac, M., Cornuet, J.M. (2001) Mol. Ecol., 9, 907–921. https://doi.org/10.1046/j.1365-294x.2000.00945.x

Muñoz, I., Dall'Olio, R., Lodesani, M., De la Rúa, P. (2009) Apidologie 40, 617–626. https://doi.org/10.1051/apido/2009041

Ruttner, F. (1988) Biogeography and Taxonomy of Honeybees; Springer: Berlin, Germany, DE.

Sinacori, A., Rinderer, T.E., Lancaster, V., Sheppard, W.S. (1998) Apidologie 29, 481–490. https://doi.org/10.1051/apido:19980601

Utzeri V.J., Ribani A., Fontanesi L. (2018) Food Control 91, 294–301. https://doi.org/10.1016/j.foodcont.2018.04.010

Utzeri, V. J., Ribani, A., Taurisano, V., Fontanesi, L. (2022) Food Control 134, 108713. https://doi.org/10.1016/j.foodcont.2021.108713

Vecchi, A. (1927) Sulla distribuzione geografica dell'*Apis mellifica ligustica* Spin. in Italia. Bollettino del Laboratorio di Zoologia Generale e Agraria Della R. Scuola Superiore d'agricoltura in Portici, 20, 150–168.

Wallberg, A., Han, F., Wellhagen, G., *et al.* (2014) Nat. Genet. 46, 1081-1088. https://doi.org/10.1038/ng.3077

609. Using genomics to predict drone quality: why are there so many 'dud' male honey bees?

G.P. Slater* and B.A. Harpur

Department of Entomology, Purdue University, 901 W. State St. Lafayette, IN 47905, USA; garett.p.slater@gmail.com

Abstract

Beekeepers are faced with severe challenges that affect their colonies' health. Among the most highly-reported of those challenges is queen failure. Queens can fail because they have mated with low-quality drones. In this study, we aim to develop a rapid, high-throughput method for beekeepers and queen producers to identify high-quality line using genomics. We use a multi-omics, field-based approach to quantify the genetic features associated with poor sperm production in drones. To do this, we compare the reproductive quality of drones sourced from lines available to beekeepers in the US and raised in a common apiary. We then created crosses of these lines and will use these crosses to identify regions of the genome associated with reproductive traits. We found variation in sperm traits, which is likely inherited. Once identified, these makers will allow beekeepers to identify reproductively-viable lines and reduce the chance of queen failure in their yards.

Introduction

One of the most pressing issues facing commercial beekeepers in the United States is the poor quality of queens they are provided from out-of-state suppliers. For example, over 220 beekeepers in North Dakota (one of the most active beekeeping states of the United States) manage approximately 667,000 colonies. Many of these colonies are re-queened during the spring, a relatively new practice that has become common commercial beekeepers because of the high rates of queen failure. Queen failure is among the top three reported causes of colony death (Steinhauer *et al.*, 2014). A shocking 30% of commercially-purchased queens used for replacement fail within six months of establishment in a colony (Pettis *et al.*, 2016). The cause of queen losses is multifarious, but poor-quality sperm is a primary culprit (Pettis *et al.*, 2016).

In the United States, where beekeepers lost an average of ~30% of their colonies over the 2016-2017 year (Steinhauer *et al.*, 2014), queen failure needs to be addressed to create a sustainable industry. Male honey bees – drones – play a critical role in determining a queen's likelihood of surviving (Slater *et al.*, 2021). Queens can fail because they have mated with males who produced low-quality sperm. If drones are not producing healthy sperm (Fisher *et al.*, 2018), their mates will not produce healthy colonies (Slater *et al.*, 2021). Beekeepers can see these effects: a symptom of poor drone sperm quality includes spotty brood pattern, slow colony build-up and high supersedure rates.

We know comparatively little about the role genetics plays in shaping drone reproductive traits (Rangel and Fisher, 2019). Genetic variation across United States populations likely contributes to low quality male bees and queen failures. Honey bees are non-native to North America and originate from centuries of importation from genetically- and geographically-distinct lineages. F_{st} ranges from 0.3 to 0.6 on average among native sub-species populations across Eurasia (Harpur *et al.*, 2014). Further, among the fastest-evolving genes in the honey bee genome are those associated with spermatogenesis (Harpur *et al.*, 2014). This rapid evolution is likely due to variation in mating behavior across the honey bee's range (Wallberg *et al.*, 2014) and may have led to variation in male reproductive senescence, strategy, and morphology (Slater *et al.*, 2021). This standing genetic variation can contribute to phenotypic differences observed among honey bee line in North America (Slater *et al.*, 2021) and contribute to the success (or failure) of

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_609, © G.P. Slater and B.A. Harpur 2022

colonies. Initial work in our lab demonstrates that there is variation in the reproductive quality of drones originating from four independent sources around the United States (unpublished). One major question for beekeepers, bee breeders, and researchers is how much do drone reproductive characters differ among stocks of honey bees? If reproductive variation among stocks does have a genetic component, we could be able to predict the quality of drones through sequencing.

Methods reproductive phenotyping. Our first goal is to establish when and how reproductive traits (flight behavior and sperm quality) vary among the honey bee lines. To accomplish this, we paired passive data collection (Figure 1) with high throughput sperm phenotyping. Purdue research apiary maintains four lines commonly used by beekeepers. These lines are often used by beekeepers and purchased at a premium because they are an 'identified' line or because they express valuable traits. For ten colonies within each line, we caged queens on drone comb for 24 hours and reared the resulting eggs until drones emerged as adults. We then marked 20 adult drones from each colony (n=1000) with RFID tags (Microsensys, Atlanta, GA; Figure 1B) and randomly placed drones into one of two colonies kept separately at two independent yards (to prevent drift). Each colony is fitted with an RF detector that will log when drones leave and return to their colonies. Drones leave colonies regularly to mate with queens and they do so at specific times of the day (Slater *et al.*, 2021). By tracking this behavior, we can uncover how much variation exists among lines for when drones leave to mate with queens and how frequently they leave.

Drones begin to sexually mature at 10-21 days old (Metz and Tarpy 2019). Beginning on day ten, we began sampling ten drones from each line from within our two host colonies. Each drone was first identified by its RFID tag and then semen was collected. Sperm was immediately stored in sperm diluent (Slater *et al.*, 2021) and then tested for viability and concentration using a standard fluorescent-based assay (Metz and Tarpy 2019). We used an Opera Phenix HCS system, a confocal robotic- and software-assisted system that can visualize 96 fluorescently-labelled samples in three dimensions and count labelled sperm cells (live, dead and total). This protocol resulted in an un-parallel look at the sperm quality, sperm morphology, and reproductive behavior (when drones leave to mate) for the four honey bee lines available to beekeepers in the United States.

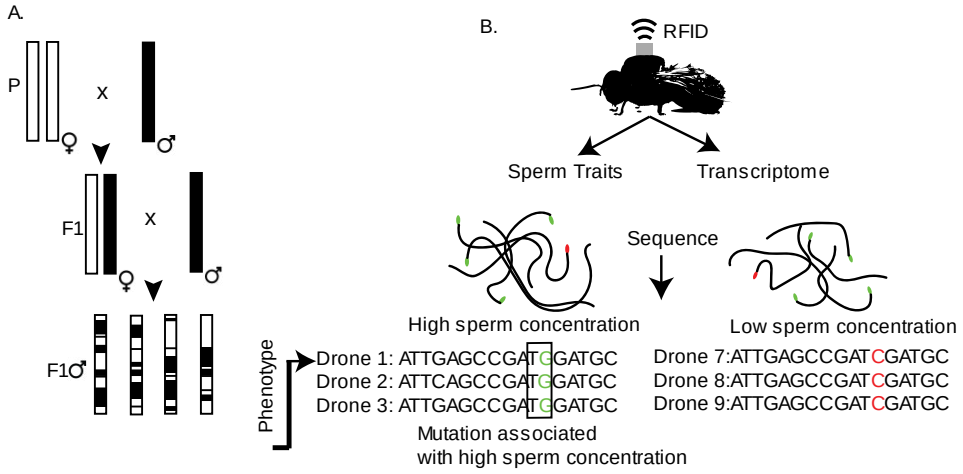


Figure 1. Experimental design. (A) Using RFID tags, we tracked male fight behavior. Beginning on day 10 (peak male reproductive age), we assessed male sperm viability and concentration. (B) We created F2 crosses from the two lines with largest differences in reproductive traits. These F2 cross were then phenotyped for reproductive traits along with the parentals.

Crossing parental stocks. In order to identify the variants associated with drone reproductive traits, we produced genetic crosses between the two most divergence lines (Figure 1A). Drones from these two lines (labeled Line 1 and Line 2) differed significantly in both flight and sperm traits. We found Line 1 drones produce more sperm and flew for longer durations at age 14 than Line 2. These significant line differences allow us to quantify how genetics contributes to reproductive variation and pinpoint the regions of the genome that contribute to this variation. Crosses were performed at Purdue university's research apiary. To produce genetic crosses, queens from each line began as mated queens from their original beekeeper source or location to ensure the queens mated with local drones. Line 1 (shown in Figure 1A as diploid chromosomes in white) queens were crossed to line 2 drones (haploid and in black) to create ten F1 queens. We then produced F2 open-mated queens from the F1 crosses. As a result of recombination, each queen produced drones that are mosaic in ancestry across the genome. These F2 drones expressed the phenotype of whichever parental allele (black or white) they inherited (Figure 1A, example highlighted in green). We then repeated our phenotyping (Figure 1B, above) with 1000 F2 drones each tagged with RFID tags.

Results

We used a Generalized Linear Mixed Model (GLMM) to test whether line had a strong effect on drone reproductive traits. Additional fixed effects include age, body weight, and age of first flight. We found breeding line and age had the most significant impact on sperm concentration (Line: $P<0.0001$, Age: $P<0.0001$) and mean flight duration at age 14 (Line: $P<0.0001$, Age: $P<0.0001$). We then isolated the most divergent lines and found that Line 1 produced drones with significantly more sperm that flew longer durations than Line 2 (Figure 2A; ANOVA; Sperm Concentration: $F_{1,161}=12.345$, $P<0.001$; Mean Flight Duration: $F_{1,167}=4.870$, $P=0.00884$). In addition, Line 1 and 2 drones differed in body weight (ANOVA $F_{1,167}=3.9554$, $P=0.04673$), sperm viability (ANOVA $F_{2,170}=7.286$, $P=0.00092$), proportion Day 14 maturity ($X^2=12.404$, $P=0.00042$), flights per day (ANOVA $F_{1,161}=10.395$, $P<0.0001$), first flight age (ANOVA $F_{1,161}=7.372$, $P=0.00086$), head length (ANOVA $F_{1,161}=10.452$, $P<0.0001$), and front wing length (ANOVA $F_{1,161}=13.564$, $P<0.0001$). Because of these results, we produced genetic crosses between only Lines 1 and 2.

F2 crosses were produced from Lines 1 and 2 (Figure 3) and phenotyped for reproductive traits. The F2 drones produced significantly more sperm than Line 2 (ANOVA $F_{2,473}=10.452$ $P<0.0001$), but less than Line 1 (ANOVA $F_{1,473}=2.462$, $P=0.117$). Thus, inherited markers are associated with sperm production.

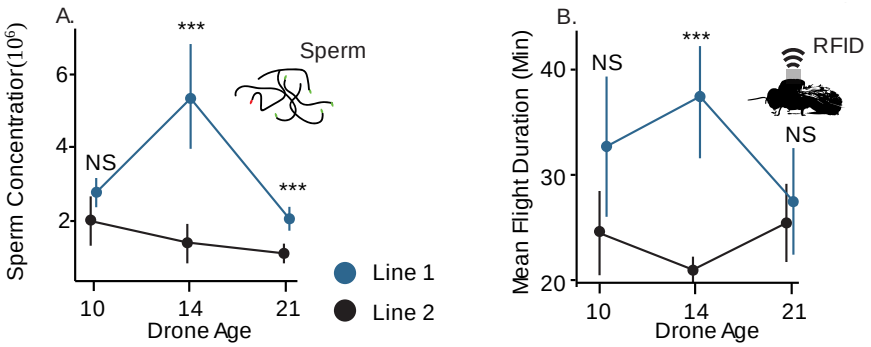


Figure 2. Line comparisons for flight and sperm traits. Line 1 and 2 displayed the most divergence in reproductive traits. (A) Drone sperm concentration from Line 1 was significantly higher than Line 2 at ages 14 and 21 ($P<0.0001$; ANOVA). (B) Line 1 drone flight duration was also higher than Line 2 at age 14 ($P<0.05$ ANOVA), but not 10 and 21.

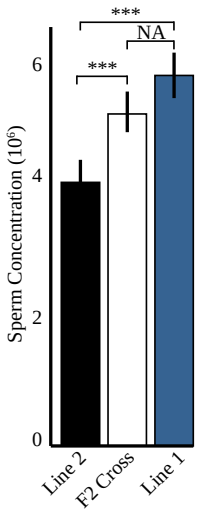


Figure 3. Parental and F2 phenotyping. Drones from Line 1 produced significantly more sperm than Line 2 ($P<0.05$; ANOVA). and the F2 cross ($P=0.117$; ANOVA). The F2 produced more sperm than Line 2.

Discussion

We found significant variation among honey bee breeding lines in both flight and sperm reproductive traits (flight duration and sperm concentration), especially between Line 1 and Line 2. We then produced genetic crosses between Line 1 and 2, and phenotyped these alongside the original parental colonies. We found the F2 colonies produced drones with sperm concentration higher than Line 2, but lower than Line 1. This signifies drones express the phenotype of the inherited parental allele (Figure 1). Thus, there is an important genetic component to this drone reproductive trait.

Next, we will identify genetic markers associated with flight behavior and sperm quality of drones (Figure 1). By integrating phenotypic and genotypic data for both parental and F2 drones, we can associate the effects of segregating genetic variation (genetic markers) on phenotypic variation. These markers will allow beekeepers to identify reproductively viable breeding lines and reduce the chance of queen failure in their yards. Furthermore, it empowers beekeepers to incorporate genetic information into breeding programs.

References

Fisher, A., Harrison, K., Love, C., Varner, D., and Rangel, J. (2018) Southwestern Entomologist 43(2):343-356. https://doi.org/10.3958/059.043.0206

Harpur, B.A., Kent, C.F., Molodtosova, D., Lebon, J.M.D., Alqarni, A.S. *et al.* (2014) Proceedings of the National Academy of Sciences 111(7):2614-2619. https://doi.org/10.1073/pnas.1315506111

Pettis, J.S., Rice, N., and Joselow, K. (2016) PloS one 11(2):e0147220. https://doi.org/10.1371/journal.pone.0147220

Rangel, J. and Fisher, A. (2019) Apidologie 50(6):759-778. https://doi.org/10.1007/s13592-019-00684-x

Slater, G.P., Smith, N.M.A., and Harpur, B.A. (2021) Genes 12(8):1251. https://doi.org/10.3390/genes12081251

Steinhauer, N.A., Rennich, K., Wilson, M.E., Caron, D.M., Lengerich, E.J. *et al.* (2014) Journal of Apicultural Research 53(1):1-18. https://doi.org/10.3896/IBRA.1.53.1.01

Wallberg, A., Han, F., Wellhagen, G., Dahle, B., Kawata, M., *et al.* (2014) Nature genetics 46(10):1081-1088. https://doi.org/10.138/ng.3077

610. Development of genomic resources in black soldier fly (*Hermetia illucens* L.) via throughput DNA pool sequencing

A.S.L. Donkpegan[1*], A. Guigue[2], F.X. Boulanger[3], S. Brard-Fudulea[1], P. Haffray[4], M. Sourdioux[1] and R. Rouger[1]

[1]SYSAAF, Centre INRAE Val de Loire, 37380 Nouzilly, France; [2]INNOVAFEED 85 rue de Maubeuge 75010 Paris, France; [3]AGRONUTRIS 35 Boulevard du libre Échange, 31650 Saint-Orens-de-Gameville, France; [4]SYSAAF LPGP/INRAE, Campus de Beaulieu, Université de Rennes 1, 35042 Rennes, France; armel.donkpegan@inrae.fr

Abstract

Single nucleotide polymorphism (SNP) arrays are a powerful genotyping tool used in genetic research and selective breeding. The black soldier fly (BSF) *Hermetia illucens* L. is becoming an economically-important entomoculture species. However, the lack of genotyping tools can imped investment in more efficient selection programs based on genomic selection. The purpose of this study was to call SNPs from DNA pool sequencing that will be used later to design high, medium and low-density SNPs chips. Pooled DNA library preparation was performed using several BSF populations from the breeding companies Innovafeed and Agronutris. Generated library pools were sequenced on Illumina platform in 2×150bp and the sequences were assigned to the samples based on their index sequence. Over 300 million of Paired-End reads and 43X of sequencing depth was obtained on average per pool. The last BSF genome with chromosome assembly level (≈1 Gb) was used as reference to call 52 million SNPs. Filters based on Quality, Minor Allele Frequency (MAF), genome coverage (depth) and non-informative sites were used to yield a final set of 9.7 million high quality SNPs that can be later used to develop the first genotyping chip for BSF.

Introduction

Use of genomic tools is now widespread in livestock and other animal species to support genetic improvement e.g. for pedigree tracing and genomic or markers assisted selection. Black Soldier Fly (BSF) is a promising candidate for the development of the insect farming industry. In this species, pioneering works led to the contig-assembly of two reference mitochondrial genomes (Qi *et al.*, 2017; Zhan *et al.*, 2020) as well as a high quality chromosomal-scale nuclear genome (Generalovic *et al.*, 2021). According to our knowledge, genomic resources as SNP markers are still limited and/or publicly available for this species, which can limit the development of selective breeding programs.

With the development of next-generation sequencing (NGS), the cost of sequencing has declined significantly, making the identification of genomic resources feasible in non-model species. Among the panoply of sequencing strategies available, whole-genome sequencing of pools of individuals (Pool-seq) is an effective strategy because it's used for the identification of high-quality SNPs (Bansal, 2010), and provides a cost-effective alternative to sequencing individuals separately (Schlotterer *et al.*, 2014).

In this work, sequencing of DNA pools was used to detect high-quality SNPs in BSF. We expect that the development of these genomic resources will provide a basis for the development of genomic tools to support selective BSF breeding in the future.

Materials & methods

The methodological approach used in this study consists of the following steps: (1) preparation of several BSF pool samples from the French breeding companies Agronutris and Innovafeed; (2) Total DNA was extracted at IRBI (CNRS, France) using the Modified NucleoMag Tissue method (Macherey-Nagel) with automation on epMotion 5075; (3) Preparation of genomic libraries using the Illumina DNA Prep kit; (4)

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_610, © A.S.L. Donkpegan *et al.* 2022

Sequencing of the libraries on the Illumina platform in 2×150 bp in order to obtain a sequencing depth of 30X per pool samples, i.e. a global output of 100 Gb of data; (5) Sorting of the sequences per pool samples. Steps 3, 4 and 5 were performed at GenoScreen.

Sequence quality was evaluated with FastQC version 0.11.15 (http://bioinformatics.babraham.ac.uk/projects/fastqc). Low quality bases and adapter contamination were removed with Trimmomatic version 0.33 (Bolger *et al.*, 2014) with the following options: ILLUMINACLIP 2:30:10, LEADING 3, TRAILING 3, SLIDINGWINDOW 4:15, MINLEN 36. Reads were mapped [BWA 0.7.5a-r405 (Li and Durbin, 2009)] to the last high-quality *H. illucens* reference genome (Generalovic *et al.*, 2021) using default parameters. The resulting alignments were converted to BAM format and PCR duplicates were removed using PicardTools (v1.12). The resulting files were realigned around indels using SAMtools 0.1.17 (Li *et al.*, 2009). Base quality score recalibration and novel SNP discovery were carried out using HaplotypeCaller algorithm of Genome Analysis Toolkit (GATK) v4.0 with standard parameters (DePristo *et al.*, 2011). VCFtools (http://vcftools.sourceforge.net/) was used to remove INDELs variation and retain only biallelic variants (SNPs).

The initial SNP set was filtered with multiple steps and parameters. First, we calculated the minor allele frequencies (MAFs) and missing rates using VCFtools, and the SNPs with MAF≤0.05 and missing rate ≥0.1 were excluded. We also filtered the SNPs using several quality criteria: Qual By Depth (QD), Fisher Strand (FS), Strand Odds Ratio (SOR), RMS Mapping Quality (MQ), depth (MIN_DEPTH=30; MAX_DEPTH=150).

Results

The sequencing run results allow us to obtain an average sequencing coverage of 43X per pool (Table 1).

These pool-seq sequencing data were aligned to the BSF reference genome, and more than 63 million variants (SNPs, INDELs, etc.) were detected. After the filtering to retain only biallellic SNPs, we obtained a preliminary VCF file with 52 million SNPs (Figure 1).

From this data, filters based on quality, missing rate, Minor Allele Frequency (MAF), genome coverage (depth), and non-informative sites were used to obtain a final set of 9.7 million high-quality SNPs (Figure 2) that can be used later to develop the first genotyping chip for BSF.

Table 1. Number of paired-end reads and sequencing depth obtained per sample pool.

Pool samples	Number of paired-end reads	Output (in Gb)	% ≥ Q30[1]	Sequencing coverage
Pool 1	306,393,544	46.27	92.46	46X
Pool 2	282,164,146	42.61	92.86	43X
Pool 3	259,851,540	39.24	92.42	39X

[1] Quality score is usually annotated QXX, where 'XX' represents the score. This is interpreted as follows: for one base, there is a probability of error (Pε), with $P\varepsilon = 10^{(-XX/10)}$ and therefore, $QXX = -10 \times \log_{10} P\varepsilon$. A Q30 score is associated with an error rate of 1/1000 (0.1%). For a 2×150 bp Illumina run, the specifications are to obtain more than 80% of reads with a score >Q30.

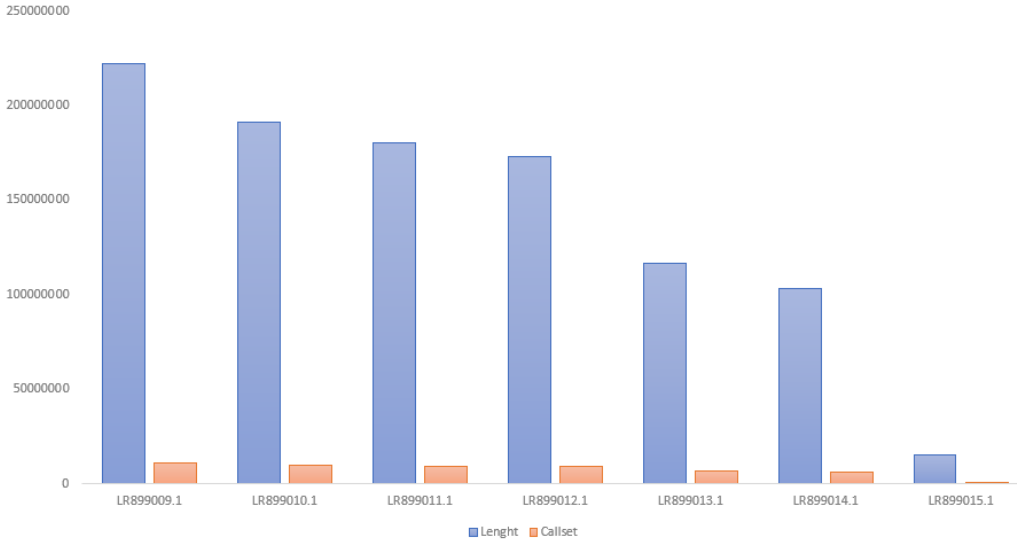


Figure 1. Distribution of (the 52 million) identified SNPs compared to seven chromosome size in BSF genome. For chromosome LR899009.1, 10,826,924 SNPs were called out of 222,122,703; For LR899010.1, 9,784,094 SNPs out of 191,142,057; For LR899011.1, 9,430,128 SNPs out of 180,357,958; For LR899012.1, 8,868,232 SNPs out of 173,077,783; For LR899013.1, 6,481,853 SNPs out of 116,824,456; For LR899014.1, 5,996,179 SNPs out of 103,444,952; For LR899015.1, 598,190 SNPs out of 15,435,017.

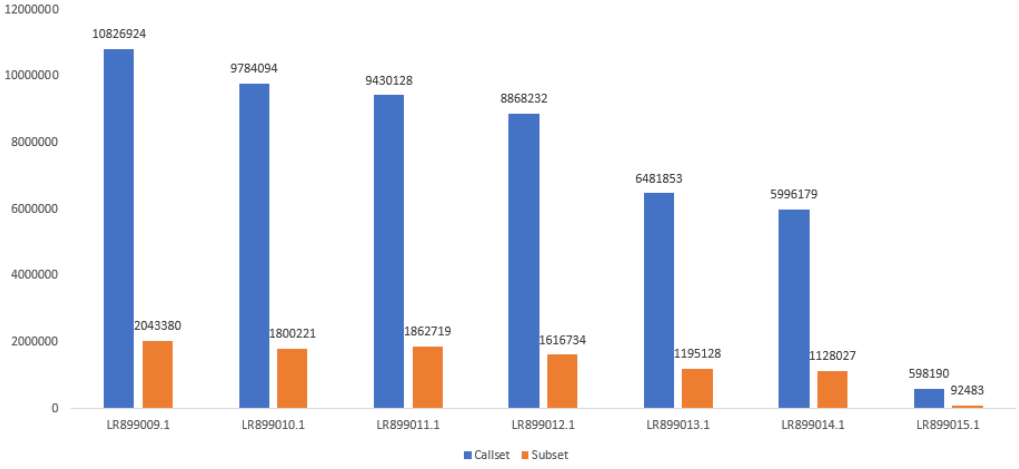


Figure 2. Distribution on the seven BSF chromosomes of the 9.7 million selected SNPs in comparison to 52 million identified in Figure 1.

Conclusions

Large-scale genome sequencing of different BSF origins by pool sequencing allowed us to identify a high-quality and diverse candidate SNP set for BSF's nascent breeding industry. Beyond their use in genomic selection, these genomic resources will be useful to improve breeding practices in BSF e.g. to manage inbreeding with adapted panels for DNA parentage assignment or pedigree tracing, to design more efficient breeding programs, to identify chromosome regions associated with performances (QTL) by association genetic studies (GWAS) and/or also speed up diffusion of broodstock with higher breeding values.

References

Bansal, V. (2010). Bioinformatics, 26 (12): i318–i324. https://doi.org/10.1093/bioinformatics/btq214

Bolger, A. M., M. Lohse, and B. Usadel. (2014). Bioinformatics 30: 2114–2120, https://doi.org/10.1093/bioinformatics/btu170

Depristo M.A., E. Banks, R. Poplin, K. V. Garimella, J. R. Maguire, *et al.* (2011). Nature Genetics 43: 491–498, https://doi.org/10.1038/ng.806

Generalovic T.N., S.A. McCarthy, I.A. Warren, J.M.D. Wood, J. Torrance, *et al.* (2021). G3 Genes|Genomes|Genetics, 11, 5, jkab085, https://doi.org/10.1093/g3journal/jkab085

Li, H., and R. Durbin (2009). Bioinformatics 25: 1754–1760. https://doi.org/10.1093/bioinformatics/btp324

Li, H., B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, *et al.* (2009). Bioinformatics 25: 2078–2079. https://doi.org/10.1093/bioinformatics/btp352

Qi Y., J. Xu, X. Tian, Y. Bai & X. Gu (2017). Mitochondrial DNA Part B2 (1): 189-90. https://doi.org/10.1080/23802359.2017.1307708

Schlötterer C., R. Tobler, R. Kofler & V. Nolte (2014). Nature Reviews Genetics, 15: 749–763, https://doi.org/10.1038/nrg3803

Zhan S., G. Fang, M. Cai, Z. Kou, J. Xu, Y. Cao, L. Bai, *et al.* (2020). Cell Research 30 (1): 50-60. https://doi.org/10.1038/s41422-019-0252-6

611. The natural history of the black soldier fly, *Hermetia illucens*: insights from complete mitochondrial genome sequences

J. Guilliet[1,2*], G. Baudouin[2], N. Pollet[1] and J. Filée[1]

[1]EGCE, CNRS – Avenue de la terrasse, Bâtiment 13, 91198 Gif Sur Yvette, France; [2]Cycle Farms, 6 boulevard de entrepreneurs, 49250 Beaufort en Anjou, France; guilliet@egce.cnrs-gif.fr

Abstract

The black soldier fly (BSF) *Hermetia illucens* is a cosmopolitan fly heavily used by industrial companies to reduce biowaste and produce protein and fat for poultry and aquaculture feed. However, the natural history and the genetic diversity of the BSF are poorly known. In this study, we analysed 677 cytochrome c oxidase 1(CO1) sequences derived from samples found all over the five continents, leading us to discover 52 haplotypes, including ten major haplotypes. We refined the definition of these haplotypes by sequencing 59 mitochondrial genomes. We could derive an estimate of the separation events of the different haplotypes at more than two million years for the oldest branches. This worldwide cryptic genetic and genomic diversity is mirrored at a local scale in France, in which we found five major haplotypes sometimes in sympatry.

Introduction

The black soldier fly (BSF) *Hermetia illucens* (Linnaeus, 1758) is a Diptera of the Stratiomyidae family. The BSF is a cosmopolitan fly found in tropical, subtropical, and tempered regions, from the 40th parallel North and the 45th South (Mason, Rozkošný, and Hauser 2009; Rozkosný 1983). Studies based on naturalistic observations were first discovered in Malta in 1926 (Lindner 1936). Currently, the BSF can be found on the whole European territory, with few exceptions. It is reasonable to assume that the rise of international traffic and the industrial use of the BSF may have led to a steady flow of introductions in France and worldwide (Craig Sheppard *et al.* 1994; Leclercq 1969). Industrial companies heavily use the BSF to reduce biowastes (Gold *et al.* 2018; Hoc *et al.* 2020), to produce protein (Ewald *et al.* 2020; Giannetto *et al.* 2019; Hoc *et al.* 2020), and fat (Ewald *et al.* 2020; Hoc *et al.* 2020; Spranghers *et al.* 2017). These products are used to make animal food (Huis 2013) or biodiesel (Leong *et al.* 2016; Surendra *et al.* 2020). BSF industries face similar issues as other farms: knowing how to master their model's biological and genetic aspects to get the best out of it. To this end, it is crucial to know the genetic diversity of BSF existing internationally, nationally, and locally. Within the framework of the BSF industry, phylogeography can allow on the one side to account for genetic diversity at different scales (global, national, and local) and, on the other side, to have keys on the geographical origin of the species present in a territory or used by companies. In this study, we quantified this diversity at different scales: worldwide by studying the variation within the CO1 gene, and more locally at the level of France. We also analysed the role of livestock BSF farming in the diversity present on the French territory.

Materials & methods

We carried out a sampling of BSF worldwide with a focus on France. We also gathered all the CO1 sequences available in the databases (NCBI and BOLD [ncbi.nlm.nih.gov/ & boldsystems.org/]). All samples were stored in absolute ethanol at -20 °C before DNA extraction. All samples were extracted under the same conditions using a Macherey-Nagel AXG100 kit. Using specific primers, we amplified and sequenced the CO1 sequences. In addition, a total of 56 BSF were used for whole-genome sequencing using the Illumina method. Mitochondrial DNA sequences were reconstructed using Unicycler software and annotated with MITOS. All bioinformatics analyses were conducted using the tools available in Geneious software (R19 geneious.com/). For phylogenies, we used MEGAX (Kumar *et al.* 2018) and BEAST (Suchard *et al.* 2018).

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_611, © J. Guilliet *et al.* 2022

The visuals were made using PopArt (Leigh and Bryant 2015) for the haplotype network and Itol (itol.embl.de/) for the phylogeny representation.

Results

We reconstructed a haplotype network from the analysis of 677 sequences spanning 658 nucleotides of the CO1 mitochondrial gene. A total of 52 distinct haplotypes were delineated from the analysis of this CO1 fragment of one to 230 fly samples (Figure 1). We visualized the actual distribution of the different haplotypes found previously on a distribution map (Figure 2).

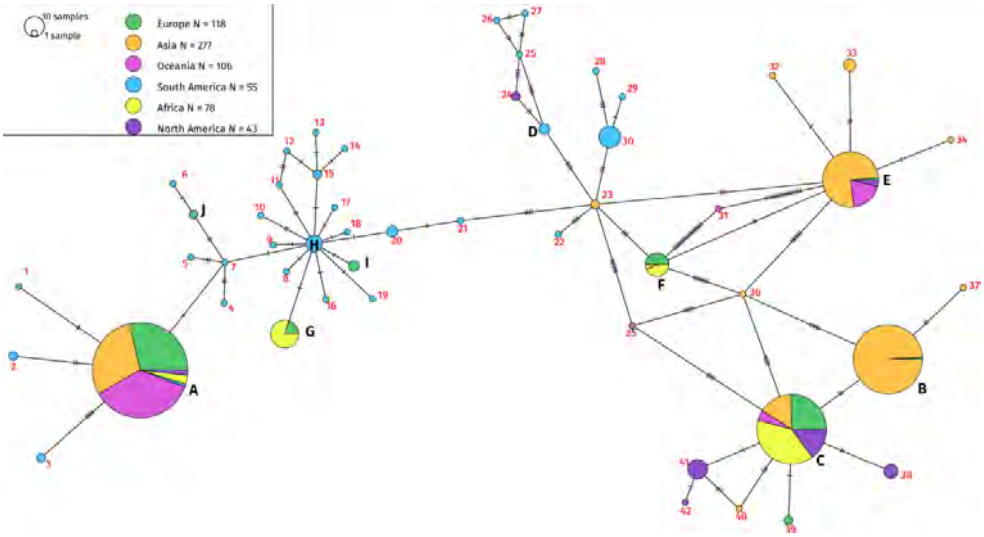


Figure 1. CO1 haplotype network of *Hermetia illucens*. Method use: minimum spanning network.

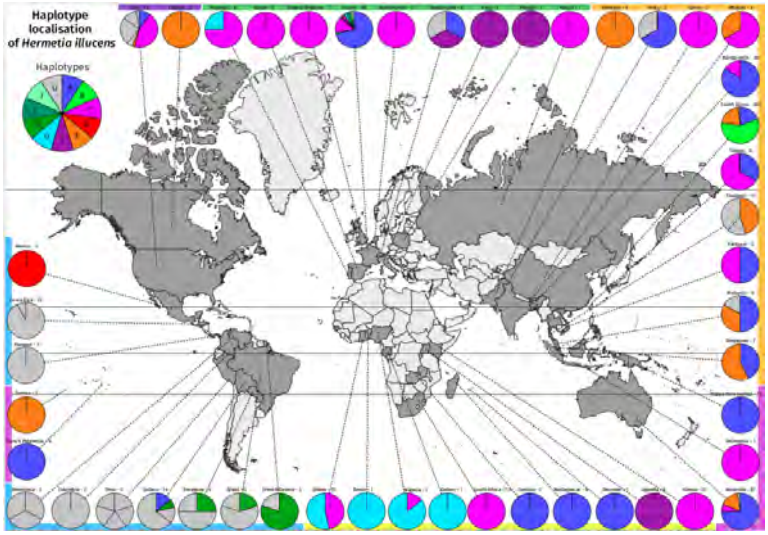


Figure 2. World CO1 haplotype distribution. Haplotype was define using the Figure 1 results.

The complete mitochondrial genome phylogeny demonstrates the very high genetic divergence among BSF populations found with the CO1 gene analysis. Two major divisions emerge within the tree: group C and others. Two subgroups emerge within the C group. The other large group contains the sequences of the A D I G and H haplotypes. The time tree (Figure 3) gives us information about the divergence time of the different mitochondrial genomes. We obtained identical Bayesian and ML trees. We noted a relatively long divergence between haplotype C and the others (2.2 million years). The others have diverged more recently (about 1.2 million years). Within the C haplotype, a divergence has been about 300,000 years, while in the other groups, the divergence times are much lower, about 20,000 years for the A haplotype.

Discussion

H. illucens is now considered a species of major interest for the food industry (Tomberlin and van Huis 2020). Thus, it became urgent to study the genetic diversity within the different populations at different scales. Analysing the different haplotypes in the same geographical area is a powerful tool to study the recent history of BSF. This study uncovered a high genetic variability of mitogenomes within the species *H. illucens*. We found that haplotype C is the most divergent. The dating of the separation events with the other haplotypes goes back more than 2 million years. We found a well-resolved separation within haplotype C (>95%); the other haplotypes are also well resolved (>95%). The analysis of complete mitogenomes supports the hypothesis of several distinct introductions in Europe and other continents. Even at a national

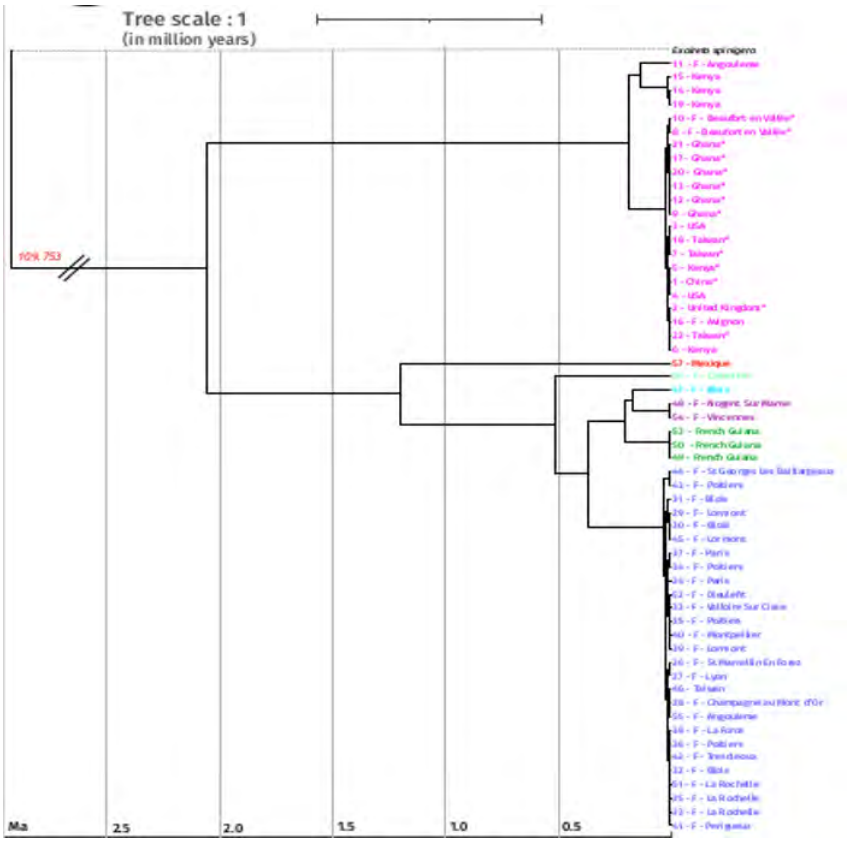


Figure 3 Phylogenetic time tree of *Hermetia illucens* based on the mitochondrial DNA. Rooted with *Exaireta spinigera*.

scale, the BSF genetic diversity is vast. By taking the example of France, we were able to show that the diversity observed at the global scale is also found at the national scale (five of the ten major haplotypes found worldwide). Moreover, we have haplotype mixtures at the local level, confirming the observations on microsatellite data (Kaya *et al.* 2021). For example, in the region of Paris, we have a mix of three haplotypes (A, F, and I). In the region of Montpellier, a mix of four haplotypes (A, B, C, and E). We also evidence that the mitochondrial genome is ideally suited to differentiate the genetic diversity and resolve phylogenetic relationships. It allows us to date the natural history of the BSF, bringing a sum of details that were impossible to distinguish based only on the CO1. Complete mitochondrial genome data allows us to make a robust and well-resolved time tree of the species. The divergence times between the different haplotypes are surprisingly long (more than two million years for the most divergent lineages), and the differences observed between the sequences prove that there is still unsuspected and hidden genomic diversity for the species. Comparing it to Drosophila, two million years correspond to many speciation events leading to different reproductively isolated species (Russo *et al.* 2013).

References

Craig Sheppard, D., G. Larry Newton, Sidney A. Thompson, and Stan Savage. 1994. 'A Value Added Manure Management System Using the Black Soldier Fly.' *Bioresource Technology* 50 (3): 275–79. https://doi.org/10.1016/0960-8524(94)90102-3.

Ewald, Nils, Aleksandar Vidakovic, Markus Langeland, Anders Kiessling, Sabine Sampels, and Cecilia Lalander. 2020. 'Fatty Acid Composition of Black Soldier Fly Larvae (Hermetia Illucens) - Possibilities and Limitations for Modification through Diet.' *Waste Management (New York, N.Y.)* 102 (February): 40–47. https://doi.org/10.1016/j.wasman.2019.10.014.

Giannetto, Alessia, Sabrina Oliva, Carlos Frederico Ceccon Lanes, Fabio de Araújo Pedron, Domenico Savastano, Cosimo Baviera, Vincenzo Parrino, *et al.* 2019. 'Hermetia Illucens (Diptera: Stratiomydae) Larvae and Prepupae: Biomass Production, Fatty Acid Profile and Expression of Key Genes Involved in Lipid Metabolism.' *Journal of Biotechnology*, October. https://doi.org/10.1016/j.jbiotec.2019.10.015.

Gold, Moritz, Jeffery K. Tomberlin, Stefan Diener, Christian Zurbrügg, and Alexander Mathys. 2018. 'Decomposition of Biowaste Macronutrients, Microbes, and Chemicals in Black Soldier Fly Larval Treatment: A Review.' *Waste Management (New York, N.Y.)* 82 (December): 302–18. https://doi.org/10.1016/j.wasman.2018.10.022.

Hoc, B., M. Genva, M.-L. Fauconnier, G. Lognay, F. Francis, and R. Caparros Megido. 2020. 'About Lipid Metabolism in Hermetia Illucens (L. 1758): On the Origin of Fatty Acids in Prepupae.' *Scientific Reports* 10 (1): 11916. https://doi.org/10.1038/s41598-020-68784-8.

Huis, Arnold van. 2013. *Edible Insects: Future Prospects for Food and Feed Security*. FAO Forestry Paper 171. Rome: Food and Agriculture Organization of the United Nations.

Kaya, Cengiz, Tomas N. Generalovic, Gunilla Ståhls, Martin Hauser, Ana C. Samayoa, Carlos G. Nunes-Silva, Heather Roxburgh, *et al.* 2021. 'Global Population Genetic Structure and Demographic Trajectories of the Black Soldier Fly, Hermetia Illucens.' *BMC Biology* 19 (1): 94. https://doi.org/10.1186/s12915-021-01029-w.

Kumar, Sudhir, Glen Stecher, Michael Li, Christina Knyaz, and Koichiro Tamura. 2018. 'MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms.' *Molecular Biology and Evolution* 35 (6): 1547–49. https://doi.org/10.1093/molbev/msy096.

Leclercq, M. 1969. 'Dispersion and transport of noxious insects: Hermetia illucens L. (Diptera Stratiomyidae).' *Bulletin des recherches agronomiques de Gembloux*. https://scholar.google.com/scholar_lookup?title=Dispersion+and+transport+of+noxious+insects%3A+Hermetia+illucens+L.%28Diptera+Stratiomyidae%29&author=Leclercq%2C+M.&publication_year=1969.

Leigh, Jessica W., and David Bryant. 2015. 'Popart: Full-Feature Software for Haplotype Network Construction.' *Methods in Ecology and Evolution* 6 (9): 1110–16. https://doi.org/10.1111/2041-210X.12410.

Leong, Siew Yoong, Shamsul Rahman Mohamed Kutty, Amirhossein Malakahmad, and Chew Khun Tan. 2016. 'Feasibility Study of Biodiesel Production Using Lipids of Hermetia Illucens Larva Fed with Organic Waste.' *Waste Management (New York, N.Y.)* 47 (Pt A): 84–90. https://doi.org/10.1016/j.wasman.2015.03.030.

Lindner, Erwin. 1936. *Die amerikanische Hermetia illucens L. im Mittelmeergebiet: (Stratiomyiidae, Dipt.)*. Akadem. Verlagsges.

Mason, Franco, Rudolf Rozkošný, and Martin Hauser. 2009. 'A Review of the Soldier Flies (Diptera: Stratiomyidae) of Sardinia.' *Zootaxa* 2318 (1): 507–30. https://doi.org/10.11646/zootaxa.2318.1.20.

Rozkosný, R. 1983. *A Biosystematic Study of the European Stratiomyidae (Diptera): Volume 2 - Clitellariinae, Hermediinae, Pachygasterinae and Bibliography*. Series Entomologica. Springer Netherlands. https://www.springer.com/gp/book/9789061931355.

Russo, Claudia A. M., Beatriz Mello, Annelise Frazão, and Carolina M. Voloch. 2013. 'Phylogenetic Analysis and a Time Tree for a Large Drosophilid Data Set (Diptera: Drosophilidae).' *Zoological Journal of the Linnean Society* 169 (4): 765–75. https://doi.org/10.1111/zoj.12062.

Suchard, Marc A, Philippe Lemey, Guy Baele, Daniel L Ayres, Alexei J Drummond, and Andrew Rambaut. 2018. 'Bayesian Phylogenetic and Phylodynamic Data Integration Using BEAST 1.10.' *Virus Evolution* 4 (1). https://doi.org/10.1093/ve/vey016.

Spranghers, Thomas, Matteo Ottoboni, Cindy Klootwijk, Anneke Ovyn, Stefaan Deboosere, Bruno De Meulenaer, Joris Michiels, Mia Eeckhout, Patrick De Clercq, and Stefaan De Smet. 2017. 'Nutritional Composition of Black Soldier Fly (Hermetia Illucens) Prepupae Reared on Different Organic Waste Substrates.' *Journal of the Science of Food and Agriculture* 97 (8): 2594–2600. https://doi.org/10.1002/jsfa.8081.

Surendra, K. C., Jeffery K. Tomberlin, Arnold van Huis, Jonathan A. Cammack, Lars-Henrik L. Heckmann, and Samir Kumar Khanal. 2020. 'Rethinking Organic Wastes Bioconversion: Evaluating the Potential of the Black Soldier Fly (Hermetia Illucens (L.)) (Diptera: Stratiomyidae) (BSF).' *Waste Management (New York, N.Y.)* 117 (November): 58–80. https://doi.org/10.1016/j.wasman.2020.07.050.

Tomberlin, J.k., and A. van Huis. 2020. 'Black Soldier Fly from Pest to 'Crown Jewel' of the Insects as Feed Industry: An Historical Perspective.' *Journal of Insects as Food and Feed* 6 (1): 1–4. https://doi.org/10.3920/JIFF2020.0003.

612. Impact of genetic selection in insect populations using different selection designs – a simulation study

R.S.C. Rikkers*, J.W.M. Bastiaansen, A.C. Bouwman and E.D. Ellen

Animal Breeding and Genomics, Wageningen University & Research, P.O. Box 338, 6700 AH Wageningen, the Netherlands; roxann1.rikkers@wur.nl

Abstract

Insects are highly nutritious and rich in proteins and could help fulfil the worldwide increasing demand for proteins. Selection can provide lasting genetic improvements. Therefore, the aim of this study was to investigate the effect of genetic parameters in four different selection designs using a simulation study. Simulated were a mixed, half sib and full sib selection design and a combination of phenotypic selection and/or breeding values as selection method. The results show that in ten generations an improvement in body mass and development time is possible in all four selection designs and only minor differences between these designs were observed. Using estimated genetic parameters from literature, an improvement of 146-150% in biomass yield per year was observed. Thus, selection can provide lasting genetic improvements in insect populations, however inbreeding should be monitored and estimation of genetic parameters is necessary to prevent unfavourable correlated responses.

Introduction

Insects are considered to be highly nutritious and rich in proteins (Rumpold and Schlüter, 2013a) and improving the production of insects could help fulfil the worldwide increasing demand for food and proteins in 2050 (FAO, 2012). Rumpold and Schlüter (2013b) concluded that in order to maximize protein production the appropriate species should be selected. These species must be suitable for mass production and should have characteristics like good health, efficient and high quality proteins. Genetic selection could improve these important characteristics within a species, however setting up an insect breeding program has some challenges. Insect populations are often kept in large groups and individual identification is not possible. Furthermore, insects often have multiple mates and the number of mates has an effect on the fecundity of females (Pai and Yan, 2005). Considering these factors the application of genetic selection in insect populations is challenging. To show the potential and impact of genetic selection in insect populations we have conducted a simulation study that mimic flour beetle (*Tribolium castaneum*) populations. Simulated traits were pupal body mass (BM), development time (DT) till first day of pupation and biomass per year (BY). The effect of genetic parameters (heritability and genetic correlation) was investigated using single trait selection on BM in four different selection designs. To show the impact of genetic parameters on response to selection.

Materials & methods

To simulate BM and DT in flour beetle populations we used estimated genetic parameters from Ellen *et al.* (2016; Table 1). A simulation program (RStudio; R Core Team, 2019) was used to simulate four selection designs: a mixed parents design (MixedSel), a Half Sib (HS) family design (HSfam) and two versions of a Full Sib (FS) family design (FSfam and FSfamBV). Of the four selection designs, MixedSel would be the least expensive and FSfamBV would be the most expensive and labour intensive design for insect producers, since animals have to be kept individually to be able to identify them.

Basic simulation program. Phenotypes for BM and DT of the founder population were simulated as $P = \mu + A + E$, where A and E were randomly drawn from a multivariate normal distribution, where the covariance matrices of A and E were based on estimates by Ellen *et al.* (2016). Phenotypes for each

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_612, © R.S.C. Rikkers *et al.* 2022

Table 1. Parameters for body mass (BM) in mg and development time (DT) in days.

Trait	Mean	h^2	σ_a^2	σ_p^2
BM	2.33	0.38	0.027	0.071
DT	27.05	0.41	2.526	6.114

consecutive generation were simulated similarly, however $A = 0.5\ A_{sire} + 0.5\ A_{dam} + MS$ and MS (mendelian sampling) is drawn from a covariance matrix adjusted for inbreeding. The simulation program was applied for a given number of batches, generations and replicates. To reduce the amount of inbreeding a batch rotation scheme was simulated where males rotate over batches.

In the MixedSel design males and females where kept together and mated at random. Per batch, selection candidates to produce the next generation were phenotypically selected on highest BM. In the HSfam design, individual males were kept with multiple females, creating HS families where sire-pedigree information was available. For each sire the BM estimated breeding values (EBV) were calculated using a univariate model in ASReml (Gilmour, *et al.*, 2015) and HS families were ranked on sire EBV. Offspring from 50% of the best performing sires were ranked on BM phenotype and selection candidates were phenotypically selected from this ranking. In the FSfam design individual males were kept with one female, creating FS families. The selection families were ranked on the average BM EBV of the sire and dam. Offspring from 50% of the best performing families were ranked on BM phenotype and selection candidates were phenotypically selected from this ranking. The FSfamBV design was identical to the FSfam design, however before hatching the eggs where kept individually making it possible to identify each insect individually and know their associated pedigree. BM EBVs for each individual were estimated using a univariate animal model in ASReml. Individuals with the highest BM EBV were selected to produce the next generation. An overview and detailed information of the four selection designs is given in Table 2.

Scenarios. We have tested the effect of: (1) genetic correlation; and (2) heritability (h^2) on response to selection in the four selection designs. In scenario 1 three genetic correlations between BM and DT were simulated, 0.15 as estimated by Ellen *et al.* (2016; Scenario 1A), a strongly positive (0.70; Scenario 1B) and a strongly negative (-0.70; Scenario 1C) genetic correlation. A high positive or high negative genetic correlation is unlikely in real life, however to study the effect of genetic parameters on the four selection designs a correlation estimated in literature and two extreme scenarios were chosen. In scenario 2 we replaced the estimated heritability for BM ($h^2 = 0.38$) with a h^2 of 0.10 to investigate the impact of a lower h^2 when performing genetic selection. Results were presented as the biomass yield per year (BY) which was calculated as BY = (BM / DT) × 365. The baseline BY based on Ellen *et al.* (2016; Table 1) is 31.44 mg BY ((2.33 / 27.05) × 365).

Table 2. Detailed information per generation on batch level.

	MixedSel	HSfam	FSfam	FSfamBV
Mating	Random	1:2	1:1	1:1
# males	20	10	20	20
# females	20	20	20	20
Selected prop.	0.100	0.075	0.100	0.100
Selection	↑ BM	↑ BM within best families	↑ BM within best families	↑ EBV BM
Pedigree	-	6 generations	6 generations	6 generations

Results

The highest BM improvement after ten generations was found in the FSfamBV design 3.72±0.04 mg, although FSfam and HSfam showed similar results, respectively 3.67±0.05 and 3.70±0.04 (Table 3). MixedSel (phenotypic selection) resulted in a total of 45.29 mg BY, while adding more pedigree information (HSfam, FSfam and FSfamBV) resulted in ~46.5 mg BY. Inbreeding was lowest in the MixedBase design and highest in the FSfamBV design. Using more pedigree information in the breeding program increased inbreeding (Table 3), even though a rotational scheme was used to reduce inbreeding.

Scenario 1A (genetic parameters estimated by Ellen *et al.* (2016)) yielded a biomass between 45.29 (MixedSel) and 46.72 (FSfam) mg per year (Table 4). Scenario 1B (high positive genetic correlation) showed a lower BY after ten generations, whereas scenario 1C (high negative genetic correlation) showed a higher BY after ten generations in the four selection designs compared to scenario 1A and 1B. Furthermore, reducing the BM h^2 (scenario 2) resulted in lower improvement of BY in all selection designs compared to scenario 1A.

Discussion

We compared four different selection designs in simulated insect populations and the effect of genetic parameters on these four methods. As expected, the genetic parameters have an impact on the selection response (how fast you can improve your population). The favourable direction of breeders would likely be a higher BM and a shorter DT, as this gives the fastest increase in BY. A strong positive genetic correlation (scenario 1B) resulted in a lower BY, because DT becomes longer reducing the number of pupa grown per year. Furthermore a lower heritability (scenario 2) resulted in less improvement of BY after ten generations, compared to a higher heritability (scenario 1A). Ellen *et al.* (2016) estimated a genetic correlation between BM and DT of 0.15±0.10 and Mwenya *et al.* (1986) estimated genetic correlations of 0.15±0.14

Table 3. Genetic level after ten generations with genetic parameters from literature (scenario 1A).

	Mean ± SD BM[1]	Mean ± SD DT[2]	BY[3]		F[5]
	Mg	Days	Mg	%[4]	%
MixedSel	3.59±0.04	28.93±0.29	45.29	146	0.047±0.007
HSfam	3.70±0.04	29.13±0.35	46.29	149	0.048±0.005
FSfam	3.67±0.04	28.71±0.72	46.72	150	0.059±0.004
FSfamBV	3.72±0.04	29.10±0.62	46.62	150	0.084±0.008

[1] Body mass (h^2=0.38).
[2] Development time.
[3] Biomass yield.
[4] Percentage of BY improvement relative to generation 0.
[5] Inbreeding level in generation 9.

Table 4. BY in mg after ten generations for each scenario and selection design.[1]

Scenario		MixedSel	HSfam	FSfam	FSfamBV
1A	r_a 0.70	45.29 (146)	46.29 (149)	46.72 (150)	46.62 (150)
1B	r_a 0.15	36.60 (118)	37.23 (120)	36.76 (118)	37.27 (120)
1C	r_a -0.70	71.66 (230)	76.42 (246)	75.63 (243)	79.10 (254)
2	h^2 0.15	34.81 (112)	35.01 (113)	36.07 (116)	35.97 (116)

[1] Percentage of improvement after 10 generations (relative to generation 0) in parentheses.

and 0.45±0.10 depending on the mating design. Our results show that selection on more than one trait (multi trait) and knowing the correct genetic parameters is important to prevent unfavourable correlated responses. In our simulation study inbreeding was relatively low. However, including pedigree information increased the inbreeding rate. So far, inbreeding has not been considered in insect production. However, Rhode *et al.* (2020) showed a black soldier fly population collapsing after five generations due to a lack of genetic diversity. Comparable to conventional livestock breeding programs, monitoring of inbreeding in insect populations is important.

Insect breeders face many challenges, conventional livestock breeders do not face, such as the difficulty to identify an individual insect. Remating, where females lay eggs from multiple males, affects identification and the number of mates has an effect on the fecundity of a female (Pai and Yan, 2003). In all scenarios the differences between the four selection designs were relatively small. Considering the costs and labour intensiveness of breeding programs where insects are kept individually (FSfamBV) and the responses observed in phenotypic selection (MixedSel) being similar, using pedigree information and/or breeding value estimation is not beneficial in the case of single trait selection.

To conclude, genetic selection could improve important characteristics in insect populations. Using single trait selection, a MixedSel or HS design, without individual identification, is favourable. However, breeding programs contain more than one trait and therefore, future research should focus on multi-trait selection. For insect populations, inbreeding should be monitored and estimation of genetic parameters is necessary to prevent unfavourable correlated responses.

Funding

This study was financially supported by the Dutch Ministry of Agriculture, Nature and Food Quality (KB-39-007-005).

References

Ellen E.D., Peeters K., Verhoeven M., Gols R., Harvey J.A., *et al.* (2016) Evolution 70-1: 207-2017. https://doi.org/10.1111/evo.12835

Food and Agriculture Organization of the United Nations (FAO), World Agriculture Towards 2030/2050 - The 2012 Revision. Available at: https://www.fao.org/global-perspectives-studies/resources/detail/en/c/411108/.

Gilmour, A., Gogel, B., Cullis, B., Welham, S., & Thompson, R. (2015). ASReml User Guide Release 4.1 Structural Specification. Hempstead, United Kingdom: VSN International Ltd.

Mwenya W.N.M, Gianola D. and Grossman M. (1986) Theor Appl Genet 72:388-394. https://doi.org/10.1007/BF00288578

Pai A. and Yan G. (2003) Can. J. Zool 81: 888-896. https://doi.org/10.1139/Z03-070

Rhode C., Badenhorst K.L., Hull M.P. Greenwood A.E. Bester-van der Merwe A.E. *et al.* (2020) Animal Genetics Volume 51 Issue 5 p. 752-762 https://doi.org/10.1111/age.12961

R Core Team (2019) R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://www.R-project.org/

Rumpold B.A. and Schlüter O.K. (2013a) Mol Nutr Food Res Volume 57, Issue 5 p. 802-823. https://doi.org/10.1002/mnfr.201200735

Rumpold B.A. and Schlüter O.K. (2013b) Innov Food Sci Emerg Technol Volume 17, p. 1-11. https://doi.org/10.1016/j.ifset.2012.11.005rstudio

613. Genetic parameters of black soldier flies estimated in full sib design

A.C. Bouwman[1*], J.E. Nugroho[2], D. Wongso[1], J. van Schelt[2], B.A. Pannebakker[3], B.J. Zwaan[3] and E.D. Ellen[1]

[1]Animal Breeding and Genomics, Wageningen University & Research, P.O. Box 338, 6700 AH Wageningen, the Netherlands; [2]Bestico B.V., Veilingweg 6, 2651 BE Berkel & Rodenrijs, the Netherlands; [3]Laboratory of Genetics, Wageningen University & Research, Droevendaalsesteeg 1, 6708 PB Wageningen, the Netherlands; aniek.bouwman@wur.nl

Abstract

The aim of this study was to design a scheme for black soldier flies (BSF, *Hermetia illuscens*) to collect data and estimate genetic parameters for the recorded traits. A full sib family design was implemented in a practically feasible way to record multiple traits on the same individuals across a number of families. Genetic parameters were estimated using a linear mixed model. Heritability estimates for body weight at prepupa, body weight at pupa, development time to prepupa was 0.25, 0.45 and 0.38, respectively. The common environmental effect was large at the larval stage and reduced after housing the BSF individually. In general, body weight showed a positive phenotypic correlation between the S2, prepupa and pupa stages, and development time was positively correlated with body weight at (pre)pupa stage. In conclusion, a full sib family design is a practically feasible way to collect data for estimation of genetic parameters in the BSF.

Introduction

Insects are an alternative protein source that can add to the challenge in the food transition we are facing today. Insects can be used as feed for pigs, chickens and fish, or directly for human consumption. In both cases, large quantities of the tiny bugs are needed to obtain sufficient product. Phenotypic selection of larger individuals is effective in insect systems to increase mass (Morales-Ramos *et al.*, 2018). Phenotypic selection for morphological traits is in general very effective, as body mass has a moderate to high heritability in most species (e.g. Visscher *et al.*, 2008). However, for most insect populations produced for edible proteins the genetic parameters are unknown, and neither are their unfavourable genetic correlations with other important traits. Insect species have the major limitation that they are difficult to track individually over time, which makes it difficult to collect suitable datasets similar to what is common in larger livestock species. Black Soldier Flies (BSF, *Hermetia illuscens*) can upcycle low-quality organic waste-streams into high-quality proteins and are currently the most used species for livestock feed. Therefore, the aim of this study was to design a scheme for BSF to collect data and estimate genetic parameters for the recorded traits.

Materials & methods

Data collection design. A full sib (FS) family design was implemented in a practically feasible way to record multiple traits on the same individuals across a number of FS families (Figure 1). Data was collected in 5 batches at a commercial BSF facility. At random, 5 to 10 fresh egg clutches were collected in each batch from the general population, assuming each clutch is one FS family. Each clutch was split in half and each half was placed in a container with standard feed. Hence each FS family was housed in two groups (i.e. two common environments) under similar conditions. Please note that the common environment is nested within family, as mixing of families was impossible due to lack of identification. After 10 days, 30 larvae were individually weighted from each family-environment combination (Body Weight Stage 1, BW_S1), and housed individually in tubes with standard feed. They were individually weighted again after 7 days (Body Weight Stage 2, BW_S2). At that moment, half of the 30 individuals per family-environment combination were freeze dried, and weighted afterwards (Dry Weight Stage 2, DW_S2), they were stored for

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_613, © A.C. Bouwman *et al.* 2022

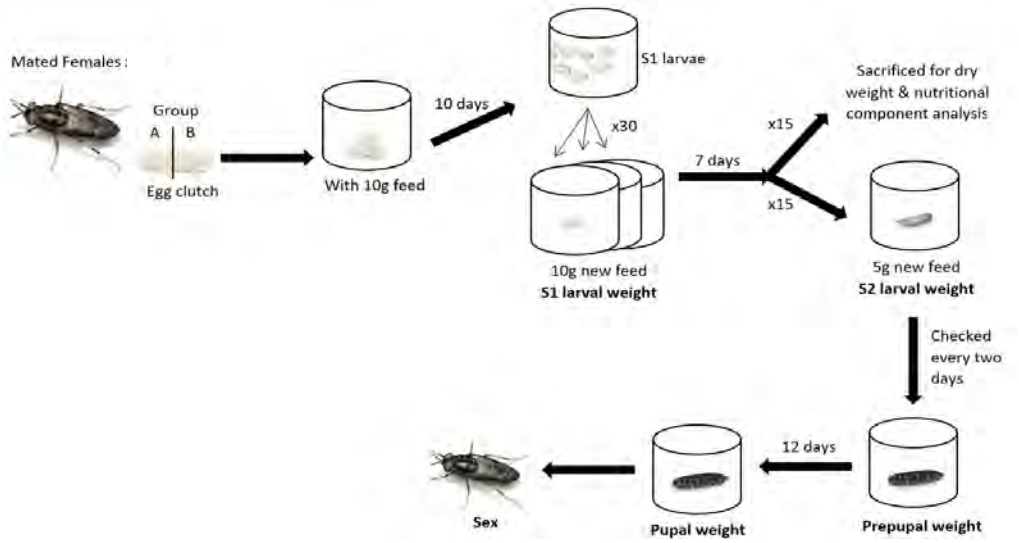


Figure 1. Schematic overview of the data collection.

other purposes. The remaining 15 individuals per family-environment combination were kept individually for further measurements. Every two days they were checked to see if they had reached the prepupa stage (black but still mobile), and if so they were weighted (Body Weight prepupa, BW_prepupa), and the day was recorded. The number of days from start of the batch till weight measurement of prepupa was included as a trait to represent development time to prepupa (Development Time, DevTime). The final pupal weight (black and immobile) was measured 12 days after reaching prepupa stage (Body Weight pupa, BW_pupa). Furthermore, it was recorded if they emerged to adults, and what their sex was based on adult morphology.

Estimation of genetic parameters. Genetic parameters were estimated using the following linear mixed model in ASReml (Gilmour *et al.*, 2015):

$$y = Xb + Zfam + Wc + e \quad (1)$$

Where *y* is the vector of observations, *b* vector of fixed effects, including the mean, sex as a fixed regression, with males being 0, females being 1, and individuals with unknow sex (the ones freeze dried and non-emerging) were included as 0.5 assuming a 50/50 sex ratio. Batch was included as a random effect. BW_S1 was not measured in batch 1. For BW_prepupa a fixed regression was included on number of days till prepupa weight measurement (i.e. DevTime). *fam* is the vector of random family effects, *c* is the vector of common environment nested within family, and *e* is the vector of random residual effects. The *X*, *Z* and *W* are the design matrices relating observations to the respective levels of fixed and random effects. Random effects were assumed to be normally distributed: $fam \sim N(0, I\sigma^2_{fam})$; $c \sim N(0, I\sigma^2_c)$; $e \sim N(0, I\sigma^2_e)$ with I as identity matrix.

The additive genetic variance (σ^2_a) is estimated as two times the family variance σ^2_{fam}. The heritability was estimated as:

$$h^2 = (2 * \sigma^2_{fam})/(\sigma^2_{fam} + \sigma^2_c + \sigma^2_e) \quad (2)$$

The contribution of the common environment to the phenotypic variation was estimated by:

$$c^2 = (\sigma_c^2)/(\sigma_{fam}^2 + \sigma_c^2 + \sigma_e^2) \quad (3)$$

Phenotypic correlations were estimated in a bivariate fashion with model 1 excluding the family and common environmental components. The dataset was too small to estimate genetic correlations.

Results

In total 2,731 larvae were recorded, of which 1,285 were followed until emergence to adulthood. A total of 6 traits were measured, their mean is given in Table 1.

Heritability estimates are given in Table 2, the three weight traits at larval stage had large standard errors, hence we were not able to declare them significantly different from zero. From Table 2 it can be seen that the common environmental effect was largest on the first body weight measure (BW_S1) and reduced substantially when individuals were housed individually and as time between group housing and trait measure increased.

The strongest phenotypic correlation (0.81) was found between prepupa and pupa body weight (Table 3), which were both unfavourably correlated with DevTime.

Table 1. Descriptive statistics of measured traits in commercial BSF.

Trait	N	Fam[1]	Mean (sd)	Min	Max
BW_S1 (mg)	2,318	39	6.62 (4.32)	0.25	39.01
BW_S2 (g)	2,731	44	0.23 (0.05)	0.01	0.35
DW_S2 (g)	1,431	44	0.07 (0.02)	0	0.12
BW_prepupa (g)	1,285	44	0.22 (0.04)	0.08	0.34
BW_pupa (g)	1,288	44	0.17 (0.03)	0.02	0.25
DevTime (d)	1,285	44	22.51 (1.25)	21	27

[1] Number of families.

Table 2. Heritability (h^2) and common environmental effect (c^2) of the traits, with their standard error in brackets.

Trait	h2	c^2
BW_S1	0.35 (0.72)	0.71 (0.13)
BW_S2	0.32 (0.25)	0.43 (0.08)
DW_S2	0.24 (0.27)	0.44 (0.09)
BW_prepupa	0.25 (0.11)	0.17 (0.04)
BW_pupa	0.45 (0.12)	0.14 (0.04)
DevTime	0.38 (0.16)	0.25 (0.06)

Table 3. Phenotypic correlation of the traits, with their standard error in brackets.

Trait	BW_S2	BW_prepupa	BW_pupa	DevTime
BW_S1	0.33 (0.02)	-0.007 (0.03)	-0.01 (0.03)	-0.38 (0.03)
BW_S2		0.39 (0.02)	0.49 (0.02)	-0.23 (0.03)
BW_prepupa			0.81 (0.01)	0.62 (0.02)
BW_pupa				0.41 (0.02)

Discussion

Insects are very small and hard to identify individually over time, which is the main limitation for the estimation of genetic parameters, and even more so for genetic selection programs. The FS family design presented here is practically feasible, as it allows to collect data on individuals from different families not knowing the parents. Ideally multiple families were housed in the same group to disentangle the common environmental effect further. A maternal effect may be present in insects as the conditions under which the mother developed and laid her eggs potentially play a role (Mousseau and Dingle 1991), but the applied design did not allow to estimate maternal effects.

For a general full sib design we estimated that with 15 individuals from 45 families we would be able to estimate heritabilities of 0.3 with a standard error around 0.08 (following Falconer and Mackay, 1996). The obtained standard errors of our experiment were higher, as there are always factors that cannot be controlled in real life. However, reaching standard errors of 0.11-0.16 for the traits measured later in the lifecycle was satisfying. Collecting data on more families will improve the heritability estimates further.

The standard errors of the heritability were higher for the traits at larval stage. At the larval stages the common environment played an important role in the phenotypic variation observed. The common environmental factor is only a short common environment from egg clutch collection till 10 days thereafter, closest to the larval stage, which explains the large common environmental effect at the larval stage. The common environmental impact reduced as the individuals were housed individually for longer time. However, at the later life stages the differences in individual environment may be confounded with individual genetic potential, even though it was averaged within family. Taken this together with the fact that the family variance also includes a potential maternal effect, the heritability estimates are likely somewhat overestimated.

The positive phenotypic correlation between body weight at S2, prepupa and pupa stages can be exploited to select for body weight at an earlier developmental stage. The positive correlation between body weight at (pre)pupa and development time is unfavourable, as this would increase the generation interval. Considering profitable protein production, body weight and development time should be weighted in a selection index to obtain an optimum in the total mass produced per year for instance.

In conclusion, a FS family design is a practically feasible way to collect data for estimation of genetic parameters. The number of families measured is an important factor in decreasing the standard error of the estimates, in practise more families are needed than estimated theoretically, due to uncontrollable factors that play a role in practical situations.

Funding

This study was financially supported by the Dutch Ministry of Agriculture, Nature and Food Quality (KB-39-007-005).

References

Falconer D.S., and Mackay T.F.C. (1996) Introduction to quantitative genetics. Longmans Green, Harlow, Essex, UK.

Gilmour, A.R., Gogel, B.J., Cullis, B.R., Welham, S.J. and Thompson, R. (2015). ASReml User Guide Release 4.1 Functional Specification, VSN International Ltd, Hemel Hempstead, UK

Morales-Ramos J.A., Kelstrup H.C., Rojas M.G., and Emery V. (2019) J Insect Sci, 19(2): 4. https://doi.org/10.1093/jisesa/iey110

Mousseau T.A., and Dingle H. (1991). Maternal effects in insects: Examples, constrains, and geographic variation. In: Dudley E.C. (eds), The unity of evolutionary biology. Dioscorides Press, Portland, OR, USA. pp. 745-761.

Visscher P., Hill W., and Wray N. (2008) Nat Rev Genet 9, 255–266. https://doi.org/10.1038/nrg2322

614. Towards selective breeding in insects – estimating genetic parameters with individual-level phenotypes and pedigree

L.S. Hansen[1*], S.F. Laursen[2], S. Bahrndorff[2], M. Kargo[1], T.N. Kristensen[2], G. Sahana[1], J.G. Sørensen[3] and H.M. Nielsen[1]

[1]*Center for Quantitative Genetics and Genomics, Aarhus University, Blichers Allé 20, 8830 Tjele, Denmark.* [2]*Department of Chemistry and Bioscience, Aalborg University, Fredrik Bajers Vej 7H, 9220 Aalborg, Denmark.* [3]*Department of Biology, Aarhus University, Ny Munkegade 116, 8000 Aarhus C, Denmark;* lsh@qgg.au.dk

Abstract

The rapid life cycle of most insects allows for the direct comparison of simulated and realised genetic gain from insect breeding schemes, enabling real-time optimisation of the breeding scheme design. However, rearing and maintenance of insects gives rise to major challenges due to specific dietary and environmental requirements, complex life cycles and reproductive patterns and this adds complexity to all aspects of the breeding plan implementation. We have developed an experimental design for large-scale individual phenotyping and tracing pedigree records and applied this to estimate genetic parameters for larval size and egg-to-adult development time using the housefly (*Musca domestica*) as a model species. These parameters contribute valuable information on the potentials for genetic improvement in this species and the methods and results presented can be used more generally in the process of insect breeding scheme implementation and optimisation.

Introduction

Using insects as food and feed has received renewed attention in the last decade, partly due to the ability of many insect species to utilise low-value waste products and convert this input into high-value protein for a more sustainable food and feed production (Van Huis and Oonincx, 2017, for a review). This has initiated an interest in the optimisation of insect farming through the manipulation of environmental factors (Tomberlin and Cammack, 2017). Selective breeding has historically been an important tool in the improvement of crop and livestock production but has so far received little attention within insect production (Eriksson and Picard, 2021). Many insect species are short-lived and highly fecund, theoretically enabling rapid and transgenerational optimisation of production efficiency through selective breeding. Variance components and heritabilities of economically important traits are key indicators of the potential for successful genetic improvement (Hill, 2010). Insects pose a challenge when attempting to estimate such parameters partly due to their rearing requirements, fragility, mode of reproduction and, occasionally, metamorphic life cycle. These fundamental challenges call for novel methods and protocols which can be applied in insect breeding schemes. In this study, we construct a two-generation pedigree by implementing a full-sib/half-sib mating design using the housefly (*Musca domestica* L.) as a model organism for holometabolous dipterans, an insect order highly suitable for converting organic waste into valuable protein products (Pastor *et al.*, 2015). We then estimate genetic parameters for two production-relevant traits, larval size (LS) and egg-to-adult development time (DT), using a large-scale laboratory experimental setup, designed to fit the unique biology and life cycle of holometabolous insects.

Materials & methods

Population. The housefly population used in the experiment was an outbred population established in June 2021 with flies collected from six cattle farms distributed across Denmark. The population was since maintained in the lab at a census size of ~15,000 individuals divided into three replicate populations of ~5,000 individuals, maintained in fly cages (30×30×30 cm) with ~1000 flies in each cage. In every generation, offspring from the five cages of a replicate population were intermixed at the pupal stage and

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_614, © L.S. Hansen *et al.* 2022

randomly distributed into new cages. Flies were reared under standard laboratory conditions at 23 °C, 40-60% relative humidity with a photoperiod of 12:12 h (light:dark) throughout their lifecycle. Larvae were reared on a standard laboratory larval medium composed of wheat bran, alfalfa meal, dry yeast, malt and tap water and adults had access to water and a petri dish filled with equal parts of granulated sugar, icing sugar and powdered milk.

Family generation. A nested paternal half-sib/full-sib design was used when generating the families in this experiment (Figure 1). The parent-generation was established by anesthetising and sorting newly emerged flies every 24 hours into sex-, age- and replicate-specific virgin cages. Virgin flies had access to granulated sugar, icing sugar, powdered milk and water.

At the age of 10 d ± 12 h, one male and five female flies were sorted into glass jars (H: 11.5 cm, Ø: 6 cm) for mating. Each jar was sealed with a foam stopper fitted with a micro centrifuge tube with 10% sugar solution. From the three replicate populations, 30, 30 and 19 mating jars were established over several days depending on when the flies reached the age of 10 days. A large number of virgin flies from replicate 3 were lost due to a technical error. Therefore, a second round of matings from replicates 1 and 2 were established with 14-day old virgin females to compensate for the reduction in family number from replicate 3. A total of 94 mating jars were established when combining all replicate populations. Flies were allowed to mate for 30 hours, where after all flies in each jar were anesthetised and females were sorted into oviposition vials (H: 9 cm, Ø: 2.5 cm) with 3 g fresh medium and kept in climate chambers at 23 °C, 60-70% relative humidity under a 12:12 h (light:dark) photoperiod. Oviposition vials were subsequently checked for eggs every 12 hours until the female had either oviposited >20 eggs in one oviposition event or died. If <20 eggs were found, females were transferred to a new oviposition vial with fresh medium. If >20 eggs were found, these were distributed into new vials with 10 g of fresh medium. The first 20 eggs were designated for the measurement of LS, and the next 20 eggs were designated for the measurement of DT, ensuring that the two traits were recorded on full siblings. Eggs for the two traits were always collected from the same oviposition event and if a female only oviposited enough eggs for one group, then no eggs were collected to measure DT. The offspring were subsequently reared under the same conditions as the females. A total of 256 families were established to measure LS (94, 107 and 55 for replicate 1, 2 and 3, respectively), and 244 families were established to measure DT (87, 103 and 54 for replicate 1, 2 and 3, respectively).

Phenotyping. LS was measured at day seven of the offspring life cycle using an automated size estimation procedure as described in Laursen *et al.* (2021). Full siblings from a vial were distributed individually into a numbered well-plate and mean larval surface area for all larvae in each family was acquired using a 60 s acquisition time. The numbering of the plate allowed individual ID-assignment. DT was estimated

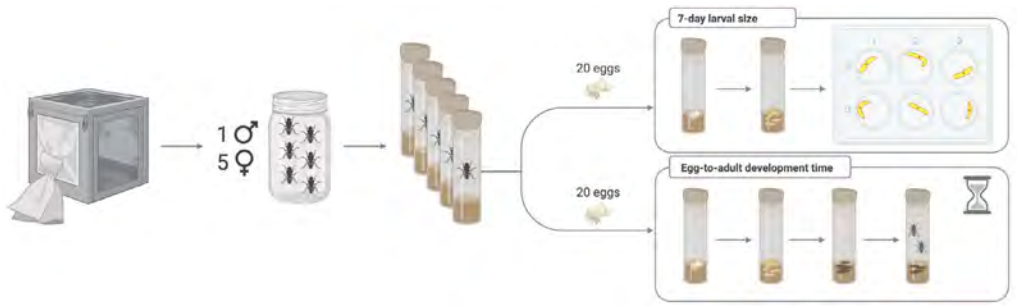


Figure 1. Schematic illustration of the experimental setup (created with BioRender.com).

every 12 hours by checking vials for newly emerged adults, which were removed from the vials and sexed. Individual IDs were assigned, and development time calculated for each emerged adult (±12 h). All vials were continuously checked until four days had passed where no flies emerged.

Statistical analysis. Variance components for LS was estimated with a univariate linear Gaussian sire-model using the full-sib family mean:

$$y = X\beta + Zs + e \quad (1)$$

and DT was analysed using a univariate linear Gaussian animal-model:

$$y = X\beta + Z_1 a + Z_2 vial + e \quad (2)$$

where y was the vector of observations of either mean LS or DT; β was a vector of fixed effects including replicate, ID of the person sorting the eggs (four groups), sex (only for DT) and dam age (2 groups, only for LS); s was the vector of random sire genetic effects, $N(0, \boldsymbol{I}\sigma_s^2)$; a was the vector of random animal additive genetic effects, $N(0, \boldsymbol{A}\sigma_a^2)$; vial was a vector of random rearing vial effects, $N(0, \boldsymbol{I}\sigma_{vial}^2)$; e was the vector of random residuals, $N(0, \boldsymbol{I}\sigma_e^2)$; I was the identity matrix, A was the additive genetic relationship matrix, and X, Z, Z_1 and Z_2 were the corresponding incidence matrices. Narrow-sense heritabilities were estimated as $h^2 = \frac{4\sigma_s^2}{\sigma_s^2+\sigma_e^2}$ when using the sire model, and as $h^2 = \frac{\sigma_a^2}{\sigma_a^2+\sigma_{vial}^2+\sigma_e^2}$ when using the animal model. Variance components were estimated using REML with the average information algorithm (AI-REML) implemented in the DMU software package (Madsen *et al.*, 2006; Madsen and Jensen, 2013).

Results

LS was measured on 1,725 larvae from 184 dams representing 73 sires, and DT was estimated for 630 flies from 97 dams representing 21 sires. The number of larvae measured in each full-sib family ranged from 1-20 larvae, and the number of adult offspring emerging ranged from 1-17 flies. Descriptive statistics for both traits are presented in Table 1.

The estimated heritabilities (Table 2) were moderate for both LS and DT. The random effect of rearing vial contributes with a large portion of the total variance for DT.

Table 1. Descriptive statistics for the traits mean 7-day larval size (based on full-sib family means) and egg-to-adult development time in houseflies (*Musca domestica*).

Trait	Unit	No. records	Mean	SD	Median	Min	Max
Mean 7-day larval size	mm^2	184	15.99	8.19	16.24	2.63	30.81
Egg-to-adult development time	Hours	630	500.29	45.17	494.0	422.0	686.0

Table 2. Estimated genetic (σ_a^2), rearing vial (σ_{vial}^2) and error (σ_e^2) variances and heritabilities (h^2) for mean 7-day larval size and egg-to-adult development time with the corresponding standard error SE(h^2) from univariate trait analysis.

Trait	σ_a^2	σ_{vial}^2	σ_e^2	$h^2 \pm SE(h^2)$
Mean 7-day larval size	6.05	-	46.70	0.13±0.33
Egg-to-adult development time	480.42	1,019.92	453.00	0.25±0.44

Discussion

This study aimed at estimating variance components and heritabilities for two production relevant traits in housefly larvae and adults, using a complex mating design frequently applied in aquacultural species (Gjerde *et al.*, 1996). Few studies have presented genetic parameters for traits relevant for protein production in holometabolous species such as the housefly (e.g. Bryant and Meffert, 1998). Estimating genetic parameters for holometabolous species is challenging, partly because physical tagging for individual identification is impossible due to metamorphosis, optimal rearing conditions and environmental susceptibility are life-stage specific and isolated rearing introduces undesired variation, all of which are required to obtain accurate phenotypes from individuals with known genetic relationships. We present moderate heritability estimates for the two investigated traits, but the challenges are reflected by the large variation around the estimates. Additional challenges will arise in coupling phenotypes from different life-stages to estimate genetic correlations. We have approached this issue by designing the experiment so that the two traits were obtained from full-sibs from the same oviposition event. The methods and results presented here constitute the first steps towards the practical implementation of insect breeding plans. Next steps include estimation of the genetic correlations between DT and LS and other traits relevant for insect production (work in progress). Subsequently we will evaluate the potential of genetic gain of selective breeding through simulation studies and compare this with the realised gain obtained from laboratory selective breeding.

Acknowledgements

We thank H. Blendstrup, S. Hansen, T. Kirwa, L. Chen and the student assistants J.W. Jensen and E.M. Battersby at Aalborg University for assistance with the experimental work. This project was funded by the Danish Council for Independent Research (DFF-0136-00171B).

References

Bryant E.H. and Meffert L.M. (1998) Evolution 52(2):626-630. https://doi.org/10.2307/2411098

Eriksson T. and Picard C.J. (2021) J. Insects as Food and Feed 7(5):661-682. https://doi.org/10.3920/JIFF2020.0097

Gjerde B., Gjøen H.M. and Villanueva B. (1996) Livest. Prod. Sci. 47(1):59-72. https://doi.org/10.1016/S0301-6226(96)01000-7

Hill W.G. (2010) Philos. Trans. R. Soc. Biol. Sci. 365(1537):73-85. https://doi.org/10.1098/rstb.2009.0203

Laursen S.F., Hansen L.S., Bahrndorff S., Nielsen H.M. *et al.* (2021) Insects 12(5):380. https://doi.org/10.3390/insects12050380

Madsen P., and Jensen J. (2013) A User's Guide to DMU: a package for analysing multivariate mixed models. Available at: https://dmu.ghpc.au.dk/dmu/DMU/Doc/Current/

Madsen P., Sørensen P., Su G., Damgaard L.H. *et al.* (2006) Proc. of the 8th WCGALP, Belo Horizonte, Brazil.

Pastor B., Velasquez Y., Gobbi P. and Rojo S. (2015) J. Insects as Food and Feed 1(3):179-193. https://doi.org/10.3920/JIFF2014.0024

Tomberlin, J.K and Cammack J.A. (2017) Black Soldier Fly: biology and mass production. In: Van Huis, A. and Tomberlin, J.K. (eds) Insects as food and feed: from production to consumption. Wageningen Academic Publishers, Wageningen, the Netherlands, pp. 230-246.

Van Huis A. and Oonincx D.G.A.B. (2017) Agron. Sustain. Dev. 37:43. https://doi.org/10.1007/s13593-017-0452-8

615. Genetic improvement program in black soldier fly for increased larval weight in industrial applications

E. Facchini[1*], K. Shrestha[2], E. van den Boer[2], P. Junes[2], G. Sader[2], K. Peeters[1] and E. Schmitt[2]

[1]Hendrix Genetics Research, Technology & Services B.V., Spoorstraat 69, 5831 CK Boxmeer, the Netherlands; [2]Protix Biosystems B.V., Industriestraat 3, 5107 NC Dongen, the Netherlands; elena.facchini@hendrix-genetics.com

Abstract

Black soldier fly farming has increased in recent years. The black soldier fly larvae convert low-grade organic materials into high-value food, feed and technical products. To meet the rising demands for proteins in the feed and food industries, there is a need to improve the efficiency of production. Selective breeding can play an important role in such a scenario. Here we present the outcomes of a breeding program for increased larval weight, that started in 2019. The performance of the selected line over the base population line was evaluated in six rounds, under different environmental conditions. Under industrial conditions, an average improvement of +39% in larval weight, +34% in wet crate yield, +26% in dry matter crate yield, +32% in crude protein crate yield and +21% crude fat crate yield was achieved. This research clearly demonstrates the potential of selective breeding for improved efficiency when farming black soldier fly.

Introduction

The black soldier fly (BSF), *Hermetia illucens* L., is globally utilized for organic waste remediation to produce various products of interest. The industrial production cycle for BSF typically lasts 40-45 days and consists of four stages: the breed (where flies mate to produce eggs), the nursery (eggs hatch into young larvae), the rearing (young larvae grow to harvestable larvae) and the pupation (only for the breeding animals; incubation of prepupae to pupae that will eclose to adult flies). The key stage of production is the rearing stage, where young larvae convert low-grade organic waste into derivable products such as protein meal, fat and frass (fertilizer).

In addition to its use in aquaculture and pet food diets, the European Commission has recently authorized the use of insects derived processed animal proteins in monogastric livestock diets (Commission Regulation (EU) 2021/1372 amending Annex IV to Regulation (EC) No 999/2001). At present, the demand for insect protein is higher than the supply, and the demand is expected to grow 50 times more by 2030 globally (Ploegmakers, 2021). One means of increasing supply is selective breeding, which has been widely applied in plants and other animal species, but not yet reported in BSF.

In mealworms, pupal body mass of *Tenebrio molitor* increased 66% after eight years of selection (Morales-Ramos *et al.*, 2019). In addition, studies reported body mass to be medium to high heritable trait in different insects' orders, with heritability estimates ranging from 0.2 to 0.6 (Dingle *et al.*, 1980; Iyengar and Eisener, 1999; Saastamoinen, 2008; Jumbo-Lucioni *et al.*, 2010; Ellen *et al.*, 2015; Facchini *et al.*, 2021). These heritability estimates show that the population mean for body mass can be improved through selective breeding.

In 2019, we started a selective breeding program for increased larval body weight in BSF. In this paper, we present how the outcomes of this three-year program translated under industrial production conditions. Our research findings are relevant as they provide an indication of the role that selective breeding may play in improving the performance of BSF.

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_615, © E. Facchini *et al.* 2022

Materials & methods

Black soldier fly lines. The aim of the study was to assess the differences in the rearing larval performance of the Body Weight line (BW) compared to the Base Population line (BP) over multiple consecutive rounds. The improved BW line was derived from the BP line in March 2019 and was selected each generation for increased larval body weight following a genetic improvement program.

Experimental setup. To compare the two lines, six sequential rounds were carried out between May 2020 and June 2021. Three different generations of the BW line were used in this study. For the first, second and third rounds, BW animals were sampled from the Genetics lab in May 2020, September 2020, and January 2021 respectively. Each round was characterized by a different production setting. Table 1 summarizes the settings which differed in feed type, rearing environment, larval density and feed ration.

For the first three rounds, a control BP group was set up next to the BW line. Six crates per line were set up in parallel. Operators were blinded to avoid bias from treating the animals differently based on the line. In round one and three, operators could feed different amounts on a crate level, according to their discretion. In round two, operators were instructed to deliberately feed the same amount on crate level, even if the animals indicated different needs.

For the last three rounds, no control BP group was setup next to the 6-10 BW line crates for logistic reasons. Therefore, from fourth round onwards, the comparisons were made with BP data derived from randomly selected production batches harvested during the 15 flanking days before and after the BW experimental batch (a minimum of 35 crates). The feeding amounts for BW line henceforth were determined based on experience from previous three rounds and with the aim of maximizing crate output per unit within the existing production system.

Rearing and data collection. On the 7th day of the experiments (harvest day at the farm), larvae were separated from leftover substrate by sieving. For all rounds and both lines, average larval weight (mg) was measured at harvest day. Larvae were randomly picked from each crate from different locations within the crate. Larvae were cleaned and 100 larvae were randomly sampled and weighed. Average larval weight was determined by dividing the sample weight (mg) by the number of larvae in the sample Wet crate yield (kg),

Table 1. Summary of the experimental setup. Density and feed ration (grams of feed per larva) are represented as deviations (Δ) in percentage from a common benchmark set at BP industrial production conditions.

Round	Feed type	Environment	Line	Δ density	Δ feed ration
1 May-20	A	1	BW	-15%	21%
			BP	-15%	4%
2 Sep-20	A	2	BW	-55%	135%
			BP	-55%	135%
3 Jan-21	B	2	BW	-27%	58%
			BP	-27%	12%
4 Mar-21	B	2	BW	0%	30%
		3	BP	0%	-2%
5 May-21	B	3	BW	0%	30%
			BP	0%	0%
6 Jun-21	B	3	BW	0%	22%
			BP	0%	0%

which reflects the total larval yield of the crate, was obtained after sieving. For nutritional composition analysis, an equivalent volume of larvae from all crates per line per round was mixed and analysed for dry matter (DM), crude protein and crude fat. The feed conversion ratio and protein conversion ratio were also determined, yet not presented as they were not significantly different between the BP and BW line.

Statistical analysis. For all traits, a paired samples t-test was used to test the difference between the two population means. All analyses were carried out in R statistical computing environment version 4.0.0 (R Development Core Team, 2020).

For average larval weight, the following linear regression was fitted to test the relationship between average larval weight and the explanatory variables through ANOVA (Equation 1):

$$y_{ijk} = \mu + L_i + F_k + R_j + \varepsilon_{ijk} \quad (1)$$

where μ represents the overall mean, L is the fixed effect of the line (i = BW, BP), F is a covariate for the adjusted feed ration (feed amount/estimated larvae density), R is the fixed effect of the round (j =1, 6) and ε represents the random error term.

Results & discussion

T-test results showed a significant difference between BW and BP average performances for larval weight (t (5) =10.988 P<0.001), wet crate yield (t (5) =6.0535, P<0.001), DM crate yield (t (5) =7.4597, P<0.001), crude protein per crate (t (5) =5.2401, P<0.01), and crude fat per crate (t (5) =5.7807, P<0.01) (Figure 1).

Table 2 reports the results from the linear regression model. All variables showed a significant effect on average larval weight. Interest is on the variance explained by genetics represented by the variable Line. The variable Line explained the largest part (47%) of the variance in the model. This reveals the inherent genetic difference in average larval weight of the BW line compared to the BP line.

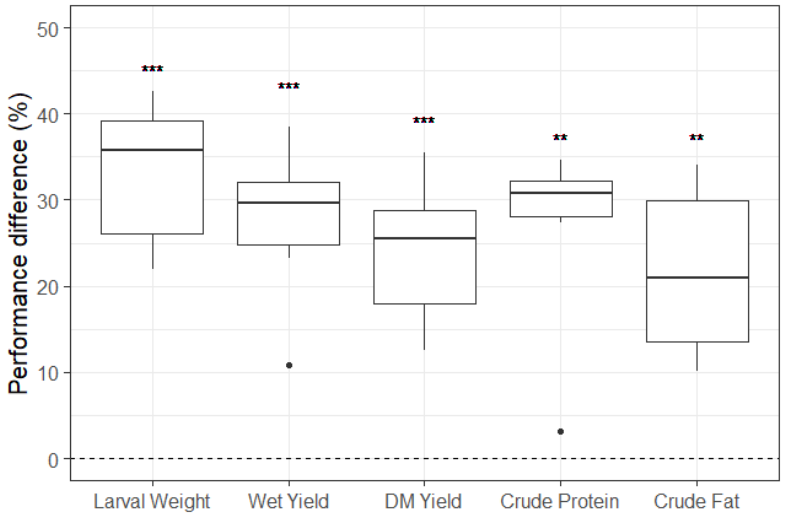


Figure 1. Performance differences in average crate outputs between BW and BP line. Boxplots represent the distribution of the results from all the six rounds for larval weight, wet yield, dry matter (DM) yield, crude protein, crude fat; ***: P< 0.001, **: P<0.01.

Table 2. Analysis of variance for average larval weight.[1]

Variable	Df	Sum sq	Mean sq	F value	Pr(>F)
Line	1	112,118	112,118	1,228.06	$P<0.001$
Adj. Feed ration	1	82,464	82,464	903.24	$P<0.001$
Round	5	15,860	3,172	34.74	$P<0.001$
Residual	302	27,572	91		

[1] Df = degrees of freedom; Sum sq = sum of squares; Mean sq = mean sum of squares.

In rounds five and six, under fully automated production settings, we found an average improvement of BW over BP of +39% in average larval weight +34% in wet crate yield, +26% in DM crate yield, +32% in crude protein per crate and +21% crude fat per crate.

Conclusions

Overall, results show that substantial genetic improvement can be realized in larvae weight in BSF which translates into improvement in crate yield. In other words, the selection under the genetic improvement program conditions translated into improvements in the industrial setting.

References

Dingle H., Blakley N. R., and Miller E. R. (1980) Evolution 34(2):371–385. https://doi.org/10.2307/2407400

Ellen E.D., Peeters K., Verhoeven M., Gols R., Harvey J.A. *et al.* (2016) Evolution 70:207-217. https://doi.org/10.1111/evo.12835

European commission (2021) Commission Regulation (EU) 2021/1372 of 17 August 2021 amending Annex IV to Regulation (EC) No 999/2001 of the European Parliament and of the Council as regards the prohibition to feed non-ruminant farmed animals, other than fur animals, with protein derived from animals

Facchini E., De Iorio M. G., Turri F., Pizzi F., Laurino D. *et al.* (2021) Animals *11*(11):3054. https://doi.org/10.3390/ani11113054

Iyengar V. K., and Eisner T. (1999) PNAS 96(16):9169-9171. https://doi.org/10.1073/pnas.96.16.9169

Jumbo-Lucioni P., Ayroles J. F., Chambers M. M., Jordan K. W., Leips J. *et al.* (2010) BMC genom *11*(1):1-13. https://doi.org/10.1186/1471-2164-11-297

Morales-Ramos J. A., Kelstrup H. C., Rojas M. G., and Emery V. (2019) J Insect Sci *19*(2):4. https://doi.org/10.1093/jisesa/iey110

Ploegmakers M. (2021) The path to a half million metric tons of insect protein. Available at: https://www.allaboutfeed.net/all-about/new-proteins/the-path-to-a-half-million-metric-tons-of-insect-protein/

R Core Team (2020). Available at: URL: https://www.R-project.org/

Saastamoinen M. (2008) Heredity 100(1):39-46. https://doi.org/10.1038/sj.hdy.6801056

616. *Tenebrio molitor* genomic structure among available populations

E. Eleftheriou[1], B. Vacherie[2], K. Labadie[2], C.G. Athanassiou[3], T. Rigaud[4], Y. Moret[4], T. Lefebvre[5] and MA. Madoui[1,6*]

[1]Génomique Métabolique, Genoscope, Institut François Jacob, Commissariat à l'Energie Atomique (CEA), CNRS, Univ Evry, Université Paris-Saclay, 91057 Evry, France; [2]Genoscope, Institut de biologie François Jacob, CEA, Université Paris-Saclay, 91057 Evry, France; [3]Department of Agriculture Crop Production and Rural Environment, School of Agricultural Sciences, University of Thessaly, Fytokou Str., 38446 Volos, Greece; [4]UMR CNRS 6282 BioGéoSciences, Équipe Écologie Évolutive, Université Bourgogne Franche Comté, 21000 Dijon, France; [5]Ynsect, 91000 Evry, France; [6]Service d'Etude des Prions et des Infections Atypiques (SEPIA), Institut François Jacob, Commissariat à l'Energie Atomique et aux Energies Alternatives (CEA), Université Paris Saclay, Fontenay-aux-Roses, France; mohammedamin.madoui@cea.fr

Abstract

Tenebrio molitor is now a promising species for insect farming to produce high protein products with low environmental impact. However, the molecular diversity and the genetic structure in available *T. molitor* populations are understudied. We analysed the genomic structure of 18 populations from 12 countries using Illumina pool-seq data. More than 5 million SNPs were selected to compute allele frequencies, then multivariate and pairwise-F_{ST} analyses were used to estimate their relative genetic distances. The results showed two very distant populations, one wild population sampled in Serbia and one from Germany, both presenting very distinct genomic structure. The populations from USA, Spain and Holland presented moderate genomic differentiation and 13 populations were closely-related. This pool-seq analysis highlighted a lack of molecular diversity between most of the sampled populations and will allow the identification of loci under selection. Current phenotyping and genotyping efforts will help guiding future breeding programs.

Introduction

By 2050, the global population will reach approximately nine billion people, increasing the demand for animal protein by 76% (Alexandratos and Bruinsma, 2012). This future demand starts to question the sustainability of our conventional production systems. Nowadays, insect production is considered as a serious alternative for food and feed production for many reasons. Essentially for the nutritional composition of edible insects (Nowak *et al.*, 2016) and for the low environmental impact of its production compared to other conventional livestock production systems (van Huis, 2013).

The yellow mealworm beetle *Tenebrio molitor* has been described as a good candidate for animal and even human nutrition (Cortes Ortiz *et al.*, 2016). Thus, genomic efforts and industrial scale production of *T. molitor* progressed during the last decade (Eleftheriou *et al.*, 2021). Despite promising results on *T. molitor* production, several aspects concerning the *T. molitor* biology remain challenging to sustain its production (Morales-Ramos *et al.*, 2019). The need to have access to the molecular and phenotypic diversity of *T. molitor* is a first step before further genetic improvement. Indeed, industrial farming of *T. molitor* will require monitoring the molecular diversity in order to better anticipate inbreeding depression that may cause decrease of biomass production. Here, we used Illumina pool-seq on 18 *T. molitor* populations and sequenced 720 individuals to identify their molecular diversity and genomic structure.

Materials & methods

Biological material. A total of 18 populations were provided by academic institutions and private companies. University of Burgundy (Dijon, France) provided three populations, one called *Dijon*, one called *Serbia* that corresponds to a wild population sampled in a disaffected mill in Serbia and one called

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_616, © E. Eleftheriou *et al.* 2022

Holland that corresponds to a population initially provided by a private Dutch company. University of Thessaly (Volos, Greece) provided seven populations, two populations from Italy called *Italy_1* and *Italy_2*, one population from Greece called *Greece*, one population from Spain called *Spain*, one population from Germany called *Germany*, one population from the United States of America called *USA* and one population from Turkey called *Turkey*. Mystik Farms (Cap Santé, Québec, Canada) provided one population called *Canada*. Papek s.r.o. (Jaroměřice nad Rokytnou, Czech Republic) provided one population called *Czech*. Two populations called *Pro_+* and *Pro_-* were provided by Pronutrix (Harelbeke, Belgium). One population called *Belgium* was provided by Nusect (Deerlijk, Belgium). The Ynsect company (Evry, France) provided one population called *AAF_Dole*. The Kinsect company (France) provided one population called *Kynsect_HOL*. The Star Food company (Barneveld, Netherlands) provided one population called *Starfood_HOL*.

Pool-seq sequencing. *Tenebrio molitor* larvae were first ground individually using ZR BashingBead Lysis Tubes (2 mm) (ZymoResearch, CA, USA) and subsequently DNA extraction was performed using the Quick-DNA 96 Plus Kit (ZymoResearch). The extracted DNA was normalized and 40 individuals from the same population were pooled equimolarly. For each pool, DNA libraries were prepared using the NEBNext® Ultra II DNA Library prep kit for Illumina (New England Biolabs, MA, USA) and sequenced on an Illumina NovaSeq6000 instrument (Illumina, San Diego, CA, USA) using 150 base-length read chemistry in paired-end mode to reach around 20x of depth of coverage by individual.

Population genomic analysis. Raw reads were cleaned and aligned against the *T. molitor* reference genome using *bwa mem* (Li and Durbin, 2009). Reads duplicates were removed and SNPs were called using samtools/bcftools. For each SNP, the alternative allele frequency was calculated, to get rid of sequencing errors, only SNPs with a minimum frequency of 5% in at least one population were kept. The relatedness between populations was first observed by principal component analysis (PCA) using the allele frequencies as input data. Then, the genomic distance between each population was estimated by computing the pairwise-F_{ST}, first we computed the F_{ST} for each allele as follow: $FST = var(p) / (mean(p) \times (1 - mean(p)))$, where p is a vector of alternative allele frequencies of two populations. Second, we used the median F_{ST} over all alleles to compute the final pairwise-F_{ST}.

Results & discussion

We obtained more than 3 million SNPs from each pool of individuals. The PCA analysis based on the allele frequencies of 5,285,733 SNPs showed that the two first PCA axes (explaining only 36% of the total variance) are strongly linked to *Serbia* and *Germany* populations while the other 16 populations are clustered into one single group located close to the origins of the axes (Figure 1.A). This suggests that *Serbia* and *Germany* have very different genomic structures while the 16 other populations may have closer genomic structures.

The pairwise- F_{ST} analysis showed that *Serbia* presented F_{ST} values ranging from 0.09 to 0.34 with the other populations, the larger value corresponding to the F_{ST} with *Germany* (Figure 1.B). *Germany* presented F_{ST} values ranging from 0.25 to 0.34. *Kynsect_HOL*, *USA* and *Spain* pairwise- F_{ST} values between 0.03 and 0.12. The last 13 populations presented low genomic differentiation between them with F_{ST} value ranging from 0 to 0.06. The F_{ST} analysis confirmed the PCA results by highlighting *Serbia* and *Germany* as two populations with high genomic differentiation compared to the 16 other populations but also validates the presence of different genomic structures between these two populations. But the F_{ST}-based analysis showed that a moderate differentiation occurred in *Spain*, *USA* and *Kynsect_HOL* and a low differentiation between the 13 other populations whose eight were sampled from private breeders. The recent interest in mealworm production by private companies may have led them to supply recently on the same or closely-related suppliers explaining possibly the observed low differentiation between their populations. On the other hand, evolutionary forces such as genetic drift and/or selection may also have occurred in the five more differentiated populations especially for *Serbia* that was sampled in a disaffected mill.

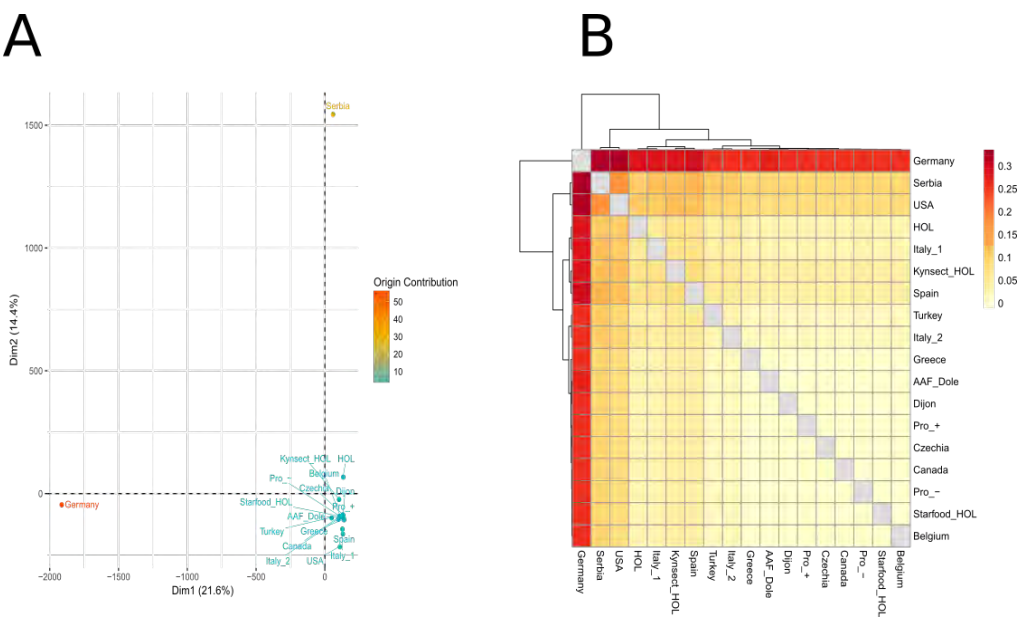


Figure 1. Genomic structure in 18 *T. molitor* populations. (A) PCA based on allele frequencies. (B) Pairwise-F_{ST} distance matrix.

Conclusions

From the pool-seq analysis presented in this study, we identified two populations having strong differentiation compared to the other 16 populations. We may consider that our results highlighted that the majority of the sampled populations presents a low molecular diversity between them. Current phenotyping of complex traits such as growth rate, fertility and immunity in available populations as well as their genotyping will help the selection of the populations having the most interesting performances for industrial production and will guide future breeding programs. Finally, to better understand molecular evolution of *T. molitor*, further analysis will be performed from the data generated in this study to identify genes under natural selection and the corresponding biological processes.

References

Alexandratos, N., and Bruinsma, J. (2012). Food and Agriculture Organisation (FAO). World agriculture towards 2030/2050: the 2012 revision. Available at: https://www.fao.org/3/ap106e/ap106e.pdf

Cortes Ortiz, J.A., Ruiz, A.T., Morales-Ramos, J. A. Thomas, M., Rojas, M. G., *et al.* (2016). Insect Mass Production Technologies. In: Dossey, A.T., Morales-Ramos, Juan A., Rojas, M. Guadalupe. (eds) Insects as Sustainable Food Ingredients. Academic Press, San Diego, USA, pp. 153–201. https://doi.org/10.1016/B978-0-12-802856-8.00006-5

Eleftheriou, E., Aury, J., Vacherie, B., Istace, B., Belser, C., *et al.* (2021). Open Res. Eur. 1. https://doi.org/10.12688/openreseurope.13987.1

Li, H., and Durbin, R. (2009). Bioinformatics 25:1754–1760. https://doi.org/10.1093/bioinformatics/btp324

Morales-Ramos, J.A., Kelstrup, H.C., Rojas, M.G., and Emery, V. (2019). J. Insect Sci. 19(2):4. https://doi.org/10.1093/jisesa/iey110

Nowak, V., Persijn, D., Rittenschober, D., and Charrondiere, U.R. (2016). Food Chem. 193:39–46. https://doi.org/10.1016/j.foodchem.2014.10.114

Van Huis, A. (2013). Annu. Rev. Entomol. 58:563–583. https://doi.org/10.1146/annurev-ento-120811-153704

617. Inferring whole-genome tree sequences and population and demographic parameters of the Western honeybee

J. Obšteter[1*], A. Marinč[1], J. Prešern[1], D. Wragg[2] and G. Gorjanc[2]

[1]Agricultural Institute of Slovenia, Hacquetova ulica 17, 1000 Ljubljana, Slovenia; [2]The Roslin Institute and Royal (Dick) School of Veterinary Studies, University of Edinburgh, Easter Bush, Midlothian, EH259RG, Edinburgh, United Kingdom; jana.obsteter@kis.si

Abstract

Although honeybees are the most important managed pollinators, their demographic history is not fully clear. Here we inferred whole-genome succinct tree sequences with *tsinfer* software to determine the honeybee population parameters and demographic history. We used a dataset of 691 whole-genome sequenced drones out of which 240 had subspecies info and 451 were hybrids. Our results confirmed that the A lineage subspecies are the least differentiated from the rest of subspecies and the most diverse. We also showed that the C lineage experienced a bottleneck followed by a large expansion related to their commercial use. Future work will use the inferred tree sequence to estimate the effective population size of the ancestral population and date the ancestral nodes and split events.

Introduction

Although honeybees are the most important managed pollinators, their demographic history remains undetermined. Previous studies agree that the genus *Apis* originated from Asia, but the origin of *Apis mellifera* is still debated with possible origins being Asia (Wallberg *et al.*, 2014), the Middle East (Cridland *et al.*, 2017), or Africa (Whitfield *et al.*, 2014). Studies agree that *A. mellifera* split from *A. cerana* around 6 million years ago (Chen *et al.*, 2018). In contrast, the estimated split time of the subspecies ranges from between 35 and 20ky ago (Wallberg *et al.*, 2014; Chen *et al.*, 2016) to around 1 million years ago (cited in Whitfield *et al.*, 2006). These differences are partially due to each study analysing different sets of subspecies. The aim of this study was to use drone whole-genome sequences to infer succinct tree sequences and shed light on demographic history of the honeybee.

Materials & methods

Inference of the ancestral alleles. We used Cactus (Armstrong *et al.*, 2020) to align the reference genomes of *Apis mellifera* (Amel_HAv3.1, GCF_003254395.2), *Apis cerana* (ACSNU-2.0, GCF_001442555.1), *Apis dorsata* (Apis dorsata 1.3, GCF_000469605.1), and *Apis florea* (Aflo_1.1, GCF_000184785.3). Next, we used a custom Python script to prepare the file for ancestral inference using *est-sfs* (Keightley and Jackson, 2018). The file contained allele counts for the focal species and the outgroups for 12,449,207 out of 18,972,875 variants in the raw VCF. We ran the *est-sfs* with three outgroups and the Kimura 2-parameter model.

Data set. We used a genomic dataset of 724 drones (Parejo *et al.*, unpublished data, provided by David Wragg). The raw VCF contained 18,972,875 variants. We first removed the variants identified as heterozygous when called diploid (Wragg *et al.*, 2021). Next, we filtered the data by genotype quality (>30), depth (>5), minor allele frequency (>0.01), and missing frequency (<0.1). We removed the samples and variants with >10% missing genotypes. The final VCF included 691 samples and 2,216,964 variants on chromosomes. 62 samples were marked as *A. m. carnica* (C lineage), 10 as *A. m. ligustica* (C lineage), 93 as *A. m. mellifera* (M lineage), 10 as *A. m. caucasica* (C or O lineage), 8 as *A. m. unicolor* (A lineage), 28 as *A. m. capensis* (A lineage), 29 as *A. m. scutellata* (A lineage), and 451 samples were considered as hybrids.

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_617, © J. Obšteter *et al.* 2022

Inference of the demographic history. We used the final VCF for the demographic inference. We included the ancestral allele where available, elsewhere the *A. mellifera* reference allele was considered to be ancestral. We used *tsinfer* (Kelleher *et al.*, 2019) to infer the genealogical history of the samples by chromosome. We used the tree sequences to compute principal component analysis (PCA), fixation index (Fst), Nei's nucleotide diversity, Tajima's D, and the genealogical nearest neighbour (GNN) between the samples. We combined the results from the chromosomes by weighting the output by the chromosome length.

Results

We inferred the ancestral allele for 7,781,679 sites from the raw VCF. This represented 1,201,346 sites from the final VCF file. For 80.8% of the sites, the inferred ancestral allele matched the major allele of our sample of drones, and the *A. mellifera* reference allele for 59.4% of the sites. In the PCA (Figure 1) the subspecies of a particular lineage grouped together. African A (lineage) subspecies (*A. m. scutellata, A. m. capensis*, and *A. m. unicolor)* separated most from the rest of subspecies. Inspecting PC3 revealed a further separation of African subspecies, with *A. m. unicolor* diverging from the other two African subspecies. The analysis also revealed that the hybrids from Reunion placed between the A and the C lineage.

We inspected the differentiation, diversity, and selection neutrality of the subspecies by computing the Fst, Nei's nucleotide diversity, and Tajima's D (Figure 2). First, we observed the lowest Fst, 0.05, between *A. m. carnica* and *A. m. ligustica,* and the highest Fst, 0.47, between *A. m. ligustica* and *A. m. unicolor* (Figure 2a). The Fst results suggested non-uniformity of the African subspecies. While the Fst between *A. m. scutellata and A. m. capensis* was 0.11, the Fst of the two with *A. m. unicolor* was higher, respectively 0.25 and 0.27. The results also suggested that the *A. m. unicolor* is the most differentiated from the non-African species. Next, we inspected Nei's nucleotide diversity (Figure 2b) and observed the highest values for the A lineage subspecies (>0.0015). The diversity for other subspecies was an order of magnitude lower and was lowest in the subspecies of the C lineage (<0.00062). Last, we observed the highest positive Tajima's D for the A lineage subspecies and the lowest negative Tajima's D for the C lineage subspecies (Figure 2c).

We computed the GNN for all the samples for the subspecies and hybrids from the island of Reunion (Figure 3). Without hybrid samples the subspecies grouped in concordance with the PCA and previous knowledge (Figure 3a). The subspecies of the lineages grouped together, and the O lineage grouped with the C lineage. Inclusion of hybrid samples from the island of Reunion changed the output: the hybrids grouped with the *A. m. unicolor*, and the clade was 'pulled' into the C lineage clade (Figure 3b).

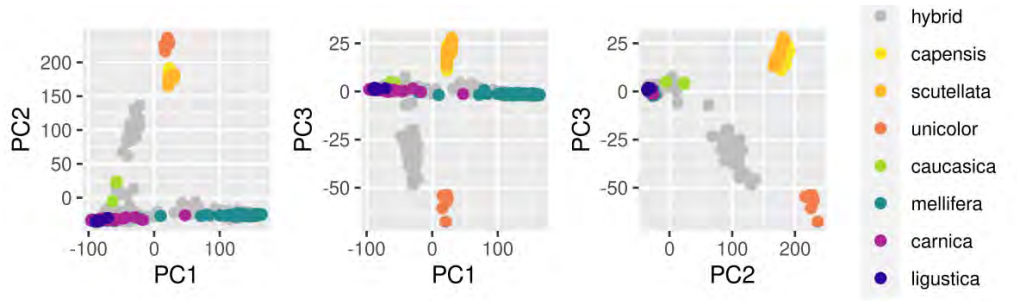


Figure 1. Principle component analysis on the tree sequences.

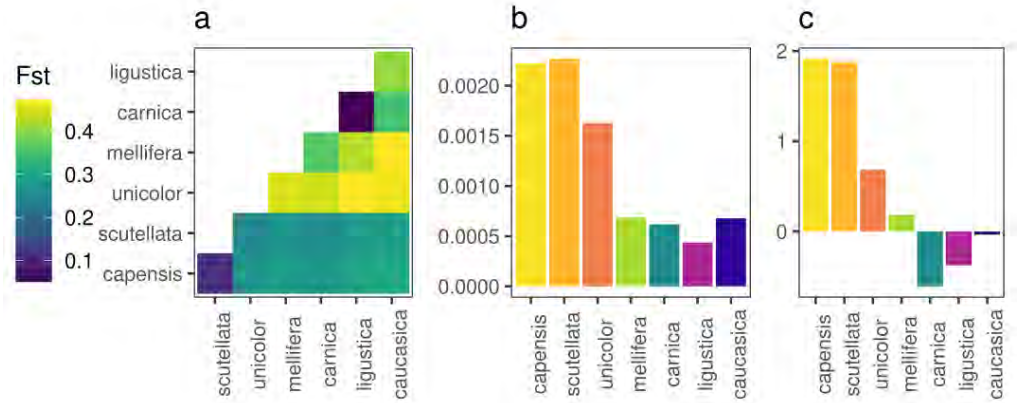


Figure 2. Fst (a), Nei's nucleotide diversity (b), and Tajima's D (c) of the subspecies.

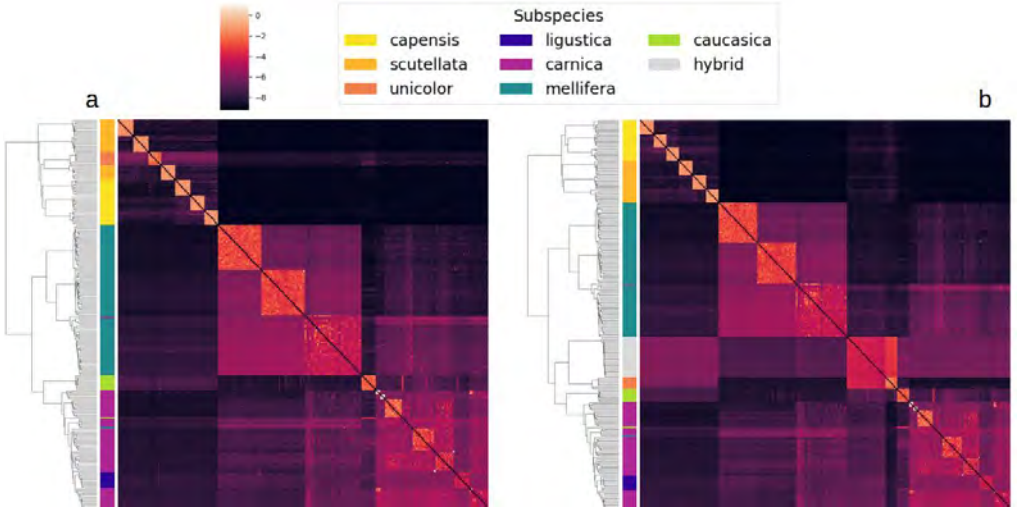


Figure 3. Genealogical nearest neighbour of (a) the subspecies samples and (b) subspecies samples and hybrid samples from the island of Reunion.

Discussion

In this study, we inferred whole-genome tree sequence to study the population parameters and demographic history of the *Apis mellifera*. Tree sequences present an efficient way to encode whole-genome information, while revealing the evolution of genomes. PCA results suggested a non-uniformity of the subspecies of the A lineage, which can be explained by the geographical location of the subspecies. While *A. m. capensis* and *A. m. scutellata* are indigenous the African mainland, *A. m. unicolor* is indigenous to Madagascar.

Our Fst pattern was in concordance with previous studies revealing that the A lineage is the least differentiated from the rest of the subspecies and that the subspecies of the M lineage are more differentiated from the C than from the A lineage (Cridland *et al.*, 2017; Dogantzis *et al.*, 2021). This supports the theory of an earlier split of the C and the M lineage and a separate colonization of Europe via two routes (Whitfield *et al.*, 2006; Cridland *et al.*, 2017). Computing the Tajima's D revealed highly negative Tajima's D for both subspecies

of the C lineage. Negative Tajima's D suggests a purifying selection or a population expansion after a bottleneck. The latter could explain the large negative Tajima's D of the *A. m. carnica* and *A. m. ligustica* since they have been widely distributed due to their commercial use. The clustering of the subspecies in the GNN structure confirmed the PCA results. When we included the hybrids from the island of Reunion, they clustered with the *A. m. unicolor*. This was expected since Reunion is geographically closest to Madagascar, the origin of *A. m. unicolor*. However, the hybrid's clade was 'pulled' into the C lineage clade. This can be explained by the fact that historically, routine imports of *A. m. carnica* and *A. m. ligustica*, have had a lasting impact on the genetic make-up of the island's honeybees – remaining long after importing of bees became prohibited (Wragg *et al.*, 2019).

The succinct tree sequence is a genealogical representation of whole-genome sequence data and are hence a powerful tool to store the data and estimate population parameters. In this paper we present only few of the metrics that can be computed from the succinct trees. Future work will include estimating the coalescent rates to determine the trajectory of effective population size of the subspecies and dating the ancestral nodes in the tree sequences to determine the time of splits between lineages and subspecies.

References

Armstrong J., Hickey G., Diekhans M., Fiddles I., and Novak A. M. (2020) Nature 587:246–251. https://doi.org/10.1038/s41586-020-2871-y

Chen C., Liu Z., Pan Q., Chen X., Wang H., *et al.* (2016) Mol Biol Evol 33(5):1337-1348. https://doi.org/10.1093/molbev/msw017

Chen, C., Wang, H., Liu, Z., Chen, X., Tang, J., *et al.* (2018) Mol Biol Evol 35(9):2260–2271. https://doi.org/10.1093/molbev/msy130

Cridland J.M., Tsutsui N.D., and Ramírez S.R. (2017) Genome Biol Evol 9(2):457-472. https://doi.org/10.1093/gbe/evx009

Dogantzis K. A., Tiwari T., Conflitti I. M., Dey A., Patch H. M., *et al.* (2021) Sci Adv 7(49):eabj2151. https://doi.org/10.1126/sciadv.abj2151

Keightley P.D., and Jackson B.C. (2018) Genetics 209(3):897-906. https://doi.org/10.1534/genetics.118.301120

Kelleher J., Wong Y., Wohns A.W., Fadil C., Albers P.K., *et al.* (2019) Nat Genet 51(9):1330–1338. https://doi.org/10.1038/s41588-019-0483-y

Wallberg A., Han F., Wellhagen G., Dahle B., Kawata M., *et al.* (2014) Nat Genet 46(10):1081-8. https://doi.org/10.1038/ng.3077

Whitfield C.W., Behura S.K., Berlocher S.K., Clark A.G, Johnson J.S., *et al.* (2006) Science 314(5799):642-645. https://doi.org/10.1126/science.1132772

Wragg D., Techer M.A., Canale-Tabet K., Basso B., Bidanel J.P., *et al.* (2019) Genome Biol Evol 10(1):220–238. https://doi.org/10.1093/gbe/evx247

Wragg D, Sonia E., Benjamin B., Kamila C.T., Emmanuelle L., *et al.* (2021) bioRxiv 2021.09.20.460798. https://doi.org/10.1101/2021.09.20.460798

618. Heritability of disease resistance to chronic bee paralysis, chalkbrood and nosemosis in the honeybee (*A.m.carnica*)

A. Hoppe*, R. Bernstein, M. Du and K. Bienefeld

Institute for Bee Research Hohen Neuendorf, Friedrich-Engels-Str. 32, 16540 Hohen Neuendorf, Germany; andreas.hoppe@hu-berlin.de

Abstract

Honeybee diseases related to viruses and fungi play a major role in colony losses. Because of the ubiquity of the pathogens, breeding for genetic resistance is highly desirable as a sustained strategy. Culling affected colonies in breeding populations has apparently not sufficed to develop resistance. Arguably, targeted selection on yield and other traits overshadowed the culling effect and explicit consideration of resistance traits is necessary. Here, we present the development of additive-genetic models of disease traits on the basis of best linear unbiased prediction in the European Carnica www.beebreed.eu breeding program. Although the data is sparse and the diseases are only reported as binary information, validation shows that disease resistance is heritable and can be used for targeted selection. According to heritability estimations and predictability of phenotypic values, chronic bee paralysis is the most promising target, followed by chalkbrood, and finally nosemosis.

Introduction

Chalkbrood, caused by the fungus *Ascosphaera apis*, is a widely distributed honeybee disease and generally connected to cold and humid weather conditions (Aronstein and Murray 2010). Increased brood hygiene has been found as a partially protecting factor (Gilliam *et al.* 1983). *Nosema apis* is an important threat to professional beekeeping (Martín-Hernanández *et al.* 2018) with a clear seasonal pattern (Fries *et al.* 2010), whereas occurrence of chronic bee paralysis (CBP) follows an erratic pattern (Amiri *et al.* 2014).

In general, three conditions must come together for the occurrence of these diseases: (1) the infectious agent must be present in the apiary in a certain quantity and infectivity; (2) the environmental conditions must be unfavourable; and (3) the colony must have susceptibility to the disease. As disease control is mostly done with medication and altered management techniques, the focus lies on the former two, whereas the hereditary component of susceptibility has received only some attention for chalkbrood (Invernizzi *et al.* 2009). While resistance breeding may be the most elaborate approach, it is potentially the most lasting.

Materials & methods

Breeding system, pedigree cutting, data preparation. Studbooks of the international Carnica population maintained in www.beebreed.eu (Hoppe *et al.* 2020) were downloaded on 29th Nov 2021. Performance testers recorded the information whether the tested colony suffered from chalkbrood, nosemosis, or CBP as a binary trait. Only apiaries/test years with both ill and healthy colonies were considered as valid phenotypes. Starting from the set of valid phenotypes as seed set, the pedigree was constructed as a subset of all 229,732 colonies in an iterative process with a cycled application of the following steps: (1) add siblings if any sister is in the set; (2) if any colony on an apiary-year combination is in the set, add the remaining colonies; (3) add any colony which is both an ancestor and descendent of colonies in the set. Finally, add all ancestors of colonies in the set. In a pedigree constructed in this way, the calculated relationships between colonies with phenotypes are identical to those calculated with the full pedigree.

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_618, © A. Hoppe *et al.* 2022

Using this set of relevant colonies, a pedigree was set up with the following items: (1) queen; (2) worker community; (3) pseudo sire; and (4) potential offspring queen as in (Hoppe *et al.* 2020). Then, PInCo (Bernstein 2018) was used to calculate the inverse relationship matrix of all pedigree items for direct input in BLUPF90 programs (Misztal *et al.* 2002).

The breeding model applied is a mixed-linear additive genetic model:

$$y = Xb + Z_W a_W + Z_Q a_Q + e \quad (1)$$

where Z_W, Z_Q and X are known incidence matrices, a_W are the worker and a_Q the queen effects of the pedigree items, b are the apiary effects, and e is the residual (Hoppe *et al.* 2020).

Genetic and residual variances were fitted with AIREMLF90 (Misztal *et al.* 2002). As recommended by (Du *et al.* 2022), the results were confirmed with subsequent REMLF90 applications.

To assess the quality of of the models, we applied: (1) the year validation as described in (Hoppe *et al.* 2020); (2) a 5-fold cross-validation of original vs predicted phenotypes; (3) a cross-validation of phenotypes where the apiary effect (calculated in an initial BLUP run) was removed; and (4) a cross-validation of prospective vs realized breeding values.

Results

As can be seen in Table 1, chalkbrood was by far the most frequently observed disease, while for nosemosis and CBP there were still sufficiently many cases to assess a familial accumulation.

AIREML convergence was fast and robust for CBP and chalkbrood, while for nosemosis it was slow. Different heritabilities are shown in Table 2, where he heritability of selection criterion is calculated from (Brascamp *et al.* 2019), while the accessible heritability is the honeybee-specific adaption to satisfy the breeder's equation (Hoppe *et al.* 2020). CBP showed a high heritability similar to gentleness (Hoppe *et al.* 2020), chalkbrood a medium heritability (similar to honey yield), while the heritability for nosemosis was low. The estimated queen/worker correlation for nosemosis was close to -1, presumably because the data was not sufficient to support separation of worker and queen effects (Du *et al.* 2022). The standard errors were lowest for chalkbrood, as the model was supported by the most data. For nosemosis and CBP, the standard errors were relatively high. But even the low end of the standard error range shows high heritability for CBP. Oppositely, the upper end of the standard error range for Nosema indicates still a low Heritability for Nosema.

See Table 3 for the results of the three types of cross-validation. Although the correlations between predicted and real phenotype resp. prospected and realized breeding values were not high compared to the regular breeding traits (see Hoppe *et al.* 2020), it underlines that a genetic predisposition for disease susceptibility exists.

Table 1. Data overview.

	CBP	Chalkbrood	Nosemosis
Disease cases	259	1,792	519
Apiaries with disease cases	115	783	186
Relevant non-disease cases	1,359	8,942	2,533
Colonies in pruned pedigree	4,820	15,962	6,414

Table 2. Heritabilities.

	CBP	Chalkbrood	Nosemosis
Accessible heritability	0.28 (±0.10)	0.18 (±0.03)	0.04 (±0.02)
Heritability of selection criterion	0.68 (±0.25)	0.45 (±0.08)	0.10 (±0.04)
Direct heritability	0.61 (±0.20)	0.20 (±0.05)	0.23 (±0.08)
Maternal heritability	0.51 (±0.23)	0.03 (±0.04)	0.18 (±0.07)
Correlation between worker and queen effect	-0.76 (±0.24)	-0.34 (±0.46)	-1.0 (±0.14)

Table 3. Results of 3 Types of cross-validation.

Validation	Parameter	CBP	Chalkbrood	Nosemosis
Cross-validation of phenotypes	Correlation predicted / real phenotype	0.48	0.29	0.38
Cross-validation, removed apiary effect	Correlation predicted / real phenotype	0.05	0.11	0.064
Cross-validation of breeding values	Correlation prospective / realized breeding value	0.53	0.82	0.66

See Figure 1 for the graphical representation of the results of the relation with relates prospective breeding values to next year's relative phenotypes. For CBP, the quarter of the lowest estimated breeding values stood apart while the rest was less differentiated. For chalkbrood, the separation of the worst and best breeding values was good, but not for the middle values. For nosemosis, there was a separation of the highest breeding values, however, the worst phenotypes could not well be predicted. The absolute difference in both breeding values and phenotypes were very low, underlining the fact that nosemosis appeared clearly as the most problematic trait for targeted selection.

Comparison of prospective breeding values with phenotypes. Estimated Breeding values are calculated without the phenotypes of the test year and later. The test year is iterated from 2014 to 2021 and the results are averaged. Phenotypes shown are standardized to an average apiary. The bars represent the adjusted average disease phenotype shown in the left y-axis (where 0 is ill and 1 is healthy) for the four quartiles of breeding values, whose averages are represented by the line and the y-axis to the right.

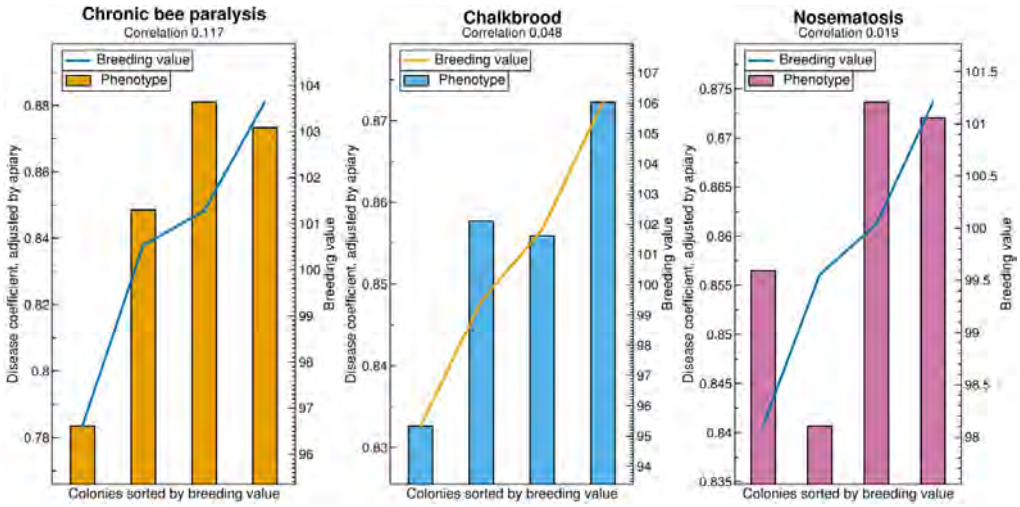


Figure 1. Year validation charts.

Discussion

As demonstrated by validations, BLUP models for disease susceptibility are a promising tool for selection for resistance. Breeding values for chalkbrood in the Carnica population (Hoppe *et al.* 2020) have been published in www.beebreed.eu for many years. Breeding progress has been slow (not shown), but chalkbrood is only a minor burden in this population. CBP has recently been a topic of intense discussion among bee breeders in Germany (Dittes *et al.* 2021) and increased awareness has led to better reporting. Here, it is shown that there is genetic variance within the Carnica population with substantial familial clustering, which suggests that CBP is a great target for resistance breeding. Compared with this, nosemosis appears to be a rather tenacious target for resistance breeding. However, this may change if awareness increases and more cases are reported.

Concerns about sustainability, feasibility and desirability of resistance breeding raised for other farm animals (Stear *et al.* 2001) must also be addressed for honeybees. The invasion of *Varroa* caused a dynamic evolutionary situation for deformed wing virus and it may also affect the sustainability of CBP resistance breeding. The feasibility of nosemosis resistance breeding is critically discussed by Villa *et al.* (2013) because, as opposed to our finding, no genetic effect was found.

References

Amiri E., Meixner M., Büchler R., and Kryger P. (2014) Viruses 6(4):1188-1201. https://doi.org/10.3390/v6031188

Aronstein, K.A., and Murray, K.D. (2010) J. Invertebr. Pathol., 103:S20-S29. https://doi.org/10.1016/j.jip.2009.06.018

Bernstein R., Plate M., Hoppe A., and Bienefeld K. (2018) J. Anim. Breed. Genet. 135:323-332. https://doi.org/10.1111/jbg.12347

Brascamp, W.E. and Bijma P. (2019) Genet. Sel. Evol. 51(1):71. https://doi.org/10.1186/s12711-019-0510-6

Dittes J., Schierling A., Aupperle-Lellbach H., Grassinger J., Mülling C. *et al.* (2021) Tierärztliche Praxis Ausgabe G. 49:326-335. https://doi.org/10.1055/a-1580-8257

Du M., Bernstein R., Hoppe A., and Bienefeld K. (2022) G3 12(2):jkab450. https://doi.org/10.1093/g3journal/jkab450

Gilliam M., Taber III S., and Richardson G. V. (1983) Apidologie 14(1):29-39.https://doi.org/10.1051/apido:19830103

Hoppe A., Du M., Bernstein R., Tiesler FK., Kärcher M. *et al.* (2020) Insects 11:768, https://doi.org/10.3390/insects11110768

Invernizzi C., Penagaricano F. and Tomasco I.H. (2009) Insectes Soc., 56(3):233-240. https://doi.org/10.1007/s00040-009-0016-2

Martín-Hernández R., Bartolomé C., Chejanovsky N., Le Conte Y., Dalmon A. *et al.* (2018) Environ. Microbiol. 20(4):1302-1329. https://doi.org/10.1111/1462-2920.14103.

Stear M.J., Bishop S.C., Mallard B.A., Raadsma H. (2001) Res. in Vet. Science, 71(1):1-7. https://doi.org/10.1053/rvsc.2001.0496

Misztal I., Tsuruta S., Strabel T., Auvray B., Druet T. *et al.* (2002). Proc of the 7th WCGALP, Montpellier, France.

Villa J.D., Bourgeois A.L. and Danka R.G. (2013) Apidologie, 44(5):511-518. https://doi.org/10.1007/s13592-013-0201-1

619. Genetic analysis of royal jelly production and behaviour traits of honeybees

B. Basso[1,2], T. Kistler[2,3], T. Gerez[4] and F. Phocas[2,3*]

[1]INRAE, Abeilles et Environnement, 84914 Avignon, France; [2]UMT PrADE, France; [3]Université Paris-Saclay, INRAE, AgroParisTech, GABI, 78350 Jouy-en-Josas, France; [4]GPGR, Agrapole, 69364 Lyon, France; florence.phocas@inrae.fr

Abstract

To improve their honeybee population, the royal jelly producers focus their selection objectives on both production and colony behaviour traits to maintain a sustainable production. To estimate genetic parameters for all these traits, a dataset from the French royal jelly producers group including records collected from 1,347 colonies was analysed using ReML methodology applied to animal models. Heritability estimates for royal jelly production traits were moderate while estimates for colony behaviour traits (gentleness, non-swarming behaviour, spring development, non-sanitary susceptibility and feed autonomy during hollow of honey flow) ranged from low estimates (0.10±0.10 for non-swarming behaviour) to very high estimates (0.68±0.10 for feed autonomy). Unfortunately, the dataset was too small to obtain reliable estimates for the genetic correlations between production and behaviour traits.

Introduction

Royal jelly production requires highly selected strains of honeybees to be profitable for beekeepers. Therefore, French royal jelly producers belonging to the 'GPGR' group conduct a collective breeding program with data recording since 2011. This very specific production requires a full-time work for a beekeeper taking care of only about a few dozen of colonies. Therefore, information to be analysed even after ten years of data collection is based on less than 1,500 colonies with performance testing, which corresponds to a very small dataset size for any analysis in quantitative genetics. Due to its very high economic value, the main trait of interest for the beekeepers is the amount of royal jelly produced per colony. However, beekeepers have also to consider various colony behaviour traits in order to maintain sustainable production conditions. First of all, the colonies being visited often, it is important that the bees are not too aggressive. The non-swarming tendency and the absence of susceptibility to disease are all the more important traits than royal jelly production relies on small size apiaries for which any colony loss has major economic impact. In addition, feed self-sufficiency becomes a major issue for colony survival and productivity in the context of climate change, the lack of floral resources being increasingly felt even during the production season. Beyond the risk of colony loss, the overall dynamics of the colonies will be lower and the production penalized for colonies with little feed autonomy. For all these reasons, it is important to develop a multicriteria breeding program considering jointly production and behaviour traits. The aim of the study was therefore to estimate genetic parameters for the main traits of interest for royal jelly producers to better orient the breeding program.

Materials & methods

Honeybee population. The French Royal Jelly Producers Group 'GPGR' is carrying out a breeding program at the national scale to select improved honeybees for their very specific production. Although data collection started in 2011, only records from years 2013 to 2021 were kept for the analysis because data have only been homogeneously recorded since 2013. Our dataset included records for 1,347 queens of which 808 were open mated and 539 were mated by artificial insemination and were potential breeding queens (BQs). Each year, a new generation of queens was produced from a pool of 9 BQs on average born the previous year. Those breeding queens came from 6 bee breeders that have been working together since

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_619, © B. Basso *et al.* 2022

2011. Each year, 8 to 18 different beekeepers have been involved in performance testing of an average of 12 BQs' daughters to select the drone-producing queens (DPQs) of the next generation. Generation interval is therefore 2 years on the sire path.

Performance testing. Queen performance test starts after the first colony wintering, when all workers are daughters of the tested queen (Büchler *et al.*, 2013). Beekeepers initiate royal jelly production for the tested queens as early as possible in the season, generally between April and May, depending on the year and the area. To do so, each hive has a queenless part in which queen cells with young larvae (stage L1) are introduced. Then, every 3 days, beekeepers harvest the royal jelly produced and introduce new queen cells with young larvae. For each colony, amount of royal jelly is recorded for the first two harvests in the season (respectively named P1 and P2) and the mean production (PROD=(P1+P2)/2) is also analysed.

Regarding behaviour traits, gentleness (GENT), non-swarming tendency (SWAR) and spring colony development (DEVE) are assess according to the COLOSS BEEBOOK recommendations (Büchler *et al.*, 2013) with a visual score ranging from 1 (worst score) to 4 (best score). In order to assess overall disease resistance (SANI) and feed autonomy (AUTO) during hollows of honey flow in the production season, beekeepers evaluate their colonies with scores ranging also from 1 to 4 almost every three days at harvest time. SANI and AUTO were analysed considering information collected the days with the worst average scores within apiary.

Genetic parameters estimation. BLUP animal genetic models were used to fit the data using BLUPf90 and its AIREMLF90 program was used to estimate genetic parameters based on the inverse of a relationship matrix derived from an R program based on Brascamp and Bijma (2014, 2019) methodology and software. To build this matrix, we assumed that each queen was mated to 8 drones bred either by a single DPQ for artificially inseminated queens or by 100 unrelated DPQs for open mated queens. Genetic parameters were initially estimated with a model considering both direct (Worker effect) and maternal (Queen effect) random genetic effects in the model describing any phenotype y_{ijk} for colony k along with the fixed effect of the comtemporary group CG_{ij} constituted by the combined environmental effects of apiary i and year j:

$y_{ijk} = \mu + CG_{ij} + Worker_k + Queen_k + e_{ijk}$, with μ the overall average and e_{ijk} the residual effect.

For all the behaviour traits, the ReML algorithm did not converge and, therefore, a simple model was used only accounting for the contemporary group fixed effect and a single random genetic effect as a colony effect combining both worker and queen effects (Du *et al.*, 2022):

$y_{ijk} = \mu + CG_{ij} + Colony_k + \varepsilon_{ijk}$, with μ the overall average and ε_{ijk} the residual effect.

For production traits (P1, P2 and PROD), heritability estimates were obtained for both the direct genetic effect (h^2_d) and the maternal genetic effect (h^2_m), as well as the genetic correlation between the two effects (r_{dm}). The simple colony effect model produced a unique estimate of heritability for a synthetic colony trait (h^2c). Genetic correlations between production and behaviour traits were calculated under this simple model.

Results

Performance statistics. Most of the queens have recorded data on all traits (Table 1). AUTO is the exception because its evaluation started in 2017 and only 733 records have been used in the analysis. During the production period considered, the average royal jelly produced was 32.9 g per harvest day and colony (Table 1).

Table 1. Performance statistics for production at first (P1), second (P2) harvests, and in average (PROD) and scores for gentleness (GENT), swarming (SWAR), sanitary status (SANI), colony development (DEVE) and feed autonomy (AUTO).

Trait	All tested queens			Open mated queens			Artificially inseminated queens		
	n	Mean	SD	n	Mean	SD	n	Mean	SD
Production traits (g)									
P1	1,293	31.0	16.2	772	27.8	15.2	521	35.7	16.4
P2	1,298	34.9	18.0	775	30.8	16.9	523	41.0	18.0
PROD	1,293	32.9	15.9	772	29.3	14.8	521	38.3	15.9
Behaviour traits (score ranging from 1 to 4)									
GENT	1,114	2.97	0.64	639	2.94	0.68	475	3.00	0.58
SWAR	1,018	3.51	0.71	593	3.45	0.80	425	3.59	0.56
SANI	1,269	3.21	0.70	755	3.09	0.68	514	3.93	0.70
DEVE	961	2.98	0.88	602	3.00	0.87	359	2.95	0.89
AUTO	733	2.03	0.94	396	2.10	0.97	337	3.11	0.51

Genetic parameters. Table 2.a shows that moderate to strong estimates of direct heritability were derived for all the production traits while estimates of maternal heritability were low and not significantly different from zero. In addition, strong negative genetic correlations between direct and maternal effects were estimated, but with extreme standard error values (ranging from 0.7 to 1.9). With a colony effect model, heritability for any production trait was moderate. The genetic correlation between P1 and P2 was strong (0.66) and their genetic correlations with PROD was not significantly different from 1 (Table 2.b).

Heritability for behaviour traits ranged from low estimates for SWAR to very high estimates for GENT and AUTO (Table 3). Apart from SWAR, all values were significantly different from zero. At the only exception of the negative estimates of the genetic correlations between DEVE and productions traits, correlations between all other colony behaviour traits and production traits could not be considered as reliable (strong difference between L1 and L2 estimates) and/or significantly different from zero.

Table 2. Genetic parameters across production traits.

a. Worker-Queen model				b. Colony model		
Trait	h^2_d	H^2_m	r_{dm}	**P1**	**P2**	**PROD**
P1	0.41±0.18	0.07±0.07	-0.65±1.90	0.32±0.11	0.66±0.34	0.88±0.33
P2	0.54±0.17	0.14±0.10	-0.66±0.74		0.41±0.10	0.94±0.14
PROD	0.52±0.17	0.12±0.09	-0.68±1.22			0.38±0.10

Table 3. Heritability of behaviour traits and genetic correlations with production traits.

	GENT	SWAR	SANI	DEVE	AUTO
Heritability	0.65±0.10	0.10±0.10	0.27±0.10	0.52±0.13	0.68±0.10
Correlation with P1	na[1]	0.31±0.86	-0.03±0.48	-0.73±0.28	-0.76±0.22
Correlation with P2	0.57±0.24	0.29±0.66	0.53±0.36	-0.58±0.39	-0.09±0.34
Correlation with PROD	0.12±0.32	0.30±1.45	0.03±0.42	-0.67±0.30	-0.31±0.33

[1] na = ReML did not converge for this parameter.

Discussion

Royal jelly production showed a relatively high heritability at the colony or direct worker genetic effect level, which confirms earlier results (Phocas *et al.*, 2017) on the same historical population and the possibility of improving production performance. Estimates of heritability for behaviour traits were less consistent across analyses since Phocas *et al.* (2017) estimated quite different values for GENT (0.40), SWAR (0.40) and SANI (0.10), but considering a different genetic model than the colony one. Nevertheless, apart from swarming tendency, our results are therefore encouraging to integrate behaviour traits in the breeding program. However, correlations between production and behaviour traits were clearly not reliable. Phocas *et al.* (2017) found a significant negative genetic correlation (-0.71±0.14) between PROD and SANI, while this correlation appeared now to be null or even positive depending on the production trait considered.

Last but not the least, it is important to notice that the estimates of genetic correlation between direct and maternal effects were strongly negative. The huge standard errors of these estimates give evidence that the structure and size of our dataset was clearly insufficient to disentangle worker and queen genetic effects in the colony performance. While it imposes caution on the true value of direct-maternal genetic correlation, it is however interesting to notice that our estimates are close to estimates derived for larger datasets on other bee populations and traits (Brascamp *et al.*, 2018). Our results confirm the ones based on simulated data from Kistler *et al.* (2022) and the importance of very large datasets to get reliable estimates of genetic parameters. Therefore, we are facing many uncertainties for implementing robust breeding programs for small honeybee populations.

References

Brascamp E.W., and Bijma P. (2014) Genet. Sel. Evol. 46:53. https://doi.org/10.1186/s12711-014-0053

Brascamp EW, and Bijma P. (2019) Proc. of the 46th Apimondia; Montreal, Canada.

Brascamp E.W., Willam A., Boigenzahn, C. Bijma P., and Veerkamp R.F. (2018) Apidologie 49: 462–463. https://doi.org/10.1007/s13592-018-0573-3

Büchler R., Andonov S., Bienefeld K., Costa C., Hatjina F., *et al.* (2013) J. Apic. Res. 52: 1–30. https://doi.org/10.3896/IBRA.1.52.1.07

Du M., Bernstein R., Hoppe A., and Bienefeld K. (2022) G3. https://doi.org/10.1093/g3journal/jkab450

Kistler T., Brascamp P., Basso B., Bijma P., and Phocas F. (2022) Proc. of the 12th WCGALP, Rotterdam, The Netherlands

Phocas F., Le Bihan C., and Basso B. (2017) EAAP 68th Annual Meeting, Tallinn, Estonia

620. Comparison of heritabilities and estimated breeding values for three honeybee traits using dam and animal models

G.E.L. Petersen[1*], F.S. Hely[1], P.F. Fennessy[1] and P.K. Dearden[2]

[1]AbacusBio Ltd, 442 Moray Place, Dunedin 9016, New Zealand; [2]Laboratory for Evolution and Development, Department of Biochemistry, University of Otago, 710 Cumberland Street, Dunedin 9016, New Zealand; gpetersen@abacusbio.co.nz

Abstract

Despite their impact on food production both in economic terms and regarding the sustainability of production, the majority of honeybee queens are currently not selected based on estimated breeding values. Although methods for the genetic evaluation of honeybees have been suggested and are in use in smaller populations across the area of endemism of the Western Honeybee (*Apis mellifera*), they have yet to penetrate into the commercial populations across e.g. the Americas and Oceania. Here we present estimates for heritabilities for three traits, honey yield, gentleness and brood in early spring, on the basis of records obtained from commercial beekeeping operations. In comparing results using an Animal and Dam Model, heritabilities for the Animal Model (0.14 for honey yield, 0.79 for gentleness and 0.004 for brood in early spring) were more closely aligned with heritabilities for these traits found in current literature.

Introduction

At the moment, the widely cultivated Western Honeybee (*Apis mellifera*) does not undergo the same rigorous genetic improvement process as other livestock species. This lack of targeted breeding exists despite the honeybee being one of the biggest contributors to the economic success and sustainability of agricultural production worldwide due to its role as the only insect pollinator which can be managed and provided to producers at a sufficient scale to satisfy pollination needs (Khalifa *et al.*, 2021). Efforts to bring honeybee breeding into the world of modern animal breeding are often driven by lone operators such as universities or government institutions, and have resulted in the development of methods for the genetic evaluation of honeybee queens (Bienefeld, Ehrhardt and Reinhardt, 2007; Brascamp and Bijma, 2014). However, these methods cannot be readily applied to commercial honeybee populations which form the backbone of large pollination-dependent production systems, even though these would benefit most from the application of genetic evaluations and are a likely target market for the provision of queen genetic evaluation as a service. One issue around the genetic evaluation of honeybee queens used as breeder queens in commercial beekeeping operations is the lack of performance records for these queens, which are usually the offspring of selected mothers, but who are managed in a way that makes the evaluation of their own performance impossible. Simultaneously, their progeny, queens that could be evaluated under field conditions, usually lack paternal pedigree and cannot be readily selected to be transferred back into the queen production part of the operation. Although both parties involved in this, the breeder queen and her queen daughters, are female, these circumstances create a situation comparable to the genetic evaluation of bulls for milk traits in their daughters. This suggests that genetic evaluation using a Dam Model could be preferable over the Animal Model previously suggested for honeybees, irrespective of whether queen and worker effects are being considered within the model.

Materials & methods

Pedigree data. The pedigree data consisted of 2,966 individuals where only maternal pedigree was available. There were 339 dams and 185 dams of dams represented in the pedigree, with the number of dam generations ranging from 1 to 17.

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_620, © G.E.L. Petersen *et al.* 2022

Performance data. Performance data from 2004 to 2021 was obtained from a specialised queen breeding operation in the South Island of New Zealand, although records were not complete for all traits due to changes in recording strategies over the years (Table 1). Through working with two commercial beekeeping operations, protocols for the performance evaluation of queen progeny in the field were developed and performance information was collected in one operation with a honey production focus and one operation with a pollination focus. Recorded traits were honey yield (kg honey produced per season per colony), brood present in early spring (at the first visit after the winter break), bees present in early spring (recorded at the same time as the brood record) and mean gentleness over the season (scored by beekeepers at each visit on a scale from 1 (aggressive) to 5 (gentle)). Gentleness was on average recorded two times per season in the commercial populations and 3.5 times per season in the queen breeding population.

Calculation of heritabilities and estimated Breeding Values (eBVs). Heritabilities were calculated and eBVs estimated using standalone ASReml (Gilmour *et al.*, 2015). A Dam and Animal model were fit with identical fixed effects, apiary and year×region. Since some individual queens had repeated records over two years, both models were run as a single record model with the later record removed and as a repeated records model. Animal model heritability was additive genetic variance divided by the sum of the additive and residual variance components or V_A / V_A + V_R. Dam model heritability was four times the additive genetic variance divided by the sum of the additive and residual variance components or $4V_A$ / V_A + V_R. The Potential maternal effects were not partitioned out in the Animal model, as this resulted in similar estimates with increased standard errors (data not shown).

Table 1. Honeybee performance records collected from 2004 to 2021.

	Queen Breeder				Commercial operations (2)			
Year	Brood spring	Bees spring	Gentle-ness	Honey yield	Brood spring	Bees spring	Gentle-ness	Honey yield
2004	71	71	71	-[1]	-	-	-	-
2005	56	56	54	-[1]	-	-	-	-
2006	64	64	55	-[1]	-	-	-	-
2007	63	63	44	22	-	-	-	-
2008	63	63	76	-[1]	-	-	-	-
2009	28	28	47	-[1]	-	-	-	-
2010	57	57	74	-[1]	-	-	-	-
2011	73	73	79	-[1]	-	-	-	-
2012	75	75	75	65	-	-	-	-
2013	48	48	49	38	-	-	-	-
2014	63	63	68	47	-	-	-	-
2015	104	106	104	83	-	-	-	-
2016	83	83	83	56	-	-	-	-
2017	69	71	71	71	-	79	199	200
2018	-	-	-	-	-	-	-	160
2019	94	94	94	-	57	56	191	-
2020	92	92	94	-	30	30	385	140
2021	-	-	20	-	-	-	208	-
Total	1,101	1,105	1,158	382	87	165	993	500

[1] Prior to 2012, honey yield had mainly been estimated using a proxy trait.

Results

Results could only be obtained for 3 out of the four recorded traits. For the trait 'bees present in early spring', all models converged, but found no genetic effect on variation. For the 3 traits honey yield, gentleness and brood in early spring, heritabilities were found to be consistent across the single record and repeated records models for the Animal Model and Dam Model, respectively. However, for both gentleness and brood in early spring, the heritability calculated with the Animal Model and Dam Model differed, with gentleness showing the biggest discrepancy between models (0.79 using the Animal Model vs 0.05 using the Dam Model; Table 2).

Discussion

The purpose of this work was to investigate the impact of the choice of model in the calculation of genetic parameters and genetic evaluation of honeybee queens. Given that honeybee pedigrees often cover only the maternal line and in our dataset, there was more data available at a dam level, it was hoped that a dam model would provide more stable heritability estimates. Heritabilities (and standard errors) were consistent between single-record and repeated record models but differed depending on the choice of Animal Model or Dam Model. The heritability for honey yield (Animal Model) of around 0.14 is consistent with previously published estimates of heritability for the queen effect on honey production (Bienefeld and Pirchner, 1990), while the heritability of gentleness (Animal Model) of 0.79 is higher than previously published estimates of around 0.16 to 0.41 (Bienefeld and Pirchner, 1990; Brascamp *et al.*, 2016). It is important to note that the assessment of gentleness was done by beekeepers during routine work and although calmness / docility was assessed at the same time, it is possible that these two traits were not separated sufficiently in these assessments. The very low heritability for brood in early spring (Animal Model) and fact that bees in early spring was found to not have a genetic contribution to variation indicates that the variability in the measurement time in spring due to weather, climatic conditions and varying beekeeper management strategies might create too much external variation to allow for a sensible analysis. This can potentially be addressed by including existing records on interventions (e.g. winter feeding) or the month or calendar week of assessment as fixed effects. Although the dataset used for the present analysis is currently still being expanded and data collection protocols are being refined, which is expected to lead to a better basis for the calculation of more reliable estimates of heritability and more accurate EBVs, one big issue in the genetic evaluation of honeybee queens remains. Due to their highly polyandrous mating habits which see one honeybee queen mated to up to 20 drones (Page Jr. and Metcalf, 1982) and mating 'on the wing' and out of the human sphere of influence, honeybee pedigree can only be observed and recorded in a one-sided, maternal-only pedigree in the absence of instrumental insemination or controlled mating in mating stations managed by breeding associations (Büchler *et al.*, 2013). Both methods are restricted in their availability and too expensive to find application in 'smallholder'-type beekeepers in developing countries and commercial beekeeping operations. However, commercial beekeepers own and manage a significant proportion of the honeybee population in developed countries, especially those with a mature beekeeping industry such as New Zealand or Canada (New Zealand Ministry for Primary Industries, 2021). As a result, absence of pedigree or dam-only pedigree is the norm for most honeybees worldwide. This is the case for

Table 2. Heritability and standard error for honey yield, gentleness and brood in early spring.

	Animal model		Dam model	
Year	Single record	Repeated records	Single record	Repeated records
Honey yield	0.14 (±0.11)	0.13 (±0.11)	0.05 (±0.07)	0.05 (±0.08)
Gentleness	0.79 (±0.07)	0.79 (±0.07)	0.05 (±0.04)	0.04 (±0.04)
Brood early	0.004 (±0.05)	0.03 (±0.05)	0.18 (±0.11)	0.29 (±0.12)

all the commercial queen records in the present dataset, with only a very small number of queens in the commercial dataset linked to the queens in the queen breeder dataset via pedigree, despite regular transfer of breeder queens from the breeding population into both commercial operations. A set of around 400 queens from both the commercial populations and the breeding population has been sampled for Whole-Genome Sequencing, the data from which is still being processed. The resulting genomic relationship matrix will be used to investigate the possibility of genomic selection using commercial 'production' queen performance records to select breeding stock both in the commercial and breeding populations. While the possibility of using genomic selection in honeybees has been suggested with the use of modelling studies (Bernstein *et al.*, 2021), the potential for reaching large commercial operations running comparatively genetically homogenous populations is currently poorly understood and to our knowledge, this will be the first attempt to improve selection decisions in bigger populations with the use of genomics.

References

Bernstein R., Du M., Hoppe A., Bienefeld K. (2021) GSE 53(1):1-17. https://doi.org/10.1186/s12711-021-00654-x

Bienefeld K., and Pirchner F. (1990) Apidologie 21:175–183. https://doi.org/10.1051/apido:19900302.

Bienefeld K., Ehrhardt K., and Reinhardt F. (2007) Apidologie 38(1):77–85. https://doi.org/10.1051/apido:2006050.

Brascamp E.W., and Bijma P. (2014) Genet. Sel. Evol. 46(1):1–15. https://doi.org/10.1186/s12711-014-0053-9.

Brascamp E.W., Willam A., Boigenzahn C., Bijma P., Veerkamp, R.F. (2016) Apidologie 47(6):739–748. https://doi.org/10.1007/s13592-016-0427-9.

Brascamp E.W., Wanders T.H.V., Wientjes, Y.C. J., Bijma P. (2018) Proc. of the 11th WCGALP, Auckland, New Zealand

Büchler R., Andonov S., Bienefeld K., Costa C., Hatjina F. *et al.* (2013) J. Apic. Res. 52(1):1–30. https://doi.org/10.3896/IBRA.1.52.1.07.

Gilmour A., Gogel B.J., Cullis B.R., Welham S.J., Thompson R. (2015) ASReml User Guide Release 4.1. VSN International, Hemel Hempstead, UK.

Khalifa S. A. M., Elshafiey E.H., Shetaia A.A., El-Wahed A.A.A., Algethami A.F. *et al.* (2021) Insects 12(8):688. https://doi.org/10.3390/insects12080688.

New Zealand Ministry for Primary Industries (2021) *Farm Monitoring Report - Apiculture Report 2020*. Available at: https://www.mpi.govt.nz/.

Page Jr. R. E. and Metcalf R. A. (1982) Am. Nat. 119(2):263–281.

621. Statistics for an accurate genome wide association study on *Varroa* resistance trait in a French honeybee

S.E. Eynard[1,2*], A. Vignal[2,3], Y. Agez[1], B. Basso[3,4], O. Bouchez[5], T. Bulach[5], Y. Le Conte[3,4], B. Dainat[6], A. Decourtye[3,7], L. Genestout[1], M. Guichard[6,8], F. Guillaume[9], E. Labarthe[2], R. Mahla[1], F. Mondet[3,4], M. Neudischko[8], F. Phocas[3,10], Y. Poquet[1], C. Sann[1], R.F. Serre[5], K. Tabet[2] and B. Servin[2,3]

[1]Labogena DNA, 78353 Jouy-en-Josas, France; [2]GenPhySe, Université de Toulouse, INRAE, ENVT, 31320 Castanet-Tolosan, France; [3]UMT PrADE, France; [4]INRAE, Abeilles et Environnement, 84914 Avignon, France; [5]INRAE, GeT-PlaGe, Genotoul INRAE, 31326 Castanet-Tolosan, France; [6]Agroscope, 3003 Bern, Switzerland; [7]ITSAP, 84914 Avignon, France; [8]Agroscope, 1725 Posieux, Switzerland; [9]Evolution XY, 35538 Noyal-Sur-Vilaine, France; [10]Université Paris-Saclay, INRAE, AgroParisTech, GABI, 78352 Jouy-en-Josas, France; sonia.eynard@labogena.fr

Abstract

The honeybee *Apis mellifera*, major pollinator worldwide, is facing multiple threats to survival with the infestation by *Varroa destructor* being a major concern. Resistance to such infestation is a crucial mechanism for the maintenance of the honeybee population. However, to date, genetic components explaining this resistance are poorly known making selection for resistant honeybees a challenge. In this study we perform a genome wide association study on about 300 *A.m. mellifera* colonies for one trait measuring varroa resistance, the recapping of varroa infested brood cells. On about 300 *A.m. mellifera* colonies we identified multiple chromosome regions showing significant effects for this polygenic trait. We reported a heritability of 0.57 (±0.21). Recapping of varroa infested brood cells could be a trait of interest for the selection of resistant honeybees.

Introduction

The honeybee, *Apis mellifera*, is the main domestic pollinator worldwide. Despite its critical value for pollination services, production of honey, royal jelly and pollen, it is currently facing considerable challenges due to land anthropication (loss of flower resources and suitable habitats, pesticide use), climate change and the spread of biological threats. The ectoparasite *Varroa destructor* is one of these threats, as it is the principal cause of honeybee colony mortality in the Western world (Le Conte *et al.* 2010). For about a decade, resistance to varroa has become a key trait of interest for beekeepers. Colonies naturally resistant to varroa infestation have been identified by natural selection when untreated colonies were left to die or survive in isolated environments (Mondet *et al.* 2020). However, the genetic mechanisms to explain the variation in varroa infestation resistance observed across honeybee colonies is poorly known. Only a few genetic markers of interest have been identified in small honeybee populations (Mondet *et al.* 2020). Moreover, it is difficult to define a relevant measure to score resistance to a parasite and estimate its genetic parameters. Multiple traits have been described, for example mite none reproduction (MNR), varroa sensitive hygiene (VSH), hygienic behaviour, recapping, and very variable genetic parameters were estimated (Guichard *et al.* 2020). Identifying genetic markers linked to resistance to varroa infestation is key to understanding the genetics behind this resistance and in the long run, to perform genomic selection for this trait. In this study, about 300 French colonies of the *A.m. mellifera* genetic background, endemic to Western Europe, were sampled for genetic material and traits expected to be linked to resistance to varroa infestation and we present results from recapping of varroa infested brood cells by adult bees. A pool of workers from each colony was sequenced, providing sequencing depth and allele counts giving allele frequencies within the colony rather than traditional genotype information. In the context of eusocial insects, a method was developed to reconstruct queen genotype from such pool data (Eynard *et al.* 2021). We applied this method to reconstruct honeybee queen genotypes for each of the colonies which were

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_621, © S.E. Eynard *et al.* 2022

used in a Genome Wide Association Studies (GWAS) for the trait of interest. Its heritability was estimated and markers and genome regions with significant effects were identified. These results contribute to the knowledge on honeybee resistance to varroa infestation and will be valuable for future implementation of genomic selection.

Materials & methods

Genetic material and sequencing. A total of 306 honeybee colonies, with a major *A.m. mellifera* genetic background, coming from beekeepers in France were sampled within the framework of the FranceAgriMer, Investissement d'Avenir BeeStrong project. Each genetic sample consisted of about 500 honeybee workers. An initial DNA extraction, using tailor made protocols, was performed. Thereafter, pool sequencing was done in order to obtain 30X raw sequencing depth. These reads were aligned on the honeybee reference genome Amel HAV3.1 (Wallberg *et al.* 2019), using BWA-MEM (Li 2013). Pool sequences were analysed using Samtools mpileup (Li *et al.* 2009) and Popoolation2 (Kofler *et al.* 2011) on 6,914,704 single nucleotide polymorphisms identified in Wragg *et al.* (2021). Markers were filtered for minor allele frequency above 0.01 and missing rate below 5% for further analysis.

Queen genotype reconstruction. Honeybee queen genotypes for all markers and each colony were reconstructed from sequencing depth and allele counts data using the method of Eynard *et al.* (2021). In short, the method is based on the likelihood of the queen genotype that can be written as:

$$x_l^c | d_l^c, f_l^c, g_l^c \sim Binomial(\frac{f_l^c + g_l^c}{2}, d_l^c) \quad (1)$$

where g_l^c is the unknown honeybee queen genotype, f_l^c is the unknown reference allele frequency in the males founder of the colony and d_l^c and x_l^c the sequencing depth and allele counts obtained from pool sequencing experiments for locus l and colony c. Assuming all colonies come from an homogeneous population, Equation 1 can be expressed as a function of a common allele frequency f_l which can be estimated by maximum likelihood and used to compute posterior probabilities of the queen genotype.

Measure of resistance to varroa infestation. One way for honeybee colonies to control *Varroa* infestation is by recapping infested brood cells as this behaviour breaks the reproductive cycle of the mite. Recapping status was thus proposed as a measure of varroa resistance (Büchler *et al.* 2017) and therefore the proportion of recapped brood cells that are infested by varroa was used as a resistance trait in this study.

Genome wide association study. We performed a GWAS using a multivariate linear model as proposed in GEMMA (Zhou and Stephens 2012). The relationship matrix between colonies was estimated on allele frequencies for 2,387,893 markers from the pool sequencing experiment, taking SNP linkage disequilibrium into account using LDAK Speed *et al.* (2012). The association of the reconstructed queen genotypes with the phenotype was tested for 2,694,182 markers. Markers' *P*-values were obtained by GEMMA and q-values and s-values were estimated using the R package 'ashr' (Stephens 2017). SNPs with a s-value less than 0.1 were considered significant.

Results

The proportion of phenotypic variance explained (pve) by the markers, a proxy for heritability, for recapping of varroa infested brood cells was 0.57 (95% confidence interval [0.16, 0.98]). The GWAS inflation factor was 1, indicating an accurate correction for population stratification (Figure 1).

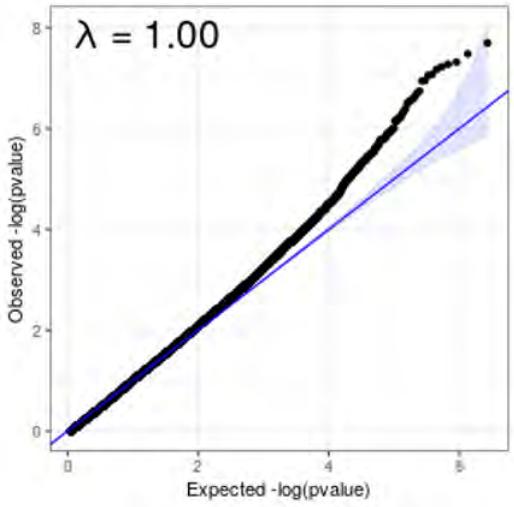


Figure 1. Quantile-quantile (QQ) plot of observed versus expected *P*-values for the GWAS with inflation factor, lambda.

We identified 18 chromosome regions significantly supported by at least one marker for both standard *P*-values and after adaptive shrinkage (Figure 2). These regions were found on 13 of the 16 chromosomes.

Discussion

Resistance to varroa infestation in honeybees is a complex trait to measure and thus far there is no consensus way to accurately measure it. One way to infer varroa resistance in a honeybee colony is to estimate the ratio of varroa infested brood cells that have been recapped. This is under the assumption that recapping of an infested brood cell by worker bees will cause a disturbance in the *Varroa destructor* reproductive cycle within the sealed brood cells and thus reducing its pressure on the colony by impairing its development (Oddie *et al.* 2018, Oddie *et al.* 2021). This trait relies highly on the initial infestation status, meaning that a colony with a low infestation rate will have a less accurate estimate of recapping than a

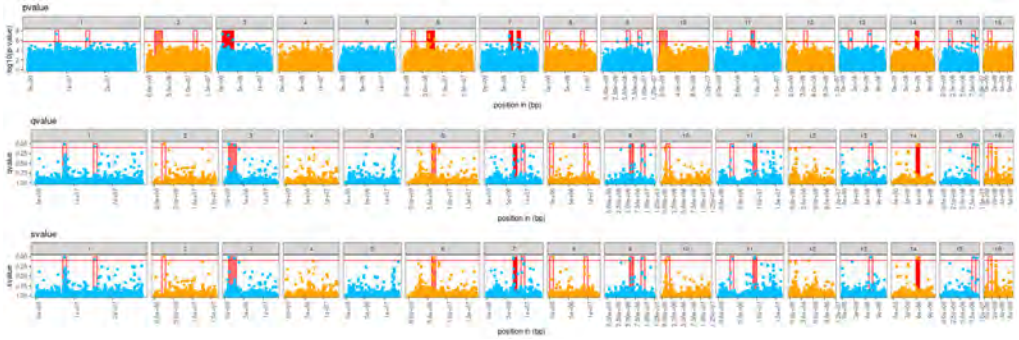


Figure 2. Manhattan plot of the GWAS results for -log10(*P*-value), q-value and s-value, estimated by adaptive shrinkage, along each of the chromosomes. Regions containing significant markers are highlighted in red, brightness of red being linked to the number of significant markers in the region. The red line marks the significance threshold.

highly infested colony, as the ratio will be based on a smaller number of observed brood cells. This causes a high variability between measures that need to be accounted for, which is done in our case by using an Empirical Bayes transformation giving weights to the phenotypes that are proportional to their estimated accuracy. Additionally, a logit transformation was needed to linearise the measured ratio and allow the use of this trait for GWAS. In our study, we identified 18 chromosome regions that appear to have a significant effect on the trait. Most chromosomes carry a region of interest meaning that the recapping of infested brood cells is a highly polygenic trait. Our results contrast with other studies in which a small number of significant markers have been observed (Guichard *et al.* 2021). We could estimate the heritability, as the phenotypic variability explained by our markers, for the trait to be 0.57 (±0.42). Although still imprecise, this estimate of heritability shows the potential there is to use this trait in a breeding program. It would now be interesting to identify genes underlying regions that appear to have a significant effect on the recapping of varroa infested brood cells in order to unravel the biological mechanisms explaining this behaviour in worker honeybees.

References

Büchler R., Costa C., Mondet F., Kezic N. and Kovacic M. (2017) Screening for low Varroa mite reproduction (SMR) and recapping in European honey bees.

Eynard S.E., Vignal A., Basso B., Le Conte Y., Decourtye A. *et al.* (2021) BioRxiv https://doi.org/10.1101/2021.11.08.467442

Guichard M., Dietemann V., Neuditschko M. and Dainat B. (2020) Genet Sel Evol 52:71 https://doi.org/10.1186/s12711-020-00591-1

Guichard M., Dainat B., Eynard S., Vignal A., Servin B. *et al.* (2021) Anim Genet https://doi.org/10.1111/age.13150

Kofler R., Pandey R.V. and Schlötterer C. (2011) Bioinformatics 27(24):3435-3436 https://doi.org/10.1093/bioinformatics/btr589

Le Conte Y., Ellis M. and Ritter W. (2010) Apidologie 41(3):353-363 https://doi.org/10.1051/apido/2010017

Li H., Handsaker B. Wysoker A., Fennell T., Ruan J. et al (2009) Bioinformatics 25:2078-2079 https://doi.org/10.1093/bioinformatics/btp352

Li H. (2013) ArXiv http://arxiv.org/abs/1303.3997

Mondet F., Beaurepaire A., McAfee A., Locke B., Alaux C. et al (2020) Int J for Parasit 50(6-7):433-437 https://doi.org/10.1016/j.ijpara.2020.03.005

Oddie M.A.Y., Büchler R., Dahle B., Le Conte Y., Locke B. et al (2018) Sci Rep 8:7704 https://doi.org/10.1038/s41598-018-26001-7

Oddie M.A.Y., Burke A., Dahle B., Le Conte Y., Mondet F. et al (2021) Sci Rep 11:9133 https://doi.org/10.1038/s41598-021-88592-y

Speed D., Hemani G., Johnson M.R. and Balding D. (2012) AJHG 91(6):1011-1021 https://doi.org/10.1016/j.ajhg.2012.10.010

R Core Team, R: A language and environment for statistical computing 2021. Available at: https://www.R-project.org/

Stephens M (2017) Biostat 18(2):275-294. https://doi.org/10.1093/biostatistics/kxw041

Wallberg A., Bunikis I., Pettersson O., Mosbech M., Childers A. et al (2019) BMC Genomics 20:275 https://doi.org/10.1186/s12864-019-5642-0

Wragg D., Eynard S., Basso B., Canale-Tabet K., Labarthe E. *et al.* (2021) BioRxiv https://doi.org/10.1101/2021.09.20.460798

Zhou X. and Stephens M. (2012) Nat Genet 44:821-824. https://doi.org/10.1038/ng.2310

622. New approach to identify Mendelian inconsistencies between SNP and pedigree information in the honey bee

R. Bernstein*, M. Du, A. Hoppe and K. Bienefeld

Institute for Bee Research Hohen Neuendorf, Friedrich-Engels-Str. 32, 16540 Hohen Neuendorf, Germany; richard.bernstein@hu-berlin.de

Abstract

In preparation of genomic selection, the genomic data is compared to pedigree information, and samples causing Mendelian inconsistencies are removed. Traditionally, opposing homozygotes and coefficients of inbreeding and relationships were used for this purpose. Here, we adapted these methods to the peculiar mating biology in honey bees. We developed a rule system for the removal of Mendelian inconsistencies in honey bees based on the empirical distributions of opposing homozygotes and coefficients of inbreeding and relationship in real and simulated data. To initiate genomic selection in the honey bee, 2,970 phenotyped queens were recently genotyped for about 100,000 markers. The samples of 207 and 135 queens were removed by the checks for opposing homozygotes and abnormalities in the relationship matrices, respectively. We suggest, that our approach should be used to prepare data for genomic selection in the honey bee.

Introduction

Genomic selection is a standard practice in many agricultural species, but not in honey bees, and methods require careful adaptation to the peculiarities of this species. Honey bees are haplodiploid, i.e. the female queens and workers are diploid and the male drones are haploid. A colony consists of several thousand workers which perform all non-reproductive tasks, and a fertilized queen which produces workers, queens, and drones. Virgin queens undertake nuptial flights to mate with several drones. To ensure lasting breeding success, the genetic quality of the drones must be controlled (Plate *et al.* 2019). Therefore, mating stations are used, where only drones of specialized drone producing queens (DPQs) are present. For unrelated DPQs, daughters of the same dam queen (full-sibs) can descend from the same drone (super-sisters), different drones but the same DPQ, or different DPQs, resulting in genetic relationships of 0.75, 0.5, and 0.25, respectively, a weighted average of which (0.32) is assumed in pedigree-based analyses. However, the DPQs on a mating station are typically full-sibs, leading to slightly closer relationships among their offspring. Recently, a marker array for 100,000 SNPs has been developed and used to genotype 2970 phenotyped queens to initiate genomic selection in the honey bee (Jones *et al.* 2020) with promising validation results for genomic breeding values (Bernstein *et al.*, submitted). Here, we focus on the quality control (QC) comparing the genotypes to the pedigree. In diploid species, the most common method to spot Mendelian inconsistencies checks for opposing homozygotes (OHs; Hayes 2011), because it is very rare that the detected genotypes of parent and offspring are homozygous for different alleles at the same locus. For the same purpose, comparisons of the pedigree-based to the genomic relationship matrix can be applied, which can also be applied to check siblings for Mendelian inconsistencies (Calus *et al.* 2011).

Materials & methods

Data. From the 2,970 genotyped queens in (Jones *et al.* 2020), 312 queens were removed since less than 90% of all loci were called in their samples. SNPs were removed, when the SNP was called in less than 90% of all samples, had a minor allele frequency (MAF) below 0.01, or showed significant deviations from Hardy-Weinberg-equilibrium after Bonferroni-correction (P-value$<0.05\times10^{-5}$). 63,240 SNPs remained in the data set.

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_622, © R. Bernstein *et al.* 2022

Quality control based on opposing homozygotes. The number of OHs was calculated for all pairs of genotyped queens. Pedigree data maintained at www.beebreed.eu (Hoppe *et al.* 2020) was downloaded on 14th Feb 2021. For pairs of dam and daughter, full-sibs, and unrelated queens, the empirical distributions of the number of OHs were considered. The appropriate threshold for QC should be higher than the frequent numbers of OHs in dam-daughter-pairs, but lower than the frequent numbers of OHs in pairs of unrelated queens, and be in between two peaks for pairs of full-sibs. In our data set, a threshold of 50 OHs satisfies these conditions (see the results section). Mendelian inconsistencies and confirmations between the pedigree and the genomic data were counted in the following way: Two samples of the same queen were counted as an inconsistency, when the number of OHs was above 50. A pair of dam and daughter counted as a confirmation when the number of OHs was below 50, otherwise as an inconsistency. Full-sibs were excluded from this test (see Figure 1 C). Other sample pairs counted as inconsistencies when the number of OHs was below 50. For each sample an inconsistency score was calculated by subtracting four times the number of confirmations from the number of inconsistencies. The sample of highest score was removed, and the scores were re-evaluated. The procedure was iterated until no sample with a positive score was left. The factor four is based on the heuristic that a dam should not be removed unless 80% of her genotyped offspring had a conflicting genotype (Calus *et al.* 2011).

Quality control based on relationship matrices. After the QC based on OHs, further steps of QC followed. The pedigree-based relationship matrix A_{22} between genotyped queens was calculated using the software PInCo (Bernstein *et al.* 2018). The genomic relationship matrix G was calculated according to Method 1 in (VanRaden 2008) using preGSf90 (Misztal *et al.* 2002). To develop a rule system for the matrix-based consistency check, matrices A_{22} and G were calculated for a simulated population of 10,000 genotyped queens, distributed over 10 years, and selected for a trait of moderate heritability (Bernstein *et al.* 2021). Based on comparisons of the simulated genomic and pedigree relationships, combinations of relationship values were defined to constitute inconsistencies or confirmations. The values of A_{22} and G for two samples i and j, denoted by a_{ij} and g_{ij}, respectively, counted as an inconsistency, if:

$$(a_{ij}<0.25 \;\&\; g_{ij}>0.5) \text{ or } (a_{ij}>0.25 \;\&\; g_{ij}<0.0625) \text{ or } (a_{ij}\geq 0.5 \;\&\; g_{ij}<0.125) \text{ or } |g_{ij}- a_{ij}|>0.5, \quad (1)$$

and as a confirmation, if:

$$a_{ij}>0.25 \;\&\; g_{ij}>0.25 \;\&\; |g_{ij}-a_{ij}|<0.5 \quad (2)$$

In a QC step for high coefficients of inbreeding, queens with diagonal entries in G>1.5 were removed. Subsequently, a score based on inconsistencies and confirmations was used in an iterative procedure as described in the subsection on the QC for OHs.

Results

Based on the empirical distribution of the number of OHs (Figure 1), the threshold of 50 was chosen, as 64% of parent-daughter-pairs, 10% of the pairs of full-sibs, and 0.1% of the pairs of unrelated queens showed 50 or less OHs, respectively. 207 queens were removed due to the QC based on OHs.

The samples were further controlled based on the empirical distributions of A_{22} and G (Figure 2). The simulated data set indicated that no QC is possible in the area ($A_{22}<0.25$ & $G<0.5$). This is due to the possibly high discrepancy between genomic and pedigree-based relationships in full-sibs, which is inherited by offspring. E.g. queens with pedigree-based relationships of 0.32 showed genomic relationships between 0.17 and 0.83. The empirical distribution of A_{22} and G in the real data were even further spread out. Possible causes include that the SNP array includes a large number of SNPs with MAF below 0.05, and diverging

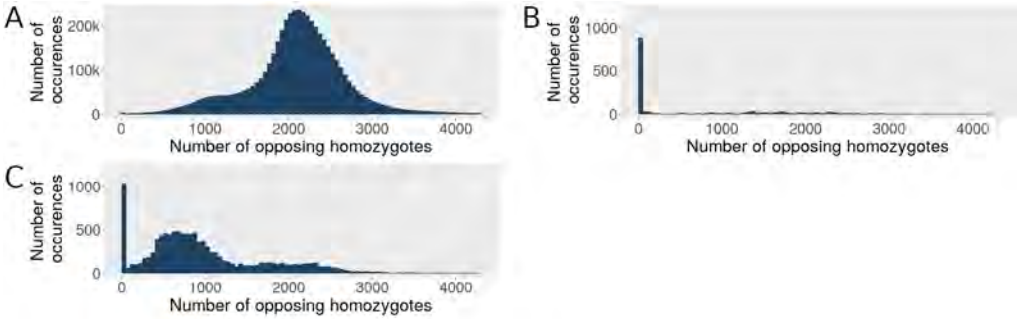


Figure 1. The number of OHs was calculated for all pairs of genotyped queens (A), dam and daughter (B), and full-sibs (C). 14,856 occurrences in (A) and 6 occurrences in (C) are not shown, as the corresponding numbers of OHs exceeded 4,300. The width of each bar covers 50 numbers of OHs.

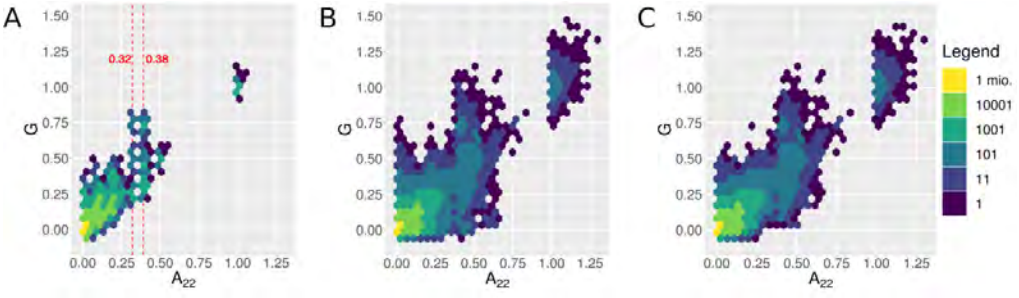


Figure 2. Comparison of the distributions of the pedigree-based and genomic relationship matrices A_{22} and G, respectively. (A) shows the matrices in a simulated data set, (B) shows the matrices after samples with high coefficients of inbreeding were removed and (C) shows the matrices after the QC was finished. Since the complete matrices are shown, coefficients of inbreeding are shown at $A_{22} \geq 1.0$. In (A), the red lines mark full-sibs sired by unrelated DPQs, and DPQs which are non-inbred full-sibs, respectively. At $A_{22}=0.32$, peaks are visible for super-sisters (G=0.75), daughters of the same DPQ (G=0.5), and of different DPQs (G=0.25). The full-sibs in (B) and (C) are visible at $A_{22} \geq 0.38$, as the corresponding DPQs were full-sibs and their offspring in some cases inbred.

breeding lines or regions. Equation 1 targets the areas ($A_{22}<0.25$ & $G>0.5$), ($A_{22}>0.25$ & $G<0.0625$), and ($A_{22} \geq 0.5$ & $G<0.125$) for QC, because they were unoccupied in the simulations, and genotyping errors are likely to show in areas were the values of A_{22} and G strongly diverge. The number of occurrences in these areas was reduced by 87% by the QC based on relationship matrices.

Nine and 135 queens were removed by the QC step for high coefficients of inbreeding, and the following iterative procedure based on a score, respectively. The quality controlled, final data set contained 2,307 queens. A stricter control is possible, but also more likely to remove correct genotypes.

As expected, the QC based on the comparison of A_{22} and G increased their similarity (Table 1). However, the simulated matrices showed considerably lower spreads of the coefficients of inbreeding and relationship. The low correlation of the coefficients of inbreeding in the simulated matrices compared to the matrices in the real data set is due to the absence of high coefficients of inbreeding in the simulated data. Based on the maximum absolute difference between A_{22} and G, the condition $|G-A_{22}|>0.5$ was included in Equation 1 to mark Mendelian inconsistency.

Table 1. Quantitative comparison of the pedigree-based and genomic relationship matrices A_{22} and G, respectively.

Version of A_{22} and G[1]	(A)	(B)	(C)
Correlation of inbreeding coefficients in A_{22} and G	0.11	0.32	0.33
Correlation of coefficients of relationship in A_{22} and G	0.63	0.49	0.51
Maximum absolute difference $\lvert A_{22} - G \rvert$	0.51	0.74	0.67

[1] (A) = Simulated matrices, (B) = matrices after correction for high inbreeding coefficients, (C) = final matrices.

Discussion

In honey bees, super-sisters have similar numbers of OHs like pairs of parent and offspring. Therefore, OHs can be used to refute parentship, but not for parentage assignment, which is possible in diploid species, if monozygotic twins can be excluded (Hayes 2011). We employed a boundary of 0.5 on the absolute difference between genomic and pedigree-based relationship, and identified three areas where Mendelian inconsistencies are likely to occur. A strict boundary of 0.2 on the absolute difference between genomic and pedigree-based relationship was employed in a real data set (Calus *et al.* 2011). Such a narrow boundary is inappropriate for honey bees, as the discrepancy between pedigree-based and genomic relationships of full-sibs must be considered in honey bees. In real data sets, unsafe mating stations can exacerbate the difference between genomic and pedigree-based values. Therefore, genomic coefficients of inbreeding are considerably more reliable than pedigree-based inbreeding coefficients for honey bees. Thus, the discrepancy between the values of A_{22} and G could be an advantage for genomic selection in the honey bee. Unknown paternity reduces the accuracy of pedigree-based selection, but is countered when genomic data is available (Bernstein *et al.* 2021).

References

Bernstein R., Plate M., Hoppe A., and Bienefeld K. (2018) J. Anim. Breed. Genet. 135:323–332. https://doi.org/10.1111/jbg.12347

Bernstein R., Du M., Hoppe A., and Bienefeld K. (2021) Genet. Sel. Evol. 53:64. https://doi.org/10.1186/s12711-021-00654-x

Calus M.P.L., Mulder H.A., and Bastiaansen J.W.M. (2011) Genet. Sel. Evol. 43:34. https://doi.org/10.1186/1297-9686-43-34

Hayes B.J. (2011) J. Dairy Sci. 94:2114-2117. https://doi.org/10.3168/jds.2010-3896

Hoppe A., Du M., Bernstein R., Tiesler F.K. *et al.* (2020) Insects 11:768. https://doi.org/10.3390/insects11110768

Jones J.C., Du Z.G., Bernstein R., Meyer M. *et al.* (2020) Ecol. Evol. 10:6246–6256. https://doi.org/10.1002/ece3.6357

Misztal I., Tsuruta S., Strabel T., Auvray B. *et al.* (2002) Proc. of the 7th WCGALP, Montpellier, France.

Plate M., Bernstein R., Hoppe A., and Bienefeld K. (2019) Genet. Sel. Evol. 51:74. https://doi.org/10.1186/s12711-019-0518-y

VanRaden P.M. (2008) J. Dairy Sci. 91: 4414-4423. https://doi.org/10.3168/jds.2007-0980

623. Investigation of parental meiotic contribution in honeybees (*Apis mellifera*)

G.E.L. Petersen[1*], A. Heywood[2], A.C.L. Walters[1], P.F. Fennessy[1] and P.K. Dearden[2]

[1]AbacusBio Ltd, 442 Moray Place, Dunedin 9016, New Zealand; [2]Laboratory for Evolution and Development, Department of Biochemistry, University of Otago, 710 Cumberland Street, Dunedin 9016, New Zealand; gpetersen@abacusbio.co.nz

Abstract

Honeybee pedigree cannot be readily recorded based on the observation of mating, since this is hindered by their reproductive behaviours, including mating in flight and extreme polyandry. Due to these circumstances, genotyping of queens with partial pedigree with the maternal line only or entirely unknown pedigree could be a valuable tool to link queens observed under field conditions back to breeding stock. However, queen management in beekeeping operations often leads to the grand-daughters of breeder queens being placed in the field, which requires pedigree reconstruction beyond dam or sire assignment. Here we investigate parental contribution to 64 queens produced in a closed mating scheme across five generations. Differences in the resulting relationships between paternal sisters as well as the relationships of sets of paternal sisters to their dam suggest variation in the paternal contribution, possibly due to the effects of the purging of deleterious haplotypes in the haploid male drone.

Introduction

Honeybees (*Apis mellifera*) differ from other livestock species in their reproduction dynamics, which impacts their patterns of inheritance. Not only do these eusocial insects display reproductive division of labour, wherein reproduction is limited to one female per colony (the queen), but also haplodiploidy, where female workers and queens are diploid and male drones are haploid. As a result, supersisters, who share both parents, are related by 0.75, while full sisters, who share a mother and whose sires come from the same source, are related by 0.5. Colony genetic diversity is maintained by extreme polyandry practised by the queen, which results in groups of full sisters coexisting with each other.

To estimate breeding values in honeybees using mixed model methodology, the relationships between individual queens need to be defined, either via recorded pedigree, which comes with its own challenges, or genotyping. Honeybee-specific methods have been defined for the efficient estimation of breeding values and inbreeding coefficients (Brascamp and Bijma, 2014, Bernstein *et al.*, 2018) based on pedigree. However, although cost-effective options for the genotyping of honeybees are available (Petersen *et al.*, 2020), the resulting relationship matrices cannot easily be transcribed into pedigree due to the skew in (sister) relationships (unpublished). This becomes particularly challenging when there is no pedigree information available for queens involved in field tests, which is often the case in commercial beekeeping operations.

In order to attempt definition of (deep) pedigree from genomic relationships, the scale of Mendelian sampling (i.e. the meiotic contribution to the genotype of a particular queen from each parent) needs to be understood. Here we present an experiment based on managed mating of honeybee queens through Artificial Insemination (AI) over five generations to investigate the contribution of the great-grand-dam to the last generation.

Materials & methods

Honeybee queens and drones. In cooperation with a commercial beekeeping business, maternal lineages were established based on Progenitor Queens, designated generation G0, in the summer of 2018

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_623, © G.E.L. Petersen *et al.* 2022

in Dunedin, New Zealand. Individual queen progeny from each Progenitor Queen were inseminated with semen from sets of drone brothers to guarantee sufficient genetic diversity within each colony to maintain colonies capable of rearing drones for the next generation (as opposed to single-drone insemination). Different drone producing queens were used for each generation (Table 1). While the original design saw all generations reared in a single beekeeping season, due to some unforeseen issues around drone rearing and weather limitations, G1 and G2 were produced and inseminated in the season 2019/20, while the rearing and insemination of G3 and G4 queens had to be delayed into the spring of 2020. All mating within the experiment was performed via AI, with drone source queens carefully selected and subsequently harvested to be sequenced. Drones were confined within the hive to guarantee their genetic origin. All queen dams were harvested at the end of their generation for genotyping (Table 1). In addition to queens, all drones used for AI were also collected to enable the determination of individual drone sires at a later date.

DNA extraction, library preparation and sequencing. DNA was extracted from individual honeybee heads. Heads were homogenised at room temperature using four 3.2 mm stainless steel ball bearings and 500µl of 1xPBS solution for two passages of 30-45s each at 20Hz. The homogenised tissue solution was then centrifuged at 4,000rpm for 10 minutes and about 150 mg of the pellet transferred to 96 deep well plate for lysis treatment. Lysis and DNA extraction were performed using the Zymo Quick-DNA Magbead Plus Kit (Zymo Research, Irvine, USA) on an ep*Motion*® 5075 (Eppendorf, Cologne, Germany). 200 ng of genomic DNA from each sample was used for library preparation using NEBNext Ultra II FS DNA kit (New England Biolabs (NEB), Ipswich, USA) at ⅓ volume with NEBNext Unique Dual Index Primer Pairs (NEB). The libraries were subsequently generated according to their respective protocols and sequenced on either a HiSeq X or NovaSeq 6000 platform (Illumina, San Diego, USA) to generate 2×200 bp paired-end libraries with ~15× read depth.

DNA quality analysis and processing. Sequence quality was checked by FastQC v0.11.8 (Andrews, 2015). Adaptors and low-quality sequences were trimmed using AdapterRemoval v2.3.2 (Schubert *et al.*, 2016). Trimmed reads were mapped using BWA-mem2 v2.2 (Li and Durbin, 2009). Mapped reads were sorted and duplicates marked using SAMtools v1.14 (Li *et al.*, 2009) and variants called using GATK4 (Van der Auwera and O'Connor, 2020). Variants for all queens in each line were combined using GLnexus v1.4.1 (Yun *et al.*, 2020) and Genetic Relationship Matrices (GRMs) were built using PLINK v1.90b6.21 (Purcell *et al.*, 2007).

Analysis of queen relationships. The relationships within each honeybee queen line were assessed by building GRMs that were based on genetic variants identified after mapping whole-genome sequencing reads to the reference genome (ENA Assembly: GCA_003314205.1). Distances between related queens were expressed as allele counts and in doing so the relative inheritance between mother / daughter queens was inferred.

Table 1. Honeybee queens collected throughout the crossing experiment.

	Queen dams	Drone sources	
Generation	No of dams per generation	No of drone sources per generation	Total queens genotyped
G0	5	0	5
G1	5	5	10
G2	6	4	10
G3	15	4	19
G4	18	3	20
Total	49	15	64

Results

Three of the five maternal lines in the dataset are represented in Figure 1 with lines P and W showing identical pedigree structure. Examination of differences in allele counts between G0–G1 across all lines showed varying levels of inheritance. However, with each generation there was a consistent trend of increasing distance from the Progenitor Queen (Figure 1A, 1B, 1D).

Comparison of identical pedigree structures between lines P an W showed that paternal sisters with the same grand-dam may be more closely related (Figure 1A) maternal sisters and that paternal sisters may be more related to their dam (Figure 1B) than maternal sisters, suggesting differences in meiotic contribution between the male and female line. With each generation, differences in genetic relatedness to the maternal line were observed in line R (Figure 1E), showing the importance in identifying specific genomic regions that may play a role in differences in genetic contribution.

Discussion and outlook

The goal of the work presented here was to gain a better understanding of Mendelian sampling in honeybees, with the intention of building a steppingstone to the development of a method to reconstruct deep pedigree based on genomic relationships in honeybees. The contribution of the paternal / sire line has

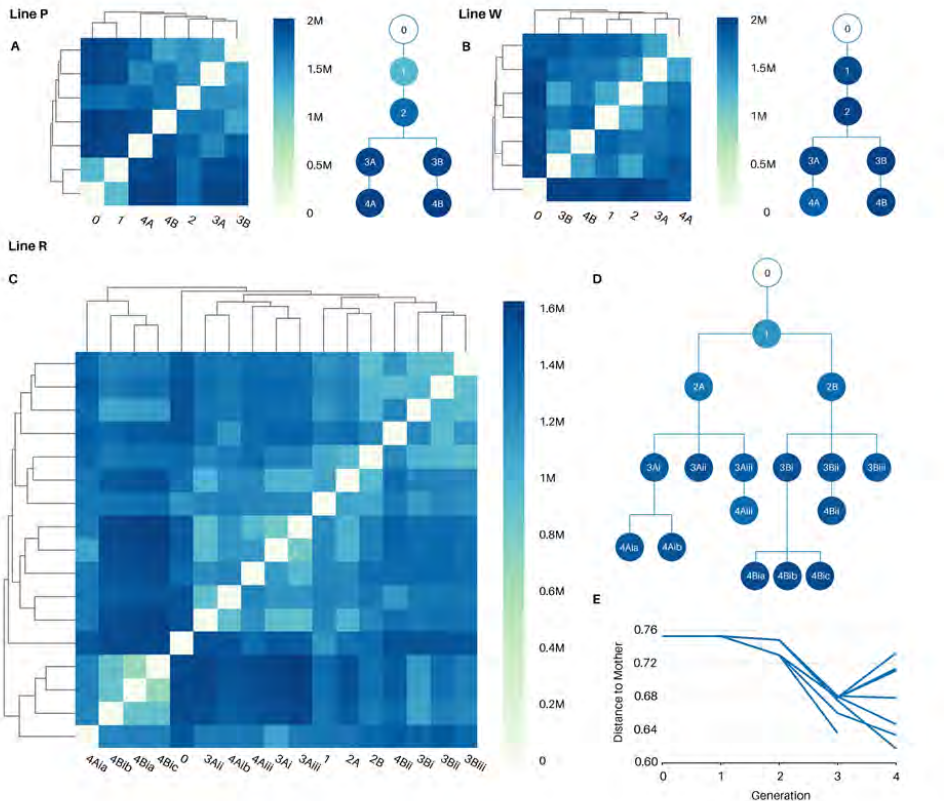


Figure 1. Relationships within honeybee lines across generations G0 to G4. (A) Heatmap and pedigree structure of line P. (B) Heatmap and pedigree structure of line W. (C) Heatmap of the distance matrix for all direct descendants of line R. (D) Pedigree for queens in line R, with each node coloured according to the relatedness of the individual to the Progenitor Queen (G0). (C) Relatedness of queens to their predecessor across Generations G0 to G4.

been assessed in isolation with a combination of long-read sequencing and deep variant calling, and the paternal contribution to inherited variation can now be removed.

Analyses are still ongoing, and by assigning individual drone sires to each queen in generations G1 to G4 and removing the paternal portion of the SNP variation in each queen, patterns of maternal contribution can now be analysed over the 4 generations where drone sires are known. With this information, not only should it be possible to assign genetic relationships derived from the inheritance of haplotypes, but also to gain a deeper understanding of the impact of meiosis and recombination on inheritance.

Purifying selection, the purging of deleterious alleles in the haploid male lineage, could compromise the resolution of meiotic contribution due to skewed inheritance of haplotypes from dam and sire, since not all haplotypes present in a drone source can be expected to be passed down in the paternal line, potentially unilaterally masking paternal contribution. The ongoing analysis of DNA sequence of individual drones from each drone source used in AI will shed light on this.

References

Andrews, S. (2015) FastQC: A Quality Control Tool for High Throughput Sequence Data. Available at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc/

Bernstein, R., Plate, M., Hoppe, A., and Bienefeld, K. (2018) J. Anim. Breed. Genet. 135(4): 323–332. https://doi.org/10.1111/jbg.12347

Brascamp, E. W., and Bijma, P. (2014) Genet. Sel. Evol. 46(1):1–15. https://doi.org/10.1186/s12711-014-0053-9

Li, H., and Durbin, R. (2009) Bioinformatics 25(14):1754–1760. https://doi.org/10.1093/bioinformatics/btp324

Li, H. *et al.* (2009) Bioinformatics 25(16):2078–2079. https://doi.org/10.1093/bioinformatics/btp352

Petersen, G. E. L. *et al.* (2020) Apidologie 51(4):545–556. https://doi.org/10.1007/s13592-020-00741-w

Purcell, S. *et al.* (2007) Am. J. Hum. Genet. 81(3):559–575. https://doi.org/10.1086/519795

Schubert, M., Lindgreen, S. and Orlando, L. (2016) BMC Res. Notes 9(1):1–7. https://doi.org/10.1186/s13104-016-1900-2

van der Auwera, G. A., and O'Connor, B. D. (2020) Genomics in the Cloud: Using Docker, GATK, and WDL in Terra (1st edition). O'Reilly Media, Newton, MA, USA.

Yun, T. *et al.* (2020) Bioinformatics 36(24):5582–5589. https://doi.org/10.1093/bioinformatics/btaa1081

624. SIMplyBee: stochastic simulator of honeybee populations and breeding programmes

L. Strachan[1*], J. Bubnič[2], J. Obšteter[2] and G. Gorjanc[1]

[1]The Roslin Institute and Royal (Dick) School of Veterinary Studies, University of Edinburgh, Easter Bush, Midlothian, EH259RG, United Kingdom; [2]Agricultural Institute of Slovenia, Hacquetova ulica 17, 1000 Ljubljana, Slovenia; l.k.strachan-1@sms.ed.ac.uk

Abstract

In recent years there has been increased interest in the simulation of honeybee breeding programmes, due to their biological and commercial importance. However, existing simulators do not provide all the functionality to replicate honeybee specificities. We developed an R package, SIMplyBee, which extends the popular AlphaSimR simulator with haplo-diploidy, polyandry, complementary sex determination, caste system, and colony quantitative genetics. We developed two classes representing operational units, Colony and Colonies, and a SimParamBee class holding global honeybee simulation parameters. Next, we assigned four levels of functions; Auxiliary, Population, Colony and Colonies to allow the simulation of both individual and colony events such as swarms or supersedures. To demonstrate the package, we simulated a yearly cycle of an isolated apiary and inspected variability at the *csd* locus. SIMplyBee represents the first flexible and holistic simulator of honeybee genomes, biology, reproduction, colony events, and quantitative values at individual and colony levels.

Introduction

The Western honeybee *(Apis mellifera)* is the most important managed pollinator contributing to food security through pollination and hive products. However, in recent decades both native and managed populations have experienced severe over-winter colony losses. These colony losses are due to numerous factors, such as habitat change, disease, pesticides and parasites (Moritz *et al.*, 2010). Due to the related economic damage, there has been an increased interest in honeybee breeding. Breeding programmes aim to improve production and behaviour, reduce labour and colony losses, and manage genetic diversity. However, certain aspects of the honeybee reproductive biology constrain creation of effective breeding programmes.

The honeybee belongs to the insect order Hymenoptera that is characterised by haplo-diploid inheritance with diploid females and haploid males (Goulet *et al.*, 1993). Drones develop from unfertilized eggs and inherit only a copy of the recombined mother's genome. The sole role of drones is to transmit the inherited genome through mating, effectively making them a flying generator of identical gametes (Winston, 1987). Another specificity of the honeybee is the complementary sex determination (*csd*) locus. Fertilised eggs that are heterozygotes for the *csd* locus develop into diploid females, while homozygotes develop into non-viable diploid drones. High rates of inbreeding hence lead to empty brood cells, also called spotty brood pattern (Hasselman *et al.*, 2008). Honeybees are one of the more extreme examples of random mating within the animal kingdom (Laidlaw Jr and Page Jr, 1984). The queen is polyandric and mates on average with about 17 drones during her mating flights. To control the mating, modern breeding programmes often utilise a close population mating structure, artificial insemination, or isolated mating stations.

The increased interest in honeybees has spurred research in quantitative genetics and the breeding of honeybees. The latter is intrinsically supported with stochastic computer simulations that are commonly used to evaluate quantitative genetic models, statistical methods, and breeding strategies (e.g. Bienefeld *et al.*, 2007). Due to biological and reproductive specificities of the honeybee, most of the existing simulators

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_624, © L. Strachan *et al.* 2022

cannot simulate honeybee populations and breeding programmes in entirety. The existing honeybee simulators handle either the colony quantitative genetic values (Du *et al.*, 2021a, 2021b; Plate *et al.*, 2019a) or the dynamics within the colony or hive (Becher *et al.*, 2014). The aim of this paper is to present the SIMplyBee R package, which is a honeybee extension of the popular AlphaSimR stochastic simulator of breeding programmes (Faux *et al.*, 2016). The package simulates the honeybee genomes, biology, reproduction, colony events, and quantitative values at individual and colony level.

Implementation

We extended AlphaSimR with functions addressing the biological peculiarities of honeybees: haplo-diploid system, polyandry, complementary sex determination, colony events, and quantitative values at individual and colony level. To this end we have developed three R classes; one holding global simulation parameters, SimParamBee, and two classes representing operational units: Colony and Colonies. We then assigned four levels of functions; Auxiliary, Population, Colony and Colonies, to allow the simulation of both individual and colony events, therefore permitting the user a greater amount of flexibility.

Classes. SimParamBee is a class holding global simulation parameters. It inherits information from the AlphaSimR SimParam class, such as the number of chromosomes, segregating sites, causal sites, simulated traits, variance components, etc. Additionally, it holds default colony size as values or functions, information on the *csd* locus, and mapping of individual queen, virgin queens, workers, or drone genetic or phenotypic values to colony phenotype values. Colony class represents a honeybee colony and contains 14 slots. The slots hold identification and location, castes (the queen, virgin queens, workers, and drones), information about past events of the colony (split, swarm, supersedure, collapse, and production), and phenotypic values of the colony. The queen slot in addition stores fathers, which are the drones the queen mated with. Colonies class combines multiple Colony objects into one to simplify simulating a whole breeding programme.

Functions. Level0 functions are Auxiliary functions that provides different parameters about populations like number of individuals, information about genome or *csd* related information. Level1 returns AlphaSimR Pop objects. These functions create individuals (virgin queens, workers, or drones), extract or pull individuals of a certain caste from the colony, cross a virgin queen with selected drones, create drone congregation area with the drones from selected colonies and pull drones groups from it, etc. Level2 returns Colony objects. These functions create a new colony, cross a colony, add, replace or remove individuals of a caste to the colony, requeen or build up a colony to full size, to simulate events in colony life (collapse, swarm, supersedure, split), to reset colony events, set location and phenotype, etc. Level3 returns Colonies objects. Functionalities of Level3 mirrors the Level2 but they work on Colonies instead of Colony.

Example. We simulated a founder population of 1000 individuals with 16 chromosomes, 1000 segregating sites per chromosomes, and 128 *csd* alleles. Next, we simulated 10 years of an isolated managed apiary with 20 or 40 colonies with 20,000 workers and 4,000 drones per colony. In the first year, we created mated-queen colonies from the founder population. Within each year, we simulated three periods resembling spring, summer, and winter. In period 1, we build up the colonies to their full size, split all the colonies with a queen of age 1 and requeened the splits, swarmed 5%, superseded 5% of the colonies with the queen of age one or two, mated the virgin colonies to available drones, and collapsed 10% of the colonies with a queen of age one or two. In period 2, we swarmed 1% and superseded 5% of the colonies, mated all the virgin colonies to available drones, and collapsed 10% of the colonies with a queen of age one or two. During period 3, we collapse 25% of colonies with a new queen (queens of that year), 30% of colonies with a queen of age 1, and all the colonies with a queen of age two. At the end of the season, we removed the splits (resembling a sell) to keep a constant number of colonies in the apiary. At the end of each year, we inspected the average number of *csd* alleles in the colony and the apiary.

Results

We created a stochastic simulator of honeybee populations and breeding programmes and ran an example session to simulate the yearly cycle of a managed apiary. We annually inspected the average number of distinct *csd* alleles per apiary and per colony (Figure 1). An isolated apiary with 20 colonies did not sample all the available 128 *csd* alleles from the start (average 119), and lost approximately half of the *csd* alleles in 10 years of isolated mating (average 57.8 in year 10). An isolated apiary with 40 colonies sampled nearly all *csd* alleles from the start (average 127) and lost a third of the alleles by year 10 (average 88.1). The dynamics within an individual colony was different. In year 1, each colony sampled on average 27.4 *csd* alleles in both scenarios. This number increased to 34.2 in year 5 with 20 colonies and 38.7 in year 6 with 40 colonies. This was followed by a step decline in the average number of 19.8 and 23.8 *csd* alleles per colony respectively with 20 and 40 colonies.

Discussion

We have developed a flexible and holistic simulator of honeybee populations and breeding programs that accounts for the biological specificities of honeybees. We provide the simulator as an R package named SIMplyBee. SIMplyBee builds upon the work of the popular AlphaSimR simulator (Gaynor *et al.*, 2021). We have used this framework to allow for the simulation of individual bees with individual genomes and quantitative values, but combining them into castes of objects of the Colony class. Here we demonstrated the functionality of SIMplyBee by modelling the yearly cycle of an isolated managed apiary. The whole script was only 306 lines long (including comments and loops) and ran 10 years in about 90.4 minutes with 20 and 163.7 minutes with 40 colonies per replicate. This shows that SIMplyBee can simulate whole breeding programmes with ease.

Future work will aim to improve the simulation of colony phenotypes, opposed to individual honeybee phenotypes. SIMplyBee already allows for the simulation of individual genetic and phenotypic values as well as their mapping to colony phenotypes. However, to map these onto the colony phenotype, the user must specify a custom function. We want to provide sensible defaults for such functions to account for the queen and worker effects and associated quantitative genetic parameters. We are also working on improving simulation of the founder genomes based on the honeybee demographic history. Currently, additive genetic relationships between colonies and caste members are not dealt with but needs to be modelled with additional external tools based on either the pedigree or genomic data obtained from the simulator. SIMplyBee is currently under active development and will be released as open source to the community.

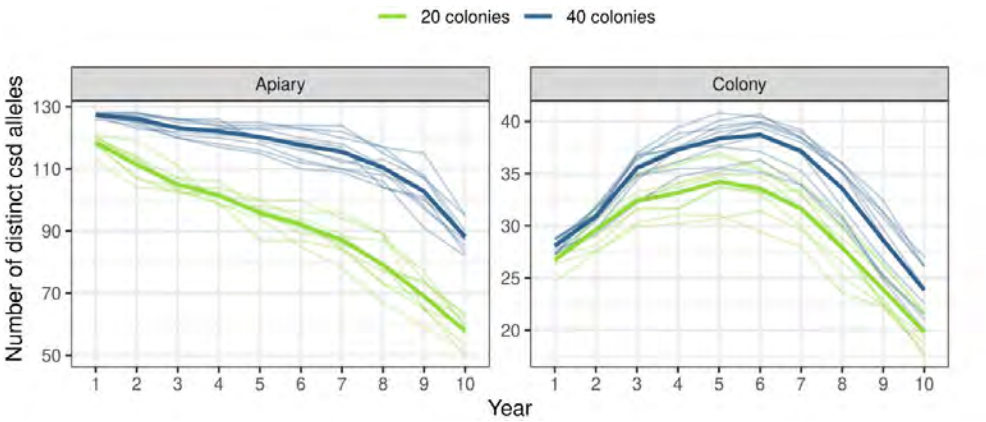


Figure 1. The average number of distinct *csd* alleles by year (a) in the apiary and (b) in a colony of the apiary.

References

Becher M. A., Grimm V., Thorbek P., Horn J., Kennedy P. J., *et al.* (2014) *Journal of Applied Ecology*, *51*(2):470–482.

Bienefeld K., Ehrhardt K., Reinhardt F. (2007) *Apidologie*, *38*(1), 77–85.

Du, M., Bernstein, R., Hoppe, A., & Bienefeld, K. (2021a). *Genetics Selection Evolution*, *53*. https://doi.org/10.1186/s12711-021-00606-5

Du, M., Bernstein, R., Hoppe, A., & Bienefeld, K. (2021b). *Heredity*. doi.org/10.1038/s41437-021-00411-2

Faux, A.-M., Gorjanc, G., Gaynor, R. C., Battagin, M., Edwards, S. M., Wilson, D. L., Hearne, S. J., Gonen, S., & Hickey, J. M. (2016). *The Plant Genome*, *9*(3), plantgenome2016.02.0013. doi.org/10.3835/plantgenome2016.02.0013

Gaynor, R. C., Gorjanc, G., & Hickey, J. M. (2021) *G3*, *11*(2), jkaa017.

Goulet, H., Huber, J. T., & Canada Agriculture Canada Research Branch. (1993). Research Branch, Agriculture Canada. *Center for Land and Biological Resources Research, Ottawa, Ontario, Canada, 680pp*.

Hasselmann M., Gempe T., Schiøtt M., Nunes-Silva C. G., Otte M., Beye M. 2008. *Nature*, *454*, 7203:, 519-522.

Laidlaw Jr, H. H., & Page Jr, R. E. (1984). *Genetics*, *108*(4), 985–997.

Moritz, R. F. A., de Miranda, J., Fries, I., Le Conte, Y., Neumann, P., & Paxton, R. J. (2010). *Apidologie*, *41*(3), 227–242. doi.org/10.1051/apido/2010010

Plate, M., Bernstein, R., Hoppe, A., & Bienefeld, K. (2019a). *PLOS ONE*, *14*(3), e0213270. doi.org/10.1371/journal.pone.0213270

625. Assessment of uncertainty in direct and maternal genetic parameters estimates for honeybee colony phenotypes

T. Kistler[1,2*], E.W. Brascamp[3], B. Basso[4,2], P. Bijma[2] and F. Phocas[1,2]

[1]Université Paris-Saclay, INRAE, AgroParisTech, GABI, 78350 Jouy-en-Josas, France; [2]UMT PrADE, France; [3]Wageningen University & Research, P.O. Box 338, 6700 AH Wageningen, the Netherlands; [4]INRAE, UR 406 Abeille et Environnement, 84914 Avignon, France; tristan.kistler@inrae.fr

Abstract

Genetic parameters in honeybees are commonly estimated using REML methodologies applied to animal models with maternal effects. These methodologies were adapted to the species' peculiarities, including: phenotypes measured on colonies, haplo-diploïdy and polyandrous mating. However, estimations' reliability is hindered by the small size of the breeding nuclei commonly used. We assessed the uncertainty of variance component estimates in small simulated honeybee populations by evaluating the impact of the breeding nucleus size, the mating strategy and the direct-maternal genetic correlation. The convergence of estimations was strongly hindered when considering a small breeding nucleus with less controlled mating and negatively correlated genetic values. Furthermore, biases could be observed in this scenario. Individual estimates deviated strongly from the real values in about 40% of the populations. When considering real breeding populations, these results highlight the caution to be taken with estimates from small populations with complex pedigree structures.

Introduction

Most of economically important traits in honeybees are not recorded on an individual scale, but at the scale of the colony. A colony consists of a single diploid queen and thousands of sterile female workers that descended from the queen and from 10 to 20 haploid drones that have mated with the queen. The performance of a colony is affected by the joint genetic effects of all workers (called worker or direct effect) and of the queen (called queen or maternal effect). Polyandry and haplo-diploid sex determination are biological specificities that require an appropriate relation matrix for honeybee genetic evaluation. Efficient methodologies based on BLUP and REML applied to animal models are widely used to estimate breeding values in farm animals and have been adapted for honeybees (Bienefeld *et al.*, 2007; Brascamp and Bijma, 2014).

The phenotypic variance of colony performance can be partitioned into an environmental variance, an additive direct genetic variance, a maternal genetic variance and a potential covariance. A reliable estimation of these genetic variances as well as of the covariance between direct and maternal effects requires to get data and pedigree records in a large population constituted from genetically well-connected apiaries. Unfortunately, most of honeybee breeding programs are applied to small nucleus populations, usually ten to a few tens of breeding queens. This is different from livestock species for which seedstock includes hundreds or thousands of animals. Therefore, the main aim of our study was to assess the uncertainty in estimates of genetic parameters for polygenic traits measured in small simulated honeybee populations. In addition, we studied the impact of two mating designs on these estimates. In the first design, the drones fertilizing a virgin queen came from a single drone-producing queen (DPQ) that can be considered as a 'single diploid sire' (SS) of the workers. In the second design, the drones came from a group of sister DPQs, and this group of DPQs can be treated as a 'dummy diploid sire', referred to as a pseudo-sire (PS).

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_625, © T. Kistler *et al.* 2022

Materials & methods

Population structure and stochastic simulations. Performance and breeding values of honeybee populations were simulated at the scale of individual queens, drones and worker groups. In the base population, queens' breeding values were drawn from a normal distribution $N(0, \Sigma^2_{BV})$ with Σ^2_{BV} the genetic variance-covariance matrix with both direct and maternal variances equal to 10 and a covariance assumed either to be 0 or -5. The base drones' breeding values were drawn from $N(0, \frac{1}{2}\Sigma^2_{BV})$. The only identified fixed effect was a year effect drawn from $N(0, 20)$ and residual effects on the performances were drawn from $N(0, 30)$.

Ten generations of within-line (one new breeding queen per maternal sister group) phenotypic selection were run, starting from non-inbred and unrelated breeding queens (either 12 or 24 BQs for a small and larger nucleus size scenario) and unrelated drones in the base population. Two-third of the same pool of maternal lines were selected each year to form DPQ families. Eight DPQs were then randomly chosen per selected family.

Two mating strategies were used: either sister DPQs formed a PS, jointly producing the drones mating a single potential dam queen with a random participation of each sister to the drone pool, or only one SS was randomly chosen to produce all the drones mating a single potential dam queen (controlled mating). All DPQs were mated to drones coming from a wild population (open-mating) with the same expected breeding values than those of the base population. Each queen (BQ or DPQ) was mated to 8 drones. Each year, wintering mortality was modelled by randomly eliminating 25% of all queens. Phenotypes were obtained after wintering, within the first year of birth and before reproduction of the dams. Both potential BQs and DPQs had phenotypes. Generation interval was 1.5 years (1 year on the maternal path and 2 years on the paternal path). Each BQ produced 24 potential BQs and 20 potential DPQs. Further details including genetic inheritance modelling can be found in Kistler *et al.* (2021). Each simulation scenario was repeated 100 times to assess the sampling standard deviations and the uncertainty of the estimates of genetic parameters.

Genetic evaluation model and estimation of genetic parameters. The vector of phenotypes y was described using a linear animal model accounting for maternal genetic effects: $y = Xb + Z_d a_d + Z_m a_m + e$, where b is the vector of fixed year effects with a corresponding incidence matrix X, a_d the vector of direct effects with incidence matrix Z_d, a_m the vector of maternal effects with incidence matrix Z_m, and e the vector of residual effects. Pedigree and performances of all simulated colonies were used for a single REML evaluation after 10 years of queen phenotypic selection. The total number of colony records was either 4,176 or 8,352, with 132 or 264 dam families for small or larger breeding nuclei, respectively.

BLUPf90 (Misztal *et al.*, 2002) with its AIREMLF90 program was used to estimate genetic parameters based on the inverse of a relationship matrix, following Brascamp and Bijma (2014, 2019). In this, it accounts for the DPQs being either a pseudo-sire, a single sire or a wild set of bees (open mating).

Results

Table 1 presents the 8 study scenarios. Around 90% of all estimations converged. However, PS mating hindered convergence, as well as a small nucleus size and a negative direct-maternal correlation (r_{dm}) to a lesser extent. Thus, only 2/3 of the 100 simulations lead to converged estimates of variance components in scenario PS12$^-$.

Among converged cases, scenario PS12$^-$ had a significantly biased σ^2_d estimate (P-value<5% using a t-test). Other scenarios showed tendencies to over- (or under-) estimate one or several (co)variances but without clear statistical significance.

Even among scenarios with mean estimations centred on the true values, individual estimates varied strongly around the real (simulated) values, with the nucleus size having the greatest impact (Table 1). Smaller nucleus scenarios had (co)-variances estimates deviating strongly in 10 to 20% more repetitions than bigger nucleus scenarios.

Strongly deviating estimates occurred more often for the direct variance (in 45% of all repetitions) than for the maternal (37%) or for their covariance (32%). To a lesser extent, PS mating compared to SS mating also increased the number of strongly deviating estimates, mostly for the maternal variance and the direct-maternal covariance.

For the large nucleus size scenarios with PS mating, Figure 1 shows the relation between direct and maternal variance deviations, coloured by the deviation of the direct-maternal covariance (strong negative deviations in lavender and strong positive ones in red). In both scenarios, a clear structure appears: when

Table 1. Summary statistics of differences between estimated and true genetic (co)variances in 100 replicates of 8 simulated scenarios. True variances σ^2_d and σ^2_m are equal to 10.

ID	r_{dm}	% converged estimates	Difference between estimated and true σ^2_d		Difference between estimated and true σ^2_m		Difference between estimated and true σ_{dm}	
			Mean (se)	% strong deviates[1]	Mean (se)	% strong deviates[1]	Mean (se)	% strong deviates[1]
PS12	0	78	0.19 (0.35)	51	-0.07 (0.33)	45	-0.20 (0.28)	40
PS24	0	89	0.12 (0.24)	38	-0.12 (0.26)	42	-0.31 (0.16)	21
SS12	0	94	-0.58 (0.30)	54	-0.05 (0.26)	37	0.17 (0.22)	39
SS24	0	98	-0.41 (0.21)	34	0.05 (0.19)	26	-0.15 (0.16)	18
PS12$^-$	-0.5	66	1.48 (0.42)	58	0.54 (0.32)	44	-0.65 (0.33)	48
PS24$^-$	-0.5	78	0.49 (0.29)	38	-0.13 (0.24)	38	-0.15 (0.24)	31
SS12$^-$	-0.5	97	-0.05 (0.31)	57	-0.22 (0.27)	42	0.20 (0.25)	40
SS24$^-$	-0.5	100	-0.36 (0.21)	30	-0.30 (0.15)	24	0.14 (0.15)	22

[1] Proportion of estimated variances deviating by more than 20% of the true values.

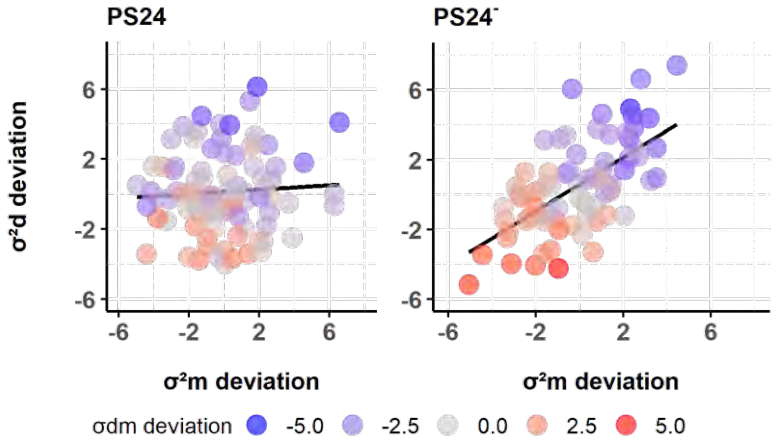


Figure 1. Estimated genetic parameters' deviations from true (simulated) values for two scenarios with a null (PS24) or negative (PS24$^-$) direct-maternal genetic correlation.

both direct and maternal variances were underestimated, the covariance was overestimated, and *vice-versa* for overestimated variances. In addition, in particular for the negative correlation scenario, when the direct variance was under- (over-) estimated, the maternal variance tended to be under- (over-) estimated as well.

Discussion

Very unreliable estimates of variance components were observed for scenarios combining a small size of the breeding nucleus, PS mating and a negative direct-maternal genetic correlation. In PS mating, true relationships between DPQs and descendants are unknown and have to be probabilistically accounted for, making estimates less precise. Additionally, direct and maternal effects are more difficult to disentangle when they are correlated, which can be seen in Figure 1. A last explanation that may explain some bias in the estimates of variance components lies in the different hypotheses considered for the contributions of various DPQs and drones to any queen's offspring in the simulation of the population and in the derivation of the relationship matrix. In our simulations, these contributions were balanced while they were drawn from a Poisson distribution in Brascamp and Bijma (2014)' relationship matrix.

Beyond the biased estimates in one scenario, the number of estimates deviating strongly from their true values was always very high. Using real datasets with very small populations sizes, Guichard *et al.* (2020) and Basso *et al.* (2022) could not estimate jointly direct and maternal effects for various honeybee traits. While 24 breeding queens in a nucleus population is currently considered as a decent size in honeybee breeding, it remains a small population for genetic parameters' estimation purposes, even with 10 years of records. Furthermore, we simulated a simplistic situation in which the whole population was composed of a unique apiary only submitted to a year effect. In real life, Clément *et al.* (2001) showed that low genetic connectedness between herds (apiaries) would induce severe biases in the estimated genetic parameters. A strategy to avoid these issues in genetic evaluation for small honeybee populations could be to use estimates of genetic parameters based on large datasets, such as in Hoppe *et al.* (2020), and to assume that they are very similar in the small population of interest.

References

Basso B., Kistler T., Gerez T. and Phocas F. (2022) Proc. of the 12th WCGALP, Rotterdam, the Netherlands, in press

Bienefeld K., Ehrhardt K., and Reinhardt F. (2007) Apidologie 38:77–85. http://dx.doi.org/10.1051/apido:2006050

Brascamp E.W., and Bijma P. (2014) Genet. Sel. Evol. 46:53. http://dx.doi.org/10.1186/s12711-014-0053-9

Brascamp E.W., and Bijma P. (2019) Proc. of the 46th Apimondia, Montreal, Canada Clément V., Bibé B., Verrier É., Elsen J.-M., Manfredi E. *et al.* (2001) Genet. Sel. Evol. 15;33(4):369. http://dx.doi.org/10.1186/1297-9686-33-4-369

Guichard M., Neuditschko M., Soland G., Fried P., Grandjean M. *et al.* (2020) Apidologie 51(5):876-891. http://dx.doi.org/10.1007/s13592-020-00768-z

Hoppe A., Du M., Bernstein R., Tiesler F. K., Kärcher M., and Bienefeld K. (2020) Insects 11(11):768. http://dx.doi.org/10.3390/insects11110768

Kistler T., Basso B., and Phocas F. (2021) Genet. Sel. Evol. 53(1):1-16. http://dx.doi.org/10.1186/s12711-021-00665-8

Misztal I., Tsuruta S., Strabel T., Auvray B., and Lee D. (2002). Proc. of the 7th WCGALP, Montpellier, France.

626. Very local genotype by environment interaction in Austrian honey bees

E.W. Brascamp[1*], M. Rubinigg[2,3], R.F. Veerkamp[1] and P. Bijma[1]

[1]Wageningen University & Research, P.O. Box 338, 6700 AH Wageningen, the Netherlands; [2]Biene Österreich, Georg-Coch Platz 3/11a, 1010 Vienna, Austria; [3]CARE, Appalachian State University, ASU, Box 32049, Boone, NC 28608, USA; pim.brascamp@wur.nl

Abstract

We investigated genotype by environment interactions (G×E) in the breeding programme of the Austrian Carnica population administered by Biene Österreich by the estimation of genetic correlations (r_G) at three levels: between geographical regions, between colonies that were bred at home or elsewhere, and between individual test locations. Traits included were honey yield, gentleness, calmness and swarming behaviour. Results suggest that honey bees are very prone to G×E, not only between regions and bred at home v. elsewhere, but also at the level of individual test locations. Despite this abundance of G×E we conclude that breeding programmes covering larger areas are justified, and will result in selection for general adaptability.

Introduction

Between 2009 and 2012 a Europe-wide experiment was carried out to study genotype by environment interaction (G×E) in honey bees (Costa *et al.*, 2012). It was triggered by the phenomenon of colony losses (Neumann and Carreck, 2010), partly associated with Varroa infestation, and the experiment intended to clarify whether G×E played a role in vitality and performance of colonies. To evaluate G×E, queens that were fertilized at their location of origin were distributed across locations in Europa and introduced in host nuclei. An overview of the published results was given by Meixner *et al.* (2014). G×E was found for vitality, performance and behavioural traits, and for vitality colonies performed best at their local environment. This lead to the notion of local adaptation, a long-term adaptation to the local environment by selection. Presently the notion of local adaptation is dominant in thinking about honey-bee genetic improvement and conservation of local breeds.

Despite the evidence of G×E there are breeding programmes that cover larger geographical areas, like the whole of Germany and Austria and a few other countries in the case of Beebreed (Hoppe *et al.*, 2020), Austria in the case of Biene Österreich (Brascamp *et al.*, 2016) and Switzerland (Guichard *et al.*, 2020).

Here we investigate G×E in the breeding programme of the Austrian Carnica population administered by Biene Österreich by the estimation of genetic correlations (r_G) at three levels: between geographical regions, between colonies that were bred at home or elsewhere, and between individual test locations. We briefly discuss the consequences of our findings for breeding-programme design.

Materials & methods

The data contained 16,924 records on honey yield, gentleness, calmness and swarming behaviour, collected in the breeding programme of Biene Österreich from performance testing year 1998 to 2020. Honey yield was weighed, and the other traits were scored on a scale of 1-4. Full pedigrees were available. For details, see Brascamp *et al.* (2016) with data until 2014.

Austria recognizes eight 'landwirtschaftliche Hauptproduktionsgebiete' (main agricultural production areas, MPA). The MPA in which colonies were tested, were known for all records. MPA8 differs from

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_626, © E.W. Brascamp *et al.* 2022

MPA1-7 in an agro-ecological sense, and shows a far higher honey yield. For this reason, we considered two geographical regions, MPA1-7 v. MPA8.

We also distinguished between colonies bred at home v. bred elsewhere. For all records, both the breeder of the queen heading a colony and the tester of the colony were known. When the breeder differed from the tester, the colony was said to be bred elsewhere. In case of colonies tested elsewhere, the identities of queens were blinded to ensure objective measurements, but testers could distinguish own colonies from colonies bred elsewhere. Table 1 summarizes the data.

Methods. We studied G×E by estimating the genetic correlation (r_G) between the traits expressed in one environment v. the other. In honey bees, the consensus statistical model includes two random genetic effects, a worker effect and a queen effect (Bienefeld and Pirchner, 1990). However, after splitting the data into groups for different environments, this model did not converge. We, therefore, included only a worker effect. The numerator relationships matrix was based upon a pedigree file including groups of workers, dams, and groups of drone-producing queens (sires), following Brascamp and Bijma (2014). All models included a fixed effect for year×tester×apiary (called *test location*), and also for bred at home v. elsewhere except when analysis was limited to bred at home. Data were analysed using ASReml (Gogel *et al.* 2015).

To evaluate G×E due to the MPA, we estimated r_G between MPA1-7 and MPA8, either using all records or only the records of colonies bred at home (9,038 for MPA1-7 and 2,703 for MPA8). To evaluate G×E due to bred at home v. elsewhere, we estimated r_G between the trait expressed in colonies tested in MPA1-7, bred at home (9,038) v. elsewhere (4,724).

To evaluate G×E at the level of individual test locations, we defined two hypothetical traits ('between test locations' scenario). We randomly selected half of all test locations present in our data, and allocated all records of these test locations to the first trait. All records of the remaining test locations were allocated to the second trait. When a tester had different test locations within a year, care was taken that these were evenly distributed over the two traits. Consequently, the genetic correlation between the two hypothetical traits originates solely from covariances between test locations, whereas information for genetic variances originates also from covariances between records belonging to the same test location. Hence, an $r_G<1$ indicates G×E between test locations.

To validate our analysis, we tested whether G×E fully disappeared (i.e. $r_G\approx1$) when there was no systematic difference between the environments for the two hypothetical traits. For this purpose, we randomly selected half of all records from each test location, and allocated these records to the first trait. The remaining records

Table 1. Numbers of colonies (N), means and standard deviations (between brackets) for colonies tested and bred in different regions (MPA1-7 v. 8).

Tested	Bred	Home/elsewhere	n	Honey	Gentleness	Calmness	Swarming
1-7	1-7	Home	9,038	39.1 (24.9)	3.67 (0.49)	3.65 (0.48)	3.66 (0.72)
1-7	1-7	elsewhere	4,724	34.5 (21.7)	3.54 (0.61)	3.54 (0.59)	3.53 (0.89)
1-7	8	elsewhere	174	39.2 (28.3)	3.62 (0.52)	3.61 (0.52)	3.46 (0.93)
8	8	Home	2,703	85.0 (38.3)	3.64 (0.46)	3.61 (0.46)	3.52 (0.85)
8	8	elsewhere	102	75.1 (28.5)	3.66 (0.50)	3.65 (0.54)	3.35 (0.85)
8	1-7	elsewhere	183	36.4 (20.0)	3.59 (0.49)	3.61 (0.50)	3.39 (0.93)

of each test location were then allocated to the second trait. Hence, each test location was split into two, including an equal division of full sib colonies within a test location ('within test locations' scenario). In this analysis, both the information for the genetic correlation and the information for the genetic variances originate from between and within test-location covariances. Hence, this analysis should yield $r_G \approx 1$, which served as a check for spurious G×E.

Results

Table 2 shows results for r_G. Even though the dataset is of considerable size, standard errors on estimated r_G were large. The estimated r_G for honey yield between MPA1-7 and MPA8 was low to moderate, but not significantly different from one. The large standard errors on these estimates indicate a lack of statistical power, and our results should not be interpreted as evidence of absence of G×E. The estimated r_G for honey yield between colonies bred at home and bred elsewhere was small (0.36±0.14), and significantly different from one, indicating clear G×E. When entire test locations were randomly divided over two hypothetical traits, then the estimated r_G between these two traits was significantly different from 1 for all four traits. Thus there was clear G×E at the level of individual test locations. In contrast, when records within test locations were randomly divided over to the two traits, all estimates of r_G were virtually unity, as expected. Hence, we did not find indications of spurious G×E in our analysis.

Discussion

Because G×E might be caused by the skewed distribution of the four traits, we looked at Box-Cox transformed data (Box and Cox, 1964) in the analysis in the full dataset of MPA1-7 v. MPA8. The transformed variables equalled $(y^{\lambda}-1)/\lambda$ where y is the observed value. For honey yield λ was 0.3 and for the other three traits it equalled 2. Results were similar to those in Table 2 and for that reason we didn't transform the data for final analyses.

The standard errors for the estimates of r_G were much lower in case test locations were divided randomly over two traits as compared to the dataset with the actual trait groups (Table 2). This was probably caused by the tighter genetic relationships between colonies assigned to both traits as compared to the actual groups.

Results suggested that honey bees are very prone to G×E. Even when dividing test locations arbitrary over two traits, there was considerable G×E between these traits. This result contrasts with results generally found in farm animals, where r_G is often high.

Genetic correlations between different regions found here were in line with Hoppe (2021, personal communication) who observed a broad range of genetic correlations between regions in the dataset of Beebreed (*re* Hoppe, 2020, for description of the data) using statistical models similar to ours. In our

Table 2. Genetic correlations with SE (between brackets) between regions, between bred home vs elsewhere tested in MPA1-7, between test locations, and within test locations.

Data used	Honey	Gentleness	Calmness	Swarming
MPA1-7 v. MPA8	0.55 (0.32)	-0.31 (0.64)	0.64 (0.28)	-0.51 (0.58)[1]
MPA1-7 v. MPA8, home bred	0.28 (0.61)	0.88 (0.39)[1]	0.84 (0.56)[1]	-0.95 (0.69)
Home v. elsewhere in MPA1-7	0.36 (0.14)	0.68 (0.17)	0.89 (0.28)	0.92 (0.70)[1]
Between test locations	0.63 (0.07)	0.85 (0.07)	0.75 (0.08)	0.41 (0.16)
Within test locations	0.99 (0.01)	0.99 (0.02)	0.99 (0.02)	0.96 (0.06)

[1] This analysis did not result in stable estimates for the additive genetic (co)variance components.

analysis this G×E wasn't statistically significant in terms of r_G, but given the significant G×E when dividing test locations arbitrary over two traits, the null-hypothesis of r_G=1 is not realistic, and therefore we consider the G×E between regions as real. This type of G×E may arise because different sets of alleles or allele frequencies may be optimal for different regions and may be expected to remain so over longer periods. On the other hand, for the contrast bred at home *v.* elsewhere one might think of short term effects, perhaps of an epigenetic nature. Yagound *et al.* (2020) demonstrated in honey bees that methylation patterns in paternal germ cells, through drones, were present in worker offspring. In our case it would involve the maternal pathway, as workers in colonies bred at home and elsewhere often share the same sire. The queens heading the colonies of course differed. If G×E for tested at home *v.* elsewhere is a short-term effect indeed, it may diminish over generations. This might have implications for experiments in which different genetic stocks are being tested at home and elsewhere, as in the Europe-wide experiment. Based on our findings, one should consider to evaluate stocks for successive generations, such that in later generations all are tested at home.

For farm-animal scenarios Mulder and Bijma (2006) addressed the question which level of r_G would justify to set up different breeding programmes for different countries or regions, or at least to work with different sets of estimated breeding values. Considering the high degree of G×E in honey bees it seems tempting to set up local or region-specific breeding programmes, suggesting splitting a breeding programme into very many small ones. However, the resulting small scale for each programme would obviously hamper genetic progress and increase rates of inbreeding. Furthermore the estimates of heritability of Brascamp *et al.* (2018), who used largely overlapping Biene-Österreich data and ignored G×E, suggest that genetic improvement across environments is realistic for the population as a whole. Hoppe *et al.* (2020) demonstrated considerable genetic improvement in Beebreed as well, where breeding value estimation ignores G×E.

We conclude that G×E is abudant in Austrian honey bees, but that this provides no basis to split breeding programmes into smaller ones, and that the solution lies in selection for general adaptability.

References

Bienefeld K., and Pirchner F. (1990) Apidologie 21:175–183. https://doi.org/10.1051/apido:19900302

Box G.E.P., and Cox D. R. (1964) J. R. Stat. Soc. B 26:211-252.

Brascamp E.W., and Bijma P. (2014) Genet. Sel. Evol. 46:53. https://doi.10.1186/s12711-014-0053-9

Brascamp E.W., Willam A., Boigenzahn C., *et al.* (2016) Apidologie 47(6):739-748. https://doi.org/10.1007/s13592-016-0427-9 and (2018) Apidology 49(4):462-463. https://doi.10.1007/s13592-018-0573-3

Costa C., Büchler R., Berg S., *et al.* (2012) J. Apicultural. Sci. 56(1):147–158. https://doi.org/10.2478/v10289-012-0015-9

Gogel B.J., Gilmour A.R., Welham S.J., *et al.* (2015) ASReml Update: What's new in Release 4.1. VSN International Ltd, Hemel Hempstead.

Guichard M., Neuditschko M., Soland G., *et al.* (2020) Apidologie 51(5):876-891. https://doi.org/10.1007/s13592-020-00768-z

Hoppe A., Du M., Bernstein R., *et al.* (2020) Insects 11:768. https://doi.org/10.3390/insects11110768

Meixner M.D., Büchler R., Costa C., *et al.* (2014) J. Apicultural Res. 53(2):183-187. https://doi.10.3896/IBRA.1.53.2.01

Mulder H.A., and Bijma P. (2006) J. Dairy Sci. 89:1727–1739. https://doi.org/10.3168/jds.S0022-0302(06)72241-X

Neumann P., and Carreck N. (2010) J. Apicultural Res. 49:1-6. https//doi.org/10.3896/IBRA.1.49.1.01

Yagound B., Remnant E.J., Buchmann G., and Oldroyd B.P. (2020) PNAS, 117(51):32519-32527. https://doi.10.1073/pnas.2017094117

Part 36

Bovine beef

627. Genetic prediction for growth traits in beef cattle using selected variants from imputed low-pass sequence data

C.A. Russell[1*], L.A. Kuehn[2], W.M. Snelling[2] and M.L. Spangler[1]

[1]Department of Animal Science, University of Nebraska, Lincoln, NE 68583, USA; [2]USDA, ARS, Roman L. Hruska U.S. Meat Animal Research Center, Clay Center, NE 68933, USA; crussell17@huskers.unl.edu

Abstract

A beef cattle population (n=2,343) was used to assess the impact of variants identified from imputed low-pass sequence on the genetic prediction of birth weight (BWT) and post weaning gain (PWG). Variants (n=1,145,892) were selected based on functional impact and were partitioned into low, modifier, moderate, and high based on predicted functional consequences. Each subset was used to construct a genomic relationship matrix (GRM) in univariate animal models. When all variants were included in a single GRM, heritability estimates for BWT and PWG were 0.41±0.05 and 0.37±0.05, respectively. Heritability estimates for BWT when the GRM was comprised of only low, modifier, moderate, or high variants were 0.36±0.05, 0.39±0.05, 0.33±0.05, and 0.10±0.03, respectively. Similar estimates for PWG were 0.33±0.05, 0.34±0.05, 0.32±0.05, and 0.10±0.03, respectively. Results suggest that despite predicted functional consequences, the high variants accounted for only ~24-27% of the genetic variance. The USDA is an equal opportunity employer and provider.

Introduction

Although advances have been made in recent years relative to the incorporation of genomic data into routine genetic evaluations of beef cattle and corresponding increases in predictive accuracy, hope remains that the inclusion of causal variants into prediction models will lead to further gains in accuracy. Such efforts require the use of whole-genome sequencing (WGS), and despite decreases in cost, wide-spread deep sequencing of seedstock cattle is cost prohibitive. The use of low-pass sequencing at much shallower depths (e.g. 0.5×) is a much more attractive alternative, when coupled with imputation, to garner genotypes of potentially causal variants on large numbers of individuals. Such an approach also represents a means of making marker subsets actually used in genetic evaluations more flexible and dynamic.

Large effect variants for complex traits such as birth weight in beef cattle are known to exist and are predictive in populations external to where they were identified (Snelling *et al.*, 2017). However, the general benefits of including putative causal variants in genetic predictions or the use of WGS for prediction more generally has had mixed results in the literature (e.g. Heidaritabar *et al.*, 2016; Veerkamp *et al.*, 2016; Warburton *et al.*, 2020). Several challenges exist relative to the use of WGS for genetic prediction, including the inability to accurately estimate the effect of all WGS SNP using currently available phenotypic datasets to allow for accurate pre-selection of variants. Additionally, currently available high-density panels may sufficiently capture enough genetic variation through LD with causal variants to limit gains from using WGS (e.g. Frischknecht *et al.*, 2018). However, evidence in the literature supports the concept of pre-selected WGS variants to improve prediction accuracy in multi-breed populations (e.g. Raymond *et al.*, 2018). Consequently, the objective of the current study was to investigate the benefit of fitting variants based on predicted functional impact in prediction models for birth weight (BWT) and post weaning gain (PWG) in a multi-breed beef cattle population using a genomic best linear unbiased prediction framework.

Materials & methods

Data. A composite beef cattle population (n=2,343) with sequence variant genotypes imputed from low-pass (~0.5×) WGS and recorded BWT and PWG was used. Animals were part of the U.S. Meat Animal

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_627, © C.A. Russell *et al.* 2022

Research Center (USMARC) Germplasm Evaluation (GPE) project. The low-coverage sequence was submitted to the Gencove pipeline for imputation with loimpute (Loimpute-Public, 2020) to a haplotype reference panel constructed from WGS of 946 cattle (598 available from NCBI Sequence Read Archive; 348 GPE sires) (Snelling *et al.*, 2020). Functional impact of the imputed variants was assessed with snpEff using the Ensembl annotation of the ARS-UCD1.2 assembly of the bovine genome (Aken *et al.*, 2016; Cingolani *et al.*, 2012; Rosen *et al.*, 2020). Genotypes for interesting variants were extracted from the imputed calls of each individual with low-pass sequence. Interesting variants included variants in exons of protein-coding genes, which may affect gene function, and variants in untranslated regions (UTR) and non-coding RNA, which may impact gene regulation. Low-pass calls were required to have a genotype probability greater than 0.95, then a 0.95 call rate filter by animal and variant was applied to the set of interesting variants. In total, a set of 1,145,892 interesting variants was identified for further analysis. Interesting variants were then partitioned into four distinct classes based on level of likely impact: Low (440,684 variants across 21,700 genes), Modifier (372,024 variants across 18,174 genes), Moderate (318,063 variants across 20,908 genes) and High (15,121 variants across 9,326 genes).

Statistical analysis. All analyses were conducted using the ASREML 4.1 software package (Gilmour *et al.*, 2015). Five GRMs were constructed, one including all variants and one for each of the four classes. Genetic parameters were obtained from univariate animal models in a GBLUP framework by fitting fixed effects of sex, contemporary group (concatenation of year, season of birth, and age of dam) and linear covariates of heterosis and breed fractions. Random effects included the additive genetic effect associated with each GRM and a residual. Models analyzed included fitting a single random effect associated with the GRM comprised of either Low, Modifier, Moderate, or High impact variants, models that included multiple random effects each associated with a different GRM (e.g. Low and Modifier, Moderate and High, all four subsets), and a model that included all variants in a single GRM (Full).

Results

Estimates of variance components and heritability for BWT and PWG are reported in Tables 1 and 2, respectively. Differences in heritability estimates were negligible with the exception of models that omitted Low and Modifier variants. Estimates from models that included a GRM based on Moderate or High impact variants, as well as the model that fitted both of these GRM, resulted in lower heritability estimates.

Table 1. Estimates of variance components and heritability (±SE) for birth weight using different genomic relationship matrices based on predicted functional impact.

GRM[1]	Additive variance (kg^2)	Residual variance (kg^2)	Heritability (h^2)
Full	21.66±3.12	31.32±2.39	0.41±0.05
Low	18.68±2.79	33.94±2.30	0.36±0.05
Modifier	20.38±2.82	32.56±2.24	0.39±0.05
Moderate	17.29±2.76	35.13±2.34	0.33±0.05
High	5.19±1.38	46.25±1.76	0.10±0.03
All subsets	21.30±2.95	31.71±2.34	0.40±0.05
Low & Modifier	21.30±2.95	31.70±2.34	0.40±0.05
Moderate & High	17.42±2.78	35.01±2.35	0.33±0.05

[1] Genomic relationship matrices, GRM; GRM containing a set of 1,145,892 variants with varying predicted impact on gene function, Full; GRM containing subset of 440,684 variants of low impact, Low; GRM containing subset of 372,024 variants of low or varying impact, Modifier; GRM containing subset of 318,063 variants of moderate impact, Moderate; GRM containing subset of 15,121 variants of high impact, High; model including four GRM corresponding to the four variant subsets, All subsets; model including a GRM for Low and a GRM for Modifier, Low & Modifier; model including a GRM for Moderate and a GRM for High, Moderate & High.

Table 2. Estimates of variance components and heritability (±SE) for post weaning gain using different genomic relationship matrices based on predicted functional impact.

GRM[1]	Additive variance (kg^2)	Residual variance (kg^2)	Heritability (h^2)
Full	320.20±49.57	555.91±41.21	0.37±0.05
Low	291.09±46.95	582.28±39.83	0.33±0.05
Modifier	299.37±47.37	578.66±39.47	0.34±0.05
Moderate	274.91±45.67	594.09±39.58	0.32±0.05
High	85.95±24.14	768.75±30.43	0.10±0.03
All subsets	320.82±49.63	555.62±41.25	0.37±0.05
Low & Modifier	316.01±49.11	561.03±40.77	0.36±0.05
Moderate & High	275.90±45.82	593.05±39.70	0.32±0.05

[1] Genomic relationship matrices, GRM; GRM containing a set of 1,145,892 variants with varying predicted impact on gene function, Full; GRM containing subset of 440,684 variants of low impact, Low; GRM containing subset of 372,024 variants of low or varying impact, Modifier; GRM containing subset of 318,063 variants of moderate impact, Moderate; GRM containing subset of 15,121 variants of high impact, High; model including four GRM corresponding to the four variant subsets, All subsets; model including a GRM for Low and a GRM for Modifier, Low & Modifier; model including a GRM for Moderate and a GRM for High, Moderate & High.

Discussion

Estimates of heritability from models fitting the Full GRM were 0.41±0.05 and 0.37±0.05 for BWT and PWG, respectively, both in the range of estimates previously reported for these traits in beef cattle (Koots *et al.*, 1994). As expected, the models with four additive genetic effects corresponding to the four GRM based on predicted functional impact provided a very similar estimate of heritability. Interestingly, only fitting the GRM comprised of low variants resulted in a negligible decrease in the additive variance compared to the Full GRM. Perhaps more interesting are results from fitting only the GRM comprised of high impact variants. These loss-of-function variants are predicted to substantially alter protein-coding genes and yet only accounted for ~24% of the additive genetic variation in BWT and ~27% in PWG in this population. Approximately 38% of annotated genes are affected by high impact variants, and the low amount of phenotypic variation explained by high impact variants could indicate that genes with similar function may compensate for genes affected by high impact alleles (El-Brolosy and Stainier, 2017). Similarly, Veerkamp *et al.* (2016) reported a marginal gain in the proportion of genetic variation explained when fitting a GRM based on full sequence data as compared to a GRM based on a common SNP array for traits in Holstein cattle despite identifying 42 variants that explained ~23% of the genetic variance when they were fitted alone. More recently, Warburton *et al.* (2020) reported results from a beef cattle population comparing the use of a single GRM with pre-selected WGS variants added to panel SNP and a multiple GRM model whereby panel SNP comprised one GRM and WGS variants comprised a second GRM. The multiple GRM model produced greater accuracy than the single GRM model in their study.

If reduced subsets of variants were to be selected for model inclusion, selecting from the High subset would be reasonable. However, additional information (i.e. pedigree or more global SNP representation) would be needed to fully capture additive genetic variation. The use of low-pass sequencing may allow for more robust molecular breeding values without increasing genotyping cost (Snelling *et al.*, 2020), and future research including different classifications for partitioning variants into GRM and expanding the current research across different traits complexes is needed.

References

Aken B.L., Ayling S., Barrell D., Clarke L., Curwen V., *et al.* (2016) Database 2016 baw093. https://doi.org/10.1093/database/baw093

Cingolani P., Platts A., Wang L.L., Coon M., Nguyen T., *et al.* (2012) Fly 6:80–92. https://doi.org/10.4161/fly.19695

El-Brolosy M.A., and Stainier D.Y.R. (2017) PLoS Genet. 13(7):e1006780. https://doi.org/10.1371/journal.pgen.1006780

Frischknecht M., Meuwissen T.H.E., Bapst B., Seefried F.R., Flury C., *et al.* (2018) J. Dairy Sci. 101:1292–1296. https://doi.org/10.3168/jds.2017-12890

Gilmour A.R., Gogel B.J., Cullis B.R., Welham S.J., and Thompson R. (2015). VSN International Ltd. Hemel Hempstead.

Heidaritabar M., Calus M.P.L., Megens H.J., Vereijken A., Groenen M.A.M., *et al.* (2016) J. Anim. Breed. Genet. 133:167–79. https://doi.org/10.1111/jbg.12199

Koots K.R., Gibson J.P., Smith C., and Wilton J.W. (1994) Animal Breeding Abstracts 62.

Loimpute-Public 2020 Available at: https://gitlab.com/gencove/loimpute-public.

Raymond B., Bouwman A.C., Schrooten C., Houwing-Duistermaat J., and Veerkamp R.F. (2018) Genet. Sel. Evol. 50:27. https://doi.org/10.1186/s12711-018-0396-8

Rosen B.D., Bickhart D.M., Schnabel R.D., Koren S., Elsik C., *et al.* (2020). GigaScience 9. https://doi.org/10.1093/gigascience/giaa021

Snelling W.M., Kuehn L.A., Keel B.N., Thallman R.M., Bennett G.L. (2017) Anim. Genet. 48:516–22. https://doi.org/10.1111/age.12579

Snelling W.M., Hoff J.L., Li J.H., Kuehn L.A., Keel B., *et al.* (2020) Genes 11:1312. https://doi.org/10.3390/genes11111312

Veerkamp R.F., Bouwman A.C., Schrooten C. and Calus M.P.L. (2016) Genet. Sel. Evol. 48:95. https://doi.org/10.1186/s12711-016-0274-1

Warburton C.L., Engle B.N., Ross E.M., Costilla R., Moore S., *et al.* (2020) Genet. Sel. Evol. 52(1):28. https://doi.org/10.1186/s12711-020-00547-5

628. *Bos indicus* genome percentage effects gene expression profile and differentially expressed genes in Brahman cattle whole blood

E.M. Ross[1*], L.T. Nguyen[1], B.N. Engle[1], A.J. Chamberlain[2], E. Kho[1] and B.J. Hayes[1]

[1]Centre for Animal Science, Queensland Alliance for Agriculture and Food Innovation, The University of Queensland, St. Lucia, QLD 4067, Australia; [2]Centre for AgriBiosciences, Agriculture Victoria, Bundoora, VIC 3083, Australia; e.ross@uq.edu.au

Abstract

RNA sequencing was used to quantify the gene expression in whole blood from 372 Brahman cattle. The percentage of the genome that was of *Bos indicus* origin (BI%) was calculated from genotype array data. BI% was associated with differential gene expression of 1,337 genes and could be significantly predicted from the gene expression matrix (prediction accuracy = 0.443). The effect of including BI% as a covariate in a differential gene expression model was then tested. Differentially expressed genes associated with weeks pregnant were identified with and without BI% in the model. 12% more genes were identified as differentially expressed when BI% was included in the model. This suggests that BI% should be included as a covariate in gene expression analysis of Brahman cattle, even when those traits are not phenotypically associated with BI%.

Introduction

RNA sequencing (RNAseq) is a sequencing method that is commonly used to perform differential gene analysis, where genes associated with phenotypic traits are identified. As sequencing costs decrease, the number of biological replicates that are used in differential gene expression studies is increasing. With this increase in biological replication more subtle phenotypic traits can be explored. The importance of appropriate co-variates for the correct identification of these traits is poorly understood.

B. indicus cattle are the result of a separate domestication event to *B. taurus* cattle. Modern Brahman cattle are a breed which was developed through the integration of both *B. indicus* and *B. taurus* breeds. This has resulted in modern Brahman cattle having a variable percentage of *B. indicus* content in their genomes, estimated on average around 9% (Koufariotis *et al.* 2018).

Here we tested the hypothesis that inclusion of *B. indicus* genome content in differential gene expression analysis of weeks pregnant would change the differentially expressed genes identified.

Materials & methods

Animals and phenotypes. Animal use in this study was approved under animal ethics number QAAFI/270/17. Animals were manually palpated by a trained veterinarian. Weeks pregnant were estimated by the veterinarian, with 0 being non-pregnant.

***Bos indicus* genome proportion.** A measure of *B. indicus* content expressed as a percentage (BI%) was calculated as an approximation of the proportion of each animals' genome that was of *B. indicus* origin, as opposed to *B. taurus* origin. BI% was calculated using the 35K SNP array by comparing each animal to a large purebred *B. indicus* dataset (Hayes *et al.*, 2019). Using a GBLUP model, phenotypes were assigned as 1 for *B. indicus* and 0 if not and the effect of each SNP was back-solved (Yang *et al.*, 2011). Prediction equations for *B. indicus* purebred then directly estimated BI% (Hayes *et al.*, 2019).

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_628, © E.M. Ross *et al.* 2022

RNA sequencing. Transcript levels were quantified in whole blood from 372 Brahman (*B. indicus*) cows and heifers collected in remote Queensland (Australia) using Qiagen RNAprotect Animal Blood Tubes and sequenced on an Illumina NovaSeq on a 2×100 bp run. RNA-seq reads were trimmed using QUADtrim v2.0.2 (Robinson and Ross 2019). Trimmed adapters, strings of Gs (an artifact of 2 dye Illumina chemistry) and bases with Q-score <20 from the start and end of reads and then filtered out reads with >5 poor quality bases (Q-score <20) and/or length <50bp. Only read pairs where both reads passed the trimming were used in the analysis.

To avoid mapping bias, the ARS-UCD1.2 (Rosen *et al.* 2020) *B. taurus* genome assembly merged with the Btau5 Y chromosome was masked at all known variant sites with an allele not present at that position. Sequencing reads were aligned to the genome with STAR v2.5.3 (Dobin *et al.* 2013) using the two-pass method. The abundance of each annotated gene was quantified using Subread v1.5.1 featureCounts (Liao *et al.* 2014) with the Ensembl_annotation v97 ARS-UCD1.2 and Btau5.0.1 Y GTF files which were merged to ensure gene identifiers on the Y chromosome were annotated correctly. Subsequent analysis was performed in R v4.1.0.

Statistics. To perform differential gene expression analysis edgeR (Robinson *et al.* 2010) was used with the glmQLFit method. Benjamini-Hochberg correction was used to control for multiple testing. Two separate models were fit to the gene expression matrix, one with BI% (model A), and one without BI% (model B). Both models also included RNA sequencing batch number (binary), blood sampling date (binary), lactation status (binary wet vs dry), whether the animal was a cow or heifer (binary), and weeks pregnant (integer).

The pseudo gene counts per million from edgeR (which includes correction for library size) were log transformed and divided by the mean log counts per million for each gene. A relationship matrix was then calculated using the *A.mat* function of the rrBLUP package (Endelman 2011). The resulting expression relationship matrix was used in BLUP model. 5-fold random cross validation was used to calculate the accuracy of the prediction.

Results

Gene expression of 372 Brahman cows and heifers was measured using RNA-seq of whole blood. Structure was visible in the relationship matrix that was generated from the expression profiles (Figure 1A), suggesting that certain animals had more similar expression profiles than others. The prediction ability of these similarities was then tested on two traits, BI% (Figure 1B) and weeks pregnant (Figure 1C). Weeks pregnant was not significantly associated with BI% ($P>0.05$). The accuracy of *B. indicus* content predicted from the expression profiles was r=0.443 ($P<0.001$); while prediction of weeks pregnant was not significant ($P>0.05$).

A differentially expressed gene analysis of 12,912 genes was then performed. BI% was associated with 1,337 total differentially expressed genes, the top three most significant genes that were annotated were D-dopachrome tautomerase (DDT), torsin family 1 member B (TOR1B), and agouti signalling protein (ASIP). Weeks pregnant was associated with 190 (model A) and 170 (model B) differentially expressed genes when BI% was included and excluded as a co-variate respectively. The two models from weeks pregnant had 163 genes in common. While the ranking according to most significant *P*-value was similar for the most significant genes in both models, genes with less significant *P*-values shows substantial re-ranking (Figure 1D). The three most significant genes for weeks pregnant were osteoglycin (OGN), mitogen-activated protein kinase 8 interacting protein 1 (MAPK8IP1), and solute carrier family 45 member 3 (SLC45A3).

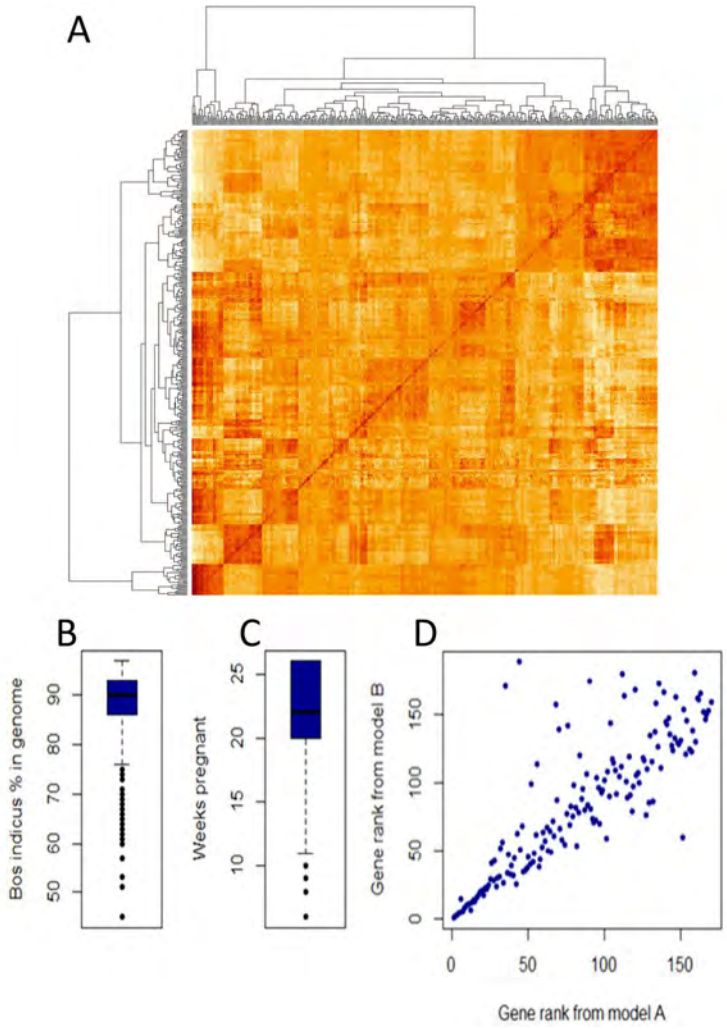


Figure 1. (A) Expression relationship matrix based on whole blood gene expression profiles of 372 Brahman females. (B) Distribution of *B. indicus* genome percentage in the animals used in the RNA-seq study. (C) Distribution of weeks pregnant in the pregnant females used in the RNA-seq study. (D) Order of significance (most significant = 1, least significant = 163) of the differentially expressed genes from both 'weeks pregnant' models: model A includes BI% as a covariate; model B does not include BI% as a covariate.

Discussion

Brahman cattle are considered to be genomically indicine cattle, however the distribution of the amount of *B. indicus* content in the genomes of the animals that were used for this study, which were all registered Brahman from Queensland Australia, was highly variable.

Here we found that BI% had a large effect on blood gene expression, to the extent that BI% could be predicted from the gene expression profile. Given the large effect of BI% on gene expression we tested its effect on our ability to identify genes that were associated with weeks pregnant. Inclusion of BI% in the model resulted in a 12% increase in the number of differentially expressed genes for weeks pregnant, which

while not dramatic, does suggest that inclusion of BI% in gene expression models, even for traits which are not phenotypically correlated with BI%, is beneficial.

Additionally, this is the first report of a BLUP based prediction being used with a gene expression relationship matrix. BLUP models have been extensively used for genomic predictions (Meuwissen *et al.* 2001), and have also been used for metagenomic predictions (Ross *et al.* 2013) and epigenetic predictions (Hayes *et al.* 2021). This finding emphasizes the flexibility of relationship matrix based predictions and could be useful in the identification of biomarkers for traits that involve gene expression variation in many genes.

References

Dobin A., Davis C.A., Schlesinger F., Drenkow J., Zaleski C. *et al.* (2013) Bioinformatics 29(1): 15-21. https://doi.org/10.1093/bioinformatics/bts635

Endelman J.B. (2011) The plant genome 4(3):250-255. https://doi.org/doi.org/10.3835/plantgenome2011.08.0024

Hayes B.J., Fordyce G. and Landmark S. (2019) Proc. Of Australian Association for Animal Breeding and Genetics, Armidale, Australia.

Hayes B.J., Nguyen L.T., Forutan M., Engle B.N., Lamb H.J. *et al.* (2021) Front. genet. 12:760450-760450. https://doi.org/10.3389/fgene.2021.760450

Koufariotis L., Hayes B.J., Kelly M., Burns B.M., Lyons R. *et al.* (2018) Sci. Rep. 8(1): 17761. https://doi.org/10.1038/s41598-018-35698-5

Liao Y., Smyth G.K. and Shi W. (2014) Bioinformatics 30(7): 923-930. https://doi.org/10.1093/bioinformatics/btt656

Meuwissen T.H.E., Hayes B.J. and Goddard M.E. (2001) Genetics 157(4): 1819-1829.

Robinson A.J., and Ross E.M. (2019) bioRxiv: 2019.2012.2018.870642. https://doi.org/10.1101/2019.12.18.870642

Robinson M.D., Mccarthy D.J. and Smyth G.K. (2010) Bioinformatics 26(1): 139-140. https://doi.org/10.1093/bioinformatics/btp616

Rosen B.D., Bickhart D.M., Schnabel R.D., Koren S., Elsik C.G. *et al.* (2020) Gigascience 9(3). https://doi.org/10.1093/gigascience/giaa021

Ross E.M., Moate P.J., Marett L.C., Cocks B.G. and Hayes B.J. (2013) PLoS One 8(9): e73056. https://doi.org/10.1371/journal.pone.0073056

Yang J., Lee S.H., Goddard M.E. and Visscher P.M. (2013) Methods Mol Biol 1019:215-236. https://doi.org/10.1007/978-1-62703-447-0_9

629. Effect of one copy of nt821 mutation in myostatin gene on ultrasound carcass traits in Senepol cattle

G.R.O. Menezes[1*], C.B. Souza[2], A.L.R. Lima[3], A. Gondo[1], F. Siqueira[1], A.A. Egito[1], L.N. Silva[4], J.A. Fernandes Junior[5] and R.A.A. Torres Junior[1]

[1]Embrapa Beef Cattle, Radio Maia Avenue, 830, 79106550, Campo Grande, Brazil; [2]Dom Bosco Catholic University, Tamandaré Avenue, 6000, 79117900, Campo Grande, Brazil; [3]State University of Mato Grosso do Sul, Graziela Maciel Barroso Highway, Km 12, 79200000, Aquidauana, Brazil; [4]Embrapa Geneplus Program, Radio Maia Avenue, 830, 79106550, Campo Grande, Brazil; [5]Grama Senepol, Marechal Rondon Highway, Km 404, 16600000, Pirajuí, Brazil; gilberto.menezes@embrapa.br

Abstract

Several mutations in the myostatin gene have been identified that explain the increased muscling in cattle, one of which is an 11-base pair deletion in the gene, known as nt821 mutation. As several studies have indicated a significant effect of this mutation on traits related to carcass and meat quality, this study aimed to evaluate the effect of one copy of nt821 mutation on ultrasound carcass traits in Senepol cattle. Records from 2,044 purebred Senepol cattle from commercial performance tests in Brazil were used. Phenotypes and expected progeny difference (EPD) for ribeye area, backfat thickness, and intramuscular fat percentage were analyzed. The effect of one copy of nt821 mutation was significant ($P<0.05$) for all traits, both for phenotypes and EPD. Heterozygous animals presented a larger area for *Longissimus dorsi*, while their carcasses were leaner with less subcutaneous and intramuscular fat compared to the ones with no copy of the mutation.

Introduction

The myostatin gene was identified in cattle as responsible for double musculature or muscle hypertrophy in 1997 (McPherron and Lee, 1997). Several distinct mutations have been identified that explain the increased muscling in cattle, one of which is an 11-base pair deletion in the third exon of the gene that results in the production of a truncated inactive protein (Casas *et al.*, 2004). This mutation, known as nt821, has been associated with muscle hypertrophy in several breeds (eg. Dunner *et al.*, 1997; Globet *et al.*, 1998).

Since 2018, genetic tests for the nt821 mutation have been carried out in the Senepol breed in Brazil, allowing the identification of heterozygous animals carrying a copy of the mutant variant. There are several studies in the literature for different beef cattle breeds that indicate a significant effect of the nt821 mutation on traits related to carcass and meat quality (e.g. Casas *et al.*, 2004; Allais *et al.*, 2010). Therefore, as the Senepol breed is a bovine genetic resource used for the efficient production of high-quality meat under tropical conditions, it is important to understand how the nt821 mutation affects traits linked to carcass and meat quality to support better selection decisions by breeders. Thus, this study aimed to evaluate the effect of one copy of nt821 mutation in the myostatin gene on ultrasound carcass traits in Senepol cattle.

Materials & methods

Records from 2,044 purebred Senepol individuals (*Bos taurus taurus*) were used. The data were obtained from a compilation of commercial performance tests coordinated by Embrapa Geneplus Program in Brazil, between 2011 and 2020. The animals finished the tests with an average weight of 437±63 kg and age of 574±65 days. The animals had *ad libitum* access to diet and water.

The studied traits were recorded at the end of the tests on live animals by trained technicians. The carcass traits collected by ultrasonography in the region between the 12th and 13th ribs in the *Longissimus dorsi* muscle were: ribeye area (REA – cm^2), backfat thickness (BF – mm) and percentage of intramuscular fat

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_629, © G.R.O. Menezes *et al.* 2022

(IMF – %). All animals had valid phenotypes for three traits. In addition to phenotypic data, expected progeny difference (EPD) for all individuals and traits were used. The EPDs for ribeye area (REAg – cm^2), backfat thickness (BFg – mm), and percentage of intramuscular fat (IMFg – %) were obtained from the 2021's Brazilian Senepol Genomic Evaluation conducted by Embrapa Geneplus Program and Brazilian Senepol Breeders Association (Geneplus Senepol, 2021). The single-step GBLUP methodology (Aguilar *et al.*, 2010) has been adopted for evaluating Senepol cattle in Brazil since 2019.

All animals were genotyped using GGP Bovine 50k from NEOGEN®, which allowed better prediction of breeding values as well as to assess the number of copies of the mutation known as nt821 in the myostatin gene (McPherron and Lee, 1997). Table 1 shows the distribution of the animals in this study according to sex and genotype for nt821 mutation.

Data were analyzed by a fixed-model ANOVA with the GLM procedure (SAS, SAS onDemand for Academics, Cary, NC). For phenotypes, fixed effects of contemporary group (CG) and genotype for nt821 mutation were included in a statistical model for all variables, as well as the age of the animal at the end of the test as a linear covariate. For EPDs, a similar model was adopted, except for not including the fixed effect of CG. The CGs were defined as birth year/season, sex, test edition, and farm of origin of the animal. The Tukey's test was used to compare the means for all variables, considering 0.05 as the significance level.

Results

In Table 2, the least square means for REA, BF and IMF are presented for animals with one or no copy of nt821 mutation in the myostatin gene. The effect of one copy of nt821 mutation was significant for all variables ($P<0.05$). Heterozygous animals presented a larger area for *Longissimus dorsi*, while their carcasses were leaner with less subcutaneous and intramuscular fat.

In Table 3, the least square means for REAg, BFg and IMFg are presented for animals with one or no copy of nt821 mutation in the myostatin gene. The effect of one copy of nt821 mutation was significant for all variables ($P<0.05$). Heterozygous animals presented greater genetic merits for producing more muscular and leaner carcasses.

Table 1. Description of studied animals according to sex and genotype for n821 mutation.

Genotype[1]	Sex		Total
	Male	Female	
mh+	92	224	316
++	494	1,234	1,728
Total	586	1,458	2,044

[1] mh+ = One copy of nt821 mutation; ++ = No copy of nt821 mutation.

Table 2. Effects of one copy of nt821 mutation on phenotypes for ultrasound carcass traits in Senepol cattle.

	Genotype[2]		*P*-value
Variables[1]	mh+	++	
REA (cm^2)	73.05±0.49	69.20±0.25	<0.0001
BF (mm)	5.02±0.09	5.45±0.05	<0.0001
IMF (%)	2.42±0.05	2.70±0.03	<0.0001

[1] REA = phenotype for ribeye area; BF = phenotype for backfat thickness; IMF = phenotype for intramuscular fat.
[2] mh+ = One copy of nt821 mutation; ++ = No copy of nt821 mutation.

Table 3. Effects of one copy of nt821 mutation on expected progeny differences (EPD) for ultrasound carcass traits in Senepol cattle.

	Genotype[2]		*P*-value
Variables[1]	mh+	++	
REAg (cm^2)	0.7473±0.0789	-0.1809±0.0337	<0.0001
BFg (mm)	-0.0689±0.0224	0.0921±0.0096	<0.0001
IMFg (%)	-0.0293±0.0072	0.0072±0.0030	<0.0001

[1] REAg = EPD for ribeye area; BFg = EPD for backfat thickness; IMFg = EPD for intramuscular fat.

[2] mh+ = One copy of nt821 mutation; ++ = No copy of nt821 mutation.

The effect of one copy of nt821 mutation was significant ($P<0.05$) on ultrasound carcass traits regardless of their types: phenotypes or EPDs (Table 2 and 3).

Discussion

Heterozygous animals for the nt821 mutation showed greater REA (Tables 2 and 3), which has been related to higher commercial cuts yield and carcass weight (e.g. Silva *et al.*, 2012). Casas *et al.* (2004) and Allais *et al.* (2010) found similar results to this study with crossbred cattle and French beef breeds, respectively. In Brazil, in the Senepol breed, homozygous individuals for the nt821 mutation are discarded from breeding, according to the recommendation of the Brazilian Senepol Breeders Association. However, heterozygotes are accepted, which may represent an opportunity to improve production systems with a focus on increasing meat productivity, as half of their progenies will be carriers of the nt821 mutation.

Regarding fat deposition in the carcass, the effects of a copy of the nt821 mutation are antagonistic to those found for REA. The results show that heterozygotes have leaner carcass with poorer carcass finishing and intramuscular fat (Table 2 and 3). Similar to REA, studies carried out by Casas *et al.* (2004) and Allais *et al.* (2010) corroborate our findings. To meet markets that want lean meat, the use of animals carrying the nt821 mutation may be desirable. In Brazil, however, the predominant demand is for carcasses with adequate fat cover, which protects against rapid cooling and consequently prevents cold-shortening that decreases meat tenderness. Additionally, there is a tendency to prioritize cuts with an adequate level of marbling, which has a direct impact on the juiciness and flavor of the meat.

The Senepol breed is a taurine genetic resource option for the efficient production of high-quality meat under tropical conditions. In Brazil, it has been used especially in crossbreeding programs with zebu breeds with the aim of improving productivity and, above all, meat quality. Therefore, the results found are important to support breeders in decision-making processes regarding the role of Senepol genetics in beef production. Additionally, there are reports of the negative effects of muscle hypertrophy on survival and reproductive traits (Fiems, 2012), which were not investigated here, that should not be overlooked.

References

Aguilar, I., Misztal, I., Johnson, D. L., Legarra, A., Tsuruta, S., *et al.* (2010). J. Dairy. Sci., 93 (2): 743–752. https://doi.org/10.3168/jds.2009-2730.

Allais, S., Levéziel, H., Payet-Duprat, N., Hocquette, J. F., Lepetit, J., *et al.* (2010). J. Anim. Sci., 88(2): 446-454. https://doi.org/10.2527/jas.2009-2385.

Casas, E., Bennett, G. L., Smith, T. P., Cundiff, L. V., (2004). J. Anim. Sci., 82 (10): 2913-2918. https://doi.org/10.2527/2004.82102913x.

Dunner, S., Charlier, C., Fanir, F., Brouwers, B., Canon, J., *et al.*, (1997). Mamm. Genome, 8 (6): 430-435. https://doi.org/10.1007/s003359900462.

Embrapa Geneplus Program for Senepol Breed (Geneplus Senepol), The Brazilian Senepol Genomic Evaluation – year 2021. Available at: https://geneplus.com.br/sumario-senepol-2021/.

Fiems, L. O. (2012). Animals, 2 (3): 472-506. https://doi.org/10.3390/ani2030472.

Grobet, L., D. Poncelet, L. J. Royo, B. Brouwers, D. Pirottin, C., *et al.* (1998). Mamm. Genome, 9(3): 210-213. https://doi.org/10.1007/s003359900727.

McPherron, A. C., Lee, S. J. (1997). Proc. of PNAS, 94 (23): 12457-12461. https://doi.org/10.1073/pnas.94.23.12457.

Silva, S. L., Tarouco, J. U., Ferraz, J. B. S., Gomes, R. C., Leme, P.R., *et al.* (2012). Braz. J. Anim. Sci., 41 (9): 2025–2031. https://doi.org/10.1590/S1516-35982012000900009.

630. Allelic diversity and effect of the intronic TBX20-191081 marker on temperament in beef cattle

V.M. Del Real-García[1], A.M. Sifuentes-Rincón[2*], F.A. Paredes-Sánchez[3], G. Ruíz-de la Cruz[2], E. Garza-Brenner[1], G.M. Parra-Bracamonte[2], H. Bernal-Barragán[1] and F.A. Rodríguez-Almeida[4]

[1]Facultad de Agronomía, Universidad Autónoma de Nuevo León, 66050 General Escobedo, Nuevo León, México; [2]Centro de Biotecnología Genómica, Instituto Politécnico Nacional, 88710 Reynosa, Tamaulipas, México; [3]Unidad Académica Multidisciplinaria Mante, Universidad Autónoma de Tamaulipas, 89840 El Mante, Tamaulipas, México; [4]Facultad de Zootecnia y Ecología, Universidad Autónoma de Chihuahua, 31453 Chihuahua, Chihuahua, México; asifuentes@ipn.mx

Abstract

Temperament in cattle is defined as the behavioural response of animals to human handling. Its expression is influenced by environmental and genetic factors. TBX20 is a candidate gene associated with bovine temperament. The objective was to determine the genetic variability of TBX20-191081 in three cattle beef breeds and evaluate its effect on temperament traits. The allelic frequencies confirmed both alleles segregation in the studied breeds and a significant allelic substitution effect for the G allele was observed, which decreases exit velocity (EV) and temperament score (TS). Thus, its association with docility of beef cattle is evident and genotyping for the marker early in life (e.g. neonates) may allow more flexibility in breeding and selection strategies.

Introduction

Cattle farming is one of the most developed and practiced livestock activities in Mexico. More than 50% of the national land is used to raise around 32 million head of cattle (SIAP, 2020). Temperament is a complex and economically relevant trait; it reflects the well-being of the animal and has been found to be related to productivity traits and job safety (Burdick *et al.*, 2010); it is closely related to gene expression at different levels (Cooke *et al.*, 2014). Few studies have explored characteristics of cattle temperament at the genomic level, mainly due to its biological complexity, difficulties in the management of animals, and statistical analyses. However, as a result of applying different approaches such as the scanning of complete genomes and the use of candidate gene strategy, now it is known that regulation of temperament involves multiple genes with divergent functions. Using genome wide association analyses, we identified candidate genes with the potential to significantly affect temperament in Brahman cattle (Paredes-Sánchez *et al.*, 2019). Among these genes, the marker TBX20-191081 (rs43398069) located within an intron of the T-box transcription factor gene was significantly associated with exit velocity, pen score and temperament score, methods commonly used to precisely score cattle temperament (Paredes-Sánchez *et al.*, 2019). This study aimed to estimate the genetic effect and allele frequencies of the genetic marker in three different beef cattle breeds: Angus, Brangus and Charolais.

Materials & methods

Hair samples and temperament records of 89 yearling bulls in a central performance test (35 Angus and 27 Brangus and 27 Charolais cows were included. The data and samples were collected by Garza-Brenner (2017) in the facilities of the Unión Ganadera Regional de Chihuahua (UGRCH), Chihuahua, México and in particular herds located at Nuevo León, Mexico, and kept in the Centro de Biotecnología Genómica-Instituto Politécnico Nacional (CBG-IPN), Reynosa, Tamaulipas, México.

Three methods were used to evaluate the animals temperament: (1) pen score (PS), which is a non-restrictive technique, in which the cattle are segregated into small groups (three to five animals) in a pen

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_630, © V.M. Del Real-García *et al.* 2022

(Burdick *et al.*, 2010) to score their reaction when approached by an evaluator; (2) exit velocity (EV), considered as an objective quantitative measure, because it is calculated by an electronic device (Schmidt *et al.*, 2014), defined as the time required for an animal to cross 6 feet (1.83 m) of distance when released from the handling press; and (3) temperament score (TS), that classifies animals into calm and temperamental combining data from PS and EV (King *et al.*, 2006; Burdick *et al.*, 2010; Schmidt *et al.*, 2014).

Genotyping. DNA isolation from hair follicles was achieved using the Genelute Mammalian Genomic DNA kit (Cat. G1N350, Sigma-Aldrich Co. LLC, St. Louis, Missouri, USA). PCR was used to obtain a 298 pb product including the TBX20-191081 marker. Allelic discrimination was achieved using the Nextera XT protocol and sequencing in the MiniSeq system (Illumina, Inc., San Diego, California, USA). The sequences obtained were subjected to bioinformatics analysis for genotype-assignation to each sample. The allele frequencies were calculated with the GENEPOP online program version 4.0.

Effect of 191081-TBX20 marker. A mixed model was fitted, considering the fixed effects of genotype (AA, GA, GG), sex of the animal (male, female) and breed (Angus, Brangus and Charolais). Since animals were from several herds, this effect was considered as a random. The least squares mean for breed, sex, and genotype were estimated for each variable and compared with the least significant difference test. Additionally, the allelic substitution effect of the markers was estimated, considering the genotype as a linear covariate. Analyses were run using the software SAS on Demand for Academics (SAS Institute Inc., Cary NC, USA).

Results

Compared with A allele, a lower frequency of the G allele, previously reported as the favourable allele for docility, was found in the three tested breeds (Table 1).

A significant effect of breed was observed, in a pooled analysis across breeds. No significant differences were found among genotype means for each behavioural trait (Table 2). However, the allelic substitution effect showed significance for EV and TS, with a negative value for the G allele (Table 2).

Table 1. Allele and genotype frequencies for the TBX20-19108 marker in the studied sample (n=89).

Breed	Genotype			Alleles	
	AA	AG	GG	A	G
Angus	0.48	0.40	0.11	0.68	0.31
Brangus	0.25	0.51	0.22	0.51	0.48
Charolais	0.40	0.29	0.29	0.55	0.44

Table 2. Least square means ± standard error and allelic substitution effect (*a*) of genotypes of 191081 locus of the TBX20 gene effects on temperament traits.[1]

Trait	AA	AG	GG	*a*
	(35)	(36)	(18)	
EV	1.73±0.15	1.91±0.15	2.14±0.18	-0.20±0.09**
PS	1.80±0.10	1.90±0.11	1.98±0.12	-0.09±0.05*
TS	1.77±0.13	1.91±0.13	2.06±0.15	-0.15±0.07**

[1] EV = exit velocity; PS = pen score; TS = temperament score. Sample size between brackets. * = $P<0.10$, ** = $P<0.05$.

Discussion

The T-box transcription factor gene (TBX20) has been studied in human and non-human species and proposed as candidate gene associated with Heart Rate (HR) (quantification of the number of times the heart beats in a given time interval, usually a minute, while the unstressed body is at rest). In cattle, some evidence has shown that temperament and HR are correlated. Curley *et al.* (2008) found that cortisol concentration in temperamental Brahman heifers was greater than in calmer animals, assessed using exit velocity. Sutherland *et al.* (2012) reported that High Responder cows had higher basal HR and cortisol concentrations compared with Low Responder cows, which may hint at the relationship between exit velocity-TBX20- and heart rate.

Here, we demonstrated in a small sample, that allelic frequencies distribution of intronic TBX20-19108 between beef breeds are similar to those found in Brahman cattle (Paredes-Sánchez *et al.,* 2019) and also that an effect of allelic substitution is observed for EV and TS temperament traits.

Once confirmed in a large cattle sample, the results of this research may be important not only for basic research but also for the application of genomic tools for the improvement of beef cattle in the medium term, given that temperament is an economically relevant trait for beef cattle production. Identification of markers that allow detection of temperamental animals early in life (e.g. neonates) will allow the producer more flexibility in their breeding and selection strategies.

References

Burdick N. C., Carroll J. A., Hulbert L. E., Dailey J. W., Willard S. T., *et al.* (2010). Livestock Science, 129(1–3): 166–172. https://doi.org/10.1016/j.livsci.2010.01.020

Cooke R. F. (2014). Journal of Animal Science, 92(12): 5325–5333. https://doi.org/10.2527/jas.2014-8017

Garza-Brenner E. (2017). Análisis genético molecular del temperamento en ganado bovino: búsqueda y asociación de polimorfismos en genes candidatos. Instituto Politécnico Nacional-Centro de Biotecnología Genómica. Reynosa, Tamaulipas, Mexico, PhD thesis.

http://rdcb.cbg.ipn.mx/handle/20.500.12273/640

Curley Jr K. O., Paschal J. C., Welsh Jr T. H., Randel R. D. (2006). Journal of Animal Science. 84: 3100-3103. https://doi.org/10.2527/jas.2006-055

King D. A., Schuehle Pfeiffer C. E., Randel R. D., Welsh T. H., Oliphint R. A., *et al.* (2006). Meat Science, 74(3): 546–556. https://doi.org/10.1016/j.meatsci.2006.05.004

Paredes-Sánchez F. A., Casas E., Parra-Bracamonte G. M., Arellano-Vera W., Riley, D. G., *et al.* (2019). Animal Science, 97 (3). https://doi.org/10.1093/jas/skz258.002

Schmidt S. E., Neuendorff D. A., Riley D. G., Vann R. C., Willard S. T., *et al.* (2014). Journal of Animal Science, 92(7): 3082–3087. https://doi.org/10.2527/jas.2013-7494

Sutherland M. A., Rogers A. R., Verkerk G. A. (2012) Journal of Animal Science, Physioly and behavior, 107: 329-337. https://doi.org/10.1016/j.physbeh.2012.07.013

631. Innovative phenotyping strategies exploiting immune response to heat stress conditions

P. Lemal*, M. Schroyen and N. Gengler

Gembloux Agro-Bio Tech, University of Liège, Passage des Déportés 2, 5030 Gembloux, Belgium; pauline.lemal@uliege.be

Abstract

Evaluating the heat stress status of an animal is a challenge. Indeed, currently, the heat stress status is determined on the basis of a significative change of a response variable known to be correlated with a temperature-humidity related variable. These response variables are, most often, milk yield or similar traits. However, it is known that heat stress has a negative effect on the immune system, but this information is still very little used in the context of heat stress phenotyping. Indeed, only a few studies highlight immune genes involved in heat stress. In addition, these studies provide very little information on the molecular mechanisms that link these genes to the observed phenotype. Immune response based deep-phenotyping strategy would provide a more comprehensive understanding of heat stress and could also serve as a basis for identifying heat stress indicators through genomics.

Introduction

Currently, phenotypes used for genetic improvement remain often defined in a very straightforward fashion, e.g. to improve production traits, these traits are recorded. Recently, interest for breeding of resilient and healthy livestock has strongly increased. Indeed, the idea has emerged that by selecting animals based on immune traits, it would be possible to reduce the occurrence of diseases and thus also increase profitability of the farm (Mallard *et al.*, 2015; Thompson-Crispi *et al.*, 2014). In addition, it is shown that, generally, healthier animals show better growth and better reproductive abilities (Mallard *et al.*, 2015). However, in this condition direct phenotyping quickly reaches its limits as disease traits are manifestations of the reaction of the animal to underlying mostly unknown solicitations. Moreover, direct recording of this data is notoriously unreliable and biased as only recognized positive cases are recorded and the quantity of false-negative cases is expected to be high.

Recently, some studies have developed technologies allowing to predict cell mediated response and humoral immune response of an animal. Even if this type of studies is still rather rare, the heritability for the immune response was estimated to be between 0.25 and 0.35 which was clearly higher than traditional disease traits and at a level where important responses to selection can be expected (Mallard *et al.*, 2015; Thompson-Crispi *et al.*, 2012). However, these technologies can hardly be used to determine the immune status of an animal in real time.

Heat stress is known to have a negative effect on the immune system (Bagath *et al.*, 2019; Dahl *et al.*, 2020). However, this knowledge is not currently used as an indicator of heat stress. The most used approach is to relate a response variable, often milk yield or similar traits, to a temperature-humidity related variable. The status of heat stress is then associated with a significant change in this response variable, i.e. decrease for production related traits, or an increase for traits like somatic cell count. The temperature at which this decrease, or increase, begins, and its slope vary from animal to animal. In current implementations, reaction norm models are used with, in general, a heat stress threshold set and only the slope being assessed for each animal. The already implemented genomic breeding value estimation for thermotolerance is based on this type of phenotypes (Nguyen *et al.*, 2017).

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_631, © P. Lemal *et al.* 2022

Only a few studies highlight the genes and pathways that could be involved in thermotolerance (Garner *et al.*, 2020; Halli *et al.*, 2021; Otto *et al.*, 2019; Sigdel *et al.*, 2019; Srikanth *et al.*, 2017). And, even if these studies provide a lot of information, few of those are looking at the molecular mechanisms which lead to the observed phenotype, and therefore could serve as basis for the definition of indicators (i.e. biomarkers) of heat stress. For this reason, the objective of the present paper was to extend the work of Otto *et al.* (2019) toward innovative potential deep-phenotyping strategies exploiting immune response in the context of heat stress.

Innovative deep phenotyping strategies for biomarkers of heat stress

From heat stress to cortisol secretion. In the event of heat stress, the paraventricular nucleus of the hypothalamus will produce the corticotropin-releasing hormone (CRH) (Mishra, 2021). This hormone will bind to the CRH receptors expressed on the surface of the anterior pituitary cells, which triggers exocytosis of the adrenocorticotrophic hormone (ACTH) stored in secretory vesicles (Rehfeld and Bundgaard, 2010). ACTH enters the adrenal gland through the bloodstream and stimulates cortisol synthesis (Feher, 2012). Once released into the bloodstream, cortisol is able to passively cross the plasma membrane of target cells. Its receptor, the glucocorticoid receptor (GR) is located in the cytoplasm where it is sequestered by several proteins including HSP90. The binding of cortisol to the GR causes its dissociation from this protein complex and its translocation to the nucleus. The GR can then bind to the glucocorticoid response element (GRE) regions of the promoter of the target genes. The GR-cortisol complex is also able to modulate the function of other transcription factors by interacting directly with them (Nicolaides *et al.*, 2020).

From cortisol secretion to greater occurrence of diseases. Among the target cells of cortisol are immune cells including macrophages. These phagocytic cells exhibit a wide variety of phenotypes *in vivo*. This panel of phenotypes has two extremes, M1 and M2. The former is associated with the initiation of inflammation and the eradication of pathogens and the latter with the inhibition of inflammation and the prevention of damages (Maciuszek *et al.*, 2020). The polarization of macrophages in M1 is initiated, among other, by lipopolysaccharide (LPS) of bacterial origin. This is because LPS binds to the CD14 receptor expressed on the surface of macrophages. The complex formed then interacts with Toll like receptor 4 (TLR4) to trigger a signaling cascade dependent on p50/p65 nuclear factor kappa B (NFκB). This leads to the activation of gene transcription encoding proteins involved in the antimicrobial activity of macrophages such as the inducible nitric oxide synthase (iNOS) as well as pro-inflammatory cytokines such as tumor necrosis factor-α (TNF-α) (Tugal *et al.*, 2013).

However, after binding to its receptor, cortisol is capable of sequestering the p65 subunit of NFκB and inducing the expression of IκB, an inactivator of NFκB (Feher, 2012; Nicolaides *et al.*, 2020). Moreover, the glucocorticoid-induced leucine zipper (GILZ) whose expression is controlled by the glucocorticoid receptor, is also capable of inhibiting NFκB (Cannarile *et al.*, 2019). In this way, heat stress appears to prevent M1 polarization and thus promote M2 polarization which is associated with reduced antimicrobial activity.

In addition, NFκB was also shown to be involved in the activation of T and B lymphocytes (Haselager *et al.,* 2020; Krawczyk and Penninger, 2014). The type of cytokines produced by dendritic cells are also modulated by the activation of NFκB (Walsh and Mills, 2013). All these data suggest that heat stress through, among others, the production of cortisol and inactivation of NFκB, may lead to a weaker immune response against pathogens by reducing the ability of immune cells to react to infection.

From heat stress to pro-inflammatory response. However, cortisol levels vary greatly with the time of exposure to heat stress (Mishra, 2021). Chronic heat stress is even associated with increased inflammation, possibly as a result of decreased cortisol levels in affected animals (Most and Yates, 2021). The increased inflammation in the absence of cortisol could be due to cell death induced by increased body temperature as well as the overexpression of heat shock proteins (HSPs). Indeed, in cells such as macrophages which express TLR4 and CD14, HSPs have been shown to be able to generate a pro-inflammatory response like other molecules released during cell damage. In this way, this pro-inflammatory response is possible via the activation of NFκB (Asea and Kaur, 2017).

From pro-inflammatory response to greater occurrence of diseases. It has been shown that with greater inflammation the symptoms of a disease can also be greater (Hunter, 2012). In this way, if the heat stress induces an increase in inflammation, it would not lead to an increase in the occurrences of infectious disease but could cause more severe symptoms which would cause more animals to pass from subclinical to clinical level.

From deep phenotyping to identification of heat stress markers. On this basis, it is expected that NFκB activity and therefore the expression of its target genes will be different between stressed and unstressed animals. This difference could then be exploited as a marker of heat stress.

From biomarkers identification to genetic progress for heat stress resilience. Translating heat stress markers into genetic progress requires several conditions. First, deep phenotypes must be acquired in sufficient quantities and on relevant animals to serve as a basis for heat stress indicators. Indeed, it is necessary to confirm the heat stress status of the animals to find reliable markers of heat stress based on these animals. Based on Otto *et al.* (2019), these markers could be target genes of NFκB. Of course, as already mentioned, significant differences must be obtained between the groups formed for the markers tested. Ideally, these markers should also be validated to determine whether they are capable, on their own, of discriminating between stressed and non-stressed individuals and be considered as biomarkers (Ou *et al.*, 2021). In the case of selection purpose, once biomarkers obtained, it is still required to link variations of these indicators to genetic variation. Moreover, this genetic variation has to be additive, therefore heritable and can be improved by selection. Many strategies are possible to implement biomarkers for selection. These can be very traditional approaches where immune response-based biomarkers are added in multi-trait settings. Indeed, these precise biomarkers not available in large quantities could be added with less precise traits but available in larger quantities. Moreover, very innovative approaches using novel methods as machine learning and these biomarkers as input could be used to generate predictions of heat stress resilience.

Conclusions

We followed the direction shown by Otto *et al.* (2019), who have identified among others, that NFκB as a transcription factor is associated with heat stress. Extending on this study, we showed how it is possible to go further and directly link NFκB to heat stress and to the higher occurrence of infectious diseases which results from it. This is far from the only opportunity to perform deep innovative phenotyping exploiting immune response to heat stress conditions. Indeed, the information given above is, of course, far from exhaustive, but it does allow to link genomic studies to molecular mechanisms, and therefore also to potential molecular phenotypes (i.e. biomarkers) associated with heat stress. Linking studies in this systematic way should allow a better understanding of the mechanisms involved at the systemic level but also facilitate the identification of biomarkers. Heritable variation of these identified biomarkers could subsequently be used in animal breeding to generate genetic progress for heat stress resilience.

References

Asea A. A., and Kaur P. (2017). Heat shock proteins in veterinary medicine and sciences (Vol. 12). Springer, Berlin, DE.

Bagaev A. V., Garaeva A. Y., Lebedeva E. S., Pichugin A. V., Ataullakhanov R. I., *et al.* (2019). Sci. Rep., 9(1), 4563. https://doi.org/10.1038/s41598-018-36052-5

Bagath M., Krishnan G., Devaraj C., Rashamol V. P., Pragna P., *et al.* (2019). Res. Vet. Sci., 126, 94-102. https://doi.org/10.1016/j.rvsc.2019.08.011

Cannarile L., Delfino D. V., Adorisio S., Riccardi C., and Ayroldi E. (2019). Front. immunol, 10, 1823. https://doi.org/10.3389/fimmu.2019.01823

Dahl G. E., Tao S., and Laporta J. (2020). Front. vet. sci, 7, 116. https://doi.org/10.3389/fvets.2020.00116

Feher J. (2012). Quantitative Human Physiology. Elsevier, Cambridge, US.

Garner J. B., Chamberlain A. J., Vander Jagt C., Nguyen T. T. T., Mason B. A., *et al.* (2020). Sci. Rep., 10(1), 19181. https://doi.org/10.1038/s41598-020-75438-2

Halli K., Vanvanhossou S. F., Bohlouli M., König S., and Yin T. (2021). PLOS ONE, 16(10), e0258216. https://doi.org/10.1371/journal.pone.0258216

Haselager M. V., Kater A. P., and Eldering E. (2020). Front. Oncol., 10, 592205. https://doi.org/10.3389/fonc.2020.592205

Hunter P. (2012). EMBO reports, 13(11), 968-970. https://doi.org/10.1038/embor.2012.142

Krawczyk C., and Penninger J. M. (2001). J. Leukoc. Biol., 69(3), 317-330. https://doi.org/10.1189/jlb.69.3.317

Maciuszek M., Klak K., Rydz L., Verburg-van Kemenade B. M. L., and Chadzinska M. (2020). Int. J. Mol. Sci., 21(23), 8954. https://doi.org/10.3390/ijms21238954

Mallard B. A., Emam M., Paibomesai M., Thompson-Crispi K., and Wagter-Lesperance L. (2015). Jpn. J. Vet. Res, 63(Supplement 1), S37-S44. https://doi.org/10.14943/jjvr.63.suppl.s37

Mishra S. R. (2021). Trop. Anim. Health Prod., 53(3), 400. https://doi.org/10.1007/s11250-021-02790-4

Most M. S., and Yates D. T. (2021). Animals, 11(12), 3539. https://doi.org/10.3390/ani11123539

Nguyen T. T. T., Bowman P. J., Haile-Mariam M., Nieuwhof G. J., Hayes B. J., *et al.* (2017). J. Dairy Sci., 100(9), 7362-7367. https://doi.org/10.3168/jds.2017-12898

Nicolaides N. C., Chrousos G., and Kino T. (2020). Glucocorticoid receptor. MDText.com, South Dartmouth, US.

Otto P. I., Guimarães S. E. F., Verardo L. L., Azevedo A. L. S., Vandenplas J., *et al.* (2019). J. Dairy Sci., 102(9), 8148-8158. https://doi.org/10.3168/jds.2018-15305

Ou F. S., Michiels S., Shyr Y., Adjei A. A., and Oberg A. L. (2021). J. Thorac. Oncol., *16*(4), 537-545. https://doi.org/10.1016/j.jtho.2021.01.1616

Rehfeld J. F., and Bundgaard J. R. (2010). Cellular Peptide Hormone Synthesis and Secretory Pathways (Vol. 50). Springer, Berlin, DE.

Sigdel A., Abdollahi-Arpanahi R., Aguilar I., and Peñagaricano F. (2019). Front. Genet., *10*, 928. https://doi.org/10.3389/fgene.2019.00928

Srikanth K., Kwon A., Lee E., and Chung H. (2017). Cell Stress and Chaperones, 22(1), 29-42. https://doi.org/10.1007/s12192-016-0739-8

Thompson-Crispi K. A., Sewalem A., Miglior F., and Mallard B. A. (2012). J. Dairy Sci., 95(1), 401-409. https://doi.org/10.3168/jds.2011-4452

Thompson-Crispi K., Atalla H., Miglior F., and Mallard B. A. (2014). Front. Immunol, 5. https://doi.org/10.3389/fimmu.2014.00493

Tugal D., Liao X., and Jain M. K. (2013). Arterioscler. Thromb. Vasc Biol., 33(6), 1135-1144. https://doi.org/10.1161/ATVBAHA.113.301453

Walsh K. P., and Mills K. H. G. (2013). Trends Immunol., 34(11), 521-530. https://doi.org/10.1016/j.it.2013.07.006

632. Epistatic interaction networks affecting feed efficiency and methane emission in Nelore cattle

T.F. Cardoso[1*], J.J. Bruscadin[2], W.J.S. Diniz[3], P. Banerjee[3], J. Afonso[1], J. Petrini[4], B.G.N. Andrade[5], J.M. Malheiros[6], T. Porto[2], A. Zerlotini[7], G.B. Mourão[4], L.L. Coutinho[4] and L.C.A. Regitano[1]

[1]Embrapa Southeast Livestock, São Carlos, São Paulo, Brazil; [2]Postgraduate Program on Evolutionary Genetics and Molecular Biology, Federal University of São Carlos, São Carlos, São Paulo, Brazil; [3]Department of Animal Sciences, Auburn University, AL, USA; [4]Department of Animal Science, 'Luiz de Queiroz' College of Agriculture, University of São Paulo/ESALQ, Piracicaba, São Paulo, Brazil; [5]Munster Technological University, Ireland; [6]Federal University of Latin American Integration, Foz do Iguaçu, Paraná, Brazil; [7]Embrapa Agricultural Informatics, Campinas, São Paulo, Brazil; tainafcardoso@gmail.com

Abstract

Genetic interactions between SNPs can lead to phenotypic changes. Herein, we used the WISH-R method to test for epistasis in residual feed intake (RFI) and residual methane emission (RME) of Nelore cattle. The genomic interaction modules identified four and ten significant SNP-SNP modules for RME and RFI traits, respectively. The region's annotation identified genes that could be impacted by SNPs modules correlated to the traits. The over-representation analysis showed that the SNPs correlated to the RME trait were close to genes involved with the ion transport, glutamate receptor signalling pathway, RAC1 GTPase cycle, and ERBB2 signalling. Regarding the RFI trait, the over-representation analysis showed genes involved in the Ca-dependent events and regulation of cation transmembrane transport. Knowing these SNP–SNP interactions could improve understanding of the genetic and physiological mechanisms contributing to feed efficiency and methane emission.

Introduction

The beef cattle industry is a fundamental and growing part of the Brazilian economy. Feed efficiency is directly associated with increased meat production and can provide information useful to reduce environmental impact, such as methane emission. Different approaches were carried out to identify genetic variants and genes associated with different traits of interest in Nelore cattle (de Lima *et al.*, 2020; de Oliveira *et al.*, 2014, 2018). However, studies exploring the epistatic effects over these traits are generally scarce. Genetic interactions between SNPs occur when the phenotypic effect given to a mutation is affected by other modifications in the genome (Papp and Pál, 2011) and are key to understanding how genes relate functionally. This study aimed to identify SNP-SNP interactions using the WISH-R method (Carmelo *et al.*, 2018) to test for epistasis in residual feed intake and methane emission traits of Nelore cattle.

Materials & methods

Animals and phenotypic data. The Nelore population used in the current work has been previously described by Andrade *et al.* (personal communications). The experimental population consisted of 52 animals born in 2014 and submitted to two diets. The animals were then sent for slaughter at 23-24 months of age, following the Humane Slaughter of Cattle guidelines, and were approved by the EMBRAPA Livestock Science Ethics Committee on Animal Experimentation, São Carlos, São Paulo (Protocol No. 09/2016). Residual feed intake (RFI, kg/d) and residual methane emission (RME) traits were measured as described elsewhere. In summary, the RFI was computed as described by Koch *et al.* (1963), and RME was obtained as described by Donoghue *et al.* (2016).

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_632, © T.F. Cardoso *et al.* 2022

Reducing genotype data dimensionality. DNA extractions were performed with TRIZOL, and all animals were genotyped using the GGP *Bos indicus* 50k beadchip. Samples were filtered according to the minimum call rate (95%) and minimum allele frequency (95%). Sexual and monomorphic SNPs were also excluded from the dataset. A total of 43,439 SNPs were available for further analysis. Genome-wide association analysis was performed to reduce genotype data dimensionality using PLINK software (Purcell *et al.*, 2007) based on the *assoc* function. Phenotypic data were corrected for diet and contemporary groups (i.e. diet, weight group, slaughter date). SNPs with $P<0.1$ were selected from each association and used to carry out the epistatic interaction calculation individually for each trait.

Epistatic networks and trait association analyses. The epistatic interactions for RFI and RME in Nelore cattle were calculated using the WISH-R package (Carmelo *et al.*, 2018), considering the default parameters. The epistatic signed weighted networks were constructed for each trait, and genomic interaction modules were created based on the WGCNA pipeline (Langfelder and Horvath, 2008). Significant modules with positive and negative correlations for each trait were selected for downstream analysis. The SNPs' genomic location annotation was performed based on the ARS-UCD1.2 bovine genome coordinates. Functional Annotation and biological interpretation were performed using the ClueGo plugin for Cytoscape (Bindea *et al.*, 2009).

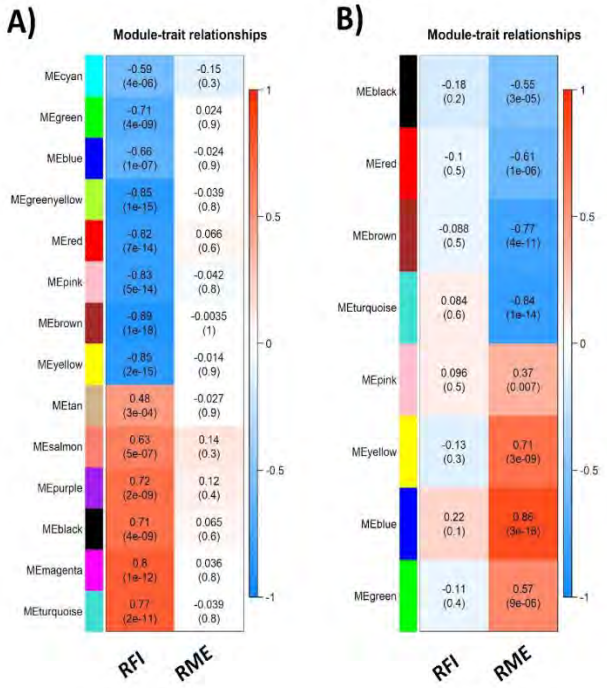
Results

SNP-SNP network analysis. From the 43,439 SNPs markers kept after QC, a total of 3,885 and 4,481 SNPs were associated ($P<0.1$) with RME and RFI traits, respectively. These SNPs were then used as input for the genome-wide epistatic analysis for each trait separately. The genomic interaction modules based on the WISH-R approach identified four and ten significant SNP-SNP modules ($P\leq0.05$ and $r>|0.7|$) for RME and RFI traits, respectively (Figure 1). In this case, the modules can be defined as clusters of highly correlated SNPs. Considering the RME trait, half of the modules showed positive and half negative correlation with the RME. For the RFI trait, six modules showed a negative correlation, while four were positively correlated. No SNP module was associated with both analyzed phenotypes.

Annotation of SNP modules and gene enrichment. For the RME trait, 52% of the hub-SNPs identified from the significantly correlated modules were harbored in intron regions, followed by 36% in intergenic regions. Regarding hub-SNPs associated with the RFI trait, 51% were in intronic variants, followed by 37% in intergenic regions. Additionally, for both traits, about 10% of the hub-SNPs were located in regulatory regions, i.e. upstream or downstream regions. The annotation of hub-SNP regions retrieved 415 and 466 genes that could be impacted by hub-SNPs of positive and negative modules correlated to RME trait, respectively. Over-representation analyses showed that the hub-SNPs of modules negatively correlated to the RME trait were involved in the *neuronal system development*, *ion transport*, and *glutamate receptor signaling pathway,* while the hub-SNPs of the positive modules were involved in the *RAC1 GTPase cycle*, *ERBB2 Activates PTK6 Signaling* and *PI3K events in ERBB2 signaling*. Regarding the RFI trait, a total of 609 and 702 genes may be impacted by hub-SNPs of positive and negative modules correlated to the trait, respectively. Over-representation analyses showed that the hub-SNPs of the negative modules were involved in the *Calmodulin induced events* and *Ca-dependent events*. Likewise, the hub-SNPs of the positive modules correlated to the RFI trait were involved in pathways related to the *regulation of cation transmembrane transport*.

Discussion

The livestock sector demands an increase in profitability and a reduction in environmental impacts through an improvement on feed efficiency and a decrease in methane emission traits (Sakamoto *et al.*, 2021). Various genetic and environmental factors can influence both traits, and this genetic variability indicates



C)

Module	SNPS	Genes
MEcyan	82	31
MEsalmon	93	35
MEtan	107	44
MEgreenyellow	199	97
MEpurple	203	87
MEmagenta	245	105
MEblack	248	88
MEpink	248	99
MEred	280	112
MEgreen	282	121
MEyellow	331	126
MEbrown	364	141
MEblue	502	189
MEturquoise	884	325

D)

Module	SNPS	Genes
MEblack	179	63
MEyellow	233	91
MEgreen	198	71
MEred	195	66
MEbrown	260	119
MEturquoise	1595	603
MEpink	95	37
MEblue	780	324

Figure 1. Module-trait correlation plots. Module trait correlation for Residual feed intake (RFI, A) and Residual methane emission (RME, B); Number of SNPs and identified genes clustered in each module for RFI© and RME (D).

that genomic selection is possible (de Oliveira *et al.*, 2014; Donoghue *et al.*, 2016). In this study, we used genome-wide SNPs to unravel the genetic contribution of the epistatic effects onto the RFI and RME variation in Nelore cattle. We found several modules associated with the analysed traits and identified genes and pathways possibly involved in phenotype determination from the network analysis. The *RAC1 GTPase cycle* and pathways involved in the *ERBB2 signaling* were over-represented in the RME positive correlated modules. Both pathways are involved in various cellular processes, including cytoskeleton organization, transcription, and cell proliferation (Bosco *et al.*, 2009, Appert-Collin *et al.*, 2015). Based on the over-representation pathway analysis, RME negatively correlated modules were involved with ion transport. Through the ion's concentration and pH regulation, the host physiology can exert a selective pressure on the gut microbiota (Engevik and Engevik, 2021). Another top pathway is involved in *glutamate receptor signaling*. Previous studies showed that the glutamatergic signaling along the microbiota-gut-brain axis may modulate the gut microbiota (Baj *et al.*, 2019). The relationship between the abundance of microorganisms and methane emission has already been reported (Bowen *et al.*, 2020; Ghanbari Maman *et al.*, 2020).

Regarding the RFI phenotype, our findings indicate that observed gene networks are primarily related to metabolic processes underlying *Ca-dependent events* and *regulation of cation transmembrane transport*. Previous studies showed the impact of ion metabolism as calcium in feed efficiency trait (de Oliveira *et al.*, 2014; Marchesi *et al.*, 2021; Shi *et al.*, 2020). Further studies will be performed with the genes and the hub-SNPs found here. The knowledge of these SNP–SNP interactions may improve our understanding of the genetic and physiological mechanisms that contribute to feed efficiency and methane emission.

References

Appert-Collin A., Hubert P., Crémel P., and Bennasroune A. (2015). Front. Pharmacol., 6, 283. https://doi.org/10.3389/fphar.2015.00283

Baj A., Moro E., Bistoletti M., Orlandi V., Crema F., et al (2019). Int. J. Mol. Sci., 20(6), 1482. https://doi.org/10.3390/ijms20061482

Bindea G., Mlecnik B., Hackl H., Charoentong P., Tosolini M., *et al.* (2009). Bioinformatics, 25(8), 1091–1093. https://doi.org/10.1093/bioinformatics/btp101

Bosco E.E., Mulloy J.C. and Zheng Y. (2009). CMLS, 66(3), 370–374. https://doi.org/10.1007/s00018-008-8552-x

Bowen J.M., Cormican P., Lister S.J., McCabe M.S., Duthie C.A., *et al.* (2020). PLOS ONE, 15(4), e0231759. https://doi.org/10.1371/journal.pone.0231759

Carmelo V.A.O., Kogelman L.J.A., Madsen M.B. and Kadarmideenet H.N. (2018). BMC Bioinformatics, 19(1), 277–277. https://doi.org/10.1186/s12859-018-2291-2

de Lima A.O., Koltes J.E., Diniz W.J.S., de Oliveira P.S.N., Cesar A.S.M., et al (2020). Front. genet., 11(March), 1–14. https://doi.org/10.3389/fgene.2020.00189

de Oliveira P.S.N., Cesar A.S., do Nascimento M.L., Chaves A.S., Tizioto P.C., *et al.* (2014). BMC genetics, 15, 100–100. https://doi.org/10.1186/s12863-014-0100-0

de Oliveira P.S.N., Coutinho L.L., Tizioto P.C., Cesar A.S.M., de Oliveira G.B., et al (2018). Sci. Rep., 8(1), 17072–17072. https://doi.org/10.1038/s41598-018-35315-5

Donoghue K.A., Bird-Gardiner T., Arthur P.F., Herd R.M. and Hegarty R.F. (2016). Sci. J. Anim. Sci., 94(4), 1438–1445. https://doi.org/10.2527/jas.2015-0065

Donoghue K.A., Bird-Gardiner T., Arthur P.F., Herd R.M. and Hegarty R.S. (2016). Anim. Prod. Sci., 56(3), 213–217. https://doi.org/10.1071/AN15573

Engevik A.C. and Engevik M.A. (2021). Comput. Struct. Biotechnol. J., 19, 134–144. https://doi.org/10.1016/j.csbj.2020.12.008

Ghanbari Maman L., Palizban F., Fallah Atanaki F., Ghiasi N.E., Ariaeenejad S., *et al.* (2020). Sci. Rep., 10(1), 4995. https://doi.org/10.1038/s41598-020-61942-y

Langfelder P. and Horvath S. (2008). BMC Bioinformatics, 9(1), 559. https://doi.org/10.1186/1471-2105-9-559

Marchesi J.A.P., Ono R.K., Cantão M.E., Ibelli A.M.G., Peixoto J.O., *et al.* (2021). Sci. Rep., 11(1), 4622. https://doi.org/10.1038/s41598-021-84125-9

Papp B. and Pál C. (2011). Cell Cycle, 10(21), 3623–3624. https://doi.org/10.4161/cc.10.21.17853

Purcell S., Neale B., Todd-Brown K., Thomas L., Ferreira M.A., *et al.* (2007). Am. J. Hum. Genet., 81(3), 559–575. https://doi.org/10.1086/519795

Sakamoto L.S., Souza L.L., Gianvecchio S.B., de Oliveira M.H.V., Silva J.A.I.V., *et al.* (2021). PLOS ONE, 16(10), e0257964. https://doi.org/10.1371/journal.pone.0257964

Shi B., Jin M., Jiao L., Betancor M., Tocher D., *et al.* (2020). Br. J. Nutr., 124(8), 773–784. https://doi.org/10.1017/S0007114520001725

633. Host-genomically influenced ruminal microbial genes are temporally stable during the finishing phase in beef cattle

J. Lima[1*], M. Martínez-Álvaro[1], J. Mattock[2], M.D. Auffret[1], C.A. Duthie[1], M.A. Cleveland[3], R.J. Dewhurst[1], M. Watson[2] and R. Roehe[1*]

[1]Scotland's Rural College, Edinburgh, United Kingdom; [2]Royal (Dick) School of Veterinary Studies, University of Edinburgh, Edinburgh, United Kingdom; [3]Genus PLC, DeForest, WI, USA; joana.lima@sruc.ac.uk; rainer.roehe@sruc.ac.uk

Abstract

The temporal stability of rumen microbiome composition is very important for microbiome-driven breeding of host performance traits because it helps to identify the number of samples needed over time and best timepoints to obtain highly informative samples. Therefore, we investigated the temporal stability of the abundances of host-genomically influenced microbial genes throughout the finishing phase in beef cattle, and their associations with estimated breeding values (EBVs) of host traits such as feed conversion efficiency and methane emissions. Our results indicate that rumen microbiome profiles generated from different timepoints throughout the finishing phase are remarkably stable, with pairwise correlations of 0.74 or higher, and that their associations with EBVs of host traits were stable and strong throughout. Based on these results, we concluded that microbiome profiles at any sampling timepoint are representative of the finishing phase and can be used to predict EBVs of performance traits at similar accuracy.

Introduction

Ruminants are unique in their ability to feed on diets rich in human-indigestible materials due to the symbiotic relationship with their rumen microbiota comprised of bacteria, protozoa, and fungi that ferment complex polysaccharides into volatile fatty acids, vitamins, and microbial proteins, the main energy and nutrient sources for the host animal. During this process, methanogenic archaea use excess hydrogen leading to methane (CH_4) emissions. Previous studies demonstrated the close association between rumen microbial genes (MG) and host traits – for example, Roehe *et al.* (2016) identified 20 and 49 rumen MG that explained 81 and 86% of the variation in CH_4 emissions and feed conversion ratio (FCR), respectively, and Lima *et al.* (2019) found that sets of fewer than 20 rumen MG explained more than 60% of the variation in traits including average daily weight gain (ADG), daily feed intake (DFI), and FCR. Host-genomically influenced microbial genes (hMG) have been proposed as potential selection criteria of beef cattle in breeding programs (Martínez-Álvaro *et al.*, 2021a, 2021b). Overall, rumen MG have been suggested as accurate predictors of host performance traits at phenotypic and genomic levels.

Studies on the association between the rumen microbiome (microbiota and their genes) and host traits in beef cattle often rely on microbiome data generated from sampling at slaughter. However, whether MG abundances generated from slaughter samples are representative of rumen microbiome profiles at earlier stages of the host animal life is still unclear.

Therefore, in this study, we aimed at assessing the temporal stability of hMG, and of the associations of hMG with host performance traits, including FCR, ADG, DFI, residual feed intake (RFI), and CH_4 yield (g/kg daily dry matter intake) throughout the finishing phase of beef cattle.

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_633, © J. Lima *et al.* 2022

Materials & methods

The experimental trial included 84 steers from breeds Charolais and Luing, fed diets with forage:concentrate ratios of 520:480 (FOR), or 84:916 g/kg dry matter basis (CON). Twenty steers (6 from each breed fed CON, and 4 fed FOR) had their rumen digesta sampled at 6 timepoints throughout the finishing phase: at 418±32 days (d) old (T1), at the start (average of 460±32 d old, T2), mid (average of 492±32 d old, T3), and end (average 520±32 d old, T4) of the performance test period, on the day the animal left the respiration chamber (T5), and at slaughter (T6). During the 56-d testing period, DFI was obtained as dry matter intake (DMI, kg/day), ADG was obtained by fitting a linear regression of body weight (measured weekly before fresh feed was offered) on test day. FCR was calculated as DMI divided by ADG. Each steer spent 3d in respiration chambers (RC) from which the final 48h were used to estimate daily CH_4 emissions, which were divided by average DMI measured within the RC, resulting in CH_4 yield. Estimated breeding values (EBV) of traits were calculated based on FCR, ADG, DFI, RFI, and CH4 records from a larger set of 359 animals (which included our 84 steers), using univariate GBLUP models corrected for systematic effects (diet, breed, and trial) using the same assumptions as in Martínez-Álvaro *et al.* (2021c).

Estimating abundances of microbial genes. Microbial DNA extracted from ruminal samples was sequenced in Illumina HiSeq and PE150 systems by Edinburgh Genomics (Edinburgh, UK). Whole metagenome sequencing reads were assembled using MEGAHIT (Li *et al.*, 2015), proteins predicted with Prodigal (Hyatt *et al.*, 2010) and identified using KofamScan (Aramaki *et al.*, 2020) and the Kyoto Encyclopedia of Genes and Genomes database (Kanehisa and Goto, 2000). MG absent from at least one animal, or with relative abundance ≤0.001% were removed from the dataset. Due to the compositionality of microbiome data, abundances were additive logratio transformed (alrMG), using the reference MG K00858, based on the criteria proposed by Greenacre, *et al.* (2021). A total of 443 MG had previously been shown to be heritable (Martínez-Álvaro *et al.*, 2021b), and therefore were used in the further analyses.

Statistical analyses. We explored the temporal stability of the alrMG within each animal by calculating the pairwise Pearson correlation (*r*) between timepoints. To assess the temporal stability of the alrMG over all the animals, we calculated Pearson correlations between the alrMG (adjusted for fixed effects breed and diet) from different timepoints. Additionally, alrMG from each timepoint (adjusted for fixed effects) were used in partial least squares models (PLS) to predict EBVs computed from GBLUP, to evaluate the stability of the associations between hMG and EBVs of host performance traits. All analyses and plots were produced in R version R-4.1.1, using packages mixOmics (Rohart *et al.*, 2017), and tidyverse (Wickham *et al.*, 2019).

Results

The within-animal correlations revealed that alrMGs generated throughout the finishing phase were mainly stable. For example, for T1-T6, correlations varied between r=0.56 (Figure 1, blue font), and r=0.90 (Figure 1, orange font). The correlations between alrMG (adjusted for fixed effects) were strong throughout the finishing phase, varying from r=0.74 (between T1 and T2) and r=0.87 (between T3 and T4) (Table 1). Host-genomically influenced alrMGs generated from each timepoint strongly predict EBVs of the traits, for example, between 63.83 and 77.92% of the variability in FCRebv was explained by alrMGs generated from sampling the rumen at T3 and T1, respectively. Additionally, 70.25±3.43% of the variability of ADGebv was explained by hMG generated throughout the finishing phase, with the lowest variance being 66.84%, explained by alrMGs generated from T2 (Table 2). Overall, the results indicate that alrMGs generated from all timepoints explained more than 56% of the variability of the EBVs of performance traits.

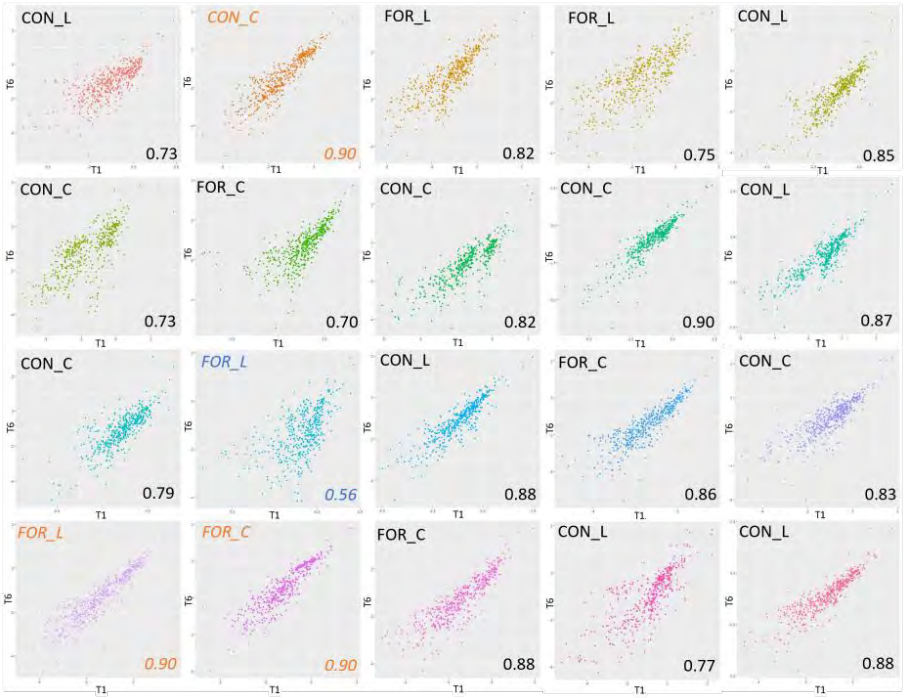


Figure 1. Temporal stability of the abundances of host-genomically influenced microbial genes. Biplots show the additive logratio abundances of microbial genes generated from sampling at timepoints T1 (x axis) and T6 (y axis) for each of the 20 animals fed concentrate (CON) or forage (FOR), of breeds Charolais (C) or Luing (L). Pearson correlations presented in the bottom-right corner of each plot.

Table 1. Pairwise Pearson correlations between abundances of host-genomically influenced microbial genes generated from samples collected at 6 different timepoints (T1 to T6) throughout the finishing phase of beef cattle.

Timepoint	T1	T2	T3	T4	T5
T2	0.74				
T3	0.75	0.80			
T4	0.75	0.79	0.87		
T5	0.77	0.81	0.78	0.79	
T6	0.79	0.81	0.84	0.81	0.80

Table 2. Variance (%) of estimated breeding values explained by abundances of host-genomically influenced microbial genes generated from samples collected at 6 different timepoints (T1 to T6) throughout the finishing phase of beef cattle.[1]

Trait	T1	T2	T3	T4	T5	T6	Average ± standard deviation
[1]FCRebv	77.92	66.55	63.83	66.98	76.07	67.66	69.84±5.73
[2]ADGebv	72.54	66.84	67.24	70.95	75.59	68.31	70.25±3.43
[3]DFIebv	81.41	70.92	72.49	60.89	81.06	70.57	72.89±7.65
[4]RFIebv	74.23	56.02	62.22	67.78	77.71	66.93	67.48±7.87
[5]CH_4ebv	70.28	71.25	78.49	65.63	64.28	68.75	69.78±5.03

[1] ebv = estimated breeding values; FCR = feed conversion ratio; ADG = average daily weight gain; DFI = daily feed intake; RFI = residual feed intake; CH_4 = methane emissions yield (g/kg dry matter intake).

Discussion

Our results indicate that hMG abundances generated from sampling throughout the seven-month long finishing phase are remarkably constant, suggesting that microbiome profiles of beef cattle are substantially stable. This result agrees with Snelling *et al.* (2019), which used microbiota data generated from 16S rRNA gene sequencing, and suggested that in the absence of major perturbations such as a change in diet rumen microbiome profiles are temporally stable. We showed that the associations between hMG abundances and EBVs of host performance traits are strong and stable, with hMG abundances generated from different timepoints explaining more than 56% of the variability of EBVs. In agreement, using microbiome profiles from slaughter samples, Lima *et al.* (2019) identified sets of rumen MG whose relative abundances explained more than 63% of the phenotypic variation of FCR, ADG, RFI, and DFI. Our results are highly relevant for the current research on the rumen microbiome, particularly for microbiome-driven breeding because they suggest that one rumen microbiome sample per animal at any time during the finishing phase of cattle is sufficient to predict EBVs of performance traits at similar accuracy.

References

Aramaki T., Blanc-Mathieu R., Endo H., Ohkubo K., Kanehisa M., *et al.* (2020) Bioinform 36(7) 2251–2252. 10.1093/bioinformatics/btz859

Greenacre M., Martínez-Álvaro M., and Blasco A. (2021) Front microbiol 12:727398. https://doi.org/10.1101/2021.05.15.444300.

Hyatt D., Chen G-L., LoCascio P.F., Land M.L., Larimer F.W., *et al.*, (2010) BMC Bioinform 11:119. https://doi.org/10.1186/1471-2105-11-119

Kanehisa M., and Goto S. (2000) Nucleic Acids Res 28(1), 27–30. 10.1016/j.meegid.2016.07.022

Li D., Liu C-M., Luo R., Sadakane K., and Lam T-W. (2015) Bioinformatics 31(10), 1674–1676. 10.1093/bioinformatics/btv033

Lima J., Auffret M.D., Stewart R.D., Dewhurst R.J., Duthie C-A., *et al.* (2019) Front genet 10:701. 10.3389/fgene.2019.00701.

Martínez-Álvaro M., Auffret M.D., Duthie C-A., Dewhurst R.J., Cleveland M.A., *et al.* (2021a) Research Square - Preprints. https://doi.org/10.21203/rs.3.rs-290150/v1

Martínez-Álvaro M., Duthie C-A., Mattock J., Dewhurst R.J., Cleveland M.A., *et al.* (2021b) Proc. of the 72nd Annual Meeting of the EAAP, Davos, Switzerland.

Martínez-Álvaro M., Zubiri-Gaitán A., Hernández P., Greenacre M., Ferrer A., *et al.* (2021c) Commun Biol 4:1246. https://doi.org/10.1038/s42003-021-02784-w

Roehe R., Dewhurst R.J., Duthie C-A., Rooke J.A., McKain N., *et al.* (2016) PLoS Genet 12(2): e1005846. 10.1371/journal.pgen.1005846.

Rohart F., Gautier B., Singh A., and Lê Cao K-A. (2017) PLoS Comput Biol 13(11): e1005752. https://doi.org/10.1371/journal.pcbi.1005752

Snelling T.J., Auffret M.D., Duthie C-A., Stewart R.D., Watson M., *et al.* (2019) Anim Microbiome 1:16. 10.1186/s42523-019-0018-y.

Wickham H., Averick M., Bryan J., Chang W., McGowan L.D-A., *et al.* (2019). J. Open Source Softw 4(43), 1686. https://doi.org/10.21105/joss.01686

634. A hypomorphic mutation in the *ATP2A1* gene increases muscle mass yet compromises meat quality of Belgian Blue cattle

T. Rombouts[1], T. Druet[2], J.L. Gualdrón Duarte[2,3], N. Ahariz[4], K. Karim[4], W. Coppieters[2,4], S. De Smet[1], M. Georges[2] and C. Charlier[2*]

[1]Department of Animal Sciences and Aquatic Ecology, Ghent University, 9000 Gent, Belgium; [2]Unit of Animal Genomics, GIGA-R, University of Liège, 4000 Liège, Belgium; [3]Walloon Breeders Association, 5590 Ciney, Belgium; [4]GIGA Genomics Platform, University of Liège, 4000 Liège, Belgium; carole.charlier@uliege.be

Abstract

We herein describe a novel phenotype related to meat quality in Belgian blue cattle breed, named 'tough and dry meat syndrome'. A genome-wide association study pinpointed a single significantly associated locus on chromosome 25, encompassing the *ATP2A1* gene coding for the SERCA1 protein. A previous reverse genetic screen identified a breed-private missense mutation *(R143W)* in this gene. It segregates at high frequency (0.15) without deviation from expected genotypic proportions and was proven to be causative for this syndrome. The prevalence of the condition is estimated at 12.5% and its severity correlates with *W143* allelic dosage. The mutation is partially dominant with a penetrance of 43 and 100% in *R/W* and *W/W* animals respectively. Retrospective examination of *W143* frequency revealed a clear linear upward trend, likely due to the fact that the *W143* variant has a positive effect on muscle mass. Direct selection against the mutation is underway.

Introduction

Belgian blue cattle (BBC) breed is known for its exceptional double-muscling phenotype due to the fixation of a loss-of-function mutation in the *myostatin* gene (Grobet *et al.*, 1997). It is directly associated with a lower collagen content in muscle resulting in a superior meat tenderness much appreciated by consumers. But, since several years, meat cutters reported the occasional observation of meat cuts that were described as 'tough and dry'. Most of the affected muscles were located in high-value cuts in the hindquarter: silverside (m. *glutaeobiceps*), topside (m. *semimembranosus*) and top rump (m. *quadriceps femoris*). In the worst cases the eye-muscle (m. *longissimus dorsi*), the underblade (m. *subscapularis*) and the shoulder tender (m. *teres major*) were also impacted. Here we report the identification of a mutation in the *ATP2A1* gene associated with a major negative effect on meat quality.

Materials & methods

Carcass phenotyping. Meat quality phenotypes were classified into six discrete categories (C) of increasing severity (0, 1, 1.5, 2, 2.5 and 3). Carcasses without any macroscopic meat defect are classified as C0. Upon cutting of the carcasses, all affected animals (C > 0) exhibited exudate at the surface of the *m. semimembranosus*, and a 'tough and dry' feeling of the *m. adductor femoris* and inner parts of the *m. glutaeobiceps* and *m. semimembranosus* (C1 and 1.5), as well as of the *m. quadriceps femoris, m. semitendinosus, m. longissimus dorsi, m. subscapularis* and *m. teres major* for the most severe cases (C2 to 3).

Genome-wide association studies (GWAS). For the association studies, DNA samples were genotyped on custom LD12K (1st cohort) or MD50K (2nd cohort) EuroGenomics SNP arrays (Illumina, Inc) as described in Charlier *et al.*, 2016. After QC, 11,088 and 47,473 autosomal SNPs were kept in the 1st and 2nd cohorts, respectively. GWAS between SNP or haplotypes and meat quality (0 = normal; 1 = 'tough and dry') were performed using GLASCOW (Zhang *et al.*, 2012). The genomic relationship matrix was estimated either based on SNP similarity or on haplotype similarity as described in Zhang *et al.* (2012). For this, haplotypes were reconstructed with Beagle 4.1 (Browning and Browning, 2007) and then grouped at each

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_634, © T. Rombouts *et al.* 2022

marker position into 10 clusters of ancestral haplotypes (hidden states; HS) based on their similarity with the model described in Druet and Georges (2010) using HiddenPHASE.

Results

Using genotypes from 38 affected meat pieces (cases) and from 1,058 unrelated breed-age-matched animals as controls, we obtained a single genome-wide significant signal on chr25 with the two models (single SNP and 10 HS) (Figure 1A and 1B). The most significant variant was a missense variant *(R143W)* in the *ATP2A1* gene encoding the SERCA1 protein (P=3.6×10^{-21}). The maximum 10 HS signal (P=6.2×10^{-23}) overlapped this variant, and was due to the enrichment in cases of a single haplotype that was in near perfect linkage disequilibrium (r^2≥0.91) with the *ATP2A1 R143W* variant (Figure 1C). The SERCA1 protein is a Ca^{2+} ATPase that is embedded in the membrane of the sarcoplasmic reticulum of muscle fibers, pumping Ca^{2+} ions from the sarcoplasm back into the reticulum thereby controlling muscle relaxation (eg. Wuytack *et al.*, 2002). Loss-of-function mutations in *ATP2A1* cause autosomal recessive Brody myopathy in humans (e.g. Brody, 1969; Odermatt *et al.*, 1996), congenital muscular dystonia I in BBC (Charlier *et al.*, 2008), and perinatal lethality in *ATP2A1* knock-out mice (Pan *et al.*, 2003). Since, missense *ATP2A1* variants have been associated with recessive pseudo-myotonia of variable severity in at least four other beef cattle breeds (OMIA 001464-9913). The *R143W* mutation is located in *ATP2A1* coding exon 5 and affects a highly conserved residue in the first cytoplasmic loop of SERCA1 (Figure 1 D-F). *R143W* is only reported in BBC. It had a frequency of 15% in the GWAS control population, yet of 55% in the 38 affected samples (P=9.5×10^{-23}). Based on these initial findings, we performed a systematic survey for 'tough and dry' meat cuts in Belgian slaughter-plants. We first examined 496 carcasses of adult Belgian Blue cows, 62 (12.5%) were considered to be affected to various extend. A striking feature was the systematic increased loss of fluid from the carcass (referred to as 'drip') during the cooling process. Upon cutting of the non-C0 carcasses, they exhibited exudate at the surface of the respective affected muscles (see M&M). Samples of the m. *glutaeobiceps* were collected from 26 affected (5 C1; 9 C1.5; 9 C2; 3 C2.5) and 21 non-affected C0 animals and analysed for standard meat quality traits. Briefly, non-C0 samples had altered colour (darker) and colour stability, texture (higher hardness, gumminess, chewiness, and shear force), and moisture content (reduced EZ-drip, thawing and cooking loss; increased dry matter content) compared to C0 controls. We genotyped 116 case (non-C0) samples as well as 127 unaffected (C0) controls. A case-versus-control GWAS confirmed the chr25 association, maximizing at the *ATP2A1 R143W* position with individual SNP (P=1.7×10^{-36}) and the HS10 (P=4.8×10^{-54}) models. The frequency of the *W143* allele was 8.2% in the 'hyper' C0 controls, while it was 55.1% in non-C0 cases. We further tested (within cases) whether there was an association between *R143W* genotype and severity of the condition by regressing dry meat score (C1 to 3) on dosage of the *W* allele (1 or 2). The average score of *R/W* cases was 1.46 while it was 1.88 for *W/W* cases (P=0.0005). We retrospectively examined the change in frequency of the *W143* variant in the general BBC population over time (>65,000 animals from 2014 to 2020).

It showed a clear linear upward trend from 14.5% in 2014 to 17.5% in 2020. To understand what may cause this trend, we studied the effect of the *W143* variant on 22 phenotypes that are systematically recorded in BBC as described in Gualdrón Duarte *et al.* (2020). Significant (Bonferroni-corrected) positive effects were observed for five traits pertaining to muscular development of the front and hindquarters (data not shown).

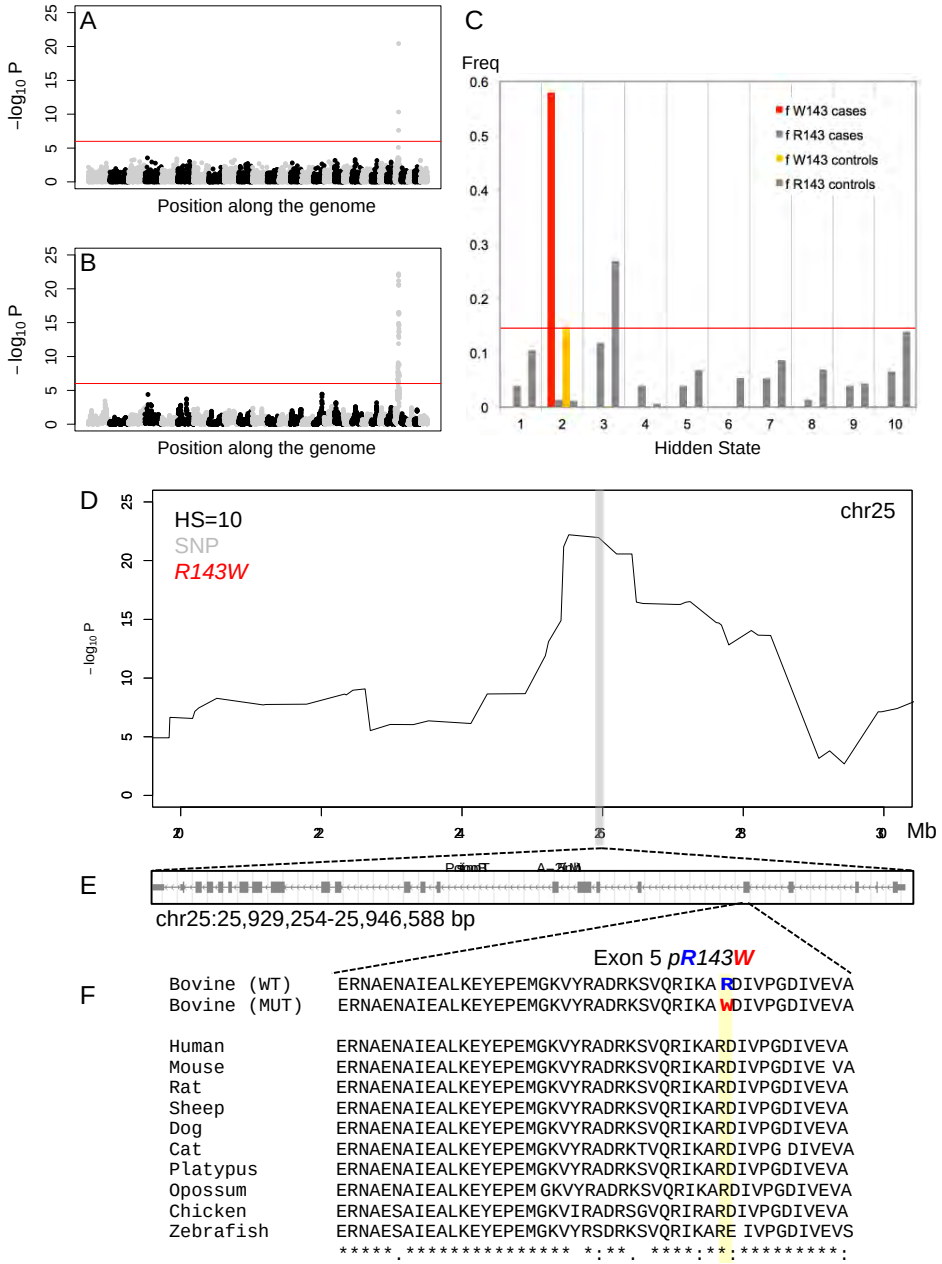


Figure 1. Mapping and fine-mapping of the 'tough and dry meat' phenotype. Manhattan plots show the respective significance of a single SNP model (A) and a ten hidden-states (HS) model (B); the horizontal red line indicates the significance threshold; alternating colors (black and grey) mark the limits between each of the 29 autosomes. (C) Frequency of the 10 HS in the 38 cases and the 1,058 controls at peak position; within each HS, frequency is given for haplotypes carrying the *W143* mutation or the wild-type allele in cases and in controls; the horizontal red line indicates the *W143* freq. in the population. (D) Association signals for a single SNP model (grey dots) and a 10HS model (black line). (E) Gene model for the *ATP2A1* gene (adapted from UCSC browser). (F) *ATP2A1* exon 5 protein alignment across vertebrates highlighting the *R143W* aa change; * marks invariant aa.

Discussion

We herein provide very strong evidence that a likely hypomorphic *ATP2A1* missense variant (*R143W*) is the cause of a newly discovered 'tough and dry meat' phenotype in BBC. Assuming that the prevalence of the condition is 12.5% in the BBC population and given the genotypic proportions observed in cases and controls, we estimate that the penetrance of the anomaly is 43% in heterozygotes and 100% in homozygotes, which at present account for 30 and 3%, respectively. We can estimate the depreciation of the non-C0 carcasses at roughly 20% by assuming that the affected cuts can only be valorized as minced meat. Further work is still needed to determine which factors affect the penetrance of the *W143* mutation. We have since confirmed that the defect is present as well in young bulls with a similar prevalence (data not shown). The frequency of the *W143* mutation appears to have been steadily increasing in the BBC population. We showed that this is likely due to the fact that the *W143* allele has a positive effect on muscle mass. It thus represents another example of the increase in frequency of a variant with a major deleterious effect because of its perceived positive impact on muscle mass – the phenotype that remains the primary selection goal in BBC. Knowing the causative mutation, will allow for effective counter selection against the *W143*. However, as heterozygotes may have a high probability to be affected (>40%), avoiding matings between carrier animals would not be sufficient to control the condition. Animals carrying the *W143* mutation should be identified by direct genotyping and progressively eliminated from the breeding stock.

References

Brody I.A. (1969) N. Engl. J. Med. 281(4):187-92. https://doi.org/10.1056/NEJM196907242810403.

Browning S.R. and Browning B.L. (2007) Am. J. Hum. Genet. 107(5):895-910. https://doi.org/10.1016/j.ajhg.2020.09.010.

Charlier C., Coppieters W., Rollin F., Desmecht D., Agerholm J.S. *et al.* (2008) Nat. Genet. 40(4):449-54. https://doi.org/10.1038/ng.96.

Charlier C., Li W., Harland C., Littlejohn M., Coppieters W. *et al.* (2016) Genome Res. 26(10):1333-1341. https://doi.org/10.1101/gr.207076.116.

Druet T. and Georges M. (2010) Genetics 184(3):789-98. https://doi.org/10.1534/genetics.109.108431.

Grobet L., Martin L.J., Poncelet D., Pirottin D., Brouwers B. *et al.* (1997) Nat. Genet. 17(1):71-4. https://doi.org/10.1038/ng0997-71.

Gualdrón Duarte J.L., Gori A.S., Hubin X., Lourenco D., Charlier C. *et al.* (2020) BMC Genomics 21(1):545. https://doi.org/10.1186/s12864-020-06921-3.

Odermatt A., Taschner P.E.M., Khanna V.K., Busch H.F.M., Karpati G. *et al.* (1996) Nat. Genet. 14(2):191-4. https://doi.org/10.1038/ng1096-191.

Pan Y., Zvaritch E., Tupling A.R., Rice W.J. *et al.* (2003) J. Biol. Chem. 278(15):13367-75. https://doi.org/10.1074/jbc.M213228200.

Wuytack F., Raeymaekers L., Missiaen L. (2002) Cell Calcium 32(5-6):279-305. https://doi.org/10.1016/s0143416002001847.

Zhang Z., Guillaume F., Sartelet A., Charlier C. *et al.* (2012) Bioinformatics 28(19):2467-73. https://doi.org/10.1093/bioinformatics/bts348.

635. Mitochondrial DNA copy number as a potential genetic indicator of growth and carcass traits in beef cattle

L.P. Sanglard[1*], L.A. Kuehn[2], W.M. Snelling[2], and M.L. Spangler[1]

[1]Department of Animal Science, University of Nebraska, Lincoln, NE, 68583, USA; [2]USDA, ARS, Roman L Hruska U.S. Meat Animal Research Center, Clay Center, NE, 68933, USA; lpereirasanglard2@unl.edu

Abstract

Mitochondrial DNA copy number (mtDNA CN) was investigated as a potential indicator of growth and carcass traits in a beef cattle population. Blood or leucocyte samples were subjected to low-pass whole-genome sequencing resulting in nuclear DNA (nuDNA) and mtDNA reads for 2,371 animals. Mitochondrial DNA CN was estimated as the ratio of mtDNA to nuDNA coverages. Growth and carcass traits were available for up 60,989 individuals. Genetic parameters were estimated from a bivariate animal model based on pedigree. Estimate of heritability of mtDNA CN was 0.11±0.07. Genetic correlation estimates of mtDNA CN were -0.30±0.16 with birth weight, -0.31±0.16 with age-adjusted weaning weight, -0.15±0.14 with post-weaning gain, -0.11±0.13 with slaughter weight, -0.14±0.13 with carcass weight, -0.07±0.14 with carcass backfat, 0.14±0.14 with carcass marbling, and -0.06±0.14 with carcass ribeye area. Thus, mtDNA CN is heritable and lowly correlated with growth and carcass traits. The USDA is an equal opportunity employer and provider.

Introduction

Mitochondrial DNA copy number (mtDNA CN), defined as the ratio of mitochondrial DNA (mtDNA) to nuclear DNA (nuDNA) coverages, could be a potential genetic indicator for economically important traits in beef cattle. Decreasing cost of low-pass whole-genome sequencing (LPS) makes this approach an attractive alternative to obtain mtDNA and nuDNA; hence mtDNA CN is relatively easy to obtain (Ding *et al.*, 2015). Mitochondrial DNA CN correlates with the size and number of mitochondria, which are critical determinants of mitochondrial function (Knez *et al.*, 2016). Mitochondrial DNA CN has been shown to be moderate to highly heritable (h^2~0.31 to 0.65) and associated with traits of interest in humans, such as semen quality (Tian *et al.*, 2014; Workalemahu *et al.*, 2017). However, studies are lacking in the literature for livestock species such as beef cattle. Therefore, the objectives of this work were to estimate genetic parameters for mtDNA CN and its genetic correlation with growth and carcass traits in a composite beef population comprising the Germplasm Evaluation (GPE) program from the U.S. Meat Animal Research Center (USMARC).

Materials & methods

Samples of blood (n=1,205) and leucocyte (n=766) were obtained from the USMARC GPE program and sent to GeneSeek (Neogen company, Lincoln, Nebraska) for DNA extraction and LPS. Reads were aligned to bosTau9 (ARS-UCD1.2) reference genome using bwa men v. 0.7 (Li, 2013). Mapped reads were split into mtDNA and nuDNA to estimate coverages (samtools v. 1.9). Mitochondrial DNA CN was calculated as the ratio of mtDNA to nuDNA coverages. Quality control included the removal of unmapped reads and reads with a quality score <20. Joint variant calling for mtDNA was performed in GATK v. 4 (Poplin *et al.*, 2017). SNP calls were then filtered as recommended by the software provider. Samples with <900,000 nuDNA reads were removed. Mitochondrial DNA CN was natural log-transformed for further analyses such that the distribution approached normality and the impact of potential outliers was decreased. Carcass and growth traits available included birth weight (BWT; n=60,989), age-adjusted weaning weight (AWWT; n=56,122), post-weaning gain (PWG; n=50,968), slaughter weight (SLWT, n=25,067); carcass weight (CWT; n=23,953); carcass backfat (FAT; n=23,586); carcass marbling (MRB; n=23,196), and carcass ribeye

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_635, © L.P. Sanglard *et al.* 2022

area (REA; n=23,828). Pedigree information was available for up to 28 generations. Breed heterozygosity was calculated as one minus the sum of the products of breed fractions obtained from pedigree of the sire and dam. A more detailed description of the breeds included in the GPE project have been previously reported (Schiermiester *et al.*, 2015; Ahlberg *et al.*, 2016; Zimmermann *et al.*, 2019).

Genetic parameters were obtained from a bivariate animal model by fitting fixed effects of sex, heterosis, breed composition, contemporary group (confounded with tissue type), age at sample collection (mtDNA CN), age of slaughter (SLWT, CWT, FAT, MRB, and REA), maternal heterosis (BW and AWWT), maternal breed compositions (BW and AWWT), and age of dam (BW and AWWT). Random effects included the additive genetic effect for all models and maternal genetic and maternal permanent environment effects for BW and AWWT. Maternal effects (genetic, heterosis, and breed composition) were also investigated for mtDNA; however, they were removed from the model given that the maternal h^2 was ~0 and the effects of maternal heterosis and breed compositions were not significant.

Results

Distribution of natural log ratio mtDNA CN by sex, age of tissue collection, direct heterosis, and maternal heterosis can be seen in Figure 1. Blood samples had lower mtDNA CN than leucocyte samples ($P \leq 0.001$). Linear and quadratic effects of age were significant ($P \leq 0.02$). The effect of heterosis was not significant ($P \geq 0.47$).

Estimates of h^2 of mtDNA CN and genetic correlation can be seen in Table 1. Estimates of h^2 for mtDNA CN was 0.11 (0.07). Estimates of genetic correlation of mtDNA CN ranged from -0.31 (0.16) with AWWT to 0.14 (0.14) with MRB.

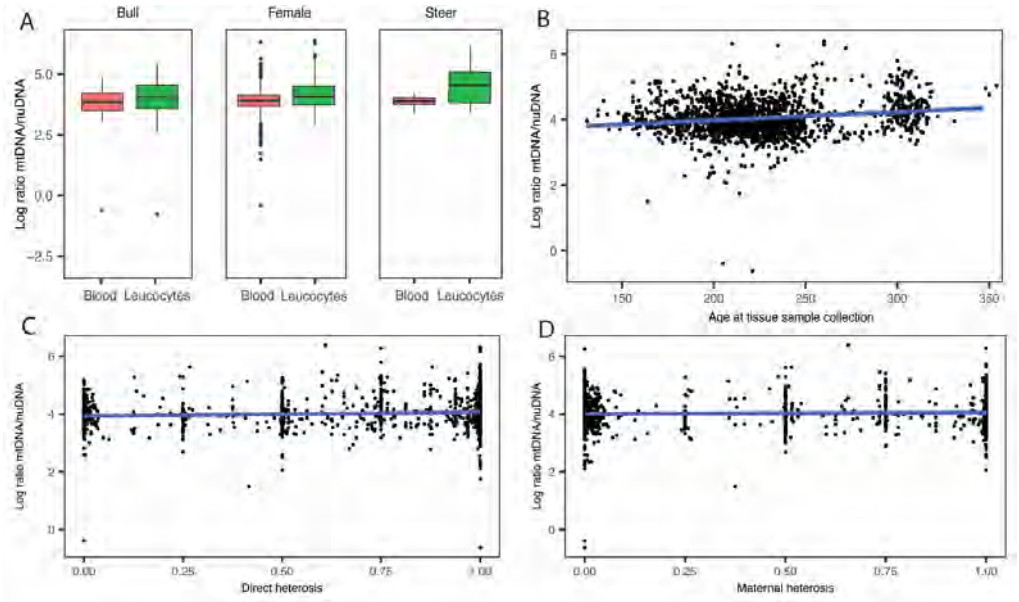


Figure 1. Distribution of natural log ratio mitochondrial DNA copy number by sex (A), age of collection (B), direct heterosis (C), and maternal heterosis (D).

Table 1. Estimates of heritability and correlation of growth and carcass traits with mitochondrial DNA copy number (mtDNA CN).

Trait[1]	Heritability	Maternal heritability	Genetic correlation	Phenotypic correlation[2]
BWT	0.44 (0.02)	0.07 (0.009)	-0.30 (0.16)	-0.06 (0.03)
AWWT	0.21 (0.01)	0.11 (0.01)	-0.31 (0.16)	-0.05 (0.03)
PWG	0.38 (0.01)	-	-0.15 (0.14)	0.009 (0.03)
SLWT	0.57 (0.02)	-	-0.11 (0.13)	-
CWT	0.56 (0.02)	-	-0.14 (0.13)	-
FAT	0.51 (0.02)	-	-0.07 (0.14)	-
MRB	0.62 (0.02)	-	0.14 (0.14)	-
REA	0.49 (0.02)	-	-0.06 (0.14)	-

[1] Birth weight, BWT; age-adjusted weaning weight, AWWT; post-weaning gain, PWG; slaughter weight, SLWT; carcass weight, CWT; carcass backfat, FAT; carcass marbling, MRB; carcass ribeye area, REA.

[2] Residual covariance was set to zero for carcass traits given no individual had both traits recorded.

Discussion

Interest in the use of whole-genome sequence data has increased, in part due to lower-cost options enabled by LPS and the possibility of including functional variants in genomic evaluations. A collateral advantage of this technology is the extraction of mtDNA and, consequently, estimation of mtDNA CN. Previous studies have shown that mtDNA CN in humans is highly heritable and associated with traits semen quality and obesity (Tian *et al.*, 2014; Ding *et al.*, 2015). Therefore, mtDNA CN could be a potential indicator trait to improve the rate of genetic gain for traits of interest in livestock. In the current study, we estimated h^2 and genetic correlations of mtDNA with growth and carcass traits in beef cattle.

Estimates of h^2 obtained in this study were lower (~0.11) than previous reports in the literature for human populations (~0.40). The differences observed could be attributed to differences in the tissue utilized to extract mtDNA, the technique to obtain mtDNA CN (PCR vs whole-genome sequencing), species (human vs cattle), the ability to quantify and fit various systematic effects in human populations, among others.

The genetic correlation estimates of mtDNA CN with growth and carcass traits were low although larger in magnitude for earlier growth traits (i.e. BWT and AWWT). The difference in sign of the correlation estimates between mtDNA CN and early growth traits vs MRB is consistent with the negative correlation between BWT and AWWT with MRB (Cundiff *et al.*, 1971; Arnold *et al.*, 1991). These results indicate that selection to decrease mtDNA CN would increase BWT and AWWT. Similarly, Reverter *et al.* (2017) reported a negative correlation between breast muscle mtDNA CN with muscle (-0.27) and carcass yield (-0.26) in broilers. Lower mitochondrial content relates to lower costs with the production and maintenance of mitochondria. Therefore, mitochondrial DNA content could be associated with metabolism efficiency (Hudson *et al.*, 2008; Hudson, 2009).

Estimates reported herein suggest that mtDNA CN is under additive genetic control of the nuclear genome in beef cattle. Mitochondrial DNA CN can be relatively easy to be obtained as a by-pass product of LPS and has the potential to be used as a genetic indicator in beef cattle. Additional research is warranted to investigate its relationship with more sparsely collected phenotypes of economic interest.

References

Ahlberg C. M., Kuehn L. A., Thallman R. M., Kachman S. D., Snelling W. M., *et al.* (2016) J. Anim. Sci. 94:1857–1864. https://doi.org/10.2527/jas.2015-0161.

Amold J., Bertrand J. K., Benyshek L., and Ludwig C. (1991) J. Anim. Sci. 69:985–992.

Cundiff L.V., Gregory K.E., Koch R.M., and Dickerson G.E. (1971) J. Anim. Sci. 33:550–555. https://doi.org/10.2527/jas1971.333550x.

Ding J., Sidore C., Butler T.J., Wing M.K., Qian Y. *et al.* (2015) PLoS Genet. 11:1–18. https://doi.org/10.1371/journal.pgen.1005306.

Hudson N.J. (2009). J. Anim. Physiol. Anim. Nutr. (Berl). 93:1–6. https://doi.org/10.1111/j.1439-0396.2007.00791.x.

Hudson N.J., Lehnert S.A., and Harper G.S. (2008) Med. Hypotheses. 70:693–697. https://doi.org/10.1016/j.mehy.2007.05.042.

Knez J., Winckelmans E., Plusquin M., Thijs L., Cauwenberghs N., *et al.* (2016). Am. J. Epidemiol. 183:138–146. https://doi.org/10.1093/aje/kwv175.

Reverter A., Okimoto R., Sapp R., Bottje W.G., Hawken R., *et al.* (2017) Biol. Open. 6:50–58. https://doi.org/10.1242/bio.022772.

Schiermiester L. N., Thallman R. M., Kuehn L. A., Kachman S. D., and Spangler M. L. (2015). J. Anim. Sci. 93:46–52. https://doi.org/10.2527/jas.2014-8493.

Tian M., Bao H., Martin F.L., Zhang J., Liu L., *et al.* (2014) Biol. Reprod. 91:1–8. https://doi.org/10.1095/biolreprod.114.122465.

Workalemahu T., Enquobahrie D.A., Tadesse M.G., Hevner K., Gelaye B., *et al.* (2017) J. Matern. Neonatal Med. 30:2433–2439. https://doi.org/10.1080/14767058.2016.1252747.

Zimmermann M. J., Kuehn L. A., Spangler M. L., Thallman R. M., Snelling W. M., *et al.* (2019) J. Anim. Sci. 97:1523–1533. https://doi.org/10.1093/jas/skz045.

636. Genomic regions in Australian cattle associated with consumer satisfaction of beef

A.M. Lynn[1*], P. McGilchrist[1], H. Aliloo[1], J.H.J. van der Werf[1], R. Polkinghorne[1,2] and S.A. Clark[1]

[1]*School of Environmental and Rural Science, University of New England, 2350, Armidale, NSW, Australia;* [2]*Birkenwood International, 45 Church street, 3122, Hawthorn, Victoria, Australia;* alynn2@une.edu.au

Abstract

Consumer satisfaction influences purchasing habits and ultimately the premiums consumers are willing to pay for beef. There are four criteria currently used to describe consumer satisfaction: tenderness, juiciness, flavour and overall liking. While previous studies have focused on the traits associated with consumer satisfaction such as marbling and shear force, few have analysed consumer traits directly. The aim of this study was to determine if genomic regions affecting consumer satisfaction could be discovered using a genome-wide association study. Phenotypic data including consumer sensory scores were collected for 1,924 animals from a breed diverse population of cattle genotyped for a minimum of 50k single nucleotide polymorphisms. A single region linked to the CAST gene on chromosome 7 (BTA7) was associated with tenderness. The CAST gene is responsible for muscle proteolysis after slaughter and has been well documented as impacting on the mechanical tenderness measurement of Warner-Bratzler shear force.

Introduction

Consumer satisfaction is a key trait in beef cattle breeding programs as consumers express a willingness to pay premiums for guaranteed quality and consistency (Shackelford *et al.*, 2001). Currently, predicting consumer satisfaction relies on objective carcass measurements (such as marbling, shear force and muscle pH) as databases for consumer satisfaction traits (tenderness, juiciness, flavour and overall liking) are too small and expensive to analyse alone. Recent global collaborations have allowed datasets to expand, therefore making consumer satisfaction traits more feasible to analyse. This also presents the opportunity to identify regions of the cattle genome linked directly to consumer derived sensory traits as genomic information can aid the prediction accuracy for objective carcass measurements (Pimentel and König, 2012; Mehrban *et al.*, 2017; Bedhane *et al.*, 2022). This study aims to identify genomic regions linked to consumer derived eating quality traits along with their associated variance components and respective heritabilities in a multi breed population of cattle.

Materials & methods

Animals. The data used in this study included 1,924 records sourced from various research projects conducted from 1997 to 2019. Cattle were sourced from 77 cohorts with a recorded breed profile consisting of five temperate beef breeds, five tropical beef breeds, two dairy breeds, crossbreds of the aforementioned breeds and some cattle having breed not identified in the data. This study included steers (n=1,481), heifers (n=403) and bulls (n=40). The data was recorded from ten commercial abattoirs in five states across Australia, with all animals graded for carcass characteristics such as pH, rib fat depth, fat colour, eye muscle area, meat colour, marbling, carcass weight and ossification by accredited graders. All consumer sensory data for tenderness, juiciness, flavour and overall liking was collected as described by Watson *et al.* (2008b) for cuts *longissimus lumborum* (STR045; n=1,591) and *longissimus thoracis* (CUB045; n=333) cooked using the grill method (Watson *et al.*, 2008). All samples were consumed by 10 untrained consumers and scored on a 0-100 point scale for each of the sensory variables. To remove extreme values, the two highest and two lowest values were removed. The remaining six values were then averaged for a 'clipped' sensory score. Animals with recorded age had their phenotypes pre-corrected by applying the effect of age observed in a linear model to the deviation from the mean age.

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_636, © A.M. Lynn *et al.* 2022

Genomic data structure, quality control and GWAS analysis. Animals were genotyped across 5 different single nucleotide polymorphism (SNP) chips (BovineSNP50 Genotyping Beadchip v1, BovineSNP50 Genotyping Beadchip v2, GGP Bovine 50K, GGP Bovine 100K and a 50k TropBeef chip). Plink v1.9 was utilised for the QC of genotypes, where SNPs with a missing rate of >0.1, a minor allele frequency (MAF) of <0.01 and those departing from the Hardy-Weinberg equilibrium at P<1E-08 were removed. The ARS-UCD 1.2 (Rosen *et al.*, 2020) map was used to merge genotypes by retaining only the markers with known genotypes on at least 60% of all individuals. This resulted in 54,728 SNPs across 29 bovine autosomes for the 1,924 cattle with phenotypic records. A GWAS was carried out using GCTA 1.93.3 *(Yang et al.*, 2011) for each consumer derived trait (tenderness, juiciness, flavour and overall liking) separately. The genomic relationship matrix (GRM) was constructed for a larger dataset of 2,567 animals through GCTA before being utilised in the GWAS. A Principal Component analysis (PCA) was carried out based on the GRM to assign individuals to breed groups in further analysis. A single SNP regression model was used to identify SNPs which were associated to the desired phenotypes using the Mixed Linear Model-based Analysis (MLMA) in GCTA software. The fixed effects in the model were cut and cohort (slaughter group), and the first three principal components of the GRM were included as covariates. The model in matrix format is stated below:

$$y = Xb + Wq + Zu + e \quad (1)$$

where y is the vector of phenotype, b is the vector of fixed effects, q is the vector of allele substitution effects, u is a vector for additive genetic effects and it was assumed to be distributed as $u|\sigma^2_g \sim N(0, G\sigma^2_g)$, where, G is the GRM as described above, X and Z are the corresponding design matrices that relates the observations to the fixed effects (b) and the random effects (u), respectively. W is the matrix of genotype codes (0,1,2) for each SNP and e is the vector of residual effects. Significance threshold was determined using the Bonferroni correction and set at $9.13e^{-07}$. Variance components and heritability were also estimated in the MLMA using GREML.

Results

The principle component analysis confirmed a large amount of background population structure in the dataset. It confirmed that the crossbred and unknown animals were likely a cross between breeds known in the population. Figure 1 illustrates the separation of the *Bos Indicus*, *Bos Taurus* beef and *Bos Taurus* dairy animals. It demonstrates the diverse population in which we attempted to find SNP associations for the four traits observed and confirmed the need to adjust for background population effects in the GWAS analysis. The phenotypic means, minimum, maximum, additive, residual variances and heritabilities (with their standard errors; SE) of the four meat quality traits are represented in Table 1. There was more phenotypic variation in consumer scores recorded for tenderness (177.7) than for flavour (106.7). Tenderness also had the highest (0.2±0.05) heritability across the four traits with juiciness having the lowest (0.13±0.05) heritability.

Table 1. Descriptive statistics, variance components and heritability of sensory traits examined with standard error (SE).

Trait	Mean	Min	Max	$V_G \pm SE$	$V_E \pm SE$	$h^2 \pm SE$
Tenderness	59.1	11.52	95.49	34.2±9.80	140.5±9.1	0.2±0.05
Juiciness	58.5	19.4	93	19.1±7.79	123.4±7.54	0.13±0.05
Flavour	60.2	21.96	92	17.1±5.55	89.6±5.55	0.16±0.05
Overall liking	59.3	12.23	96	24.2±7.40	106.3±6.86	0.19±0.05

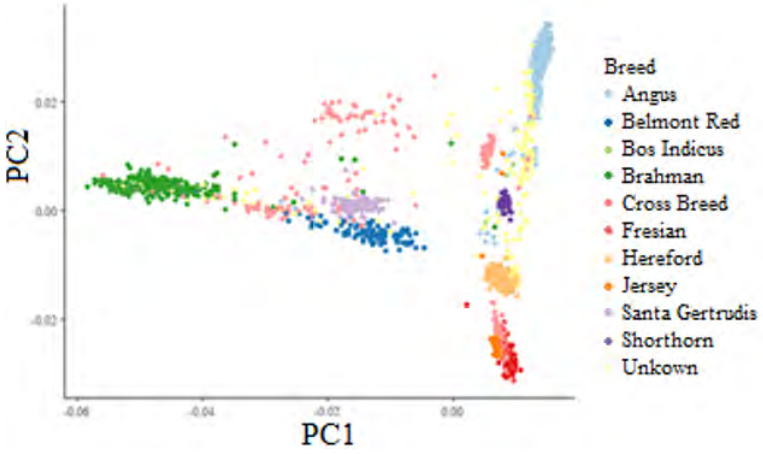


Figure 1. Plot of first two principal components depicting the genetic structure and relatedness of the population. High content *Bos Indicus* cattle stretched to the left with composite breeds between *Bos Taurus* cattle on the right.

Tenderness and overall liking were the only two traits to have observable peaks above the threshold of 5-$\log_{10}$, however only tenderness had SNPs (rs137601357 and rs109677393) on BTA7 surpassing the Bonferroni threshold (Figure 2). With rs137601357 having an effect of -3.3 tenderness units. This same region had the highest significance for other traits, however, neither juiciness or flavour passed the significance threshold. A downstream analysis of the position of the two SNP was undertaken and the two SNP were closely located to the CAST gene on BTA7.

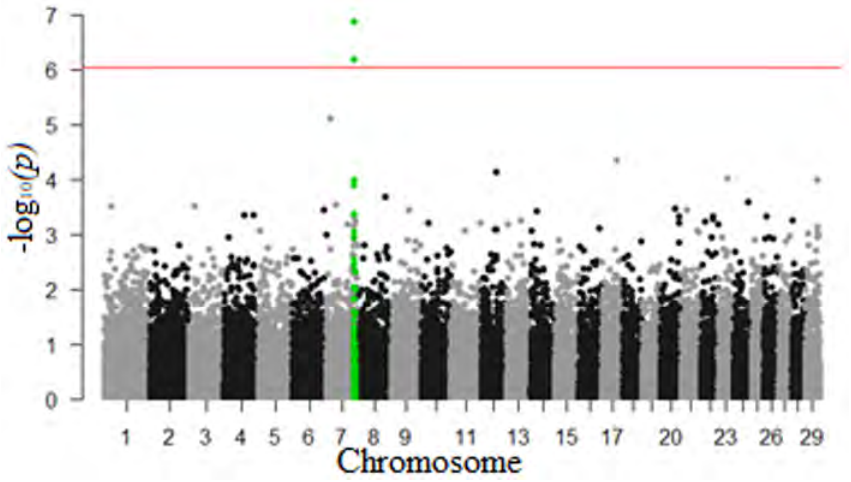


Figure 2. Manhattan plot of association between common SNPs in the population and tenderness. The genomic position is represented on the x axis and -$\log_{10}$(p) on the y axis. The red line indicates P=9.13e^{-07} to represent the Bonferroni threshold.

Discussion

This study found a low to moderate heritability for the traits included, these heritabilities are within the range of those observed by Johnston *et al.* (2003), who estimated heritability for the same traits in tropical and temperate breeds separately. Tenderness was the only trait to pass the more stringent threshold of the Bonferroni correction, both rs137601357 and rs109677393 have previously been linked to the CAST gene (on BTA7) which regulates calpastatin production which is known to influence proteolysis after slaughter and, therefore affect tenderness (Mateescu *et al.*, 2017; Abo-Ismail *et al.*, 2018). A unique feature of the current study is that it utilised a very unstructured and semi commercial dataset, achieving similar results to previous studies with more structured datasets. Future research could observe if the separation of the dataset between tropical and temperate breeds of cattle as noted in Johnston *et al.* (2003) would impact on observable peaks for consumer satisfaction traits. Alternatively the dataset could be divided based on the date of collection to observe if peaks have moved over the 23 year period of collection as important quantitative trait loci are known to move or even disappear over generations (Meuwissen *et al.*, 2016). The challenges experienced in this dataset have highlighted the importance of improved uniformity in data collection for future analysis.

This preliminary study has identified a region of the cattle genome which has been previously identified and linked to tenderness. This is supportive of the theory that muscle proteolysis due to variation in calpastatin production influenced by the CAST gene is significant for tenderness when judged by consumers. Further research is required to ascertain if this same region is evident in other populations of cattle. The next stage of this research is to incorporate a wider data set from overseas to examine if the same results are replicated on a larger scale.

References

Abo-Ismail M., Lansink N., Akanno E., Karisa B., *et al.* (2018) J. Anim. Sci. 96(2):375-397. https://doi.org/10.1093/jas/sky020

Bedhane M., van der Werf J., de las Heras-Saldana S., Lim D., *et al.* (2022) Anim. Prod. Sci. 62(1) https://doi.org/10.1071/AN20659

Johnston D.J., Reverter A., Ferguson D.M., Thompson J.M., Burrow H.M. (2003) Aust. J. of Agric. Res. 54(2):135-147. https://doi.org/10.1071/AR02087

Mateescu R.G., Garrick D.J., and Reecy J.M. (2017) Front. Genet. 8:171. https://doi.org/10.3389/fgene.2017.00171

Mehrban H., Lee D.H., Moradi M.H., IlCho C., *et al.* (2017) Genet. Sel. Evol. 49(1):1. https://doi.org/10.1186/s12711-016-0283-0

Meuwissen T., Hayes B., and Goddard M. (2016) Anim. Front. 6(1):6-14. https://doi.org/10.2527/af.2016-0002

Pimentel E., and König S. (2012). J. Anim. Sci. 90(10):3418-3426. https://doi.org/10.2527/jas2011-5005

Rosen B.D., Bickhart D.M., Schnabel R.D., Koren S., *et al.* (2020). Gigascience 9(3) https://doi.org/10.1093/gigascience/giaa021

Shackelford S., Wheeler T., Meade M., Reagan J., *et al.* (2001). J. Anim. Sci. 79(10):2605-2614. https://doi.org/10.2527/2001.79102605x

Watson R., Gee A., Polkinghorne R., and Porter M. (2008) Aust. J. Exp. Agric. 48(11):1360-1367. https://doi.org/10.1071/EA07176

Yang J., Lee S.H., Goddard M.E., and Visscher P.M. (2011) Am J Hum Genet 88(1):76-82. https://doi.org/10.1016/j.ajhg.2010.11.011

637. A genome-wide scan for signatures of selection in Hanwoo and Angus cattle using whole-genome sequence data

H.J. Kim, S. de las Heras-Saldana, S. Clark and J.H.J. van der Werf*

School of Environmental & Rural Science, University of New England, Armidale, NSW, 2351, Australia; jvanderw@une.edu.au

Abstract

Signature of selection in livestock is a method used to identify genomic regions and candidate genes related to economic traits when phenotypic information is unknown. This study used 406 beef cattle (203 Hanwoo and 203 Angus) with a whole-genome sequence to find selection signatures. As a result, we identified 298 significant genomic regions within Hanwoo, including the candidate genes such as *LPL* related to the lipid metabolism and 33 genomic regions within Angus with *HK1* and *ACTC1* genes linked with glucose metabolism and muscle formation. From the between populations results, 36 significant genomic regions were detected, including the *CCKBR* gene associated with the feed efficiency. This study can assist in understanding the history of these breeds and identifying the genomic regions under selection for breeding programs in beef cattle.

Introduction

A signature of selection refers to genomic regions selected under natural or artificial selection (Qanbari and Simianer, 2014). The process of selection signature provides opportunities to find target selected genomic regions and genes related to economic traits, and understand the history of populations (Qanbari and Simianer, 2014). Many methods have been proposed for detecting signatures of selection. A common method is the extended haplotype homozygosity (EHH) method, which assesses the probabilities that two randomly chosen chromosomes carry a core haplotype homozygous for a region from the core SNP to a distant locus (Sabeti *et al.*, 2002). The use of haplotypes for signatures of selection studies has shown a higher potential to detect selection than other methods (Ma *et al.*, 2015). When searching for selection signatures, it is also important to select breeds with divergent attributes such as high meat yield in Angus and good meat quality in Hanwoo. It is then hypothesised that such breeds have been subjected to different breeding objectives and environments (Taye *et al.*, 2018). This study aimed to identify the signature of positive selection in Hanwoo and Angus cattle using whole-genome sequence data.

Materials & methods

A total of 406 animals with whole-genome sequence data, including 203 Hanwoo and 203 Angus, were used in this study. A quality control assessment of the genotypes in PLINK v.1.9 was applied to remove the SNPs with a genotype call rate of less than 95%, less than 1% minor allele frequency, and a *P*-value for Hardy Weinberg equilibrium <0.00001. After the quality control step, we extracted the SNPs that overlapped between both data sets resulting in the use of 7,868,872 SNPs from *Bos Taurus* autosome (BTA). All individuals were phased using Eagle v.2.4 to infer haplotypes in each population. The extended haplotype homozygosity (EHH) method was used with the R packages '*rehh*'. In addition, the integrated haplotype score (iHS) for within a population (Voight *et al.*, 2006) and the Rsb statistic (Tang *et al.*, 2007) for between populations assessment were used. The values of iHS and Rsb were standardized using a mean of 0 and a standard deviation of 1 for allele frequency at the core SNP and transferred from each value to the *P*-value using the Gaussian cumulative distribution function. The significance threshold for each analysis was used for iHS ($-\log_{10}$*P*-value (0.0001)>4) and for Rsb ($-\log_{10}$*P*-value(0.05)>5.48). The significant genomic regions were defined within 1 Mb of significant SNPs, and extended until there were no significant SNPs. We used the Ensembl database with the ARS-UCD1.2 bovine reference genome

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_637, © H.J. Kim *et al.* 2022

to identify the candidate genes in significant genomic regions. These genes were compared with those previously identified in the Cattle QTL database (https://www.animalgenome.org/cgi-bin/QTLdb/BT/search).

Results

We constructed Manhattan plots with the *P*-value of iHS scores for each population and Rsb between populations (Figure 1). As a result, 298 significant genomic regions ($-\log_{10}$(*P*-value)>4) were detected for Hanwoo, but only thirty-three regions were identified in Angus. The most significant value of iHS was located on BTA6 (iHS=6.95) in Hanwoo and BTA10 (iHS=4.87). For the signature of selection results of the between population analysis using Rsb, 36 significant regions were detected and the most significant value of Rsb was -6.55 on BTA12.

Nine candidate genes surrounding the top 5 significant regions within the Hanwoo population were detected on BTA8 and BTA16. For Angus, there were 50 candidate genes detected in the top 5 significant regions on BTAs 6, 10, 20, 28 and 29. From the between population results, 32 candidate genes were identified in the top 5 significant genomic regions. Interestingly, the region on BTA28 was significant and shared between populations (Rsb) and the iHS results in Angus (Table 1).

Discussion

The EHH values were used to detect signatures of selection in each population. iHS, the EHH method for within population analysis, compares the ancestral allele and derived allele, while Rsb compares between populations. The iHS method found 298 and 33 significant genomic regions for Hanwoo and Angus, respectively. Genes related to the immune system and intramuscular fat were previously reported in Hanwoo (Lim *et al.*, 2016) on BTAs 2, 3, 9, 10, 11, 12, 22 and 26. But genes involved in body size, body structure and meat quality were identified for Angus on BTAs 2, 3, 4, 14, 15 and 16 (Paim *et al.*, 2020). Some significant genomic regions that overlapped with previous studies were on BTA2 and BTA11 in Hanwoo and on BTA14 in Angus.

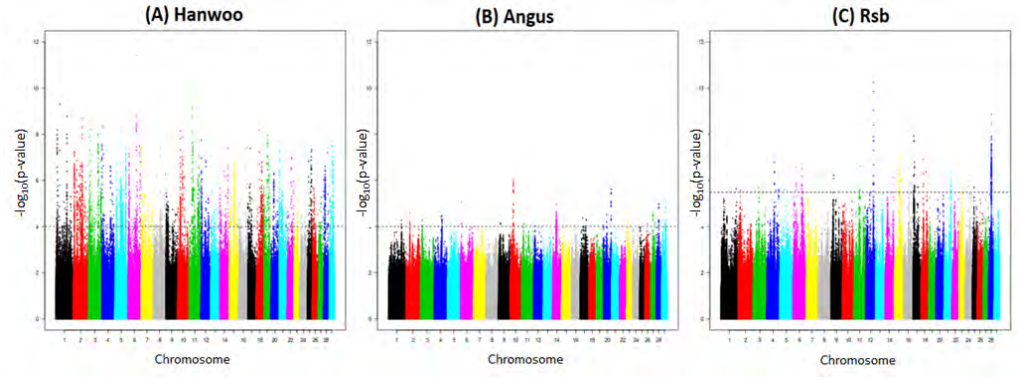


Figure 1. The Manhattan plot of $-\log_{10}$(*P*-value) of (A) iHS in Hanwoo, (B) iHS in Angus, and (C) Rsb between Hanwoo and Angus.

Table 1. The top 5 significant genomic regions and associated genes.[1]

Method	Population	Chr	Regions (Mb)	n SNP	Genes
iHS	Hanwoo	6	75.2-76.2	18	-
		8	66.1-67.1	34	*LPL*
		11	34.2-35.2	126	-
		16	75.2-76.2	12	*PLXNA2, CD34, CD46, ASPM, ZBTB41, F13B, CRB1*
		1	36.5-37.5	3	*EPHA3*
	Angus	10	29.9-30.9	27	*GREM1, SCG5, GJD2, ACTC1, AQR, ZN770, DPH6*
		29	38-39.5	13	*PAG7, PAG15, PAG4, PAG14, PAG16, PAG20, PAG21, PAG1, PAG19, PAG17, MGC157408, MGC157405, PAG9, PAG3, PAG6, PAG11*
		20	70.4-71.5	14	*IRX4, NDUF56, LPCAT1, SLC6A3, CLPTM1L, TERT, SLC6A19, SLC6A18, NKD2, TRIP13, BRD9, TPPP, CEP72*
		6	7.84-8.87	3	*TRAM1L1*
		28	24.9-25.9	12	*TET1, CCAR1, STOX1, DDX50, DDX21, KIFBP, SRGN, VPS26A, SUPV3L1, HKDC1, HK1, TACR2, TSPAN15*
Rsb	Between	12	69.9-70.9	18	-
		28	24.8-25.9	240	*SLC25A16, TET1, CCAR1, STOX1, DDX50, DDX21, KIFBP, SRGN, VPS26A, SUPV3L1, HKDC1, HK1, TACR2, TSPAN15*
		17	6.6-7.9	8	*SH3D19, RPS3A, LRBA, MAB21L2, DCLK2*
		17	13.5-14.7	16	*HHIP, GYPA, GYPB, SMARCA5, GAB1, USP38*
		15	46.7-49.1	33	*OR52fam, CAVIN3, CCKBR, CNGA4, C15H11orf42, HBE2, TRIM34*

[1] Chr = chromosome; Mb = mega bases; n SNP = The number of SNPs in the region.

The results of iHS in Hanwoo were stronger than Angus. We assumed that Hanwoo might have had less selection pressure than Angus because the effective population size of Angus decreased faster than Hanwoo in this study. iHS is a more powerful method to detect intermediate selection sweeps than complete selection signatures. A population with weak selection pressure can have a higher number of intermediate allele frequencies than a population with strong selection pressure due to relatively taking more time for complete sweeps under a weak selection pressure population (Lopez *et al.*, 2019). This might be the reason why more significant regions were found with iHS in Hanwoo. In contrast with iHS, Rsb has the power to detect complete selective sweeps (Tang *et al.*, 2007). Most significant genomic regions from Rsb were related to Angus since this breed shows more evidence of strong selection. The nine candidate genes located close to the top 5 significant regions for the within population (iHS) assessment in Hanwoo were *LPL* and *ZBTB41*, which have been associated with the intramuscular fat in Qunchaun cattle (Oh *et al.*, 2013), and residual feed intake in Holstein (Zhou *et al.*, 2018), respectively. For Angus, the *HK1* and *ACTC1* genes were linked to glucose metabolism (Ebara *et al.*, 2013) and muscle formation (Qanbari *et al.*, 2011). These results confirm that Hanwoo has been mainly selected for meat quality traits (lipid metabolism), while Angus has been selected for growth and meat yield. For Rsb, 32 candidate genes were detected from which the *LRBA* gene was associated with the kinase A related to the immune effector of molecules (Mapholi *et al.*, 2016), and the *CCKBR* gene was associated with feed efficiency in cattle (Abo-Ismail *et al.*, 2013). These results can be helpful to confirm the evidence of selection and pre-selecting these SNPs could help to increase the genomic prediction accuracy (Ye *et al.*, 2020) for economically important traits in beef cattle breeding programs.

References

Abo-Ismail M.K., Kelly M.J., Squires E.J., Swanson K.C., Bauck S *et al.* (2013) J Anim Sci 2512-2529. https://doi.org/10.2527/jas2012-5756

Ebara F., Inada S., Morikawa M., Asaoka S.H., Isozaki Y *et al.* (2013) J Anim Physiol Anim Nutr (Berl) 97(4), 684-693. https://doi.org/10.1111/j.1439-0396.2012.01310.x

Iyengar V.K., Reeve H.K., and Eisner T. (2002) Nature 419(6909), 830-832. https://doi.org/10.1038/nature01027

Lim D., Choi B.H., Cho Y.M., Chai H.H., Jang G.W *et al.* (2016) BMB Rep 49(9), 514-519. https://doi.org/10.5483/bmbrep.2016.49.9.074

Lopez M.E., Linderoth T., Norris A., Lhorente J.P., Neira R *et al.* (2019) Front Genet, 10, 901. https://doi.org/10.3389/fgene.2019.00901

Ma Y., Ding X., Qanbari S., Weigend S., Zhang Q *et al.* (2015) Heredity (Edinb) 115(5), 426-436. https://doi.org/10.1038/hdy.2015.42

Mapholi N.O., Maiwashe A., Matika O., Riggio V., Bishop S.C *et al.* (2016) Ticks Tick Borne Dis 7(3), 487-497. https://doi.org/10.1016/j.ttbdis.2016.02.005

Oh D., La B., Lee Y., Byun Y., Lee J. *et al.* (2013) Mol Biol Rep 40(4), 3155-3163. https://doi.org/10.1007/s11033-012-2389-y

Paim T.D.P., Hay E.H.A., Wilson C., Thomas M.G., Kuehn L.A *et al.* (2020) Front Genet 11, 710. https://doi.org/10.3389/fgene.2020.00710

Qanbari S., Gianola D., Hayes B., Schenkel F., Miller S *et al.* (2011) BMC Genomics 12:318. https://doi.org/10.1186/1471-2164-12-318

Qanbari S., and Simianer H. (2014) Livestock Science 166, 133-143.

Shi Y., Zhao P., Dang Y., Li S., Luo L *et al.* (2021) Biol Reprod 105(2), 359-370. https://doi.org/10.1093/biolre/ioab081

Strillacci M.G., Moradi-Shahrbabak H., Davoudi P., Ghoreishifar S.M., Mokhber M *et al.* (2021) BMC Genomics 22(1), 305. https://doi.org/10.1186/s12864-021-07604-3

Tang K., Thornton K.R., and Stoneking M. (2007) PLoS Biol 5(7), e171. https://doi.org/10.1371/journal.pbio.0050171

Voight B.F., Kudaravalli S., Wen X., and Pritchard J.K. (2006) PLoS Biol 4(3), e72. https://doi.org/10.1371/journal.pbio.0040072

Ye S., Song H., Ding X., Zhang Z., and Li Z. (2020) Animal 14(8):1555-64. https://doi.org/10.1017/S1751731120000506

Zhou Y., Connor E.E., Wiggans G.R., Lu Y., Tempelman R.J *et al.* (2018) BMC Genomics 19:314. https://doi.org/10.1186/s12864-018-4699-5

638. Genetic analyses reveal the genetic architecture of the thermotolerance – production complex in beef cattle

R.G. Mateescu[1*], F.M. Rezende[1], K.M. Sarlo Davila[2], A.N. Nunez Andrade[1], A. Hernandez[1], M.A. Elzo[1] and P.A. Oltenacu[1]

[1]University of Florida, Department of Animal Sciences, Gainesville, FL 32608, USA; [2]ORISE/NADC, 1920 Dayton Ave, Ames, IA 50010, USA; raluca@ufl.edu

Abstract

Heat stress is a principal factor limiting production of animal protein in subtropical and tropical regions, and its impact is expected to increase dramatically. Development of effective strategies to improve the ability to cope with heat stress is imperative to enhance productivity of the livestock industry and secure global food supplies. However, selection focused on production and ignoring adaptability results in beef animals with higher metabolic heat production and increased sensitivity to heat stress. The heritabilities estimated in this study in an Angus-Brahman multibreed population demonstrate genetic variation, which supports the hypothesis that selection for improved thermal tolerance is possible. Moreover, the estimated genetic correlations are favorable and indicate the opportunity to develop genomic tools for simultaneous improvement of tolerance to heat stress as well as production.

Introduction

In tropical and subtropical regions where more than half of the world cattle are maintained, climatic stress is a major limiting factor of production efficiency. This stress is expected to increase due to predicted changes in climate. Beef cattle when exposed to environmental high temperature and humidity, exhibit significant declines in feed intake, growth, fertility and welfare. Selection to increase productivity disregarding the genotype × environment interaction is likely to increase susceptibility to climatic stress. This makes the quest for heat-tolerant cattle with increased efficiency of production and reproduction increasingly important. Beef producers in tropical and sub-tropical environments are incorporating a certain proportion of 'indicus' genes in their herds but, without knowledge of genes associated with thermotolerance, this also brings along negative aspects of indicus cattle. Research is needed to uncover the phenotypic and genetic relationships underlying this thermotolerance-production complex and subsequently identify the functional variants for thermotolerance without an antagonistic pleiotropy on production and reproduction. This will allow the incorporation of the G×E interaction in genomic selection programs for improvement of economically important traits in a predicted hotter world.

Animals vary in their ability to dissipate heat and, therefore, in their ability to cope with heat stress, and this variability has a genetic component. The goal of this research is to describe novel traits which can be used to characterize genetic pathways for thermotolerance which are independent or positively associated with production performance.

Materials & methods

Animal population. The University of Florida Institutional Care and Use Committee approved the research protocol (Approval no. 201203578). The population consisted of 330 heifers from the University of Florida multibreed herd (Elzo and Wakeman, 1998; Elzo *et al.*, 2016, 2017) over 2 years in 2017 and 2018. For mating purposes, animals in the multibreed herd are assigned to 6 breed groups based on breed composition: 100% Angus = 100 to 80% Angus; 75% Angus = 79 to 60% Angus; Brangus = 62.5% Angus; 50% Angus = 59 to 40% Angus; 25% Angus = 39 to 20% Angus; and 100% Brahman = 19 to 0% Angus.

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_638, © R.G. Mateescu *et al.* 2022

Heifers were managed similarly across both years. DNA was extracted from blood samples from all animals and genotyped with the Bovine GGP F250 array (Illumina Inc., San Diego, CA, United States).

Skin biopsies. Skin samples were taken during summer (July 17, 2017 and August 7, 2018) between 07:00 and 11:00 h. Skin samples were collected from the back, 4 inches down from spine and halfway along horizontal axis. The skin was cleaned and disinfected with 70% ethanol and chlorhexidine (Clorhexidine 2%; VetOne, Boise, ID). A skin biopsy sample was collected using a 0.6 cm diameter punch biopsy instrument (Biopsy Punch, Miltex Inc., PA) and fixed in 10% formalin for approximately 24 h. Samples were dehydrated in 70% ethanol and infiltrated in liquid paraffin and stored until sectioned and stained at the UF Molecular Pathology Core. Sections were cut on a microtome with a thickness of 7 um, and sections were placed on slides, then stained with Harros-Eosin Hematoxylin. All histological sections were analyzed from digitized images obtained from a Nikon T3000 inverted phase microscope equipped image capture equipment (DMZ1200F with NIS Image Elements software). Images were obtained with the microscope in 40 X, and analyzed with ImageJ software. Sweat gland area (mm^2) and sweat gland depth as the distance from the top of the sweat glands to the skin surface (mm) were determined from a constant 4.6 mm^2 cropped image area.

Hair samples. Hair samples were collected from the shoulder, 4 inches down from spine and halfway along horizontal axis of each animal, as described in Hamblen *et al.* (2018). Hair samples were measured for length using ImageJ software. Five long and 5 short hairs from each individual were measured to evaluate the length of the topcoat and undercoat, respectively. The averages of the 5 short and long hairs were used in the statistical analysis.

Body temperature. Core body temperature was measured as vaginal temperature at 15-min intervals for 5 d using an iButton data logger (Dikmen *et al.*, 2014) inserted into a blank CIDR device and then into the vagina of each animal. Each iButton was calibrated before the study started and pre-programmed to record body temperature at 15-min intervals on a 24-h cycle. Ambient environmental conditions were monitored using HOBO data loggers which continuously record temperature, humidity dew-point temperature with HOBO-U23 data logger (Onset Computer Corp., Bourne, MA), and black globe temperature by using HOBO-U22 data logger. The temperature humidity index (THI) was calculated as in (1):

$$THI = (1.8 \times T + 32) - [(0.55 - 0.0055 \times RH) \times (1.8 \times T - 26)] \quad (1)$$

where T = air temperature (°C) and RH = relative humidity (%). This equation has been shown to be a good indicator of heat stress (Dikmen and Hansen, 2009). Only body temperatures from the 3 continuous days when cattle were on pasture undisturbed were analyzed, as described in Sarlo Davila *et al.* (2019). Based on the thresholds defined by the livestock weather hazard guide and the THI level encountered during our experiment, THI conditions between 84 and 86 were considered high THI. Body temperatures at high THI for each individual were calculated by averaging all measurements collected during the time that the THI windows occurred. This was accomplished for each heifer by averaging the body temperature from all 15-minute windows when the heifer was exposed to a high THI interval.

Carcass traits. A certified technician recorded ultrasound images from yearling calves using an Aloka 500 ultrasound system (Hitachi Aloka Medical, Ltd., Wallingford, Connecticut, USA). Analysis of the ultrasonic images with UICS Scanning Software by Walter and Associates, LLC (Ames, 106 Iowa, USA) yielded yearling ultrasound backfat (UFAT, cm) and yearling ultrasound percent intramuscular fat (UPIMF, %) phenotypes.

Statistical analyses. Average information restricted maximum likelihood (AIREML) variance components, heritabilities, additive genetic correlations, and phenotypic correlations were estimated using single-trait and pairwise two-trait animal linear mixed models. The statistical model for both analyses included the direct additive genetic and residual as random effects, breed group (based on genomic breed composition) and group of data collection as class effect, except for short hair length and skin biopsy records, where group was not significant, and age at measurement as a covariate. The pedigree file consisted of 2,327 individuals, 715 sires and 1,286 dams. All analyses were performed using the airemlf90 package from BLUPF90 software (Misztal *et al.*, 2002).

Results

Heritability estimates for skin histology characteristics, hair characteristics, body temperature under high THI conditions, and ultrasound carcass traits are provided in Table 1. A high heritability of 0.69 was estimated for the sweat gland area while the sweat gland depth had a low heritability estimate of 0.09. Heritability was estimated to be 0.33 for short hair length (undercoat) and 0.16 for long hair length (top coat). Heritability for coat score has been estimated to be 0.6, (Turner and Schleger, 1960) and McEwan Jenkinson *et al.* (1975) estimated the heritability of hair follicle measurements to range from 0.15 to 0.76. The heritability for body temperature under high THI conditions was estimated to be 0.13 which is similar the heritability estimated reported for rectal temperature in a Brahman × Angus crossbred population (0.19; Riley *et al.*, 2012) and dairy cattle (0.17; Dikmen at al., 2012). Both studies utilized cattle located in Florida. High heritability estimates were obtained for backfat (0.76) and intramuscular fat (0.37) ultrasound measures.

Two-trait AIREML estimates of direct additive genetic and phenotypic correlations between skin histology characteristics, hair characteristics, body temperature under high THI conditions, and ultrasound carcass traits are presented in Table 2. Sweat gland area had a negative genetic correlation with sweat gland depth (-0.49), short and long hair length (-0.45 and -0.28, respectively), and body temperature under high THI conditions (-0.65). These negative correlations suggest a similarity in the genetic control underlying these traits which would allow for selection of animals with large sweat glands, short hair (both topcoat and under coat), and able to maintain a lower body temperature under high THI conditions. More importantly, although weak, the genetic correlations between sweat gland area and the two production traits (backfat and intramuscular fat) were favorable (0.22 and 0.20, respectively). Similarly, there was a medium negative genetic correlation between the body temperature under high THI and the two ultrasound carcass traits, suggesting animals able to maintain a lower body temperature would be more productive.

Table 1. Additive genetic variance (σ^2_a), residual variance (σ^2_e), and heritability (h^2) estimates for skin histology characteristics (sweat gland area and depth), hair characteristics (short and long hair length), core body temperature under high THI conditions, and ultrasound carcass traits (backfat thickness and intramuscular fat) with approximate sampling errors (in parentheses).

Trait[1]	σ^2_a	σ^2_e	h^2
Sweat gland area (mm^2)	2.03 (0.62)	0.89 (0.49)	0.69 (0.18)
Sweat gland depth (mm)	0.002 (0.004)	0.02 (0.004)	0.09 (0.15)
Short hair length (mm)	1.95 (1.07)	3.97 (0.99)	0.33 (0.18)
Long hair length (mm)	3.21 (3.39)	16.82 (3.42)	0.16 (0.17)
Temperature at high THI (°C)	0.02 (0.02)	0.10 (0.018)	0.13 (0.15)
UFAT (cm)	0.001 (0.0003)	0.0003 (0.0002)	0.76 (0.19)
UPIMF (%)	0.22 (0.12)	0.38495 (0.11)	0.37 (0.19)

[1] UFAT = ultrasound backfat (cm); UPIMF = ultrasound intramuscular fat (%).

Table 2. Two-trait AIREML estimates of phenotypic (above diagonal) and direct additive genetic (below diagonal) correlations between skin histology properties, hair characteristics, and carcass traits.

Trait[1]	SWGA	SWGD	SHL	LHL	THighTHI	UFAT	UPIMF
SWGA	0.69	-0.18	-0.22	0.02	-0.23	-0.05	-0.13
SWGD	-0.49	0.09	0.32	0.26	0.12	0.08	0.22
SHL	-0.45	0.27	0.33	0.75	0.23	0.07	0.17
LHL	-0.28	0.02	1.00	0.16	0.23	0.04	0.11
THighTHI	-0.65	-0.61	-0.28	-0.45	0.13	-0.17	0.04
UFAT	0.22	-0.57	-0.34	-0.60	-0.38	0.76	0.23
UPIMF	0.20	0.49	0.08	0.09	-0.33	0.42	0.37

[1] SWGA = sweat gland area (mm^2); SWGD = sweat gland depth (mm); SHL = short hair length (mm); LHL = long hair length (mm); THighTHI = temperature at high THI (°C); UFAT = ultrasound backfat (cm); UPIMF = ultrasound intramuscular fat (%).

Discussion

The values of heritability estimated in this study indicate a large, exploitable genetic variance which can be used in selection programs to improve heat tolerance in cattle. Novel traits describing the thermotolerance phenotype such as sweat gland area, short hair length and body temperature under high THI had medium to high heritabilities. More importantly, the genetic correlations estimated in this population are encouraging, indicating favorable relationships between the thermotolerance phenotypes and the production traits. This would suggest that genetic programs to improve resilience to environmental stress could be successful and opportunities exist for simultaneous improvement of production related traits.

References

Dikmen, S., and Hansen, P. J. (2009). J. Dairy Sci. 92, 109–116. https://doi.org/10.3168/jds.2008-1370

Dikmen, S., Khan, F. A., Huson, H. J., Sonstegard, T. S., Moss, J. I., Dahl, G. E., *et al.* (2014). J. Dairy Sci. 97. https://doi.org/10.3168/jds.2014-8087

Elzo, M. A., Mateescu, R. G., Johnson, D. D., Scheffler, T. L., Scheffler, J. M., Carr, C., *et al.* (2017). Livest. Sci. 202, 58–66. https://doi.org/10.1016/j.livsci.2017.05.027

Elzo, M. A., Mateescu, R., Thomas, M. G., Johnson, D. D., Martinez, C. A., Rae, D. O., *et al.* (2016). Livest. Sci. 190, 104–112. https://doi.org/10.1016/j.livsci.2016.06.011

Elzo, M. A., and Wakeman, D. L. (1998). J. Anim. Sci. 76, 1290–1302. https://doi.org/10.2527/1998.7651290x

Hamblen, H., Hansen, P. J., Zolini, A. M., Oltenacu, P. A., & Mateescu, R. G. (2018). J. Anim Sci, 96(8), 3131–3137. https://doi.org/10.1093/jas/sky224

Misztal, I., S. Tsuruta, T. Strabel, B. Auvray, T. Druet, and D. H. Lee. (2002) Proc. 7th World Congr. Genet. Appl. Livest. Prod. Montpellier, France. Communication No. 28–07.

Riley, D. G., Chase, J. C., Coleman, S. W., and Olson, T. A. (2012). Livest. Sci. 148, 109–118. https://doi.org/10.1016/j.livsci.2012.05.017

Sarlo Davila, Kaitlyn M., Hamblen, H., Hansen, P. J., Dikmen, S., Oltenacu, P. A., & Mateescu, R. G. (2019). J. Anim Sci, 97(8), 3246–3252. https://doi.org/10.1093/jas/skz188

Turner, H. G., and Schleger, A. V (1960). Aust. J. Agric. Econ. 11, 645–663. http://dx.doi.org/10.1071/AR9600645

639. The effect of allele ancestry on hip height and body weight of tropical beef cattle

L.R. Porto-Neto[1*], S. McWilliam[1], P.A. Alexandre[1], M.R.S. Fortes[2] and A. Reverter[1]

[1]CSIRO, Agriculture and Food, Queensland Bioscience Precinct, 306 Carmody Rd., St Lucia, Brisbane, QLD 4067, Australia; [2]The University of Queensland, School of Chemistry and Molecular Bioscience, St Lucia, QLD 4072, Australia; laercio.portoneto@csiro.au

Abstract

Bos taurus and *Bos indicus* are the two main types of domestic cattle bred worldwide. Due to their genetic distance to a common ancestor and history of natural and human-oriented selection, they evolved different beneficial characteristics relating to animal production. It is known, in crossbred animals, that the ancestral composition affects several production traits. We defined the ancestral origin of SNP alleles and associated it with two production traits (hip height and body weight) in bulls and cows of tropical composite cattle. Four genomic regions with ancestry-related effects were identified at previously described quantitative trait loci, adding information on their origin. The methodology applied here warrants further investigation using other cattle populations and traits. A better understanding of ancestry effects on production traits opens opportunities for guided selection to accumulate favourable alleles in crossbred cattle.

Introduction

Most beef cattle production systems are based on *Bos taurus*, *Bos indicus*, or their crosses. These two ancestral cattle types (taurine and zebu or indicine) have a long history of separation, different domestication events, and distance to a common ancestral population over 280,000 years ago (Achilli *et al.* 2008, Murray *et al.* 2010). The local environmental adaptation plays a major role in the cattle type distribution worldwide, being the taurine cattle more prevalent in temperate environments, while zebu cattle are primarily bred in the (sub)tropics.

Local adaptation is highly beneficial from an animal production viewpoint. Nonetheless, it often carries productivity trade-offs. For instance, zebu cattle tend to become sexually mature later than taurine cattle. However, they are also more resistant to external parasites. Crossbreeding between these two types of cattle is used to combine the positive attributes of each ancestral type into a single animal that often outperforms their pure-bred counterparts. The proportion of taurus vs indicus ancestry in crossbred cattle has long been known to affect production traits (Porto-Neto *et al.* 2014). However, the precise location of genomic regions driving these effects needs further exploration. This article aims to map genomic regions with variable ancestry patterns and their impact on production traits.

Materials & methods

Animals and traits. The Cooperative Research Centre for Beef Genetic Technologies established the population used in this experiment to explore the genetic links between adaptation and profitability in northern Australia (Barwick *et al.* 2009). We used males and female tropical composite cattle and two trait observations, hip height (HH) and body weight (WT), measured at a mature age (cows) or post-yearling (bulls) (Table 1).

SNP genotypes. The cattle were genotyped using either the Bovine SNP50 v1, v2 or BovineHD (Illumina Inc, San Diego). Animals genotyped using the lower density arrays had their genotypes imputed to HD using an industry reference on 1,220 animals; then up to whole-genome sequence (~22.6 million SNP) using a reference set from 1000 Bull Genome Project run6 (Bouwman *et al.* 2018), 472 animals of 18 breeds.

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_639, © L.R. Porto-Neto *et al.* 2022

Table 1. Summary statistics for hip height (HH, cm) and body weight (WT, kg) in cows (adult) and bulls (post-yearling).

Trait	n	Mean	Standard deviation	Min	Max
HHcow	1,034	137.30	4.68	119.30	154.69
HHbull	1,126	125.20	5.29	103.00	139.00
WTcow	1,037	523.49	55.58	282.40	737.12
WTbull	1,450	325.74	43.76	174.00	550.00

For both imputation rounds, the genotypes were first phased using Eagle (Loh *et al.* 2016), then imputed using Minimac3 (Das *et al.* 2016).

Ancestry calling of SNP alleles and genome-wide association. The haplotype-based software ChromoPaint (Lawson *et al.* 2012) was used to estimate, for each SNP in each animal, the likelihood of those alleles having *Bos taurus* or *Bos indicus* ancestry. Similarly, to the SNP imputation procedure, ChromoPaint uses a reference population to define ancestry. Here we used an extract from the 1000 Bull Genome Project run6 (Bouwman *et al.* 2018) as reference (taurine n=203 of six breeds, and zebu n=101 of three breeds). Then, the likelihood of being taurine that varies from 0 to 1 was recoded to a binary format using hard thresholds. If the likelihood was >0.75 the allele was called taurine (or 'A'), if <0.25 zebu (or 'B'), and between 0.25 and 0.75 undefined (or '0'). Only SNP with both alleles defined were considered for the genome-wide association analyses. The bi-allelic form of ancestry was used in a mixed linear model using GCTA (Yang *et al.* 2011). The mixed models included the genomic relationship matrix based on ~50K SNP genotypes (BovineSNP50), and the population and trait specific fixed and covariate effects.

Results & discussion

The considerable genetic distance between taurine and zebu cattle allowed ChromoPaint (Lawson *et al.* 2012) to define the allele ancestry of most SNP sites. On average per animal, ancestry was defined for 92% (SD=8%) of SNP. Ancestral definition was not constant along the genome, with some alleles having its ancestry definition on as low as 9% of the animals. Some genomic regions had poor ancestral definition such as parts of BTA5 and BTA15 (data not shown). These lowly differentiated regions might represent genomic regions that are less differentiated between taurine and zebu cattle either by a common selection or simply due to no selective pressure.

After recoding the likelihood of ancestry for each allele into the binary form (*taurus* vs *indicus*), and removing those undefined sites, we run genome-wide associations for each of the four traits. Inspecting the results, two chromosome regions were highly significant for most traits, a region on BTA5 and BTA14 (Figure 1).

Applying the significance threshold of $P<1.0E-7$ to the association analyses, four ancestry-related quantitative trait loci could be defined, with some regions encompassing many SNP (Table 2). If the significance threshold is relaxed, the three regions on BTA5 would be merged into one that previously has been associated with several production traits (Bolormaa *et al.* 2011, Porto-Neto *et al.* 2014, Aguiar *et al.* 2018). Our analyses also confirmed reported results showing the relevance of a region on BTA14. This region had been associated with human and bovine stature (Pryce *et al.* 2011, Bouwman *et al.* 2018) and linked to a taurine-derived allele effect on traits (Fortes *et al.* 2013), among several other associations to production traits. However, the exact gene that was suggested before (*PLAG1*) is close but outside the critical region on our analyses.

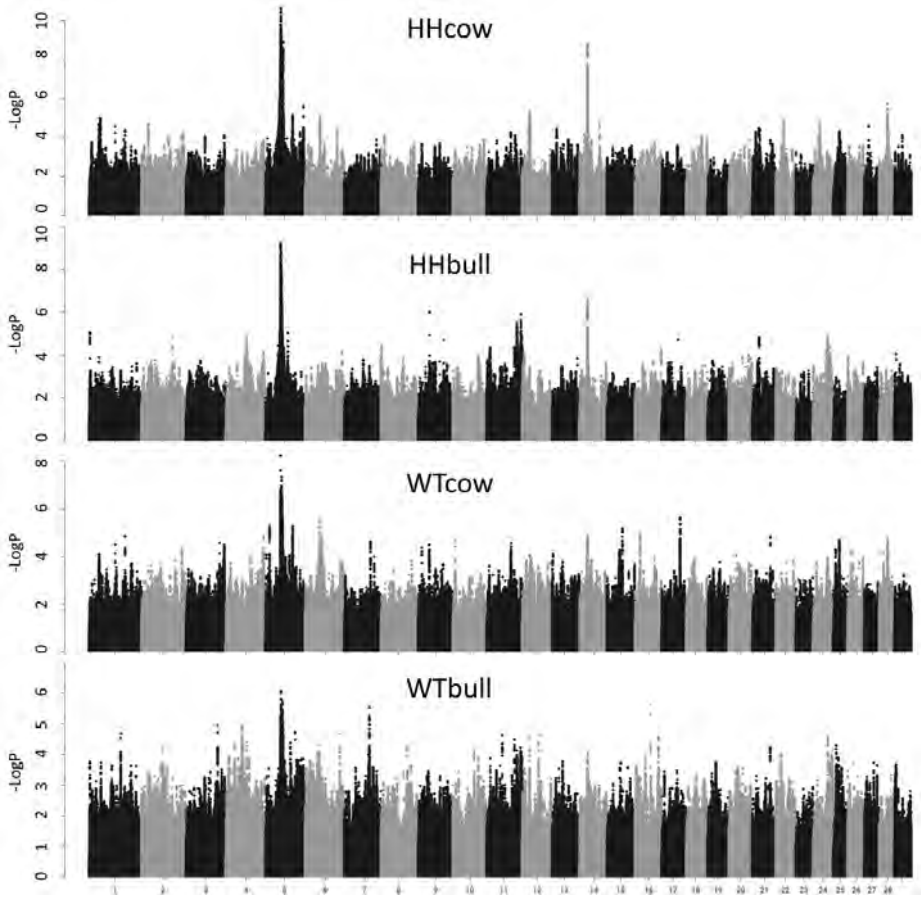


Figure 1. Manhattan plot of genome-wide association of ancestry on hip height (HH) and Weight (WT) of cows and bulls. Autosomes are in alternate colours from 1, on the left, to 29.

Table 2. Summary of the genomic regions' ancestry associated to production traits.

BTA	Main peak			N SNP <1.0E-7	Transcript type (number)	Trait[1]
	Start (Mbp)	End (Mbp)	Lowest *P*-value			
5	46.222	46.339	6.20E-10	1,902	protein coding (1)	HHbull
5	47.289	47.362	2.40E-11	4,741	0	HHcow
5	49.285	49.285	4.80E-08	14	protein coding (1)	WTcow
14	24.899	24.995	1.60E-09	272	protein coding (3), snoRNA (1), snRNA (1)	HHcow

[1] Hip height and body weight in bulls and cows.

It is worth noting that some alleles related to differential ancestry have been selected in composite cattle populations. Bolormaa *et al.* (2011) observed the accumulation of genomic segments carrying the favourable alleles in tropical composite cattle. A similar observation was done in Brahman cattle for the region on BTA14 linked to stature and BTA1 at the polled locus (Koufariotis *et al.* 2018).

The methodology applied here confirmed the relevance of considering the ancestral origin of alleles in association analyses. Further investigation could include other cattle populations and more production traits, under natural and human-guided selection. Further, it will be informative to discern the direction of each effect, to determine the ancestral origin conferring the desirable production trait. A better understanding of the effects of ancestry on production traits could assist in breeding crossbred cattle that accumulate favourable alleles, improving the sustainability of animal production systems.

References

Achilli A., Olivieri A., Pellecchia M., Uboldi C., Colli L. *et al.* (2008) Curr. Biol. 18(4):R157-8. https://doi.org/10.1016/j.cub.2008.01.019.

Aguiar T.S., Torrecilha R.B.P., Milanesi M., Utsunomiya A.T.H., Trigo B.B. *et al.* (2018) Front. Genet. 9:627. https://doi.org/10.3389/fgene.2018.00627.

Barwick S.A., Johnston D.J., Burrow H.M., Holroyd R.G., Fordyce G. *et al.* (2009) Anim. Prod. Sci. 49(5-6):367-382. https://doi.org/10.1071/EA08273.

Bolormaa S., Hayes B.J., Hawken R.J., Zhang Y., Reverter A. *et al.* (2011) J. Anim. Sci. 89(7):2050-60. https://doi.org/10.2527/jas.2010-3363.

Bouwman A.C., Daetwyler H.D., Chamberlain A.J., Ponce C.H., Sargolzaei M. *et al.* (2018) Nat. Genet. 50(3):362-367. https://doi.org/10.1038/s41588-018-0056-5.

Das S., Forer L., Schonherr S., Sidore C., Locke A.E. *et al.* (2016) Nat. Genet. 48(10):1284-1287. https://doi.org/10.1038/ng.3656.

Fortes M.R., Kemper K., Sasazaki S., Reverter A., Pryce J. *et al.* (2013) Anim. Genet. 44(6):636-47. https://doi.org/10.1111/age.12075.

Koufariotis L., Hayes B.J., Kelly M., Burns B.M., Lyons R. *et al.* (2018) Sci Rep 8(1):17761. https://doi.org/10.1038/s41598-018-35698-5.

Lawson D.J., Hellenthal G., Myers S., and Falush D. (2012) PLoS Genet. 8(1):e1002453. https://doi.org/10.1371/journal.pgen.1002453.

Loh P.R., Palamara P.F., and Price A.L. (2016) Nat. Genet. 48(7):811-6. https://doi.org/10.1038/ng.3571.

Murray C., Huerta-Sanchez E., Casey F., and Bradley D.G. (2010) Philos. Trans. R. Soc. Lond., B, Biol. Sci. 365(1552):2531-9. https://doi.org/10.1098/rstb.2010.0103.

Porto-Neto L.R., Reverter A., Prayaga K.C., Chan E.K.F., Johnston D.J. *et al.* (2014) Plos One 9(11):e113284. https://doi.org/10.1371/journal.pone.0113284.

Pryce J.E., Hayes B., Bolormaa S. and Goddard M.E. (2011) Genetics 187(3):981-4. https://doi.org/10.1534/genetics.110.123943.

Yang J.A., Lee S.H., Goddard M.E., and Visscher P.M. (2011) Am. J. Hum. Genet. 88(1):76-82. https://doi.org/10.1016/j.ajhg.2010.11.011.

640. Mining of producer recorded data; using beef calf and cow live-weight data as a case study

S. Walsh[1,2*], N. McHugh[1], R.D. Evans[3], C. Murphy[2] and D.P. Berry[1]

[1]Teagasc, Animal & Grassland Research and Innovation Centre, Moorepark, Fermoy, P61 P302, Co. Cork, Ireland; [2]Department of Biological Sciences, Munster Technological University, Bishopstown Campus, T12 P928, Co. Cork, Ireland; [3]Irish Cattle Breeding Federation, Highfield House, Shinagh, Bandon, P72 X050, Co. Cork, Ireland; shauna.walsh@teagasc.ie

Abstract

The objective of this study was to develop a novel editing criteria for anomaly detection of beef cow and calf live-weight data. Methods included generation of within-herd regression estimates and partial correlations between cow and calf live-weight records; five sources of live-weight data (i.e. professionally-recorded, owned-scales, borrowed-scales, scales hired from a depot, other) were investigated. Results suggest anomaly detection should use both within-herd regression and partial correlation estimates jointly. A contingency table was formed with herds grouped into five classes according to their herd-level regression and partial correlation estimates between cow and calf live-weight. The distribution of herds within each level of the table was the same irrespective of the source of the live-weight data, suggesting that data source did not impact the data quality. This editing criteria developed can nonetheless be used to quality control data at an individual herd-level.

Introduction

Animal live-weight contributes substantially to profitability in beef herds and is a key determinant of the overall efficiency of the beef sector (McHugh *et al.*, 2012). Calf live-weight provides an indication how soon an animal will reach slaughter, while the live-weight of the cow can directly affect its feed intake and maintenance requirements and, by extension, the herd stocking rate. Live-weight of the beef calf, cow or a combination of both has been studied extensively with both phenotypes used frequently in beef genetic evaluations globally. Live-weight can be recorded through various media such as through livestock marts (McHugh *et al.*, 2010), recorded by professionals or by producers themselves or even inferred from carcass data. Irrespective of the source of the data, errors will exist and these can impact downstream analyses. Detection of outliers or anomalies in farmer-recorded data can, for example, improve the predictability and accuracy of genetic evaluations (McHugh *et al.*, 2017; Ribeiro *et al.*, 2021). The objective of the present study was to develop a novel editing criteria for anomaly detection of beef cow and calf live-weight data. Results from this study will contribute to the formation of a specific editing criteria for live-weight data in national genetic evaluations.

Materials & methods

Data. A total of 754,118 live-weight records from 371,212 beef cows and 382,906 beef calves collected from 17,008 herds in the year 2019 were available. The live-weight data for both calf and cow were recorded on the same day; several options existed to record the live-weight data and these were categorised into five data sources (Table 1). Beef calves in this study were defined as singleton-born, aged between 50 and 365 days and weighing between 70 and 700 kg. Only beef dams from parities 1 to 15 and weighing between 300 and 1,200 kg were retained. Dam parities were subsequently categorised as 1, 2, 3, 4 or ≥5.

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_640, © S. Walsh *et al.* 2022

Table 1. Description of sources of live-weight data and number (N) of records in each.

Type	Description	n
Professional	Professionally trained individuals record weights for producers on-farm	17,740
Owned Scales	Producers that own their own scales	94,573
Borrowed Scales	Producers that borrow scales	76,246
Depot Scales	Producers that rent scales	87,493
Other	Data recorded without a defined weighing method	67,167

Only herds that recorded at least 5 calf-cow pairs of live-weight records, recorded on the same day were retained; this resulted in the removal of 4,304 herds. In all instances, professionally recorded live-weight data were assumed to represent the gold standard. Live-weight records more than ±3 standard deviations (SD) from the mean of the professionally recorded weights were initially discarded; 1,759 total calf-dam pairs were removed. Following all edits, 327,354 cow-calf weight pairs remained.

Anomaly detection. A series of strategies were used to detect live-weight anomalies. Firstly, to ensure the greatest possible quantity of data was retained, a mixed regression model was fitted to the professionally recorded live-weight data only:

$$Y = \mu + Age + Age\times Parity + Month + Sex + Sex\times Age + Sex\times DamWt + Parity\times DamWt + SirePTA + SirePTA\times sex + DamWt\ (Herd) + e \quad (1)$$

where Y is calf live-weight, μ is the population mean, Age is calf age at weighing in days, Parity is the parity of the dam, Month is calendar month of weighing, Sex is calf sex, DamWt is the covariate of dam live-weight, SirePTA is the covariate of the PTA of the calf's sire for carcass weight, DamWt is the random term for DamWt by Herd of weighing and e is the residual term. From the resultant random regression herd estimates, only herds with a random regression coefficient within ±3 standard deviations (SD) of the mean random regression estimates of the professionally recorded live-weight data were retained; this resulted in the removal of four herds.

Following the removal of the four outlier herds, the mixed regression model as described in Equation 1 was reran, but without the random terms of dam weight and herd. The fixed effect solutions from the mixed model were then subtracted from the calf live-weight data in the entire dataset.

A similar approach was taken to derive an adjusted dam live-weight, in that the mixed model which was fitted to the professionally recorded live-weight data only:

$$Y = \mu + Days + Days\times Parity + Month + Sex + Sex\times Days + Parity + e \quad (2)$$

where Y is dam live-weight, μ is the population mean, Days is days since calving, Parity is dam parity, Month is calendar month of weighing, Sex is calf sex and e is the residual term.

For the newly derived adjusted calf live-weight trait, a random regression mixed model was fitted to generate a random regression estimate per herd using the model:

$$Y = \mu + DamWt\ (Herd) + e \quad (3)$$

where Y is the adjusted live-weight, μ is the population mean, DamWt(Herd) is the linear random regression on dam weight by herd and e is the residual term.

In a second set of analyses, two generalised linear models were used with (a) the newly derived adjusted calf live-weight as the dependent variable, with no fixed effects included and (b) the newly derived adjusted dam live-weight as the dependent variable, with no fixed effects included to determine the residuals. The within herd partial correlations were then estimated between the residuals of the adjusted calf live-weight and the residuals of the adjusted dam live-weight. From the estimated herd-level random regression estimates and partial correlations between the adjusted calf and dam live-weight, a contingency table was constructed whereby herds were categorised into five distinct classes for the random regression estimates and five classes for the partial correlations (Table 2). The five classes were determined based on the distance of each herd from the standard deviation of the professional recorded data for both the regression and partial correlation estimates separately.

Results

Across all sources of live-weight data, the average age and weight of calves was 174 days and 258 kg (SD=74 kg), respectively; the average calf weight of the professionally recorded data was 253 kg. Across the five live-weight sources, the average herd-level random regression estimates for calf live-weight regressed on dam live-weight ranged from 0.009 to 0.035 (Figure 1a). The average partial correlation estimated per herd between the calf and dam live-weight residuals ranged from 0.130 to 0.324 across the five live-weight source (Figure 1b).

From the contingency table (Table 2), 10.21% of herds ranked in the very low and very high class for both the herd-level random regression and partial correlation estimates between calf and dam live-weight residuals; 3.51% herds ranked within the average class for each. Across both the herd-level random regression and partial correlation estimates, the proportion of herds within each of the five live-weight sources that ranked within very low or very high classes varied from 0.56% (professionals) to 3.04% (other).

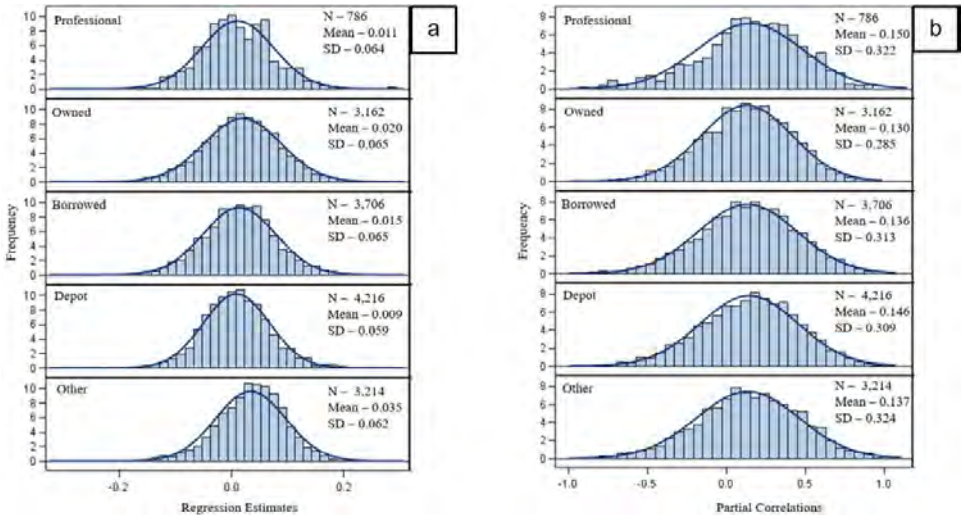


Figure 1. Frequency distribution of herd-level (a) random regression estimates of calf live-weight regressed on dam live-weight and (b) partial correlations between calf and dam live-weight residuals for each of the five data sources.

Table 2. Contingency table representing the percentage of herds within each of the five categories (very low, low, average, high and very high) for both the random regression estimates and partial correlations between calf and dam live-weight residual per herd.

		Partial correlation				
		Very low	Low	Average	High	Very high
Regression estimates	Very low	1.29	1.62	2.35	3.08	4.02
	Low	1.60	1.95	2.92	3.98	5.69
	Average	1.84	2.72	3.51	4.98	6.93
	High	2.07	3.08	4.47	5.76	8.16
	Very high	2.61	3.98	5.52	6.95	8.92

Discussion

Ensuring the integrity of the growing quantity of data being generated from multiple sources is a growing challenge. This is important not only for management purposes, but also for use in genetic evaluations. Approaches of anomaly detection in large datasets are therefore required. Although Ribeiro *et al.* (2021) investigated methods of anomaly detection using simulated data, no such analysis has been investigated using cattle live-weight on a nationally recorded dataset. Regression estimates and partial correlations between calf and dam live-weight were used to detect atypical data or indeed herds with atypical data. The contingency table revealed that only 21% of herds ranked within the same category using the partial correlations and random regression estimates, both calculated per herd for calf live-weight regressed on dam live-weight. This implies that anomaly detection using calf and dam live-weight data will vary depending on whether herd-level random regressions or partial correlations between calf and dam live-weight residuals are used separately or jointly. These results are not unexpected given that the regression estimate can be derived once the correlation and standard deviations are known for both traits (Bewick *et al.*, 2003). Regardless of the source of live-weight data, the percentage of herds that ranked below or above average in the contingency table (Table 2) (i.e. relative to the professional scored herds) was 28 and 51%, respectively. However, based on the contingency table a similar number of herds irrespective of live-weight source ranked across classes; this indicates that the sources of live-weight did not determine the quality of the data recorded by each herd. The use of this novel editing criteria has the potential to highlight herds that are providing poor data. Better data recording can be enhanced through herd-level inspections, providing higher data quality for national genetic evaluations.

Acknowledgements

Funding from the Irish Department of Agriculture, Food and Marine Research Fund DAIRY4BEEF (19/R/553) is gratefully acknowledged.

References

Berry D.P., Evans R.D., and Kelleher M.M. (2021) J. Dairy Sci. 104(6):6885-6896. https://doi.org/10.3168/jds.2021-20154

Bewick V., Cheek L., and Ball J. (2003) Crit. Care 7:451-459. https://doi.org/10.1186/cc2401

Mc Hugh N., Fahey A., Evans R., and Berry D.P. (2010) Anim. 4(8):1378-1389. https://doi.org/10.1017/S1751731110000297

Mc Hugh N., Evans R.D., Fahey A.G., and Berry D.P. (2012) Livest. Sci. 144(1–2):11-19. https://doi.org/10.1016/j.livsci.2011.10.006

McHugh N., Pabiou T., Wall E., McDermott K., and Berry D.P. (2017) J. Anim. Sci. 95(5): 1926–1938. https://doi.org/10.2527/jas.2016.1344

Ribeiro A., Golden B.L., and Spangler M.L. (2021) J. Anim. Sci. 99(3): https://doi.org/10.1093/jas/skab053

641. Sexual precocity of purebred Charolais heifers

S. Taussat[1,2*], A. Vinet[2], D. Maupetit[3], D. Dozias[4], L. Chanteloup[4], G. Renand[2] and P. Martin[2]

[1]Allice, 149 rue de Bercy, 75012 Paris, France; [2]Université Paris-Saclay, INRAE, AgroParisTech, GABI, Domaine de Vilvert, 78350 Jouy-en-Josas, France; [3]UE0332 Domaine Expérimental Bourges-La Sapinière, INRAE, La Sapinière, 18390 Osmoy, France; [4]UE0326 Domaine Expérimental du Pin-au-Haras, INRAE, Le Pin-au-Haras, 61310 Exmes, France; sebastien.taussat@inrae.fr

Abstract

Improving sexual precocity can reduce unproductive lifetime and thus may increase farm profitability. The objective of this study was to estimate the genetic variability of sexual precocity and its relationship with pre and post-weaning performance traits. A total of 633 purebred Charolais heifers was used in this study. Sexual precocity was defined as month class from 1 (first oestrus at 14 months) to 8 (at 21 months or later). Genetic parameters were estimated using WOMBAT software. Sexual precocity was moderately heritable (0.20±0.09). Genetically, sexual precocity was positively correlated with birth weight (0.29±0.21) and average daily gain at 14 months (0.41±0.26) and negatively with backfat thickness (-0.27±0.26). Environmentally, all traits were negatively correlated with sexual precocity. Combining genetic selection of sexual precocity with good environmental conditions for raising animals seems to be the solution for reducing early unproductive lifetime of females.

Introduction

Reducing the duration of unproductive life in beef cattle could lead to an increase farm profitability. A key component of this unproductive duration in early life is the sexual precocity: early puberty enables early first mating. On one hand, age at puberty is influenced by environmental conditions, such as diet at juvenile stage (Cardoso *et al.*, 2014). On the other hand, genetic variability of age at puberty exists among French Charolais heifers (Mialon *et al.*, 1999b; Phocas and Sapa, 2004). To understand the genetic determinism of sexual precocity, an experiment was conducted with purebred Charolais heifers to estimate genetic relationship with several pre and post-weaning performance traits.

Materials & methods

Animal management and phenotypes. Between 2011 and 2015, 633 purebred Charolais heifers were born and raised in two experimental farms of the French National Institute for Agriculture, Food and Environment (INRAE): 305 animals at Bourges-La Sapinière farm (https://doi.org/10.15454/1.5483259352597417E12) and 328 animals at Le Pin-au-Haras farm (https://doi.org/10.15454/1.5483257052131956E12). Two seasons of birth were defined: winter (October to January) and spring (February to May). All females were weighted at birth (BW), double weighted at the weaning (WW) and every months during the experiment. Two average daily gains were calculated with regression of weights over time with R software (version 3.5.0): a pre-weaning one between BW and WW (ADGW) and a post-weaning one between WW and 14 months of age (ADG14). The last calculation was used to establish the weight at 14 months (W14) for everyone, the animals being not always weighted at the exact same age.

At 14 months of age, metric body measurements were performed on all females such as wither height (WH), trochanter width (TW) and body length (BL). At the same time, backfat thickness (BF) was assessed by ultrasound measurements at the fourth lumbar and body condition score (BCS) was assessed from palpation by trained technicians on a 5-point scale, from 0 (leanest animals) to 5 (fattest animals) (Agabriel *et al.*, 1986). Blood samples were collected at every 10 days from weaning to 20 months of age, and blood progesterone concentration was determined by radioimmunoassay (Terqui and Thimonier, 1974) to assess

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_641, © S. Taussat *et al.* 2022

puberty. That trait was defined as month puberty class (PC) from 1 (first oestrus at 14 months) to 8 (first oestrus at 21 months or later).

Statistical analysis. Genetic parameters were estimated using the restricted estimation of maximum likelihood (REML) method with WOMBAT software (Meyer, 2007). The model included a contemporary group, combination of farm, year and season of birth (18 groups), being born twin or singleton and age of dam at birth (classes of 3, 4, 5, 6, 7 and 8 years and older) as fixed effect for all traits, age at the weaning as covariate for WW, age at the last weight for W14, age at the mensuration as covariate for WH, TW and BL and random effect of dam (432 animals) for BW, WW and ADGW. Phenotypic correlations were calculated with R software (version 3.5.0) with corrected phenotypes using the same fixed effects and covariates as the genetic model.

Results

Mean PC was 4.5 (±2.1), however a large effect of season could be observed (Figure 1). Figure 1 presents the evolution of number of females that reached puberty at a given age regarding the season of birth. Females born in spring season were in oestrus earlier than heifers born in winter season (3.8±2.0 vs 5.5±1.7 months). As a consequence, most of the spring-born heifers were in oestrus at 17 months, while that age corresponds to the mean age at puberty of the winter-born heifers.

Descriptive statistics and heritability for all traits and phenotypic and genetic correlations with PC are presented in Table 1. The highest heritability estimated was for BW (0.51±0.12) and the lowest was for BL (0.13±0.08). Heritability of PC was intermediate (0.27±0.10). It should be noted that all traits were negatively environmentally correlated with PC. Phenotypically, all traits were also negatively correlated with PC, except BW which was independent. Genetically, low correlations were estimated except positive correlations with ADG14 (0.41±0.26) and BW (0.29±0.21) and a negative correlation with BF (-0.27±0.26).

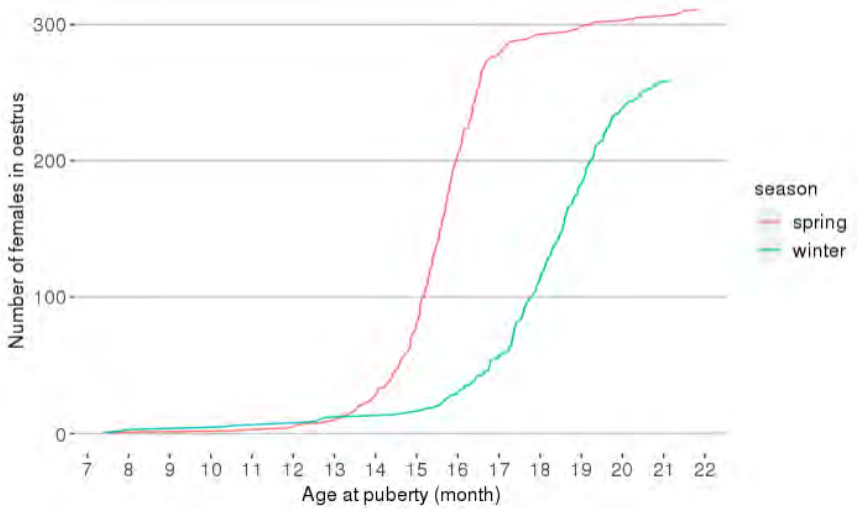


Figure 1. Number of heifers that reach puberty according to the age and the season of birth.

Table 1. Mean, standard deviation and heritability (± SE[1]) of traits and phenotypic, genetic (± SE) and environmental (± SE) correlations with month puberty class.

Trait[2]	Number	Mean	Std	h^2	r_p[3]	r_g	r_e
PC	617	4.5	2.1	0.27±0.10	-	-	-
BW	633	46	6	0.51±0.12	0.02	0.29±0.21	-0.18±0.16
WW	633	266	41	0.30±0.10	-0.34	0.03±0.28	-0.54±0.09
ADGW	633	960.6	174.2	0.30±0.10	-0.37	-0.02±0.28	-0.58±0.08
W14	632	369	51	0.38±0.11	-0.38	0.23±0.25	-0.69±0.09
ADG14	632	546.1	184.8	0.26±0.09	-0.25	0.41±0.26	-0.52±0.09
WH	626	114.1	3.9	0.32±0.11	-0.30	0.15±0.26	-0.48±0.10
TW	627	41.3	3.0	0.37±0.11	-0.31	0.00±0.36	-0.40±0.07
BL	627	129.8	13.7	0.13±0.08	-0.33	0.18±0.35	-0.47±0.07
BF	555	6.9	1.4	0.29±0.11	-0.30	-0.27±0.26	-0.31±0.09
BCS	627	2.2	0.4	0.26±0.10	-0.29	0.11±0.28	-0.42±0.08

[1] SE = standard error; Std = standard deviation; h^2 = heritability; r_p = phenotypic correlation; r_g = genetic correlation; r_e = environmental correlation.
[2] PC = month puberty class; BW = birth weight (kg); WW = weaning weight (kg); ADGW = average daily gain at weaning (g/d); W14 = weight at 14 months (kg); ADG14 = average daily gain at 14 months (g/d); WH = wither height (cm); TW = trochanter width (cm); BL = body length (cm); BF = backfat thickness (mm); BCS = body condition score.
[3] $|rp|<0.03$: P not significant; $|rp|>0.24$: $P<0.0001$.

Discussion

The average age at puberty found in this study was higher than values previously reported in the literature in the same breed. For example, 56% of Charolais heifers had a positive progesterone test at 15 months in the study of Mialon *et al.* (1999a) and 60% were observed in oestrus at 15 months in the study of Phocas et Sapa (2004). This difference may be due to the existence of two birth seasons in this experiment which had an impact on the onset of puberty: heifers born in spring were in oestrus earlier. All animals were kept indoor in winter and started grazing in March. Therefore, environmental conditions were better in spring and thus, females began to be in oestrus at this moment. Heifers born in winter were older when the environmental conditions became favourable. In the study of Mialon *et al.* (1999a), all heifers were born from mid-January to mid-April, and their results were similar to our heifers born in spring. Moreover, Cardoso *et al.* (2014) tested effect of different diet treatments on sexual precocity and highlighted that a critical window existed between 4 and 6.5 months of age. In our study, spring-born heifers were raised outdoor between birth and weaning. This practice would therefore improve the sexual precocity of heifers by optimizing the development of animals before weaning.

These results influenced the estimation of genetic parameters: for environmental correlations, all traits were indeed negatively correlated with PC. If the environment is favourable for heifers, they will have better growth and thus will be in oestrus earlier. Despite this strong effect of environmental conditions, some genetic variability was highlighted in this study, with a moderate heritability estimated for PC. Interestingly, this trait was unfavourably genetically correlated with ADG14, meaning that animals genetically driven to have higher growth between weaning and 14 months of age were also driven to an older puberty. Same trend was observed by Johnston *et al.* (2009) with an unfavourable genetic relationship between dry-season growth rate and age at corpus luteum in Brahman breed. However in the study of Mialon *et al.* (1999b), they found favourable genetic correlation between age at puberty and ADG between weaning and 12 months of -0.39 (±0.18). Sexual precocity was negatively correlated with BF, therefore heifers with higher backfat in 4th lumbar were in oestrus earlier. However, no genetic relationship was found with BCS in our study. Johnston *et al.* (2009) also highlighted negative relationship between fat trait and age at corpus luteum.

This study highlighted that sexual precocity was influenced by both environmental conditions and genetics. A genetic selection for sexual precocity combined with favourable environmental conditions can help decreasing age at puberty and therefore contribute to reduce unproductive life of females.

References

Agabriel J., Giraud J. M., Petit M., Barboiron C, Coulaud G., *et al.* (1986) Bull. Tech. CRZV Theix 66:43–50.

Cardoso R.C., Alves B.R.C., Prezotto L.D., Thorson J.F., Tedeschi L.O., *et al.* (2014) J. Anim. Sci. 92:2942–2949. https://doi.org/10.2527/jas.2014-7713

Johnston D.J., Barwick S.A., Corbet N.J., Fordyce G., Holroyd R.G., *et al.* (2009) Anim. Prod. Sci. 49:399–412. https://doi.org/10.1071/EA08276

Meyer K. (2007) J. Zhejiang Univ. Sci. B 8:815–821. https://doi.org/10.1631/jzus.2007.B0815

Mialon M.-M., Renand G., Krauss D., and Ménissier F. (1999a) Ann. Zootech. 48:413–426. https://doi.org/10.1051/animres:19990601

Mialon M.-M., Renand G., Krauss D., and Ménissier F. (1999b) Ann. Zootech. 48:427–434. https://doi.org/10.1051/animres:19990602

Phocas F., and Sapa J. (2004) Anim. Sci. 79:41–48. https://doi.org/10.1017/S1357729800054515

Terqui M., and Thimonier J. (1974) CR. Acad. Sc. 279:1109–1112.

642. Estimation of variance components for female longevity in Australian Angus cattle using random regression models

H. Aliloo* and S. Clark

School of Environmental and Rural Science, University of New England, Armidale, NSW 2350, Australia; haliloo@une.edu.au

Abstract

Cow longevity is an economically important trait for beef cattle breeding. In this study, random regression models were used to estimate genetic parameters for female longevity in Australian Angus cattle. Longevity was defined using 10 binary records (0 for 'not in the breeding herd' and 1 for 'active in the breeding herd') between age 2 to 11. A Bayesian random regression model was fitted separately to 5 different groups of disposal reasons supplied by breeders. The heritability estimates were generally low with peaks ranging from 0.03 to 0.12. The highest and lowest heritability estimates were obtained for performance and fertility datasets, respectively. Genetic correlations were high between ages 3 to 5 and age 6 for all groups except for the structural problems group in which age 3 showed a low genetic correlation. This illustrates that selection for longevity could be undertaken early in cow's life.

Introduction

Cow longevity is a complex, economically important trait for beef cattle breeding programs. It directly relates to lifetime productivity and farm profitability by decreasing the ratio of rearing costs to production profit per animal. Improving cow longevity leads to improved animal welfare and a more sustainable beef production. Improving cow longevity also impacts the breeding program by reducing the number of replacement heifers required, for a constant herd size, which provides more opportunities for voluntarily culling based on animal performance.

Different definitions of longevity have been used in livestock breeding programs. In cattle, the most common measurements are lifetime (e.g. length of productive life) and stayability traits. Survival analysis using a proportional hazard model (PHM) is a popular method used for genetic evaluation of lifetime traits (Ducrocq *et al.*, 1998). Veerkamp *et al.* (2001) showed the equivalence between a PHM and a random regression model (RRM) when applied to binary representation of survival data (i.e. stayability). RRM were further compared to PHM by other studies where some found advantageous features of RRM (Jamrozik *et al.*, 2008) and suggested them as the method of choice for the genetic analysis of longevity traits.

The objective of this study was to apply RRMs for estimation of variance components and genetic parameters for female longevity in Australian Angus cattle population and to inform an optimal age for selection for cow longevity.

Materials & methods

Data. Raw culling and calving information of 1,438,363 cows were extracted from the Angus Australia database. For data cleaning, animals with missing date of birth, missing date of disposal or without a specific disposal reason were removed. Cows born before 2000, and those with an age at disposal of less than 19 months or greater than 20 years were excluded. The progeny records with a missing date of birth, sire and dam or those bred via an embryo transfer program were removed. Only calving records from cows with an age at first calving between 19-30 months were kept. A maximum of 2 missing calvings was permitted between two reported calving events and between the last reported calving and culling for each cow. This resulted in a total of 249,711 real calving events reported for 69,316 cows with known disposal

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_642, © H. Aliloo and S. Clark 2022

reasons. Individual disposal reasons were categorized into 4 groups based on disposal codes supplied by breeders. This included disposal because of fertility, structural problems, animal performance and natural death (Table 1). All disposal codes were also combined into a single longevity analysis. Longevity was defined as stayability to consecutive ages. Animals present in the herd after first calving were coded as 1 at age *t* if the animal avoided disposal from age *t-1* to age *t* and as 0 if the animal was disposed from the breeding herd before age *t*. Initial investigation showed almost no variation in data after age 11. Therefore, for each animal 10 binary records were assigned from age 2 to 11 years.

Statistical model. A single-trait random regression model was fitted to estimate variance components using a Bayesian Gibbs sampler from gibbs3f90 software (Misztal *et al.*, 2018):

$$y=Xb+Hh+Za+Wp+e \quad (1)$$

where y is the vector of phenotypes, b is the vector of estimates for fixed class effect of embryo transfer and fixed regression coefficients for year-season of cow birth, h, a and p are random regression coefficients for herd of calving within year-season of cow birth, animal additive genetic and permanent environment deviations, respectively, and e is the vector of random residuals. Initial analyses were undertaken to determine the appropriate number of Legendre polynomials to use in the model. The same order of 4 Legendre polynomials was then used for fixed and all random regression terms. Random residual variance was assumed to be heterogeneous (i.e. 10 classes 1 each for each age). A total of 150,000 rounds with a burn-in of 50,000 and a thinning of 10 were used for the MCMC chain. The variance components, genetic correlations, breeding values and heritabilites were calculated using the appropriate formulations (Oliveira *et al.*, 2020).

Results

The number of animals in each disposal group are included in Table 1. The largest number of animals with known disposal reasons (~52%) were removed from the breeding herd because of fertility problems with natural death (~35%) being the second largest group. Only around 6% of cows were disposed because of voluntarily culling i.e. performance.

The estimates of additive genetic variance and heritabilites are shown in Figure 1. For all groups of disposal reasons additive genetic variance was small resulting in low heritability estimates across all ages. The highest additive genetic variance was found at age 4 for fertility, structure and performance datasets while for natural death and all reasons, the highest additive genetic variance was found at age 5 and 7, respectively. Between all groups of disposal reasons, the group with voluntarily culling reasons (i.e. performance) resulted in the highest heritability estimate while the heritability for the fertility dataset was the lowest. The heritability estimates for all disposal reasons were similar to those from natural death. High heritability values were estimated at extreme ages for animals culled due to structural problems. This was because of low amounts of phenotypic variation at extreme ages.

Table 1. The number of cows within each culling reason group.

Culling group	Number of cows
Fertility	36,442
Natural death	24,497
Performance	3,847
Structural problems	4,530
All reasons	69,316

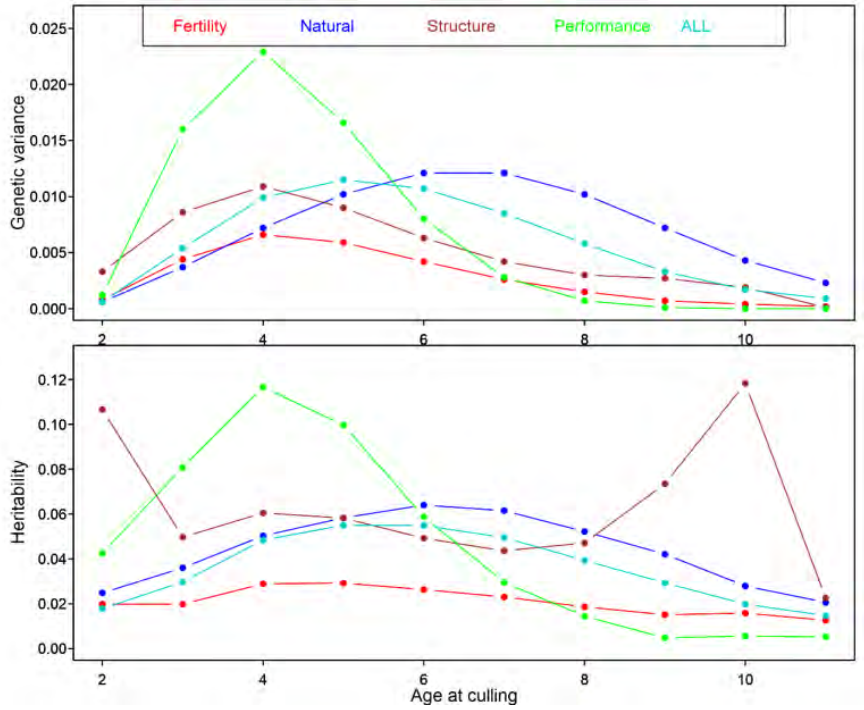


Figure 1. The estimated additive genetic variance (top) and heritabilites (bottom) for fertility (red), natural death (blue), structural problems (brown), performance (green) and all reasons (pink) disposal groups.

The genetic correlations between ages 3 to 5 and age 6 for all groups of disposal reasons are given in Table 2. As expected, for all groups of disposal reasons higher genetic correlations were observed between age 6 and more adjacent ages. Genetic correlations were lower in group of structural problems and higher in groups of performance and all reasons.

Discussion

Improving longevity can increase profitability by decreasing the involuntarily culling and therefore enabling more voluntarily culling. It is, therefore, very important to provide more specific reasons for culling animals to be able to distinguish between voluntarily and involuntarily culling. In the current study, the largest number of cows were culled due to fertility. This agreed with previous studies investigating the specific reasons of culling cows in beef cattle populations (e.g. Oliviera *et al.*, 2020).

Table 2. The genetic correlations between ages 3 to 5 and age 6.

Culling group	Age at culling		
	3	4	5
Fertility	0.8050	0.9062	0.9734
Natural death	0.7302	0.8909	0.9778
Performance	0.9019	0.9422	0.9788
Structural problems	0.5665	0.7953	0.9409
All reasons	0.8728	0.9442	0.9863

The size of estimated additive genetic variance was largest from the performance dataset and lowest from fertility. This resulted in highest and lowest heritability estimates from performance and fertility disposal groups, respectively. This may indicate that longevity from performance and fertility disposal groups can be deemed as a proxy for the actual performance and fertility traits, respectively. The estimated heritabilities from the current study are lower than those reported by Oliviera *et al.* (2020) but agree with previous studies on longevity traits in cattle (e.g. van Pelt *et al.*, 2015; Heise *et al.*, 2018).

Similar to Oliviera *et al.* (2020), we also found considerable variation in heritability from different groups of disposal reasons. This suggest that the expected genetic progress is likely to be higher from selection based on disposal reasons with higher heritability and combining all groups of disposal reasons will likely weaken the genetic progress for longevity because of a lower overall heritability. Nevertheless, the magnitude of heritability from all groups of disposal reasons imply that direct selection for longevity could improve the longevity in Australian Angus cattle. Higher number of observations are needed for all groups of culling reasons to accurately estimate variance components especially at extreme ages.

Random regression models enable the inclusion of repeated measurements of longevity at different ages to be used for estimation of genetic variance and correlations. This leads to a better use of all available information as records at adjacent ages are highly genetically correlated, resulting in an increase in stability and accuracy of EBVs compared to when only records from a single age is used. Therefore, if the aim is to select for longevity at the traditional age of 6, the inclusion of records before and after this age can help in predication of longevity at age 6. Given the high genetic correlation between age 6 and earlier ages of 3-5, selection can be done at an earlier age to maximize the genetic gain for longevity by choosing an age with the highest heritability and by shortening the generation interval.

This study showed that female longevity has a low heritability in Australian Angus cattle population and that the estimates of heritability obtained from different culling reason groups differ from each other. The peak of heritability for different groups aligned with ages where there was a balance between disposed and active animals resulting in largest variation across all ages. High genetic correlation between ages at close vicinity provides the opportunity to select at an earlier age even if target age is a later age. This lined up with ages with heritability peaks can maximize the genetic gain from selection for longevity in Australian Angus cattle.

Acknowledgements

We thank Angus Australia for providing data and funding for this study. We acknowledge Andrew Byrne, Peter Parnell and Julius van der Werf for their valuable feedbacks.

References

Ducrocq V., Quaas R.L., Pollak E.J. and Casella G. (1988) J Dairy Sci. 1(11):3061-3070. https://doi.org/10.3168/jds.S0022-0302(88)79906-3

Heise J., Stock K.F., Reinhardt F., Ha N.T. and Simianer H. (2018) J Dairy Sci. 101(1):425-432. https://doi.org/10.3168/jds.2017-12957

Misztal I., Tsuruta S., Lourenco D.A.L. Masuda Y., *et al.* (2018) Manual for BLUPF90 family programs. University of Georgia. Available at: http://nce.ads.uga.edu/wiki/doku.php

Olivieria H.R., Brito L.F., Miller S.P. and Schenkel F.S. (2020) Animals 10(12):2410. https://doi.org/10.3390/ani10122410

Van Pelt M.L., Meuwissen T.H.E., De Jong G. and Veerkamp R.F. (2015) J Dairy Sci. 98(6):4117-4130. https://doi.org/10.3168/jds.2014-9090

Veerkamp R.F., Brotherstone S. and Engel B. (2001) Anim. Sci. 72(1):1-10. https://doi.org/10.1017/S1357729800055491

643. Eye muscle area and lean muscle area are genetically different traits in high marbling Wagyu carcasses

R.A. McEwin[1*], M.L. Hebart[1], R. Tearle[1], J. Grose[2] and W.S. Pitchford[1]

[1]Davies Livestock Research Centre, School of Animal and Veterinary Sciences, University of Adelaide, SA 5371, Australia; [2]3D Genetics Pty LTD, 939 Pukawidgi Rd, Bukkulla, NSW 2360, Australia; *rudi.mcewin@adelaide.edu.au*

Abstract

The relationship between marbling (fat) and muscle (protein plus some fat) is generally antagonistic in beef cattle. Both traits are key components of breeding objectives with marbling determining meat quality aspects while muscle, particularly loin eye muscle area (EMA), has implications for yield especially of high value cuts. However, this antagonistic relationship does not appear to hold true in highly marbled carcasses, with high levels of marbling (intramuscular fat, IMF) pushing muscle fibre bundles apart and increasing EMA. The genetic correlation between IMF and EMA, measured using Meat Image Japan camera technology, was 0.23. Lean muscle area, calculated as EMA/(1-IMF), had a genetic correlation of -0.48 with IMF, which is more aligned to expectation from lower marbling breeds. In conclusion, caution is advised when including EMA breeding values in the breeding objectives of highly marbled animals as it is a function of IMF in addition to muscling.

Introduction

Wagyu, particularly Japanese Black Wagyu, are a breed of cattle originating from Japan that are famed for their high propensity to produce highly marbled beef. Marbling, is defined as the accumulation of adipocytes within muscle fibre bundles resulting in visible white flecks of fat when a cross section of the muscle is taken (Harper and Pethick 2004). This trait has become a strong focus in many beef breeding programs due to its association with meat eating quality through tenderness, juiciness and flavour (Thompson 2004). This is particularly important for high value cuts, such as those obtained from the *longissimus dorsi* muscle, more commonly known as the loin or eye muscle. The size of the eye muscle measured as cross-sectional area (EMA) and amount of intramuscular fat (IMF) are large profit drivers in Wagyu cattle.

Generally, accumulation of muscle and accumulation of fat are antagonistic traits, that is animals with more muscle generally have less fat due to energy partitioning. While this is particularly true when comparing muscle versus subcutaneous fat, marbling is a late maturing trait with genetic correlations against EMA usually negative and/or close to zero (Gregory *et al.* 1995, Torres-Vázquez *et al.* 2018). Distinctly different to this is the genetic correlation between EMA and marbling score reported in Wagyu which is moderate and positive (0.43, Oyama 2011).

The aim of this paper is to demonstrate that EMA is describing both marbling and muscle in highly marbled cattle and discuss implications for breeding programs, providing alternate measures.

Materials & methods

Genotype data. Genotype information on a subset of 1,091 Full-Blood Wagyu individuals with carcass information was utilised for this study. These animals were genotyped with the 30K GGP-LD (NeoGen Australasia: GeneSeek Genomic Profiler Low-Density) SNP chip. Animals were not excluded on sample call rate, however, SNPs were excluded if they possessed ≥5% missing data (95% call rate) and/or a minor allele frequency (MAF) less than 0.05. 20,955 of 29,547 SNPs were retained for the analysis. Homozygous

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_643, © R.A. McEwin *et al.* 2022

genotypes were coded as -1 and 1, with heterozygous genotypes as 0. After genotype cleaning, a genomic relationship matrix (GRM) was constructed as per VanRaden's first method (VanRaden 2008);

$$\boldsymbol{G} = \frac{\boldsymbol{zz'}}{2\sum_{i=1}^{n} p_i(1-p_i)} \quad (1)$$

where Z denotes a centred matrix of allele effects with a mean of zero, p_i is the frequency of the second (minor) allele at locus i and division by $\mathbf{2}\sum p_i(1-p_i)$ scales the G matrix to be similar in magnitude (diagonal elements average 1) to the numerator relationship matrix constructed from pedigree (VanRaden 2008).

Phenotype data. A subset of phenotype data on 1,079 individuals was provided. The traits under investigation were hot standard carcass weight (HSCW, kg) and Meat Image Japan (MIJ) camera image traits; image eye muscle area (EMA, cm^2) and image intramuscular fat (IMF, %), measured at the 5^{th}-6^{th} ribbing site (Kuchida *et al.* 2006). A fourth trait, image lean muscle area (LMA, cm^2) was estimated from the camera data as:

$$\text{LMA} = ((100 - \text{IMF}) / 100) \times \text{EMA} \quad (2)$$

Model development and statistical analysis. Datum was first analysed as a series of univariate models with a general linear mixed model using ASReml-R 4.0 (Butler *et al.* 2017) such that;

$$\boldsymbol{y} = \boldsymbol{Xb} + \boldsymbol{Zu} + \boldsymbol{e} \quad (3)$$

where, b is the vector of fixed effects, u is the vector of random effects (with X and Z their respective design matrices) and e is the vector of residuals. All traits included fixed effects of dam age (Maiden; <2 year of age, Mature; 3-9 years or Old; >10 years), heterozygosity (calculated as the proportion of heterozygous genotypes, 0.36±0.02) and birth group based on a predefined age slice that grouped animals born within the same year and calving period. A birth date co-variate was also fitted nested within management group. Management groups were defined as the combination of weaning group (200 days), slaughter group and slaughter date of which there were 23 levels. EMA, LMA and IMF were not adjusted for HSCW in the models.

Having identified the appropriate models from the univariate analysis, a bi-variate analysis between all traits was conducted incorporating the univariate models. Let $\boldsymbol{y} = (\boldsymbol{y_1'}, \boldsymbol{y_2'})'$, be the combined vector of data between two traits. The mixed model for the bivariate analysis is given by:

$$\boldsymbol{y} = \boldsymbol{X^*b} + \boldsymbol{Z^*u} + \boldsymbol{e} \quad (4)$$

Table 1. Summary statistics and number of records in the subset provided for the four traits under investigation.

Trait	Units	Mean	Min	Max	SD[1]
HSCW	kg	437.8	302.0	618.5	46.8
EMA	cm^2	45.4	25.0	82.2	7.8
LMA	cm^2	32.1	17.4	65.2	6.1
IMF	%	29.1	13.5	60.3	6.6

[1] SD refers to standard deviation.

Where $\boldsymbol{b} = (\boldsymbol{b_1'}, \boldsymbol{b_2'})'$ is the $2m \times 1$ vector of fixed effects with $\boldsymbol{X^*} = \boldsymbol{I_2} \otimes \boldsymbol{X}$ the associated design matrix; $\boldsymbol{u} = (\boldsymbol{u_1'}, \boldsymbol{u_2'})'$ is the $2n \times 1$ vector of random effects with $\boldsymbol{Z^*} = \boldsymbol{I_2} \otimes \boldsymbol{Z}$ the associated design matrix and $\boldsymbol{e} = (\boldsymbol{e_1'}, \boldsymbol{e_2'})'$ the vector of residual variance ordered as for the data vector. The variance assumptions for the random effects are:

$$var(\boldsymbol{u}) = var\begin{pmatrix}\boldsymbol{u_1}\\ \boldsymbol{u_2}\end{pmatrix} = \begin{bmatrix}\sigma_1^2 & \\ \sigma_{12} & \sigma_2^2\end{bmatrix} \otimes \boldsymbol{G} \quad (5)$$

where σ_1^2 is the genetic variance at trait 1, σ_2^2 is the genetic variance at trait 2 and σ_{12} is the genetic covariance between the two traits. Univariate and bi-variate models were fitted for all traits utilising a genomic relationship matrix and heritabilities, genetic and phenotypic correlations were calculated.

Results

All traits were highly heritable (Table 2). The heritability of eye muscle area was similar to lean muscle area (0.38 vs 0.43).

The genomic phenotypic and genetic correlations were determined between all four traits (Table 3). HSCW was moderately and positively correlated with traits describing muscling, EMA and LMA respectively. However, correlations between HSCW and IMF were very low and close to zero. IMF presented differing correlations against the two muscling traits, namely the genetic correlation with EMA is lowly to moderately positive (0.23) while the genetic correlation to LMA is strongly negative (-0.48). The phenotypic correlations among these traits were similar to the genetic correlations.

Discussion

The key finding is the genetic correlation between EMA and IMF which is moderately positive (0.23, Table 3). This agrees with Oyama (2011), whom presented an unweighted mean genetic correlation of 0.43, averaged across 6 studies, and McEwin (2016) who reported a genetic correlation of 0.24 in a different data set from that herein. McEwin (2016) demonstrated that this positive correlation was explained by excessive intramuscular fat depots pushing muscle fibre bundles apart, thereby increasing EMA. This hypothesis was tested by creating a new trait which estimates LMA using a similar method to that used herein. The

Table 2. Variance components and heritabilities (with standard errors; se) of traits from univariate analysis.

Trait	σ_A^2	σ_P^2	h^2	se
HSCW	32.9	1,928	0.56	0.05
EMA	20.7	54.9	0.38	0.06
IMF	5.21	40.1	0.68	0.05
LMA	13.2	30.8	0.43	0.06

Table 3. Phenotypic (r_p above diagonal) and genetic (r_g below diagonal) correlations between traits[1]

	HSCW	EMA	LMA	IMF
HSCW		0.32	0.26	0.10
EMA	0.33		0.85	0.19
LMA	0.30	0.75		-0.39
IMF	0.02	0.23	-0.48	

[1] Standard errors range 0.01-0.04 for r_p and 0.05-0.10 for r_g.

MIJ data herein is expected to provide a more accurate description of IMF %. McEwin (2016) reported the genetic correlation between marble score and LMA as -0.37, in agreement with -0.48 reported herein for camera data. This relationship is substantially more severe than the genetic correlations reported close to zero in other breeds (Gregory *et al.* 1995, Torres-Vázquez *et al.* 2018), highlighting a potential biological preference to partition energy towards marbling over muscle in Wagyu.

The genetic and phenotypic correlations between LMA and EMA are high (0.75, Table 3), indicating IMF was removed with some accuracy and lean muscle area was truly estimated. The current breeding objective is a function of muscle weight and quality determined by IMF. It is possible that modelling the components separately may more accurately describe genetic variation in this trait of such high value. Selection index theory will be utilized to investigate this further.

References

Butler D., Cullis B., Gilmour A.R., Gogel B., Thompson R. (2017) ASReml-R Reference Manual Version 4. *VSN International Ltd, Hemel Hempstead, HP1 1ES, UK*. Available at: https://asreml.kb.vsni.co.uk/wp-content/uploads/sites/3/ASReml-R-Reference-Manual-4.pdf

Gregory K., Cundiff L., Koch R. (1995) J. Ani Sci. 73:1920-1926. https://doi.org/10.2527/1995.7371920x

Harper G., Pethick D. (2004) Ani Prod Sci. 44:653-662. https://doi.org/10.1071/EA02165

Kuchida K., Osawa T., Hori T., Kotaka H., Maruyama S. (2006) J Ani Gen. 34:45-52. https://doi.org/10.5924/abgri2000.34.2_45

McEwin R. (2016) Mythbusters: Non-additive Genetic Effects had Negligible Impact on Wagyu Carcass Quality. Honors Thesis. University of Adelaide.

Oyama K. (2011) Ani Sci J. 82:367-373. https://doi.org/10.1111/j.1740-0929.2011.00895.x

Thompson J. (2004) Ani Prod Sci. 44:645-652. https://doi.org/10.1071/EA02171

Torres-Vázquez J.A., van der Werf J.H., Clark S.A. (2018) J Ani Sci. 96:4521-4531. https://doi.org/10.1093/jas/sky325

VanRaden P.M. (2008) J Dairy Sci. 91:4414-4423. https://doi.org/10.3168/jds.2007-0980

644. Selection on yearling days to calving improves mature fertility traits in tropically adapted cattle

M.L. Facy[1*], M.L. Hebart[1], H. Oakey[1,2], R.A. McEwin[1],G.I. Popplewell[3] and W.S. Pitchford[1]

[1]Davies Livestock Research Centre, School of Animal and Veterinary Sciences, University of Adelaide, Roseworthy, SA 5371, Australia; [2]Robinson Research Institute, Adelaide Medical School, University of Adelaide, Adelaide, SA 5006, Australia; [3]Popplewell Composite, 5 Cedar St, Maleny QLD 4552, Australia; maddie.facy@adelaide.edu.au

Abstract

Female fertility is essential in any beef breeding program but is often treated with low priority in genetic evaluations due to low heritability, long generation intervals and difficulty of measurement. Days to calving is a female fertility trait that is both easy to measure and can be measured early in life i.e. first joining as a yearling. Utilising early in life fertility measure allows for selection at an earlier age thus it will lead to greater genetic change and improvement in the breeding objective of lifetime weaning rate. This paper reported that treating early in life measures as a different trait increased heritability estimates and variance components of days to calving, increasing accuracy. The genetic correlation between early in life measures and mature measures was high (0.81), thus selection on early in life days to calving will improve lifetime traits.

Introduction

Female fertility is the most important profit driver in beef production systems but is often ignored in genetic evaluations. This is due to female fertility traits having a low heritability, coupled with difficulty to measure, few correlated traits and longer generation intervals (Johnston *et al.* 2013). Fertility is particularly important in tropical environments however little to no genetic progress has been made. A potential solution is using early in life female fertility traits, early in life being measured in the first few years of life. This potential solution allows for genetic improvement on female fertility.

Days to calving is the number of days from the start of joining (day the bull goes in) to subsequent calving (Meyer *et al.* 1990). This trait is a function of both anoestrus and gestation length. Days to calving is a trait that can be measured multiple times over a cows lifetime, typically days to calving is evaluated as a repeated measure trait (Graser *et al.* 2005). Days to calving is a female fertility trait that uses yearling joining and subsequent joinings that measures both attainment of puberty and pregnancy rate which can produce a better selection criterion as it is cheaper and easier to measure than using multiple ovarian scans to determine age at puberty.

The present paper evaluates the effects of treating the first joining independent to mature joinings through genetic correlation, heritability estimates and variance components to allow for more genetic gain. A unique feature is that yearling joinings are rare in tropically adapted cattle so the first joining is a year earlier than typical Northern beef systems (first joined at two years).

Materials & methods

Data. Phenotype and genotype data used was provided by Popplewell Composites; a bull breeding program with a focus on breeding composite bulls adapted for Northern Australia. The breeding objective of Popplewell Composites is utilisation of retained heterosis, increasing fertility, reducing age of puberty and good growth that are well adapted for tropical environments such as Northern Australia.

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_644, © M.L. Facy *et al.* 2022

Approximately 2,600 composite cattle were genotyped across the breeding program, this was conducted over multiple years which resulted in different SNP chips being used. The SNP chips used were Illumina, GGPLD V3 30K, Illumina GGPLD V4 30K and Weatherbys Scientific Versa50K. There were 10,830 common SNPs across all SNP chips. The common SNPs were imputed to GGPLD30K (27,755 SNPs) using FImpute 2.2 (Sargolzaei *et al.* 2011) for all genotypes. Genotype data was then filtered to exclude SNPs with <1% minor allele frequency and duplicate animals. The genomic relationship matrix (GRM) was constructed from a total of 2,609 animals and 27,638 SNPs.

Phenotype data was used to calculate days to calving which is defined as the time from the day the bull goes in with the cow to the day the cow gives birth. The data included bull in date, bull out date, pregnancy test results (fetal age) and calf date of birth. Cattle in this dataset were first joined at 12-15 months old. There was a total of 649 joining one days to calving records (J1_DC) and a total of 1,593 mature (second joining onwards) days to calving records (JM_DC) with 903 unique animals. For cows that did not conceive a calf were given a corrected value. The corrected value was defined as the maximum days to calving value within each joining group plus 32 days (1.5 × 21-day oestrus cycles). The assumption is that the cows would have conceived in the next 1.5 cycles, on average, if given the opportunity. This penalty was deemed by the breeder as more appropriate than the commonly used 21 days due to some influence of indicus in the genetic makeup of the composite.

Genomic model. A linear mixed model to estimate breeding values can be written as:

$$y = X\beta + Zu + e \quad (1)$$

where y is the vector of phenotypic values for days to calving; β represents fixed effects with design matrix X; the fixed effects included in all models are contemporary group, joining number, lactation and heterozygosity fraction; Z is the design matrix for animal additive effects u; and e is the vector of residuals.

The genomic relationship matrix was constructed following VanRaden's method 1 (VanRaden 2008).

Statistical model. A linear mixed model was fit for joining one days to calving (J1_DC) with variance components and breeding values estimated. The fixed effects for J1_DC were contemporary group (which included birth group and all lifetime management groups), dam age, heterozygosity fraction covariate. Birth group was defined by year of birth and birth time: A for early in the season (1st of June to 4th of September), B for mid-season (5th of September to 30th of November) and C for late in the season (anything from 1st of December onwards. Two repeatability linear mixed models were run for days to calving: one with all joinings included (DTC) and another with all mature joinings (JM_DC; where J1_DC was excluded). The fixed effects for JM_DC was joining number, lactation status, heterozygosity fraction covariate, days postpartum (DPP) covariate nested within contemporary group. The fixed effects for DTC was joining number, DTC management group, lactation status, heterozygosity fraction covariate and DPP covariate nested within contemporary group. DPP was calculated as the number of days from last calf date of birth to next bull in date. Animal tag was fit alongside the genomic relationship matrix as a random term to account for multiple records on the same animal, thus estimating the permanent environmental variance. All models were run using ASReml-R 4.0 (Butler *et al*., 2017). A genetic correlation was estimated between J1_DC and JM_DC.

Results

Table 1 lists variance components and heritability estimates for all models. The heritability estimates for J1_DC was 0.20 with a heterozygosity coefficient of -1.90 days/%. The heritability estimates for J1_DC (0.20) was higher than DTC and JM_DC which were 0.13 and 0.14 respectively. J1_DC had the smallest heterozygosity coefficient with JM_DC having the largest heterozygosity coefficient of -2.55d/%. The residual variance was lower for JM_DC than DTC and J1_DC. The repeatability of DTC was 0.09 which was higher than JM_DC (0.03). The genetic correlation between J1_DC and JM_DC was high (0.81±0.21).

Discussion

The current study reported DTC to have a heritability of 0.13. Meyer *et al.* (1990) reported days to calving heritability in a zebu cross to be 0.086 which is lower then what was reported in the current study. However, Johnston *et al.* (1996) reported a heritability of 0.12 in an Angus herd, similar to the estimate in this study. Olasege *et al.* (2021) estimated a higher heritability for days to calving for the first five joinings, with a heritability of 0.31 for tropical composites. JM_DC had a similar heritability estimate to DTC of 0.14. There is variation in heritability estimates across different populations for days to calving.

J1_DC had a heritability of 0.20 which is high for a fertility trait. Johnston *et al.* 2013 have previously reported a heritability of joining one days to calving to be 0.22 in Brahmans and 0.13 in tropical composites. Johnston *et al.* (2013) cattle were first mated at two-year-old whereas herein they were mated as yearlings. The Brahman estimates are similar to the current study estimate of J1_DC however the tropical composites are more similar to DTC. This is due to the difference of age at first mating between Brahman and tropical composite since tropical composite mature at an early age and joining one days to calving should be measured around age of puberty.

The heritability of days to calving was lower when all joinings were included compared to the other two models (JM_DC and J1_DC). This demonstrates that when joining one is removed and treated as a different trait, more genetic variance can be accounted for thus more genetic gain can be made. This increase allows for greater genetic gain and accuracy.

Days to calving for yearling (J1_DC) and mature (JM_DC) joinings were highly correlated which suggests they could be treated as the same trait. Johnston *et al.* (2013) also reported a high genetic correlation between J1_DC and lifetime fertility traits (0.75). However, given that there was more than double the genetic variation in J1_DC than combined traits and that yearling joining success is a trait of major focus for the breeding program, it is worth reporting J1_DC and JM_DC GEBVs separately. Using early in life joining allows selection on cows after first joining, increasing selection intensity and genetic gain. Johnston *et al.* (2013) also concluded that selecting cattle based on early lifetime days to calving will improve lifetime fertility.

Table 1. Summary of variance components (standard error in parentheses), heritability, repeatability and heterozygosity coefficient (Het) estimate for days to calving models.[1]

Models	Additive variance[a]	Permanent env variance[a]	Residual variance[a]	Heritability	Repeatability	Het coefficient[b]
DTC	124 (43)	82 (43)	811 (38)	0.13 (0.04)	0.09 (0.05)	-2.45
JM_DC	115 (39)	26 (42)	741 (42)	0.14 (0.04)	0.03 (0.05)	-2.55
J1_DC	257 (122)	-	1,049 (110)	0.20 (0.09)	-	-1.90

[1] DTC = days to calving; JM_DC = mature days to calving, J1_DC = joining one days to calving; [a] days^2; [b]days/%.

References

Butler D., Cullis B., Gilmour A., Gogel B. and Thompson R. (2017) ASReml-R Reference Manual Version 4. VSN International Ltd, Hemel Hempstead, UK.

Graser H.-U., Tier B., Johnston D.J. and Barwick S.A. (2005) Aus J of Exp Agri 45(8): 913-921. https://doi.org/10.1071/EA05075

Johnston D.J., Barwick S.A., Fordyce G., Holroyd R.G., Williams P.J. *et al.* (2013) Ani Pro Sci 54(1): 1-15.

Johnston D.J., and Bunter K.L. (1996) Livestock Pro Sci 45(1): 13-22. https://doi.org/10.1016/0301-6226(95)00088-7

Meyer K., Hammond K., Parnell P.F., Mackinnon M.J. and Sivarajasingam S. (1990) Livestock Pro Sci 25(1-2): 15-30. https://doi.org/10.1016/0301-6226(90)90038-8

Olasege B.S., Tahir M.S., Gouveia G.C., Kour J., Porto-Neto L.R. *et al.* (2021) Ani Pro Sci. https://doi.org/10.1071/an21097

Sargolzaei M., Chesnais J.P. and Schenkel F. (2011) J of Dairy Sci 94 (E-Suppl. 1):421.

Vanraden P.M. (2008) J of Dairy Sci 91(11): 4414-4423. https://doi.org/10.3168/jds.2007-0980

645. Estimation of maternal variance for slaughter traits

A.J. Twomey[1*] and T. Pabiou[2]

[1]Teagasc, Moorepark, Animal and Grassland Research and Innovation Centre, Moorepark, Fermoy, P61 P302, Co. Cork, Ireland; [2]Irish Cattle Breeding Federation, Highfield House, Shinagh, Bandon, P72 X050, Co. Cork, Ireland; alan.twomey@teagasc.ie

Abstract

The objective of the present study was to quantify the maternal variance contributing to carcass traits. The dataset compromised of a random sample of 50,000 beef animals. Direct and maternal variances were estimated for each of the carcass traits using a linear mixed animal model. The direct genetic standard deviation for carcass weight, carcass conformation, carcass fat score and age at slaughter was 21 kg, 0.77 units, 0.99 units and 42 d, respectively; direct heritability estimates ranged from 0.42 to 0.52. Maternal variance was significantly different from zero for carcass weight and age at slaughter; maternal heritability estimates were 0.03 for age at slaughter and 0.04 for carcass weight. There was a strong negative correlation between direct and maternal effects for age at slaughter (-0.89) and carcass weight (-0.61). This study clearly shows the potential to improve maternal traits in beef cows to improve carcass performance of their progenies.

Introduction

Animals for beef production originate from suckler-based system, where cows rear calves for beef (Hocquette and Chatellier, 2011) and surplus animals from dairy systems. In suckler-based systems, weight gain of progeny in early life is influenced by the milk availability from the dam. Twomey *et al.* 2020 documented that early life differences in milk availability can influence carcass performance, where beef dams originating from dairy herds produced heavier and younger cattle at slaughter compared to beef dams originating from dairy herds, assuming the same progeny had the same genetic merit for growth traits. The present study will be the first, to our knowledge, to quantify the maternal variance contributing to carcass traits in beef cattle. The objective of this study was to estimate the direct and maternal genetic variance, for carcass weight, carcass conformation, carcass fat score and age at slaughter.

Materials & methods

Carcass records for 5,362,210 animals were obtained from the Irish Cattle Breeding Federation database 2016 to 2020, inclusive; ancestry and breed composition was also available. Carcass weight (kg), conformation score (5-score as per European Council regulations 1208/81 and 2930/81), fat score (score as per European Council regulations 1208/81 and 2930/81) and age of animal at slaughter were available. Carcass weight was measured, on average, 1 h after slaughter following the removal of the head, legs, thoracic and abdominal organs, and internal fats and hide. As described by Pabiou *et al.* (2011), carcass conformation and fat scores were graded using video image analysis and were graded under the European Union beef carcass classification system (EUROP). The resulting EUROP classification grades were transformed into a 1 to 15 scale as outlined in Englishby *et al.* (2016). Age at slaughter in days was also available for all animals.

Data edits. Dam parity was categorized as 1, 2, 3, 4, 5+; parities >10 were not considered further. Only steers and nulliparous females (i.e. heifers) slaughtered between 14 and 36 months of age and young bulls slaughtered between 12 and 24 months of age were retained. Carcass records were discarded from animals that resided in more than 3 herds during their lifetime, and/or were present in the herd they were slaughtered from for <100 days prior to slaughter. Animals with an unknown sire or dam were not considered further. Only animals originating from a beef herd and beef dam were retained. Beef herds were defined as herds

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_645, © A.J. Twomey and T. Pabiou 2022

where the average dairy (Holstein-Friesian, Jersey, Norwegian Red, Ayshire, Mountbellard, and Brown Swiss) breed percentage of the cows, calving between the years 2015 and 2020, was ≤55%.

Animals were allocated to a contemporary group for subsequent use in the statistical model. All contemporary groups were formed using algorithm described by Berry and Evans (2014). The algorithm is based on grouping animals together, within herd that have dates in close proximity. Contemporary group of slaughter was defined as animals grouped based on date of slaughter, herd prior to slaughter and gender (i.e. heifer, steer, young bull). Contemporary group of entry in to the herd was defined as animals grouped based on gender, age when purchased into the same herd (from which they would be eventually slaughtered from) and the date they entered that herd. Prior to the estimation of variance components a random sample of contemporary groups was chosen, to result in a dataset with approximately 50,000 records per trait. Only contemporary groups with ≥10 animals and >1 sire represented were retained.

Statistical analysis. Variance components were estimated for carcass weight, conformation, fat score and age at slaughter using univariate animal linear mixed models in ASReml (Gilmour *et al.*, 2009). The fitted model was

$$y = X\beta + Za + Wm + Tme + e$$

where y is the vector of the trait of interest (carcass weight, conformation, fat score and age at slaughter); β is the vector of fixed effects; a is the vector of direct additive genetic effects of the animal; m is the vector of direct additive genetic effects of the animal's dam; me is a vector of permanent environmental effects of the animal's dam; e is the vector of residual effects; X is the incidence matrix for fixed effects; Z is the incidence matrix relating to the random direct additive genetic effects; W is the incidence matrix relating to the random maternal additive genetic effects; and T is the incidence matrix relating to the random maternal permanent environmental effect. The covariance structure for the random effects included in the model were assumed to be as follows:

$$\text{var}\begin{bmatrix} a \\ m \\ me \\ e \end{bmatrix} = \begin{bmatrix} A\sigma_a^2 & A\sigma_{am} & 0 & 0 \\ A\sigma_{am} & A\sigma_m^2 & 0 & 0 \\ 0 & 0 & I\sigma_{me}^2 & 0 \\ 0 & 0 & 0 & I\sigma_e^2 \end{bmatrix}$$

where A is the numerator relationship matrix; I are identity matrices with their order equal to the number of observations; σ_a^2 is the additive genetic variance; σ_m^2 is the maternal genetic variance, σ_{am} is the direct-maternal genetic covariance, σ_{me}^2 is the maternal environmental variance and σ_e^2 is the residual variance. The fixed effects included in the model for all dependent variables were a general heterosis coefficient (0.0%, >0.0 to <0.1%, ≥0.1 to <0.2%, …, ≥0.9 to <100.0%, 100.0%), a general recombination loss coefficient (0.00%, >0.00 to <0.05%, ≥0.05 to <0.10%, …, ≥0.45 to <0.50%, 0.50%, >0.50%) and dam parity. When the dependent variable was carcass weight, conformation or fat, age in months and contemporary group of slaughter were also included as fixed effects. Only when the dependent variable was age at slaughter, the fixed effects for carcass fat score and contemporary group of entry in to the herd were included in the model. The pedigree of each animal was traced back to the founder population, which was allocated to 11 genetic groups on the basis of breed.

Results

The direct genetic standard deviation for age at slaughter, carcass weight, conformation and fat score was 42 d, 21 kg, 0.77 units, and 0.99 units, respectively (Table 1). The direct heritability estimates ranged from 0.42 and 0.53. The maternal heritability estimate for age at slaughter and carcass weight was significant;

Table 1. Mean performance, genetic standard deviation for direct (σ_D), genetic standard deviation for maternal effects (σ_M), standard deviation for maternal permanent environmental effects (σ_{PE}), phenotypic standard deviation (σ_P), direct heritability estimates (h^2_D; SE in parentheses), maternal heritability estimates (h^2_m; SE in parentheses), and genetic correlation between direct and maternal effects (corr; SE in parentheses) for each carcass trait.

Trait	mean	σ_D	σ_M	σ_{PE}	σ_P	h^2_D	h^2_m	corr
Age at slaughter (days)	733	42	10	18	58	0.53 (0.024)	0.03 (0.012)	-0.89 (0.110)
Carcass weight (kg)	388	21	7	0	32	0.43 (0.022)	0.04 (0.014)	-0.61 (0.085)
Conformation (1-15)	9.36	0.77	0.06	0.00	1.18	0.43 (0.024)	0.00 (0.009)	-0.34 (0.517)
Fat score (1-5)	8.08	0.99	0.23	0.24	1.53	0.42 (0.024)	0.02 (0.013)	-0.24 (0.153)

maternal heritability for age at slaughter was 3% and for carcass weight was 4%. Maternal variance was not significant for both conformation and fat score. The genetic correlation between the direct age at slaughter and maternal age at slaughter was -0.89 (SE=0.110). For carcass weight the genetic correlation between the direct and the maternal effect was -0.61 (SE=0.085).

Discussion

In Irish suckler beef systems, calves are predominantly weaned at 5 to 9 months of age (Lorenz *et al.*, 2011). It is well documented that there is a significant maternal variance for weaning weight in beef cattle, with heritability estimates ranging from 0.20 to 0.67 (Cantet *et al.*, 1988; Weik *et al.*, 2021). In addition, Twomey *et al.*, 2020 did document a significant difference between carcass performance from progeny of beef X beef cows compare beef X dairy cows (known to have a higher milk yield) when direct genetic effects for growth were the same. Nevertheless, breeding evaluations for carcass traits in beef cattle currently do not consider the maternal variance. Thus, the objective of this study was to estimate the direct and maternal genetic variance for carcass traits.

The smaller contribution of maternal variance to carcass weight compare to weaning weight is somewhat expected as there is, on average, 15 to 19 months between weaning and slaughter in the present study and other environmental factors would have diluted the contribution of maternal variance (e.g. compensatory growth). The low maternal heritability estimates in the present study for carcass weight in cattle was similar to other species. The maternal heritability in carcass weight for Mukota pigs was 0.04 (Chimonyo *et al.*, 2007), for Brazilian commercial broilers it ranged from 0.04 to 0.10 (Grosso *et al.*, 2010) and for Swedish sheep it ranged from 0.08 to 0.12 (Näsholm 2004). Although heritability estimates are low, similar to fertility traits (Berry and Evans, 2014), there was considerable maternal genetic variability observed for age at slaughter and carcass weight which suggests that there is large potential to genetically improve dams maternal characteristics to subsequently improve their progeny carcass performance. The lack of maternal variance observed in the present study for carcass conformation and carcass fat score suggest that milk yield of the dam has no influence on muscle and fat deposition for slaughter later in life. This is likely due to the majority of growth in muscle and fat occurring later in life and closer to the time of slaughter (Wegner *et al.* 2000).

The strong negative association between direct and maternal effects reported in the present study is similar to growth traits in a large number of species, sheep (Fitzmaurice *et al.*, 2020), broilers (Grosso *et al.*, 2010) and cattle (Weik *et al.* 2021). Genetic correlations between most terminal and cow traits tend to be antagonistic (Phocas, 2009; Berry and Evans, 2014). The negative correlation observed in the present study between direct and maternal effects for carcass weight was similar to what is reported for weaning

weight, where correlations between direct and maternal effects for weaning weight range from -0.52 to -0.79 (Weik *et al.*, 2021; Cantet *et al.*, 1988). Although there is a moderate to strong genetic correlation between maternal and terminal traits, the combination of both direct and maternal traits for carcass weight and age at slaughter in a selection index can potentially achieve genetic gain in both simultaneously.

References

Berry D.P., and Evans R.D. (2014) J. Anim. Sci. 92:1412–1422. https://doi.org/10.2527/jas.2013-6723

Cantet R.J., Kress D.D., Anderson D.C., Doornbos D.E., Burfening P.J., *et al.* (1988) J Anim Sci. 66(3):648-60. https://doi.org/10.2527/jas1988.663648x

Chimonyo M. and Dzama, K. (2007) Animal. 1(3):317-323. https://doi.org/10.1017/S1751731107661849

Englishby T.M., Banos G., Moore K.L., Coffey M.P., Evans R.D., *et al.* (2016) J. Anim. Sci. 94:1354–1364. https://doi.org/10.2527/jas.2015-0246

Gilmour A.R., Gogel B.J., Cullis B.R., and Thompson R. 2009. ASReml user guide release 3.0 VSN International Ltd, Hemel Hempstead, UK.

Grosso J.L., Balieiro J.C., Eler J.P., Ferraz J.B., Mattos E.C., *et al.* (2010) Genet. Mol. Res 9(2):908-918.

Hocquette J.F. and Chatellier V. (2011) Anim. Front. 1(2):20-28. https://doi.org/10.2527/af.2011-0014

Lorenz I., Earley, B., Gilmore, J., Hogan, I., Kennedy, E., *et al.* (2011). Ir Vet J., 64(1):1-9. https://doi.org/10.1186/2046-0481-64-14

Näsholm A., 2004. J. Anim. Breed. Genet. 121(1):66-75. https://doi.org/10.1046/j.1439-0388.2003.00440.x

Pabiou T., Fikse W.F., Amer P.R., Cromie A. R., Näsholm A., *et al.* (2012) Animal. 6:1389–1397. https://doi.org/10.1017/S1751731112000705

Phocas F. (2009) J Anim Sci. 87(6):1865-1871. https://doi.org/10.2527/jas.2008-1426

Twomey A.J., Ring, S.C., McHugh, N. and Berry, D.P. (2020) J. Anim. Sci. 98(11):1-11. https://doi.org/10.1093/jas/skaa321

Wegner J., Albrecht, E., Fiedler, I., Teuscher, F., Papstein, H.J., *et al.* (2000) J Anim Sci. 78(6):1485-1496. https://doi.org/10.2527/2000.7861485x

Weik F., Hickson R.E., Morris S.T., Garrick D.J. and Archer J.A. Animals. 2021; 11(9):2509. https://doi.org/10.3390/ani11092509

646. Biological types, direct and maternal heterosis and recombination effects on weaning weight of composite beef cattle

J.B.S. Ferraz[1*], R. Espigolan[1], F.S. Baldi[2], J.P. Eler[1], E.C.M. Oliveira[1], R. Nuñez-Dominguez[3] and L.T. Gama[4]

[1]Research Center of Animal Breeding, Biotechnology and Transgenesis, College of Animal Science of the University of São Paulo, Brazil, Rua Duque de Caxias Norte, 225, 13635-900 Pirassununga, SP, Brazil; [2]FCAV/UNESP, Rodovia Carlos Tonani, s/n, 14884-900 Jaboticabal, SP, Brazil; [3]Universidad Autónoma de Chapingo, Km. 38.5, Carretera México – Texcoco, Chapingo, Estado de México CP 56230, Mexico; [4]Faculdade de Medicina Veterinária, Universidade de Lisboa, Avenida da Universidade Técnica – Polo Universitário – Alto da Ajuda – 1300-477 Lisboa, Portugal; jbferraz@usp.br

Abstract

A tropical beef composite population was studied to investigate the effects of biological type, heterosis and recombination on weaning weight, one of the most important and complex traits for the beef industry. Biological types used to group the breed composition were N (*Bos indicus* breeds), A (*Bos taurus* adapted to tropics), B (*Bos taurus* of British origin) and C (*Bos taurus* of Continental origin). The pedigree and breed composition files had 943,334 observations, while the weaning weight file had information from 230,567 animals. Biological types differed in direct and maternal genetic effects. The A group had the poorest performance in both direct and maternal effects, while the C group excelled in direct effects and the B group was better in maternal effects. The N group was intermediate for both genetic effects. The direct and maternal heterosis were both positive and important, but recombination effects acted in the opposite direction.

Introduction

The Montana Composite beef population is a project that started, in Brazil, in 1994, with the objective of producing beef bulls that should be used in natural mating, producing calves with high level of heterosis to be raised under tropical conditions. The founders were purebred Nellore cows and their crosses, and the genetics from founding sires were semen bulls from several *Bos taurus* breeds adapted to these conditions, or from British or Continental origin. Since the early stages, this population has been under full control, with pedigree and breed composition information recorded for each animal, as well as phenotypic data. In recent years, genomic data based on SNP arrays has been obtained in a broad representation of Montana cattle, being recently introduced in the genetic evaluation of the population.

The analyses of data obtained in composite populations represents a challenge, particularly when several breed groups are included. Indeed, with this data structure, the fixed part of the mixed model analyses should include, in addition to the common fixed factors (contemporary group, sex, age of dam, etc.), also the fixed effects of biological type for direct and maternal genetic effects, as well as the direct and maternal heterosis, as appropriate for the trait under consideration. However, the important effects of recombination, as proposed by Dickerson (1969), also studied by VanRaden and Sanders (2003), and, more recently by Berry *et al.* (2021) should also be considered, but were seldom studied.

The objective of this work was to investigate how weaning weight in Montana beef cattle is affected by biological type of cattle, and by direct and maternal heterosis for that trait. In addition, we wanted to assess the magnitude of maternal and individual recombination effects in a trait of economic relevance such as weaning weight, using a large data set obtained in a composite beef population.

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_646, © J.B.S. Ferraz *et al.* 2022

Materials & methods

Description of population, pedigree and breed composition control. Since the beginning of the project, all founders and animals with records had pedigree and breed composition controlled. The pedigree and breed composition data sets for this analysis included 943,334 animals, with 230,567 weaning weights recorded, by 1,386 sires and 110,291 dams and belonging to 1,497 contemporary groups.

As many breeds were involved, the breed composition of each animal was expressed as biological type, as suggested by Gregory *et al.* (1985), Cundiff and Gregory (1999) from the Germplasm Project of USDA-MARC types. The groups were: N, grouping the *Bos indicus* breeds; A, for *Bos taurus* adapted for tropical conditions; B, for *Bos taurus* of British origin; and C for *Bos taurus* of continental origin. The description of the population and genetic analysis were presented in Ferraz, Eler and Golden (1999a, b). The major breed of biological type N was Nellore; in type A many breeds were included, with Bonsmara and Senepol being the most important; in group B, Red Angus and, more recently, Aberdeen Angus; in group C, Simmental was the major breed. The program was controlled by the Brazilian Ministry of Agriculture, under a program named CEIP, Special Certificate of Identification and Production, where animals can be certified and sold as breeders only if they belong to the best 30%, under a selection criterion specified by the program. To be considered a Montana Composite, the animal must have the biological type as described in Table 1.

Genetic analysis. The genetic analyses were performed with AIREML from the BLUPF90 family of programs (Misztal *et al.* 2002, 2014). The variance of contrasts among biological types was obtained from MTDFREML, which estimates standard errors of prediction of estimable functions, reported in the MTDF67 output file (Boldman *et al.*, 1995). Those analysis included in the model the fixed effect of contemporary groups (including farm, year, season, sex) and, as covariates, age of dam (linear and quadratic), age at weaning, the effects of N, A, B and C direct and maternal genetic effects, as well as the direct and maternal heterosis and recombination effects for weaning weight. In addition, the model included the random direct and maternal animal genetic effects and the permanent environment of the dams.

Significance of effects and their contrasts. The statistical significance of contrasts among biological types, and of the estimates of direct and maternal heterosis and recombination effects was tested by the ratio of the estimate relative to its standard error, assuming a t-distribution with very large degrees of freedom.

Results

Descriptive statistics. For the 230,567 weaning weight records analysed, the mean was 199.42±38.54 kg, ranging from 84.0 to 315.0 kg, with a coefficient of variation of 19.32%, with an average age at weaning of 217.18±26.93 days. The average age of dam was 67.01±35.46 months.

Biological types and their constraints, direct and recombination effects. The estimates of the direct and maternal effects for biological types N, A, B and C, and their contrasts, as well as direct and maternal

Table 1. Breed composition, grouped in biological types of Montana seedstock.

Biological type	Minimum	Maximum
Number of biological types	3	No limit
N	0	6/16
A	2/16	14/16
N + A	4/16	16/16
B	0	12/16
C	0	12/16
N + C	0	12/16

heterosis and recombination effects, expressed in kg and in percentage of the global mean weaning weight are reported in Table 2.

Discussion

The various biological types differed in their direct and maternal genetic effects for weaning weight, which is a trait of high economic relevance for the beef industry. However, more important than the biological effects, their contrasts presented very interesting results, with a significant advantage in direct genetic effects of the *Bos indicus* cattle (type N), over the *Bos taurus* adapted to tropical condition (A) and a small, but not statistically significant advantage over *Bos taurus* of British origin (B) and even a disadvantage when compared to animals of Continental origin (C). The contribution to weaning weight performance was negative for the taurine breeds adapted to the tropics, and Continental breeds had a better contribution in direct genetic effects than any other biological type.

As related to maternal genetic effects, the N group was better than A, similar to C but not as good as B. Group A had important negative effects when compared with the British and Continental groups and those European breeds did not differ significantly from each other.

Table 2. Estimates of biological type effects and their differences and direct and maternal heterosis and recombination effects on weaning weight of the Tropical Montana beef cattle population.[1]

Effect	Estimate, kg	Percentage of mean weaning weight
Biological types[2]		
N	-25.29	-12.68
A	-29.94	-29.94
B	-25.72	-12.89
C	-17.28	-8.67
Maternal N	65.53	32.86
Maternal A	56.85	28.50
Maternal B	68.35	34.28
Maternal C	66.42	33.31
N-A	4.65***	2.33
N-B	0.43 NS	0.22
N-C	-8.00***	-4.01
A-B	-4.22***	-2.12
A-C	-12.66***	-6,35
B-C	-8.44***	-4.23
Maternal N – Maternal A	8.68***	4.35
Maternal N – Maternal B	-2.82**	-1.42
Maternal N – Maternal C	-0.90 NS	-0.45
Maternal A – Maternal B	-11.50***	-5.76
Maternal A – Maternal C	-9.58***	-4.81
Maternal B – Maternal C	1.92 NS	0.96
Direct Heterosis	13.03***	6.53
Maternal Heterosis	9.44***	4.73
Direct Recombination	-10.11***	-5.07
Maternal Recombination	-5.91***-	-2.96

[1] Statistical significance levels: * = $P \leq 0.05$; ** = $P \leq 0.01$; *** = $P \leq 0.001$; NS = $P \geq 0.05$.

[2] Biological type effect: N = *Bos indicus*; A = *Bos taurus* adapted for tropical environments; B = *Bos taurus* from British origin; C = *Bos taurus* from Continental origin.

Overall, the A group had the poorest performance in both direct and maternal effects for weaning weight while the C group excelled in direct effects and the B group was better in maternal effects. The N group performed intermediately for both direct and maternal genetic effects.

Although the effects of direct and maternal heterosis for weaning weight were not of a very large magnitude, they were very important nevertheless. As expected, the direct and maternal recombination effects acted in the opposite direction to heterosis, possibly cancelling part of the positive effects of crossbreeding in the performance of calves at weaning. These results suggest that heterosis and recombination effects must be included in the analysis of composite populations, in addition to direct and maternal breed or biological type effects. Ignoring direct and maternal heterosis and recombination effects can have an impact in the rank of select seedstock when genetic merit is used as selection criteria.

References

D.P. Berry, J.M. Coyne, J. Doyle, R.D. Evans. (2020) Livest.Sci.,234:103984, doi.org/10.1016/j.livsci.2020.103984.

K. Boldman, L.A. Kriese, L.D. Van Vleck, C.P. Tassell, S. Kachman. (1995). USDA, ARS.

L.V. Cundiff, K.E. Gregory. In: Cattlemen's College, (1999) Cattle Industry Annual Meeting and Trade Show, National Cattlemen's Beef Association. Charlotte, NC.

G. Dickerson. Animal Breeding Abstracts, (1969). 37, 191-202.

J.B.S. Ferraz, J.P. Eler, B.L. Golden. (1999a). Rev. Bras. Repr. Anim., v.23, p.115-117.

J.B.S. Ferraz, J.P. Eler, B.L. Golden. (1999b). Rev. Bras. Repr. Anim.., v.23, p.111–113.

K.E. Gregory, J.C.M. Trail, H.J.S Marples, J. Kakonge. (1985). J. Anim. Sci, 60:1165–1174. https://doi.org/10.2527/jas1985.6051165x

I. Misztal *et al.* (2014). University of Georgia, Athens. Available: http://nce.ads.uga.edu/wiki/lib/exe/fetch.php?media=blupf90_all7.pdf

I. Misztal. *et al.* (2002). Proc of the 7th WCGALP, Montpellier, France.

P.M. VanRaden, A.H. Sanders, A.H. J Dairy Sci. (2003) 86(3):1036-44, https://doi.org/10.3168/jds.S0022-0302(03)73687-X.

647. Breed differences, heterosis and recombination effects for calving interval in a composite beef cattle breed

L.T. Gama[1,2*], R. Espigolan[3], F.S. Baldi[4], J.P. Eler[3], E.C.M. Oliveira[3], R. Nuñez-Dominguez[5] and J.B.S. Ferraz[2]

[1]CIISA, Faculty of Veterinary Medicine, University of Lisbon, Av. Universidade Técnica, 1300-477 Lisboa, Portugal; [2]Associate Laboratory for Animal and Veterinary Sciences (AL4AnimalS), Portugal; [3]Research Center of Animal Breeding, Biotechnology and Transgenesis, College of Animal Science of University of São Paulo, 13635-900, Pirassununga, SP, Brazil; [4]FCAV/UNESP, 14884-900, Jaboticabal, SP, Brazil; [5]Universidad Autónoma de Chapingo, Estado de México CP 56230, Mexico; ltgama@fmv.ulisboa.pt

Abstract

Montana is a composite cattle breed developed in Brazil in the 1990's, established with the goal of performing well in tropical climates. The Montana breed was developed with a contribution to its gene pool of nearly 52% from *indicus* cattle (N), 25% from British breeds (B), 13% from adapted *taurus* breeds (A) and 10% from continental breeds (C). Nearly 64% of the maximum heterozygosity has been retained in Montana cattle. Our results indicate that important differences in calving interval exist between the biological groups contributing to the formation of Montana cattle, with the lowest calving interval associated with the contribution of the B group, followed by N. The maximum estimated heterosis for calving interval in this population is about 12 days, corresponding to 3% of the mean. The effect of recombination was expressed as a 'recombination gain', revealing the possible existence of negative epistatic combinations retained in purebred animals.

Introduction

Crossbreeding is a well-established strategy in animal breeding, with the goals of capitalizing on heterosis and taking advantage of breed differences in various production and adaptation traits (Dickerson 1969, Dickerson, 1973). In tropical regions, crosses with B. *indicus* are very common, given their ability to adjust to the climate-associated challenges prevailing in those regions (Ferraz and Felicio, 2010). Among the different crossbreeding strategies possible in livestock species, the development of composite breeds is one approach that is particularly appealing, because it is essentially managed as a purebred system while enabling the combination of the additive genetic effects of various breeds, though maintaining some of the benefits of heterosis (Koch *et al.*, 1985). Nevertheless, in composite breeds the parents used are crossbred, and any favourable epistatic combinations that were maintained in purebred animals tend to be broken, with a consequence known as 'recombination loss' (Dickerson, 1969, Kinghorn, 1980). The consequences of recombination may, to some extent, offset the benefits arising from heterosis in these crosses (Brotherstone and Hill, 1994); however, it is also possible that recombination may result in a gain rather than a loss if, for example, epistatic interactions resulted in the maintenance of genetic combinations in purebred animals with a negative net effect on some trait(s). The Montana is a breed of cattle which was developed in Brazil in the 1990's, with the goal of combining various breed sources in one composite type of cattle, with the ability to perform well in tropical climates (Ferraz *et al.*, 1999).

The objective of this work was to investigate how calving interval in Montana beef cattle is affected by the genetic contributions of different biological types of cattle, and by heterosis and recombination effects, with the goal of assessing in retrospect the best way to combine the available breed resources.

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_647, © L.T. Gama 2022

Materials & methods

In this study, we have analysed the data set comprising pedigree and performance records for the Montana breed since it was established in 1994.

Animals. The Montana composite breed was established from four biological groups of cattle, as follows: N group, including *Bos indicus* breeds (mostly Nellore cattle); A group, including *Bos taurus* adapted to tropical conditions (Bonsmara and Senepol); B group, including *Bos taurus* of British origin (mostly Aberdeen Angus and Red Angus); C group, including *Bos taurus* of continental origin (mostly Simmental). Throughout the development of the Montana composite, those breeds were combined in different proportions, with the goal of maintaining a predominant contribution of the N group, as it is expected to have better adaptation to tropical climates. Over time, animals were selected, mostly for growth and reproductive efficiency.

Pedigrees were recorded, such that the expected contribution of each biological group to a given animal could be computed from the genetic composition of its parents, in a recursive procedure. The level of heterosis and recombination effects expected for each animal were estimated from the genetic composition of each individual and its parents, according to the model of Dickerson (1969), and computed with the modification proposed by VanRaden (2003).

Assuming that α_i^s and α_i^d are, respectively, the proportional contribution of the ith breed to the sire and dam of the jth individual, the expected heterosis in that individual is calculated as:

$$Het_j = 1 - \sum_{i=1}^{n} (\alpha_i^s \, \alpha_i^d)$$

and the expected recombination effects can be computed as:

$$Rec_j = 1 - \sum_{i=1}^{n} \frac{1}{2} \left[(\alpha_i^s)^2 + (\alpha_i^d)^{\,2} \right]$$

Data and statistical analyses. Calving intervals were calculated for all the cows in the data set, and validated if they were between 280 and 499 days. After this validation, a total of 213,038 intervals by 79,750 cows with a mean calving interval of 413.5±113.5 days were included in a mixed model analysis. The model included the fixed effects of contemporary group (418 groups), age of cow (quadratic covariate with a mean of 68.8±35.5 months), the linear effect of the contribution by each of the four biological type groups and the linear effect of the expected heterosis and recombination component for the individual. The random effects in the mixed model were the breeding value of the animal and the permanent environmental effect of the cow, using a genetic relationship matrix with 181,407 animals.

Data were analysed and variance components estimated using the BLUPF90 package and the AIREML procedure (Misztal *et al.*, 2014). After convergence, the estimated variance components were used in analyses with MTDFREML (Boldman *et al.*, 1993), to estimate the significance of differences between the solutions for biological group effects, by specifying the appropriate estimable functions to test various contrasts among biological group solutions.

Results & discussion

The calculated contributions of each of the four biological groups to the population analysed and the mean level of heterosis (relative to F1) and of recombination effects are in Table 1.

Table 1. Estimated contribution (mean % ± standard deviation) of different biological groups to the Montana population analysed and proportional heterosis and recombination effect achieved.

Contribution to current population[1]	
N	52.2±28.2
A	12.6±21.2
B	25.0±24.7
C	10.2±16.9
Heterosis	0.637±0.384
Recombination effect	0.212±0.233

[1] N = *indicus* breeds; A = *taurus* adapted breeds; B = British breeds; C = Continental breeds.

These results indicate that approximately one-half of the genetic pool of Montana cattle is contributed by the N group, one quarter by the B group, and nearly one quarter is shared between the A and C groups, in similar proportions. The level of realized heterosis was quite high, achieving nearly 64% of the maximum heterosis observed in F1 individuals. On the other hand, the amount of recombination loss achieved is over 20%, which represents the expected loss of the possible epistatic superiority present in purebreds.

Differences between biological groups for calving interval were large (Table 2), and the estimates indicate that B is the group with the shortest calving interval (i.e. best reproductive efficiency), followed by N, which was better than the A and B groups. The A group resulted in the longest calving interval.

The estimated heterosis for calving interval (Table 3) indicates that full heterozygosity, such as that observed in F1 females, is expected to result in a reduction of about 12 days in calving interval, which corresponds to about 3% of the global mean. The recombination effect had a negative value associated, i.e. it reveals the possible existence of a negative epistatic combination retained in purebred animals, which would be broken and result in a 'recombination gain' when crossbred parents are used.

Even though the impact on calving interval of both heterosis and recombination effects appears to be small in Montana cattle, these are nevertheless significant and larger than those reported for dairy cattle (Wall *et al.*, 2005). Further studies should be conducted to investigate the levels of heterosis and recombination effects realized when specific combinations of biological groups are considered in the statistical model.

Table 2. Estimated difference between biological types for calving interval.[1]

	N-A	**N-B**	**N-C**	**A-B**	**A-C**	**B-C**
Estimate (days)[2]	-19.47***	10.31***	-6.79*	29.77***	12.67***	-17.10***
Standard error	2.67	2.60	3.05	2.56	2.97	2.23

[1] Biological types: N = *indicus* breeds; A = *taurus* adapted breeds; B = British breeds; C = Continental breeds.

[2] Statistical significance: * = $P \leq 0.05$; ** = $P \leq 0.01$; *** = $P \leq 0.001$.

Table 3. Estimated effects of heterosis and recombination on calving interval in Montana cattle.

Parameter	**Estimate (days)**	**Standard error**	**% of mean**
Heterosis	-12.37***	1.33	3.0
Recombination effect	-4.48***	1.77	1.1

In conclusion, important differences in calving interval ability were detected between the biological groups contributing to the formation of Montana cattle, with advantage for the British group, followed by *indicus*. Heterosis for calving interval was small but significant, and a recombination gain was observed, possibly reflecting the existence of negative epistatic combinations in purebred animals.

Acknowledgements

The authors thank the support received from the Centre for Interdisciplinary Research in Animal Health – CIISA, through funding from Project UIDB/00276/2020 financed by FCT – Fundação para a Ciência e a Tecnologia, I. P.

References

Boldman K., Kriese L.A., Van Vleck L.D., Tassell C.P., and Kachman S. (1995). A Manual for Use of MTDFREML – a Set of Programs to Obtain Estimates of Variances and Covariances. USDA, ARS. Available at: https://data.nal.usda.gov/dataset/mtdfreml-multiple-trait-derivative-free-restricted-maximum-likelihood.

Brotherstone S., and Hill W.G. (1994) Livest. Prod. Sci. 40(2):115–120. https://doi.org/10.1016/0301-6226(94)90041-8

Dickerson G.E. (1973) Proc. of an Animal Breeding Symposium in Honor of J.L. Lush. American Society of Animal Science: 54-77. https://doi.org/10.1093/ansci/1973.Symposium.54

Dickerson G.E. (1969) Animal Breeding Abstracts 37:191-202.

Ferraz J.B., and Felicio P.E. (2010) Meat Science 84(2):238-243. https://doi.org/10.1016/j.meatsci.2009.06.006.

Ferraz J.B.S., Eler J.P., and Golden B.L. (1999) Revista Brasileira de Reprodução Animal 23(2):115-117.

Kinghorn B. (1980). J. Anim. Breed. Genet. 97:138-143. https://doi.org/10.1111/j.1439-0388.1980.tb00919.x

Koch R.M., Dickerson G.E., Cundiff L.V., and Gregory K.E. (1985). J. Anim. Sci. 60(5):1117-1132. https://doi.org/10.2527/jas1985.6051117x.

Misztal I., Tsuruta S., Lourenco D., Masuda Y., Aguilar I., *et al.* (2014). Manual for BLUPF90 family of programs. Available at: http://nce.ads.uga.edu/wiki/lib/exe/fetch.php?media=blupf90_all7.pdf.

VanRaden P.M., and Sanders A.H. (2003). J. Dairy Sci. 86(3):1036-1044. https://doi.org/10.3168/jds.S0022-0302(03)73687-X

Wall E., Brotherstone S., Kearney J.F., Woolliams J.A., and Coffey M.P. (2005). J. Dairy Sci. 88(1):376–385. https://doi.org/10.3168/jds.S0022-0302(05)72697-7.

648. Prediction ability for growth and maternal traits using SNP arrays based on different marker density in Nellore cattle

F. Baldi[1*], J.D. Rodríguez[1], R. Espigolan[2], R.B. Lôbo[3], R. López-Correa[4] and I. Aguilar[5]

[1]São Paulo State University, Faculty of Agricultural and Veterinary Sciences, Department of Animal Science, Jaboticabal 14884-900, Brazil; [2]University of São Paulo, Faculty of Animal Science and Food Engineering, Department of Veterinary Medicine, 13535-900 Pirassununga, Brazil; [3]National Association of Breeders and Researchers, Ribeirão Preto, Brazil; [4]University of the Republic, Faculty of Veterinary Medicine, Department of Genetics and Animal Improvement, Montevideo, Uruguay; [5]National Institute of Agricultural Research, Montevideo, Uruguay; fernando.baldi-rey@unesp.br

Abstract

The prediction ability for growth and maternal traits using different low-density customized SNP arrays selected by informativeness and distribution of markers across the genome employing single-step genomic BLUP (ssGBLUP) was performed. The SNPs arrays were customized based on different length size windows per chromosome to select proportions of 10.0, 5.0, 2.5, 1.25 and 0.625% of markers from the 427k panel. The SNPs arrays were customized with densities of 40k, 20k, 10k, 5k and 2k, respectively. Two criteria were used to select the SNPs within a chromosome window: (1) highly informative SNPs; and (2) evenly distributed SNPs. The linear regression method (LR) was used to evaluate the prediction ability, inflation and bias of GEBV. The prediction ability was higher as the array density increased and it was similar with customized arrays higher than 10k SNPs. The genomic selection with low-density customized arrays could be feasible and cost-effective in Nellore cattle.

Introduction

To predict the GEBV for complex traits, the highest the number of markers, the increased is the chance to pick-up more markers in LD with a quantitative trait locus (QTL), improving the prediction ability of the markers. However, the use of denser marker arrays does not guarantee better predictions since the additive genetic variance is captured by a limited number of markers (Lourenco *et al.*, 2020). In a simulated study, Pocrnic *et al.* (2016) demonstrated that the number of markers required to explain the variation of the genomic matrix increased as a function of the Ne and concluded that the optimal dimension of the genomic matrix should include markers explaining 95 to 98% of the additive genetic variation to obtain reliable predictions with a less computational cost. The customization of low-density SNPs arrays has been a topic of interest to reduce the genotyping costs as well as the computational requirements for the genomic evaluations. It is well understood that *Bos taurus indicus* breeds need a different subset of informative SNPs than *Bos taurus taurus* breeds to adequately perform genomic selection in such breeds (Nayee *et al.*, 2018). The Illumina BovineHD BeadChip opened opportunities to develop customized or *in silico* SNPs arrays for *Bos taurus indicus* beef cattle breeds to be applied on genomic evaluations. Therefore, the knowledge of the population structure and the use of low-density SNP arrays that prioritize informative markers is an interesting strategy to obtain reliable and unbiased GEBV predictions with a lower cost and higher computational optimization (Wongpom *et al.*, 2019). The minor allele frequency (MAF) is an important informative propriety of the population's background to predict the GEBV (Zhu *et al.*, 2017), and the estimation of the genomic matrix allows the inclusion of more markers with high MAF (VanRaden, 2008). The objective of this study was to evaluate the prediction ability for growth and maternal-related traits using SNPs arrays with different marker densities based on SNP informativeness employing the ssGBLUP methodology in a Nellore beef cattle population.

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_648, © F. Baldi *et al.* 2022

Materials & methods

Phenotypic and genotypic records from the Nellore Brazil breeding program coordinated by the National Association of Breeders and Researchers (ANCP, Ribeirão Preto-SP, Brazil) were used. A total of 1,195,831 records for adjusted weight at 210 days were used. Weight records were measured between 1974 and 2018. The phenotypic and genomic data of the population was divided into partial and complete datasets using information of animals born from 1974 to 2008 and 1974 to 2018, respectively. A total of 945 representative sires of the main Nellore lineages were genotyped with the Illumina BovineHD BeadChip (Illumina Inc., San Diego, CA, USA). Quality control was performed on PreGSf90 program (Aguilar *et al.*, 2014). After quality control, 427,495 markers (427k) and 914 animal samples were available for analyses. The descriptive statistic for adjusted weight at 210 days is presented in Table 1. The SNPs arrays were customized based on different length size windows per chromosome to select proportions of 10.0, 5.0, 2.5, 1.25 and 0.625% of markers from the 427k panel. As a result, the SNPs arrays were customized with densities of 40k, 20k, 10k, 5k and 2k, respectively. Two criteria were used to select the SNPs within a chromosome window: (1) highly informative SNPs (H-I); and (2) evenly distributed SNPs (E-D). For the first criteria, markers with MAF close to 0.5 were selected by chromosome for each customized marker density. To select evenly distributed SNPs, the average distance between adjacent SNPs by chromosome for each customized marker density was considered. A total of 11 scenarios were evaluated considering the customized SNP arrays (427K, 40K_H-I, 40K_E-D, 20K_H-D, 20K_E-D, 10K_H-I, 10K_E-D, 5K_H-I, 5K_E-D, 2K_H-I, and 2K_E-D). The contemporary group (CG) were composed by farm, sex, year and season of birth, and management group at weaning (n=36,196). Records within ±3.5 standard deviations range from the CG mean, and CG with at least ten animals were considered in the analysis. The ssGBLUP was implemented to obtain the genomic predictions (Aguilar *et al.*, 2010). The model included the fixed effects of CG, calving dam age and the animal age at weaning as covariable (linear effect). The random effects of direct genetic additive, maternal genetic and maternal permanent environmental were also included in the model. The prediction ability, inflation, and bias of the genomic prediction for each scenario was evaluated with the Linear Regression (LR) method (Legarra and Reverter, 2018). The genotyped animals born after 2008 were selected for the validation dataset (n=267) and those before 2008 for the training dataset (n=647). The correlation (r) between the GEBVs obtained with the complete dataset and 427k array (GEBVc_427k) with the GEBVs predicted with the partial dataset (GEBVp) for each scenario (GEBVp_427k, GEBVp_40k, GEBVp_20k, GEBVp_10k, GEBVp_5k, GEBVp_2k) was calculated. The inflation was evaluated through the regression of GEBVc_427k on GEBVp for each scenario considering only the animals of the validation dataset. The bias was measured as the difference of the average GEBVp with the average GEBV of complete dataset.

Table 1. Number of records and descriptive statistics for the complete and partial dataset for adjusted weight at 210 days.

Dataset	Genotypes	Offspring	Total records	Mean±SD
Partial[1]	647	84,510	523,935	181.49±31.14
Complete[2]	914	266,546	1,195,831	189.10±31.10

[1] Partial= dataset with phenotypic information from 1974 to 2008.

[2] Complete= dataset with phenotypic information from 1974 and 2018.

Results

The prediction ability or correlation (r), inflation (b1) and bias (d) between the GEBVc_427k and the GEBV obtained with customized SNP densities in the validation subset (partial dataset) is shown in Table 2.

Discussion

The prediction ability improved as the marker density increased. The prediction ability was similar and moderate to high for marker density arrays higher than 10k. Comparing the criteria to select SNPs (H-I or E-D), similar prediction abilities were obtained for marker densities higher than 2k arrays. For 2k arrays, higher prediction abilities for W210 were observed for E-D markers. The inflation (b) of the predictions increased as the marker density decreased. The higher inflation (b) was observed for marker densities arrays of 2k and 5k, while the 10k, 20k, and 40k displayed similar and less inflated predictions. Minor differences between H-I and E-D criteria were detected when considering marker densities arrays below 5k. As the SNP density decays, the low heritability trait (0.12), such as the maternal effect for adjusted weight at 210 days, displayed the highest inflation (b), and the inflation was maximum at the lowest SNP density arrays. The bias (d) tends to decrease as the SNP densities array decrease, and there were no major differences among the criteria applied to select such SNPs. Despite there was a trend in decreasing the bias (d) as the SNP density array decreased, the dispersion or variance of the predictions increased, elucidated by the standard deviation. For all customized SNP arrays, the bias was higher for direct effects than did for the maternal. The results obtained in this study shown that moderate to low density in silico arrays, prioritizing

Table 2. Prediction ability (r), inflation (b), and bias (d) for the validation partial dataset for direct (W210) and maternal effect (MW210) considering different customized SNP arrays.

Trait	Array	Criteria[1]	r	b±SD	d±SD
W210	427k	-	0.78	0.95±0.05	3.47±4.65
	40k	H-I	0.78	0.95±0.05	3.47±4.66
		E-D	0.78	0.94±0.05	3.50±4.68
	20k	H-I	0.78	0.94±0.05	3.37±4.70
		E-D	0.77	0.93±0.05	3.48±4.74
	10k	H-I	0.77	0.94±0.05	3.36±4.76
		E-D	0.77	0.92±0.05	3.47±4.75
	5k	H-I	0.76	0.90±0.05	3.29±4.85
		E-D	0.77	0.88±0.05	3.47±4.79
	2k	H-I	0.72	0.82±0.05	3.38±5.18
		E-D	0.74	0.86±0.05	3.39±5.03
MW210	427k	-	0.48	0.83±0.08	0.48±2.95
	40k	H-I	0.47	0.80±0.08	0.48±2.98
		E-D	0.47	0.81±0.08	0.46±2.96
	20k	H-I	0.46	0.79±0.08	0.47±2.99
		E-D	0.47	0.81±0.08	0.50±2.90
	10k	H-I	0.46	0.79±0.08	0.44±3.00
		E-D	0.47	0.81±0.08	0.45±2.98
	5k	H-I	0.45	0.77±0.08	0.45±3.01
		E-D	0.45	0.74±0.08	0.41±3.04
	2k	H-I	0.43	0.69±0.08	0.43±3.11
		E-D	0.42	0.67±0.08	0.43±3.15

[1] H-I= selected markers with high informativeness; E-D= markers evenly distributed.

informative markers, were capable to estimate genomic relationship coefficients allowing the estimation of reliable genomic predictions for growth and maternal related traits in Nelore cattle. The results revealed that genomic selection with low-density customized arrays could be feasible and cost-effective in Nellore beef cattle. Therefore, there is an important margin to reduce high or moderate to low-density SNPs arrays without compromising the predictive capacity of the genomic information using the ssGBLUP method. This fact opens the opportunity for indicine beef cattle breeding programs develop *in silico* SNPs markers arrays for GEBVs predictions, minimizing the number of SNPs to be imputed. At least 10,000 informative SNPs obtained from the Illumina BovineHD BeadChip SNPs are necessary to adequately predict the GEBVs for growth and maternal related traits of young candidates with the ssGBLUP method.

References

Aguilar, I., Misztal, I., Johnson, D. L., Legarra, A., Tsuruta, S., *et al.* (2010). J. Dairy Sci. 93(2) 743–752. https://doi.org/10.3168/jds.2009-2730

Lourenco, D., Legarra, A., Tsuruta, S., Masuda, Y., Aguilar, I., *et al.* (2020) Genes, 11(7) 1–32. https://doi.org/10.3390/genes11070790

Nayee, N., Sahana, G., Gajjar, S., Sudhakar, A., Trivedi, K., *et al.* (2018). J. Anim. Breed. Genet., 135(6) 432–441. https://doi.org/10.1111/jbg.12356

Pocrnic, I., Lourenco, D. A. L., Masuda, Y., Legarra, A., & Misztal, I. (2016) Genetics, 203(1) 573–581. https://doi.org/10.1534/genetics.116.187013

VanRaden, P. M. (2008) Journal of Dairy Science 91(11) 4414–4423. https://doi.org/10.3168/jds.2007-0980

Wongpom, B., Koonawootrittriron, S., Elzo, M. A., Suwanasopee, T., Jattawa, D. (2019). Asian-Australas. J. Anim. Sci., 32(9) 1340–1348. https://doi.org/10.5713/ajas.18.0816

Zhu, B., Zhang, J. J., Niu, H., Guan, L., Guo, P., *et al.* (2017) J. Integr. Agric. 16(4) 911–920. https://doi.org/10.1016/S2095-3119(16)61474-0

649. Applying single step genomic evaluation and integrating local with foreign EBVs to increase calving performance

S. Naderi*, T. Pabiou and R.D. Evans

The Irish Cattle Breeding Federation, Highfield House, Newcestown Road, Bandon, Cork, Ireland; snaderi@icbf.com

Abstract

The objective of this study was to validate single step genomic breeding values (ssGEBV) for direct calving difficulty compared with non-genomic breeding values (EBV) and integrated (ssG)EBV. Calving difficulty genetic evaluation are undertaken for dairy heifers (DH), dairy cows (DC), beef heifers (BH) and beef cows (BC) calving scores with birth weight and birth size included as predictor traits. Single step GTBLUP approach was applied for a 6-trait calving difficulty evaluation on 1.77 million genotyped animals. To incorporate the international EBVs into domestic evaluation a joint evaluation was run where deregressed EBVs (DRP) were used as extra records weighted by their associated effective record contributions (ERC). The correlation between (G)EBV from well proven sires and their adjusted phenotypes was considered as the accuracy. Including genomic information and applying ssGTBLUP improved the accuracy for all six traits. In addition, blending the international EBVs into the domestic evaluation further improved the accuracy.

Introduction

The current routine GEBV estimation for Irish dairy and beef calving performance traits are based on a two-step approach, in which a reference population is first built to estimate the SNP marker effects, which are then used to calculate GEBVs for young animals. Incorporating the genomic information into traditional animal models used in national evaluations and applying a single-step genomic evaluations (ssGBLUP) could reduce potential biases in the GEBV estimation by simultaneously combining genomic, pedigree and all local phenotypic information. ssGBLUP has not been broadly implemented in large-scale national dairy or beef cattle evaluations due to large computational cost. Mäntysaari *et al.* (2017) developed an efficient computational approach for ssGBLUP called ssGTBLUP which requires less computation than ssGBLUP in iteratively solving by a preconditioned conjugate gradient (PCG) method. Moreover, integrating foreign information (i.e. MACE) into the national evaluation is advantageous but poses further challenges. Implementing a joint evaluation with national phenotypes in a single evaluation would enable the propagation of foreign information into domestic animals' evaluations even if they did not participate in the international evaluation itself. Therefore, the objective of this study was to explore and validate a ssGTBLUP approach comparing it with a traditional evaluation but also exploring the benefit of integrating foreign EBVs into the Irish genetic evaluation for calving performance traits.

Materials & methods

Calving difficulty in Ireland is assessed on a scale from 1 (i.e. no assistance) to 4 (i.e. with veterinary assistance). Four different calving traits are defined: dairy heifers (DH), dairy cows (DC), beef heifers (BH) and beef cows (BC) were defined. Birth weight (BW; kg) and birth size (BSIZE; 1 to 5 scale) are included in the multi-trait genetic evaluation as predictor traits. The number of records as well as some statistics for each trait is in Table 1. The total number of animals in the pedigree was 25,957,178 while the number of genotyped animals in the evaluation was 1,775,185.

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_649, © S. Naderi *et al.* 2022

Table 1. Number of records and the statistics for six different traits of dairy heifers (DH), dairy cows (DC), beef heifers (BH), beef cows (BC), birth weight (BW; kg) and birth size (BSIZE).[1]

Traits	Mean	SD	Min	Max	No.[2]
DH	1.31	0.60	1.00	4.00	1,894,408
DC	1.21	0.51	1.00	4.00	6,347,887
BH	1.40	0.71	1.00	4.00	924,801
BC	1.26	0.57	1.00	4.00	5,889,128
BSIZE	3.16	0.68	1.00	5.00	1,932,005
BW	41.47	7.91	20.00	115.00	199,759

[1] Max = maximum; Min = minimum; SD = standard deviation.

[2] Number of phenotypic records (No.) for each trait.

Traditional breeding values (EBVs) as well as genomic breeding values (GEBVs) were estimated for each of the six traits using the following multiple linear mixed model:

$$y_i = \mathrm{Xb} + Z_a u_a + Z_m u_m + Z_{mpe} u_{mpe} + \mathrm{e}$$

where y_i is a vector of phenotypes record for all animals with records and for trait i (i =1: 6), X is the incidence matrix for fixed effects, b is the vector of fixed effects consisting of breed proportions and specific heterosis effects, birth year, sex of calf and age of dam, Z_a, Z_{am} and Z_{mpe} are the incidence matrices relating the random effects u_a for direct additive genetic, u_m for maternal genetic and u_{mpe} for maternal environmental effects to observations in y; and e is a vector of random residuals. For the traditional genetic evaluation, it was assumed that $\boldsymbol{u_{a(m)}} \sim N\left(0,\ \boldsymbol{A}\sigma^2_{ua(m)}\right)$ where $\sigma^2_{ua(m)}$ are the direct(maternal) additive genetic variances and A is pedigree relationship matrix, and $\boldsymbol{u_{mpe}} \sim N\left(0,\ \boldsymbol{I}\sigma^2_{umpe}\right)$ where σ^2_{umpe} are the maternal environmental variance For the estimation of GEBV, an efficient computational approach presented by Mäntysaari *et al.* (2017) called ssGTBLUP was used where the genomic relationship matrix was constructed based on the following equation:

$$\mathrm{G_C} = \mathrm{G_0} + \mathrm{C}$$

where $\mathrm{C} = w\mathrm{A}_{gg}$, w is the residual polygenic proportion, A_{gg} is the pedigree relationship matrix among the genotyped animals and $\mathrm{G_0}$ is the genomic relationship matrix constructed based on VanRaden (2008) method 1. The matrices and vectors which are describing the linear equations system as well as efficient computation method with low memory cost for solving the equations in ssGTBLUP is outlined in detail by Mäntysaari *et al.* 2020.

To integrate direct Interbull EBVs and reliabilities into the Irish dairy evaluation as well as direct and maternal Interbeef EBVs and reliabilities, direct and maternal DRP were included into the model as extra traits weighted with their corresponding ERC. The ERC for both direct and maternal EBVs were computed using the reversed Harris *et al.* (1988) algorithm based on the reliability, pedigree and the traits' heritability. Direct and maternal EBVs were deregressed into direct and maternal DRP using the methods proposed by Jairath *et al.* (1998). The direct ERC and DRP for Interbull information and ERC and DRP for direct and maternal effects using Interbeef information were computed independently. Details on avoiding double-counting of Irish and foreign information and calculating adjusted ERCs and DRPs as well as the modelling are described by Vandenplas *et al.* (2014) and Pabiou *et al.* (2018). For all the computations, the MiX99 software (MiX99 development team, 2020) was used.

Validation. To assess the model prediction accuracy a group of 'well' proven sires with at least 100 progeny in the domestic evaluation were identified for each trait. The progeny records of these sires were masked in all evaluations. The correlation between the (G)EBVs of those proven sires with their daughter yield deviation (DYD) reflected the prediction performance. DYDs were calculated based on the average yield deviation from progenies of each sire including all phenotypic records (full data set). The correlation between (G)EBV from validation animals (progeny of the proven sires) and their phenotypes which were corrected for all nongenetic effects was considered as prediction accuracy at animal level. The validation of the integrated (G)EBV was performed by comparing integrated (G)EBV with non-integrated (G)EBV for the corresponding evaluations. The slopes of the regressions of DYD on (G)EBV of the validation sires was computed to measure bias of the evaluations.

Results & discussion

Validation results from four different models at sire level as well as number of validation sires for each trait are presented in Table 2. The results demonstrate that including genomic information and applying the single step genomic approach improved the prediction accuracy for all six different traits. For instance, the prediction accuracy for DH increased from 0.40 to 0.59 by incorporating the genomic information into the model.

The greatest improvement in accuracy was for BW where the traditional approach with only pedigree information had a correlation of only 0.27 which improved to 0.49 when the ssGBLUP approach was applied. Across all traits, the slope of DYD on (G)EBV of validation sires where the expectation of the slope was 0.5, improved with the inclusion of genomic information into the model indicating ssGBLUP could reduce potential biases.

The integration of Interbull and Interbeef information into the domestic evaluation also improved prediction accuracy at both the EBV or GEBV level and across all six different traits; however, the gain in prediction performance was not huge and in context of bias reduction only small improvement was observed. Small number of sires in the validation set which had a foreign sire with a MACE or an Interbeef proof could explain this modest improvement (i.e. only 256 sires for DC). Prediction performance at the animal level

Table 2. Correlations and slope for direct (G)EBV of validation sires with(out) foreign EBVs for six different traits of dairy heifers (DH), dairy cows (DC), beef heifers (BH), beef cows (BC), birth weight (BW; kg) and birth size (BSIZE).

Trait	No. of sires	Statistic	EBV	EBV inc. foreign EBVs	GEBV	GEBV inc. foreign EBVs
DH	1,143	Cor.	0.4	0.43	0.59	0.63
		Slope	0.14	0.15	0.45	0.47
DC	5,194	Cor.	0.44	0.47	0.60	0.62
		Slope	0.26	0.28	0.57	0.59
BH	343	Cor.	0.59	0.62	0.63	0.66
		Slope	0.28	0.27	0.41	0.42
BC	9,289	Cor.	0.37	0.38	0.43	0.44
		Slope	0.20	0.19	0.48	0.47
BSIZE	2,235	Cor.	0.28	0.29	0.47	0.50
		Slope	0.11	0.10	0.38	0.38
BW	363	Cor.	0.27	0.25	0.49	0.52
		Slope	0.08	0.11	0.29	0.30

followed the same pattern as observed at sire level with the inclusion of genomic information or foreign information improving the accuracy (Table 3). The gain in prediction accuracies was more pronounced for incorporating the genomic information rather than integrating foreign EBVs into the model.

Table 3. Correlations for direct (G)EBV of validation animals with(out) foreign EBVs for six different traits of dairy heifers (DH), dairy cows (DC), beef heifers (BH), beef cows (BC), birth weight (BW; kg) and birth size (BSIZE).

Trait	EBV	EBV inc. foreign EBVs	GEBV	GEBV inc. foreign EBVs
DH	0.17	0.18	0.2	0.22
DC	0.14	0.16	0.17	0.18
BH	0.15	0.16	0.17	0.19
BC	0.13	0.14	0.2	0.21
BSIZE	0.12	0.12	0.2	0.2
BW	0.18	0.18	0.26	0.27

Conclusions

The result from present study is encouraging in that the single step genomic approach and using foreign breeding values in the Irish domestic genetic evaluation improved prediction performance and reduced the bias in the evaluations. However, before implementing, the validation results should be compared with the current routine two-step approach.

References

Harris B. and Johnson D. (1998). J. Dairy Sci. 81(10): 2723–2728. https://doi.org/10.3168/jds.S0022-0302(98)75829-1.

Jairath L., Dekkers J. C. M., Schaeffer L. R., Liu Z., Burnside E.B., and *et al.* (1998). J. Dairy Sci. 81(2): 550– 562. https://doi.org/ 10.3168/jds.S0022-0302(98)75607-3.

MiX99 Development Team, (2020) MiX99: Natural Resources Institute Finland (Luke). Jokioinen, Finland. URL: http://www.luke.fi/mix99.

Mäntysaari E.A., Evans R. D., and Strandén I. (2017). J. Anim. Sci. 95(11): 4728:4737.
https://doi.org/ 10.2527/jas2017.1912

Mäntysaari E.A., Koivula M., and Strandén I. (2020). J. Dairy Sci. 103(6):5314-5326.
https://doi.org/ 10.3168/jds.2019-17754.

Pabiou T., Pitkanen T., Evans R. D., Herpje E., and Vandenplas J. (2018). Proc. Of WCGALP, New Zealand, Auckland.

Vandenplas J., Colinet F.G., and Gengler N. (2014). Genet. Sel. Evol. 46 (1): 46-59. https://doi.org/ 10.1186/s12711-014-0059-3.

VanRaden P.M. (2008). J. Dairy Sci. 91 (11): 4414-4423. https://doi.org/ 10.3168/jds.2007-0980.

650. Understanding the genetics of fertility and temperament in Northern beef cattle using genomic technologies

J.P. Copley[1*], N.J. Corbet[2], J.M. Allen[3], A.R. Laing[4], G. Fordyce[1], M.R. McGowan[5], B.M. Burns[6], R.E. Lyons[7] and B.J. Hayes[1]

[1]Queensland Alliance for Agriculture and Food Innovation, Centre for Animal Science, University of Queensland, 306 Carmody Rd, St Lucia, QLD, 4072, Australia; [2]CQ University, Rockhampton North Building 361 – G20 CQIRP, 4700 Rockhampton, Australia; [3]University of New England, Agricultural Business Research Institute, Armidale, 2351, NSW, Australia; [4]Queensland Government, Department of Agriculture and Fisheries, Ayr, 4087 QLD, Australia; [5]University of Queensland, School of Veterinary Science, University of Queensland Gatton Campus, 4343 Lawes QLD, Australia; [6]Queensland Government, Department of Agriculture and Fisheries, Rockhampton, 4702 QLD, Australia; [7]Neogen Australasia, University of QLD, Gatton Campus, 4343 Lawes, Australia; james.copley@uq.edu.au

Abstract

Fertility and temperament are important drivers of productivity in the beef industry of northern Australia. The purpose of this study was to determine the heritability of temperament and fertility traits and to estimate the genetic correlation between these traits using a gBLUP analysis. Fertility phenotypes, Post-partum anoestrus interval (PPAI) and age at puberty (AGECL, Tscore) were recorded in Brahman (n=936), Tropical composite (n=1,097) and Smart Futures Heifers (n=3,696). Flighttime records totalled 4,645. The genotype data was imputed from 35k to 800k SNP. Heritability was 0.56(0.08), 0.37(0.08), 0.44(0.11), 0.24(0.08), 0.19(0.03) and 0.33(0.03) for AGECL(Brahman), AGECL (Tropical composite), PPAI (Brahman), PPAI (Tropical composite), Tscore (Smart Futures Heifers) and Flighttime respectively. Genetic correlation was -0.01(0.10), -0.06(0.11), -0.03(0.07) and 0.32(0.09) between AGECL and Flighttime, PPAI and Flighttime, Tscore and Flighttime and PPAI and AGECL respectively. The result suggests that selecting for improved temperament would not improve fertility.

Introduction

Fertility of beef breeding females, measured by pregnancy, weaning rates and age at puberty is poor in northern Australia. Whilst environment is a significant contributing factor, the heritability of fertility traits indicates that substantial gains could be made in fertility through targeted selection. Fertility traits are, however, time-consuming, expensive and in some cases, completely unfeasible to measure in the extensive environments of northern Australia. Temperament is also heritable and is comparatively simple to measure. A potential, beneficial genetic correlation between temperament and fertility would allow for fertility to be improved through measurement of temperament only. The first objective of this project was to verify the heritability of individual temperament and fertility traits. The second and primary objective was to probe a potential genetic correlation between temperament and fertility.

Materials &methods

Animal data. The data utilised in the project stems from research undertaken by the Cooperative Research Centre for Cattle and Beef Quality (CRC) as part of the Northern Australian breeding project (Johnston *et al.*, 2009). The animals were located on seven commercial beef operations located across Northern Australia – four for BRAH and three for TROPCOMP, as well as data from the 'Belmont' research station in Central Queensland (Johnston *et al.*, 2009). Data from the SmartFutures (SF) project was also included (Burns *et al.* 2013).

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_650, © J.P. Copley *et al.* 2022

Age at first *Corpus Luteum* (AGECL) was used to measure age at puberty. AGECL measurements in BRAH and TROPCOMP cohorts were obtained through multiple ovarian ultrasound scans of heifers conducted at 4-6 week intervals, commencing when heifers reach ~200 kg at 10-12 months (Johnston *et al.*, 2009). Length of Post-partum anoestrus interval (PPAI) measured the period of non-cyclicity following calving. Records in the BRAH and TROPCOMP cohorts were obtained similarly to AGECL, scanning commenced post-calving and continually every four weeks up until weaning (Johnston *et al.*, 2009). Tscore was an analogous measure to AGECL taken from a single ovarian ultrasound scan conducted at ~600 days of age using the procedures outlined by Corbet *et al.* (2017) in the SF cohort. Temperament was measured by Flighttime, using the standard practice outlined by Fordyce *et al.* (1988).

Statistical analysis. The statistical model fitted to the AGECL data set (separately for TROPCOMP and BRAH) was:

$$y=X\beta+Zu+e$$

Where; y = vector of phenotypes X= design matrix allocating phenotypes to fixed effects, β = vector of fixed effects, including age as a covariate, contemporary group, and percentage *Bos indicus*, Z= design matrix allocation records to animal effects, u = vector of (random effect) additive breeding values, one for each animal, distributed Normal $(0,G\sigma_u^2)$, e = vector of residual errors, distributed $N(0,\sigma_e^2)$. The G matrix is the genomic relationship matrix among the animals, calculated from the SNP data, following the formulas of VanRaden *et al.* (2008). GCTA was the software used to calculate the G matrix and to estimate the variance components, σ_u^2 and σ_e^2 (Yang *et al.* 2010). The heritability of the traits were calculated as $h^2 = \sigma_u^2 /(\sigma_u^2+ \sigma_e^2)$.

Results

The heritability of individual temperament and fertility traits are presented in Table 1.

Genetic correlations amongst fertility and temperament traits are given in Table 2.

Table 1. Heritability of fertility and temperament traits from the G-REML analysis. Standard errors are given in brackets.

Trait	n	Range	Std	h^2
AGECL (BRAH)	922	423-1,169	138.53	0.56 (0.08)
AGECL (TROPCOMP)	999	344-945	120.40	0.37 (0.08)
PPAI (BRAH)	582	17-484	108.44	0.44 (0.11)
PPAI (TROPCOMP)	822	19-473	109.89	0.24 (0.08)
Tscore	3,696	0-5	1.29	0.19 (0.03)
Flighttime	4,645	39-528.8	54.90	0.33 (0.03)

Table 2. Genetic correlations among fertility and temperament traits from G-REML analysis. Standard errors are given in brackets.

Trait	AGECL	PPAI	Tscore	Flighttime
AGECL	-	0.58 (0.10)	-	-0.01 (0.10)
PPAI	0.58 (0.10)	-	-	-0.06 (0.11)
Tscore	-	-	-	-0.03 (0.07)
Flighttime	-0.01 (0.10)	-0.06 (0.11)	-0.03 (0.07)	-

Discussion

The heritability estimates of AGECL across the BRAH and TROPCOMP datasets indicated that the trait was moderately to highly heritable. This confirmed that direct selection for AGECL is possible. The results aligned closely to those derived by Johnston *et al.* (2009) for the same trait (0.57) and Vargas *et al.* (1998) (0.45). The heritability in TROPCOMP was lower than the estimate by Johnston *et al.* (2009) (0.52). The minor divergence between these results was potentially due to differences in the fixed effect modelling in the respective datasets. The Tscore measurement used for the SF dataset was a related, but less precise measure of age at puberty compared with AGECL. It interpreted the development of the uterine tract of heifers from a single measurement taken at 600 days. The obtained result for Tscore aligned closely with estimates made by Corbet *et al.* (2017) (0.18-0.32). Tscore, while a less precise measurement of puberty, is potentially a viable alternative to the multiple scannings required for AGECL, allowing puberty traits to be recorded in larger herds from more extensive, commercial businesses. This approach has been utilised in studies such as Hayes *et al.* (2019).

The estimated heritability of PPAI in both the BRAH and TROPCOMP datasets suggested PPAI was moderately to highly heritable. The results compared favourably to estimates by Johnston *et al.* (2013) who found heritability of PPAI to be 0.52(0.14), 0.20(0.12), 0.24(0.14) for *Bos indicus* females aged three, four and five years respectively. PPAI was found to be sufficiently heritable as to warrant direct selection in targeted breeding programs. In productive beef herds where breeding females should ideally produce one calf every twelve months, reducing PPAI is vital to improving reproductive success (Shatz *et al.*, 2011).

Flighttime was estimated to be moderately heritable. Numerous other results by Burrow *et al.* (2001), Burrow *et al.* (1988) and Burrow and Corbet (2000), all using pedigree rather than genomic data in their estimates, found the heritability of Flighttime to be 0.40, 0.26, 0.35 and 0.48 respectively. Estimates by Valente *et al.* (2015) and Valente *et al.* (2017) found the heritability to be lower, 0.27(0.07) and 0.22(0.02). The estimate obtained by this study emphasises Flighttime is a low to moderately heritable trait suggesting that selection would be possible. Selection for any trait relies on the collection of precise data using repeatable and objective methods. Temperament has been subjectively measured using metrics such as 'crush score' and 'temperament score', however, using the objective measure of Flighttime avoids issues stemming from human error and unconscious bias (Curley *et al.*, 2006). The potential for selection of improved temperament would be welcomed in Northern Australia, where temperament impacts directly on productivity (Fordyce *et al.*, 1988).

Estimated genetic correlation between PPAI and AGECL was positive and high, 0.58(0.10), suggesting the possibility of indirect selection. The previous results indicated that both these traits are highly heritable on an individual basis. The estimated correlation between PPAI and AGECL for TROPCOMP, closely matched results obtained by Johnston *et al.* (2014) who found correlation to be 0.72(0.07) and 0.31(0.18) for Tropical composite and Brahman females respectively. This result suggests females with earlier puberty tend to have shorter PPAI and are thus more likely to reconceive and calve again within the ideal 12-month cycle. Puberty, either AGECL or Tscore, can be measured earlier in an animals life, this combined with the favourable genetic correlation means that reduced PPAI can be reliably selected for without waiting for the animals to conceive and subsequently calve.

The second objective was to investigate if superior temperament, expressed as longer Flighttime in female beef cattle, can be used as a proxy trait to select indirectly for fertility. The estimated genetic correlation between Flighttime, and all fertility traits was effectively zero. This result of low, negative correlations between temperament and fertility traits was similarly observed by Valente *et al.* (2015) who found the correlation between Flighttime and early puberty in males and early pregnancy in females to be -0.09(0.09)

and -0.07(0.08). Burrow *et al.* (2001) found no genetic correlation between Flighttime and fertility traits including pregnancy status and early puberty. Cooke *et al.* (2011) and Burrow *et al.* (1988) found a positive relationship, between temperament and success rate of artificial breeding programs, however this was evaluated to be caused by easier management of quieter cattle rather than these cattle being inherently more fertile. A significant genetic correlation would have allowed direct selection for temperament traits to lead to indirect selection for fertility traits. Temperament and fertility appear to be independent traits. To improve fertility, temperament cannot be used as indicator trait, rather fertility must be measured and selected for directly.

References

Burns B.M., Corbet N.J., Allen J.M., Laing A., Sullivan M.T. (2013) Proc. of Northern Beef Research Update Conference, Cairns, Australia.

Burrow H.M., and Corbet N.J. (2000) Aus. J. Agri. Res. 51(1):155-62. https://doi.org/10.1071/AR99053

Burrow H.M., Seifert G.W., Corbet N.J. (1988) Aus. Soc. Anim. Prod. 17:154-157.

Burrow H.M. (2001) Liv. Prod. Sci. 70(3):213-33. https://doi.org/10.1016/S0301-6226(01)00178-6

Cooke R.F., Bohnert D.W., Meneghetti M, Losi T.C., and Vasconcelos J.L.M. (2011) Liv. Sci.142(1):108-13. https://doi.org/10.1016/j.livsci.2011.06.024

Corbet N.J., Allen J.M., Laing A.R., Fordyce G, McGowan M.R. and Burns B.M. (2017) Anim. Prod. Sci. 58(9):1735-1742. https://doi.org/10.1071/AN16616

Curley K.O., Paschal J.C., Welsh T.H. and Randel R.D. (2006) J. Anim. Sci. 84(11):3100-3. https://doi.org/10.2527/jas.2006-055

Fordyce G, Wythes J.R., Shorthose W.R., Underwood D.W. and Shepherd R.K. (1988) J. of Exp. Agri. 28(6):689-93. https://doi.org/10.1071/EA9880689

Johnston D.J., Barwick S.A., Corbet N.J., Fordyce G, Holroyd R.G., Williams P.J. and Burrow H.M. (2009) Anim. Prod. Sci. 49(6):399-412. https://doi.org/10.1071/EA08276

Johnston D.J., Barwick S.A., Fordyce G, Holroyd R.G., Williams P.J., Corbet N.J. and Grant T (2013) Anim. Prod. Sci. 54(1):1-15. http://dx.doi.org/10.1071/AN13043

Johnston D.J., Corbet N.J., Barwick S.A., Wolcott M.L. and Holroyd R.G. (2014) Anim. Prod. Sci. 54(1):74-84. https://doi.org/10.1071/AN13044

Schatz T 2011 'Understanding and improving Heifer fertility in Northern Australia', Charles Darwin University, Darwin.

Valente T.S., Albito O.D., Sant'Anna A.C., Carvalheiro R, Baldi F, Albuquerque L.G. and da Costa M.J.R.P. (2017) Liv. Sci. 206:45-50. https://doi.org/10.1016/j.livsci.2017.10.010

Valente T.S., Sant'Anna A.C., Baldi F, Albuquerque L.G. and da Costa M.J. (2015) J. Appl. Genet. 56(3):349-54. https://doi.org/10.1007/s13353-014-0259-0

Yang J., Lee S.H., Goddard M.E. and Visscher P.M. (2011) Amer. J. Hum. Genet. 88(1):76-82. https://doi.org/10.1016/j.ajhg.2010.11.011

651. Genomic prediction with non-additive genetic effects for carcass weight and beef marbling in Japanese Black cattle

K. Inoue[1*], M. Nishio[2], Y. Inoue[3], M. Takeda[1] and H. Hirooka[4]

[1]National Livestock Breeding Centre, Nishigo, 961-8511 Fukushima, Japan; [2]NARO Institute of Livestock and Grassland Science, Tsukuba, 305-0901 Ibaraki, Japan; [3]Tottori prefectural Livestock Research Centre, Kotoura, 689-2503 Tottori, Japan; [4]Kyoto University, Sakyo, 606-8502, Kyoto, Japan; k1inoue@nlbc.go.jp

Abstract

In total, 1,908 genotyped Japanese Black cattle were used to infer non-additive genetic effects for carcass weight and beef marbling score (BMS) and to validate accuracy of the genomic predictions for the traits. Genomic models including dominance, imprinting, or both effects were used in this analysis. Ratios of the dominance and imprinting variance to the phenotypic variance for the traits were small and ranged from 0.08 to 0.10 and from 0.05 to 0.06, respectively. In addition, any genomic models including non-additive genetic effects did not improve accuracy of the genomic predictions. In contrast to the previous pedigree-based studies which showed large imprinting variances and effects for BMS, small imprinting variances and variance ratios for BMS were found in this study using genomic information. To find out the differences between pedigree and genomic based studies, further studies are required.

Introduction

Non-additive genetic effects, such as dominance and genomic imprinting effects are important as well as an additive genetic effect in animal breeding systems because they could lead to bias in genetic parameters and breeding values. Neugebauer *et al.* (2010) and Tier and Meyer (2012) reported imprinting effects for carcass traits using pedigree-based imprinting models in beef cattle. In addition, large relative imprinting variances for beef marbling score (BMS) were found in different Japanese Black cattle populations (Inoue *et al.*, 2020, 2021; Okamoto *et al.*, 2019). In recent years, the genotypic coding and construction of a genomic relationship matrix for an imprinting effect using single nucleotide polymorphism (SNP) data from animal-parents trios were developed by Nishio and Satoh (2015) in the context of genomic prediction. Using their methods, some studies have been reported dominance and imprinting effects and predictive ability of the models (e.g. Guo *et al.*, 2016). However, there are no reports on the effects of using SNP information in Japanese Black cattle. The objectives of this study were to infer dominance and imprinting effects for carcass weight (CW) and BMS and to validate predictive ability of the genomic models for these traits in Japanese Black cattle.

Materials & methods

Data. The data used in this study were obtained from the carcass records (i.e. Animal Care and Use Committee approval was not necessary) of 1,908 Japanese Black steers and heifers fattened in Japan. The carcass records included CW and BMS, which is scored from 1 (poor) to 12 (abundant). BMS was evaluated between the sixth and seventh rib in accordance with the Japan Meat Grading Association (1988). Those 1,908 animals with records and their parents were genotyped using the GeneSeek GGP LD v4 SNP chip (Neogen GeneSeek Operations, NE, USA). The call rate of these animals was over 0.95. The SNP genotype data were imputed to the Illumina BovineSNP50 v2 BeadChip (Illumina, CA, USA) and then were phased by using Beagle ver. 4.0 (Browning and Browning, 2007) with pedigree information to distinguish the parental origin of the alleles. Finally, 34,481 SNPs on the 29 cattle autosomes were used.

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_651, © K. Inoue *et al.* 2022

Statistical models. The following four statistical models were fitted to phenotypic values. The full model including an additive genetic effect in addition to dominance and imprinting effects (model 1) can be written as follows:

$$y = X\beta + Z_a a + Z_d d + Z_i i + e \quad (1)$$

where y is a vector of the observed traits CW and BMS; β is a vector of fixed effects, including the sex of animals (two levels; steer and heifer), slaughter year (13 levels; 2006 to 2018), slaughter month (12 levels), carcass market (five levels), linear and quadratic covariates for age at slaughter (average 28.6 months); a, d and i are vectors of additive genetic, dominance and imprinting effects, respectively, and distributed as $\mathbf{a} \sim N(0, \mathbf{G_a}\sigma_a^2)$, $\mathbf{d} \sim N(0, \mathbf{G_d}\sigma_d^2)$ and $\mathbf{i} \sim N(0, \mathbf{G_i}\sigma_i^2)$, respectively; e is a vector of random residual effects as $\mathbf{e} \sim N(0, \mathbf{I}\sigma_e^2)$; and X, Z_a, Z_d and Z_i are known incidence matrices. G_a, G_d and G_i are the additive genomic, dominance genomic and imprinting genomic relationship matrices, respectively; I is an identity matrix; σ_a^2 is the additive genetic variance; σ_d^2 and σ_i^2 are variances due to the dominance and imprinting effects, respectively; and σ_e^2 is the residual variance. All other covariances between random effects were assumed to be zero. The genomic relationship matrices G_a, G_d and G_i were calculated according to VanRaden (2008), Nishio and Satoh (2014) and Nishio and Satoh (2015), respectively, as follows:

$$\mathbf{G_a} = \frac{\mathbf{M_a M_a'}}{\sum_{j=1}^{m} 2q_j(1-q_j)},$$

$$\mathbf{G_d} = \frac{\mathbf{M_d M_d'}}{\sum_{j=1}^{m} \{2q_j(1-q_j)\}^2},$$

$$\mathbf{G_i} = \frac{\mathbf{M_i M_i'}}{\sum_{j=1}^{m} 2q_j(1-q_j)},$$

where M_a, M_d and M_i are $n \times m$ matrices (n is the number of genotyped individuals and m is the number of SNPs). Considering two alleles A_1 and A_2 with frequencies $1 - q_j$ and q_j, respectively, in the population, the elements of M_a, M_d and M_i for the i^{th} individual at the j^{th} SNP were calculated as follows (Nishio and Satoh, 2015):

$$\mathbf{M_a}_{ij} = \begin{cases} 2q_j & A_1A_1 \\ 2q_j - 1 & A_1A_2 (A_2A_1), \\ 2q_j - 2 & A_2A_2 \end{cases}$$

$$\mathbf{M_d}_{ij} = \begin{cases} -2q_j^2 & A_1A_1 \\ 2q_j(1-q_j) & A_1A_2 (A_2A_1), \\ -2(1-q_j)^2 & A_2A_2 \end{cases}$$

$$\mathbf{M_i}_{ij} = \begin{cases} 0 & A_1A_1 \\ 1 & A_1A_2 \\ -1 & A_2A_1 \\ 0 & A_2A_2 \end{cases},$$

The three reduced models, which excluded either imprinting (model 2) or dominance (model 3), or both (model 4), from the full model were as follows:

$$y = X\beta + Z_a a + Z_d d + e \quad (2)$$

$$y = X\beta + Z_a a + Z_i i + e \quad (3)$$

$$y = X\beta + Z_a a + e \quad (4)$$

The RKHS function in BGLR statistical package (Pérez and de los Campos, 2014) was used for obtaining posterior distributions of variance components. For each analysis, a single Markov chain with 15,000 samples was generated, with the initial 5,000 samples discarded as burn-in. Finally, variance components were determined based on the remaining 10,000 samples.

Cross-validation of genomic predictions. Predictive ability of the models was examined by 10-fold cross-validation. The animals with phenotypes were randomly split into 10 folds. Accuracy of the predictions was evaluated based on Pearson's correlation coefficient between the phenotypic values (y) and predicted phenotypic values ($\hat{y}$) which were the sum of the estimated fixed effects and predicted genetic effects.

Results

Posterior means of the variance components and ratios of variance for carcass traits are shown in Table 1. Additive genetic variances (σ_a^2) and ratios of σ_a^2 to the phenotypic variance (h_a^2) for CW in models including non-additive effects (model 1, 2 and 3) were slightly smaller than those in model including only an additive effect (model 4). Similarly, for BMS, σ_a^2 and h_a^2 in model 1 and 3 were slightly smaller than those in model 2 and 4. The ratios of dominance variance to the phenotypic variance (h_d^2) for CW and BMS were ranged from 0.08 to 0.10. The ratios of imprinting variance to the phenotypic variance (h_i^2) for CW and BMS were 0.06 and 0.05, respectively.

Correlations between phenotype and predicted phenotype for different models of analysis are shown in Table 2. The correlations for CW and BMS in models including non-additive effects (model 1, 2 and 3) were slightly higher than those in model including only an additive effect (model 4), in particular, the correlations in models including dominance effects (model 1 and 2) were highest among them but the differences were all small and negligible.

Table 1. Posterior means of the variance components and ratios of variance for carcass traits.

Trait[1]	Model[2]	Parameter[3]							
		σ_p^2	σ_e^2	σ_a^2	σ_d^2	σ_i^2	h_a^2	h_d^2	h_i^2
CW	1	2,417.7	850.2	1,246.5	187.7	133.3	0.52	0.08	0.06
(kg)	2	2,399.5	944.9	1,252.9	201.7		0.52	0.08	
	3	2,409.1	1,013.1	1,249.9		146.1	0.52		0.06
	4	2,404.0	1,127.8	1,276.2			0.53		
BMS	1	4.238	1.829	1.825	0.387	0.197	0.43	0.09	0.05
(no.)	2	4.225	1.930	1.863	0.432		0.44	0.10	
	3	4.220	2.171	1.823		0.227	0.43		0.05
	4	4.207	2.339	1.868			0.44		

[1] CW = carcass weight; BMS = beef marbling score.

[2] 1 = full model includes additive, dominance and imprinting effects; 2 = model includes additive and dominance effects; 3 = model includes additive and imprinting effects; 4 = model includes additive effect.

[3] σ_p^2 = phenotypic variance; σ_e^2 = residual variance; σ_a^2 = additive genetic variance; σ_d^2 = dominance variance; σ_i^2 = imprinting variance; $h_a^2 = \sigma_a^2/\sigma_p^2$; $h_d^2 = \sigma_d^2/\sigma_p^2$; $h_i^2 = \sigma_i^2/\sigma_p^2$.

Table 2. Correlations between phenotype and predicted phenotype for different models of analysis.

Trait[1]	Model[2]			
	1	2	3	4
CW	0.694	0.695	0.692	0.692
BMS	0.541	0.542	0.538	0.537

[1] CW = carcass weight; BMS = beef marbling score.

[2] 1 = full model includes additive, dominance and imprinting effects; 2 = model includes additive and dominance effects; 3 = model includes additive and imprinting effects; 4 = model includes additive effect.

Discussion

In studies using pedigree-based imprinting models, Neugebauer *et al.* (2010) reported a significant imprinting effect for paternal gametes on fat score in German Simmental. Tier and Meyer (2012) reported paternal and maternal gametic effects for intramuscular fat percentage in Australian Angus. In Japanese Black cattle, large imprinting variances for BMS were inferred in different populations (Inoue *et al.*, 2020, 2021; Okamoto *et al.*, 2019). In contrast to the previous pedigree-based studies, small imprinting variances and h_i^2 for BMS were inferred by using genomic relationship matrices in the present study. In addition, any genomic models including non-additive genetic effects did not improve the accuracy of the genomic predictions. Some previous studies also reported that non-additive genetic variances and effects were small and models including non-additive genetic effects did not improve predictive ability, probably due to overfitting of the models with more parameters (Hu *et al.*, 2016).Garcia-Baccino *et al.* (2020) reported h_d^2 of 0.01 for growth traits and no improvement of predictive ability on the models including non-additive genetic effects in American Angus. In Danish Duroc pigs, h_d^2 for dairy gain and backfat thickness were ranged from 0.040 to 0.049 and h_i^2 for the traits were ranged from 0.013 to 0.019 (Guo *et al.*, 2016). As described above, there were large differences between the imprinting effects for BMS from pedigree-based imprinting models and from genomic information. Therefore, further studies are required to find out the differences between pedigree and genomic based studies.

References

Browning S.R., and Browning B.L. (2007) Am. J. Hum. Genet. 81: 1084–97. https://doi.org/10.1086/521987

Garcia-Baccino C.A., Lourenco D.A.L., Miller S., Cantet R.J., and Vitezica Z.G. (2020) J. Anim. Sci. 98: skz384. https://doi.org/10.1093/jas/skz384

Guo X., Christensen O.F., Ostersen T., Wang Y., Lund M.S., *et al.* (2016) Genet. Sel. Evol. 48: 67. https://doi.org/10.1186/s12711-016-0245-6

Hu Y., Rosa G.J.M., and Gianola D. (2016) Genet. Sel. Evol. 48: 34. https://doi.org/10.1186/s12711-016-0213-1

Inoue K., Nishio M., Shoji N., and Hirooka H. (2020) J. Anim. Breed. Genet. 138: 45-55. https://doi.org/10.1111/jbg.12493

Inoue K., Inoue Y., Oe T., and Nishimura M. (2021) Anim. Sci. J. 92: e13504. https://doi.org/10.1111/asj.13504

Japan Meat Grading Association. (1988) Japan Meat Grading Association, Tokyo, Japan.

Neugebauer N., Rader I., Schild H.J, Zimmer D., and Reinsch N. (2010) J. Anim. Sci. 88: 523-532. https://doi.org/10.2527/jas.2009-2026

Nishio M., and Satoh T. (2014) PLoS One 9:e85792. https://doi.org/10.1371/journal.pone.0085792

Nishio M., and Satoh T. (2015) Genet. Sel. Evol. 47: 32. https://doi.org/10.1186/s12711-015-0091-y

Okamoto K., Oishi K., Nakamura R., Abe A., Inoue K., *et al.* (2019) J. Anim. Breed. Genet. 136: 190-198. https://doi.org/10.1111/jbg.12382

Pérez P., and de los Campos G. (2014) Genetics 198: 483-495. https://doi.org/10.1534/genetics.114.164442

Tier B., and Meyer K. (2012) J. Anim. Breed. Genet. 129: 359-368. https://doi.org/10.1111/j.1439-0388.2012.00996.x

VanRaden P.M. (2008) J. Dairy. Sci. 91: 4414–23. https://doi.org/10.3168/jds.2007-0980

652. Reaction norm models contribute to the investigation of environmental sensitivity for weaning weight in Nellore cattle

I. Carvalho Filho[1*], D.A. da Silva[1], C.S. Teixeira[1], T.L. Silva[1], L.G. Albuquerque[1,2] and R. Carvalheiro[1,2]

[1]São Paulo State University (UNESP), School of Agricultural and Veterinarian Sciences, Jaboticabal, Brazil; [2]National Council for Scientific and Technological Development (CNPq), Brasília, Brazil; ivancf08@gmail.com

Abstract

Reaction norm models (RNM) can be used to assess environmental sensitivity of animals for traits affected by genotype by environment interaction. Three RNM were tested for weaning weight (WW) in Nellore cattle. Phenotypic data from 772,855 animals were obtained from the Aliança Nelore dataset. The best model, according to the AIC criteria, was a spline linear-linear heteroscedastic RNM. Results suggest existence of a non-linear component in the reaction norm of WW and that the slope of the first and second segments are highly and negatively correlated. For animals selected in the low environment gradient (EG), the EBV average decreased in the medium and high EG. Trajectories along the EG when animals were selected in the medium and high gradients were similar. To reduce animal sensitivity to environmental variation, the selection could be based on genetic performance and environmental sensitivity, which can lead to a better genetic response for all EG.

Introduction

In genotype by environment interaction (G×E) studies, reaction norm models (RNM) can be used for the prediction of genetic values as a function of environmental conditions (Knap and Su, 2008). For traits that require additional random effects, such as maternal genetic and maternal permanent environment effects, the RNM can become highly parameterized and difficult to apply in practice. A possible solution is to fit a RNM only for the direct genetic effect and model the other effects for an average environment (Carvalho Filho *et al.*, 2021).

Finding the best model that fits the data is of utmost importance to promote more accurate genetic evaluations prediction. Thus, the objective was to compare different RNM for weaning weight in Nellore cattle considering goodness of fitting measures.

Materials & methods

Phenotypic data for weaning weight (WW) were obtained from the Aliança Nelore dataset. After the criteria used of keeping only animals with known parents, contemporary groups (CG) with at least 20 animals and WW between 100 and 280 kg, 772,855 animals from 15,448 CG were considered for the analysis. The CG was comprised by animals born in the same herd, year and season, of the same sex and raised in the same management group from birth to weaning. The CG effects estimated previously using a regular mixed model (without modelling G×E) were standardized to have mean and variance equal to 0 and 1, respectively, and were used in RNM as environmental gradients (EG). Genetic sensitivity to environmental variation was assessed by fitting three different RNM.

The first (RNM_homo) assumed residual variance homogeneity and is described by the equation:

$$y_{ijk} = \mathbf{X}_j\boldsymbol{\beta} + \emptyset\widehat{EG}_i + b_{0_j} + b_{1_j}\widehat{EG}_i + m_j + mpe_k + e_{ijk} \quad (1)$$

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_652, © I. Carvalho Filho *et al.* 2022

where: y_{ijk} is the phenotypic data of animal j in a given environment i, β is the vector of fixed effects and covariates; X_j is the incidence matrix related to β; ϕ is the overall linear fixed regression coefficient of y_{ijk} in $\widehat{EG}_i$; $\widehat{EG}_i$ is the standardized EG solution of environment i, estimated previously; b_{0_j} is the random direct additive genetic effect or the intercept of animal j for an average EG, b_{1_j} is the coefficient of random regression or slope of the animal j in the environment represented by $\widehat{EG}_i$; m_j is the maternal genetic effect, mpe_k is the maternal permanent environmental effect and e_{ijk} the residual effects. The following assumptions were assumed for RNM_homo:

$[b_{0_j}, b_{1_j}] \sim N(0, \mathbf{A} \otimes \begin{bmatrix} \sigma^2_{b_0} & \sigma_{b_0 b_1} \\ \sigma_{b_0 b_1} & \sigma^2_{b_1} \end{bmatrix})$, $m_j \sim N(0, \mathbf{A}\sigma^2_m)$, $mpe_k \sim N(0, \mathbf{I_d}\sigma^2_{mpe})$ and $e_{ij} \sim N(0, \mathbf{I_n}\sigma^2_e)$, where:

A is the relationship matrix based on pedigree information (⊗ is the Kronecker product); $\sigma^2_{b_0}$, $\sigma^2_{b_1}$ and $\sigma_{b_0 b_1}$ are the variances of the intercept, the slope and their covariance, respectively; σ^2_m is the variance of the maternal genetic effect, σ^2_{mpe} is the variance of the maternal; I_d is an identity matrix of order equal to the number of dams, I_d is an identity matrix of order equal to the number of animals with phenotypic records and σ^2_e is the residual variance.

The second model (RNM_hete) differed in relation to RNM_homo by assuming that the residual variance is heterogeneous across EG, using a linear regression on $\widehat{EG}_i$, with the intercept and slope coefficients being modeled using the log-residual function (Foulley and Quaas, 1995). The third model (RNM_l-l) also assumed heteroscedasticity, and used a linear-linear spline function to model the genetic merit of animals across the EG, using the same knot in the average EG (knot = 0) for all animals, according to the following equation:

$$y_{ijk} = \mathbf{X_j}\boldsymbol{\beta} + \emptyset_1 \widehat{EG}_i + \emptyset_1^* \widehat{EG}_i^* + b_{0_j} + b_{1_j} \widehat{EG}_i + b_{1_j}^* \widehat{EG}_i^* + m_j + mpe_k + e_{ijk} \quad (2)$$

where: $\emptyset_1^*$ is the difference of the regression coefficient between the first and the second segments of the linear-linear spline function of y_{ij} on $\widehat{EG}_i^*$, where $\widehat{EG}_i^*$ is equal to zero if $\widehat{EG}_i$ is less than or equal to zero, if $\widehat{EG}_i$ is greater than zero, $\widehat{EG}_i^*$ is equal to $\widehat{EG}_i$; $b_{1_j}^*$ is the difference between the first and the second segments of the linear–linear spline function of the random additive genetic effect of animal j on $\widehat{EG}_i^*$ expressed as a deviation from $\emptyset_1^*$; the other terms were as previously described.

(Co)variance component estimates for all models were obtained by restricted maximum likelihood using the AIREMLF90 software (Misztal *et al.*, 2002). The different models were compared based on the Akaike information criterion (AIC).

In all, 508 sires remained after the filter of having at least fifty offspring and among these at least ten offspring in the low EG (-3 to -1) and ten in the high EG (1 to 3). Ten sires (Top10) with the highest EBV in each environmental gradient (low, medium and high) were selected to have their reaction norm described along the environmental gradient. Additionally, forty sires with the highest EBV in the medium environmental gradient were selected. Of these, the ten most robust sires (top10 robust), i.e. with the slope closer to zero, will be shown.

Results

Based on the AIC criterion, both heteroskedastic models outperformed the homoscedastic model, and RNM_l-1 presented the best fit, suggesting the existence of a non-linear component in the trajectory of the WW reaction norm in Nellore cattle (Table 1).

Table 1. Estimates of variance (diagonals), covariance (upper triangular; in italic) and correlation (lower triangular; in bold) among coefficients of reaction norm models (RNM) for the additive genetic effect of weaning weight in Nellore cattle, with respective maternal (mat), maternal permanent environment (mpe) and residual variance estimates, and Akaike information criterion (AIC).

Model[1]	Coefficient[2]	b0	b1	b2	mat	mpe	e[3]	n[4]	AIC
RNM_homo	b0 (int)	106.65	*15.10*		48.16	64.20	212.72	6	4,624,242
	b1 (slp)	**0.63**	5.42						
RNM_hete	b0 (int)	101.62	*9.82*		49.52	64.13	5.37	7	4,624,219
	b1 (slp)	**0.46**	4.41				0.04		
RNM_l-l	b0 (int)	106.87	*16.40*	*9.46*	49.57	64.12	5.36	11	4,624,089
	b1 (slp1)	**0.47**	11.27	*9.22*			0.04		
	b2 (slp2)	**-0.29**	**-0.88**	9.64			0.02		

[1] RNM_homo: linear homoscedastic; RNM_hete: linear heteroscedastic; RNM_l-l: spline linear–linear heteroscedastic.

[2] b0–b2 coefficients of the RNM for the additive genetic random effect [int: intercept; slp: slope; slp1(2): slope segment 1(2)].

[3] Residual variance (RNM_homo) or residual coefficients associated with parameters of heteroscedastic RNM that were modelled using a log-residual function (Foulley and Quaas, 1995).

[4] Number of estimated parameters.

Figure 1 illustrates the expected consequence of different selection strategies, showing the reaction norms and the average of the ten best sires EBV for different EG. It is possible to observe the different trajectories along the EG according to the environment in which the animal is selected.

When observing the reaction norms trajectory for the 10 best selected sires in the low EG (Figure 1A), there is a decrease in EBV average and greater dispersion in the medium and high EG. Trajectories along the EG when animals were selected in the medium and high gradients are similar (Figure 1B and 1C). In the robust Top10, the EBV varies little along the EG (Figure 1D).

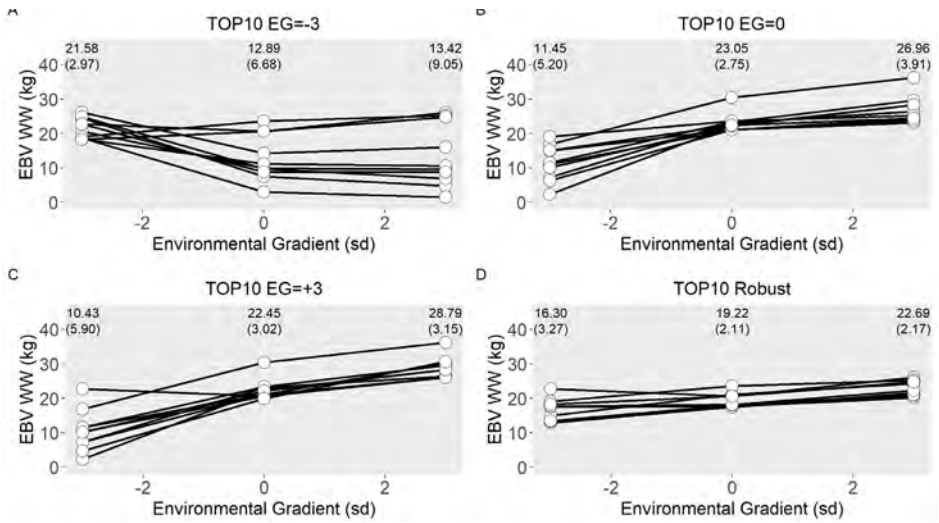


Figure 1. Estimated breeding value (EBV) reaction norms for weaning weight (WW) of Nellore cattle sires classified as top 10 for low (A), medium (B) and high environmental gradient (C), and top 10 robust (D), i.e. with slopes closer to zero that are among the best forty sires in the medium EG, considering the reaction norm model (RNM) fitted with a spline linear–linear function and assuming heteroscedasticity (RNM_l-l). The numbers above the reaction norms trajectories represent the mean and in parentheses, the standard deviation.

Discussion

RNM_homo and RNM_hete showed moderate to high correlation between intercept and slope, unlike Ribeiro *et al.* (2015), who found a high correlation between intercept and slope for weaning weight in Nellore cattle. However, this moderate correlation suggests that animals are reclassified in different environments and that it is possible to carry out joint selection in order to increase performance and reduce sensitivity (Santana *et al.*, 2013; Su *et al.*, 2006). The correlation estimates between the coefficients presented by RNM_l-1 suggest that this association is moderate only for the first segment. For the second segment, the correlation between intercept and slope was low and negative, showing that there may be different patterns of animals environmental sensitivity according to the EG level.

In Figure 1A, among the top ten sires for low EG, there is a reduction in the mean EBV value in the medium and high gradients, indicating that the selection response would be lower for these environments if the selection was based on animals with higher EBV for low EG. On the other hand, if the selection was based on superior animals for medium and high EG, the expected response in these environments would be higher for these EG, at the expense of worse expected responses for low EG (Figures 1B and 1C, respectively), when compared to the top ten sire averages for low EG.

Robust sires, i.e. sires that are less sensitivity to environmental variation, are shown in Figure 1D. These sires are among the top 40 sires in medium EG and in the top 50 sires in extreme EG (low and high). The selection of these robust animals could lead to an increase in performance and a decrease in environmental sensitivity.

Therefore, we concluded that reaction norm models can help in the selection of animals with high weaning weight and less sensitivity to environmental variation.

Acknowledgements

To 'Coordenação de Aperfeiçoamento de Pessoal de Nível Superior' (CAPES, finance code 001), National Council for Scientific and Technological Development (CNPQ 305435/2017-5), and São Paulo Research Foundation (FAPESP – #17/10630-2, #2018/20026-8, #2019/06361-1 and #2020/14846-2) for financial support, and to the Genetic Breeding Programs participating in the Aliança Nellore (www.gensys.com.br) for providing the data.

References

Carvalho Filho I., Silva D. A., Teixeira C. S., Albuquerque L. G., Silva F. F., *et al* (2021) XIV Simpósio Brasileiro de Melhoramento Animal, Santa Catarina, Brasil. http://sbmaonline.org.br/anais/xiv/pdf/4DPC.pdf

Foulley J. L., and Quaas R. L. (1995). Genet. Sel. Evol, 27(3), 211. https://doi.org/10.1186/1297-9686-27-3-211

Knap P. W., and Su G. (2008) Animal, 2(12), 1742–1747. https://doi.org/10.1017/S1751731108003145

Misztal I., Tsuruta S., Strabel T., Auvray B., Druet T., *et al.* (2002). 7th World Congress on Genetics Applied to Livestock Production, Montpellier, France.

Ribeiro S., Eler J. P., Pedrosa V. B., Rosa G. J. M., Ferraz J. B. S., *et al.* (2015) Livestock Science, 176, 40–46. https://doi.org/10.1016/j.livsci.2015.03.032

Santana M. L., Eler J. P., Cardoso F. F., Albuquerque L. G., and Ferraz J. B. S. (2013) Animal, 7(2), 202–210. https://doi.org/10.1017/S1751731112001711

Su G., Madsen P., Lund M. S., Sorensen D., Korsgaard I. R., *et al.* (2006). Journal of Animal Science, 84, 1651–1657. https://doi.org/10.2527/jas.2005-517

653. Investigation on inheritance transmission modes using breeding values estimated from genomic or pedigree information

H. Hirooka[1*], M. Nishio[2], K. Inoue[3], M. Takeda[3] and Y. Inoue[4]

[1]Kyoto University, Sakyo, 606-8502 Kyoto, Japan; [2]NARO Institute of Livestock and Grassland Science, Tsukuba, 305-0901 Ibaraki, Japan; [3]National Livestock Breeding Center, Nishigo, 961-8511, Fukushima, Japan; [4]Tottori Prefectural Livestock Research Center, Kotoura 689-2503, Tottori, Japan; hirooka.hiroyuki.3s@kyoto-u.ac

Abstract

Inheritance transmission modes of five carcass traits (carcass weight, rib eye area, rib thickness, subcutaneous fat thickness and marbling score) of Japanese Black cattle were investigated using pedigree-based breeding values and genomic breeding values based on genotypic information from triplet sources of sires, dams and offspring. The phenotypic and genotypic information of 1,908 feedlot animals and genotypic information of their parents were used in this study. The inheritance transmission modes (genetic contributions of sires and dams) were quantitatively represented by path coefficients of sires and dams estimated breeding values on offspring breeding values. A large difference of genetic contribution between sires and dams was found for marbling score (0.490 for sires and 0.681 for dams) when pedigree-based breeding values were used, indicating the maternally expressed genomic imprinting. In contrast, there was no such difference between genetic contributions (0.546 for sires and 0.568 for dams) when genomic breeding values were used.

Introduction

Japanese Black cattle (a breed of Wagyu) has been utilized for draft before 1950's but became a beef breed since 1960's. A genetic evaluation program for carcass traits of the breed have started in 1980's using BLUP method (Sasaki, 2001). In recent years, genomic selection programs using GBLUP have been implemented for carcass traits of the breed (Onogi *et al.* 2014) and the new breeding program has become a standard procedure in genetic evaluation for Japanese Black cattle. The genetic evaluation programs in livestock are generally based on the assumption that the genetic contribution of parents is equal between sires and dams. Recently, however, Okamoto *et al.* (2019) first found the existence of maternally expressed genomic imprinting for marbling score of the breed using sire-MGS model based on the theory by Blunk *et al.* (2017). Subsequently, Inoue *et al.* (2020,2021) also reported the maternally expressed genomic imprinting for the marbling score using sire-dam gametic relationship matrix and sire-MGS model for different Japanese Black populations. As the next step, such findings should be assessed using genomic information. In this study, 'trio' genotyped data which consist of genotypes from sire-dam-offspring triplets were used for the purpose. The objective of this study was to investigate the inheritance transmission modes on carcass traits in Japanese Black cattle using pedigree-based breeding values and genomic breeding values estimated from the trio genotyped data.

Materials & methods

Data. The phenotypic data set in this study were obtained from the carcass records of 1,908 Japanese Black feedlot steers and heifers with 8,312 pedigree ancestors reared in National Breeding Centre and Tottori Prefecture, Japan. The carcass traits were carcass weight, rib eye area, rib thickness, subcutaneous fat thickness and marbling score (scores 1-12, with 12 as the most abundant). The carcass traits were evaluated between the sixth and seventh rib in accordance with the Japan Meat Grading Association (JMGA, 1988). The feedlot animals and their sires and dams were genotyped using the GeneSeek GGP LD v4 SNP chip (Neogen GeneSeek Operations, NE, USA). The SNP markers on autosomes were pruned with call rate (>0.95), minor allele frequency (>0.01) and Hardy-Weinberg equilibrium ($<10^{-4}$). The SNP genotype data

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_653, © H. Hirooka *et al.* 2022

were imputed to the Illumina BovineSNP50 v2 BeadChip (Illumina, CA, USA) and then were phased by using Beagle ver. 4.0 (Browning and Browning, 2007) with pedigree information to distinguish the parental origin of the alleles. Finally, 34,481 SNPs on the 29 cattle autosomes were used in this study. The genotypic data were characterized as 'trio' data consisting of genotypes from sire-dam-offspring triplets.

Statistical models. The pedigree-based breeding values and genomic breeding values were estimated by BLUP and GBLUP methods, respectively, using BLUPF90 package (Misztal 2015). In this study, the GBLUP method without pedigree information instead of single step GBLUP method (Misztal *et al.*, 2009) was used to make clear the differences of utility of pedigree information and genomic information in estimating breeding values and inferring genetic transmission modes of sires and dams. The fixed effects were the sex of animals, slaughter year, slaughter month and linear and quadratic covariates for age of slaughter. The differences between pedigree-based and genomic breeding values were evaluated with Pearson's correlation coefficient.

In this study, the inheritance transmission modes (i.e. genetic contributions of sires and dams to their sons and daughters) were quantitatively represented based on path coefficients (Wright 1934, 1960). The model considered was:

$$EBV_o = b_s \times EBV_s + b_d \times EBV_d + \phi_o \quad (1)$$

where EBV_o, EBV_s, EBV_d are estimated breeding values for feedlot animals (offspring), sire and dam, respectively, estimated from pedigree or genomic information and b_s and b_d are the partial regression of offspring on sire and dam, respectively. As noted by Cantet and Vitezica (2014), the error term ϕ_o represents the deviation from the mid parental estimated breeding values and originates in random Mendelian sampling of the grand-parental gametes present in the parents' segregation.

The path coefficient (standardized regression coefficients) for sires (β_s) and dams (β_d) were calculated as:

$$\beta_s = b_s \times \sigma_s / \sigma_o \quad (2)$$

$$\beta_d = b_d \times \sigma_d / \sigma_o \quad (3)$$

where σ_o, σ_s and σ_d are the standard deviation of pedigree-based or genomic breeding values for feedlot animals, sire and dams, respectively. The path coefficients represent the direct effects of sires or dams on the breeding values of the offspring animals.

Results & discussion

The estimated breeding values from pedigree and genomic information of sires, dams and offspring are shown in Table 1. The higher breeding values of offspring over average of sires and dams for all traits except subcutaneous fat thickness may result from genetic improvement in one generation toward desirable direction for the carcass traits by breeding programs in Japan. Furthermore, it should be noticed that the breeding values of carcass weight of dams were lager that those of sires derived from both procedures, indicating higher genetic improvement of carcass weight for dams (female side) than sires (male side). In Japan, carcass traits have been evaluated for dams as well as sires using breeding values estimated from pedigree-based BLUP method, although dams did not have phenotypic records for carcass traits. Historically, selection for sires have been more emphasized for marbling score than other carcass traits through progeny testing programs.

Table 1. Means of genomic and pedigree estimated breeding values for carcass traits.

	Genomic breeding value			Pedigree-based breeding value		
	Sire	Dam	Offspring	Sire	Dam	Offspring
No. animals	234	1,509	1,908	234	1,509	1,908
Carcass weight (kg)	-8.396	1.692	0.022	-0.716	5.369	6.134
Rib eye area (cm^2)	-1.149	-1.459	-0.007	2.112	1.786	3.338
Rib thickness (cm)	-0.105	-0.020	0.000	0.151	0.196	0.240
Subcutaneous fat thickness (cm)	0.004	0.039	0.000	0.043	0.068	0.036
Marbling score	-0.283	-0.513	0.002	1.259	1.084	1.587

The correlation between pedigree-based breeding values and genomic breeding values can compare performance of genomic prediction with pedigree-based prediction. The correlations between pedigree-based and genomic breeding values (Table 2) were largest (0.895 to 0.927) for offspring with phenotypic records, followed by sires (0.893 to 0.911) and dams (0.771 to 0.839). The result indicates that ranking of dams based on pedigree-based breeding values differed from those on genomic breeding values. In general, it is suggested that genomic breeding values of dams were much more reliable than pedigree-based ones especially when dams were genotyped. On the other hand, it should be noted that genomic prediction did not more reliable and advantageous for sires than pedigree-based prediction, when sires have many progeny and the breeding values were estimated with high reliabilities.

The path coefficients from sires and dams to offspring for the carcass traits are given in Table 3. From theoretical point of view, as an offspring receive half of the genetic materials from each parent, genetic contribution of each parent should be equal according to Mendelian law of segregation. In fact, the genetic contributions of sires and dams were almost equal for all traits except marbling score. However, for pedigree-based breeding values, the different genetic contributions between sires and dams were found for marbling score (0.490 for sires and 0.681or dams). This result may indicate evidence of existence of maternally expressed genomic imprinting for marbling score of Japanese Black cattle and is well in agreement with previous studies (Okamoto *et al.*, 2019; Inoue *et al.*, 2020, 2021). In pedigree-based analysis, the consistent results were obtained for marbling score of Japanese Black cattle regardless of different populations in Shimane Prefecture (Okamoto *et al.,* 2019), Yamagata Prefecture (Inoue *et al.*, 2020) and Tottori Prefecture (Inoue *et al.*, 2021). In contrast, however, there was no difference of genetic contributions for marbling scores when using genomic estimated breeding values in the present study.

Table 2. Correlations of genomic and pedigree-based estimated breeding values for sires, dams and progenies.

	Sire	Dam	Offspring
Carcass weight (kg)	0.911	0.789	0.927
Rib eye area (cm^2)	0.844	0.777	0.924
Rib thickness (cm)	0.877	0.794	0.895
Subcutaneous fat thickness (cm)	0.897	0.771	0.913
Marbling score	0.893	0.839	0.924

Table 3. Path coefficients (standardized regression coefficients) from sire or dam to offspring for carcass traits.

	Genomic breeding value		Pedigree-based breeding value	
	Sire	Dam	Sire	Dam
Carcass weight (kg)	0.567	0.549	0.674	0.645
Rib eye area (cm^2)	0.571	0.532	0.598	0.618
Rib thickness (cm)	0.607	0.565	0.616	0.650
Subcutaneous fat thickness (cm)	0.616	0.529	0.644	0.633
Marbling score	0.546	0.568	0.490	0.681

The 'missing' genomic imprinting effects in using genomic breeding values may be caused by the two possibilities: (1) the spurious effects resulted from ambiguous genetic transmission modes assumed in pedigree-based analysis; and (2) rigid assumption of additive effects of each SNP in genomic analysis. In our companion paper, Inoue *et al.* (this proceeding) also reported small and negligible imprinting variance for marbling score of Japanese Black cattle in genomic analysis in spite of the fact that large imprinting variance were inferred in pedigree-based analysis. Further research would be required to explain the results.

References

Blunk I., Mayer M., Hamann H., and Reinsch N. (2017) J. Anim, Sci. 95:559-571. https://doi.org/10.2527/jas.2016.0997

Browning S.R., and Browning B.L. (2007) Am. J. Hum. Genet. 81: 1084–97. https://doi.org/10.1086/521987

Cantet, R.J.C., and Vitezica, Z.G. (2014) Proc of the 10th WCGALP, Vancouver, Canada.

Inoue K., Inoue Y., Oe, T., and Nishimura M. (2021) Anim. Sci. J. 92: e13504. https://doi.org/10.1111/asj.13504

Inoue K., Nishio M., Shoji N., and Hirooka H. (2020) J. Anim. Breed. Genet 138(1):45-55. https://doi.org/10.1111/jbg.12493

Japan Meat Grading Association. (1988) Japan Meat Grading Association, Tokyo, Japan.

Misztal I., Legarra A., and Aguilar I. (2009) J. Dairy Sci. 92:4648-4655. https://doi.org/10.3168/jds.2009-2064

Misztal I., Tsuruta S., Lourenco D.A., Masuda Y., Aguilar I., *et al.* (2015) Manual for BLUP90 family programs. Retrieved from http://nce.ads.uga.edu/wiki/doku.php?id=documentation

Okamoto K., Oishi K., Nakamura R., Abe A., Inoue K., and Hirooka H. (2019) J. Anim. Breed. Genet. 136:190-198. https://doi.org/10.1111/jbg.12382

Onogi A., Ogino A., Komatsu T., Shoji N, Simizu K. *et al.* (2014) J. Anim. Sci. 92(5):1931-1938. https://doi.org/10.2527/jas.2014-7168

Sasaki Y. (2001) Asian-Australian J. Anim. Sci. 14(Special issue):111-122.

Wright S. (1934) Annals Math. Stat. 5:161-215. https://www.jstor.org/stable/2957502

Wright S. (1960) Biometrics 16:189-202. https://doi.org/10.2307/2527551

654. Genetic parameters for fatty acid traits of beef in Australian Angus

C.J. Duff[1,2*], A.M. Samaraweera[1], P. McGilchrist[2] and S.A. Clark[2]

[1]Angus Australia, 86 Glen Innes Road, Armidale, New South Wales, Australia; [2]School of Environmental and Rural Science, University of New England, Armidale, New South Wales, Australia; christian@angusaustralia.com.au

Abstract

Preliminary genetic and phenotypic parameters were estimated for fatty acid traits in feedlot finished Australian Angus cattle. This included the relationships between saturated fatty acids (SAT), monounsaturated fatty acids (MONO) and polyunsaturated fatty acids (PUFA), with carcase weight (CWT), carcase rump fat (CP8) and carcase intramuscular fat (CIMF). All fatty acid traits showed phenotypic variation, with a considerable portion attributed to genetics as shown by heritabilities for SAT, MONO and PUFA of 0.64±0.11, 0.67±0.12 and 0.25±0.09 respectively. SAT had moderate negative genetic correlation with CWT, CP8 and CIMF (-0.30±0.16, -0.21±0.17 and -0.38±0.14 respectively), while MONO and PUFA have moderate positive genetic correlation with CWT, CP8 and CIMF ranging from 0.14±0.17 to 0.46±0.22. The results show that the proportion of MONO and PUFA can be increased, and SAT decreased, in intramuscular fat through selection. Importantly, this can be achieved without being detrimental to carcase weight, carcase rump fat or total intramuscular fat.

Introduction

Beef delivers a valuable source of nutrition to the consumer by providing macronutrients such as fat and protein, vitamins such as B12 and minerals such as iron and zinc. However, there is often a focus in society on the role of fat in the diet, particularly from red meat sources like beef. Fat from beef cattle has two main forms – saturated and unsaturated fats. It is well understood that any increase in saturated fats in the diet is associated with detrimental effects on consumer health such as chronic heart disease. Conversely, some unsaturated fatty acids (e.g. Linoleic and Linolenic acid) are essential in the diet and have potential positive effects on the aetiology of chronic disorders (National Health and Medical Research Council, 2006).

To understand the potential role of selection on fatty acid composition in beef cattle, several previous studies have generated genetic parameters for fatty acid traits (Malau-Aduli *et al.*, 2000; Inoue *et al.*, 2011; Saatchi *et al.*, 2013; Kelly *et al.*, 2012). However, few studies have examined the relationship between meat fatty acid composition traits to eating quality traits such as intramuscular fat (IMF) and quantity traits such as carcase weight, particularly within Australian Angus cattle.

Therefore, the purpose of this study was to estimate preliminary phenotypic and genetic parameters for fatty acid traits that are prevalent in beef products, and to determine their relationship with carcase weight (CWT), carcase rump fat (CP8) and intramuscular fat (CIMF) in Australian Angus cattle.

Materials & methods

All phenotypes, associated fixed effects and pedigree data used in this study were generated from the Angus Sire Benchmarking Program (ASBP) as described by Parnell *et al.* (2019). The animals involved (n=927) were born in the 2015, 2018 and 2019, and consisted of straight-bred steers (n=810) or heifers (n=117) that are progeny of Angus sires (n=129) and Angus dams (n=847) from seven co-operator herds located in New South Wales, Australia. Of the dams included in this study, the majority had a single progeny represented, while 71 had two progeny. Nine animals with phenotypes did not have a dam recorded.

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_654, © C.J. Duff *et al.* 2022

All animals undertook an initial feedlot phase of approximately 100 days at the University of New England Tullimba Feedlot, Kingstown, NSW, Australia. From feedlot entry, all animals were fed a conditioning ration for approximately 21 days, following which they had *ad libitum* access to a ration composed of 74.8% tempered barley, 4.6% cotton hulls, 6% cottonseed, 5% mill run, 4.6% chopped hay and 5% liquid mineral supplement. This ration being a common formulation used in Australian feedlots with Angus cattle. Following the Tullimba Feedlot feeding phase, around half the animals (n=440) were slaughtered, based on herd of origin, to reflect the Australian mid-fed feeding program. The remaining animals (n=487) were relocated to Rangers Valley Feedlot, Rangers Valley, NSW for a second phase (i.e. further 170 days on feed) to reflect the Australian long-fed feeding program. Here they were fed a similar ration as in phase one, under a normal, controlled commercial feeding program. The mean age at slaughter was 651.3 (SD 66.0) days for all animals, with the long-fed animals being older than the mid-fed animals at 700.7 (SD 39.1) days and 596.7 (SD 42.5) days respectively.

At the end of the feeding periods all animals were slaughtered and their carcases weighed on a hot standard carcase basis and rump fat measured at the p8 site. The following day meat samples (50 grams) were collected on chilled carcases from the *longissimus thoracis* muscle at the carcase grading site, in this case between the 12th and 13th rib. The samples were assessed for IMF using near infrared spectrophotometry (NIR), described by Perry *et al.* (2001) and seven fatty acid composition traits using the gas chromatography-mass spectrometry (GC-MS) method. The fatty acid composition traits included C14:0 (Myristic Acid), C16:0 (Palmitic Acid), C16:1 (Palmitoleic Acid), C18:0 (Stearic Acid), C18:1 (Oleic Acid), C18:2 (Linoleic Acid) and C18:3 (Linolenic Acid). Three fatty acid groups were formed by combining the saturated fats (SAT = C14:0, C16:0 and C18:0), the monounsaturated fats (MONO = C16:1, C18:1) and the polyunsaturated fats (PUFA = C18:2 and C18:3). All fatty acid traits were measured on a grams per 100 grams of dry beef product basis, which were converted to a proportion of total fatty acids on a wet beef product basis for this study.

ASReml software (Gilmour *et al.*, 2021) was used to fit the animal model to each trait to estimate parameters based on univariate and bivariate mixed model analysis using three generations of pedigree. Fixed effects fitted in all models included the contemporary group and dam age. Age at slaughter was fitted for CWT, while CWT was fitted as a co-variate for CIMF, CP8 and the fatty acid traits. The contemporary group included herd, year of birth, sex, birth type (twin vs single), breeder-defined management group and slaughter date. Feeding period in the feed lot was allocated based on herd of origin therefore accounted for in the contemporary group formation. Heritabilities, as well as phenotypic and genetic correlations, were calculated from the resulting variance components.

Results

The characteristics of the traits included in this study, their heritability estimates and phenotypic variances are listed in Table 1, with the genetic and phenotypic correlations given in Table 2. The results, particularly genetic correlations, should be considered preliminary due to the large standard errors. The mean proportion of the individual fatty acid composition traits ranged from 0.22% (Linolenic acid) to 39.2% (Palmitic acid). SAT accounted for the highest proportion (mean of 58.9%) of total fatty acids, ranging from 48.9 to 67.8%. MONO was the next most prevalent (mean of 37.3%), ranging from 27.7 to 47.1%. PUFA had the lowest concentration (mean of 3.8%), ranging from 1.1 to 8.4%.

Table 1. Descriptive statistics, heritability (h^2 with standard error in parenthesis) and phenotypic variance (σ^2_p) for carcase weight, rump fat, intramuscular fat and fatty acid traits.

Trait[1]	n	Mean	Min	Max	SD	CV %	h^2	σ^2_p
CWT (kg)	926	384.3	244.0	585.0	68.5	17.8	0.54 (0.12)	781.800
CP8 (mm)	924	20.0	7.0	40.0	6.9	34.6	0.48 (0.12)	27.066
CIMF (%)	911	9.3	2.0	29.0	4.5	49.0	0.81 (0.13)	7.491
C14:0 (%)	903	3.0	1.0	5.9	0.9	30.1	0.59 (0.11)	0.435
C16:0 (%)	907	39.2	25.3	52.1	4.7	11.9	0.36 (0.11)	8.562
C16:1 (%)	905	3.3	1.7	5.1	0.6	16.9	0.40 (0.10)	0.264
C18:0 (%)	899	16.5	9.6	24.3	2.1	12.7	0.62 (0.12)	3.379
C18:1 (%)	905	34.0	23.9	44.1	3.4	9.9	0.27 (0.09)	4.311
C18:2 (%)	898	3.6	0.3	8.2	1.3	35.6	0.33 (0.10)	1.004
C18:3 (%)	826	0.22	0.02	0.55	0.08	36.0	0.32 (0.10)	0.002
SAT (%)	875	58.9	48.9	67.8	3.7	6.3	0.64 (0.11)	5.585
MONO (%)	875	37.3	27.7	47.1	3.6	9.5	0.67 (0.12)	4.423
PUFA (%)	875	3.8	1.1	8.4	1.3	33.1	0.25 (0.09)	0.957

[1] CWT = Carcase weight, CP8 = Rump fat, CIMF = Carcase intramuscular fat, C14:0 = Myristic acid, C16:0 = Palmitic acid, C16:1 = Palmitoleic acid, C18:0 = Stearic acid, C18:1 = Oleic acid, C18:2 = Linoleic acid, C18:3 = Linolenic acid, SAT = Saturated fats (C18:0+C16:0+C14:0), MONO = Monounsaturated fats (C16:1+C18:1), PUFA = Polyunsaturated fats (C18:2+C18:3). n = Number of records, SD = Standard Deviation, CV % = Coefficient of variation.

Table 2. Genetic (above diagonal) and phenotypic (below diagonal) correlations for carcase weight, carcase rump fat, intramuscular fat and fatty acid groups (standard error in parenthesis).

Trait	CWT	CP8	CIMF	SAT	MONO	PUFA
CWT	-	0.51 (0.15)	0.02 (0.16)	-0.30 (0.16)	0.32 (0.16)	0.27 (0.22)
CP8	0.26 (0.04)	-	0.04 (0.16)	-0.21 (0.17)	0.14 (0.17)	0.46 (0.22)
CIMF	0.10 (0.05)	0.02 (0.04)	-	-0.38 (0.14)	0.41 (0.14)	0.17 (0.20)
SAT	-0.05 (0.05)	0.01 (0.04)	-0.08 (0.04)	-	-0.96 (0.17)	-0.67 (0.15)
MONO	-0.01 (0.05)	-0.02 (0.04)	0.10 (0.04)	-0.90 (0.01)	-	0.44 (0.22)
PUFA	0.16 (0.04)	0.05 (0.04)	-0.02 (0.04)	-0.49 (0.03)	0.09 (0.04)	-

All traits showed phenotypic variation, with a considerable portion attributed to genetics. The heritabilities for SAT, MONO and PUFA were 0.64, 0.67 and 0.25 respectively, while the heritabilities for the individual fatty acid traits ranged from 0.27 (Oleic acid) to 0.62 (Stearic acid). The genetic correlations between the fatty acid composition traits to the production traits show contrasting results. SAT has moderate negative genetic correlation with CWT, CP8 and CIMF (-0.30, -0.21 and -0.38 respectively), while MONO and PUFA have moderate positive genetic correlation with CWT, CP8 and CIMF (ranging from 0.14 to 0.46). The genetic correlations between the fatty acid composition traits followed expectations, with strong negative genetic correlations between SAT to MONO and PUFA at -0.96 and -0.67 respectively. The present study also showed a moderate positive genetic correlation between MONO and PUFA of 0.44.

Discussion

The preliminary genetic parameters from this study indicates that selection can be applied in Australian Angus beef cattle breeding programs, with a feedlot finishing component, to change the proportion of SAT to healthier unsaturated fatty acids (MONO and PUFA) within the fatty acid profile of IMF.

Like previous studies (Malau-Aduli *et al.*, 2000; Inoue *et al.*, 2011; Saatchi *et al.*, 2013; Kelly *et al.*, 2012), a portion of the phenotypic variation observed in fatty acid traits can be attributed to genetics. However, the heritability estimates across the fatty acid traits were generally higher from this study (0.25 to 0.67) compared to those reported by Kelly *et al.* (2012), Malau-Aduli *et al.* (2000) and Saatchi *et al.* (2013) which ranged from 0.05 to 0.54, 0.05 to 0.27 and 0.06 to 0.55 respectively. In contrast, they were lower than those reported by Inoue *et al.* (2011) which ranged from 0.34 to 0.86. The differences observed across studies probably being a function of variation in dataset size and structure (764 to 2,110 animals), production systems (length of feeding or ration composition) and breed composition (Angus, Wagyu or multibreed).

The results from the present study also show that the proportion of MONO and PUFA can be increased in IMF through selection, without being detrimental to CWT, CP8 or CIMF. Conversely, the results also indicate that selection for heavier CWT, more CIMF or more CP8 will indirectly increase the proportions of MONO and PUFA, and beneficially reduce the proportion of SAT. These findings are consistent with Kelly *et al.* (2012) who found a negative genetic correlation between carcase rump fat (P8) to the saturated fatty acids (ranging from -0.19 to -0.46) and a positive genetic correlation to the monounsaturated fatty acids (ranging from 0.25 to 0.56). Malau-Aduli *et al.* (2000) also found similar results for marbling (which is highly related to IMF), with a moderate negative genetic correlation to saturated fatty acids (-0.42) and moderate positive genetic correlation to monounsaturated fatty acids (0.45).

To further develop breeding program advice on selection for fatty acid composition in beef, research is required to understand their genetic relationships to a range of selection traits including calving ease, live weight, fertility, temperament, feed efficiency, immune competence and carcase traits, including consumer eating preference traits such as tenderness, juiciness and flavour. Through the ASBP and industry collaborations more fatty acid phenotypes are being collected to confirm the results of this study and reduce the associated uncertainty as indicated by the large standard errors for the estimated genetic parameters.

Acknowledgements

The authors thank Meat and Livestock Australia (MLA) for providing co-funding support for the ASBP (PSH.0528 and P.PSH.1172). Tyson Burey and Xuemei Han from the University of New England are also gratefully acknowledged for the provision of fatty acid composition data.

References

Gilmour A. R., Gogel B. J., Cullis B. R., Welham S. J., and Thompson R (2021). ASReml User Guide Release 4.1 Functional Specification. Available at: https://asreml.kb.vsni.co.uk/wp-content/uploads/sites/3/ASReml-4.2-Functional-Specification.pdf

Inoue K., Shoji N., Honda T., and Oyama K. (2017). Anim Sci J, 88: 11– 18. https://doi.org/10.1111/asj.12613

Kelly M. J., Tume R. K., Newman S., and Thompson J. M. (2012). Anim. Prod. Sci. 53: 129-133. https://doi.org/10.1071/AN12154

Malau-Aduli E. M., Siebert B., Bottema C. D., and Pitchford W. (2000). J. Anim. Phys. and Anim. Nutrit. *83*(2): 95–105. https://onlinelibrary.wiley.com/doi/abs/10.1046/j.1439-0396.2000.00254.x

National Health and Medical Research Council, Australian Government Department of Health and Ageing, New Zealand Ministry of Health. Nutrient Reference Values for Australia and New Zealand. (2006). Available at: https://www.nhmrc.gov.au/sites/default/files/images/nutrient-refererence-dietary-intakes.pdf

Parnell P. F., Duff C. J., Byrne A. I., and Butcher N. M., (2019). Proc. Assoc. Adv. Anim. Breed. Genet. 23:492-495

Perry D., Shorthose, W. R. Ferguson, D. M., and Thompson J. M. (2001). Ani. Prod. Sci.41: 953-957. https://doi.org/10.1071/EA00092

Saatchi M., Garrick D.J., Tait R.G., Mayes M.S., Drewnoski M., *et al.* BMC Genom 14: 730 (2013). https://doi.org/10.1186/1471-2164-14-730

655. Integration of beef cattle international estimated breeding values in the Italian evaluation

R. Bonifazi[1*], M.P.L. Calus[1], J. ten Napel[1], R.F. Veerkamp[1], S. Biffani[2], M. Cassandro[3,4], S. Savoia[5] and J. Vandenplas[1]

[1]Animal Breeding and Genomics, Wageningen University & Research, P.O. Box 338, 6700 AH Wageningen, the Netherlands; [2]Istituto Di Biologia E Biotecnologia Agraria, Consiglio Nazionale Delle Ricerche, Via Edoardo Bassini 15, 20133 Milano, Italy; [3]Department of Agronomy, Food, Natural resources, Animals and Environment, University of Padova, Viale dell'Università 16, 35020 Legnaro, Italy; [4]National Federation of National Breeders Associations (FedANA), XXIV Maggio 43, 00187 Roma, Italy; [5]Interbull Centre, Department of Animal Breeding and Genetics, SLU, P.O. Box 7023, 75007 Uppsala, Sweden; renzo.bonifazi@wur.nl

Abstract

Beef cattle international evaluations, led by Interbeef, combine data from different countries to compute international estimated breeding values (EBV_{INT}) which are expressed on the same country scale as national ones, allowing participating countries to compare national and foreign sires. These two EBV might be in disagreement due to differences in the information used in national and international evaluations. This can be overcome by integrating the EBV_{INT} into the national evaluations to obtain a 'blended' EBV. Here we tested and validated a general procedure to integrate publishable sires EBV_{INT} at the national level. Using evaluations for Limousin weaning weight and the integration for Italy as a case study, we show that the integration procedure accurately blends international information at the national level while avoiding double-counting of national information which are used as input to compute EBV_{INT} in Interbeef evaluations. This procedure can also be used with other traits and breeds.

Introduction

International evaluations allow comparing estimated breeding values (EBV) of foreign elite animals with national animals by computing international EBV (EBV_{INT}) and associated reliabilities (REL_{INT}). EBV_{INT} usually would have higher reliabilities than national EBV (EBV_{NAT}) as international evaluations take into account data from relatives recorded in other countries (Bonifazi *et al.*, 2020b). In beef cattle international evaluations as performed by Interbeef, each participating country receives a list of EBV_{INT} and REL_{INT} for: (1) all animals in the national pedigree; and (2) so-called 'publishable sires', i.e. sires that meet Interbeef publication rules, as described in Bonifazi *et al.* (2021). Thus, an individual could have two EBV at the national level: EBV_{NAT} and EBV_{INT}, which can differ due to differences in information between national and international evaluations (e.g. national evaluations are usually multi-trait and in some cases multi-breed while international evaluations are trait-specific and performed within-breed). To overcome this issue, the EBV_{INT} and its associated measure of precision (e.g. REL_{INT}) can be integrated into the national evaluations (Vandenplas and Gengler, 2015), resulting in a single 'blended' EBV that includes and propagates international information to all animals in the national evaluation. However, double-counting of national information needs to be removed from the EBV_{INT} as Interbeef uses national phenotypes as input for international evaluations. Thus, the objectives of this study were to test and validate the accuracy of a general procedure to integrate publishable sires' EBV_{INT} at the national level using Limousin weaning weight and Italy (ITA) as a case study for national evaluations.

Materials & methods

Data and software. A total of 441,691 Limousin males and females individual age-adjusted weaning weights (AWW) were available. AWW were recorded in eight different countries for a total of six populations: Denmark, Finland and Sweden (modelled as one population), Czech Republic, Ireland,

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_655, © R. Bonifazi *et al.* 2022

Germany, Switzerland, and Italy. A total of 7,506 AWW (of which 2,304 from ITA) were recorded from animals born between January and May 2019. The international pedigree included 683,317 animals. All analyses were performed using MiXBLUP (ten Napel *et al.*, 2020): the convergence criterion was 'CD≤10^{-5}, and reliabilities were computed using the Tier and Meyer (2004) algorithm.

International and national evaluations. The direct and maternal EBV_{INT} and REL_{INT} were estimated using the Interbeef model which is equivalent to a multi-trait animal model with maternal effects where each population national model was fitted as a different correlated trait (see Bonifazi *et al.* (2020a) for a complete description). Direct and maternal genetic effects were modelled using an across-populations genetic (co)variance matrix following the Interbeef procedure as described in Bonifazi *et al.* (2021). Random environmental, maternal permanent environmental, and residual effects were modelled using block-diagonal variance matrices. Both genetic and environmental variances were the same as those used in the national genetic evaluations of participating countries. The direct and maternal EBV_{NAT} and REL_{NAT} for ITA were obtained by running a single-trait evaluation using only ITA submitted phenotypes and the same national model as the one used for the international evaluations.

Integration. After EBV_{INT} and EBV_{NAT} and their associated REL (i.e. REL_{INT} and REL_{NAT}) on a 0 to 1 scale were obtained following the aforementioned models, publishable sires' international information was integrated into national evaluations in four steps as follows.

Step 1. For each publishable sire *i*, direct and maternal effective record contributions (ERC) associated with REL_{NAT} and REL_{INT} (ERC_{NAT} and ERC_{INT}, respectively) were computed as: $ERC_i = \lambda \cdot REL_i/(1-REL_i)$, where $\lambda = \sigma^2_{residual} / \sigma^2_{genetic}$ with $\sigma^2_{genetic}$ being either the direct or maternal genetic variance.

Step 2. For each publishable sire *i*, direct and maternal de-regressed proofs for both national and international EBV (DRP_{NAT} and DRP_{INT}, respectively) were computed following Garrick *et al.* (2009): $DRP_i = PA_i + (EBV_i - PA_i)/ REL_{i\,(o+p)}$, where PA_i is the parent average EBV of the individual computed as $(EBV_{sire} + EBV_{dam})/2$, and $REL_{i\,(o+p)}$ is the reliability due to the individual own performance (o) and its progeny (p) computed as $dERC_i / (dERC_i + \lambda)$. The de-regressed ERC for the individual *i*, $dERC_i$, is computed as $ERC_i - ERC_{PA}$, with ERC_{PA} being the ERC calculated from parent average reliability defined as $(REL_{sire} + REL_{dam})/4$. If the $dERC_i$ was ≤0, both the $dERC_i$ and its associated DRP_i were set to 0.

Step 3. For each publishable sire *i*, to avoid double-counting of national information, the direct and maternal adjusted DRP (DRP*) and its associated weight (dERC*) adjusted for national data were computed based on Vandenplas *et al.* (2014): $DRP^*_i = ((dERC_{INTi} \cdot DRP_{INTi}) - (dERC_{NATi} \cdot DRP_{NATi}))/ dERC^*_i$, where $dERC^*_i = dERC_{INTi} - dERC_{NATi}$. If $dERC^*_i$ was ≤0 or if the gain in reliability (defined as the difference between REL_{INT} and REL_{NAT}) was smaller than 0.01, both the $dERC^*_i$ and its associated DRP^*_i were set to 0.

Step 4. The direct and maternal blended EBV were computed with a national evaluation using national phenotypes and publishable sires' direct and maternal DRP* as pseudo-phenotypes. The dERC* were used as weights for the DRP*. The maternal DRP* were associated with the maternal effect of the animal itself and not of its dam. The direct and maternal DRP* were modelled as additional phenotypes for the same trait and fitted using one general mean for each DRP*.

Scenarios and validation. We applied the above procedure for ITA where national and international data differ due to differences in publication calendars. We integrated EBV_{INT} from an international evaluation in January 2019 into a national evaluation in May 2019. All analyses used the complete international pedigree. Hereafter, we will refer to ITA phenotypes as 'national', and to phenotypes from other countries

as 'foreign'. First, the two following evaluations were performed and their outputs were used as inputs for the integration procedure:

- INT_{JAN}. An international evaluation using both national and foreign phenotypes of animals born up to January 2019. Publishable sires to be integrated are selected from this evaluation.
- NAT_{JAN}. A national evaluation using only national phenotypes of animals born up to January 2019. EBV and REL from this evaluation are used during the integration procedure to avoid double-counting of national information.

Then, three scenarios were implemented and validated:

- NAT_{MAY}. A national evaluation using only national phenotypes of animals born up to May 2019. This scenario is used for comparison with $BLEND_{MAY}$ and GOLD to illustrate the benefits of the integration procedure over a national evaluation without any integration.
- $BLEND_{MAY}$. A blended national evaluation using national phenotypes of animals born up to May 2019 with the integration of publishable sires' information from INT_{JAN}.
- GOLD. As $BLEND_{MAY}$ but integrating information of publishable sires from an international evaluation using only foreign phenotypes of animals born up to January 2019. This international evaluation avoids double-counting of national information when integrating publishable sires' information in GOLD, providing a 'gold standard' for comparison with $BLEND_{MAY}$.

We validated the integration procedure by comparing publishable sires' EBV obtained with either NAT_{MAY}, $BLEND_{MAY}$, or GOLD with their EBV from a reference scenario (REF) that is an international evaluation using national phenotypes of animals born up to May 2019 and foreign phenotypes of animals born up to January 2019. Scenario REF uses the same sources of information as in $BLEND_{MAY}$ and GOLD, but without approximating foreign information by DRP* and dERC*. Therefore, the more accurate the integration procedure, the closer the EBV are to those of REF. As validation metrics, we computed the Pearson correlations (ρ) between EBV, and the slopes (b_1) of the regression of REF EBV on the EBV of the validated scenarios. Publishable sires were divided into three groups: (A) domestic publishable sires with ≥15 recorded offspring in ITA; (B) domestic publishable sires with <15 recorded offspring in ITA; and (C) foreign publishable sires with no recorded offspring in ITA.

Results

The number of publishable sires in groups A, B and C was 1,382, 94 and 3,470 for direct EBV, and 491, 51, and 1,165 for maternal EBV, respectively. The mean dERC* of all publishable sires and the number of dERC*>0 were similar between $BLEND_{MAY}$ and GOLD (Table 1). As expected, for direct EBV, group A of publishable sires had a smaller amount of international information to be integrated (mean dERC* equal to 0.5 for $BLEND_{MAY}$) compared to sires in groups B and C (mean dERC* equal to 5.2 and 2.5, respectively, for $BLEND_{MAY}$) (Table 1). For all groups, direct EBV ρ and b_1 improved in both $BLEND_{MAY}$ (0.98 and 1.02, respectively, on average across groups) and GOLD (0.98 and 1.06 on average) compared to NAT_{MAY} (0.60 and 0.77 on average) (Table 1). For maternal EBV, $BLEND_{MAY}$ and GOLD resulted in similar ρ and b_1 as NAT_{MAY} for all groups except for foreign sires (Table 1). For foreign sires, $BLEND_{MAY}$ and GOLD had ρ closer to 1 but worse b_1 compared to NAT_{MAY}. Overall, $BLEND_{MAY}$ gave similar ρ and similar or slightly worse b_1 than GOLD: on average across groups and effects, ρ and b_1 were 0.93 and 0.88 for $BLEND_{MAY}$, and 0.93 and 0.94 for GOLD.

Table 1. Validation results for direct and maternal EBV of publishable sires.[1] All scenarios are validated against EBV from scenario REF.[2]

Sire group	Scenario[2]	Direct EBV				Maternal EBV			
		ρ	b_1	Mean dERC*	n. dERC* >0	ρ	b_1	Mean dERC*	n. dERC* >0
A (Domestic ≥15 off)	NAT_{MAY}	0.95	0.96	-	-	0.99	1.01	-	-
	$BLEND_{MAY}$	0.99	0.96	0.5	97	0.98	0.96	0.1	52
	GOLD	0.99	1.01	0.3	121	0.98	1.02	0.1	70
B (Domestic <15 off)	NAT_{MAY}	0.63	0.66	-	-	0.86	0.80	-	-
	$BLEND_{MAY}$	0.97	0.96	5.2	94	0.84	0.72	0.7	51
	GOLD	0.97	1.03	4.0	94	0.86	0.87	0.7	51
C (Foreign)	NAT_{MAY}	0.24	0.69	-	-	0.51	0.98	-	-
	$BLEND_{MAY}$	0.97	1.13	2.5	3,469	0.83	0.57	0.5	1,127
	GOLD	0.97	1.14	2.4	3,470	0.81	0.58	0.4	1,165

[1] ρ = Pearson correlation; b_1 = slope; dERC* = adjusted de-regressed ERC; n. = number; off = recorded offspring in ITA.

[2] REF = reference international evaluation; NAT_{MAY} = national evaluation without integration; $BLEND_{MAY}$ = blended evaluation with integration of publishable sire's information; GOLD = as $BLEND_{MAY}$ but integrating publishable sire's information computed without national data.

Discussion

Our results show that a blended evaluation is especially beneficial for sires in groups B and C which have only a few or no recorded offspring at the national level. Also, double-counting of national information during blending was properly removed as shown by the good agreement of $BLEND_{MAY}$ and GOLD. Scenario GOLD completely avoids double-counting of national information but would require Interbeef to run and distribute for each population EBV and REL from an international evaluation from which national phenotypes are removed. Instead, $BLEND_{MAY}$ can be directly applied by countries, assuming that the publishable sires' parental EBV_{INT} and REL_{INT} are known and that national information submitted for the international evaluation (here NAT_{JAN}) are available. Overall, maternal EBV showed lower ρ and b_1 compared to direct EBV, which could be due to the lower gains in REL associated with maternal EBV compared to direct EBV. The outlined integration procedure enables the integration of publishable sires' EBV_{INT} at the national level in a single blended EBV. Results show that the integration procedure yields accurate results for both direct and maternal EBV.

References

Bonifazi R., Vandenplas J., ten Napel J., Matilainen K., Veerkamp R.F., *et al.* (2020) Genet. Sel. Evol. 52(32):1-16. https://doi.org/10.1186/s12711-020-00551-9

Bonifazi R., Vandenplas J., ten Napel J., Cromie A., Veerkamp R.F., *et al.* (2020b) Acta Fytotech. Zootech. 23:144-155. https://doi.org/10.15414/afz.2020.23.mi-fpap.144-155

Bonifazi R., Vandenplas J., ten Napel J., Veerkamp R.F., and Calus M.P.L. (2021) J. Anim. Sci. 99(9):1-14. https://doi.org/10.1093/jas/skab222

Garrick D.J., Taylor J.F., and Fernando R.L. (2009) Genet. Sel. Evol. 41(55):1-8. https://doi.org/10.1186/1297-9686-41-55

Ten Napel J., Vandenplas J., Lidauer M., Stranden I., Taskinen M., *et al.* (2020) MiXBLUP, user-friendly software for large genetic evaluation systems.

Tier B., and Meyer K. (2004) J. Anim. Breed. Genet. 121(2):77-89. https://doi.org/10.1111/j.1439-0388.2003.00444.x

Vandenplas, J., Colinet, F.G., and Gengler, N. (2014) Genet. Sel. Evol. 46(59):1-15. https://doi.org/10.1186/S12711-014-0059-3

Vandenplas J., and Gengler N. (2015) Livest. Sci. 181:121-130. https://doi.org/10.1016/j.livsci.2015.09.012

656. A genomic evaluation of genotype-by-environment interactions for foot scores in American and Australian Angus cattle

A.B. Alvarenga[1*], K. Retallick[2], A. Garcia[2], H.R. Oliveira[1,3], S.P. Miller[2,4], A. Byrne[5], L.F. Brito[1*]

[1]Purdue University, 270 S Russell Street, West Lafayette, IN 4790,7 USA; [2]Angus Genetics Inc., 3201 Frederick Avenue, Saint Joseph, MO 64506, USA; [3]Lactanet, N1G 2W1, Guelph, Canada; [4]AGBU, The University of New England, 2351, Armidale, Australia; [5]Angus Australia, 86 Glen Innes Road, NSW 2350, Armidale, Australia; alvarena@purdue.edu; britol@purdue.edu

Abstract

The main goal of this study was to investigate the feasibility of an across-country genomic evaluation for claw-set and foot angle scores in North American and Australian Angus populations, by analysing within and across-country genotype-by-environment interactions (G×E). In total, 123,322 and 1.13M animals had phenotype and genotype information. Reaction norm models fitting environmental gradients based on the effects of management group and bivariate animal models were used to evaluate genetic correlation within- and across-countries, respectively. Similar heritability estimates were obtained for both North American and Australian populations (0.22-0.25), and moderate-to-high genetic correlations between foot scores (claw-set: 0.73; foot angle: 0.51) were observed between countries. On average, high genetic correlations (>0.90) were observed among pairs of environmental gradients within countries. However, moderate correlations were observed among contrasting environmental gradients (e.g. 0.50), indicating moderate G×E. In general, gains in both theoretical and prediction accuracy of genomic estimated breeding values from a joint- versus within-country genomic evaluation were observed.

Introduction

Animal soundness has been ranked within the top five traits to be selected in a breeding program based on a survey taken by American and Australian Angus stakeholders (Santos *et al.*, 2019; Bell *el al.,* 2019). Furthermore, structural problems affecting foot angle (FA) and claw set (CS) are important culling reasons, representing around 6% of animals culled from North American Angus farms (Oliveira *et al.*, 2020). Foot problems have a direct impact on animal longevity, welfare, and the profitability of the production system. Low-to-moderate heritability estimates and clinical trials highlight the influence of environmental conditions on the phenotypic expression of foot scores (Wang *et al.*, 2017; Giess *et al.*, 2021). With the globalization of germplasm, genotype-by-environment interactions (G×E; Lynch and Walsh, 1998) can be a concern. With significant G×E, cattle ranking highly in one environment may not be the best in a different environment resulting in losses of revenue and jeopardizing genetic improvement.

The overarching goal of this study was to characterize the genetic sensitivity of foot scores within and among the United States of America (USA) and Australian (AU) Angus populations to allow an optimal across-country genomic evaluation. More specifically, the main objectives were to: (1) estimate the genetic correlation between countries for two foot score traits (i.e. FA and CS); (2) evaluate G×E within-country based on the Best Linear Unbiased Estimates (BLUE) of management group effect as the environmental gradient; and (3) assess the predictive ability of genomic prediction of breeding values based on single- and across-country training populations. This study comprises the largest beef cattle dataset for foot score collected in a wide range of climatic conditions and production systems.

Materials & methods

Data. American Angus Association (Saint Joseph, USA) and Angus Australia (Armidale, Australia) provided the datasets used in this study. The scores for CS and FA, respectively were recorded on a scale

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_656, © A.B. Alvarenga *et al.* 2022

from one (i.e. spread or steep) to nine (i.e. scissor or weak), in which five is the optimal score (www.angus.org/performance/Documents/footscorebrochure.pdf). Measurements of foot scores are collected annually from yearling through older ages. The information on the worst foot at specific ages was considered as the phenotypic record of each animal. The incidence of one through four scores were very low, therefore, only scores from five to nine were included in the analyses. The quality control applied included removing records with unknown management group, management groups with fewer than three records, or groups without phenotypic variation. Finally, 123,322 (50,621 USA and 72,701 AU) CS records from 109,055 (46,864 USA and 62,191 AU) animals and 120,730 (48,442 USA and 72,288 AU) FA records from 106,333 (44,485 USA and 61,848 AU) animals remained for further analyses. The USA and AU animals were born from 2002 to 2020 and from 1993 to 2020, respectively. The animals had from one to six repeated records. In total, 1.13 million animals (996K USA and 132K AU), which were genotyped and/or imputed to a 54K marker density panel, were available for this study, including 36,786 USA and 23,023 AU animals with both phenotypic and genomic information.

Population stratification. Principal components were estimated based on the genomic relationship matrix to investigate the genomic similarities of the Angus populations. The linkage disequilibrium for each population and consistency of gametic phase among populations were also investigated. Both SNP1101 and ADMIXTURE (Alexander and Lange, 2011) were utilized for the analysis of population structure.

Statistical models. A repeatability animal model was used, which included management group and age in days (linear covariate) as fixed effects; and animal, permanent environment, and residual as random effects. The results from threshold and linear models were highly correlated (results not shown), therefore, a linear model was used in all analyses.

Genetic parameter estimation. A single-trait animal model was used to estimate the heritability for each trait (CS and FA) per country (USA and AU). A two-trait animal model was used to calculate the genetic correlation between countries for each trait separately. To evaluate the genetic correlation within countries, two scenarios were evaluated: (1) randomization of farms into two groups to be considered as different traits, and (2) a reaction norm approach, in which the effect of each management group was used as the environmental gradient. A two-trait model was used for scenario 1. For the reaction norm analyses, a similar approach as proposed by Carvalheiro *et al.* (2019) was employed. Accurate estimation of management group effect (environmental gradient) is crucial for these analyses. Therefore, additional quality control on the data was performed requiring a minimum of 20 animals per management group. The first step was to estimate the management group effects (i.e. BLUE), which were then rounded to one decimal digit, and standardized to have mean equals to zero and variance equals to one. Then, the reaction norm was defined as:

$$y_{ij} = x_j'\beta + b\hat{\theta}_i + n_{0_u} + n_{1_u}\hat{\theta}_i + n_{0_{pe}} + n_{1_{pe}}\hat{\theta}_i + e_i \quad (1)$$

Where y_{ij} is the phenotypic observation on the animal j in the environmental gradient i; β is the vector of fixed effects, and x_j' is its corresponding incidence vector; $\hat{\theta}_i$ is i environmental gradient (i.e. management groups' BLUE-standardized), and b is the overall fixed regression coefficient of y_{ij} on $\hat{\theta}_i$; n_{0_u} and n_{1_u} are the reaction norm intercept and slope of animal j regressed on $\hat{\theta}_i$ for the additive genetic random effect, while $n_{0_{pe}}$ and $n_{1_{pe}}$ refer to the permanent environment random effect; and e_i is a homogeneous residual. For all genetic parameters' estimation, a pedigree-based relationship matrix was implemented, in which four generations were traced back from the animals with phenotypic records (1.9M animals). The (co) variance components were estimated using thrgibbsf90 and airemlf90 software (Misztal *et al.*, 2014a) for the reaction norm step and remaining analyses, respectively.

Forward genomic prediction. We used a single-step GBLUP two-trait model (CS and FA) and implemented the Algorithm for Proven and Young with 22k core animals (Misztal *et al.*, 2014b) to estimate the Genomic Estimated Breeding Values (GEBV). Animals born in 2020 (4,755 animals) and 2019-2020 (4,366 animals) were considered as the validation group in the USA and AU datasets, respectively. Different scenarios were evaluated alternating the training population such as (1) USA as training and validation; (2) AU as training and validation; (3) joint genomic evaluation of USA and AU (as training and validation). The predictive performance of the scenarios was evaluated using the Legarra and Reverter (2018) method.

Results & discussion

Population structure. The USA and AU Angus populations are highly linked at the genomic level based on principal components, admixture, and consistency of gametic phase (correlation equal to 0.84 for a distance of markers up to 5.0 Mb). Similar patterns were observed among Angus populations raised in Canada and Brazil (Cardoso *et al.,* 2020).

Genetic parameters estimation. Heritability estimates are similar among countries and foot score measurement, being moderately heritable [0.22 (0.01) and 0.25 (0.01) for CS, and 0.23 (0.01) and 0.24 (0.01) for FA using USA and AU animals, respectively; values are mean (SE)]. Even with exchange of germplasm and the populations being similar based on genomic information, moderate genetic correlations were observed between countries [0.73 (0.08) for CS; 0.51 (0.12) for FA]. Some hypotheses justifying these lower correlations are: (1) differences in the trait recording (i.e. cattlemen record foot score in the USA while a certified technician is responsible for the recording in AU); (2) differences in age at recording (i.e. ~60% of USA recording happens at yearling while ~77% of recordings in AU is equally distributed up to 2 years of age); (3) subpopulation stratification or G×E due to management system that could be even observed within country; and (4) across-country G×E.

To investigate the potential source of the moderate genetic correlation between countries, farms within a country were randomly assigned to two different groups and each group treated as a different trait, similar how USA and AU were treated as different traits. Moderate-to-high genetic correlations (from 0.66 to 0.85) were observed within country by randomization of herds to separate traits. Further investigation of within-country G×E based on reaction norms considering BLUE of management groups as the environmental gradient was performed. The management groups represent sex and the location of the farm, which could potentially reflect the production environment. Figure 1 represents the heatmap of genetic correlations among all environmental gradients. High averages of genetic correlation were observed in all scenarios (from 0.91 to 0.99). However, the correlation among extreme environments varied from 0.50 to 0.94. This suggests that among extreme environmental gradients within countries, lower genetic correlations could be observed, indicating G×E.

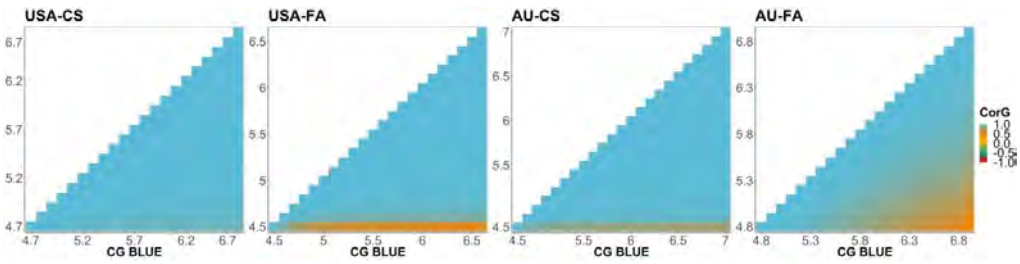


Figure 1. Estimates of genetic correlations across environmental gradients.

Forward genomic predictions. An improvement in predictive accuracy was observed while performing a joint genomic evaluation for the USA (from 0.59 to 0.60 for CS; 0.56 to 0.58 for FA). In the AU data, in which a larger training population was available, no gain in predictive accuracy was observed while adding the USA dataset into the evaluation. However, a reduction in dispersion (from 0.94 to 0.98 for FA) was observed. Small improvements in the predictive ability could be because the data is already large enough to accurately estimate GEBVs. Even though small or no improvements were observed in the overall predictive ability combining USA and AU datasets, a joint genetic evaluation is of great interest in terms of the exchange of genetic material. Furthermore, the joint genetic evaluation allows improvement in the accuracy of individual estimated breeding values compared to a within-country evaluation. Individual accuracy increased by up to 38% while performing a joint genetic evaluation.

References

Alexander D.H., and Lange K. (2011) BMC Bioinform. 12. doi.org/10.1186/1471-2105-12-246

Bell A.M., Byrne A.I., Duff C.F., and Dominik S. (2019) Proc. of the 23rd AAABG, Armidale, AU.

Cardoso D.F., Fernandes Júnior G.A., Scalez D.C.B., Alves A.A.C., Magalhães A.F.B., *et al.* (2020) Sci. Rep. 10. doi.org/10.1038/s41598-020-65565-1

Carvalheiro R., Costilla R., Neves H.H.R., Albuquerque L.G., Moore S., *et al.* (2019) Genet. Sel. Evol. 51. doi.org/10.1186/s12711-019-0470-x

Giess L.K., Jensen B.R., Bormann J.M., Rolf M.M., and Weaber R.L. (2021) J. Anim. Sci. 99. doi.org/10.1093/jas/skab256

Legarra A., and Reverter A. (2018) Genet. Sel. Evol. 50. doi.org/10.1186/s12711-018-0426-6

Lynch M., and Walsh B. (1998) Genetics and Analysis of Quantitative Traits. Sinauer Associates, Sunderland, MA.

Magrin L., Brscic M., Armato L., Contiero B., Cozzi G., *et al.* (2018) Prev. Vet. Med. 161: 83-89. doi.org/10.1016/j.prevetmed.2018.10.018

Misztal I., Legarra A., and Aguilar I. (2014a) J. Dairy Sci. 97. doi.org/10.3168/jds.2013-7752

Misztal I., Tsuruta S., Lourenco D.A.L., Masuda Y., Aguilar I., *et al.* (2014b) Manual for BLUPF90 family of programs. Available at: http://nce.ads.uga.edu/wiki/lib/exe/fetch.php?media=blupf90_all2.pdf

Oliveira H.R., Brito L.F., Miller S.P., and Schenkel F.S. (2020) Animals 10. doi.org/10.3390/ani10122410

Santos B., Archer J.A., Martin-Collado D., Quinton C.D., Crowley J., *et al.* (2019) Proc. of the 23rd AAABG, Armidale, AU.

Sargolzaei M. (2014) SNP1101 User's guide (V1.0). HiggsGene Solutions Inc. Guelph, CA.

Wang L., Miller S.P., Retallick K.J., and Moser D.W. (2017) J. Anim. Sci. 95. doi.org/10.2527/asasann.2017.202

657. Single-step GBLUP evaluation for categorical traits for beef cattle

A. Nazari-Ghadikolaei[1*], S. Eriksson[1] and W.F. Fikse[2]

[1]Swedish University of Agricultural Sciences, P.O. Box 7023, 75007 Uppsala, Sweden; [2]Växa, P.O. Box 288, 75105 Uppsala, Sweden; anahit.nazari@slu.se

Abstract

In order to study the performance of single-step genomic prediction for a categorical trait such as calving difficulty in beef cattle, data was simulated and single-step genomic breeding values (GEBV) and pedigree-based breeding values (PEBV) were estimated. Different number of trait categories with and without normal score transformation, and including different information sources, were studied. For a trait with a heritability of 0.30 and four categories, the increase in accuracy using GEBV compared with PEBV was 28.6% with, and 27.9% without normal score transformation, respectively. Moreover, in a scenario without phenotypes for males, the accuracy of GEBV was 26.6% higher than for PEBV for a binary trait with a heritability of 0.05. All results indicated that single-step genomic evaluation of categorical traits is beneficial to use in beef cattle breeding programs.

Introduction

Striking improvement of accuracies by using genomic breeding values (GEBV) has been seen in dairy cattle (Hayes *et al.*, 2009). GEBV accuracies are generally lower for beef cattle, due to smaller reference populations, fewer tested progeny and limited use of artificial insemination (Meuwissen *et al.*, 2016). Single-step genomic prediction (ssGBLUP) has been applied successfully for e.g. Angus in the U.S. (Lourenco *et al.*, 2015). The ssGBLUP that allows adding genomic, phenotype and pedigree information at the same time is well suited for beef populations (Legarra *et al.,* 2014), and was shown to improve breeding value accuracies for growth traits in simulated and real data (Lourenco *et al.*, 2013; 2015). Pedigree-based genetic evaluation of categorical traits is often done using linear models, and breeding values tend to have satisfactory properties despite ignoring the categorical nature of the observations (Meijering and Gianola, 1985). However, for single-step GBLUP evaluation experiences with categorical traits are scarcer.

In this study, we simulated a population mimicking the structure of the most common beef breeds in Sweden to compare breeding value accuracies using pedigree-based BLUP (PBLUP) and ssGBLUP in different scenarios. We studied traits with two or four categories, similar to calving ease, with and without normal score transformation and with different information sources.

Materials & methods

Simulated population. The QMsim software (Sargolzaei *et al.*, 2009) was used to create a population with genomic information, a single quantitative trait with a heritability of 0.30, and phenotypic variance of 1. True breeding values (TBVs) were calculated as the sum of polygenic and QTL effects, both effects explaining 10 and 20% of the phenotypic variation. For the genome, 750 QTL and 40K SNP were simulated across 29 autosomal chromosomes. Marker mutation rate was assumed to be zero, and neither missing marker genotypes nor marker genotyping errors were considered. In order to create LD, a historical population was created consisting of 1000 individuals in generation zero and kept constant in number for 1,020 generations. In generation 1,020, the population was decreased to 100 males and 100 females. From the last generation of the historical population, the population BeefA consisting of 100 males and 100 females was created using the option random union gamete mating system, allowing females to mate with more than one male per generation, and five offspring per dam were simulated. At this stage, the population was expanded with a factor five for each of five generations. Furthermore, population BeefB was simulated

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_657, © A. Nazari-Ghadikolaei *et al.* 2022

from the last generation of BeefA, using 200 males and 3,300 females, followed by random mating system and sire and dam replacement rate of 0.30 and 0.20, respectively. The simulation was continued for 25 generations with one offspring per dam. Finally, the genotypes and phenotypes of 86,000 individuals in population BeefB were available for further analyses.

Phenotypes. We generated an additional phenotype with a low heritability (LH) of 0.05 by adding random variation to the simulated quantitative trait. Binary traits were created with thresholds at: (1) the 0.75 quantile ($C2_{constant}$); and (2) quantile 0.70 in males and 0.75 in females in generation zero and with a slight decrease of 0.002 in each generation ($C2_{changed}$). The latter was made to mimic better calving ease for female calves that tend to be smaller compared to male calves and to mimic an environmental improvement in calving ease over time. In addition, binary traits with four categories were created with thresholds at and 0.94, 0.97, and 0.99 ($C4_{constant}$). Normal score (NS) transformation was done to reduce extreme phenotypes and generate traits closer to a normally distributed trait (Table 1).

Breeding value estimation. Different approaches for estimating breeding values were conducted using the BLUPF90 family programs (Aguilar *et al.*, 2018). Pedigree-based breeding values (PEBV) and GEBV were estimated using blupf90 for the different traits. All traits were analysed using linear models (LM), and binary traits were also analysed using threshold models (TM) in Thrgibbs1f90 with 100,000 iterations, and the first 2,000 as burn-in; results of the Gibbs sampling were summarized using postgibbsf90.

Table 1. Scenarios, traits, models, heritability (h^2) and data for breeding value estimation.

Scenario	Trait and Model type[1]	h^2†	Generations with phenotype data		Generations with genotype data	
			Males	Females	Males	Females
A	Quant LM	0.30	1-23	1-23	19-25	19-25
	$C2_{constant}$ LM	0.30	1-23	1-23	19-25	19-25
	$C2_{changed}$ LM	0.30	1-23	1-23	19-25	19-25
	$C2_{constant}$ NS LM	0.30	1-23	1-23	19-25	19-25
	$C2_{changed}$ NS LM	0.30	1-23	1-23	19-25	19-25
	$C2_{constant}$ TM	0.30	1-23	1-23	19-25	19-25
	$C2_{changed}$ TM	0.30	1-23	1-23	19-25	19-25
	$C4_{constant}$ LM	0.30	1-23	1-23	19-25	19-25
	$C4_{constant}$ NS LM	0.30	1-23	1-23	19-25	19-25
	QuantLH LM	0.05	1-23	1-23	19-25	19-25
	$C2_{constant}$ LH LM	0.05	1-23	1-23	19-25	19-25
	$C2_{changed}$ LH LM	0.05	1-23	1-23	19-25	19-25
	$C2_{constant}$ NSLH LM	0.05	1-23	1-23	19-25	19-25
	$C2_{changed}$ NSLH LM	0.05	1-23	1-23	19-25	19-25
B (a)	QuantLH LM	0.05	–	1-23	19-25	24-25
	$C2_{constant}$ LH LM	0.05	–	1-23	19-25	24-25
	$C2_{changed}$ LH LM	0.05	–	1-23	19-25	24-25
B (b)	QuantLH LM	0.05	–	1-23	24-25	19-25
	$C2_{constant}$ LH LM	0.05	–	1-23	24-25	19-25
	$C2_{changed}$ LH LM	0.05	–	1-23	24-25	19-25

[1] Quant = quantitative; C2 and C4 = categorical with 2 and 4 categories, respectively; constant and changed = refer to trait thresholds; NS = normal score transformation; LH = low h^2 (0.05). LM = linear model; TM = threshold model.

† h^2 on the underlying scale for categorical observations.

Scenarios studied. We created several different scenarios with respect to data included in the genetic/genomic evaluation (Table 1). In all cases, we removed phenotypes for animals in the last two generations (24 and 25). In Scenario A, we used all genotype information (approximate 40K SNPs) from the last seven generations; more generations could not be considered due to program limitations. In Scenario B, we include phenotypes of females only and used genotypes in two different settings: (1) for males in the last seven generations and for females in the last two generations; (2) for females in the last seven generations and for males in the last two generations. Accuracies for PEBV and GEBV were estimated as the correlation with TBV. Five replicates were made for each simulation, and average accuracies of these were used.

Results

Using ssGBLUP instead of PBLUP in scenario A, improved accuracies with 22.4 to 29.6% for animals in the last two generations and without progeny (Table 2). Overall, the increase in accuracy was higher for the last generation (gen. 25) compared to the previous generation (gen. 24). In Scenario A, the largest improvement of accuracy by using ssGBLUP instead of PBLUP was for the trait C4 with heritability 0.30, followed by the corresponding normal score transformed trait C4NS, for individuals with no progeny in the last two generations. The accuracy improvement from using ssGBLUP instead of PBLUP for Scenario B, was generally higher when the animals with phenotypes (females) were genotyped (22.6–28.2%), than when males without phenotypes, were genotyped (14.2-20.5%). Relatively small effects of the normal score transformation were seen in scenario A (Table 2).

Discussion

The increase in accuracy from using ssGBLUP compared to PBLUP was higher for the trait C4, than for the binary C2 and C2NS traits, probably due to the lower initial accuracy for C4 (using PBLUP). For the low heritability traits (LH), all accuracies were as expected lower compared to the corresponding traits with higher heritability. More records for predicting accurate GEBV are needed for traits with lower heritability, such as calving ease, compared to traits with higher heritability (Hayes *et al.*, 2009). However, the use of genomic information improved the accuracies for binary (C2) traits in this study. This shows that adding genomic information and using ssGBLUP is beneficial also for categorical traits with lower heritability, such as calving, fertility and health traits. When phenotypes only for females were included, using genotypes for females resulted in higher accuracies (Scenario B (b)) compared to using genotypes of males (Scenario B (a)).

To conclude, genomic selection has opened a window for estimating more accurate breeding values also for categorical traits in beef cattle, especially for traits such as calving ease with low heritability. Although the response to selection for such a trait is not as fast as for other traits with higher heritability, having an appropriate phenotype recording along with adding genotype data could improve accuracy of selection. Moreover, genotyping of animals with poorly defined phenotypes like for categorical traits appeared better than having disparate phenotyped and genotyped populations. The current simulated results showed that using ssGBLUP could improve accuracies within different scenarios in simulated Swedish beef cattle population.

Table 2. Accuracies and accuracy improvements when using ssGBLUP instead of PBLUP for animals in generations 24 and 25 without phenotypes and progeny.

Scenario[1]	Trait type[2]	Accuracy					
		PBLUP		ssGBLUP		Improvement	
		Gen 24	Gen 25	Gen 24	Gen25	Gen 24	Gen 25
A	Quant	0.544	0.494	0.674	0.640	23.9%	29.6%
	$C2_{constant}$	0.484	0.446	0.598	0.566	23.6%	26.9%
	$C2_{changed}$	0.482	0.440	0.592	0.562	22.8%	27.7%
	$C2_{constant}$NS	0.482	0.446	0.598	0.566	24.1%	26.9%
	$C2_{changed}$NS	0.482	0.442	0.590	0.562	22.4%	27.1%
	$C2_{constant}$TM	0.488	0.448	0.600	0.568	23.0%	26.8%
	$C2_{changed}$TM	0.482	0.442	0.594	0.562	23.2%	27.1%
	$C4_{constant}$	0.378	0.351	0.479	0.452	27.0%	28.6%
	$C4_{constant}$NS	0.385	0.358	0.488	0.458	26.7%	27.9%
	QuantLH	0.362	0.328	0.444	0.414	22.7%	26.2%
	$C2_{constant}$LH	0.300	0.276	0.372	0.346	24.0%	25.4%
	$C2_{changed}$LH	0.294	0.270	0.362	0.338	23.1%	25.2%
	$C2_{constant}$NSLH	0.300	0.276	0.370	0.346	23.3%	25.4%
	$C2_{changed}$NSLH	0.292	0.270	0.362	0.340	24.0%	25.9%
B (a)	QuantLH	0.310	0.278	0.354	0.322	14.2%	15.8%
	$C2_{constant}$LH	0.248	0.232	0.294	0.270	18.5%	16.4%
	$C2_{changed}$LH	0.234	0.222	0.282	0.260	20.5%	17.1%
B (b)	QuantLH	0.310	0.278	0.380	0.348	22.6%	25.2%
	$C2_{constant}$LH	0.248	0.232	0.314	0.292	26.6%	25.9%
	$C2_{changed}$LH	0.234	0.222	0.300	0.282	28.2%	27.0%

[1] See Table 1.

[2] Quant = quantitative; C2 and C4 = categorical with 2 and 4 categories, respectively; constant and changed = refer to trait thresholds; NS = normal score transformation; LH = low h^2; TM = threshold model.

Acknowledgements

This study was financed by the Swedish Farmers' Foundation for Agricultural Research.

References

Aguilar I., Tsuruta S., Masuda Y., Lourenco D., Legarra A., *et al.* (2018) Proc. of the 11th WCGALP, Auckland, NewZealand.

Hayes B.J., Bowman P.J., Chamberlain A.J. and Goddard M.E. (2009) J Dairy Sci 92(2): 433–443. https://doi.org/10.3168/jds.2009-92-3-1313

Legarra A., Christensen O.F., Aguilar I., and Misztal I. (2014) Livest. Sci 166:54-65. https://doi.org/10.1016/j.livsci.2014.04.029

Lourenco D.A.L., Misztal I., Wang H., Aguilar I., Tsuruta S., *et al.* (2013) J Anim Sci 91(9):4090-4098. https://doi.org/10.2527/jas.2012-5826

Lourenco D.A.L., Tsuruta S., Fragomeni B.O., Masuda Y., Aguilar I., *et al.* (2015) J Anim Sci 93(6):2653-2662. https://doi.org/10.2527/jas.2014-8836

Meijering A., and Gianola D. (1985) Genet Sel Evol 17:115. https://doi.org/10.1186/1297-9686-17-1-115

Meuwissen T., Hayes B., and Goddard M. (2016) Anim Front 6(1):6-14. https://doi.org/10.2527/af.2016-0002

Sargolzaei M., and Schenkel F.S. (2009) Bioinformatics 25(5):680-681. https://doi.org/10.1093/bioinformatics/btp045

658. Genetic parameters for mature cow body condition scores and fat and muscle traits in finishing beef cattle

F. Weik[1,2*], R.E. Hickson[1,3], S.T. Morris[1], D.J. Garrick[1] and J.A. Archer[4]

[1]School of Agriculture and Environment, Massey University, Private Bag 11222, Palmerston North 4442, New Zealand; [2]AbacusBio Ltd., P.O. Box 5585, Dunedin 9058, New Zealand; [3]Focus Genetics, 17c Mahia St, Ahuriri 4144, New Zealand; [4]Beef + Lamb NZ Genetics, P.O. Box 5501, Dunedin 9054, New Zealand; f.weik@massey.ac.nz

Abstract

Traits related to muscle and fat deposition are economically relevant for beef cattle finishing but may also influence reproductive performance in mature beef cows. The aim of this study was to examine the genetic relationships of mature cow body condition scores (BCS) with fat and muscle traits in finishing cattle. Genetic correlations among BCS at different seasonal time points were high, with the lowest estimate of 0.84 between BCS at weaning and BCS prior to calving. Generally, muscle and fat traits in both mature cows and finishing cattle were lowly genetically correlated but similarly so across all measured time points. There is scope to identify animals with favourable genotypes for both maternal and finishing fat and muscle traits. Improving genetic merit for muscling or fat content in finishing cattle may be achieved without greatly impacting on body energy reserves in mature cows.

Introduction

Body energy reserves are key components determining the reproductive performance of beef cows in the breeding herd (Osoro and Wright 1992). The optimum condition of cows may differ throughout the production cycle and for varying production systems. Cow body condition scoring is a tool to subjectively assess stored energy reserves in cows, made up of fat and muscle (Hickson *et al.* 2017). Currently, in New Zealand breeding values are commonly available for ultrasound (US) or carcass fat and muscle traits in finishing cattle but not for body condition score (BCS) in mature cows. Some breeders are selecting carcass fat traits with the objective of maintaining BCS in cows. Although extensive research exists on the genetic relationships among economically relevant traits in beef cattle, there is only limited information on the relationship of subjective measures of body condition in cows with objectively measured muscling and fat traits in slaughtered cattle (De Lacerda *et al.* 2019). Weik *et al.* (2022) showed that genetic correlations were low between cow BCS and finishing traits in New Zealand beef cattle, assuming that BCS measured at different time points was genetically the same trait. The aim of this study was to determine whether distinct genetic relationships exist for cow BCS at different seasonal time points, considered as separate traits, with muscle and fat traits in finishing beef cattle.

Materials & methods

Dataset. All manipulations were approved by the AgResearch Animal Ethics Committee. Data originated from an ongoing beef progeny test by Beef + Lamb NZ Genetics conducted on commercial New Zealand hill country farms since 2014 (Weik *et al.* 2021). Measures on several performance and production traits were recorded for the original population of Angus or Hereford cows and their crossbred progeny. This study used data on traits related to muscling and fat composition measured on cows in the breeding herd (Angus, Hereford or Stabilizer sired) and males and females as part of the finishing operation (Angus, Hereford, Simmental, Stabilizer or Charolais sired).

Traits. Body condition score was obtained through visual assessment of body energy reserves of mature cows from three years of age on a 1 to 10 scale (1= emaciated, 10 = obese; Hickson *et al.* (2017)) for a

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_658, © A. Nazari-Ghadikolaei *et al.* 2022

total of 4,590 cows. Scoring was conducted by a trained scorer or by the farmer under regular supervision and training. Measures were taken at three seasonal time points: prior to mating, at weaning and prior to calving. Scanning of the US traits eye-muscle area (UEMA) and rib fat depth (URF) was conducted on live animals between 475 and 616 days of age by a Breedplan-accredited technician on 1,900 steers and 2,049 heifers. The traits carcass eye-muscle area (CEMA) and carcass rib fat depth (CRF) were recorded between 541 and 1,029 days of age by trained graders using the Meat Australia chiller assessment system (AUS-MEAT 2018) on the day following slaughter of 1,615 steers and 354 heifers.

Data editing. A contemporary group (CG) was assigned to each animal and trait. For BCS, CG comprised herd, recording date and management group at the time of measurement. For US and carcass traits, CG also included the sex of the animal and management groups leading up to the US scan or slaughter date, respectively. Phenotypes were removed from the dataset when only one individual was recorded in a CG or when all animals within a CG had equal phenotypes.

Cow BCS was pre-adjusted to five years of age by fitting a fixed effects model with cow age in years (cows≥12 years of ages in a single class) and CG as class effects in the model. Finishing traits were pre-adjusted to either a constant live weight of 432 kg for US traits or a constant carcass weight of 310 kg for traits measured at slaughter (representing the mean weight at US scanning or slaughter, respectively) to compare muscularity and fatness rather than absolute size. Adjustments on an individual animal basis were obtained using the four-step procedure presented by Reverter *et al.* (2000) and fitting either live weight at US scanning or carcass weight as linear and quadratic covariates as well as CG as a factor in that model. Following pre-adjustments, observations above or below three standard deviations (SD) from the CG mean were removed. All traits were tested for heterogeneity of variances. Traits with a significant regression ($P<0.05$) of the CG SD on the CG mean were scaled to homogenize the variances (Pickering *et al.* 2012, Weik *et al.* 2022).

Statistical analysis. Data cleaning and pre-adjustments of the phenotypes were done using R version 3.6.0. Genetic parameters were estimated using restricted maximum likelihood in ASREML 4.1 (Gilmour *et al.* 2015). Variance components were obtained for BCS and US traits using univariate animal models. For carcass traits bivariate models were used by including live weight at US scanning as a correlated trait to account for any selection prior to slaughter due to drafting of animals in weight-dependent slaughter management groups following US scanning. Genetic and phenotypic correlations were estimated from bivariate animal models, or trivariate models when a carcass trait together with live weight at US scanning was included.

A CG class effect, and covariates for breed percentage and heterosis (0 = purebred; 1 = first-cross) were fitted as fixed effects for all traits, and an age of dam class effect was included for US traits only. A random animal and residual error term were fitted for each trait. Additionally, a random permanent environmental effect was fitted for each BCS trait to account for correlated residuals on repeated measures from the same animal. Variance structures for the random effects were assumed as follows: $var(a)=A\sigma_a^2$, $var(pe)=I\sigma_{pe}^2$ and $var(e)=I\sigma_\varepsilon^2$. The numerator relationship matrix A comprised a total of 14,240 animals including 424 sires and 4,473 dams.

Results

Among seasonal BCS traits, the phenotypic mean was lowest prior to calving but the associated SD were similar and small across seasons (Table 1). Variance components for BCS traits were in a similar range across seasons for the additive genetic (0.13-0.17) and the residual (0.33-0.37) variances, however, the permanent environmental variance was slightly higher at weaning than at other seasons, and this is

Table 1. Means (±standard deviations), variance components (σ_a^2=direct additive genetic, σ_{pe}^2=permanent environmental, σ_ε^2=residual), heritabilities (h^2±standard errors) and repeatabilities (t±standard errors) in New Zealand beef cattle.[1]

Trait	Mean	σ_a^2	σ_{pe}^2	σ_ε^2	h^2	t
BCS mating	7.25 (0.91)	0.15	0.04	0.37	0.27 (0.04)	0.34 (0.01)
BCS weaning	6.93 (0.91)	0.17	0.10	0.36	0.27 (0.04)	0.43 (0.01)
BCS calving	6.46 (0.93)	0.13	0.05	0.33	0.26 (0.06)	0.35 (0.01)
UEMA[2]	63.24 (5.03)	5.84		10.26	0.36 (0.05)	
URF[2]	3.53 (0.98)	0.53		0.45	0.54 (0.05)	
CEMA[2]	74.58 (6.75)	13.91		18.79	0.43 (0.11)	
CRF[2]	5.35 (2.59)	2.01		5.18	0.28 (0.07)	

[1] BCS = mature cow body condition score; UEMA = eye-muscle area (ultrasound); URF = rib fat depth (ultrasound); CEMA = eye-muscle area (carcass); CRF = rib fat depth (carcass).

[2] Finishing cattle; parameters first published by Weik *et al.* (2022).

reflected in the higher repeatability. Heritabilities for BCS traits were the same across seasons (0.26-0.27). Heritabilities of finishing fat and muscle traits were higher overall compared to BCS at any season. At US scanning, heritability was higher for URF compared to UEMA whereas CEMA was more heritable than CRF at slaughter.

Genetic correlations among BCS traits were high across all seasons, ranging from 0.84 to 0.99 (Table 2). The lowest estimate was observed between BCS at weaning and BCS before calving.

Genetic correlations were generally low among BCS at all seasonal time points with finishing traits (Table 3), but moderate correlations were obtained for BCS prior to calving with either UEMA (0.31) or URF (0.32). Genetic correlations of finishing traits with BCS tended to be highest at calving and lowest at weaning with the exception of CEMA. Across all time points BCS traits were slightly higher genetically correlated with fat traits compared to muscling.

Table 2. Genetic (±standard errors; below diagonal) and phenotypic (±standard errors; above diagonal) correlations for body condition scores (BCS) at different seasons in mature New Zealand beef cows.

	BCS mating	BCS weaning	BCS calving
BCS mating		0.65 (0.01)	0.47 (0.01)
BCS weaning	0.95 (0.03)		0.54 (0.01)
BCS calving	0.99 (0.06)	0.84 (0.09)	

Table 3. Genetic (r_g) and phenotypic (r_p) correlations (±standard errors) among BCS in mature New Zealand beef cows and muscle and fat traits in finishing cattle.[1,]

	UEMA		URF		CEMA[2]	CRF[2]
	r_g	r_p	r_g	r_p	r_g	r_g
BCS mating	0.18 (0.07)	0.06 (0.04)	0.27 (0.06)	0.10 (0.04)	-0.07 (0.11)	0.08 (0.12)
BCS weaning	0.14 (0.08)	-0.01 (0.04)	0.16 (0.06)	0.08 (0.04)	-0.05 (0.12)	-0.06 (0.12)
BCS calving	0.31 (0.09)	0.14 (0.05)	0.32 (0.08)	0.21 (0.05)	0.03 (0.12)	0.17 (0.13)

[1] BCS = mature cow body condition score; UEMA = eye-muscle area (ultrasound); URF = rib fat depth (ultrasound); CEMA = eye-muscle area (carcass); CRF = rib fat depth (carcass).

[2] No r_p identifiable between traits not expressed within the same animal (carcass and mature cow BCS traits).

Discussion

Heritabilities for BCS were similar across seasons but overall higher compared to results presented by Arango *et al.* (2002). Arango *et al.* (2002) reported slightly higher permanent environmental variation at weaning compared to measures taken prior to mating and calving, which aligns with results found in the current study. Results indicate potential to improve BCS through direct selection at any season, but genetic progress may be slow due to low phenotypic variation. A selection response is likely to be stronger for muscling and fat traits in finishing animals compared to BCS traits in cows due to higher heritabilities (Wolcott *et al.* 2013).

In agreement with Arango *et al.* (2002) and Wolcott *et al.* (2013), cow BCS traits were highly genetically correlated across seasons. Thus, animals with a higher genetic merit for BCS at any season are likely to also exhibit greater body energy reserves at other seasons in the annual production cycle. The lower genetic correlation between BCS at weaning and BCS prior to calving compared to BCS at mating may be attributable to the physiological status of the cows during lactation or pregnancy which is likely to impact on fat deposition (Arango *et al.* 2002).

Generally, genetic associations of fat and muscle traits in finishing cattle with mature cow BCS traits were low and similar across time points such that treating seasonal BCS measures as different traits may not be required. The BCS before calving, however, tended to be slightly higher genetically correlated to all finishing traits compared to BCS at other seasons and this may also be induced by physiological differences during late pregnancy. Direct selection on carcass traits with the objective of increasing cow BCS throughout the annual cycle is likely to be inefficient. Ultrasound rib fat and UEMA offered the best options for selecting on finishing traits to generate a correlated response in cow BCS and results were similar to those presented by De Lacerda *et al.* (2019). However, genetic progress in maternal and finishing fat and muscle traits is likely to be more efficient when animals are measured for both trait groups.

References

Arango J.A., Cundiff L.V., and Van Vleck L.D. (2002) J. Anim. Sci. 80(12): 3112-3122. https://doi.org/10.2527/2002.80123112x

AUS-MEAT. (2018) Australian beef carcase evaluation. Beef and veal chiller assessment language. Available at: https://www.ausmeat.com.au/WebDocuments/Chiller_Assessment_Language.pdf.

De Lacerda V.V., Campos G.S., Silveira D.D., Roso V.M. *et al.* (2019) Anim. Prod. Sci. 59(7): 1209-1217. https://doi.org/10.1071/AN17873

Gilmour A.R., Gogel B.J., Cullis B.R., Welham S.J., and Thompson R. (2015) ASReml user guide release 4.1 structural specification. VSN international Ltd.: Hemel Hempstead, UK.

Hickson R.E., Morris S.T., and Thomson B.C. (2017) Beef cow body condition scoring. Available at: https://www.beeflambnz.com/knowledge-hub/PDF/beef-cow-body-condition-scoring.

Osoro K., and Wright I.A. (1992) J. Anim. Sci. 70(6):1661-1666. https://doi.org/10.2527/1992.7061661x

Pickering N.K., Dodds K.G., Blair H.T., Hickson R.E. *et al.* (2012) J. Anim. Sci. 90(5):1411-1420. https://doi.org/10.2527/jas.2011-4163

Reverter A., Johnston D.J., Graser H.-U., Wolcott M.L., and Upton W.H. (2000) J. Anim. Sci. 78(7):1786-1795. https://doi.org/10.2527/2000.7871786x

Weik F., Hickson R.E., Morris S.T., Garrick D.J., and Archer J.A. (2021) Animals. 11(9): 2509-2525. https://doi.org/10.3390/ani11092509

Weik F., Hickson R.E., Morris S.T., Garrick D.J., and Archer J.A. (2022) Animals. 12(1): 25-43. https://doi.org/10.3390/ani12010025

Wolcott M.L., Johnston D.J., Barwick S.A., Corbet N.J., and Williams P.J. (2013) Anim. Prod. Sci. 54(1): 37-49. http://dx.doi.org/10.1071/AN12427

659. Decision support tools to support a more sustainable beef-on-dairy industry

S.C. Ring[1*], R.D. Evans[1], J. McCarthy[1], A. Cromie[1], and D.P. Berry[2]

[1]Irish Cattle Breeding Federation, Highfield House, Shinagh, Bandon, Co. Cork, P72 X050, Ireland; [2]Animal & Grassland Research and Innovation Centre, Teagasc, Moorepark, Fermoy, Co.Cork, Ireland; sring@icbf.com

Abstract

The global agri-industry is challenged with improving sector sustainability. Cattle production systems are not exempt, especially as dairy cow genetic merit becomes more specialized for dairy production, often to the determinant of beef value. Using national measures of genetic merit, the Irish industry has developed and deployed three interdependent strategies to complement the sustainable productivity and profitability of the dairy and beef sectors. The first tool, the Dairy Beef Index (DBI), was developed to identify genetically elite beef bulls for mating to dairy females with the goal of simultaneously improving calving performance and beef merit. Adoption of the DBI will be incentivized through: (1) sire mating advice to aid in mating decisions; and (2) a *Commercial Beef Value* per animal to aid beef producers in valuing cattle.

Introduction

Dairy cow reproductive performance in Ireland has progressively improved reducing the requirement for replacement dairy females and, in turn, create an opportunity to increase surplus calf value. Consequently, the frequency of mating beef sires with dairy females has increased 10% in the previous five years, a trend expected to continue. Beef sire selection by dairy producers has traditionally focussed on breed, easy expected calving, and short expected gestation length. The antagonistic genetic correlation between calving performance and beef merit will contribute to a progressive deterioration in beef merit of the beef sires available. Therefore, decision support tools are required to: (1) identify suitable beef sires for mating with dairy females that possess both desirable calving and beef attributes; (2) recommend optimum matings of beef bulls with dairy females taking cognisance of genetic and phenotypic characteristics of the female; and (3) value surplus cattle destined for slaughter. These three tools must be inter-dependent in that suitable beef bulls must be available from which to suggest the recommended matings and the value of the resulting progeny must, on average, reflect the expected value at the time of mating. Combined, these decision support tools will help to improve the profitability and sustainability of the intrinsically linked dairy and beef systems. The objective here is to provide an overview of these decision support tools which can be relatively easily extended to other jurisdictions.

Materials & methods

Measures of genetic merit for calving, beef, and docility performance traits (Evans, 2007, 2008, 2009, 2020) from the national multibreed genetic evaluations were available to construct each of decision support tools.

Dairy beef index. The DBI was derived to rank beef bulls for mating to dairy females (Berry *et al.*, 2019) as:

$$DBI = direct\ dystocia + direct\ gestation + mortality + feed\ intake + carcass\ weight + carcass\ conformation + carcass\ fat + desirable\ carcass\ specification + docility + polledness$$

where each trait was a function of the sire's predicted transmitting ability (PTA) multiplied by the economic value and cumulative discounted genetic expression (Berry *et al.*, 2019). The economic value applied to direct dystocia was non-linear (Berry *et al.*, 2019) whereby bulls with a higher dystocia PTA were penalised more severely.

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_659, © S.C. Ring *et al.* 2022

Mating advice. The mating advice decision support tool ranks matings between each member of a team of beef bulls (selected by the dairy producer) to each dairy female in a herd where easier calving bulls are more favoured towards females more prone to calving difficulty while also considering the carcass merit of the bull to correct the deficiencies in carcass merit of the female (e.g. very light, poor conformation). Each sire × female mating receives an index based on the expected profit of the resulting progeny as:

$$INDEX = CALVING + (1\text{-}Pred_{MORT_BIRTH}) \times POLLED + (1\text{-}Pred_{MORT_FINISH}) \times BEEF$$

where $Pred_{MORT_BIRTH}$ is the predicted calf mortality at birth based on the phenotypic mean of the base population plus the sum of the PTA for perinatal morality of both parents, $Pred_{MORT_FINISH}$ is simply $Pred_{MORT_BIRTH}$ + 0.03 where the 0.03 is the mean mortality from birth until slaughter and POLLED is the economic value of whether or not the beef bull is polled. The CALVING and BEEF components of the index are described as:

$$CALVING = Pred_{dystocia} \times COST_{dystocia} + Pred_{gestation} \times COST_{gestation}$$

$$BEEF = Pred_{CARCASS} \times PRICE_{CARCASS} + Pred_{docility} \times EW_{docility} + Pred_{feed\ intake} \times EW_{feed\ intake} + breed_bonus$$

The prediction of dystocia ($Pred_{dystocia}$) considers the direct PTA of the parents for dystocia plus twice the female's maternal PTA for dystocia all estimated from the national genetic evaluations. Non-linearity in the economic impact of calving difficulty is accounted for. Prediction of gestation length ($Pred_{gestation}$) is simply the sum of the parental direct PTAs. The prediction of carcass and carcass price per kg is a function of the expected EBV of the progeny for carcass weight, conformation and fat score individually but also includes a penalty if the resulting carcass does not adhere to abattoir carcass specifications for carcass weight and conformation. Predictions for docility and feed intake are based on the sum of the parental PTAs; a breed bonus was applied to Angus and Hereford bulls.

Commercial beef value. The Commercial Beef Value (CBV) values cattle for slaughter (Dunne *et al.*, 2021) as:

$$CBV = carcass\ weight + carcass\ conformation + carcass\ fat + docility + feed\ intake$$

where each trait was a function of the animal's estimated breeding value multiplied by the economic value and cumulative discounted genetic expression; non-genetic effects are likely to be added to the CBV in later versions.

Results & discussion

Growing consumer awareness of farming practices has strengthened the need to re-examine the inter-dependency between the dairy and beef sectors. Assuming 20-25% of calves born in the dairy herd are required as replacements, approximately 75-80% of calves are destined for slaughter; of the latter in Ireland, many are purchased <6 weeks old, weaned, and finished typically at 22-24 months of age by beef producers. Because dairy farmers are often not directly impacted by the sale value of the animal at slaughter, this can lead to a lack of consideration and appreciation of the importance of beef bull selection. In addition, the purchaser of the resulting calf <6 weeks of age cannot have any great knowledge of the carcass potential of the animal at such a young age. The decision support tools outlined here should incentivise stakeholders to work in unison towards the same common goal, a sustainably profitable dairy beef system. All data flowing into these tools originate from the one source, the ICBF database. The ICBF database maintains and utilises data from an array of sources including farmers, abattoirs, marts, etc.; all data used in these decision

support tools are updated bi-monthly as additional phenotypes become available; decision support tools are also updated.

The DBI rewards bulls that produce a live calf with minimal calving assistance following a short gestation; docile, polled progeny that attain an acceptable carcass weight and both a desirable carcass conformation and fat grade are also favoured. The ability of this index to deliver phenotypically has been validated (Berry and Ring, 2020). The mean DBI for some beef breeds are provided in Figure 1. The Dairy Beef Index is comparable across beef breeds since the genetic evaluations used to develop the index are based on a multibreed population where all beef breeds are included in the same genetic evaluation. The DBI has highlighted the desirability of some breeds which would not traditionally have been used in dairy herds (e.g. Aubrac, Saler).

Although dairy producers remain cautious on selecting beef bulls that are acceptable on calving performance, the beef merit of beef bulls used in dairy herds has improved since the inception of the DBI, and this has also translated to a favourable change in CBV trends; trends in DBI and CBV pre and post the DBI launch are in Figure 2. Future versions of the DBI are likely to consider meat eating quality, animal health, age at slaughter and direct measures of methane.

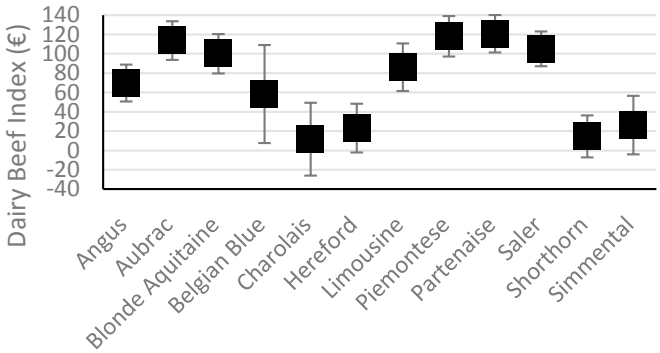


Figure 1. Mean and SD (at each side of SE bar) Dairy Beef Index value for the main beef breeds used in Ireland.

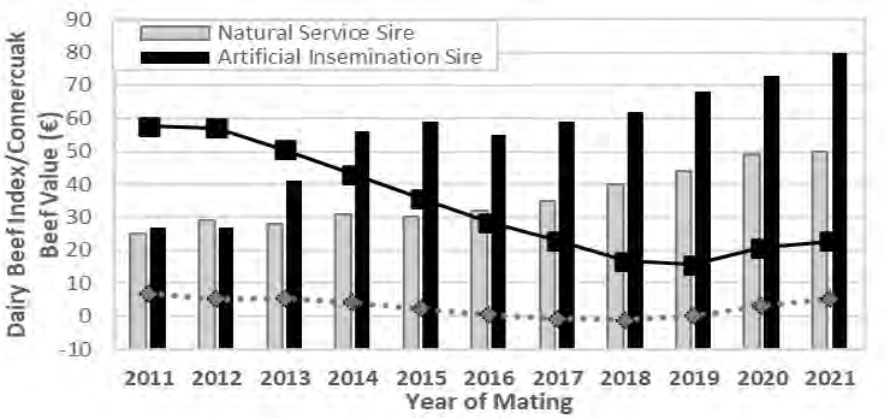


Figure 2. Trend in Dairy Beef Index (columns) and Commercial Beef Value (markers) for natural service sires and artificial insemination sires pre and post the availability of the Dairy Beef Index in 2019.

The framework of the sire mating advice algorithm is built on that of the DBI thus exploiting the gains in beef seedstock producing cattle of increasing DBI. Dairy producers must first select a team of beef bulls with a separate team selected for heifers and cows; the maximum usage per bulls is inputted. Future versions will recommend a bull team and usage. Using the sire advice, an older cow with good direct and maternal genetic merit for dystocia may be allocated the more difficult of the bulls in the team in the pursuit of achieving a better calf value through improved carcass credentials. The producer decides which females are to be served to dairy bulls, beef bulls, and which females are to be culled (and not bred) based on the COW index (Kelleher *et al.*, 2015); it is common practice in Irish dairy herds to use dairy semen for the first 6 weeks of the 15-week breeding season after which beef bulls are used. The sire mating advice program mates each bull in the team to each female; the three sub-herds are treated separately so that the intended proportional representative use of bulls should be close to the eventual usage in: (1) heifers; (2) cows that the farmer has indicated should be served with a beef bull; and (3) all other cows which are not destined to be culled. An index value for each candidate bull × female combination is generated within a sub-population; then linear programming is used to maximise the expected index value of the resulting progeny within the constraints of the desired proportional representation of each bull. A check is also included to avoid some predefined matings like those which might result in very difficult calvings; this is at the discretion of the user. Some 8,000 dairy herds (out of a total of 17,000 nationally) already use sire mating advice to rank dairy bulls for use on dairy females; the user interface for the beef-on-dairy mating system is close to a mirror representation of the dairy-on-dairy sire advice. The output to the farmer is a user-friendly report which identifies the female and the three highest ranked bulls on the index. Users can save the results for their own use as well as onto their AI technicians handheld, enabling the immediate identification of preferred bull by technicians on the day of service.

Potential purchases need a reliable method of identifying profitable progeny from the beef-on-dairy matings. Assuming no/minimal market failure, this should logically facilitate higher market demand for the seller thus encouraging the dairy producer to use more elite beef bulls and closing the loop. The Commercial Beef Value was developed to support decisions on how much to pay for a calf destined for prime beef production. Dunne *et al.* (2021) validated that knowledge of the breed and Commercial Beef Value of the animal was a better indication of than knowledge of the animal breed alone. Each of these decision support tools should have a substantial impact on the future sustainability of both dairy and beef sectors alike; methods applied in the present study can be replicated in other jurisdictions to improve the profitability of beef progeny originating from the dairy herd.

References

Berry D.P., Amer P.R., Evans R.D., Byrne, T., Cromie A.R., *et al.* (2019) J. Dairy Sc. 102(11):10056-10072. https://doi.org/10.3168/jds.2019-16912

Berry D.P., and Ring S.C. (2020) J. Dairy Sc. 103(13):2523-2533. https://doi.org/10.3168/jds.2019-17431

Dunne F.L., Evans R.D., Kelleher M.M., Walsh S.W., and Berry D.P. (2021) Animal 15(2):100077 https://doi.org/10.1016/j.animal.2020.100077

Evans, R.D., Pabiou T., Cromie A.R., Kearney F., Veerkamp R., *et al.* (2007) Proc. of INTERBULL Technical Workshop, Paris, France. Bulletin 36, p. 7

Evans, R.D., Pabiou T., Cromie A.R., Kearney F., Veerkamp R., *et al.* (2008) Proc. of INTERBULL Meeting, Niagara Falls, NY. Bulletin 38, p. 13.

Evans, R.D., Pabiou T., Cromie A.R., Kearney F., Veerkamp R., *et al.* (2009) Proc. of INTERBULL Meeting, Barcelona, Spain. Bulletin 40, p. 127.

Evans R.D. (2020) Proc. of ICBF and Sheep Ireland Genetics Conference, Laois, Ireland

Kelleher M.M., Amer P.R., Shalloo L., Evans R.D., Byrne T.J., *et al.* (2015) J. Dairy Sc. 98(6):4225-4239. https://doi.org/10.3168/jds.2014-9073

Part 37

Bovine dairy – genetic evaluation methods

660. Breed and heterosis effects for fatty acid composition of milk from dairy cows milked once- and twice-a-day

I. Sanjayaranj[1*], N. López-Villalobos[1], H.T. Blair[1], P.W.M. Janssen[2], S.E. Holroyd[3] and A.K.H. MacGibbon[3]

[1]Animal Science, School of Agriculture and Environment, Massey University, Private Bag 11 222, Palmerston North 4442, New Zealand; [2]School of Food and Advanced Technology, Massey University, Private Bag 11 222, Palmerston North 4442, New Zealand; [3]Fonterra Research and Development Centre, Palmerston North 4442, New Zealand; i.sanjayaranj@massey.ac.nz

Abstract

The objective of this study was to estimate the effects of breed and heterosis on the gross composition and fatty acid composition of milk from cows milked once-a-day (OAD) and twice-a-day (TAD). Milk samples were taken from 39 Holstein-Friesian (F), 27 Jersey (J) and 34 F×J crossbred cows from a OAD milking herd, and 104 F and 83 F×J crossbred cows from a TAD milking herd. Holstein-Friesian cows were superior for the production traits compared to J cows in both milking frequencies. Jersey cows milked OAD produced significantly higher percentage of fat and a higher proportion of C18:0 fatty acid. The heterosis effects for the production traits were positive and significant in cows milked OAD. The heterosis effects for fatty acids were not significant in both milking frequencies. The butter produced from F milk would likely be softer than of that of J milk, mainly from F cows milked OAD. These results can be used to assist with selecting breeds and animals that are suitable for either OAD or TAD milking and this can allow closer alignment with milk processing needs.

Introduction

The New Zealand dairy industry is pasture-based, with most of the cows calving in spring. Twice-a-day (TAD) milking during the whole lactation is practiced in about 55% of the herds and once-a-day (OAD) during the whole lactation in about 10% of herds, with the remainder using a mixture of TAD and OAD (Edwards, 2018). Once-a-day milking has been adopted by several farmers due to its benefits of reducing the cost of labour, feed and management and creating alternative employment opportunities (Edwards, 2018). Milk composition, especially milk fat and fatty acids (FAs), affects the processability of milk into products such as cheese and butter (MacGibbon, 1996; Auldist *et al.* 2004). The variation in milk composition and fatty acid composition is caused by several factors, including genetics, breed, feed, season, stage of lactation and age of cows. Heterosis is defined as the superiority of the first crossbred cows compared to the average performance of the parents. Heterosis is considered as an important factor which affects the economically important traits. There are no studies reporting the effect of milking frequency on the fatty acid composition of cows under grazing conditions. The objective of this study is to estimate the effects of breed and heterosis on gross composition and fatty acid composition of cows milked OAD and TAD.

Materials & methods

The experiment was carried out at Massey University dairy farms from July 2020 to March 2021. Massey University No. 1 Dairy farm is a low-input system with cows milked OAD and cows grazing on ryegrass (*Lolium perenne*) and white clover (*Trifolium repens*) pasture with limited supplements. Massey University No. 4 Dairy farm is managed as a high-input system with cows milked TAD and cows grazing on rye grass and white clover pasture with a higher proportion of supplements.

Sampling and determination of milk composition and fatty acids. Morning milk samples were collected from 39 Holstein-Friesian (F), 27 Jersey (J) and 34 F×J crossbred cows from a OAD milking

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_660, © I. Sanjayaranj *et al.* 2022

herd. Morning and evening milk samples were collected separately from 104 F and 83 F×J crossbred cows from a TAD milking herd. Milk samples were collected from first and second parity cows on both farms. The analysis of milk composition was carried out within two days of sampling. Gross milk composition and FA composition were determined using a Milkoscan FT1 (Foss Analytical, Hillerød, Denmark) for individual samples. Prediction equations for FAs were developed in the FTIR calibrator software (Foss Analytical) using a data set with mid infrared spectral data and gas chromatography reference values. The The coefficient of determination values of the prediction equations for all the FAs were greater than 0.8. Composition of morning and evening milk samples was analysed separately, and an average weighted by volume was calculated for daily yield to get the final values for daily yield and composition.

Statistical analysis. Data were analysed using the MIXED procedure of the SAS package version 9.4 (SAS Institute Inc., Cary, NC, USA). The model included the fixed effects of milking frequency, lactation number, stage of lactation, proportion of F and F×J heterosis within milking frequency, and deviation from median calving date with linear and quadratic effect, and the random effect of cow and residual error.

Results

The proportion of saturated fatty acids was lower ($P<0.05$), and the proportion of unsaturated fatty acids was higher in cows milked OAD compared to cows milked TAD (Table 1). The proportions of short-chain fatty acids (SCFA) and long-chain fatty acids (LCFA) were significantly lower, and the proportion of medium-chain fatty acids (MCFA) was significantly higher in cows milked OAD. Breed effects for daily milk yields of milk, protein and lactose, percentages of fat and protein and proportions of SFA, C18:0 and omega6 in cows milked OAD were significantly different ($P<0.05$) from zero (Table 2). Breed effects for the proportions of C18:0 and LCFA were significantly different from zero in cows milked TAD. The breed effects for the percentages of fat and the proportions of SFA and C18:0 were negative in cows milked OAD and TAD. Friesian cows produced milk with an average 1.08% less SFA than J cows milked OAD. The breed effects for the percentage of protein and proportions of C16:0 and LCFA were significantly different between miking frequencies. The heterosis effects for the daily yield of milk, fat, protein and lactose of cows milked OAD were positive and significantly different from cows milked TAD.

Discussion

Holstein-Friesian cows had greater yields of milk, protein, and lactose than J cows, but J cows produced milk with higher concentration of fat in both milking frequencies as previously reported in New Zealand (Lambeye *et al.* 2015). The F×J heterosis effects for milk, fat and protein yield were greater in OAD than TAD milking, suggesting that heterosis effects are expressed differently according to the environment. The breed effect for the proportion of C18:0 was negative and significantly different from zero in both milking frequencies.

Friesian cows produced milk with a lower proportion of C18:0 compared to J cows which suggests that the butter produced from F milk is likely to be softer than the butter from the J milk. This is in agreement with MacGibbon (1996), who reported that the SFC_{10} (solid fat content at 10 °C) was positively correlated with the proportions of high melting point long-chain saturated fatty acids in the milk and negatively correlated with the proportion of C18:1n_9. Similarly, the F cows milked OAD would produce softer butter than F cows milked TAD. For bulk milk, cows milked OAD produce harder butter than cows milked TAD due to the higher proportions of MCFA (MacGibbon, 1996). However, the seasons of the year also have significant effect on the hardness of the butter as Auldist *et al.* (1998) reported that the butter produced during the spring was softer than the butter produced in summer due to lower proportions of long-chain saturated fatty acids. This suggests that F cows milked OAD will produce softer butter in spring.

Table 1. Least-square means and standard errors of daily production traits, gross milk composition and fatty acid composition of cows milked once-a-day (OAD) and twice-a-day (TAD).

Variable[1]	Milking frequency				*P*-value
	n	OAD	n	TAD	
Yield (kg/day)					
Milk	289	16.04±0.3	543	21.31±0.19	<0.0001
Fat	289	0.78±0.02	543	0.90±0.01	<0.0001
Protein	289	0.61±0.01	543	0.73±0.01	<0.0001
Lactose	289	0.72±0.02	543	1.00±0.01	<0.0001
Fat%	296	5.03±0.08	543	3.99±0.05	<0.0001
Protein%	296	4.12±0.03	543	3.76±0.02	<0.0001
Lactose%	296	4.80±0.02	543	5.02±0.01	<0.0001
Fatty acid (% of the total fatty acids)					
SFA	296	69.47±0.21	543	70.47±0.13	<0.0001
UFA	296	31.03±0.17	543	29.59±0.11	<0.0001
PUFA	296	3.11±0.02	543	2.83±0.02	<0.0001
C4:0	296	3.83±0.02	543	4.12±0.01	<0.0001
C6:0	296	2.76±0.01	543	2.84±0.01	<0.0001
C8:0	296	1.53±0.01	543	1.51±0.01	0.0190
C10:0	296	3.57±0.03	543	3.36±0.02	<0.0001
C12:0	296	4.00±0.02	543	3.74±0.02	<0.0001
C14:0	296	12.83±0.08	543	12.72±0.05	0.2493
C16:0	296	31.31±0.18	543	32.39±0.12	<0.0001
C18:0	296	12.62±0.11	543	13.24±0.07	<0.0001
C18:1 *cis*-9	296	19.9±0.17	543	20.21±0.11	0.1310
Omega6	296	1.67±0.02	543	1.53±0.01	<0.0001
SCFA	296	8.12±0.04	543	8.46±0.03	<0.0001
MCFA	296	20.4±0.13	543	19.82±0.08	0.0002
LCFA	296	65.5±0.19	543	67.37±0.12	<0.0001

[1] LCFA = long chain fatty acids (sum of C16:0, C18:0 and C18:1n_9 and Omega6); MCFA = medium chain fatty acids (sum of C10:0, C12:0 and C14:0); PUFA = polyunsaturated fatty acid; SCFA = short chain fatty acids (sum of C4:0, C6:0 and C8:0); SFA = saturated fatty acids; UFA = unsaturated fatty acid.

The findings of this study reveal that the *de novo* synthesised FAs, the SCFA and MCFA were significantly affected only by milking frequency but not by the breed. The breed effects for percentages of fat and protein of cows milked OAD would be more important in the processability of milk than the fatty acid composition. These results suggest further studies should be undertaken to study the effects of milking frequency, with the prospect that better decisions could be made when selecting breeds that are suitable for either OAD or TAD milking frequency according to the processability of milk.

Table 2. Estimates of breed and heterosis effects for daily production traits, milk composition and fatty acids composition of milk from cows milked once-a-day (OAD) and twice-a-day (TAD).

Variable[1]	Breed effect[2]				Heterosis effect			
	OAD		TAD		OAD		TAD	
	Estimate	s.e	Estimate	s.e	Estimate	s.e	Estimate	s.e
Yield (kg/day)								
Milk	2.27*[a]	0.60	2.29[a]	1.95	3.45*[a]	1.25	-1.22[b]	1.29
Fat	-0.03[a]	0.04	0.01[a]	0.11	0.18*[a]	0.07	0.04[a]	0.08
Protein	0.04*[a]	0.02	0.09[a]	0.07	0.13*[a]	0.04	-0.01[b]	0.04
Lactose	0.10*[a]	0.03	0.12[a]	0.10	0.18*[a]	0.06	-0.04[b]	0.07
Fat%	-1.09*[a]	0.15	-0.89[a]	0.51	-0.3[a]	0.32	0.31[a]	0.33
Protein%	-0.40*[a]	0.05	0.03[b]	0.18	-0.03[a]	0.11	0.14[a]	0.12
Lactose%	-0.05[a]	0.03	0.04[a]	0.11	0.11[a]	0.07	0.08[a]	0.08
Fatty acid (% of the total fatty acids)								
SFA	-1.08*[a]	0.45	-1.88[a]	1.46	-0.5[a]	0.94	0.05[a]	0.96
UFA	0.59[a]	0.37	1.48[a]	1.19	0.69[a]	0.76	-0.07[a]	0.78
PUFA	0.07[a]	0.05	0.24[a]	0.17	-0.11[a]	0.11	0.04[a]	0.11
C4:0	-0.005[a]	0.04	-0.16[a]	0.14	0.02[a]	0.09	-0.06[a]	0.09
C6:0	-0.03[a]	0.03	-0.10[a]	0.10	-0.01[a]	0.06	-0.06[a]	0.06
C8:0	-0.01[a]	0.02	-0.02[a]	0.07	-0.01[a]	0.04	-0.04[a]	0.04
C10:0	0.08[a]	0.06	0.10[a]	0.20	0.02[a]	0.13	-0.07[a]	0.13
C12:0	0.06[a]	0.06	0.05[a]	0.18	-0.01[a]	0.12	-0.08[a]	0.12
C14:0	-0.13[a]	0.18	0.15[a]	0.59	-0.25[a]	0.38	0.15[a]	0.39
C16:0	0.05[b]	0.38	-2.24[a]	1.23	-0.83[a]	0.79	-0.45[a]	0.81
C18:0	-1.27*[a]	0.25	-1.59*[a]	0.80	-0.14[a]	0.52	0.37[a]	0.53
C18:1 *cis*-9	0.40[a]	0.37	0.60[a]	1.21	0.16[a]	0.78	0.06[a]	0.80
Omega6	0.14*[a]	0.04	0.24[a]	0.14	0.06[a]	0.09	0.03[a]	0.09
SCFA	-0.04[a]	0.08	-0.28[a]	0.26	-0.001[a]	0.17	-0.16[a]	0.17
MCFA	0.01[a]	0.28	0.31[a]	0.91	-0.23[a]	0.58	-0.01[a]	0.60
LCFA	-0.64[b]	0.41	-3.07*[a]	1.33	-0.67[a]	0.86	-0.02[a]	0.88

[1] LCFA = long chain fatty acids (sum of C16:0, C18:0 and C18:1n_9 and Omega6); MCFA = medium chain fatty acids (sum of C10:0, C12:0 and C14:0); PUFA = polyunsaturated fatty acid; SCFA = short chain fatty acids (sum of C4:0, C6:0 and C8:0); SFA = saturated fatty acids; UFA = unsaturated fatty acid.
[2] * = estimates were significantly ($P<0.05$) different from 0; [a,b] estimates with different superscripts were significantly different between milking frequencies ($P<0.05$).

References

Auldist M.J., Johnston K.A., White N.J., Fitzsimons W.P., and Boland M.J. (2004) J. Dairy Sc. 71(1):51-57. https://doi.org/10.1017/s0022029903006575

Auldist M.J., Walsh, B.J. and Thomson, N.A. (1998) J. Dairy Res. 65(3):401-411. https://doi.org/10.1017/S0022029998002970

Edwards J.P. (2018) Anim. Prod. Sci. https://doi.org/10.1071/AN17484

Lembeye F., López-Villalobos N., Burke J.L., and Davis S.R. (2015) Proc. N.Z. Soc. Anim. Prod. 75:60-63.

MacGibbon, A.K.H. (1996) Proc. N.Z. Soc. Anim. Prod. 56:224-227.

661. Long term effect of growth in rearing phase on reproduction and production trait in Murrah buffalo

P. Joshi[1*], A. Verma[1], G.R. Gowane[1], R. Alex[1], I.D. Gupta[1] and D. Worku[2]

[1]ICAR, National Dairy Research Institute, Karnal, India; [2]Department of Animal Science, Salale University, Salale, Ethiopia; joshi33333puja@gmail.com

Abstract

The purpose of the study was to determine the effect of average daily gain (ADG) on reproductive and milk production traits of Murrah buffalo. For this purpose, animals were divided into 3 groups based on ADG during different growth phase *viz* low, medium and high, respectively. Murrah heifers in the high ADG group (9-18month of age) reached age at first calving (AFC) 96 days earlier than those in low and medium ADG group ($P<0.05$). A significant increase of 206 kg of milk was found in the animals of higher ADG group ($P<0.01$).

Introduction

Dairy farmers face a complex dilemma in minimizing costs associated with rearing heifers while ensuring the lifetime economic productivity. Decisions about heifer management are influenced by underlying biological aspects of growth, thereby influencing future profitability of the herd (Mourits *et al.*,1999). Simplest approach to reduce costs is to shorten the non-productive period of dairy heifers, which can be accomplished by breeding the heifers early. It is already understood that puberty in heifers occurs when females achieve approximately 45 to 55% of their mature body weight (BW) (Freetly *et al.*, 2011). This goal requires a rapid growth rate during pre-pubertal and post pubertal phase.

Effect of growth during rearing phase on future performance of beef cattle has been studied by many researchers. However, there is no comprehensive study that has investigated the association between growth rate during rearing in Murrah buffalo with their productive and reproductive performance later in life. Therefore, the aim of this work was to evaluate the effect of average daily gain (ADG) before and after puberty on first lactation production and reproduction traits in Murrah heifers.

Materials & methods

The dataset used in this study consisted of growth, first lactation and reproduction records of Murrah buffaloes maintained at ICAR-National Dairy Research Institute, Karnal during 1974 to 2019. Body weights were measured at monthly intervals starting from time of birth up to 24 month of age. Average daily gain (ADG) was calculated using the difference in body weight (BW) between 2 consecutive months during the period analysed. ADG was calculated for different growth phases like birth to 3 month of age, birth to 6 month of age (up to birth to 24 month of age), 3 month to 6 month, 3 month to 9 month (up to 3 to 24 month of age) up to 18-24 month of age. Animals with missing birth date, body weight and calving date were pruned. Milk yield data of first parity were recorded and reproductive parameters like age at first calving (AFC), first service period (SP) and calving interval (CI) were generated.

Statistical analysis. The explanatory variable i.e. ADG of the heifers was divided into 3 (low, medium, high) classes for each growth trait in particular growth phase and the effect of each ADG class was studied on dependent variable (Table 1). The MIXED linear model, corrected for other significant effects, was applied to determine the effect of the heifer rearing period on production and reproduction during subsequent lactation. The Bonferroni method was used for means comparison.

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_661, © P. Joshi *et al.* 2022

$$Y_{ijklm} = \mu + A_i + Sj + BV_k + ADG_l + b(AFC_{ijkl}) + e_{ijklm}$$

where, Y_{ijklm} = the responding variable as mentioned in Table 1 below; μ = the overall mean; A_i = effect of the i^{th} year of calving; S_j = the effect of j^{th} season of calving (j= spring, summer, autumn or winter); BV_k = the effect of k^{th} estimated sire breeding value for respective trait; ADG_l = the effect of l^{th} ADG class on trait; $b(AFC_{ijkl} - AFC)$ = regression coefficient of age at first calving in days (exclude a scenario, when AFC is a responding variable); e_{ijklm} =the random error.

Results

ADG during different growth phase. During early rearing phase, highest ADG of 0.51 kg/day and 0.52 kg/day was found during the period from birth to 12 month and birth to 24 month of age. After 3 months of age, mean ADG was between 0.49 to 0.65 kg/day. Growth of nearly 0.52 kg/day was seen in 6 month to 12 month of age. Higher ADG of 0.60 kg/day was seen during the phase 9 to 12 month of age. After 1 year of growth period, mean ADG was 0.48 kg/day.

ADG and subsequent AFC performance. A significant difference of days (P<0.05) between the highest and the lowest ADG group in evaluated time period was observed in relation to the AFC(Figure 1). Studying ADG association revealed that the highest ADG was associated with the earliest AFC. A reduction of 67 days in AFC was achieved in the high ADG group of heifers as compared to low ADG group during the growth phase of birth to 18 months. The earlier AFC by 73 and 66 days was achieved with the highest ADG group 9-18 month and 12 to 18 month of age, respectively.

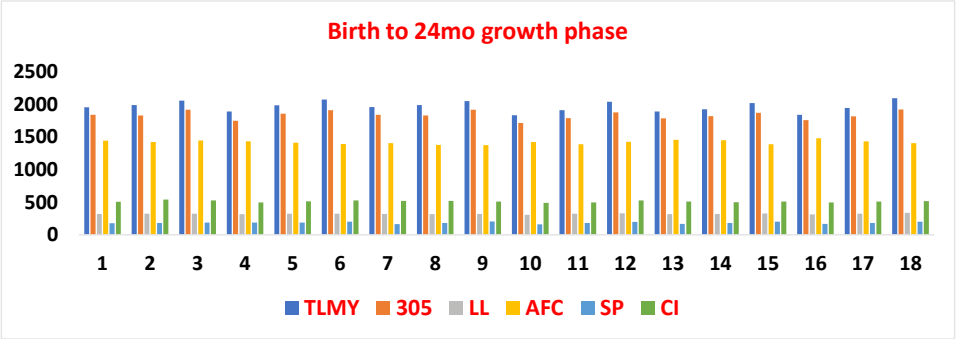
ADG and subsequent total lactation milk yield, 305day milk yield and lactation length. Higher ADG during various growth period was associated with increase in the TLMY. Significant increase (P<0.01) of 207 and 255 kg milk in TLMY was found in heifers of the high ADG group birth to 12 month and 12 month to 24 month of age, respectively. Significant increase (P<0.01) of 466 kg in TLMY was found in the 3 month to 24 month high ADG group. Similar trend was observed in the 305d milk yield. Increase of 26 days(P<0.05) in LL of heifer was observed in higher ADG group of 3 month to 24 month age.

ADG and subsequent service period and calving interval. The groups with increased service period were observed to have higher ADG during the evaluation period starting since birth. Similar results of ADG effects were observed between other evaluated periods. Significant increase of 38 days in Service period (P<0.05) was observed in higher ADG group of 9 month to 12 month of age. The results for calving interval revealed that elongated CI was associated with higher ADG group during entire rearing phase except 3 month to 6 and 9 month of age where CI was less in high ADG group.

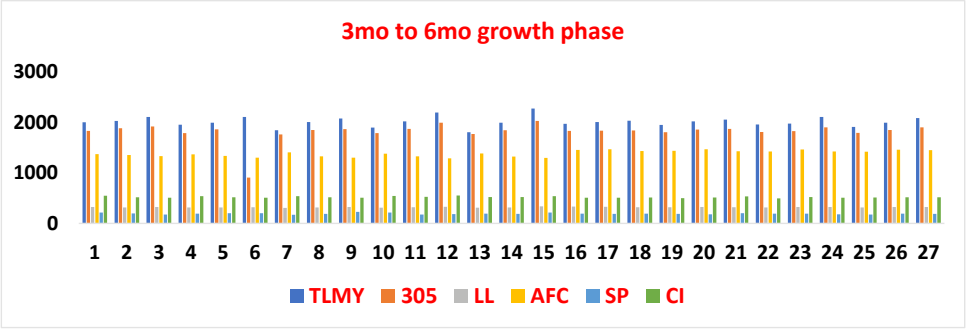
Table 1. Evaluated parameters in heifer and first lactation (dependent variable).[1]

Trait	Mean ± S.E.
Age at first service	975.45±4.93 (1,354)
Age at first calving	1,354.55±5.00 (1,946)
Total lactation milk yield (TLMY)	1,987.42±16.80 (1,655)
305d milk yield	1,841.37±14.36 (1,655)
Lactation length	314.84±2.00 (1,655)
Service period	163.47±2.81 (1,179
Calving interval	475.12±2.84 (1,249)

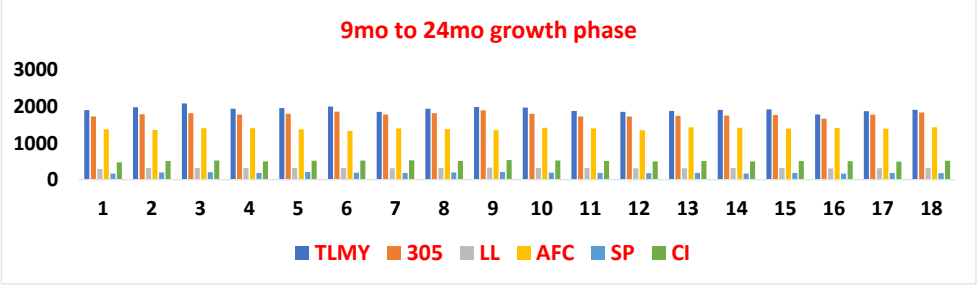
[1] Values in the parenthesis represents number of observation under each trait.



Where **1,2,3** are low medium and high ADG group respectively during birth -3mo growth phase; **4,5,6** during birth -6mo growth phase;**7,8,9** during birth -9mo growth phase ;**10,11,12** during birth -12mo growth phase;**13,14,15** during birth -18mo growth phase:**16,17,18** during birth -24mo growth phase. are low, medium and high ADG group respectively



Where **1,2,3** are low medium and high ADG group respectively during 3mo-6mo growth phase; **4,5,6** during 3mo -9mo growth phase;**7,8,9** during 3mo -12mo growth phase ;**10,11,12** during 3mo -18mo growth phase;**13,14,15** during 3mo-24mo growth phase; **16,17,18** during 6mo-9mo growth phase; **19,20,21** during 6mo-12mo growth phase; **22,23,24** during 6mo-18mo growth phase; **25,26,27** during 6mo-24mo growth phase.



Where **1,2,3** are low medium and high ADG group respectively during 9mo-12mo growth phase; **4,5,6** during 9mo -18mo growth phase;**7,8,9** during 9mo -24mo growth phase ;**10,11,12** during 12mo -18mo growth phase;**13,14,15** during 12mo-24mo growth phase; **16,17,18** during 18mo-24mo growth phase.

Figure 1. Effect of ADG during different time period on dependent traits.

Discussion

In this study, faster growth led to earlier AFC, which was related to early maturity of these animals. Likewise, Lopez *et al.* (2018) and Atashi (2021) also reported that heifers with a rapid daily gain calved at a younger age than those which shows delayed growth. This association can be explained by assuming that those animals with higher pre pubertal and post pubertal growth rate reach the weight required for a successful pregnancy at a younger rate. Our results showed positive effect of ADG on milk yield of first lactation. A few studies have also documented similar findings where greater pre-weaning -ADG)- had a positive effect on the future production (Moallem *et al.* 2010; Soberon *et al.* 2012; Soberon and Van Amburgh, 2013),

mainly due to higher growth rate during early life. On the contrary, Ettema and Santos (2004) and Daniels (2010) reported negative effect of high ADG during the heifer rearing period on milk production in first lactation due to the reduction of parenchyma growth in the mammary gland.

Conclusions

Overall results suggested that better ADG during early life has significant beneficial effect on future reproduction and production performance.

References

Atashi H., Asadi A., and Hostens M (2021). PLoS One 0244825 https://doi.org/10.1371/journal.pone.0244825

Byerley D. J., Staigmiller R.B., Berardinelli J.B., and Short R.E. (1987) J. Anim. Sci. 65:645–650. https://doi.org/10.2527/jas1987.653645

Daniels K. M. (2010). Proc. 19th Annu. Tri-State Dairy Nutrition Conf., Ft. Wayne,IN: 69–76

Ettema J., and Santos J (2004). J Dairy Sci 87: 2730–2742. https://doi.org/10.3168/jds.S0022-0302(04)73400-1

Freetly H.C., Kuehn L.A., and Cundiff L.V. (2011). J. Anim. Sci. 89:2373–2379. https://doi.org/10.2527/jas.2011-3847

Hoseyni F., Mahjoubi E., Zahmatkesh D., and Yazdi M (2016). J Dairy Res. 83(4):453-455. https://doi.org/10.1017/S0022029916000558

López E., Véliz F.G., Carrillo E., Santiago Á.D., García J.E., Mellado M. (2018). Slov Vet Res. 55 (1): 35. https://doi.org/10.26873/SVR-394- 2017

Moallem U., Werner D., Lehrer H., Zachut M., Liveshitz L., Yakoby S., and Shamay A. (2010). J. Dairy Sci. 93:2639–2650. https://doi.org/10.3168/jds.2009-3007

Mourits M.C.M., Huirne R.B.M., Dijkhuizen A.A., Kristensen A.R., and Galligan D.T. (1999). Agric. Syst. 61:17–31.

Soberon F., and Amburgh M.E. (2013). J. Anim. Sci. 91: 706–7012 https://doi.org/10.2527/jas2012-5834

Soberon F., Raffrenato E., Everett R.W., and Amburgh M.E. (2012). J Dairy Sci. 95: 783–793 https://doi.org/10.3168/jds.2011-4391

662. Integrating parent-of-origin effects in random regression models for milk traits

R.E. Jahnel*, I. Blunk and N. Reinsch

Research Institute for Farm Animal Biology (FBN), Institute of Genetics and Biometry, Wilhelm-Stahl Allee 2, 18196 Dummerstorf, Germany; jahnel@fbn-dummerstorf.de

Abstract

The importance of parent-of-origin effects (POEs) for the genetic variability of quantitative traits is frequently addressed by current research in farm animals, humans and laboratory animals and may have implications for animal breeding. An epigenetically induced non-equivalence of identical alleles when inherited either from sire or dam is assumed as the underlying mechanism. In order to restrict the number of necessary (co-)variance components in random regression models to a minimum we suggest that POEs are considered for lower-order terms of Legendre-polynomials only. Using 256,089 milk samples from 31,345 German Holstein cows we attributed shares between 5.9 and 12.6% of the total additive genetic variance to POEs (yield and composition traits; protein percentage: not significant). Consequences for selection and genetic evaluation remain to be investigated in follow-up studies.

Introduction

The potential action of parent-of-origin effects (POEs, including genomic imprinting) can be investigated by (generalised) gametic models (Schaeffer *et al.* 1989; Reinsch *et al.*, 2021), where each individual with a phenotype is represented through four gametic effects: two with paternal and two with maternal expression patterns. Such models can account for both full and partial imprinting (Neugebauer *et al.*, 2010 a,b; Meyer and Tier, 2012). The non-equivalence of paternal and maternal inheritance at these loci is summarised in a single parameter called the imprinting variance, which can be interpreted as the variance of the difference of gametic effects under both parental expression patterns (Neugebauer *et al.*, 2010a,b; Blunk *et al.*, 2017a,b). Including POEs into random regression models (RRM; Jamrozik and Schaeffer, 1997) increases the number of genetic (co-)variance parameters drastically. Even with only 2nd order Legendre-polynomials, as many as 21 parameters need to be estimated in a POE-RRM, compared to six parameters under the assumption of a RRM without POEs (Mendelian inheritance). Therefore, we assume POEs only for the lower order parts of the Legendre-polynomials, i.e. at least the intercept, and neglect POEs for all the higher-order terms. We demonstrate and discuss our new POE-RRM by analysing milk traits measured in 31,345 Holstein cows from Mecklenburg-West Pomerania, Germany.

Materials & methods

Cow data. Data for milk yield (MY), fat yield (FY) and -percentage (FP), and protein yield (PY) and percentage (PP) from first lactation dairy cows in Mecklenburg-West Pomerania were collected between 2008 and 2019 and were provided by the German Evaluation Centre (VIT, Verden/Germany). Cows with known parents and grandparents of German Holsteins were kept. The calving age of these cows was between 20 and 40 months. Samples taken between 5 and 330 days in milk (DIM) were taken into account for cows with ≥3 records or one record ≤DIM 80 or two records ≤DIM 100. Cows were either milked by an automated milking system or in a traditional milking parlour. This resulted in 31,345 cows with 256,089 milk samples and 125,041 individuals in the pedigree.

POE-RRM. The model was implemented in Echidna MMS (Gilmour, 2021) and can be written as:

$$\boldsymbol{y}_{irjd} = \boldsymbol{HTD}_i + \sum_{k=0}^{3} \boldsymbol{\phi}_{tk} \boldsymbol{\beta}_{kr} + \boldsymbol{u}^{s}_{j1} + \boldsymbol{u}^{d}_{j2} + \sum_{k=1}^{2} \boldsymbol{\phi}_{tk} \boldsymbol{u}_{kj1} + \sum_{k=1}^{2} \boldsymbol{\phi}_{tk} \boldsymbol{u}_{kj2} + \sum_{k=0}^{3} \boldsymbol{\phi}_{tk} \boldsymbol{pe}_{jk} + \boldsymbol{e}_{irjd} \quad (1)$$

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_662, © R.E. Jahnel *et al.* 2022

where y_{irjd} is the test day yield of the d^{th} test day of cow j, HTD_i is the fixed effect of the i^{th} herd-test-day x milking frequency, ϕ_{tk} are coefficients of Legendre-polynomials of order two or three at DIM (t); β_{kr} represents the k^{th} regression coefficient of the fixed lactation curve of the fixed effect r representing calving age-year-season of cow j; $\boldsymbol{u}^s_{j1}$ ($\boldsymbol{u}^d_{j2}$) is the paternal (maternal) gametic effects of cow j as sire (as dam), superscripts s and d indicate paternal and maternal expression patterns of the imprinted random intercept effect; u_{kj1} and u_{kj2} are the un-imprinted linear and quadratic random regression coefficients for paternal (1) and maternal (2) gametes of cow j assuming mendelian inheritance; pe_{jk} is the k^{th} random regression coefficient of cow j for permanent environment effects and e_{irjd} is the residual. It was assumed, that the covariance structure of the random genetic effects is:

$$\text{var}\begin{bmatrix}\boldsymbol{u}^s\\ \boldsymbol{u}^d\\ \boldsymbol{u}_1\\ \boldsymbol{u}_2\end{bmatrix} = \left[\begin{array}{cc|cc} \sigma^2_{\boldsymbol{s}} & \sigma_{\boldsymbol{sd}} & \sigma_{\boldsymbol{s},\boldsymbol{u}1} & \sigma_{\boldsymbol{s},\boldsymbol{u}2} \\ \sigma_{\boldsymbol{sd}} & \sigma^2_{\boldsymbol{d}} & \sigma_{\boldsymbol{d},\boldsymbol{u}1} & \sigma_{\boldsymbol{d},\boldsymbol{u}2} \\ \hline \sigma_{\boldsymbol{s},\boldsymbol{u}1} & \sigma_{\boldsymbol{d},\boldsymbol{u}1} & \sigma^2_{\boldsymbol{u}1} & \sigma_{\boldsymbol{u}1,\boldsymbol{u}2} \\ \sigma_{\boldsymbol{s},\boldsymbol{u}2} & \sigma_{\boldsymbol{d},\boldsymbol{u}2} & \sigma_{\boldsymbol{u}1,\boldsymbol{u}2} & \sigma^2_{\boldsymbol{u}2} \end{array}\right] = \begin{bmatrix}\boldsymbol{V}_{1,1} & \boldsymbol{V}_{1,2}\\ \boldsymbol{V}'_{1,2} & \boldsymbol{V}_{2,2}\end{bmatrix} = \boldsymbol{V} \tag{2}$$

where $V_{1,1}$ and $V_{2,2}$ represent gametic (co-)variance components for imprinted (here: the intercept) and non-imprinted (linear and quadratic) components of the Legendre-polynomials, respectively, and $V_{1,2}$ comprises all covariances between the two kinds of gametic effects. Furthermore, it is assumed that $\text{var}(\boldsymbol{u}) = \boldsymbol{V} \otimes \overline{\boldsymbol{G}}$, $\text{var}(\boldsymbol{pe}) = \boldsymbol{C} \otimes \boldsymbol{I}_{\boldsymbol{N}}$ and $\text{var}(\boldsymbol{e}) = \boldsymbol{I}_{\boldsymbol{n}}\boldsymbol{\sigma}^2_{\boldsymbol{e}}$. $\overline{\boldsymbol{G}}$ is the generalised gametic relationship matrix (Reinsch *et al.*, 2021) and $\otimes$ denotes the Kronecker product. I_N is an identity matrix of order N of the cows and n being the sample size. C is a symmetric matrix, containing variance components for permanent environmental effects, corresponding to the Legendre-polynomials (4×4 matrix). Significance of the imprinting variance (see below) was tested through a restricted maximum log-likelihood ratio test with four degrees of freedom.

Genetic Parameters over lactation. Estimates were derived using variance components for each DIM t ϵ {5, …, 330}, similar to Jahnel *et al.* (2021) using self-written programs in R (R Core Team, 2020). If the vector of Legendre-polynomials for the paternal gamete is $\phi_{s,t} = (\phi_{t,0},0,\phi_{t,1},\ldots,\phi_{t,n})$ and maternal gamete is $\phi_{d,t} = (0,\phi_{t,0},\phi_{t,1},\ldots,\phi_{t,n})$, the additive genetic variance σ^2_a was obtained at DIM t as: $\sigma^2_{a,t} = (\phi_{s,t}V\phi_{s,t}') + (\phi_{d,t}V\phi_{d,t}')$. The imprinting variance at DIM t is defined as:

$$\boldsymbol{\sigma}^2_{\boldsymbol{i},\boldsymbol{t}} = (\boldsymbol{\phi}_{\boldsymbol{s},\boldsymbol{t}}\boldsymbol{V}\boldsymbol{\phi}_{\boldsymbol{s},\boldsymbol{t}}') + (\boldsymbol{\phi}_{\boldsymbol{d},\boldsymbol{t}}\boldsymbol{V}\boldsymbol{\phi}_{\boldsymbol{d},\boldsymbol{t}}') - 2(\boldsymbol{\phi}_{\boldsymbol{s},\boldsymbol{t}}\boldsymbol{V}\boldsymbol{\phi}_{\boldsymbol{d},\boldsymbol{t}}') \tag{3}$$

If no imprinting effect is observed $(\phi_{s,t}V\phi_{s,t}') = (\phi_{d,t}V\phi_{d,t}') = (\phi_{s,t}V\phi_{d,t}')$ and $\sigma^2_i = 0$. The relative imprinting variance can be estimated as:

$$\boldsymbol{i}^2_{\boldsymbol{t}} = \frac{(\boldsymbol{\phi}_{\boldsymbol{s},\boldsymbol{t}}\boldsymbol{V}\boldsymbol{\phi}_{\boldsymbol{s},\boldsymbol{t}}') + (\boldsymbol{\phi}_{\boldsymbol{d},\boldsymbol{t}}\boldsymbol{V}\boldsymbol{\phi}_{\boldsymbol{d},\boldsymbol{t}}') - 2(\boldsymbol{\phi}_{\boldsymbol{s},\boldsymbol{t}}\boldsymbol{V}\boldsymbol{\phi}_{\boldsymbol{d},\boldsymbol{t}}')}{(\boldsymbol{\phi}_{\boldsymbol{s},\boldsymbol{t}}\boldsymbol{V}\boldsymbol{\phi}_{\boldsymbol{s},\boldsymbol{t}}') + (\boldsymbol{\phi}_{\boldsymbol{d},\boldsymbol{t}}\boldsymbol{V}\boldsymbol{\phi}_{\boldsymbol{d},\boldsymbol{t}}')} \tag{4}$$

Results & discussion

To account for the complex nature of traits with repeated measurements, the new POE-RRM was tested on 256,089 milk samples of German Holstein cows. Investigating POEs can lead to a large system of equations and reducing equations in RRM is crucial. For the sake of simplicity, we assumed that most of the potential POEs can be explained by an imprinted intercept, neglecting POEs for all higher polynomial terms and reducing the variance component matrix V from 21 unknown parameters to 10. In other words, a constant imprinting variance σ^2_i over time was assumed. The model, in principle, could however be expanded up to a full POE-RRM by assuming POEs for every variance component (For 2nd order Legendre-polynomials: paternal gamete vector $\phi_{s,t} = (\phi_{t,0},\phi_{t,1},\phi_{t,2},0,0,0)$ and maternal gamete vector $\phi_{sd,t} = (0,0,0,\phi_{t,0},\phi_{t,1},\phi_{t,2})$.

With the new POE-RRM we estimated a potential action of POEs on five milk production traits: MY, FY, FP, PY and PP and found significant imprinting effects in the yield traits (MY, FY, PY) and FP. The error probability for PP was slightly above the 5% significant threshold (Table 1). This was also reflected by the genetic correlations between gametic effects as sire and dam. The lowest genetic correlation between both gametes in PY is 0.8918, whereas both gametes in PP are expressed almost equally (r_g>0.99). When averaged over lactation i^2 ranged from 5 to 12% (Table 1) where σ^2_i was significant. Our model is unique in that it can account for the differences in i^2 over time. Thus, we can demonstrate highest i^2 estimates in the early lactation stage before DIM 70, peaking at DIM 60 with 14.9% for MY, at DIM 57 with 9.3% for FP and at DIM 69 with 18.7% for PY (Figure 1). Only i^2 for FY peaked at the end of the early lactation stage (<DIM 120) at DIM 115 with 9.4%. This suggests a stronger influence and importance of POEs in the expression of the traits during peak lactation. Generally, research on POEs in milk traits is still scarce, to the best of our knowledge with only three previous studies (Kaiser *et al.* (1998): significant imprinting effect in MY, Essl and Voith (2002): significant imprinting effect in PP, Jiang *et al.* (2017): imprinting in MY, FY and PY). However, none of these studies estimated POEs through a RRM approach.

Follow-up research will, of course, be needed for a confirmation of our results. Details of potential applications in breeding programs will depend on wether few imprinted loci with larger effects can be identified or if many chromosomal regions with small contributions to POEs exist. At least in the latter case our parsimonious version of a POE-RRM may also be useful for the purpose of genetic evaluation.

Acknowledgements

R.E. Jahnel gratefully recognises financial support from the H. Wilhelm Schaumann-Stiftung.

Table 1. *P*-values of restricted maximum log-liklihood ratio test, imprinting variance (σ^2_i), average total additive variance (σ^2_a), genetic correlation (r_g) between paternal and maternal gametic effect and average relative imprinting variance (i^2).

Trait	*P*-value	σ^2_i	σ^2_a	r_g	i^2
milk yield (kg)	0.0040	0.8211	7.5130	0.9069	0.1133
fat percentage (%)	<0.001	0.0069	0.1319	0.9690	0.0589
fat yield (kg)	0.0127	0.0005	0.0083	0.9379	0.0741
protein percentage (%)	0.0621	0.0002	0.0269	0.9990	0.0073
protein yield (kg)	0.0127	0.0008	0.0069	0.8918	0.1259

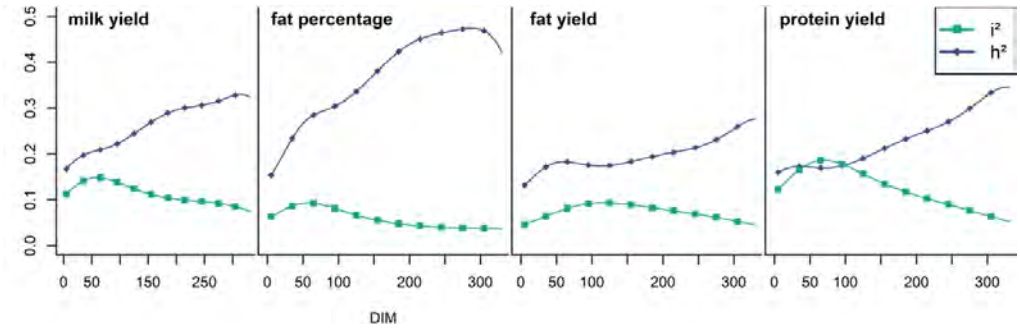


Figure 1. Estimates for heritability (h^2) and relative imprinting variance (i^2) for each DIM.

References

Blunk, I., Mayer, M., Hamann, H. and Reinsch, N. (2017a) Animal 11:1096-1106. https://doi.org/10.1017/S1751731116002391

Blunk, I., Mayer, M., Hamann, H. and Reinsch, N. (2017b) J Anim Sci 95:559-571. https://doi.org/10.2527/jas.2016.0997

Essl, A. and Voith, K. (2002) J Anim Breed and Genet 119:182-189. https://doi.org/10.1046/j.1439-0388.2002.00334.x

Gilmour, A.R. (2021) Echidna Mixed Model Software. Available at: www.EchidnaMMS.org

Jahnel, R.E., Wittenburg, D., Mayer, M., Täubert, H. And Reinsch, N. (2021) Züchtungskunde 93:157-168.

Jamrozik, J. and Schaeffer, L.R. (1997) J Dairy Sci 80:762-770. https://doi.org/10.3168/jds.S0022-0302(97)75996-4

Jiang, J., Shen, B., O'Connell, J.R., VanRaden, P.M., Cole, J. B. *et al.* (2017) BMC Genomics 18:425. https://doi.org/10.1186/s12864-017-3821-4

Kaiser, C., Goddard, M., and Reverter, A. (1998) Proc. of the 6th WCGALP, Armidale, Australia.

Meyer, K. and Tier, B. (2012) J Anim Breed Genet 129:359-368. https://doi.org/10.1111/j.1439-0388.2012.00996.x

Neugebauer, N., Luther, H. and Reinsch, N. (2010a) Animal 4:672-681 https://doi.org/10.1017/S1751731109991625

Neugebauer, N., Rader, I., Schild, H.J., Zimmer, D. and Reinsch, N. (2010b) J Anim Sci 88:523-532 https://doi.org/10.2527/jas.2009-2026

R Core Team (2020): R: A language and environment for statistical computing, version 3.6.1. Available at: https://www.R-project.org/

Reinsch, N., Mayer, M. and Blunk, I. (2021) G3 11:jkab064. https://doi.org/10.1093/g3journal/jkab064

Schaeffer, L.R., Kennedy, B.W. and Gibson, J.P. (1989) J Dairy Sci 72:1266-1272. https://doi.org/10.3168/jds.S0022-0302(89)79231-6

663. Recombination rate and hotspots in Holstein cattle from Mexico

E.I. Mejía-Melchor[1*], A. García-Ruiz[2], C. Vásquez-Peláez[3], H.O. Toledo-Alvarado[3] and F.J. Ruiz-Lopez[2]

[1]Posgrado en Ciencias de la Producción y de la Salud Animal-UNAM, Circuito de Posgrados, Edificio B, 1° Piso, Ciudad Universitaria, Alcaldía Coyoacán, C.P 04510, Ciudad de México, México; [2]CENID FyMA-INIFAP, km. 1, Carretera a Colón, Col. Ajuchitlán Colón, C.P. 76280, Querétaro, México; [3]Facultad de Medicina Veterinaria y Zootecnia-UNAM, Circuito escolar s/n, Ciudad Universitaria, C.P. 04510 Alcaldía Coyoacán, Ciudad de México, México; enoc.2189@comunidad.unam.mx

Abstract

The present research aimed to describe recombination rates, their patterns through autosomes, as well as the identification of high-frequency recombination sites in the genome of Mexican Holstein cattle. Recombination, like functional mutations, can be considered as the raw material of the evolution of species, since both, mutation and recombination change the configuration of paternal alleles during meiosis. Genotypes (98,619 markers) of 652 animals from 298 families (father, mother and offspring) were included in the study. Based on the ~2.4 million recombination events identified from the families studied, the recombination frequencies in pairs of single nucleotide polymorphisms (SNP) were calculated in a permutation process, obtaining an average frequency of 0.08, with a maximum frequency of 0.14 and a minimum of 0.03. Those SNP whose recombination value exceeded 2.5 standard deviations of the mean were assumed as hotspots, finding 2,230 that represent the 2% higher than the genotype studied.

Introduction

Species need an adequate amount of genetic variation to cope with environmental conditions and their changes. During recombination, parental genotypes are mixed resulting in new gene combinations, and phenotype expression. Recombination begins with a double break in the deoxyribonucleic acid (DNA) strand, the break site as a first step is degraded and later repaired. During repair, recombination events known as crosses can occur, since DNA strands of homologous chromosomes are occupied (Petes, 2001; Úbeda and Wilkins, 2011).

Estimation of recombination frequencies across genomes has evolved; In the beginning, crossover events were detected by counting phenotypes, parental and recombinant, which allowed the calculation of linkage disequilibrium (LD) and gene mapping; At present, with the advancement of technologies, such events can be identified using molecular markers and statistical methodologies, which allow the estimation of parameters for the studied populations, thus obtaining high-density maps with information on LD regions, patterns recombination frequencies and hotspots, which are small regions of genome where recombination events occur most frequently and can explain a large part of the recombination that occurs throughout the entire genome from one generation to the next.

Evidence has been found that the increase in genetic variation allows an increase in the response to selection in the short term, as well as the management and use of knowledge of meiotic recombination, increasing genetic variation between generations, would allow a response to the selection with greater intensity in the long term(Gonen *et al.*, 2017).

Knowledge of recombination rates and the presence of high-frequency recombination sites could help to understand the origin of genetic variability in a population and assist in the design of optimal selection programs for that population.

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_663, © E.I. Mejía-Melchor *et al.* 2022

Materials & methods

Data set and genotypes. In the present study included 652 animals, in 298 families with information of father, mother and offspring. The genotype used comes from three different chips: 1) Bovine HD® Beadchip (777,962 SNP), 2) Genomic Profiler HD-V3 (139,376 SNP) and 3) GGP HDv02 (138,892 SNP). The study included only those SNPs in common in all panels, leaving a final markers number of 98,619 (for each individual).

Construction of the high-density maps. The construction of the high-density maps was carried out with the information of the frequency of the recombination events observed in the population, as well as the physical position between adjacent SNPs and the genetic distance.

Estimation of recombination rates. For this estimation was used the direct recombinant counting methodology θ (Ott, 1999; Garbe and Da, 2003), θ is a vector with the estimated frequency of recombination, which consist in divide off the observed number of recombination events found (n), by the total number of meiosis, the offspring (T) (these frequencies vary from 0 to 0.5) for each pair of markers (i), obtaining the estimated recombination frequency by Equation 1:

$$\theta_i = n_i / T_i \quad (1)$$

The methodology is an iterative process, where recombination events are estimated through a permutation analysis between a marker with its adjacent ones, obtaining the frequencies of the recombination events for θ.

Identification of the hotspots. Was carried out according to Sandor *et al.* (2012), which designate those sites where the recombination frequencies are 2.5 standard deviations greater than the mean of θ, computed in the LocusMap® package (Ma *et al.*, 2015; Shen *et al.*, 2017).

Visualization of the recombination patterns throughout the genome. The θ data was smoothed by fitting the data with a 'smooth spline' model and a relative physical position was set for each of the marker intervals, dividing the original physical position by the length of the corresponding chromosome, the information obtained was plotted and compared with the information of each chromosome using the 'ggplot2' package of the statistical software R version 4.0.5

Results

Around 2.4 million recombination events were identified in the 298 families studied, and recombination frequencies were calculated with a mean of 2.4 crosses, with a maximum value of 146 and a minimum of 47 for each pair of markers, with a mean of 58 crosses.

The position of the recombination windows was standardized by dividing the position of the first marker in the window by the size of the chromosome. Recombination rates in an average chromosome were graphed, as well as the division of the bovine karyotype into three groups (large, medium and small) with respect to the size of the chromosomes measured in bp (*Bos taurus* ARS-UCD1.2 2016), as show in Figure 1.

Based on the frequencies of crossing or recombination of the 98,619 SNPs, those SNPs were assigned as hot spots whose recombination value exceeded 2.5 standard deviations of the means of θ, finding 2,230 hot spots that represent the upper 2% of the recombination frequencies of the genotypes studied, as can be seen in Figure 2 (hotspots in red points).

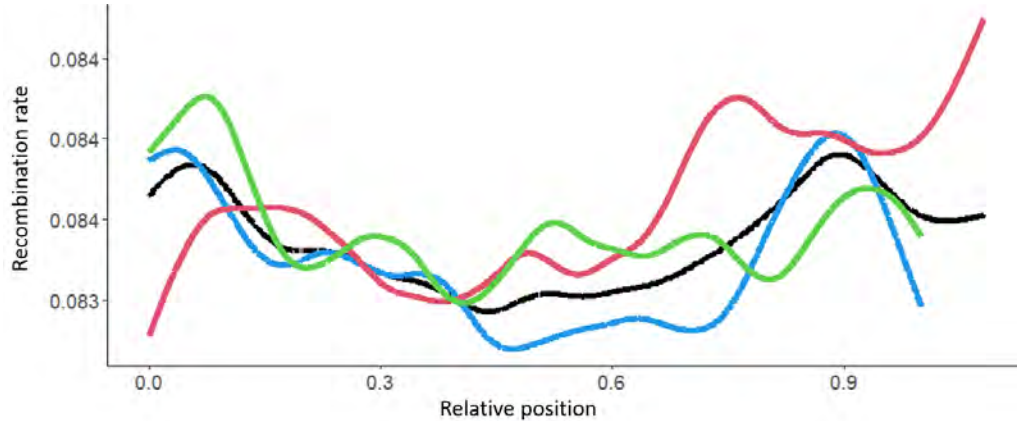


Figure 1. Recombination rate patterns in different autosomes sizes (general, large, medium and small autosomes in black, red, blue and green respectively).

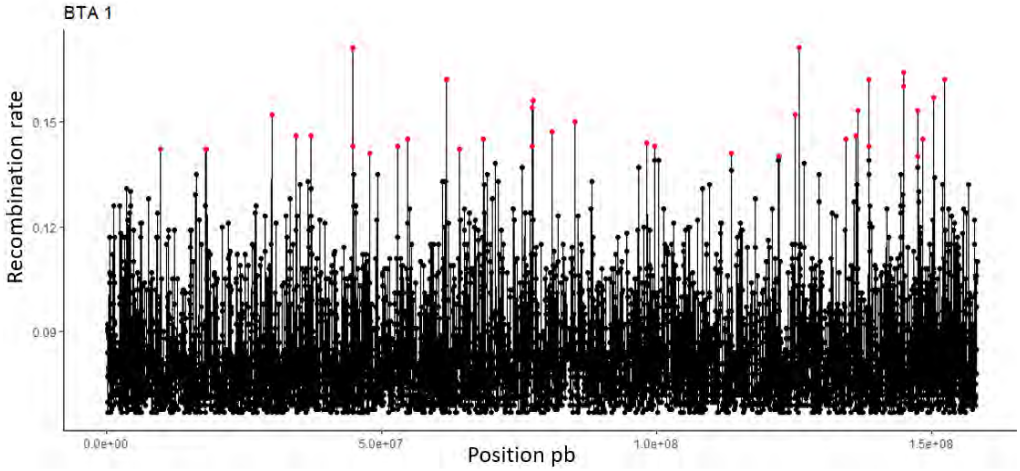


Figure 2. Hotspots (red points) in autosome 1.

Discussion

The average behaviour of the recombination frequencies through all the chromosomes can be seen in Figure 1, where the values of θ are usually higher in the telomeres of the chromosomes, showing a depression of the same in the physical center. This is similar to results presented by Li Ma *et al.* (2015) with 1.6 million recombination events and Shen *et al.* (2017) with 3.4 million, with an average of 24 recombination events per SNP window.

For the hot spots results, we also found differences from those reported by Li Ma in 2015 and Qanbari in 2020 (Qanbari and Wittenburg, 2020), and we attribute these differences to the different density of markers used. In this work we used 98K SNPs, being almost twice as many compared with the aforementioned studies, who used 50k SNP's.

References

Garbe J., and Da Y. (2003). Locusmap User Manual Version 1.1. Department of Animal Science. Univ. Minnesota, Minneapolis/St. Paul.

Gonen, S., M. Battagin, S. E. Johnston, G. Gorjanc, and J. M. Hickey. (2017). Genet. Sel. Evol. 49:1–12. https://doi.org/10.1186/s12711-017-0330-5.

Ma, L., J. R. O'Connell, P. M. VanRaden, B. Shen, A. Padhi, *et al.,* (2015).PLoS Genet. 11:e1005387. https://doi.org/10.1371/journal.pgen.1005387

Ott, J. 1999. Analysis of human genetic linkage. JHU Press.

Petes, T. D. (2001). Nat. Rev. Genet. vol. 2, no 5, p. 360-369.

Qanbari, S., and D. Wittenburg. (2020). Nat. Rev. Genet. 2:360–369. https://doi.org/10.1186/s12711-020-00593-z

Shen, C., X. Li, R. Zhang, and Z. Lin. (2017).Genet. Sel. Evol. 52:1–11. https://doi.org/10.1371/journal.pone.0188682

Úbeda, F., and J. F. Wilkins. 2011.Evol. Biol. 24:541–553. https://doi.org/10.1111/j.1420-9101.2010.02187.x.

664. Longevity in genetic evaluation of Holstein cows in Uruguay

N.F. Silveira*, O. Ravagnolo and I. Aguilar

Instituto Nacional de Investigación Agropecuaria (INIA), Montevideo, CP 11500, Uruguay; nferraz@inia.org.uy

Abstract

The objective of this study was to estimate variance components and to analyze the possibility of carrying out a genetic evaluation of the longevity trait in animals of the Uruguayan dairy production system of the Holstein breed. Longevity was defined as a binary trait with values of 1 and 0 whether or not the cow survived a lactation, respectively. For the analysis of the variance components, programs of the Blupf90 package were used, using the REML methodology and the Bayesian methodology. Heritability (with values between 0.03 to 0.13) was increasing according to the lactation. These results of the present study shows that even though it has a high influence of environmental factors, there is additive genetic variance available and therefore there is the possibility of inclusion of longevity in the genetic evaluation of Holstein cows in Uruguay.

Introduction

Stayability, is defined as the probability of a female to remain in production at a specific age, given that she had the opportunity to reach this age (Hudson and Van Vleck 1981). Therefore, the selection for longevity has the objective of identifying those individuals with the greatest capacity to remain productive, that is, with the greatest probability of survival due to voluntary and involuntary culling within the herd.

With the increase in the lifespan of the cow, it is possible to achieve a greater dilution of the maintenance and replacement costs of females, as demonstrated by Gabler, Tozer, and Heinrichs (2000). They showed that the highest lactation yield occurs after the third parturition, so keeping the animal in the system for a longer time will result in an increase in the dairy production of the establishment.

For the evaluation of longevity, different ways to define the trait to be included were proposed (Imbayarwo-Chikosi *et al.* 2015). The chosen form and the statistical model used depend on different factors related to the final objective of the analysis and the data available.

The objective of this study was to estimate variance components and analyze the possibility of carrying out a genetic evaluation of the longevity trait in animals of the Uruguayan dairy production system of the Holstein breed.

Materials & methods

Data. Records were available from the 'Instituto Nacional para el Control y Mejoramiento Lechero' (MU). The database was composed of milking cow's historical information, including reproduction, milk production and disposal information. Genealogical data was available from the 'Asociacion Rural del Uruguay' (ARU).

For variance components analyses, animals were required to have a known sire and dam and with the year of first calving between 1990 and 2014, giving the cow the possibility of completing at least five lactations. After editing the final dataset include records of 828,246 animals.

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_664, © N.F. Silveira *et al.* 2022

Trait definition. Longevity was defined as a binary trait with values of one and zero whether or not the cow survived the lactation, respectively. Thus, analyzing the probability that the animal remains in the dairy during the first five lactations.

Statistical model. Variance components analyses were performed using programs from the BLUPF90 package (Misztal *et al.*, 2002), using the REML methodology (REMLF90 and AIREMLF90) and the Bayesian analysis methodology through Gibbs sampling (GIBBS2F90, THRGIBBS1F90 and POSTGIBBSF90). Multiple trait linear models and non-linear models were analyzed for the binary longevity traits at different lactations.

The following multi-trait model used was:

$$y = Xb + Zu + Zp + e$$

Where y is the survival observation vector (1/0), X the incidence matrix that links the observations with the fixed effects, b the fixed effects vector (contemporary group effect, composed of herd-year of first calving and age at first calving), incidence matrix Z for the animal additive genetic effect, $u = N(0,G_0 \otimes A)$ with genetic covariance matrix (G_0) and additive genetic relationship matrix (A), p = vector corresponding to the permanent environment effect and the residual effect, $e = N(0,R_0 \otimes I)$, with a residual covariance matrix (R_0).

Breeding values were estimated for the linear and non-linear models in order to establish the correlations between the predictions and to asses the re-ranking of the animals according to the methodology used.

Results & discussion

Variance component analysis. The results of the variance components performed by the linear model (REML and Bayesian methodology) and nonlinear (Bayesian) are shown in Table 1, 2 and 3, respectively. The estimates of additive genetic variance for the first five lactations are highlighted, confirming that there is genetic variability for longevity, allowing genetic improvement for the trait.

Table 1. Additive genetic variance, heritability and genetic correlations for the first five lactations – linear (REML).

Additive genetic variance / genetic correlation[1]					Heritability
0.004	0.004	0.005	0.006	0.007	0.03
0.97	**0.004**	0.005	0.007	0.008	0.03
0.86	0.94	**0.007**	0.009	0.012	0.04
0.87	0.93	0.98	**0.011**	0.015	0.05
0.82	0.01	0.96	0.97	**0.020**	0.08

[1] Upper triangle: additive genetic variance; lower triangle: genetic correlation between the first 5 lactations.

Table 2. Additive genetic variance, heritability and correlations for the first 5 lactations – linear (Gibbs sampling).

Additive genetic variance / genetic correlations[1]					Heritability
0.004	0.004	0.006	0.009	0.012	0.03
0.93	**0.005**	0.008	0.011	0.015	0.03
0.89	0.94	**0.012**	0.017	0.023	0.06
0.86	0.89	0.96	**0.027**	0.034	0.09
0.87	0.90	0.95	0.93	**0.049**	0.13

[1] Upper triangle: additive genetic variance; lower triangle: genetic correlation between the first 5 lactations.

Table 3. Additive genetic variance, heritability and correlation for the first 5 lactations – nonlinear.

Additive genetic variance / genetic correlations[1]					Heritability
0.313	0.225	0.188	0.225	0.245	0.06
0.92	**0.197**	0.153	0.198	0.220	0.06
0.88	0.89	**0.152**	0.199	0.199	0.07
0.78	0.85	0.94	**0.298**	0.282	0.08
0.77	0.85	0.86	0.88	**0.341**	0.09

[1] Upper triangle: additive genetic variance; lower triangle: genetic correlation between the first 5 lactations.

Heritability increased according to the lactation analyzed, showing that the survival of the cow at a greater number of lactations depends more on genetic components. The methodologies present similar values for heritability and high genetic correlation values between the lactations.

The estimated values for heritability by the different models were similar to previous studies that showed values between 0.03 and 0.25 (Essl and Voith 2002; Jamrozik, Fatehi, and Schaeffer 2008; Sewalem *et al.* 2005).

The correlations obtained in the present study were similar with the previous studies for dairy cattle from Germany and Canada, (Heise *et al.* 2016 and Sewalem *et al.* 2007, respectively), finding a high correlation between the different lactations.

The high correlation between traits will allow prediction of longevity at later ages from analysis carried out using the first lactations.

Correlations of the breeding values. The breeding values for the linear and non-linear model were estimated using BLUPF90 and THR1GIBBSF90 respectively, using the estimated variance components. Reliabilities were calculated using prediction error variance from the inverse of the mixed model (linear model) and from the posterior standard deviation of breeding values (non-linear models).

The correlation between the breeding values obtained through the linear model and the non-linear model for the first five lactations were greater than 0.98, which will not cause re-ranking of animals from the alternative genetic evaluation models. The correlation of the reliabilities between linear and non-linear models were greater than 0.99. No difference was found between the use of statistical analysis by direct analysis (blup) and analysis by Gibbs sampling (Bayesian).

Considering the computational cost and the greater complexity of performing a non-linear model and considering the correlation between the different models, the linear model can be used to estimate genetic values.

These results of the present study shows that even though it has a high influence of environmental factors, there is additive genetic variance available and therefore there is the possibility of inclusion of longevity in the genetic evaluation of Holstein cows in Uruguay.

References

Essl, A. and Voith K. 2002. *Journal of Animal Breeding and Genetics* 119(3):182–89. https://doi.org/10.1046/j.1439-0388.2002.00334.x

Gabler, M.T., Tozer P.R., and Heinrichs A.J.. 2000. *Journal of Dairy Science* 83(5):1104–9. https://doi.org/10.3168/jds.S0022-0302(00)74975-7

Heise J., Liu Z., Stock K.F., Rensing S., Reinhardt F., et. al. 2016. *Journal of Dairy Science* 99(2):1253–65. https://doi.org/10.3168/jds.2015-10163

Hudson, G.F.S. and Van Vleck L.D. 1981. *Journal of Dairy Science* 26:2246–50. https://doi.org/10.3168/jds.S0022-0302(81)82836-6

Imbayarwo-Chikosi, V.E., Dzama K., Halimani T.E., Van Wyk J.B., Maiwashe A., et. al. 2015. *South African Journal of Animal Sciences* 45(2):105–21. http://dx.doi.org/10.4314/sajas.v45i2.1

Jamrozik, J., Fatehi J., and Schaeffer. 2008 L.R.. *Journal of Animal Breeding and Genetics* 125(2):75–83. 'https://doi.org/10.1111/j.1439-0388.2007.00712.x'

Sewalem, A., Kistemaker G.J., Ducrocq V., and Van Doormaal B.J. 2005. *Journal of Dairy Science* 88(1):368–75. https://doi.org/10.3168/jds.S0022-0302(05)72696-5

Sewalem, A., Miglior F., Kistemaker G.J., Sullivan P., Huapaya G., et. al. 2007. *Journal of Dairy Science* 90(4):2025–28. https://doi.org/10.3168/jds.2006-719

665. Comparison of two models for dairy herd improvement data in Taiwan

Y.H. Lan[1*], K.H. Lin[1], S.H. Ting[1], C.L. Su[2], P.A. Tu[3] and E.C. Lin[1]

[1]Dept. of Animal Science & Technology, National Taiwan University, No. 50, Ln. 155, Sec. 3, Keelung Rd., 10673 Taipei, Taiwan; [2]Dairy Association Taiwan, 3F, No. 177-4, Sec. 1, Heping E. Rd., 10644 Taipei, Taiwan; [3]Hsinchu Branch, Livestock Research Institute, Council of Agriculture, No. 207-5, Bitomein, Wuhu Village, Xihu Township, 36843 Miaoli County, Taiwan; jeff10688@gmail.com

Abstract

The objective of this study is to establish a model suitable for the Dairy Herd Improvement (DHI) records under tropical/subtropical climate in Taiwan. There are 62,068 lactation records of 28,722 cows obtained from Taiwan the DHI program during 2009~2019. A model with the physical situation of lactation (*PSL*) as a fixed effect which accounts for abnormal physiology and subclinical mastitis is compared to a model without that effect. The heritability of milk yield and somatic cell score are decreased and the number of sires with the accuracy of estimated breeding values greater than 75% would also change in the model with *PSL* as a fixed effect. Moreover, AIC and BIC in the model with *PSL* are lower than the model without *PSL*, which indicates the former model is more appropriate for this DHI dataset.

Introduction

Taiwan is located in the area under tropical/subtropical marine climate with high temperature and humidity. Such climate has a great impact on milk production, milk quality, reproduction and health of dairy cows. Mastitis is the most common disease in dairy cows with great impact on milk production and profit of dairy farm. There were 14.7% of cows with clinical mastitis in Taiwan (Chen *et al.*, 2013). Besides, there are approximately 10% of lactation records of cows with subclinical mastitis according to Taiwan DHI program. A metabolic disease, ketosis, is mainly caused by systemic dysfunction resulted from carbohydrate and volatile fatty acid metabolism disorders in the body. The incidence of clinical ketosis accounts for 18% of postpartum illnesses in Taiwan, which mainly occurs between 2-7 weeks postpartum (Chen *et al.*, 2008). Because considerable proportion of dairy cows in Taiwan may have abnormal physiology of lactation, it should be considered in the model of genetic evaluation for the DHI data in Taiwan. The objective of this study is to establish a model suitable for the DHI records under tropical/subtropical climate.

Materials & methods

Data. The milk production data and pedigree during 2009~2019 were collected from the Taiwan DHI program. Those records having: sire ID in blank or not identifiable; phenotypic and tested values collected less than 106 in the first lactation and 136 days in the later lactations were eliminated. Thus, 62,068 lactation records of 28,722 cows with 2,356 sires and 39,965 animals in the pedigree were obtained. Production traits include milk yield (MY), milk fat percentage (MF%), milk fat yield (MFY), milk protein percentage (MP%) and milk protein yield (MPY) are standardized to 305d-2X yield. Somatic cell count is transformed to somatic cell score (SCS) to achieve normality of this trait and homogeneity of its variance (Ali and Shook, 1980).

Statistical analysis. Based on the above data and pedigree of lactating cows, the following multiple-trait animal models were used to estimate the variance and covariance components as well as heritability and correlation. The first model is shown as following:

$$y_{ijlm} = \mu + HYS_i + L_j + pe_l + a_{ijlm} + e_{ijlm} \quad (1)$$

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_665, © Y.H. Lan *et al.* 2022

where, y_{ijlm} is the record for each trait of the cow; μ is the overall population mean; HYS_i is the fixed effect of i^{th} herd-year-season; L_j is the fixed effect of j^{th} lactation; pe_l is the random permanent environmental effect; a_{ijlm} is the random additive genetic effect; and e_{ijlm} is the random residual effect.

In addition to the effects in the model (1), an extra fixed effect, PSL_k, is included for physical situation of lactation k, where k=1 represents the cows might have subclinical mastitis (average lactation SCS>4.585 or somatic cell counts $>3\times10^5$), k=2 represents the cows might have abnormal physiology (305d-2X milk yield <5,713 kg), k=3 represents the cows might have subclinical mastitis and abnormal physiology simultaneously, and k=4 represents the cows were in normal physiological situation. The second model is shown as following:

$$y_{ijklm} = \mu + HYS_i + L_j + PSL_k + pe_l + a_{ijklm} + e_{ijklm} \quad (2)$$

The number of lactation records for PSL_k with k=1~4 were 6,249 (10.0%), 4,948 (8.0%), 1,261 (2.0%) and 49,610 (80.0%), respectively. The statistical analyses were conducted using ASReml software 4.1 (Gilmour *et al.*, 2009).

Results

The estimates of heritability, genetic correlations and phenotypic correlations for different traits using model (1) are shown in Table 1. The heritability estimated for MY, MF%, MFY, MP% and MPY are 0.249, 0.588, 0.284, 0.504 and 0.275, respectively. These estimates are similar to those obtained from DHI data analysis in different countries (Cassell, 2009; Roman *et al.*, 2000; Suzuki and Van Vleck, 1994). However, the heritability of SCS estimated by the model (1), 0.274, is higher than the usual estimates, but few studies had similar estimates to those from the model (1) (Roman *et al.*, 2000; Sabedot *et al.*, 2018). The estimates of genetic correlation among different traits are similar to those from several studies except for those negative ones between SCS and production traits (Rincón *et al.*, 2015; Sabedot *et al.*, 2018).

However, when a physical situation of lactation (PSL_k) was used as a fixed effect in the model (2), estimates of heritability for SCS, MY and MPY were significantly decreased 46.7, 12.0 and 9.5%, respectively (Table 2). It may due to the fixed effect which accounts for the abnormal lactation physiology and subclinical mastitis in Taiwan dairy herds. Those estimates of genetic correlation between SCS and production traits changed from 26.5% (SCS and MFY) to 227.8% (SCS and MP%). The estimate of genetic correlation between SCS and MY is closed to zero, which is similar to some studies (Kheirabadi and Razmkabir, 2016; Rincón *et al.*, 2015).

Table 1. Estimates with standard errors of heritability (on diagonal), genetic correlations (above the diagonal) and phenotypic correlations (below the diagonal) using model (1) from multiple-trait model analyses.[1]

	MY	MF%	MFY	MP%	MPY	SCS
MY	0.249±0.008	-0.372±0.017	0.550±0.015	-0.268±0.018	0.849±0.006	-0.181±0.023
MF%	-0.288±0.005	0.588±0.007		0.632±0.009		-0.116±0.018
MFY	0.782±0.002		0.284±0.008		0.726±0.010	-0.264±0.022
MP%	-0.342±0.005	0.507±0.004		0.504±0.008		-0.036±0.019
MPY	0.931±0.001		0.827±0.002		0.275±0.008	-0.198±0.023
SCS	-0.214±0.005	0.005±0.006	-0.203±0.005	0.104±0.005	-0.187±0.005	0.274±0.008

[1] MY = milk yield; MF% = milk fat percentage; MFY = milk fat yield; MP% = milk protein percentage; MPY = milk protein yield; SCS = somatic cell score.

Table 2. Estimates with standard errors of heritability (on diagonal), genetic correlations (above the diagonal) and phenotypic correlations (below the diagonal) using model (2) from multiple-trait model analyses.[1]

	MY	MF%	MFY	MP%	MPY	SCS
MY	0.219±0.008	-0.366±0.018	0.378±0.020	-0.270±0.019	0.759±0.009	-0.026±0.032
MF%	-0.253±0.005	0.590±0.007		0.625±0.009		-0.191±0.023
MFY	0.708±0.003		0.291±0.008		0.633±0.013	-0.194±0.028
MP%	-0.281±0.005	0.495±0.004		0.526±0.008		-0.118±0.024
MPY	0.902±0.001		0.771±0.002		0.249±0.008	-0.091±0.030
SCS	-0.134±0.005	-0.014±0.005	-0.128±0.005	0.053±0.005	-0.114±0.005	0.146±0.007

[1] MY = milk yield; MF% = milk fat percentage; MFY = milk fat yield; MP% = milk protein percentage; MPY = milk protein yield; SCS = somatic cell score.

Discussion

An additional fixed effect of PSL_k in the model (2) would reduce the accuracy of estimated breeding value (EBV) in some traits. The number of sires with accuracy of EBV greater than 75% decreased in MY, MF%, MPY and SCS, with 7.37, 0.33, 5.34 and 37.06%, respectively (Table 3). The degree of decrease in accuracy of EBV would be in accordance with the decrease of heritability.

The percentage of changes in the rank of those sire EBV with accuracy greater than 75% are 49.5, 43.1 and 80.1% for MY, MPY and SCS, respectively (Figure 1). For analyzing data in the tropics/subtropics climate, it is necessary to reconsider the effects in the model, otherwise the accuracy of EBV for sires might be overestimated.

The Akaike Information Criteria (AIC) and Bayesian Information Criteria (BIC) of MY in the model (1) are 920,993.11 and 921,020.05, while those in the model (2) are 900,280.55 and 900,307.49. The AIC and BIC of SCS in the model (1) are 96,309.85 and 96,336.80, while those in the model (2) are 60,845.68 and 60,872.62. It shows that the AIC and BIC in Model (2) are all smaller than those in the model (1). The decreases of AIC and BIC for both MY and SCS show that the model (2) might be a more appropriate model for this DHI dataset.

Conclusions

The model (2) with the fixed effect, PSL_k, accounts for abnormal physiology and subclinical mastitis in dairy cows, has a lower AIC and BIC than the model (1) without the fixed effect, which indicates the model (2) is better in the analysis. The heritability of MY and SCS are decreased largely than other traits, and the number of sires with accuracy of EBV greater than 75% also reduced. To analyze DHI data in tropical/subtropical climate, we need to be more careful in the effects accounting for physical situation of lactation in the model, otherwise, it might lead to some bias in estimation of genetic parameters and breeding values.

Table 3. Number of sires with the accuracy of estimated breeding values greater than 75% for different traits in the two models.[1]

	MY	**MF%**	**MFY**	**MP%**	**MPY**	**SCS**
Model (1)	543	906	609	858	580	564
Model (2)	503	903	627	864	549	355
Change	-7.37%	-0.33%	2.96%	0.70%	-5.34%	-37.06%

[1] MY = milk yield; MF% = milk fat percentage; MFY = milk fat yield; MP% = milk protein percentage; MPY = milk protein yield; SCS = somatic cell score.

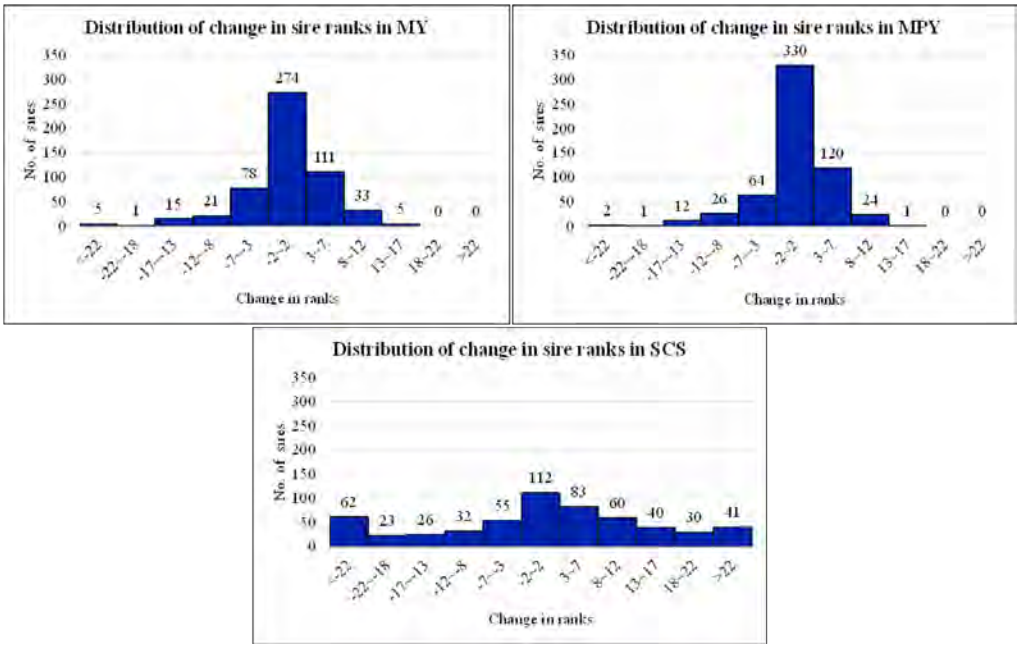


Figure 1. The amount of changes in EBV rank of sires in different traits from the model (1) to model (2).

References

Ali A. K. A., and Shook G. E. (1980). An optimum transformation for somatic cell concentration in milk. J. Dairy Sci. 63(3): 487-490.

Cassell B. (2009). Using Heritability for Genetic Improvement. Virginia Cooperative Extension Publication, pp. 404-084.

Chen J. Y., Lee K. H., and Chang C. L. (2013). Study on risking factors of production life for DHI. Project Report of Livestock Research Institute. Available at: https://www.grb.gov.tw/search/planDetail?id=2518086.

Chen S. K., Chan J. P.W., and Chuang S.T. (2008). Investigation of Postpartum Diseases in Dairy Cows in Taiwan. Taiwan Veterinary Journal. 34(1): 41-48.

Gilmour A. R., Gogel B. J., Cullis B. R., and Thompson R. (2009). ASReml User Guide Release 3.0. VSN International Ltd, Hemel Hempstead, UK.

Kheirabadi K. and Razmkabir M. (2016). Genetic parameters for daily milk somatic cell score and relationships with yield traits of primiparous Holstein cattle in Iran. J. Anim. Sci. Technol. 58: 38.

Rilanto T., Reimus K., Orro T., Emanuelson U., Viltrop A., and Mõtus K. (2020). Culling reasons and risk factors in Estonian dairy cows. BMC Vet. Res. 16(1):1-16.

Rincón J. F., Zambrano J. A., Echeverri J. (2015). Estimation of genetic and phenotypic parameters for production traits in Holstein and Jersey from Colombia. Rev. MVZ Córdoba 20(Supl.): 4962-4973.

Roman R. M., Wilcox C. J. and Martin F. G. (2000). Estimates of repeatability and heritability of productive and reproductive traits in a herd of Jersey cattle. Genet. Mol. Biol. 23(1): 113-119.

Sabedot M. A., Romano G. S., Pedrosa V. B., and Pinto L. F. B. (2018). Genetic parameters for milk traits, somatic cell, and total bacteria count scores in Brazilian Jersey herds. R. Bras. Zootec. 47: e20160351.

Suzuki M. and Van Vleck L. D. (1994). Heritability and repeatability for milk production traits of Japanese Holsteins from an animal model. J. Anim. Sci. 77(2): 583-58

666. Comparison of random regression test-day models for production traits of South African Jersey Cattle

M.G. Kinghorn[1*], E.D. Cason[1], V. Ducrocq[2] and F.W.C. Neser[1]

[1]University of the Free State, Bloemfontein, South Africa; [2]Université Paris Saclay, INRAE, AgroParisTech, GABI, 78350 Jouy-en-Josas, France; matthew.kinghorn566@gmail.com

Abstract

This study aims to identify the most appropriate random regression test-day model (RRTDM) for the South African Jersey dairy breed, using Legendre Polynomials (LP) to model the lactation curve. The data comprised of production test-day records for milk yield (MY), fat yield (FY) and protein yield (PY) from 33,852 multiparous cows. Genetic parameters were estimated using single trait RRTDM using AIREML90. The fixed effects fitted were Herd × Test-day × Birth Year of the Cow × Calving year × Parity × Number of Milkings (HTDBCPNM), production system (PROD) (TMR, pasture or both), age at first calving (AFC) and inter-calving period (ICP), and fixed regressions on age class-season-district. Random effects were direct additive, permanent environmental and residual effects. The estimated heritabilities for MY, FY and PY were 0.254, 0.182 and 0.170, respectively. The genetic correlations between total milk production and persistency were low (0.06 to 0.24). Overall, the heritability estimates of all the traits analysed were low to moderate.

Introduction

The accurate estimation of genetic parameters and (co)variance components associated with RRTDM is important for the estimation of accurate breeding values (Danell, 1982). The genetic parameter estimates obtained are largely dependent on the regression functions used to model the lactation curve, as well as the covariance structure for additive, permanent environmental, and residual effects (Bignardi *et al.*, 2011). The most common functions used to model the lactation curve include the ones proposed by Ali and Schaeffer (1987), Wilmink (1987), splines (White *et al.*, 1999), and orthogonal polynomials (Kirkpatrick *et al.*, 1990). Parametric functions such as those of Ali and Schaeffer (1987) and Wilmink (1987) have mostly been abandoned because the estimated covariance matrices show high correlations between random regression coefficients which can hinder the estimation process (Schaeffer, 2004). Therefore, orthogonal polynomials such as the Legendre Polynomials (LP) introduced by Kirkpatrick are more commonly used since their parameters tend to be less correlated than the aforementioned functions (Guo *et al.*, 2002), and they provide estimates of covariance matrices that are more robust over different datasets (Schaeffer, 2004). This study makes use of LP to estimate lactation curves, based on additional benefits outlined by Pool (2000).

In South Africa, dairy cattle breeding values are estimated for production traits based on a fixed regression test-day model (Mostert, 2007). Dzomba *et al.* (2010), outlines how the current South African model is criticised for assuming a standard lactation curve for cows in similar contemporary groups, homogenous additive genetic variances across lactations, and for its inability to account for persistency. The adoption of a RRTDM would overcome these limitations and has other advantages such as improved accuracy of breeding value estimates, extra possibility to decrease generation interval, and consequently improved genetic response to selection. This study aims to identify the most appropriate RRTDM for the South African Jersey dairy breed, using Legendre Polynomials to model the lactation curve.

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_666, © M.G. Kinghorn *et al.* 2022

Materials & methods

Data ranging from 1988 to 2017 were obtained from the South African Jersey Association. A total of 429,940; 420,691 and 420,691 test-day records were analysed for MY, FY, and PY respectively. The data analysed was from 33,852 multiparous cows, of 1,558 sires, in 51 herds located across South Africa. Age at calving ranged between 15-36, 26-53, and 38-67 months for first, second, and third parity respectively (Interbull, 2015). Test-day yields from 5 to 305 days of lactation were considered. Lactations included at least five test-day records per cow. Contemporary groups were defined as a concatenation of Herd × Test-day × Birth Year of the Cow × Calving year × Parity × Number of Milkings (HTDBCPNM), with a total number of 12,033 groups containing at least 10 individuals and 2 sires per group. Only animals containing information on production system (i.e. pasture, total-mixed ration, or a combination of these systems) were included. Test-day yields for milk, fat, and protein were adjusted according to the ICAR guidelines (2017).

Analyses were performed using single trait RRTDM, including the direct additive, permanent environmental, and residual effects. The fixed effects included in the model were contemporary group (as defined above), PROD, AFC, and ICP, and age average lactation curve modelled with fifth-order LP within age class × season × district classes where age classes were assigned per lactation according to Mostert (2007). The chosen statistical model is as follows:

$$y_{tijklmn} = \mathrm{HTDBCPNM}_i + \mathrm{PROD}_j + \mathrm{AFC}_k + \mathrm{ICP}_l + \sum_{r=1}^{5} ASD_{mn0}\, x_n + \sum_{r=1}^{5} a_{mn} x_n + \sum_{r=1}^{5} pe_{mn} x_n + e_{tijklmn} \quad (1)$$

where $y_{tijklmn}$ is the t^{th} test day record (milk yield, fat and protein contents) of the m^{th} cow in i^{th} contemporary group (HTDBCPNM) effect, j^{th} production system and k^{th} age at first calving, l^{th} inter-calving period; ASD_{mn0} is the n^{th} fixed regression coefficient of the o^{th} class of cows age class-season-district; a_{mn} and pe_{mn} are regression coefficients n^{th} for additive genetic and permanent environmental effects on m^{th} cow respectively; x_n is the n^{th} Legendre Polynomial for the t^{th} day; $e_{tijklmn}$ random residual effect associated with $y_{tijklmn}$.

For each trait, a single trait RRTDM was run using Average Information Restricted Maximum Likelihood methodology with the AIREMLF90 program (Misztal *et al.*, 2002). Numerous models with varying fixed effects were tested, and AIC values and variance components were compared for each model.

Results

The first coefficient of LP can be related to the total yield in each lactation and the second regression coefficient can be viewed as a persistency measure. The first two coefficients tend to account for most of the genetic variation (Strabel and Jamrozik, 2006). The first two coefficients accounted for approximately 94% of the genetic variation for MY, whereas for FY and PY the first four coefficients accounted for approximately 90% of the genetic variation, only results of the first two coefficients shown in Table 1.

Discussion

Assessment of AIC values showed that Model 1 produced the lowest values, and thus best fit the data for all traits (results not shown). The first random regression coefficient (intercept) of orthogonal polynomials describes the total yield in lactation, thus the heritability of this coefficient is equal to the heritability of 305-day lactation yield. The second random regression coefficient (linear) can be viewed as a persistency measure (Strabel and Jamrozik, 2006). The largest genetic variances were associated with the intercept and linear coefficients, while all other variances (not shown) were close to zero for all traits. The genetic correlation for MY and persistency was low and positive (0.1669), however not as low as the genetic correlation between total FY and persistency of fat production (0.0678). The genetic correlation between PY and persistency of protein production, although still lowly correlated (0.2410), appeared to have the strongest genetic relationship between the first two random regression coefficients.

Table 1. Genetic variances of the random coefficients (diagonal, bold face), correlations (upper diagonal), covariances (lower diagonal), phenotypic variances (σ_p), residual error variances (σ_r) and direct ratios.

Direct additive effect					
	Random regression coefficients		σ_P	σ_r	h^2
Trait	**0**	**1**			
Milk yield	5.4297	0.1669	21.4090	6.1314	0.2536
	0.4757	1.4957			0.0699
Fat yield	0.0167	0.0678	0.0920	0.0300	0.1816
	0.0008	0.0073			0.0792
Protein yield	0.0163	0.2410	0.0960	0.0100	0.1700
	0.0034	0.0123			0.1286
Permanent environmental effect					
	Random regression coefficients		σ_P	σ_r	**Ratio**
Trait	**0**	**1**			
Milk yield	7.1516	0.0444	21.4090	6.1314	0.3341
	0.0825	0.4829			0.0226
Fat yield	0.0117	0.0660	0.0920	0.0300	0.1272
	0.0005	0.0047			0.0511
Protein yield	0.0102	0.2398	0.0960	0.0100	0.1072
	0.0021	0.0075			0.0785

The variances obtained in this study were different to those obtained by Interbull (2015) for the South African Jersey population, likely do due to differences in the model, as well as the data used. Despite this, the heritabilities for the direct additive effects appeared to be similar to those obtained in this study; 0.23, 0.14 and 0.20 for MY, FY and PY respectively. The Interbull variance ratios for permanent environmental effects also appeared to be similar, ranging between 0.27-0.31, 0.13-0.18 and 0.24-0.26 for MY, FY and PY respectively. Heritabilities of 0.18 and 0.11, and variance ratios of 0.44 and 0.88 for the first two random regression coefficients for total milk production were obtained by Strabel and Jamrozik (2006). The same study also found a genetic correlation between the intercept and linear coefficients of first lactation milk production to be 0.06. Other studies have obtained heritabilities of 0.23-0.25 for milk yield (Li *et al.*, 2020). Variance ratios obtained by a fixed regression model used on South African Jersey data were 0.39, 0.21 and 0.34 for additive genetic effects of milk, fat and protein respectively, and 0.23,0.16 and 0.20 for permanent environmental effects of milk, fat and protein yields respectively. The current South African test-day model is a fixed regression model, and therefore does not have a comparative genetic measure for persistency of production traits (Dzomba *et al.*, 2010).

References

Ali, T. E., & Schaeffer, L. R. 1987. Accounting For Covariances Among Test Day Milk Yields In Dairy Cows. Can. J. Anim. Sci. 67, 637–644.

Bignardi, A. B., El Faro, L., Torres Junior, R. A. A., Cardoso, V. L., Machado, P. F., *et al.,* 2011. Random regression models using different functions to model test-day milk yield of Brazilian Holstein cows. Genetics and Molecular Research 10, 3565–3575 http://dx.doi.org/10.4238/2011.October.31.4.

Danell, B. 1982. Studies on lactation yield and individual test-day yields of Swedish dairy cows. Acta. Agric. Scand. 32, 82–92 https://doi.org/10.1080/00015128209435733.

Dzomba, E. F., Nephawe, K. A., Maiwashe, A. N., Cloete, S. W. P., Chimony, M., *et al.*, 2010. Random regression test-day model for the analysis of dairy cattle production data in South Africa: Creating the framework. S.Afr.J.Anim.Sci. 40, 273–284.

Guo, Z., Lund, M. S., Madsen, P., Korsgaard, I., & Jensen, J. 2002. Genetic Parameter Estimation for Milk Yield over Multiple Parities and Various Lengths of Lactation in Danish Jerseys by Random Regression Models. J Dairy Sci 85, 1596–1606.

ICAR. 2017. Section 2 - Guidelines for Dairy Cattle Milk Recording. ICAR The Global Standard For Livestock Data. Available at: https://www.icar.org/Guidelines/02-Overview-Cattle-Milk-Recording.

Interbull. 2015. Description of National Genetic Evaluation Systems. Interbull.Available at: https://interbull.org/ib/geforms.

Kirkpatrick, M., Lofsvold, D., & Bulmer, M. 1990. Analysis of the Inheritance, Selection and Evolution of Growth Trajectories. Genetics 124, 979–993.

Li, J., Gao, H., Madsen, P., Li, R., Liu, W., *et al.*, 2020. Impact of the Order of Legendre Polynomials in Random Regression Model on Genetic Evaluation for Milk Yield in Dairy Cattle Population. Frontiers in Genetics 11, 1–8. https://doi.org/10.3389/fgene.2020.586155.

Misztal, I., Tsuruta, S., Strabel, T., Auvray, B., Druet, T., *et al.* 2002. BLUPF90 and Related Programs. Proc. 7th WCGALPP, Montpellier, France.

Mostert, B. E. 2007. The suitability of test-day models for genetic evaluation of dairy cattle in South Africa. PhD Thesis, University of Pretoria, South Africa http://hdl.handle.net/2263/25127.

Pool, M. H. 2000. Test-day Models: Breeding value estimation based on individual test-day records.Universal Press, Veenendaal, the Netherlands.

Schaeffer, L. R. 2004. Application of random regression models in animal breeding. Livest. Prod. Sci. 86, 35–45. https://doi.org/10.1016/S0301-6226(03)00151-9.

Strabel, T., & Jamrozik, J. 2006. Genetic Analysis of Milk Production Traits of Polish Black and White Cattle Using Large-Scale Random Regression Test-Day Models. J. Dairy Sci. 89, 3152–3163.https://doi.org/ 10.3168/jds.S0022-0302(06)72589-9

White, I. M. S., Thompson, R., & Brotherstone, S. 1999. Genetic and Environmental Smoothing of Lactation Curves with Cubic Splines. J Dairy Sci 82, 632–638. https://doi.org/10.3168/jds.S0022-0302(99)75277-X.

Wilmink, J. B. M. 1987. Adjustment of Lactation Yield for Age at Calving in Relation to Level of Production. Livest. Prod. Sci. 16, 321–334. https://doi.org/10.1016/0301-6226(87)90002-9.

667. ROH describes the small holder dairy systems with Holstein cattle in Mexico

A. García-Ruiz[1*], J.G. Cortés-Hernández[2] and F.J. Ruiz-López[1]

[1]National Institute for Forestry, Agriculture and Livestock Research (INIFAP), Km 1 carretera Ajuchitlán – Querétaro, Qro. CP 76280, México; [2]National Autonomous University of Mexico (UNAM) Av. Universidad #3000, Colonia, C.U., Coyoacán, 04510 México City, México; garcia.adriana@inifap.gob.mx

Abstract

The objective of this study was to identify traits that have been incorporated into the selection processes in Holstein cows of the small holder production system (SHS) through the presence of Runs of Homozygosity (ROH) in the genome. The study included 214 animals genotyped with the GeneSeek GGP 50K® chip. Minimum ROH length and size were established (500 kb and 25 SNP), allowing 5 missing genotypes and one heterozygous. The most frequent ROHs were identified and annotations and genes identified in other populations were obtained. 10,798 ROH were detected with an average, minimum and maximum length of 4.79, 0.43 and 91.90 Mb, respectively. In the region where the most frequent ROH were found in the population (≥20 animals), QTL and genes have been reported that are related to milk production and composition, fertility parameters, susceptibility to diseases, body conformation, feed efficiency and some characteristics of the carcass composition.

Introduction

Signatures of selection are regions of the genome that have been conserved for generations in populations due to natural or artificial selection and are related to functionally important traits. Their detection helps to identify candidate genes that have been favored in selection processes to which populations have been exposed naturally or induced, and to identify beneficial mutations.

Among the methods to identify signatures of selection, those that are formed by the reduction of genetic variation stand out, such as, for example, runs of homozygosity (ROH), which are contiguous regions of homozygous genotypes that occur within an individual when two haplotypes share a common ancestor, those formed in recent generations being shorter.

The formation of ROH in the genome is influenced, among other factors, by the degree of intensity of selection in the populations, which in turn is determined by the availability of replacements and the adoption of technological and reproductive tools, such as artificial insemination (AI) and embryo transfer (ET). In dairy cattle, the intensity of selection is very high and the selection of genetic material is restricted to a limited number of families, a factor that contributes to the formation of ROH.

In Mexico, more than 30% of the national milk production was produced in the small holder system (SHS), which includes approximately 78% of dairy farms. Holstein cattle predominates in the small holder farms, although Brown Swiss animals can be found, as well as their crosses (Sainz-Sánchez *et al.*, 2016). Currently there is no reliable information on production records per animal in the SHS and on rare occasions it is possible to collect genealogical information, which makes it very difficult to carry out genetic evaluations of the animals found in this system. The genetic improvement of these animals has been carried out by the selection made by the farmer within his herd or by the genetic material introduced as AI semen, but there is no evidence of matings directed with a specific selection purpose, or which traits have been a priority at the time of selection. Genomic tools could be used to identify possible selection signatures that would allow identification of the priorities in the SHS.

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_667, © A. García-Ruiz *et al.* 2022

Materials & methods

Genotypes and quality control. Genotypes of 214 Holstein cows from four SHS herds located in the region of Tepatitlán Jalisco, Mexico were used. Animals were genotyped with the GeneSeek GGP 50K® chip, which contains 43,285 single nucleotide polymorphism markers (SNP). Genotypes were filtered by genomic quality control which consisted of excluding animals with a call rate <0.90; exclude SNP with minor allele frequency (MAF) <0.02, or a call rate <0.95, or with a Hardy-Weinberg *P*-value<0.0001.

Identification of signatures of selection. Signatures of selection were detected by the presence of runs of homozygosity (ROH) in the genome. ROH runs that had a minimum length of 500 kb and a minimum number of 25 SNP were included. In addition, 1 heterozygous SNP and 5 missing genotypes were allowed per run. With those parameters, the risk on including short ROH was avoided (Purfield *et al.*, 2012). ROH analysis was performed with the bioinformatics platform SNP & Variation Suite v7.6.8 Win64 (Golden Helix, Bozeman, MT, USA).

Frequent ROH were classified as those that were present in 20 animals or more of the population and, according to their physical position in the genome, annotations in the genome that showed their association with genes were searched. In order to find genomic information that could help identify possible traits that have been selected in the population of the SHS in Mexico, QTLs or traits of economic interest found in the Animal QTLdb Release 43 database were examined (Zhi-Liang *et al.*, 2019).

Results

In the population studied, 10,798 ROH were detected with an average length of 4.79 Mb, finding a maximum length of 91.90 Mb and a minimum of 0.43 Mb. 70.65% of the ROH were unique in the population; that is to say, they were found in a single animal; while 29.35% (3,169) were found in more than one animal, some with the same length (identical) (10.42%) and others with variable lengths (18.93%), coinciding only in the final position.

The amount of ROH was lower compared to that reported in animals which come from the specialized production system in Mexico, indicating a lower intensity of selection. For example, in Holstein cattle from the intensive production system in Mexico, a total of 88,529 ROH were found, although the population size was larger (4,432 animals), the average length was greater than 8.95 Mb and a proportion of the genome covered of 179.57 Mb (Cortés-Hernández *et al.*, 2021). In other studies, in Holstein cattle from the specialized system in USA, an average of 299.6 Mb of the covered genome by ROH was reported (Forutan *et al.*, 2018).

The most frequent ROH in the population (≥20 animals) were found on chromosomes 1, 2, 10, 13 and 15, in regions where QTL and genes related to milk production and composition have been reported, as have traits related to fertility, susceptibility to diseases, body conformation, feeding efficiency and some carcass composition traits (Animal QTLdb Release 43; Zhi-Liang *et al.*, 2019; Chen *et al.*, 2020). To search for ROH annotations with variable length, the shortest ROH with respect to the final position was taken as a reference (Table 1).

Table 1. Genome annotations found in the regions where the most frequent ROH were detected in the small holder dairy population.[1]

Chromosome	Initial position (pb)	Minor length (pb)	Associations / QTL	Gene ID associated
1	156,777,075	2,078,048	ADG, CALEASE, CONCEPT, CONCRATE, DPR, DYF, FANG, FP, FSC, FY, MAINEF, MATRATE, MCIEB, MKCASP, NM, PL, PP, PY, PY, RFI, RLEGS, SB, SCS, STR, T3, TLGTH y UC	
2	84,235,779	1,824,453	BVDV, CONCRATE, MUGKCASP, MSPD, CONCEPT, BTBS, FY, BD, CALEASE, PTAT, FTPL, UA, NM, PL, RTPL, UHT, RUMWD, SCS, SB, STA, UC, UDPTH, LMY, BW, EY, MARBL, CWT, MTBR, NRR y BRDS	504677, 526800, 521004, Candidate: 19122, Candidate: 526800
10	102,405,153	1,848,427	BWG, CONCEPT, CONCRATE, FA-C22:6, FSC, FY, MAG, MCREAT, MFA-C18:0, MFA-C4:0, MPOT, MUGKCASP, MY, PHOS, SF y SOD	517053, 787307, 616909, 540780, 540232
13	83,153,486	1,076,496	BW, CALEASE, CONCRATE, DMI, FY, HPG, MFA-C6:0, MY y SCS	

[1] ADG: Average Daily Gain, BD: Body Depth, BRDS: Susceptibility to Bovine Respiratory Disease, BTBS: Susceptibility to Bovine Tuberculosis, BVDV: Susceptibility to Bovine Viral Diarrhea, BW: Body Weight at Birth, BWG: Weight Gain, CALEASE: Calving ease, CONCEPT: Number of inseminations per conception, CONCRATE: Conception rate, CWT: Carcass weight, DMI: Dry matter consumption, DPR: Daughter pregnancy rate, DYF: Dairy character, EY: Energy milk production, FA-C22: 6: Docosahexaenoic acid content, FANG: Angle of the hoof, FP: Percentage of fat in milk, FSC: First service conception, FTPL: Anterior teats placement, FY: Production of fat in milk, HPG: Heifer pregnancy, LMY: Lean meat yield, MAG: Muscle magnesium content, MAINEF: Maintenance efficiency, MARBL: Marbling score, MATRATE: Maturity rate, MCIEB: Meat color, MCREAT: Muscle creatinine content, MFA-C18: 0 : Milk stearic acid content, MFA-C24: 0: Milk tetracosanoic acid content, MFA-C4: 0: Milk butyric acid content, MFA-C6: 0: Milk caproic acid content, MKCASP: M Kappa Casein percentage, MPOT: Muscle Potassium Content, MSPD: Milking Speed, MTBR: Bone-Meat Ratio, MUGKCASP: Kappa Casein percentage in Non-Glucose Milk, MY: Milk Production, ML: Net Merit, NRR: Non return rate, PHOS: Muscle phosphorus content, PL: Productive lifespan, PP: Milk protein percentage, PTAT: Conformation final score, PUBAGE: Age at puberty, PY: Milk protein, RFI: Residual feed intake, RLEGS: Lateral view of posterior legs, RTPL: Posterior teats placement, RUMWD: Rump width, SB: Stillbirths, SCS: Somatic cell score, SF: Shear force, SOD: Muscle sodium content, STA: Height, STR: Strength, T3: Triiodothyronine level, TLGTH: Teat length, UA: Udder attachment, UC: Udder cleft, UDPTH: Udder depth, UHT: Hight of udder insertion, WWT: body weight.

Discussion

The results show that although the lengths of the ROH in this population (~4.8 Mb) suggest a crossbreeding of related animals of approximately ~16 generations (Ferenčaković *et al.*, 2013), the conserved regions could indicate that genetic selection in this population has been aimed at improving milk yield, its composition, fertility and health in this population. In some cases, associations with carcass traits were observed, findings that could suggest possible crosses with other breeds.

Studies in other specialized Holstein populations, have reported around 82.3±9.83 ROH per animal in USA Holstein cattle (Forutan *et al.*, 2018) and 81.7±9.7 runs per animal in Italian Holstein cattle (Marras *et al.*, 2014). These large differences may be due to the high degree of selection in the populations of specialized production systems in USA and Italy, as well as the availability of genetic material from highly selected stallions compared to small holder production system that was analyzed in where the selection objectives are not so focused.

Genomic regions (ROH) that have been maintained as a result of the selection process of the population of the family system were found. These conserved regions were found in positions within the genome where the annotations are numerous and include associations of SNP-like markers, QTL, genes, and signatures of selection; all of which are mostly related to traits of economic interest in the dairy industry, such as milk yield and composition, fertility parameters, susceptibility to diseases, body conformation, feed efficiency and some characteristics of the composition of the carcass.

These results indicate that the above-mentioned traits have been modified by the selection processes of the genetic material in the SHS population.

References

Chen N., Fu W., Zhao J., Shen J., Chen Q., Zheng Z., and Jiang Y. (2020). Genomics Proteomics & Bioinformatics *18*(2): 186-193. https://doi.org/10.1016/j.gpb.2019.03.007

Cortes-Hernández J., García-Ruiz A., Vásquez-Peláez C. G., and Ruiz-Lopez F. D. J. (2021). *Animals*, *11*(11), 3234. https://doi.org/10.3390/ani11113234

Ferenčaković M., Hamzić E., Gredler B., Solberg T. R., Klemetsdal G., *et al.* J. (2013). Journal of Animal Breeding and Genetics 130(4), 286-293. https://doi.org/10.1111/jbg.12012

Forutan M., Mahyari S. A., Baes C., Melzer N., Schenkel, F. S., *et al.* (2018). *BMC genomics*, *19*(1), 1-12. https://doi.org/10.1186/s12864-018-4453-z

Golden Helix, I., & Bozeman, M. (n.d.). SNP & Variation Suite ™ (Version 8) [Software]. Bozeman, MT: Golden Helix, Inc. Available from http://www.goldenhelix.com.

Marras G., Gaspa, G., Sorbolini S., Dimauro C., Ajmone-Marsan., *et al.* (2015). *Animal genetics*, *46*(2), 110-121. https://doi.org/10.1111/age.12259

Purfield D. C., Berry D. P., McParland S., and Bradley D. G. (2012). BMC genetics 13(1), 1-11. https://bmcgenomdata.biomedcentral.com/articles/10.1186/1471-2156-13-70

Sainz-Sánchez P. A., López-González F., Estrada-Flores J. G., Martínez-García C. G., and Arriaga-Jordán, C. M. (2016). Trop Anim Health Prod, 49(1), 179-186. https://doi.org/10.1007/s11250-016-1178-3.

668. How ssGBLUP became suitable for national dairy cattle evaluations

I. Misztal[1*] D. Lourenco[1], S. Tsuruta[1], I. Aguilar[2], Y. Masuda[3], M. Bermann[1], A. Cesarani[1] and A. Legarra[4]

[1]University of Georgia, Department of Animal and Dairy Science, Athens, GA 30602, USA; [2]Instituto Nacional de Investigación Agropecuaria, 11500, Montevideo, Uruguay; [3]Rakuno Gakuen University, 069-8501, Ebetsu, Hokkaido, Japan; [4]Institut national de recherche pour l'agriculture, l'alimentation et l'environnement, UMR1388 GenPhySE, 31326 Castanet Tolosan, France; ignacy@uga.edu

Abstract

National dairy evaluations use mostly multistep methods, where a nongenomic BLUP evaluation is followed by extraction of pseudo-phenotypes, genomic analyses, and merging nongenomic and genomic evaluations. With genomic preselection and many females genotyped, pseudo-phenotypes are biased, and it is hard to accommodate female genotypes fully. Such problems were successfully solved in other species by ssGBLUP, but large biases and expensive computing were experienced in dairy. This paper documents several improvements that made ssGBLUP feasible for national dairy evaluations. The pedigree relationships account for inbreeding. The formulas for unknown parent groups involve relationships for genotyped animals, and the number of groups is small. Pedigrees and phenotypes are truncated. Large data is accommodated by the APY algorithm. Multibreed evaluation is possible with judicious choice of APY core animals. In tests with the US national data and up to 3.5 M genotypes, analyses by ssGBLUP showed minimal bias and superior stability of GEBV.

Introduction

Dairy genomic evaluations at the national level use mostly multistep methods (VanRaden *et al.*, 2009), where a nongenomic BLUP evaluation is followed by extraction of pseudo-phenotypes for genotyped animals, genomic evaluation for genotyped animals, and an index to merge genomic and nongenomic values. Accurate creation of the index is difficult, and many countries use an approximate index, e.g. assuming 20% of polygenic and 80% of genomic component.

Over time as animals were genomically selected, BLUP became biased by pre-selection (Patry and Ducrocq, 2011; Masuda *et al.*, 2018). Subsequently, pseudo-phenotypes are biased and require ad-hoc corrections. Accommodating genotypes of females without double counting is difficult, potentially creating problems as now most genotyped animals are females. One alternative to the multistep method is ssGBLUP (Aguilar *et al.*, 2010; Misztal *et al.*, 2020) that is used in almost all species except dairy. It automatically accounts for genomic preselection, avoids calculating pseudo-phenotypes, and accommodates male and female genotypes without double counting. In dairy, ssGBLUP was applied as early as 2010 by Aguilar *et al.* (2010) and more recently by Masuda *et al.* (2018) and Bradford *et al.* (2019); however, the evaluations were biased in all the cases. Another problem was a relatively large computing time due to slow iteration. Recently, two studies looked at an evaluation using Holsteins only (Cesarani *et al.*, 2021) and multibreed using 5 breeds (Cesarani *et al.*, 2022), and evaluations were nearly unbiased in both cases. An additional study compared evaluations in the last study with the official CDCB evaluations (Mota *et al.*, 2022). That study found that the reliability of ssGBLUP was up to 0.07 higher, and stability was up by 0.12 higher. Subsequently, the current implementation of ssGBLUP seems beneficial for the routine evaluation in dairy cattle. The purpose of this paper is to describe steps that made ssGBLUP unbiased, accurate, and computationally efficient.

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_668, © I. Misztal *et al.* 2022

Materials & methods

Data. Recently, several ssGBLUP evaluations for milk, fat, and protein were run with the CDCB data, only for Holsteins (Cesarani *et al.*, 2021) or for five dairy breeds altogether that included Holstein, Jersey, Ayrshire, Brown Swiss, and Guernsey animals (Cesarani *et al.*, 2022). Genotyped animals used in the analyses ranged from 800k to 3.9 million. The developments and improvements required to make ssGBLUP a suitable method for dairy evaluations are described below.

Compatibility between pedigree and genomic relationships. To avoid biases, pedigree and genomic relationships need to be compatible. Standard scaling is by a linear transformation of the genomic relationship matrix (G) for equal mean with the pedigree relationship matrix (A) (Vitezica *et al.*, 2011). While A is affected by the completeness of pedigree, G is not. Therefore, an important part of scaling is accounting for inbreeding in the pedigree relationships. If pedigrees are missing, the inbreeding of animals with missing parents is calculated as 0, so accounting for nonzero inbreeding of missing parents (VanRaden, 1992) could provide additional compatibility.

Unknown parent groups (UPG). Under selection, the merit of missing parents over time needs to be accounted for. Early studies in ssGBLUP accounted for UPG only in A, resulting in slow convergence and biases. Later studies found that a reasonable option is also considering UPG for pedigree relationships for genotyped animals (A_{22}). Many BLUP evaluations use 100s of UPG. With UPG in A_{22}, there is not enough information to account for many groups, and considerably reducing the number of groups was a good option to reduce bias.

Interaction among pedigree, data truncation, and UPG. In general, the additive information decays fast with subsequent generations, and old data can be discarded without losses in accuracy for new animals but with gains in computing time. Including more than 2 generations of data with additional 2 generations of pedigree did not improve the accuracy for young animals and sometimes even lowered it (Lourenco *et al.*, 2014; Cesarani *et al.*, 2021). The broiler industry uses at most 3 generations of data despite many more available. Potential reasons for lower accuracy with older data are changing definitions of traits, model adjustments, and poor convergence for the solutions for young animals. Additionally, age adjustments vary over time as cows are indirectly selected for early calving. Truncation of old pedigrees eliminates the effect of missing pedigrees prior to truncation and reduces the number of needed UPG. If the number of generations is small, accounting for inbreeding is less or not critical.

Metafounders (MF) in ssGBLUP. MF are generalizations of UPG with scaling and inbreeding plus covariances accommodated (Legarra *et al.*, 2015). In regular scaling, G is scaled for compatibility with A_{22}. With MF, G is created with 0.5 gene frequencies, and pedigree relationships are adjusted for compatibility with G. Variance and covariances across MF potentially account for inbreeding of unknown parents and relationships among unknown parents. While MF did very well in simulations, problems with parameter estimation are still under investigation in real populations.

Large number of genotypes. The dairy populations accumulated a large number of genotypes, and many algorithms exist that can support that number in single-step methods (Misztal, 2018). All of them exploit the fact that the rank of genomic computations (based on SNP or G) is limited by the number of SNP (usually 60k) and by the small dimensionality of G in populations with small effective population size. This dimensionality is about 15k in dairy cattle (i.e. 15k core animals). With the APY algorithm (Misztal, 2016), which relies on the limited dimensionality of G, calculations have linear computing and storage costs past the number of core animals. An advantage of APY is the low cost and easy application of UPG. A disadvantage is some fluctuations of GEBV with the choice of core animals, which can be handled by

keeping the core constant. An important part of the evaluation is pruning genotypes. While the number of genotyped animals may be very high, those without progeny or phenotypes can be eliminated from ssGBLUP evaluations, reducing computing costs. Predictions for omitted animals by indirect predictions can be as accurate with large data sets as when they are included in the model (Tsuruta *et al.*, 2021). Another factor influencing computations is the blending of G, used to avoid the singularity of G. Blending with A_{22} was costly; however, computing coefficients separately for core and noncore animals is highly efficient. Blending with 0.01I is also efficient without affecting accuracy.

Other issues. Many other issues are remaining in the dairy ssGBLUP evaluations that are not discussed here but have been addressed elsewhere. They include using external information from Interbull (Guarini *et al.*, 2019), exploiting putative QTN (Misztal *et al.*, 2020), using purebred and crossbred information combined (Misztal *et al.*, 2022), and approximating accuracies of GEBV (Bermann *et al.*, 2022).

Results

Holsteins (Cesarani et al., 2021). Phenotypes were truncated for cows recorded before 1980, 1990, or 2000, and pedigrees were truncated to 2 or 3 generations beyond phenotyped cows. Two UPG formulations were used, the QP transformation for A (UPG1) or A and A_{22} (UPG2). With forward prediction, reliabilities were calculated for bulls and predictivity for cows. Only genotypes for contributing animals were used (about 800k). With UPG2, reliabilities and accuracies were the same with all truncations. Inflation was minimal as the b_1 parameter was close to 1.0. With simple UPG, reliabilities were smaller and varied with truncation.

Five breed analyses (Cesarani et al., 2022). Holstein, Jersey, Ayrshire, Brown Swiss, and Guernsey data were analyzed separately and altogether in a model with UPG2, truncation of phenotypes recorded before 2000, and 3 generations beyond phenotypes. The total number of genotyped animals was 3.9 million. When the multibreed evaluation used 15k randomly selected core animals in the APY algorithm, the reliabilities were lower for the smaller breeds than in the single-breed analyses. When core animals were selected to include 5k animals of each smaller breed and 15k of each Holstein and Jersey, the reliabilities were close to those from single breeds, and they were higher for Jerseys. One explanation is that the Jersey population included Holsteins in the past. In all cases except for the smallest breeds, the inflation was low (i.e. b_1 close to 1.0).

Official comparisons. Evaluations for Holstein bulls from the previous analyses were compared with official multistep evaluations (MS) by CDCB (Mota *et al.*, 2022); the latter were computed using different weights for SNP in a nonlinear approach, whereas ssGBLUP assumed equal weights for SNP. For milk, R^2 based on DYD for raw MS was 0.07 lower than for ssGBLUP. After age adjustments, this difference reduced to 0.01. Stability of evaluations was calculated by correlations of GEBV derived with 2017 and 2021 data. The stability for MS was 0.86 (0.89 with correction for age). The stability for ssGBLUP was 0.90 (0.92 after age correction). Regression coefficients were close to 1.0 for both ssGBLUP and MS predictions.

Discussion

The results for reliability and stability indicate a dilemma on how to judge an evaluation system. If DYD is derived from BLUP, it is biased even for bulls with many daughters. Also, GEBV by unadjusted MS are biased. Subsequently, it is not clear whether adjustments for age compensate GEBV, DYD, or both. In terms of stability, ssGBLUP was ahead of MS. After many developments as described, ssGBLUP is finally ready for implementation at the national level. This method provides evaluations that are possibly more accurate than the current one, does not require adjustments, and is simpler to run. Further improvements of accuracy can be expected with the inclusion of external Interbull data and possibly weighing SNP differently.

References

Aguilar I., I. Misztal, D. L. Johnson, A. Legarra, S. Tsuruta, *et al.* (2010) J. Dairy Sci. 93:743-752. https://doi.org/10.3168/jds.2009-2730

Bermann M., D. Lourenco, and I. Misztal. (2022) J. Anim. Sci. skab353. https://doi.org/10.1093/jas/skab353

Bradford H. L., Y. Masuda, J. B. Cole *et al.* (2019) J. Dairy Sci. 102:2308-2318. https://doi.org/10.3168/jds.2018-15419

Cesarani A., Y. Masuda, S. Tsuruta, *et al.* (2021) J. Dairy Sci. 104:5843-5853. https://doi.org/10.3168/jds.2020-19789

Cesarani A., Y. Masuda, S. Tsuruta, *et al.* (2022) J. Dairy Sci. (Under review)

Guarini A.R., D.A.L. Lourenco, L.F. Brito, *et al.* (2019) J. Dairy Sci. 102:8175-8183. https://doi.org/10.3168/jds.2018-15819

Lourenco D., I. Misztal, S. Tsuruta, *et al.* (2014) J. Dairy Sci. 97:3930-3942. https://doi.org/10.3168/jds.2013-7769

Legarra A., O. F. Christensen, Z. G. Vitezica, *et al.* (2015) Genetics 200:455-468. https://doi.org/10.1534/genetics.115.177014

Mota R.R, A. Cesarani, and P.M. VanRaden. Proc. of the 12th WCGALP, Rotterdam, the Netherlands.

Masuda,Y., P. M. VanRaden, I. Misztal, *et al.* (2018). J. Dairy Sci. https://doi.org/10.3168/jds.2017-13310

Misztal I. (2018) 11th WCGALP, Auckland, New Zealand.

Misztal I., D. Lourenco, and A. Legarra (2020) J. Animal Sci. 98:1-14. https://doi.org/10.1093/jas/skaa101

Misztal I. (2016) Genetics 202:401–409. https://doi.org/10.1534/genetics.115.182089

Misztal I., Y. Stein, and D.A.L. Lourenco (2022) J. Dairy Sci. Comm. (Accepted)

Patry C., and V. Ducrocq. (2011) Genet. Sel. Evol. 43:30. https://doi.org/10.1186/1297-9686-43-30

Tsuruta S., D. Lourenco, Y. Masuda, *et al.* (2021) J. Dairy Sci. Commun. 2:356-360. https://doi.org/10.3168/jdsc.2021-0097

VanRaden P. M. (1992) J. Dairy Sci.75:3136–3144. https://doi.org/10.3168/jds.S0022-0302(92)78077-1

VanRaden, P. M., C. P. Van Tassell, G. R. Wiggans, *et al.* (2009) J. Dairy Sci. 92:16-24. https://doi.org/10.3168/jds.2008-1514

Vitezica Z., I. Aguilar, I. Misztal, and A. Legarra (2011) Genet. Res. Camb. 93:357-366. https://doi.org/10.1017/S001667231100022X

669. Options for evaluating multiple breeds in a single-step GBLUP for US dairy population

A. Cesarani[1*], D. Lourenco[1], P.M. VanRaden[2], E.L. Nicolazzi[3], A. Legarra[4], S. Tsuruta[1] and I. Misztal[1]

[1]Department of Animal and Dairy Science, University of Georgia, Athens, GA 30602, USA; [2]Animal Genomics and Improvement Laboratory, Agricultural Research Service, USDA, Beltsville, MD, USA; [3]Council on Dairy Cattle Breeding, Bowie, MD 20716, USA; [4]INRA, UMR1388 GenPhySE, Castanet-Tolosan, France; acesarani@uniss.it

Abstract

This work aimed to develop single-step multibreed genomic predictions for US dairy population. About 45M yield phenotypes recorded between 2000 and 2020 in purebred Ayrshire, Brown Swiss, Guernsey, Holstein, and Jersey animals were analyzed. Genotypes at 79,294 SNPs were available for 3.9M animals. A 3-traits model with milk, fat, and protein yields was applied to four different scenarios: each breed separately; all five breeds together; Aushire, Brown Swiss, and Guernsey together; Holstein and Jersey together. For each scenario, BLUP and single-step genomic BLUP were carried out and were validated using the correlation between adjusted phenotypes (for cows) or daughter yield deviations (for bulls) and (genomic) breeding values. As expected, all ssGBLUP scenarios led to higher values compared to BLUP. The ssGBLUP multibreed evaluation for US dairy population is feasible, but model fine-tuning is needed to achieve good validation results.

Introduction

In 2007, a multibreed genetic evaluation was implemented for US dairy cattle, and it was extended to a multistep genomic evaluation in 2009 (VanRaden *et al.*, 2007; 2009). Currently, US dairy cattle are evaluated with a two-step method: a multibreed BLUP model followed by SNP effects estimation and direct genomic values computation within each breed. Single-step genomic BLUP (ssGBLUP) has been adopted in lieu of the multistep method for genetic evaluations in several farm animal species over the years. ssGBLUP has been applied with satisfactory results in almost all livestock species (e.g. Macedo *et al.*, 2020; Pimentel *et al.*, 2021). Genomic models have been tested also in a multibreed context (e.g. Winkelman *et al.*, 2015; Khansefid *et al.*, 2020). The aim of this work was to develop ssGBLUP multibreed genomic predictions for US dairy cattle. In particular, we used the algorithm for proven and young (APY) to compute the inverse of the genomic relationship matrix.

Materials & methods

Data. Official data used in the multibreed genomic evaluations for US dairy cattle breeds were provided by the Council on Dairy Cattle Breeding (CDCB, Bowie, MD). Only purebred Ayrshire (AY), Brown Swiss (BS), Guernsey (GU), Holstein (HO), and Jersey (JE) animals were considered. Phenotypes for milk (MY), fat (FY), and protein (PY) yields recorded from 2000 to 2020 were considered. The 305-d yields (about 45M recorded on 19.4M cows) for the first 5 lactations included projected records for the final lactation and for lactations not yet completed by June 2020. About 3.9M animals were genotyped at various densities and imputed to 79,294 selected SNPs within each breed. Only purebred animals were included in the evaluation.

Analysis. Two evaluation methods were considered: (1) traditional BLUP, with unknown parent groups (UPG); (2) ssGBLUP, with QP transformation of unknown parent groups applied to A and A_{22}, implemented as in Tsuruta *et al.* (2019) and Cesarani *et al.* (2021). UPG were defined based on breed, sex, and year of birth for a total of 8 UPG for each breed. The two methods were carried out for: (1) each breed

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_669, © A. Cesarani *et al.* 2022

separately (SINGLE); (2) the 5 breeds together (5_BREEDS); (3) AY, BS, and GU (AY_BS_GU); (4) HO and JE together (HO_JE). The APY algorithm was applied with 15k core animals in SINGLE (for HO and JE), 30k for AY_BS_GU and HO_JE, and 45k core animals for 5_BREEDS. A 3-trait repeatability animal model was applied with herd-management, age-parity, inbreeding coefficient and heterosis covariates, and UPG as fixed effects and herd-sire combination, animal, and permanent environment as random effects. For the multibreed models, all effects were made breed-specific by adding the breed code to each level of the effects.

Validation. Each scenario was run two times: FULL, with all phenotypes; REDUCED, removing the phenotypes of cows born in the last 4 years (2015-2018). Validation of each model was assessed by the correlation between adjusted phenotypes in FULL and (genomic) estimated breeding values ((G)EBV) in REDUCED for cows (predictive ability), and by the correlation of daughter yield deviation (DYD) in FULL on the (G)EBV in REDUCED for bulls. Validation cows were genotyped females born in 2015-2018 with no records in REDUCED, whereas validation bulls had at least 10 (AY, BS, GU), or 50 (HO, JE) daughters in FULL and no daughters in REDUCED.

Results

Correlations between raw EBV from BLUP estimated in all the different models were 1.00 for all breeds and traits, whereas correlations between raw GEBV estimated in ssGBLUP SINGLE and ssGBLUP 5_BREEDS ranged from 0.95 (MY for AY and BS) to 1.00 (HO and JE). Moreover, we also computed the slope of the regression of (G)EBV in 5-BREEDS on (G)EBV in SINGLE for each breed: slopes for BLUP were always 1, confirming that in practice each breed was independently evaluated, as breeds were not connected by the pedigree. Acceptable slope values were also found for ssGBLUP as values ranged from 0.93 (MY and PY for BS) to 1.02 (MY for AY). Table 1 shows the validation results for both cows and bulls. As expected, ssGBLUP SINGLE gave better results compared to BLUP for all considered breeds and traits. Within ssGBLUP, higher values were observed for 5_BREEDS compared to SINGLE for all breeds, expect for BS bulls, for which a slight decrease was observed in the multibreed scenario. Results in AY_BS_GU and HO_JE were similar to SINGLE. However, results for AY, and GU are strongly affected by the very low number of bulls.

Table 2 reports the slope of DYD or adjusted phenotypes on (G)EBV for bulls and cows, respectively. Results were similar to those in Table 1, where ssGBLUP SINGLE led to better results, on average (i.e. values closer to 1) compared to BLUP. All regression coefficients estimated for HO and JE were within 15% of the unit, considered acceptable values (Tsuruta *et al.*, 2011). In this case, results for AY and GU were also affected by the limited number of validation animals.

Regarding computing time, about 7 hours were needed to solve the MME with ssGBLUP in HO SINGLE (3.4M genotypes); this time increased by a factor of 9 for the 5_BREEDS scenario.

Discussion

The use of genomic information is expected to increase the reliability of breeding value prediction in livestock. ssGBLUP gives higher validation accuracies in several livestock, but applications of this method in large-scale dairy evaluations have been limited. Recently, Cesarani *et al.* (2021) showed the superior performance of ssGBLUP compared to BLUP in US Holsteins. The present study gave insights into using this genomic model in a multibreed context. As expected, using genomic information within each breed improves validation statistics for both bulls and cows. Results for Holsteins were similar in all scenarios involving this breed because of the greatest number of genotypes and records. This study demonstrates that large-scale multibreed genomic evaluations are feasible, as we obtained results in the multibreed scenarios

Table 1. Correlations between adjusted phenotypes for cows or daughter yield deviations for bulls and EBV from BLUP or GEBV from ssGBLUP in the five considered breeds.

Breed	Trait	N animals				ssGBLUP							
				BLUP		SINGLE		5_BREEEDS		AY_BS_GU		HO_JE	
		Cows	Bulls	Cows	Bulls	Cows	Bulls	Cows	Bulls	Cows	Bulls	Cows	Bulls
	milk			0.45	0.45	0.47	0.52	0.55	0.85	0.47	0.58	–	–
AY	fat	181	17	0.51	0.45	0.54	0.56	0.52	0.86	0.53	0.58	–	–
	prot			0.52	0.55	0.54	0.66	0.55	0.93	0.54	0.69	–	–
	milk			0.23	0.39	0.39	0.63	0.40	0.61	0.39	0.63	–	–
BS	fat	2,423	107	0.27	0.37	0.42	0.55	0.41	0.55	0.42	0.58	–	–
	prot			0.27	0.46	0.44	0.67	0.43	0.65	0.44	0.67	–	–
	milk			0.24	0.42	0.35	0.54	0.44	0.64	0.35	0.51	–	–
GU	fat	750	28	0.27	0.45	0.33	0.56	0.39	0.63	0.34	0.53	–	–
	prot			0.24	0.42	0.34	0.50	0.39	0.53	0.34	0.48	–	–
	milk			0.31	0.53	0.55	0.88	0.55	0.88	–	–	0.55	0.88
HO	fat	577,340	3,278	0.34	0.56	0.55	0.88	0.55	0.87	–	–	0.55	0.88
	prot			0.33	0.55	0.52	0.86	0.52	0.86	–	–	0.52	0.86
	milk			0.32	0.66	0.5	0.84	0.51	0.87	–	–	0.52	0.87
JE	fat	90,666	471	0.29	0.60	0.46	0.80	0.47	0.82	–	–	0.47	0.82
	prot			0.35	0.67	0.51	0.83	0.52	0.86	–	–	0.52	0.86
Mean				0.33	0.50	0.46	0.69	0.48	0.76	0.42	0.58	0.52	0.86
SD				0.09	0.09	0.08	0.15	0.06	0.14	0.08	0.07	0.03	0.02

Table 2. Regression coefficients (b1) of bull daughter yield deviations (DYD) on EBV from BLUP and on GEBV from ssGBLUP for validation bulls and regression coefficients of cow adjusted phenotypes on EBV from BLUP and on GEBV from ssGBLUP for validation cows.

Breed	Trait	N animals				ssGBLUP							
				BLUP		SINGLE		5_BREEEDS		AY_BS_GU		HO_JE	
		Cows	Bulls	Cows	Bulls	Cows	Bulls	Cows	Bulls	Cows	Bulls	Cows	Bulls
	milk			1.03	0.94	0.96	0.88	0.88	1.37	0.98	0.96	–	–
AY	fat	181	17	0.95	0.79	0.89	0.86	0.76	1.25	0.90	0.92	–	–
	prot			0.98	1.13	0.92	1.04	0.80	1.34	0.93	1.06	–	–
	milk			0.72	0.64	0.96	0.94	0.90	0.81	0.97	0.93	–	–
BS	fat	2,423	107	0.72	0.55	0.91	0.80	0.84	0.72	0.92	0.80	–	–
	prot			0.73	0.70	0.96	0.93	0.89	0.82	0.96	0.92	–	–
	milk			0.75	0.75	0.91	1.04	1.07	1.14	0.91	1.03	–	–
GU	fat	750	28	0.71	0.67	0.74	0.84	0.93	0.89	0.77	0.85	–	–
	prot			0.71	0.71	0.84	0.9	1.05	1.02	0.85	0.92	–	–
	milk			1.01	0.86	1.10	1.03	1.11	1.03	–	–	1.10	1.04
HO	fat	577,340	3,278	1.03	0.91	1.03	0.99	1.03	0.97	–	–	1.02	0.99
	prot			1.02	0.87	1.08	1.03	1.08	1.03	–	–	1.07	1.02
	milk			1.02	1.06	1.03	1.01	1.07	1.05	–	–	1.09	1.08
JE	fat	90,666	471	0.89	0.89	0.88	0.86	0.93	0.92	–	–	0.97	0.97
	prot			1.01	1.06	0.99	0.99	1.05	1.06	–	–	1.07	1.08
Mean				0.89	0.84	0.95	0.94	0.96	1.03	0.91	0.93	1.05	1.03
SD				0.14	0.17	0.09	0.08	0.11	0.19	0.07	0.08	0.05	0.05

that were at least similar to the ones in single-breed evaluations. Fine-tuning of models and a better representation of each considered breed in the core animals are needed to avoid reducing the reliability of smaller breeds when evaluated together with numerically dominant breeds. This is the first step towards the multi breed evaluation, in which crossbreed and external information will be considered.

References

Cesarani A., Masuda Y., Tsuruta S., Nicolazzi E.L., VanRaden P.M., *et al.* (2021) J. Dairy Sci. 104:5843-5853. https://doi.org/10.3168/jds.2020-19789

Khansefid M., Goddard M.E., Haile-Mariam M., Konstantinov K.V., Schrooten C., *et al.* (2020) Front. Genet. 11:598580. https://doi.org/10.3389/fgene.2020.598580.

Macedo F.L., Christensen O.F., Astruc J.M., Aguilar I., Masuda Y., *et al.* (2020) Genet. Sel. Evol. 52:1-10. https://doi.org/10.1186/s12711-020-00567-1

Pimentel E.D.C.G., Edel C., Krogmeier D., Emmerling R., and Götz K.U. (2021) Interbull Bulletin, (56), 121-124.

Tsuruta S., Lourenco D.A.L., Masuda Y., Misztal I., and Lawlor T.J. (2019) J. Dairy Sci. 102:9956-9970. https://doi.org/10.3168/jds.2019-16789

Tsuruta S., Misztal I., Aguilar I., and Lawlor T.J. (2011) J. Dairy Sci. 94:4198–4204. https://doi.org/10.3168/jds.2011-4256.

VanRaden P.M., Tooker M.E., Cole J.B., Wiggans G.R., and Megonigal Jr J.H. (2007) J. Dairy Sci. 90:2434-2441. https://doi.org/10.3168/jds.2006-704.

VanRaden P.M., Van Tassell C.P., Wiggans G.R., Sonstegard T.S., Schnabel R.D., *et al.* (2009) J. Dairy Sci. 92:16–24. https://doi.org/10.3168/jds.2008-1514

Winkelman A.M., Johnson D.L., and Harris B.L. (2015) J. Dairy Sci. 98:659-675. https://doi.org/10.3168/jds.2014-8560.

670. Integrating MACE breeding values into multiple-trait single step random regression test day evaluations of Aussie Reds

V. Boerner[1,2*], T. Nguyen[3] and G. Nieuwhof[3]

[1]GHPC Consulting and Services PTY. LTD., Armidale 2351, Australia; [2]Centre for Quantitative Genetics and Genomics, Aarhus University, Denmark; [3]DataGene, AgriBio, 5 Ring Road, Bundoora VIC 3083, Australia; vinzent.boerner@gmx.de

Abstract

Interbull's multiple across country evaluation provides national breeding organisations with breeding values for internationally used bulls which must be integrated into the domestic evaluation. For that purpose several methods have been proposed which either model Interbull breeding values as prior information in a Bayesian approach, or as additional pseudo data points. Single step random regression models add complexity to the integration because of the dimensionality discrepancy between the number of Interbull breeding values and the number of modelled genetic effects. This paper presents the results from integrating 16,063 Interbull breeding values into the domestic single step random regression test day model for milk, fat and protein yield for Australian red dairy cattle breeds. Results suggest that the integration was successful with regard to alignment of Interbull breeding values with their domestic equivalent as well as with regard to the individual and population-wide increase in reliabilities.

Introduction

National breeding programs can only benefit from Interbull's multiple across country evaluation if the breeding values reported by Interbull are integrated into the national evaluation system. This is even more relevant for single step genetic evaluation as the number of breeding values affected by the integration is expected to increase. External breeding values can be integrated as prior information in a Bayesian approach (Vandenplas *et al.* 2014) which requires manipulation of the mixed model equation system's right-hand side and a special co-variance matrix between animals, or as additional pseudo data (Pitkänen *et al.* 2020) which requires derivation of one or several additional data points and manipulation of the residual co-variances matrix. The integration for production traits like milk, fat and protein yield is complicated due dimensionality discrepancies between the Interbull model and the domestic multiple-trait random regression test day models. In this paper we present results from the integration of Interbull breeding values for milk fat and protein yield into the domestic single step random regression test day model genetic evaluation for Australian Red Dairy breeds.

Materials & methods

The data set for genetic evaluation of production traits of Australian red dairy breeds consisted of test day observation on litre milk, kilogram fat and kilogram protein in the first three lactations with a total of 10,000,614 observations across all nine traits. The pedigree consisted of 73 phantom parents and 975,532 individuals of which 8,191 were genotyped on various platforms. Genotypes underwent internal quality control and imputation procedures and after being imputed to a common set of about 73,000 markers were used to construct a genomic relationship matrix. The total number of bulls to be integrated was 16,063 where 466 bulls had Australian information provided to Interbull (*A*bulls), and 15,597 bulls may have Australian information but it was not provided to Interbull or did not have Australian information at all (*B*bulls). Breeding values of *A*bulls for integration were identified by a positive difference between the Interbull reliability and the local reliability of at least 0.01. If the difference was only achieved for a subset of an *A*bulls's breeding values, e.g. for milk yield, but not for fat and protein yield, only the subset was integrated. All breeding values of *B*bulls were integrated regardless of the Interbull reliabilities. The genetic evaluation

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_670, © V. Boerner *et al.* 2022

model into which the Interbull breeding values were integrated modelled phenotypic observations as linear functions of fixed herd-year, season, age and lactation curve effects, where the latter used Legendre polynomials of order 1 to 4, of correlated direct random genetic effects using Legendre polynomials of order 0 to 2, of of correlated direct random permanent environment effects of the animal using Legendre polynomials of order 0 to 2, and correlated random residuals. Vectors of direct genetic effects ~$N(0,\Gamma\otimes\Sigma_a)$ and vectors of permanent environmental effets ~$N(0,I\otimes\Sigma_q)$, where Σ_q is co-variance matrix of the permanent environmental effects of dimension 27×27 and Σ_a is the genetic co-variance matrix of dimension 27×27, and $vec(\Gamma) = [\Theta, Q\Theta, \Theta Q', QQ\Theta' + H]$ where Θ is a diagonal matrix with elements equal to 0.62, Q is the pedigree derived genetic group regression matrix, and H is the single step relationship matrix (Christensen and Lund 2010) constructed using a polygenic weight of 0.2. Interbull breeding values were modeled as additional data points with the data point specific residual variance as the single tuning parameter. For aligning within animal dimensions Interbull breeding values u_{MACE} and reliabilities r_{MACE}, and sent breeding values u_{SENT} and reliabilities r_{SENT} were mapped from $\Re_3$ to $\Re_9$ using a block-diagonal matrix K of dimension 9×3 containing 3×1 column blocks of ones. Further Σ_a was mapped from $\Re_{27}$ to $\Re_9$ using a 9×27 block diagonal matrix $L = t\otimes I_9$ with $t = [1'f_0(d), 1'f_1(d), 1'f_2(d))$ where I_9 is an identity matrix of dimensions 9×9, t is a vector of length 3, d is a vector containing a consecutive sequence from 6 to 305, f_i is a vector-valued function applying the Legendre polynomial order i to the argument, and 1 is a vector of ones. Therefore $u_{:,m} = Ku_{:}$ and $r_{:,m} = Kr_{:}$, where ':' was either $_{MACE}$ or $_{SENT}$ and $\Sigma_{a,m} = LS_aL'$. In case only a subset of breeding values within animal met the integration requirement the above dimension were adjusted accordingly. Within animal a vector of prediction error variances was calculated by $p_{:} = diag(\Sigma_{a,m})(1 - r_{:,m})$, where $diag(\ )$ denotes a vector of diagonal elements of a squared matrix. The system for deriving a diagonal matrix $D_{:}$ of inverse residual variances across traits within animal was $diag\left(\left(D_{:} + (\Sigma_{a,m})^{-1}\right)^{-1}\right) \equiv diag(p_{:})$ where $D_{:}$ was be obtained directly only if the system's dimension is 1. Otherwise $D_{:}$ was established iteratively using a procedure similar to (Vandenplas 2012). Within animal vectors of pseudo-phenotypes $y^x_{:}$ were calculated for *A*bulls as $D_r = D_{MACE} - D_{SENT}$ and $y^x = D_r^{-1}\left(\left(D_{MACE} + (\Sigma_{a,m})^{-1}\right)u_{MACE,m} - \left(D_{SENT} + (\Sigma_{a,m})^{-1}\right)u_{SENT,m}\right)$ and for *B*bulls as $y^x = D_{MACE}^{-1}\left(D_{MACE} + (\Sigma_{a,m})^{-1}\right)u_{MACE,m}$.

A maximum of nine pseudo-phenotypes y^x of animal i were added to the evaluation.

The genetic evaluation model was solved and reliabilities were deterministically calculated using MiX99 and Apax, respectively(Lidauer *et al.* 2017).

Results

Generally, results for all three traits were very similar. Therefore, only results for milk yield are summarised in Figure 1. Correlations between Interbull breeding values and domestic evaluation breeding values increased from about 0.05, 0.05 and 0.04 pre-integration to 0.99, 0.98 and 0.99 post-integration for milk, fat and protein yield respectively. The regression intercept changed from 392, 12.4 and 11.3 to 261, 14.3 and 7.11, and the regression slope from 0.05, 0.06 and 0.06 to 1.01, 1.01 and 1.01 for milk fat and protein yield, respectively. Results were similar for the reliabilities, but correlations and regression parameters are weaker and, due to the nature of reliabilities, do not provide the same goodness-of-fit information as for the breeding values. However, post-integration reliabilities of all bulls were at least as high as the Interbull reliabilities. For *A*bulls, the post-integration reliabilities aligned very well with the Interbull reliabilities. Further, for *B*bulls breeding values with a pre-integration reliability of zero the procedure returned a post-integration reliability similar to the Interbull reliability. For the entire evaluation population, but excluding the integrated bulls, about 40% of the reliabilities of all three traits were not affected by the integration, and only 5% of the reliabilities increased by more than 0.1. Contrarily, every single breeding value changed as a result of the integration with 27, 29, and 29% of the breeding values for milk, fat and protein yield, respectively, having changed by more than 10%.

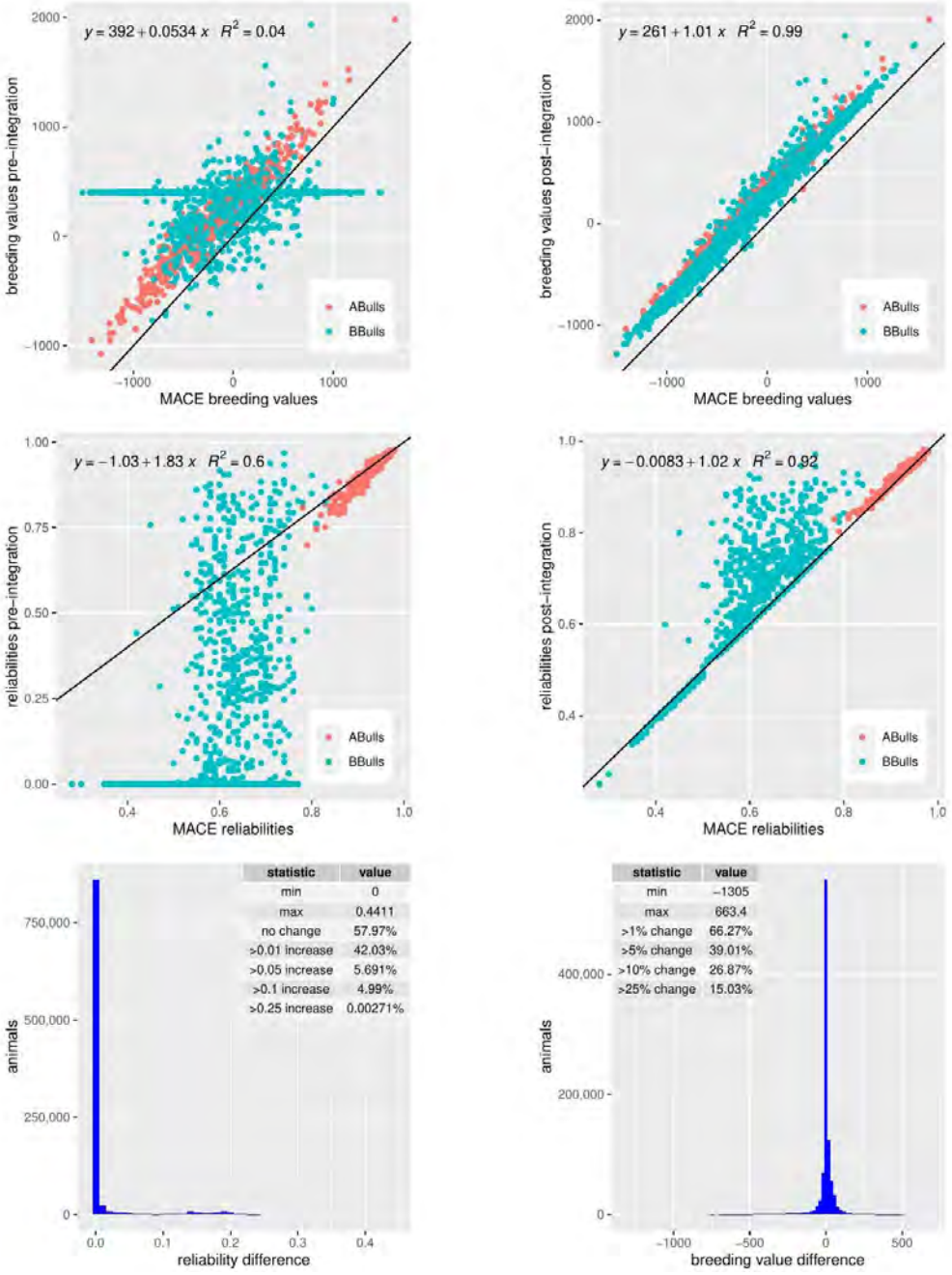


Figure 1. Single Step H-BLUP breeding values (first row) and reliabilities (second row) for litremilk per 305 day lactation pre-integration (left) and post-integration (right) for bulls with (Abulls) and without (Bbulls) Australian information in the Interbull evaluation, respectively. Histograms of the integration impact on breeding values and reliabilities of the entire evaluation population (third row).

Discussion

Results shown here align very well with the expectations and therefore the integration of Interbull breeding values into the domestic single step random regression test day evaluation is regarded as successful. The current integration approach ignores the relationships between integrated bulls when deriving individual specific residual variances. Empirically, ignoring the relationships between integration animals does not pose a problem if the reliabilites are high and integration candidates are not closely related. Accounting for relationships poses a computational problem (Vandenplas *et al.*, 2014). For the approach presented here the procedure for deriving D_i required up to tens of thousand of inversions. Extending the approach to system of $\Sigma_a \otimes \Gamma_{:,:}^{-1}$ will certainly affect inversion time and most likely also the required number of iterations, and will also require block extractions from a complex Γ matrix. An intermediate approach is to approximate some matrix $K = C_{a_i,a_i} - C_{a_i,-a_i}\left(C_{-a_i,-a_i}\right)^{-1} C_{-a_i,a_i}$, where C is the mixed model coefficient matrix and a_i is a vector indexing the block in C related to animal i. Subsequently K is used instead of Σ_a in $diag((D_{:} + K)^{-1}) \equiv diag(p_{:})$ (Pitkänen, personal communication). However, it might be computationally prohibitive to apply such procedure to a single step system.

References

Christensen, Ole F., and Mogens S. Lund. 2010. Genet. Sel. Evol. 42: 2.

Lidauer M., Matilainen K., Mäntysaari E., Timo Pitkänen T., Taskinen M. et. al. 2017.

Pitkänen T. J., Koivula M., Strandén I., Aamand G.P., and Mäntysaari E.A. 2020 In Proc. Of the 71th Annual Meeting of the European Association for Animal Production, Porto, Portugal, 582.

Vandenplas J., Colinet F.G., and Gengler N.. 2014. Genet. Sel. Evol. 46.

Vandenplas, J., and Gengler N. 2012. J. Dairy Sci. 95: 1513–26.

671. Optimum contribution selection in a dairy cattle population with different relationship matrices

E. Gautason[1*], G. Sahana[1], B. Guldbrandtsen[2] and P. Berg[1,3]

[1]Aarhus University, Center for Quantitative Genetics and Genomics, Blichers Allé 20, 8830 Tjele, Denmark; [2]University of Copenhagen, Dept. of Veterinary and Animal Sciences, Grønnegårdsvej 3, 1870 Frederiksberg C, Denmark; [3]Norwegian University of Life Sciences, Department of Animal and Aquacultural Sciences, Oluf Thesens vei 6, 1433 Ås, Norway; egill@au.dk

Abstract

We compared the effects of using three genomic relationship matrices and a pedigree based relationship matrix for optimum contribution selection (OCS) with a constraint on number of sires selected, on genetic gain, kinship and additive genetic variance. This was done by simulating a genomically selected small dairy cattle population. The cattle population underwent genomic selection for 11 generations. In total, 6,000 females and 2,500 males were genotyped per generation. We compared a pedigree-based matrix as well as VanRaden's method 1 to construct genomic relationship matrices using either allele frequencies of the base population, of the current generation, or allele frequencies among all genotyped animals. The pedigree-based OCS resulted in higher kinship and more loss of genetic variance than using genomic OCS based on base population frequencies. Thus, genomic OCS with genomic relationship matrices using base population frequencies are preferable compared with pedigree-based OCS.

Introduction

Genomic selection has increased genetic gain in dairy cattle in recent years, but in some cases at the cost of higher inbreeding rates (Makanjuola *et al.* 2020). Small local breeds that are not genomically selected have become increasingly less competitive to the transboundary breeds. However, simulations have shown that small dairy cattle populations can also benefit from genomic selection (Thomasen *et al.* 2014; Obšteter *et al.* 2019), but research is lacking on ways to balance genetic gain and inbreeding in small populations. To achieve genetic gains through selective breeding in future generations, genetic diversity must be preserved. Inbreeding has to be managed to avoid inbreeding depression. This is not least the case in small populations. Optimum contribution selection (OCS; Meuwissen 1997) maximizes genetic gain while restraining inbreeding. OCS has been implemented using pedigree relationships (POCS), but genomic OCS (GOCS) can be an alternative by using a genomic relationship matrix (GRM). A common way to form GRMs is VanRaden's method one (VanRaden 2008). This method scales the cross-product of centered genotype scores by $\Sigma 2p(1\text{-}p)$, where p is the marker allele frequency. The choice of reference allele frequencies to use is an important parameter to consider when constructing GRMs. VanRaden (2008) recommended the use of base population frequencies. Berg *et al.* (subm.) found important differences in results depending on the type of reference used for GOCS. The aim of this study was to compare four scenarios, three different allele frequencies to construct GRMs and the pedigree based relationship matrix, on genetic gain, kinship and additive genetic variance in a small dairy cattle population undergoing genomic selection.

Materials & methods

The simulations were modelled to mimic the breeding program of the Icelandic Cattle population (Sigurdsson and Jonmundsson 2011). QMSim (Sargolzaei and Schenkel 2009) was used for simulating a base population. The R package Modular Breeding Program Simulator (MoBPS; Pook *et al.* 2020) was used for breeding program simulations. We used GMATRIX (Su and Madsen 2012) to construct GRMs. EVA (Berg *et al.* 2006) was used to optimize genetic contributions of selection candidates. DMU (Madsen and Jensen 2013) was used to estimate breeding values.

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_671, © E. Gautason *et al.* 2022

Historical population. We simulated a genome of 29 chromosomes each with 2,000 evenly spaced loci. The mutation rate was 2.5×10^{-5} per locus per generation. For each replicate of scenarios, a new historical population was simulated. The population size was 1000 individuals for 2,000 generations and increased over 200 generations to 12,000. Genotype data were converted into Plink *ped* format (Purcell *et al.* 2007) and read into MoBPS. We randomly sampled 2,000 segregating loci to be quantitative trait loci (QTL), which were not used for construction of GRM. We assigned effects to the QTL by drawing from a gamma distribution with shape 0.4 and scale parameter 1.66. We discarded very low frequency marker loci so that the genotyping markers reflected the allele frequency distribution of commercial SNP chips. The number of polymorphic genotyping marker loci were around 39,000 at the start of the genomic selection, with some variation across replicates. We simulated a sex-based trait with a constant heritability of 0.4. We used three replicates for each scenario.

Breeding program structure and breeding value prediction. We simulated four discrete generations of pedigree-based best linear unbiased prediction (PBLUP) selection before genomic selection started. We then simulated three generations of genomic prediction using single-step genomic BLUP (ssGBLUP), and subsequently eight generations of GBLUP. Each generation contained 6,000 males and 6,000 females. Each cow had exactly two calves. In each generation of genomic selection, 2,500 male and 6,000 female selection candidates were selected for genotyping based on their parent average and their genomic breeding values (GEBVs) were estimated. Selection was according to optimum contribution selection, only on the sire side, with additional constraints described below. Dams of selection candidates had phenotypes at the stage of selection. The model for genetic evaluation included the additive genetic animal effect, an intercept and a residual: y = Xb + Za + e, where y was a vector of animal phenotypes, b was a vector of fixed effects, a was a vector of (genomic) estimated breeding values where a followed $N(0, A\sigma_g^2)$, $N(0, H\sigma_g^2)$ and $N(0, G\sigma_g^2)$ for PBLUP, ssGBLUP and GBLUP, where A was the numerator relationship matrix, H was a combined pedigree and genomic relationship matrix (Christensen and Lund 2010) and G was a GRM computed using VanRaden's method one. X and Z were design matrices and e was a vector of random residuals following $N(0, I\sigma_e^2)$. We used the true values of variance components.

Optimum contribution selection. To constrain the number of sires selected, we constructed a 'pseudo-female' out of all female selection candidates. The pseudo-female had a relationship with itself that was equal to the mean relationship among female selection candidates, and a relationship with each bull that was equal to the mean relationship of all female candidates to that bull. The pseudo-female was assigned 30 matings, and each bull was allowed one mating. The 30 selected bulls were then randomly mated to females and used equally. The genetic contributions were optimized to achieve the same target rate of inbreeding, while maximizing genetic gain, and selecting a fixed number of bulls, and using them equally. Genetic contributions *c* were optimized in each generation to maximize the genetic level in the offspring generation, *G*: G=c'â where â was a vector of GEBVs, subject to the constraint: c'Rc=C where R was a relationship matrix, and the target rate of average coancestry was set to 0.005, and the constraint that the sum contributions of males and females each equaled ½ (Meuwissen 1997; Woolliams *et al.* 2015).

We used the first method of VanRaden (2008) to compute GRMs: $G = ZZ^T/(2\Sigma p_j(1-p_j))$, where Z = M – P where M is the genotypic matrix for genotyped animals for which rows are genotypes with values 2 and 0 for the two homozygotes, and 1 for heterozygotes; P is a matrix where all elements in the j^{th} column are $2p_j$, where p_j is the frequency of the allele at locus *j* that was counted in M. We used three estimates of p_j: *Base* used the allele frequencies in the base population, *Recent* used genotyped animals in the current generation, and *All* used all genotyped animals to estimate p_j. Additionally, we used a pedigree-based relationship matrix (*Pedigree*). Thus the study composed of four scenarios of OCS: *All*, *Base*, *Recent* and *Pedigree*.

Statistical analysis. We computed the mean true breeding value and estimated mean identical-by-descent (IBD) based genomic kinship for the selection candidates in each generation. Kinship 'represents the value of chromosome segments of two different individuals to be IBD at a random position when randomly drawing one of the two haplotypes of each individual' (T. Pook, pers. comm.). We used the following linear model for generations six to 16:

$$y = Replicate + Scenario + Scenario \times Generation + error \quad (1)$$

where y was breeding value, kinship or additive genetic variance. We used the standard errors of the linear regression coefficients for significance testing. The interaction term of scenario and generation gave a regression coefficient nested within scenario.

Results

Figure 1 shows the average results of the simulation from generation 1 to 16. Genetic gain was 0.827, 0.846, 0.818 and 0.835 genetic standard deviations; kinship rate was 0.81, 0.84, 0.73 and 0.98%; and loss of additive genetic variance was -3.0, -3.0, -2.9 and -3.6% per generation, for *All*, *Recent*, *Base* and *Pedigree*, respectively, according to the linear model. Pedigree-based OCS resulted in significantly higher rate of increase in kinship than GOCS ($P<0.01$), and *Base* resulted in significantly lower kinship increase than *Recent* and *Pedigree* ($P<0.05$). *Pedigree* resulted in significantly more loss of genetic diversity than genomic OCS ($P<0.05$). The difference in genetic gain among scenarios was not statistically significant.

Discussion

The results suggest that genomic OCS can outperform pedigree-based OCS, contrary to the conclusion of Henryon *et al.* (2019), but agreeing with the recommendation of Sonesson *et al.* (2012). Henryon *et al.* (2019) assessed inbreeding rate at a set of IBD loci not used for prediction, while we assessed true kinship using recombination points and founder chromosome segments. The faster build-up of kinship, and more rapid loss of genetic diversity observed with POCS, suggests that genomic OCS is preferable to pedigree-based OCS under the conditions in this simulation. Our results agree with results of Berg *et al.* (*subm.*).

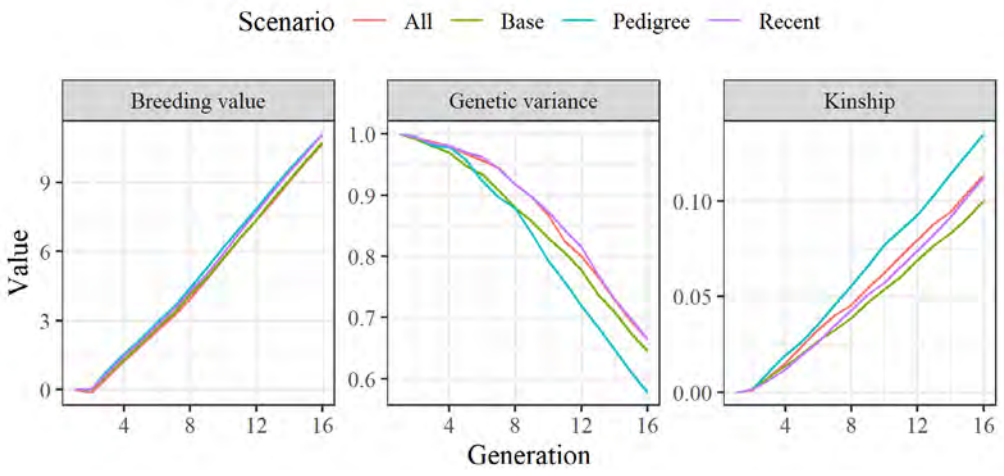


Figure 1. Mean breeding value, additive genetic variance and kinship. Genomic selection started in generation six. Breeding value is in units of additive genetic standard deviation, genetic variance is the proportion of variance in the base population and kinship is based on genomic identity-by-descent.

They found that distant reference frequencies for GOCS resulted in less build-up of kinship than using recent frequencies. We simulated a population with number of markers and animals resembling a feasible breeding scheme for the Icelandic dairy cattle population. This included simulating a breeding scheme that used a fixed number of bulls and equal use of these bulls. While this is a deviation from the theory of optimum contribution selection, we believe that this resembles a practical and cost-effective way for genomic breeding of a small dairy cattle population, since number of AI bulls is a limiting factor in a small closed population. Our results suggest that in such settings, GOCS should be conducted, using old animals to compute reference allele frequencies for the GRM.

References

Berg P., Nielsen J., and Sørensen M.K. (2006) Proc. of the 8th WCGALP. Belo Horizonte, Minas Gerais, Brazil.

Christensen, O.F., and Lund M.S. (2010) Genet. Sel. Evol. 42:2. https://doi.org/10.1186/1297-9686-42-2

Henryon M., Liu H., Berg P., Su G. *et al.* (2019) Genet. Sel. Evol. 51(1):39. https://doi.org/10.1186/s12711-019-0475-5

Madsen P., and Jensen J. (2013) DMU: A Package for Analysing Multivariate Mixed Models.

Makanjuola B.O., Miglior F., Abdalla E.A., Maltecca C. *et al.* (2020) J. Dairy Sci. 103(6):5183–5199. https://doi.org/10.3168/jds.2019-18013

Meuwissen T.H.E. (1997) J. Anim. Sci. 75(4):934–940. https://doi.org/10.2527/1997.754934x

Obšteter J., Jenko J., Hickey J.M., and Gorjanc G. (2019) J. Dairy Sci. 102(11):9971–9982. https://doi.org/10.3168/jds.2019-16853

Pook T., Schlather M., and Simianer H. (2020) G3: Genes Genom. Genet. 10(6):1915–1918. https://doi.org/10.1534/g3.120.401193

Purcell S., Neale B., Todd-Brown K., Thomas L. *et al.* (2007) Am. J. Hum. Genet. 81(3):559–575. https://doi.org/10.1086/519795

Sargolzaei M., and Schenkel F.S. (2009) Bioinformatics. 25(5):680–1. https://doi.org/10.1093/bioinformatics/btp045

Sigurdsson A., and Jonmundsson J.V. (2011) Icel. Agric. Sci. 24(2011):55–64.

Sonesson A.K., Woolliams J.A., and Meuwissen T.H. (2012) Genet. Sel. Evol. 44:27. https://doi.org/10.1186/1297-9686-44-27

Su G., and Madsen P. (2012) User's guide for GMATRIX Ver2.

Thomasen J.R., Sørensen A.C., Lund M.S., and Guldbrandtsen B. (2014) J. Dairy Sci. 97(9):5822–5832. https://doi.org/10.3168/jds.2014-7906

VanRaden P.M. (2008) J. Dairy Sci. 91(11):4414–23. https://doi.org/10.3168/jds.2007-0980

Woolliams J.A., Berg P., Dagnachew B.S., and Meuwissen T.H.E. (2015) J. Anim. Breed. Genet. 132:89–99. https://doi.org/10.1111/jbg.12148

672. The rising popularity of embryo transfer in U.S. dairy cattle and implications for national fertility evaluations

A.M. Miles*, J.L. Hutchison and P.M. VanRaden

Animal Genomics and Improvement Laboratory, Agricultural Research Service, United States Department of Agriculture, Beltsville, MD 20705-2350, USA; asha.miles@usda.gov

Abstract

This research was motivated by the exponential increase in embryo transfer (ET) use with >5 times as many ET calves born in the U.S. in 2021 compared to just 5 years earlier. A survey of the national database revealed that only 1% of ET calves correspond to ET breeding events, 2% are incorrectly reported as artificial insemination, and 97% have no associated breeding event. Herdyears reporting >10% of calves born by ET but less than half of the expected number of ET breeding events were removed, and sire conception rate recalculated with this new dataset. Subsequent analysis showed that censoring herdyears with inconsistent ET reporting has little effect, except in the case of elite young bulls popular for ET use. Improved ET reporting will be critical for providing accurate fertility evaluations, especially as the popularity of these advanced reproductive technologies continues to rise.

Introduction

Commercial embryo transfer (ET) began in the 1970s and today has become common practice in herds desiring to increase their rate of genetic progress (Hasler 2014; Moore and Hasler 2017). Part of the attraction of ET can be explained by substantially reduced generation intervals among habitual ET users. For example, consider a top genetic merit heifer whose embryos were collected starting at 7 months old (Table 1). The maternal line continued this rapid turnover resulting in a remarkable 7 generations in 10 years (average generation interval 16.7 months). Embryos purchased in high enough volumes could cost as little as $125 per embryo, putting ET at a comparable price level to artificial insemination (AI) services (Trans Ova, personal communication). As transfer technologies become more affordable, ET use has grown exponentially with 11% of U.S. dairy calves born in 2021 attributable to ET (Figure 1). However, the trend in ET breeding event reporting (red line) does not parallel the ET calf birth rate (blue line) and this lack of congruence can interfere both with national evaluations and on-farm management of fertility. This discrepancy could be explained by a number of things, including embryo donation or implantation not being reported at all or ET being incorrectly reported as AI. Some of these errors can be eventually corrected when breed associations provide pedigree records, but they arrive 9 months too late as unofficial fertility evaluations are provided to U.S. producers on a weekly basis. Previously, the USA imposed edits for ET donors and reported recipients in our evaluations, and only 3 Interbull-participating countries account for ET by excluding ET records (Interbull Centre, 2021). Completely censoring ET-associated records is not necessarily the desired approach, as these represent the most elite animals and herds, and unreported ET could also bias fertility trait evaluations in the population. In this paper we propose an edit to account for incorrect ET reporting and explore its impact on sire conception rate (SCR), as an example.

Table 1. Maternal line and genotyped embryo of dam HOCAN13913420.

	Maternal generation							
	0	**1**	**2**	**3**	**4**	**5**	**6**	**7**
Birth year	2022	2020	2019	2018	2016	2015	2013	2012
Birth month	3	11	6	3	11	6	12	6
Age at embryo donation (months)	--	7	8	6	7	8	9	9
Generation interval[1]	16	17	15	16	17	18	18	--

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_672, © A.M. Miles *et al.* 2022

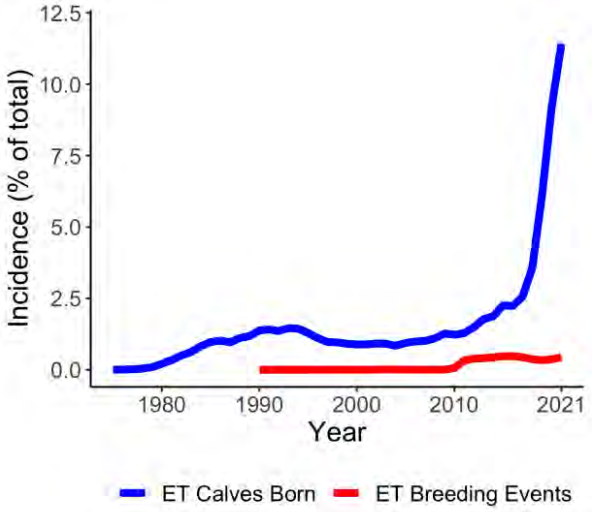


Figure 1. ET use in the USA.

Materials & methods

Data were extracted from the National Cooperator Database maintained by the Council on Dairy Cattle Breeding in December 2021, representing the most current information available on ET use in the USA. Reported mating types were matched with recorded birth types to assess ET reporting error rates. A new dataset was created by removing herdyears that reported >10% of calves born by ET but less than half of the expected ET breeding events given the number of ET calves born. This edit maximizes data preservation while removing the records most likely to introduce confounding bias. All summaries were performed in SAS version 9.4 (SAS Institute, Cary NC, USA) and data visualization performed with R version 4.1.1 (R Foundation for Statistical Computing, Vienna, Austria). This new dataset was used to compute SCR with the same methodology used for the official evaluations. SCR is computed with an all-breed animal model, including Ayrshire, Brown Swiss, Guernsey, Holstein, Jersey, and Milking Shorthorn, and using only the most recent 4 years of data (Council on Dairy Cattle Breeding, 2021). Pearson correlation coefficients between the new and old values were calculated, and the differences in SCR (SCR_{DIFF}) were computed for each bull by subtracting the official values from the new values generated with the proposed edit.

Results

For calves born by ET, breeding events reported 9 months earlier revealed only 1% of recipients correctly reported as ET, 2% incorrectly reported as AI, and 97% with no breeding event reported at all. The ET edit resulted in the removal of 252 herdyears for SCR, accounting for a 1.2% reduction in the total number of records included in the calculations. The number of publishable bulls was reduced for all breeds except Milking Shorthorn, which do not appear to be utilizing ET and only have one publishable bull (Table 2). The median SCR_{DIFF} is close to zero for all breeds and the biggest changes were observed in Holstein bulls (range of -1.9 to 1.3). A wider spread in SCR_{DIFF} is observed among bulls more frequently used for ET (Figure 2A), though SCR_{DIFF} converges on zero as more matings are reported (Figure 2B). ET usage is not strongly correlated with total number of matings, as many of the highest merit young bulls with comparatively few matings are now being used almost exclusively for ET (Figure 2C).

Table 2. Summary statistics of the impact of ET edits on sire conception rate.

		SCR_{DIFF}[2]				
	Publishable bulls[1]	Minimum	Median	IQR[3]	Maximum	Correlation[4]
Ayrshire	6 (7)	-0.3	-0.1	0.4	0.3	0.98*
Brown Swiss	44 (48)	-0.9	-0.1	0.3	0.6	0.99*
Guernsey	14 (17)	-0.6	0.1	0.6	0.6	0.96*
Holstein	1,379 (1,454)	-1.9	0.0	0.2	1.3	0.99*
Jersey	132 (141)	-1.4	-0.1	0.3	0.9	0.99*
Milking Shorthorn	1 (1)	0.0	0.0	0.0	0.0	--

[1] Publishable bulls with ET edit (publishable bulls without edit).

[2] Difference in SCR, computed as new values – old values.

[3] Interquartile range.

[4] Correlation between the new and old values. *$P<0.0001$.

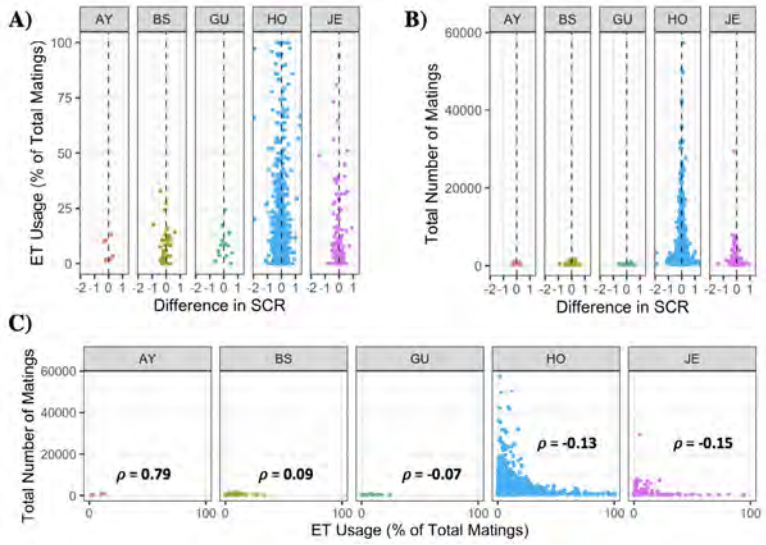


Figure 2. By-breed differences in SCR plotted against (A) percent of matings related to ET for each bull, (B) total number of matings for each bull, and (C) the correlation of total number of matings and popularity of each bull for ET. AY (Ayrshire), BS (Brown Swiss), GU (Guernsey), HO (Holstein), JE (Jersey), ρ = Pearson correlation coefficient.

Discussion

The new SCR computed from data filtered for ET was consistently highly correlated to the official values as shown in Table 2. The small effect of this edit can also be appreciated by examining the SCR_{DIFF} for individual bulls. The wider range of SCR_{DIFF} values among popular ET bulls was expected given that records were censored on the basis of ET usage. By plotting SCR_{DIFF} against the total number of reported matings for each bull, we see that as more records are added SCR_{DIFF} converges on zero, suggesting that the true effect of ET on SCR for proven bulls is nearly zero. This trend is most obvious in Holstein and Jersey, due to the larger amount of data available. It is notable that even for non-zeros, the difference is small (1-2 percentage points) and will have overall little effect. However, most breeds indicate a negative correlation

between the total number of matings and the amount that bull is used for ET, suggesting that popular, proven bulls are not being prioritized for ET over young bulls. This explains anecdotal reports of young bulls whose SCR estimates may change a lot as more records are added or edited out, especially if their daughters are nearly all born by ET. Similar edits for unreported ET had larger effects on heifer conception rate than cow conception rate or SCR and might be needed for other fertility-related traits such as daughter pregnancy rate, gestation length, and early first calving. Acquiring high-quality ET records (comprising details like multiple ovulation ET versus *in vitro* fertilization, fresh versus frozen, embryo grade and stage, recipient synchrony, follicle stimulating hormone protocol, numbers of degenerate embryos, etc.) can help partition genetic effects among fertility, conception rate, and pregnancy rate. The Beef Improvement Federation recently approved guidelines for using ET records and necessary considerations in evaluation models, such as accounting for the effects of both the donor (e.g. calf genetics) and recipient (e.g. birth weight) dams. These principles could also be adopted by the dairy industry with the establishment of new data pipelines. Currently, unstandardized ET reporting is having minimal effect, but that is not a guarantee if the ET calf birth rate continues to rise without a parallel increase in accurate ET breeding event reporting.

Conclusions

The rapid increase of ET is likely to continue as advanced reproductive technologies become more affordable. An investigation of SCR shows that censoring herdyears without consistent ET reporting has overall negligible effect, except for young bulls. Even though the SCR correlations are high with small effects, the biggest changes are observed in elite new bulls which have a huge influence on the breeding program. There is an urgent need to improve ET reporting to facilitate the delivery of accurate fertility evaluations.

References

Beef Improvement Federation, Embryo Transfer (ET): Data Collection and Utilization 2021-04-13. Available at: http://guidelines.beefimprovement.org/index.php/Embryo_Transfer_(ET):_Data_Collection_And_Utilization

Council on Dairy Cattle Breeding (CDCB), Code of Practice, Male Fertility 2021-04-01. Available at: https://queries.uscdcb.com/reference/Form_PE_MFertility-SCR_1204.pdf

Hasler, J.F. (2014) Theriogenology 81:152-69. https://doi.org/10.1016/j.theriogenology.2013.09.010

Interbull Centre, National Genetic Evaluation Forms Provided by Countries 2021-12-16. Available at: https://interbull.org/ib/geforms

Moore, S.G., and J.F. Hasler (2017) J Dairy Sci 100:10314-31. https://doi.org/10.3168/jds.2017-13138

673. Simulating technical, economic and environmental performance of a dairy herd under selection in future climate conditions

J.G. Burns[1,2*], V. Eory[1], R. Mrode[1,3], G. Simm[2] and E. Wall[1]

[1]Scotland's Rural College (SRUC), Edinburgh, Scotland, EH9 3JG, United Kingdom; [2]The Global Academy of Agriculture and Food Security, University of Edinburgh, Roslin, Scotland, EH25 9RG, United Kingdom; [3]International Livestock Research Institute, Kenya, Box 30709-01001 Nairobi, Kenya; jay.burns@sruc.ac.uk

Abstract

Climate change will have direct effects on performance of livestock. Simulations that mimic livestock systems risk misrepresenting performance if these effects are not accounted for. Here, the technical, economic and environmental performance of a Scottish dairy herd under genetic selection (for 7-correlated traits) was investigated using an original simulation model. There were two scenarios: in the control scenario, the technical performance of individuals was not affected by daily weather; in the treatment scenario, technical performance was dynamic to daily weather conditions. Herds were simulated for 20-years. By the final year of the simulation, for herds in the treatment scenario relative to the control scenario, annual gains in gross margin were reduced by 2.1% and gains in greenhouse gas emissions intensity were reduced by 1.7%. Failure to account for weather effects in models that simulate herds into the future will likely cause over-estimation of technical, economic and environmental performance, even in temperate regions such as southwest Scotland.

Introduction

It is likely that global temperatures will continue to rise. This, together with increasing magnitude and frequency of extreme weather events, will present major challenges to livestock production in the coming years (IPCC, 2021). The production of ruminant livestock, such as dairy cows, is under particular pressure given that some negative externalities of the system (e.g. greenhouse gas – GHG – emissions, which not only include enteric emissions but also emissions from manure management and managed soils – for details, see Gavrilova *et al.*, 2019) will exacerbate existing environmental challenges. In the face of such challenges, dairy production must remain efficient, which means the system and its components must be resilient (Friggens *et al.*, 2017). This resilience may be improved with artificial selection (e.g. Poppe *et al.*, 2020), however, the value of resilience at the animal-level is not well established, and is likely to be proportionate to the degree of challenge expected for a given system. Against this back-drop, we present the application of an individual-based model (IBM) under different scenarios of future climate, where the objective is to demonstrate the value of herds if individuals are resilient – or not – to variable daily weather conditions.

Methods

Using an original, daily time-step, IBM, we mimic the daily biological performance and management of a herd of individuals, where individuals are represented according to a set of underlying genetic characteristics, covering production and functional traits. As in real life, individuals differ in their genetic ability, and, over time, the genetics of the herd improves, driven by breeding dams to genetically improved sires and subsequent selection of genetically superior replacements. This, together with stochastic functions in the model, generates variability in the performance of individuals, resulting in realistic variation in the performance of the herd. The performance of individuals can also be dynamic to key weather variables, based on scenarios of future climate change. In terms of model outputs, individual performance data were collected daily and used to calculate raw technical performance, economic outcomes and GHG emissions at the herd level.

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_673, © J.G. Burns *et al.* 2022

The model was parameterised to be representative of a specialist dairy herd in southwest Scotland. Therefore, the simulated herd consisted of approximately 200 lactating cows; they were housed for 6 months per year (October – March, inclusive), and housed part of the day for the remainder, as is common in the United Kingdom (March *et al.*, 2014). The herd was simulated for 20-years into the future (2021-2040).

The daily performance of herd-members was affected by 7-traits of economic importance, these were: milk yield, fat yield and protein yield (production), dry matter intake (maintenance), probability of mastitis infection (health), number of inseminations to conception (fertility), and days of life (longevity). The daily phenotypic performance of each individual across these traits was a function of their genetic ability and environmental factors (including daily weather). According to these phenotypic functions, each day individuals would yield a given quantity and quality of milk, would consume a given quantity of dry matter, may succumb to mastitis infection, may conceive, and/or may be culled, respectively. Main interactions of productivity, reproduction and health occurring during the productive lifetime were integrated. However, the simulation was not an optimisation model, therefore, herd management rules remained static between climate scenarios and across time. In addition to this, we assume no adoption of climate mitigation technologies or practices by the farmer.

In one scenario of the simulation (herein, the treatment scenario), the explicit weather-dependence of daily phenotypes enabled prediction of the degree and extent of weather effects on herd performance. Here, weather affected daily yields, feed intake, and rates of mastitis, conception and mortality. This allowed us to capture the cumulative impact of seasonal and long-term weather variations on herd performance. As well as simulation runs that accounted for climate change in this way, the simulation was also run with static weather variables that did not reflect climate change, which acted as a baseline where animals were resilient and therefore unperturbed by variable weather (herein, the control scenario). Weather variables came from UK Climate Projections 2018 (UKCP18), downscaled for the target area.

The technical output and production characteristics of the herd were used to calculate economic and environmental performance. Gross margin was calculated as income (e.g. from milk sales) minus variable costs (e.g. pregnancy testing). For the environmental performance, we calculated enteric and manure methane, and direct, volatilised and leached nitrous oxide from manure; these emissions were calculated according to the IPCC 2019 refinement to the IPCC 2006 national GHG inventory (Gavrilova *et al.*, 2019). We did not directly calculate emissions from spreading manure on managed soils, as we assumed these emissions are captured by feed component emissions factors (EFs). Dietary assumption and associated EFs were provided by consultants at SRUC as representative for the modelled system. Emissions intensity is reported as total emissions of the lactating herd per unit of energy-corrected milk (kg CO_2-eq/kg ECM). Total emissions of the herd were calculated as kg CO_2-eq, using the IPCC AR5 GWP100 conversion. ECM accounts for yield, fat and protein content of milk (Sjaunja *et al.*, 1990).

Results

Results show that estimates of technical, economic and environmental performance were significantly different between simulation runs for control and treatment scenarios (statistical model outputs not shown). Accounting for the effects of climate change on the technical performance of individuals led to unfavourable changes in several key herd parameters: for example, an overall decrease in productive lifespan and increased incidence of mastitis (see Figures 1a-c). In addition, we found similarly unfavourable results for age at first calving, calving index, conception rate and ECM.

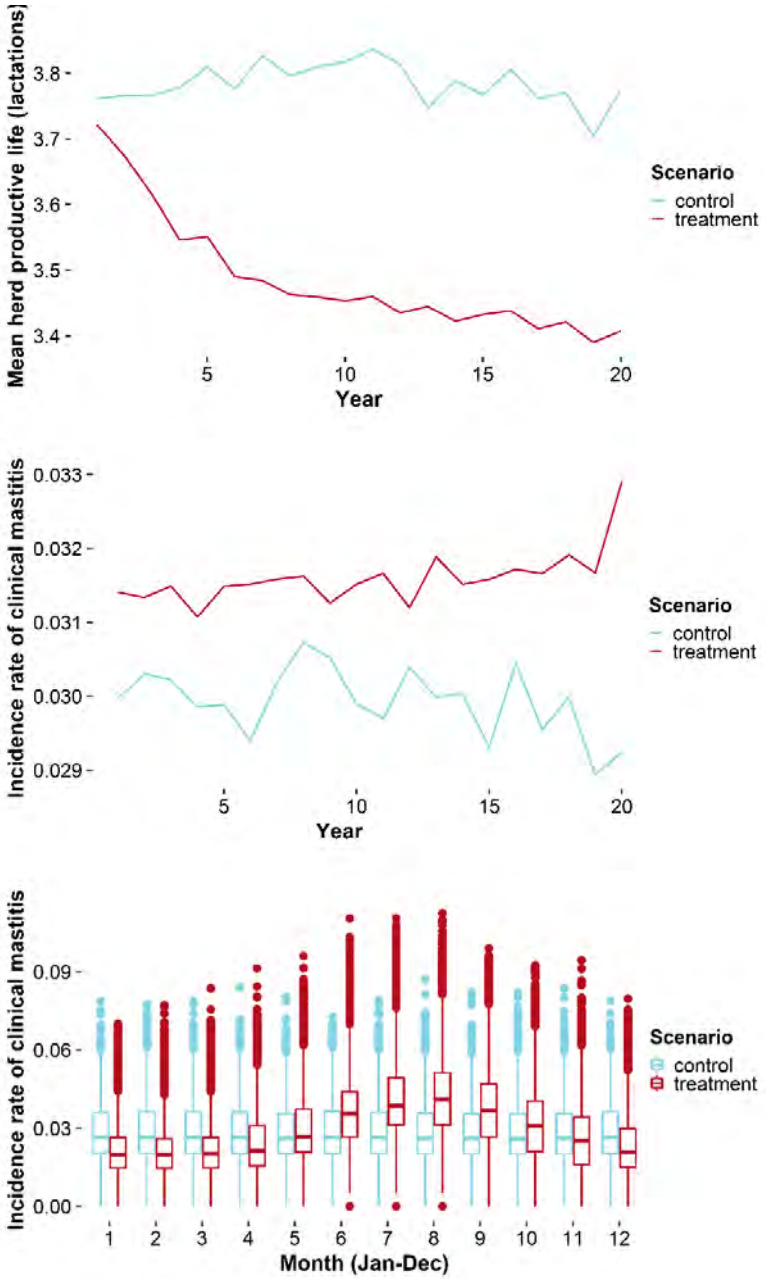


Figure 1 Plots to demonstrate the impacts of key weather variables on (a) productive lifespan across years, (b) incidence rate of clinical mastitis across years, and (c) incidence rate of clinical mastitis within years. Treatment scenario refers to simulation runs that account for climate change, and control scenario refers to simulations run with static weather variables. Please carefully note the y-axis scales when interpreting these plots.

Accordingly, relative to the control scenario, where individuals were robust to variation in daily weather conditions, gains in gross margin were reduced by 2.1%, and gains in GHG emissions intensity were reduced by 1.7% in the treatment scenario. With 900 herds in the Scottish dairy system, this reduction in gross margin could amount to £15m for the industry – and the difference in emissions intensity could lead to an extra 25,000 tonnes CO_2-eq– resulting from less efficient production over the next 20-years, even in the temperate climate of southwest Scotland.

Discussion

Without cattle who are resilience to the impacts of projected weather over the coming decades, we can expect negative impacts on the technical, economic and environmental performance of dairy herds, even in temperate regions. Resilience to these direct effects may be introduced through selection for traits that promote individuals who are able to maintain production in the face of environmental perturbations by, for example, accounting for genetic-by-environment interactions in breeding decisions. Animals who are resilient will not only be more profitable but also produce better quality milk in greater quantities at a lower emissions intensity. As research into breeding for resilient animals continues (e.g. www.GenTORE.eu), a critical question is what value should be placed on these novel traits. Towards this question, the current paper is an important contribution.

However, also important are the methodological implications, which suggest that failing to account for the effects of climate at the animal-level in such models will cause the misrepresentation of key performance indicators, especially as this type of model becomes increasingly popular and simulated scenarios can run for many years into the future. Therefore, if researchers aim to mimic the biological performance of a herd, it is increasingly difficult to overlook relationships between climate and herd functioning, even in temperate regions.

Conclusions

This paper presents the results of a novel approach to estimating the value of resilience at the animal-level in a way that can be tailored to specific regions within production systems. We report substantial differences in individual and herd performance between production scenarios.

References

Friggens, Blanc, F., Berry, D.P., Puillet, L., 2017. Animal. 11, 1–15. https://doi.org/10.1017/S175173111700088X

Gavrilova, O., Leip, A., Dong, H., MacDonald, J.D., *et al.* 2019. Emissions from livestock and manure management, in: 2019 Refinement to the 2006 IPCC Guidelines for National Greenhouse Gas Inventories. Available at: https://www.ipcc-nggip.iges.or.jp/

IPCC, 2021. Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge Univ. Press. UK

March, M.D., Haskell, M.J., Chagunda, M.G.G., Langford, F.M., Roberts, D.J., 2014. J. Dairy Sci. 97, 7985–7994. https://doi.org/10.3168/jds.2014-8265

Poppe, M., Veerkamp, R.F., van Pelt, M.L., Mulder, H.A., 2020. J. Dairy Sci. 103, 1667–1684. https://doi.org/10.3168/jds.2019-17290

Sjaunja, L. O., Baeve, L., Junkkarinen, L., Pedersen, J. and Setälä, J. 1990. Twenty-seventh session of the International Committee of Recording and Productivity of Milk Animals, Paris, pp. 156-157

Part 38

Bovine dairy – genome informed breeding

674. Understanding the genetic architecture of claw horn lesions in different lactation stages in Holstein cattle

B. Li[1*], M. Barden[2], V. Kapsona[1], E.S. Molano[3], A. Anagnostopoulos[2], B.E. Griffiths[2], C. Bedford[2], X. Dai[4], M. Coffey[1], A. Psifidi[4], G. Oikonomou[2] and G. Banos[1*]

[1]Scotland's Rural College (SRUC), Roslin Institute Building, Easter Bush, Midlothian EH25 9RG, United Kingdom; [2]Department of Livestock and One Health, Institute of Infection, Veterinary and Ecological Sciences, University of Liverpool, Leahurst Campus, CH64 7TE, Neston, Liverpool, United Kingdom; [3]The Roslin Institute and R(D) SVS, University of Edinburgh, Easter Bush, Midlothian EH25 9RG, United Kingdom; [4]Department of Clinical Science and Services, Royal Veterinary College, Hawkshead Lane, AL9 7TA, Hatfield, Hertfordshire, United Kingdom; bingjie.li@sruc.ac.uk; georgios.banos@sruc.ac.uk

Abstract

Lameness in dairy cattle is primarily caused by foot lesions including sole haemorrhage (SH) and sole ulcers (SU). This study investigated the genetic architecture of SH and SU in different lactation stages based on accurate phenotyping, genetic parameter estimation, genome-wide association (GWA) and functional enrichment analyses. Foot lesion records were collected from 2,353 Holstein cows on four herds at four time-points: prior to calving, immediately after calving, early lactation, and late lactation stages. Heritability estimates were 0.11-0.20 for SH and 0.05-0.13 for SU across stages, and genetic correlations between stages were high for both traits. Candidate genes associated with SH and SU link to immune functions such as complement activation and inflammation, nervous system and muscle functions. To conclude, SH and SU are under genetic control and foot health can be improved with selective breeding.

Introduction

Lameness in dairy cattle is primarily caused by foot lesions including sole haemorrhage (SH) and sole ulcers (SU) as two common claw horn lesions (CHL). Recent studies have shown genetic variation in resistance to foot lesions (Heringstad *et al.*, 2018). A few genome-wide association (GWA) analyses have shown a complex genetic background for foot health (e.g. Swalve *et al.*, 2014; Sánchez-Molano *et al.*, 2019; Butty *et al.*, 2021; Lai *et al.*, 2021), but results generally differ between studies. Most previous studies utilized foot-trimming data that may be subject to variability between recorders and populations and may miss some mild lesion cases.

In the present study we closely monitored cohorts of cows across lactation and collected detailed foot lesion records. The objectives were to: (1) estimate genetic parameters of CHL in different lactation stages; and (2) characterise the genomic architecture of CHL.

Materials & methods

Animals and data recording. A total of 2,353 Holstein cows from four herds in the United Kingdom were included. The studied cows calved between April and September 2019. Individual animals were assessed for SH and SU at four time points: prior to calving (pre-calving stage, mean=55.3 days before calving), immediately after calving (calving stage, mean=5.4 days in milk), at peak yield time (early lactation stage, mean=83.9 days in milk), and in mid-late lactation (late lactation stage, mean=199.6 days in milk).

Foot lesions were recorded by qualified veterinary surgeons using case definitions in the ICAR claw health atlas (Egger-Danner *et al.*, 2020). Over 90% of foot lesion identification and recording was performed by a single researcher. All four feet were examined for each lesion, which were scored as 0 (no lesion), 1 (mild lesion), 2 (moderate lesion) and 3 (severe lesion). The pedigree spanning seven generations was available for all phenotyped animals.

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_674, © B. Li *et al.* 2022

Phenotypes. The phenotype of each lesion severity was calculated as the average score across all assessed feet. The phenotype was calculated for each studied time-point, so each animal had repeated records across lactation stages.

Genotypes. Genotypes were available for 2,250 of the studied animals. The genotypes had been imputed to 80K SNPs within the national dairy cattle genomic evaluation. Chromosome locations of the 80K SNP panel were based on ARS-UCD 1.2. Genotype quality control kept SNP and animal call rates higher than 90%, SNPs with minor allele frequency higher than 0.05, and animals with no parent-progeny Mendelian conflicts.

Genetic parameter estimation. Variance components and heritability for SH and SU were estimated for each lactation stage using the airemlf90 software (Misztal, 2013). Multivariate analyses were conducted to estimate correlations between stages. The model used was:

$$y = Xb + Za + e \quad (1)$$

where y is the phenotype; b is the fixed effect for parity and herd-year-season of recording; age effect was not included because variance explained by age effect has been explained by its correlated effect of parity; a is additive genetic effect with var(a) ~ $N(0, H\sigma_a^2)$, where σ_a^2 is the additive genetic variance and H is the relationship matrix incorporating pedigree and genomic information (Legarra *et al.*, 2009); e is random residual with var(e) ~ $N(0, I\sigma_e^2)$; and X and Z are incidence matrices for b and a, respectively.

Subsequently, repeated records of individuals across lactation were combined and analysed with an animal repeatability model. The model was as model (1) with the addition of a fixed effect for lactation stage and a random permanent environmental effect. Single-step GBLUP animal GEBVs were estimated by the animal repeatability model and blupf90 software (Misztal, 2013).

Genome-wide association analyses. Animal GEBVs were back-solved to obtain SNP effects for each trait, using the postGSf90 software (Wang *et al.*, 2012). In addition, sliding genomic windows with a size of 0.65 Mb were constructed to calculate the proportion of genetic variance explained by each genomic window. The window size of 0.65 Mb was determined based on the average distance in the genome where linkage disequilibrium between markers halved. The proportion of genetic variance explained by sliding windows was calculated with the method described by Wang *et al.* (2014) with postGSf90 software.

Candidate genes located close to the large-effect variants within 0.2 Mb upstream and downstream or within the candidate regions were examined with functional enrichment analyses using the DAVID bioinformatic resource.

Results & discussion

Incidence rates of SH and SU. The incidence rate of SH in the study population (83.0%) was higher than that of SU (13.9%) over the whole lactation. The higher prevalence of SH than SU can be partly explained by the fact that SU cases may be developed from severe SH cases. The incidence rates of SH and SU were at the highest at early lactation stage (57.5% for SH, and 6.1% for SU). Our findings infer a higher risk of SH and SU in early lactation with high yield. Unfavourable genetic correlations between foot lesions and milk yield were reported previously (Koenig *et al.*, 2005).

Genetic parameters for CHL. The heritability estimates ranged from 0.11 to 0.20 for SH, and 0.05 to 0.13 for SU across stages, with no significant differences between stages (Table 1). Although the phenotypic correlations between stages were generally lower than 0.50 for both traits, their genetic correlations were

much higher, especially for SU (Table 1). The high genetic correlations suggest that CHL may have similar genetic background across lactation.

Single-marker GWA results. Significant SNPs were identified on BTA1 for SH, and on BTA8 and BTA5 for SU (Table 2). Genes located closest to these SNPs were *PGM5, TAC3, and BACE2*. The *PGM5* gene was associated with human lower-limb muscle injury (Aguilar *et al.*, 2015). The *TAC3* gene links to nervous system function, smooth muscle contraction, and leukocyte gene expression after burn injury or trauma-haemorrhage in mice (Topaloglu *et al.*, 2009). The *BACE2* gene is related to bovine muscle formation (Lee *et al.*, 2012).

Genomic windows associated with CHL. For SH, highest proportions of genetic variance were explained by genomic regions on BTA20, BTA22 and BTA3 (Table 3). The region on BTA3 was associated with digital dermatitis (Sánchez-Molano *et al.*, 2019). Genes within the region of BTA3 (*C8B, C8A*) link to complement activation. Genes within the regions on BTA20 and BTA22 link to immune functions including inflammation regulation (*OTULIN, OTULINL*), leukocyte migration (*TRIO),* and T cells cytokine production (*BHLHE40*).

Respective regions for SU were found on BTA18 and BTA14 (Table 3). The *CHD9* gene within the region on BTA18 is involved in osteogenesis. Genes within the regions on BTA14 link to neutral system function (*KCNQ3*) and inhibition of cell growth and proliferation (*KHDRBS3*).

Gene enrichment. Enrichment of the above-mentioned genes link to Gene Ontology terms of DNA transcription regulation (*OTULIN, BHLHE40*), KEGG pathways of complement and coagulation cascades (*C8A, C8B*), cholinergic synapse (*ITPR1, KCNQ3*), and Reactome pathway of regulation of complement cascade (*C8A, C8B*).

Table 1. Heritability estimates (diagonal, in bold) for sole haemorrhage (SH) and sole ulcer (SU) at pre-calving stage, calving stage, early and late lactation stages, and genetic correlations (upper-diagonal) and phenotypic correlations (below diagonal) between stages; standard errors of estimates are in brackets.

Trait		Pre-calving	Calving	Early	Late
	Pre-calving	**0.11** (0.03)	0.98 (0.02)	0.78 (0.05)	0.80 (0.04)
SH	Calving	0.35 (0.02)	**0.14** (0.03)	0.74 (0.05)	0.72 (0.07)
	Early	0.27 (0.02)	0.21 (0.02)	**0.20** (0.04)	0.98 (0.02)
	Late	0.25 (0.02)	0.16 (0.02)	0.36 (0.02)	**0.13** (0.04)
	Pre-calving	**0.13** (0.04)	0.97 (0.01)	0.96 (0.01)	0.96 (0.02)
SU	Calving	0.38 (0.02)	**0.07** (0.03)	0.92 (0.03)	0.97 (0.01)
	Early	0.29 (0.02)	0.21 (0.02)	**0.05** (0.02)	0.96 (0.01)
	Late	0.37 (0.02)	0.39 (0.02)	0.44 (0.02)	**0.08** (0.04)

Table 2. Significant markers for sole haemorrhage (SH) and sole ulcer (SU) in single-marker GWA analyses with respective chromosome (BTA) and position (Pos), minor allele frequency (MAF), *P*-value, closest gene, and distance to the gene.

Lesion	BTA	Pos	MAF	*P*-value	Gene	Distance (bp)
SU	8	44,652,431	0.08	6.76E-06	*PGM5*	39,594
	8	44,735,178	0.08	8.21E-06	*PGM5*	122,341
	5	56,420,099	0.06	1.37E-05	*TAC3*	2,087
SH	1	141,365,527	0.21	2.22E-05	*BACE2*	16,070

Table 3. Genomic regions that explain the highest proportions of genetic variance (Var%) for sole haemorrhage (SH) and sole ulcer (SU), including chromosome (BTA) and positions (Pos_start, Pos_end) and genes located within.

Trait	BTA	Pos_start	Pos_end	Var%	Genes
SH	20	58,161,594	58,795,440	0.78	*ANKH, OTULIN, OTULINL, TRIO*
	22	21,284,764	21,911,533	0.63	*BHLHE40, ITPR1, SUMF1*
	3	88,750,416	89,386,272	0.50	*C8B, C8A*
SU	14	8,954,477	9,597,429	0.93	*KCNQ3, EFR3A*
	18	21,454,669	22,081,129	0.69	*CHD9, RBL2, AKTIP, FTO*
	14	5,922,777	6,526,644	0.53	*KHDRBS3*

Conclusions

SH and SU are under genetic control and animal resistance to CHL may be improved with selective breeding. The CHL may have highly similar genetic background in different lactation stages with high genetic correlations. As polygenic traits, SH and SU were associated with immune functions such as complement activation and inflammation, nervous system and muscle functions, DNA transcription regulation, and cell growth and proliferation.

Acknowledgements

Research was funded by the Biotechnology and Biological Sciences Research Council.

References

Aguilar, C. A., Shcherbina, A., Ricke, D. O., Pop, R., Carrigan, C. T., *et al.* (2015). *Sci. Rep.* 5, 13885. https://doi.org/10.1038/srep13885.

Butty, A. M., Chud, T. C. S., Cardoso, D. F., Lopes, L. S. F., Miglior, F., *et al.* (2021). J Dairy Sci. 104, 8050-8061. https://doi.org/10.3168/jds.2020-19879.

Egger-Danner, C., Nielsen, P., Fiedler, A., Muller, A., Fjeldaas, T., *et al.* (2020). ICAR. Available at: https://www.icar.org/ICAR_Claw_Health_Atlas.pdf.

Heringstad, B., Egger-Danner, C., Charfeddine, N., Pryce, J. E., Stock, K. F., *et al.* (2018). *J. Dairy Sci.* 101, 4801–4821. https://doi.org/10.3168/jds.2017-13531.

Koenig, S., Sharifi, A. R., Wentrot, H., Landmann, D., Eise, M., and Simianer, H. (2005). *J. Dairy Sci.* 88, 3316–3325. https://doi.org/10.3168/jds.S0022-0302(05)73015-0.

Lai, E., Danner, A. L., Famula, T. R., and Oberbauer, A. M. (2021). *Front. Genet.* 12, 1–15.

Lee, E.J., Lee, H.J., Kamli, M.R., Pokharel, S., *et al.* (2012). *Genomics* 100, 195–202.

Legarra, A., Aguilar, I., and Misztal, I. (2009). *J. Dairy Sci.* 92, 4656–4663.

Misztal, I. (2013). Available at: http://nce.ads.uga.edu/html/projects/blupf90.pdf.

Sánchez-Molano, E., Bay, V., Smith, R. F., Oikonomou, G., and Banos, G. (2019). *Front. Genet.* 10, 1–9. https://doi.org/10.3389/fgene.2019.00926.

Swalve, H. H., Floren, C., Wensch-Dorendorf, M., Schöpke, K., Pijl, R., *et al.* (2014). *J. Dairy Sci.* 97, 507–519. https://doi.org/10.3168/jds.2013-6997.

Topaloglu, A. K., Reimann, F., Guclu, M., Yalin, A. S., Kotan, L. D., *et al.* (2009). *Nat. Genet.* 41, 354–358. https://doi.org/10.1038/ng.306.

Wang, H., Misztal, I., Aguilar, I., Legarra, A., *et al.* (2014). *Front. Genet.* 5, 1–10.

Wang, H., Misztal, I., Aguilar, I., Legarra, A., Muir, W. M. (2012). *Genet. Res. (Camb).* 94, 73–83. https://doi.org/10.1017/S0016672312000274.

675. Shared response pattern in the mammary transcriptome after intramammary challenge with *E. coli* or *S. aureus* pathogen

D. Becker[1*], A. Heimes[1], R. Weikard[1], M. Hoedemaker[2], M.M. Meyerholz[3,4], H.J. Schuberth[3], W. Petzl[4], H. Zerbe[4], M. Schmicke[5], S. Engelmann[6,7] and C. Kühn[1,8]

[1]Research Institute for Farm Animal Biology (FBN), Institute of Genome Biology, Germany; [2]Clinic for Cattle, University of Veterinary Medicine Hannover, Germany; [3]Institute for Immunology, University of Veterinary Medicine Hannover, Germany; [4]Clinic for Ruminants with Ambulatory and Herd Health Services, Centre for Clinical Veterinary Medicine, Ludwig-Maximilians-University Munich, Germany; [5]Faculty of Natural Sciences III, Martin-Luther University Halle-Wittenberg, Germany; [6]Technical University Braunschweig, Institute for Microbiology, Germany; [7]Helmholtz Centre for Infection Research, Microbial Proteomics, Germany; [8]Agricultural and Environmental Faculty, University of Rostock, Germany; becker.doreen@fbn-dummerstorf.de

Abstract

In order to evaluate the mammary gland transcriptome response to an early lactation intramammary challenge with viable clinical mastitis pathogen isolates, 24 and 12 German Holstein cows in their early first lactation were challenged with *Staphylococcus aureus* or *Escherichia coli*, respectively. RNA from challenged and control mammary gland tissue was isolated and after depletion of milk and whey protein gene transcripts subjected to RNA sequencing using the Illumina HiSeq 2500 system. A differential expression analysis of the transcriptome of challenged and control mammary gland samples within animal was conducted. An Ingenuity Pathway Analysis (IPA) was performed using the detected differentially expressed genes (DEGs). Both pathogen challenges showed an overlap regarding inactivation of pathways related to energy and lipid metabolism and activating and deactivating effects on immune response pathways.

Introduction

The infection and inflammation of the mammary gland is one of the most frequent and costly diseases in dairy cows (Halasa *et al.* 2007), which can seriously damage the health of affected animals, if left untreated. The course of the disease is crucially influenced by the pathogen species. Mastitis can be caused by gram-positive and gram-negative bacteria. Of particular interest are *Staphylococcus aureus*, mostly inducing chronic and subclinical mastitis, and *Escherichia coli*, predominantly inducing an acute and clinical mastitis. While most previous studies investigated mid-lactation cows of different lactations and used inactivated pathogens or bacterial endotoxins, we focussed on the early first lactation period and challenged the cows with a viable pathogen. In a comprehensive multi-tissue project, a holistic approach was used in order to gain deeper insights into the underlying biological networks affected by an intramammary challenge via monitoring the local (mammary gland), intermediate (blood) and peripheral (liver) response. The aim of this part of the study was to characterize the local (mammary gland) response to an intramammary challenge with *S. aureus* or *E. coli* in primiparous German Holstein cows.

Materials & methods

Animals and study design. For this study, thirty-six healthy German Holstein heifers shortly before first calving were selected. At approximately 36 days after parturition, 24 of these animals were challenged intramammarily with *S. aureus* in both hind quarters of the mammary gland. Twelve animals were challenged with *E. coli* in one hind quarter of the mammary gland. In both groups, one udder quarter of each animal was infused with sterile sodium chloride solution (NaCl) as a control. This challenge scheme had already been established and successfully applied in previous studies (Jensen *et al.* 2013, Petzl *et al.* 2012). Animals were sacrificed 96 hours or 24 hours after challenging with *S. aureus* or *E. coli*, respectively. At these time points, peaks of inflammation of the mammary gland had been observed in previous mid-

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_675, © D. Becker *et al.* 2022

lactation challenge experiments (Petzl *et al.* 2008, Petzl *et al.* 2012). Mammary gland parenchymal tissue samples of all udder quarters were collected during necropsy of the animals, immediately frozen in liquid nitrogen and stored at -80 °C. All ethical evaluations were performed as required by the German Animal Care Law and associated legislative regulations.

Transcriptome analysis. The effect of the intramammary challenge on global gene expression was monitored by comparative transcriptome analysis of challenged and NaCl-infused udder quarters. Therefore, total RNA was extracted as described earlier (Weikard *et al.* 2012). Prior to library preparation, highly abundant milk and whey protein gene transcripts were depleted using a previously established protocol (Brodhagen *et al.* 2019). Stranded indexed libraries were prepared using a TruSeq RNA Library Preparation Kit (Illumina) including polyA-selection. After quality control using a Bioanalyzer 2100 (Agilent), RNA libraries were subjected to paired-end sequencing on the Illumina HiSeq 2500 system.

Bioinformatic analysis. For data processing and analysis, SAMtools (Li *et al.* 2009) as well as Linux and R-based scripts were used. After quality trimming, processed reads were aligned to the bovine reference genome ARS-UCD1.2 with Ensembl 95 reference annotation using Hisat2 v2.1.0 (Pertea *et al.* 2016). Additionally, a guided transcript assembly was performed using StringTie v1.3.5 (Pertea *et al.* 2016) to allow the assembly of reads for unannotated transcripts. Differential expression analysis based on count matrices established via featureCounts v.1.6.2 (Subread package) (Liao *et al.* 2014) was performed using DESeq2 v1.18.1 (Love *et al.* 2014). The identified differentially expressed genes (DEGs) were submitted to IPA (Qiagen) using a threshold for significance of $P_{adj}<0.05$.

Results

Using the milk protein transcript depletion method, a total of 14,546 gene transcripts annotated in the Ensembl database could be detected. In both pathogen groups, clearly different expression patterns were found between the challenged and control udder quarters.

***E. coli* challenge.** The mammary transcriptome analysis revealed a total of 6,727 differentially expressed genes between challenged and control udder quarter (3,453 genes higher and 3,274 genes lower expressed). Among genes with highest expression differences (>8 log fold change) in response to pathogen challenge were acute-phase protein (*SAA2*, *HP*, *DEFB10*) and chemokine genes (*CCL20*, *CXCL2*). As the expression of 46% of all annotated genes was modulated by *E. coli* challenge, we only included DEGs with a $P_{adj}<0.01$ in IPA to highlight the most severely affected pathways. IPA revealed 317 significantly enriched canonical pathways in the mammary transcriptome in response to intramammary *E. coli* challenge. The pathways with the lowest activation score relate to lipid metabolism, whereas the most significantly activated pathways affect the immune system (Table 1). In addition, *SOCS1* and *IL10RA,* known to be repressors of the immune response, were identified as inactivating upstream regulators.

***S. aureus* challenge.** After intramammary challenge with *S. aureus,* 4,637 genes were differentially expressed (2,529 higher, 2,108 lower expressed). Similar to the *E. coli* challenge, genes of the acute-phase response (*SAA3*, *CFB*) as well as chemokine genes and genes encoding their receptors (*CXCL8*, *CXCR1*) were upregulated. Using IPA, we obtained 293 significantly enriched canonical pathways after the *S. aureus* challenge. Analogous to the *E. coli* challenge, pathways related to energy metabolism showed inactivation, whereas a variety of immune response pathways were activated (Table 2). Micro-RNAs and *SOCS1* were identified as inactivating upstream regulators.

Table 1. Five canonical pathways with the highest and lowest activation score (*Escherichia coli* challenge).

Ingenuity canonical pathways	-log(*P*-value)	z-score
PPAR signaling	6.87E+00	-6.091
White adipose tissue browning pathway	1.42E+00	-4.352
LXR/RXR activation	4.25E+00	-3.772
PPARα/RXRα activation	1.04E+01	-3.474
Dopamine-DARPP32 feedback in cAMP signaling	2.23E+00	-3.395
IL-6 signaling	8.47E+00	5.485
Neuroinflammation signaling pathway	9.49E+00	5.438
Systemic lupus erythematosus in B cell signaling pathway	1.29E+01	5.274
Hepatic fibrosis signaling pathway	1.19E+01	5.225
TREM1 signaling	5.73E+00	5.209

Table 2. Five canonical pathways with the highest and lowest activation score (*Staphylococcus aureus* challenge).

Ingenuity canonical pathways	-log(*P*-value)	z-score
PPAR signaling	6.58E+00	-4.768
Sumoylation pathway	1.68E+00	-2.828
Inhibition of angiogenesis by TSP1	2.24E+00	-2.111
Dopamine-DARPP32 feedback in cAMP signaling	1.89E+00	-2.082
LXR/RXR Activation	1.01E+01	-1.897
Dendritic cell maturation	1.05E+01	5.842
Neuroinflammation signaling pathway	1.21E+01	5.511
TREM1 signaling	6.71E+00	5.209
Production of nitric oxide and reactive oxygen species in macrophages	9.35E+00	4.662
IL-6 signaling	9.49E+00	4.423

Discussion

The exceptionally strong response to an intramammary challenge with *E. coli* and *S. aureus* is consistent with the substantial systemic response of the heifers observed in the periphery at transcriptional level, as described in the liver transcriptome of these animals in response to both pathogens (Heimes *et al.*, 2020). In spite of substantial differences in the clinical outcome of the challenge (Rohmeier *et al.*, 2020), both local pathogen challenge responses of the mammary gland share several similar affected pathways, e.g. inactivation of the PPAR signaling and LXR/RXR Activation pathways, indicating an impaired lipid metabolism. In addition, selected activating and deactivating effects on the innate immune response were affected by both pathogens, though their modulation was stronger in response to *E. coli*.

Compared to previous studies (Petzl *et al.* 2008, Petzl *et al.* 2018), the number of DEGs in our approach is higher than in previous studies. This might be due to the early lactation stage, in which cows encounter a negative energy balance (Meyerholz *et al.* 2019) and are particularly vulnerable to infectious diseases. The advantages of our study design are a significantly higher number of animals studied and the tailored transcriptome analysis (depletion of milk and whey protein gene transcripts), which may also contribute to revealing previously undetected transcriptomic responses due to an increased sensitivity of the transcriptomic analysis. The intramammary gland challenge experiments with two divergent mastitis pathogens showed that both pathogens have activating and deactivating local effects on the expression of genes involved in the innate immune response directly in the mammary tissue. However, *E. coli* challenge caused a more pronounced modulation of the gene expression.

Acknowledgements

This work was supported by Funds of the German Government's Special Purpose Fund held at Landwirtschaftliche Rentenbank. We thank the Förderverein Bioökonomieforschung for continuous support.

References

Brodhagen J., Weikard R., Thom U., Heimes A., Gunther J. *et al.* (2019) BMC Genomics 20:400. https://doi.org/10.1186/s12864-019-5781-3

Halasa T., Huijps K., Osteras O., Hogeveen H. (2007) Vet Q 29:18–31. https://doi.org/10.1080/01652176.2007.9695224

Heimes A., Brodhagen J., Weikard R., Meyerholz M.M., Petzl W. *et al.* (2020) Front Immunol 11:715. https://doi.org/10.3389/fimmu.2020.00715

Jensen K., Gunther J., Talbot R., Petzl W., Zerbe H. *et al.* (2013) BMC Genomics 14:36. https://doi.org/10.1186/1471-2164-14-36

Li H., Handsaker B., Wysoker A., Fennell T., Ruan J. *et al.* (2009) Bioinformatics 25:2078–9. https://doi.org/10.1093/bioinformatics/btp352

Liao Y., Smyth G.K., Shi W. (2014) Bioinformatics 30:923–30. https://doi.org/10.1093/bioinformatics/btt656

Love M.I., Huber W., Anders S. (2014) Genome Biol 15:550. https://doi.org/10.1186/s13059-014-0550-8

Meyerholz M.M., Rohmeier L., Eickhoff T., Hülsebusch A., Jander S. *et al.* (2019) J Dairy Sci 102(11):9983-9994. https://doi.org/10.3168/jds.2018-16171

Petzl W., Gunther J., Pfister T., Sauter-Louis C., Goetze L. *et al.* (2012) Innate Immun 18:467–77. https://doi.org/10.1177/1753425911422407

Petzl W., Zerbe H., Gunther J., Yang W., Seyfert H.M. *et al.* (2008) Vet Res 39:18. https://doi.org/10.1051/vetres:2007057

Petzl W., Zerbe H., Gunther J., Yang W., Seyfert H.M., Hussen J. *et al.* (2018) Res Vet Sci 116:55–61. https://doi.org/10.1016/j.rvsc.2017.12.012.

Rohmeier, L.; Petzl, W.; Koy, M.; Eickhoff, T.; Hülsebusch, A. *et al.* (2020) BMC Vet Res 16:33. https://doi.org/10.1186/s12917-020-2251-8

Weikard R., Goldammer T., Brunner R.M., Kuehn C. (2012) Physiol Genomics 44:728–39. https://doi.org/10.1152/physiolgenomics.00007.2012

Pertea M., Kim D., Pertea G.M., Leek J.T., Salzberg S.L. (2016) Nat Protoc 11:1650–67. https://doi.org/10.1038/nprot.2016.095

676. Epistatic interaction between *KIT* and *MITF* causes splotchy face in Hereford × dairy crossbred calves

S. Jivanji[1*], A. Yeates[1], C. Harland[1], C. Gray[1], C. Couldrey[1], G. Worth[1], L. McNaughton[1], T. Fransen[1], T. Monehan[1], R. Spelman[1], D. Garrick[2] and M. Littlejohn[1,2]

[1]Livestock Improvement Corporation, Newstead, Hamilton, New Zealand; [2]School of Agriculture and Environment, Massey University, Palmerston North, New Zealand; swati.jivanji@lic.co.nz

Abstract

Hereford bulls are being used in dairy farming systems to sire calves that can be finished as beef cattle, bred to cows that would not otherwise be used for replacement cows. Their characteristically white faces offer simple markers for parentage, however we observed that some dairy × Hereford crossbred calves have splotchy faces. We conducted a genome-wide association analysis for face colour in 128 Hereford-cross calves using genotype data from the GMK50kv1 chip and found that the adulteration of the white face trait is caused by an epistatic interaction between mutations at the *KIT* and *MITF* loci. We found that the one copy of the G allele at rs209784468 was sufficient to rescue pigmentation in the face in Hereford-cross calves. These results present a novel gene × gene interaction and suggest that the white face trait is not technically dominant.

Introduction

Public demand for higher animal welfare standards is rising across the globe, and in particular, there have been rising concerns about bobby calves (Coleman 2020). Bobby calves are surplus calves that are not required by farmers to replace cows in their current herd and are slaughtered when they are less than 30 days old. Coat colour and patterning traits can be utilised to reduce the number of bobby calves produced annually by serving as an inexpensive, easily identifiable marker of beef parentage. Cows in the herd that will not be bred for replacement cows can be mated to a beef bull so that the progeny will grow and perform like a beef animal, thus giving an advantage over 'pure' dairy calves. For this reason, calves that are born from dairy dams and beef sires receive a higher premium when sold if they have visible markers that can prove beef parentage (Coleman 2020). Hereford bulls are often used for this purpose due to their characteristically white face, however some dairy × Hereford calves have splotchy faces. The splotchy face trait is undesirable to farmers, and reduces the value of these calves.

The aim of our study was to investigate the genetic architecture of the splotchy face trait in New Zealand Hereford crossbred cattle.

Materials & methods

Study population and genotype data. Tissue samples were obtained from ear tissue biopsies from 21 Angus × Hereford calves, and 107 Holstein-Friesian × Jersey × Hereford calves with either white, or splotchy faces. These calves were born on commercial farms in the Waikato region, New Zealand. DNA extraction was conducted at GeneMark (Hamilton, New Zealand), using the Qiagen BioSprint kit, and genotyping was carried out using the GeneMark 50kv1 (GMK50kv1) single nucleotide polymorphism (SNP) chip (manufactured by Illumina). All procedures involving animals were approved by the AgResearch Animal Ethics Committee (15174).

Phenotypes, population structure adjustments, and association analyses. Face colour (i.e. white or splotchy; Figure 1) was reported by farmers or scored from photographs taken by farmers. A 'leave one chromosome out' approach was used to construct 29 genomic relationship matrices (GRM), where each

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_676, © S. Jivanji *et al.* 2022

Figure 1. Image of a splotchy-faced calf (left) and a white-faced calf (right).

GRM lacked one autosome to avoid double fitting when testing the effect of candidate variants from that autosome. The GRM were calculated using variants from the GMK50kv1 platform used to genotype these cattle (n=21,159 variants), where variants had been filtered based on minor allele frequency (those with a minor allele frequency <0.02 were removed), deviation from Hardy Weinberg equilibrium (those with a $P<0.15$ were removed), missing genotype rates (those with a genotyping rate <0.01 were removed), and high linkage disequilibrium (LD) with another marker on the panel (those with pairwise $R^2>0.9$ were removed). The GCTA software (v1.93.2beta) (Yang *et al.* 2011) was then used to perform a mixed linear model-based association analysis (MLMA) using the GMK50kv1 resolution genotype data. For significance testing, we used a *P*-value of 5×10^{-8} as the significance threshold to account for multiple hypothesis testing.

Results

A strong association signal was detected on chromosome 22, with the top variant mapping to Chr22 g.31651404T>C, with a *P*-value of 2.41×10^{-20} (Figure 2). This variant mapped to intron 2 of the skin-expressed isoform of *MITF* (Koufariotis *et al.* 2015; Zerbino *et al.* 2018), a transcription factor that is involved in pigment cell (i.e. melanocyte) survival, maintenance, and differentiation (D'Mello *et al.* 2016). The third top associated variant (Chr22 g.31651379A>G; rs209784468) was near-equivalently associated ($P=5.97\times10^{-20}$), and was previously identified to be a candidate causal mutation in white spotting in Holstein-Friesian, Jersey, and crossbred cattle (Jivanji *et al.* 2019). When the rs209784468 SNP was fitted as a fixed effect in the association model, all significance at the chromosome 22 locus was removed, suggesting that the signal was likely representative of a single biallelic effect (smallest residual genome-wide $P=0.00088$ for Chr22 g.50885680C>G).

Candidate causative regulatory variant at the MITF locus. The rs209784468 variant maps to a region immediately upstream of the annotated *MITF* transcription start site (Zerbino *et al.* 2018), and intron 2 of the skin-expressed *MITF* gene isoform (Jivanji *et al.* 2019; Koufariotis *et al.* 2015; Zerbino *et al.* 2018). The variant appears to be highly conserved, mapping to a constrained element identified from 32-way amniote vertebrate alignments, and having a site-wise genome evolutionary rate profiling (GERP) score of 1.21 (Zerbino *et al.* 2018). The putative mutation appeared to operate in a Mendelian fashion, where nearly all calves carrying the 'G' allele manifested splotchy faces, and nearly all pure white-faced animals were homozygous for the 'A' variant (Figure 3). None of the Hereford crossed calves in our population were homozygous for the rs209784468 G allele.

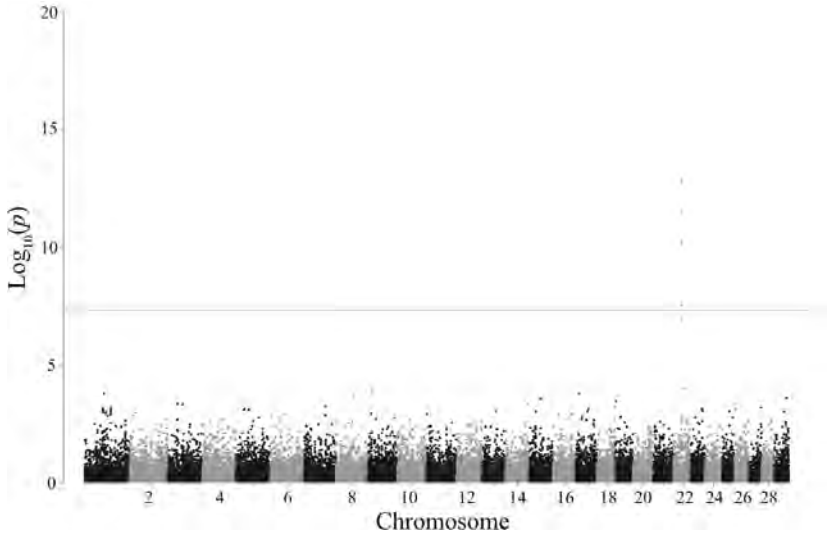


Figure 2. Manhattan plot based on the GWAS result for the splotchy-face trait. A single association signal is observed on chromosome 22, with the top variant mapping to Chr22 g.31651404T>C, $P=2.41\times10^{-20}$. The red line indicates the genome-wide significance threshold $P=5\times10^{-8}$.

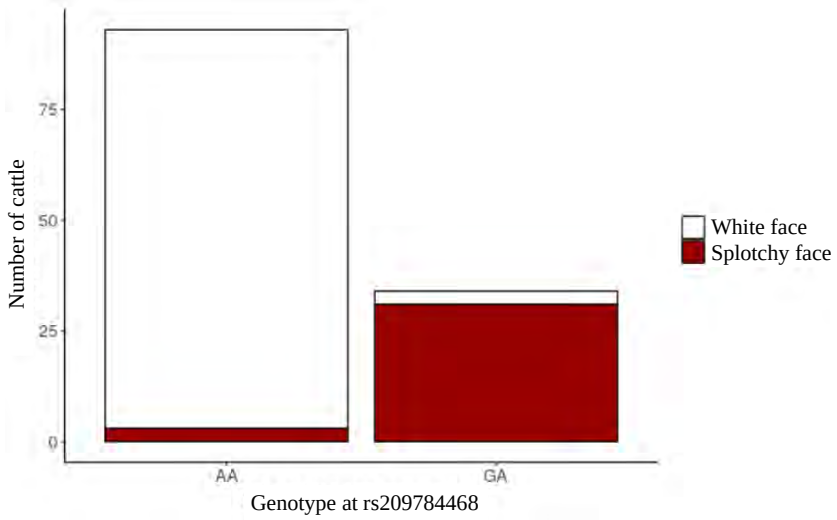


Figure 3. The number of cattle with white faces (white) or splotchy faces (red) plotted by their genotype at candidate causal mutation rs209784468 (Chr22 g.31651379A>G).

Discussion

We present a novel epistatic interaction between variants at the *KIT* and *MITF* loci that cause a splotchy-face phenotype in calves with Hereford parentage. The G allele at the *MITF* rs209784468 site appears to rescue pigmentation, presumably through interaction with the serial duplication upstream of *KIT* proposed to be causal for the white face trait characteristic of Hereford cattle (Whitacre 2014).

The Hereford white-face mutation has long been believed to be a dominant effect mutation, caused by a serial duplication upstream of the *KIT* gene. Inheritance of one copy of this mutation appears sufficient to cause the white-face trait (Whitacre 2014). This mutation is proposed to modulate *KIT* expression and result in a complete lack of pigmentation in the face. However, when Hereford cattle are crossed with some typically solid-coloured cattle (e.g. Angus or Jersey), or crosses thereof, their calves sometimes have a broken white-face, or 'splotchy' face. Inheritance of one copy of the rs209784468 G allele, mapping to intron 2 of the skin-expressed *MITF* gene isoform, appears to be sufficient to cause the splotchy-face trait in Hereford cross calves. The rs209784468 G allele has also been implicated in reduced white-spotting of the coat in Holstein-Friesian, Jersey, and crosses thereof (Jivanji *et al.* 2019). We hypothesise that the rs209784468 G allele has a positive effect on *MITF* expression and/or function, enabling survival, differentiation, and/or migration of functional melanocytes that are otherwise impacted by the white-face *KIT* mutation, and perhaps other white-spotting mutations. Our observation that inheritance of one copy of the rs209784468 G allele appears to be sufficient to cause a splotchy face also suggests that the Hereford white face trait is not technically dominant. Although we did not have any individuals in our study that had two copies of the rs209784468 G allele, future studies could aim to investigate how face colour may change if calves had different combinations of genotypes across the *KIT* and *MITF* loci.

The results presented here provide valuable insight into pigment biology and epistasis. Knowledge of how face colour is modulated could also enable breeders to produce dairy × Hereford claves with pure white faces for beef rearing, thus enhancing animal welfare by minimising the number of calves that need to go on the bobby truck per season.

References

Coleman LW. (2020). The Use of High Genetic Merit Angus and Hereford Bulls in a New Zealand Dairy Herd. Massey University, Palmerston North, New Zealand.

D'Mello SA, Finlay GJ, Baguley BC, and Askarian-Amiri ME. (2016) Int J Mol Sci 17(7):1144. https://doi-org.ezproxy.massey.ac.nz/10.3390/ijms17071144

Jivanji S, Worth G, Lopdell TJ, Yeates A, Couldrey C *et al.* (2019) Genet Sel Evol 51(1):1–18. https://doi-org.ezproxy.massey.ac.nz/10.1186/s12711-019-0506-2

Koufariotis LT, Chen YP, Chamberlain A, Vander Jagt C, and Hayes BJ. (2015) PLoS ONE 10(10):1–24. https://doi-org.ezproxy.massey.ac.nz/10.1371/journal.pone.0141225

Whitacre, L. (2014). Structural Variation at the KIT Locus Is Responsible for the Piebald Phenotype in Hereford and Simmental Cattle. University of Missouri, Columbia, USA.

Yang J, Lee SH, Goddard ME, and Visscher PM (2011) Am J Hum Genet 88(1):76–82. https://doi.org/10.1016/j.ajhg.2010.11.011

Zerbino DR, Achuthan P, Akanni W, Amode MR, Barrell D et al (2018) Nucleic Acids Res 46(D1): D754–61. https://doi-org.ezproxy.massey.ac.nz/10.1093/nar/gkx1098

677. Searching for genes involved in heat stress response in lactating cows through transcriptomics

C. Diaz[1], M. Ramón[2], C. Gonzalez-Verdejo[1] and M.J. Carabaño[1*]

[1]INIA-CSIC, P.O. Box 28040 Madrid, Spain; [2]Instituto Regional de Investigación y Desarrollo Agroalimentario y Forestal de Castilla-La Mancha (IRIAF), Av. del Vino s/n, 13300 Valdepeñas, Spain; mjc@inia.es

Abstract

Understanding the genetic mechanisms that regulate the response of animals to heat stress (HS) events is of interest when developing selection strategies towards more thermotolerant animals. In this study, we have addressed, in dairy cattle, the study of genes that are differentially expressed (DE) under heat stress compared to a comfort situation. A total of 407 genes DE between heat stress and comfort conditions were found, many of them being upregulated (230) under HS conditions. The study of the biological processes to which these genes are associated has shown processes associated with response to external aggressors, immune response or cellular repair.

Introduction

Heat stress (HS) is one of the most significant sources of distress for animals and of economic loss for famers. In this aspect, dairy cattle farming is one of the livestock sectors most severely affected by HS, since highly productive cows are under a high metabolic burden. Unveiling the genetic basis of heat stress response should help in understanding the complex cellular, metabolic and physiological mechanisms triggered by HS and in providing farmers with tools for selection of heat tolerant animals. However, although a number of studies have dealt with the goal of finding genes and/or genetic markers, only a few genes, such as the heat shock protein (HSP) and heat shock factor (HSF) genes (see, e.g. review by Collier *et al.*, 2008) and the slick hair gene (Dikmen *et al.*, 2014) are widely accepted as associated with HS response and HS tolerance, respectively. Genome-wide information both from DNA or RNA sequencing is expected to provide useful and complementary results when searching for the genetic basis of complex traits. In this study RNA-seq was used to investigate the genetic background of HS response in lactating Holstein cows.

Materials & methods

This study was carried out under a Project License from the INIA Scientific Ethic Committee. Animal manipulations were performed according to the Spanish Policy for Animal Protection RD 53/2013, which meets the European Union Directive 86/609 on the protection of animals used in experimentation. The INIA Scientific Ethic Committee, which is named IACUC, specifically approved this study.

Data and experimental design. Nineteen multiparous lactating cows were blood sampled twice in HS (on July1 and July 7) and once in Comfort (October 10). Cows were all from the same farm and had the same management all along the experiment. All cows were multiparous and started the experiment on the first third of their lactation. Cows participating in the study did not have common parents in order to allow for a wider genetic variability. On each blood collection date, respiration rate (rpm) and rectal temperature (°C) were measured individually. Milk yields from the official milk recording of the farm were also available. Weather data (daily maximum, average and minimum temperature) from one week before the beginning of the study until the last sampling were provided by the National Meteorological Agency (AEMET).

RNA extraction and samples preparation. RNA was isolated from blood samples using RiboPure™ -Blood Kit protocol. After, performing RNA quality controls, contamination removal was carried out in 56

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_677, © C. Diaz *et al.* 2022

samples with Agilent RNA 6000 Nano Kit (Agilent Technologies, Germany) and RQ1 RNase-Free DNase (Promega Corporation, USA). RNA-seq libraries were constructed with the SMARTer® Stranded Total RNA-Seq Kit for High-input strand specific total RNAseq for Mammalian samples (Takara Bio USA, Inc.) according to the manufacturer's instructions, which is compatible with input ranges from 100 ng to 1 µg of total RNA. Sequencing of stranded mRNA was performed in a NovaSeq 6000 SP Illumina platform to obtain 50bp long paired-end reads.

Statistical analyses. Quality control of the raw and trimmed paired-end reads were performed with FastQC v.0.11.7 (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/). Trimming was carried out with Trimmomatic v.0.38 (Bolger *et al.* 2014) to remove low quality reads and adaptors. Trimmed reads were mapped to the Bovine reference genome (Bos taurus ARS-UCD1.104) with BWA (Li and Durbin, 2009) and mapped transcripts were assembled individually with Stringtie v.1.3.3b (Pertea *et al.* 2015). Differential abundance (DA) analysis was carried out using DESeq2 (Love *et al.*, 2013). Genes were considered to be DA when the significance level (adjusted *P*-value for multiple comparisons) was lower than 0.05, after the Benjamini-Hochberg correction and the absolute value of the Log2FoldChange was higher or equal to 2. Annotation of the DE genes between the two climate conditions was retrieved with BiomaRt (https://www.ensembl.org/biomart). Functional Enrrichment Analysis and Gene annotation for Biological process was performed via PANTHER (Mi *et al.*, 2020) based on Biological Proceses (BP) and pathway analysis.

Results

Average daily milk yield and fat and protein contents of cows were 51.5 kg; 3.34%; 3.07% (sd=11.1; 0.7; 0.3). Daily maximum/average/minimum temperatures on the three collection dates were 35.2/26.1/17.0 °C, 34.9/28.8/22.8 and 18.3/13.0/7.6 °C. Figure 1 represents the distribution of rectal temperature and respiration rates physiological measures of cows for the HS and comfort samplings, showing clear differences between periods for both variables. Basal levels of both variables as well as the dispersion changed from Comfort to HS. Some cows showed only small changes between periods, pointing at a wide variability in physiological response to heat stress across animals.

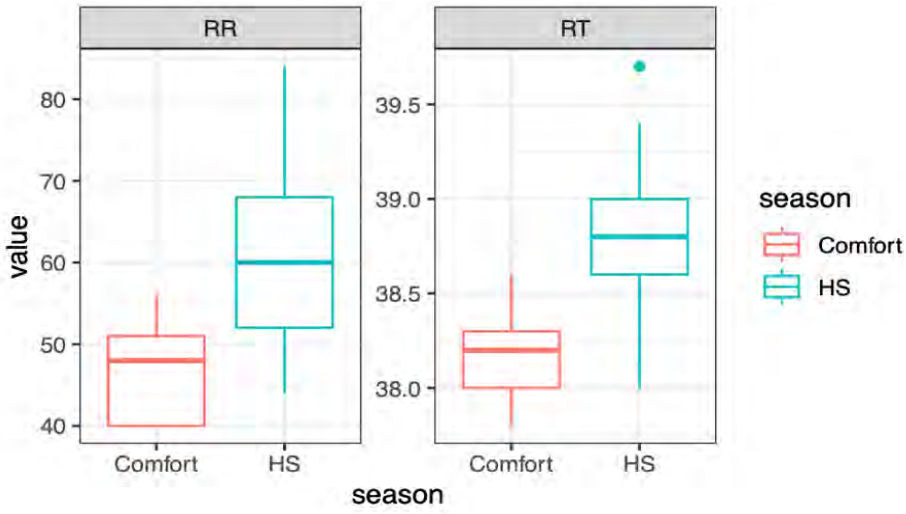


Figure 1. Differences on respiration rate (RR, rpm) and rectal temperature (RT, °C) between heat stress and comfort conditions.

A total of 407 genes DE between heat stress and comfort conditions in blood samples of dairy cows were found. From those DE genes, 230 genes were upregulated under HS conditions and 177 were downregulated. Among all DE genes, we identified 80 Non-coding RNAs in several forms (Small nucleolar RNAs (3), ribosomal RNAs (9), spliceosomal RNAs (43), Y-RNAs (3), 7SK_RNA (7), Metazoa_SRC signal recognition particles (7) and miRNAs (8) that were mostly (70 out of 80) upregulated under HS conditions. Those elements are of potential great interest due to their regulatory role in the mammalian response to heat shocks (Place and Noonan, 2014). Moreover, some miRNAs are being proposed as biomarkers of HS (Liu *et al*., 2020). For this first analysis, those non-coding elements were excluded for the Functional Enrichment Analysis and Gene Annotation. Thus, a total list of 327 genes were used, 160 and 167 genes upregulated and downregulated in HS, respectively.

For the analysis of the main Biological Process (BP) represented in gene ontologies (GO) only 308 genes were uniquely mapped. Figure 2 presents the main BP that were found as well as the number of genes classified into each of them. According to the GO, DE genes were classified in 18 different BPs. The BPs more represented by PANTHER GO-Slim classification involved cellular process (GO:0009987) that included cellular stress, biological regulation (GO:0065007), metabolic process (GO:0008152, immune-system process (GO:0002376), response to stimulus (GO:0050896), signaling (GO:0023052). We also searched BP significantly associated with the gene lists based on their FDR (P<0.05) using as a reference list GOdatabase (http://geneontology.org). Most of the biological process are conserved between lists but some other appealing groups appear such as Nervous system process, sensory receptors as well as regulation groups (Figure. 2).

Discussion

Livestock production systems, and in particular dairy cattle are facing the impact of climate change. Breeding for thermotolerant animals is a strategy that will help to the dairy sector to cope with this challenge. In order to establish a selection strategy, it is of great interest to understand the physiological mechanisms underlying HS since so far selection criteria are based on slopes relating production traits with THI (Carabaño *et al.*, 2020). With this purpose, we have performed a first analysis of the changes in the transcriptome profiles of dairy cows between comfort and HS conditions. In this analysis, we have not

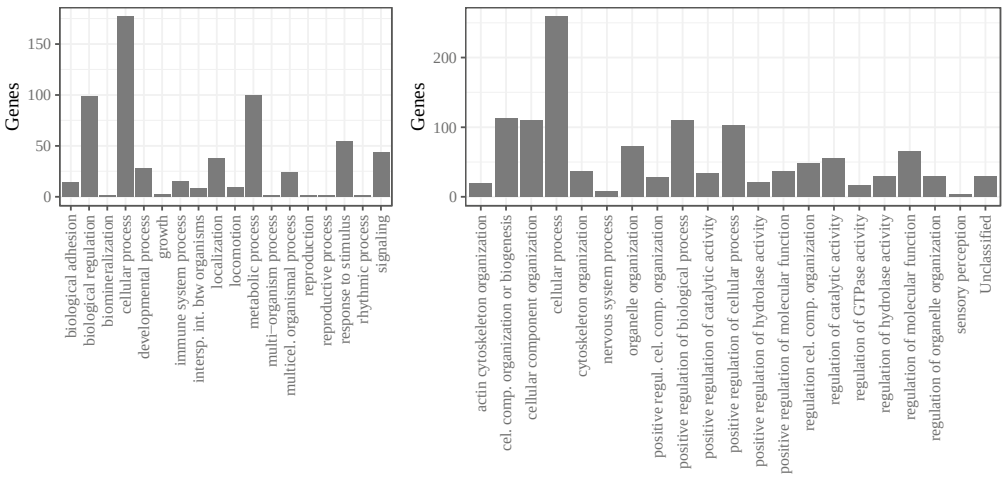


Figure 2. GO classification of genes into different biological processes based on GO-Slim (PANTHER; left) the GO database (right).

distinguished between short and long exposure to HS or what is to say acute versus chronic HS provided that thermal loads, RR and RT were similar between sampling dates under HS. The results showed how under heat stress conditions some genes involved in important biological processes such as response to external aggressors, immune response or cellular repair were up-regulated providing thermotolerant animals with an advantage in responding to the effects of heat stress. Biological processes identified were very general and involved interleukins, cytokines or chaperones among others, clearly associated in the response to stress events, and specifically, to heat stress.

Acknowledgements

The authors acknowledge the financial support of the Spanish Ministry of Science and Innovation national funded project RTA2015-00035 (Madrid, Spain). They also thanks to COVAP (Cooperative of Valle de los Pedroches) for its contribution obtaining the biological samples.

References

Bolger A.M., Lohse M., Usadel B. 2014;30(15):2114-20. https://doi.org/10.1093/bioinformatics/btu170

Carabaño M.J, Ramón M., Menendez-Buxadera A., Molina A., and Díaz C. (2019) Animal Frontiers. Vol9 (1). https://doi.org/10.1093/af/vfy033

Collier R.J., Collier J.L., Rhoads R.P., and Baumgard L.H. (2008). J. Dairy Sci. 91:445–454. https://doi.org/10.3168/jds.2007-0540.

Dikmen S., Khan F.A., Huson H.J., Sonstegard T.S, Moss J.I., Dahl G.E. and Hansen P.J. (2014). J. Dairy Sci. 97:5508–5520. https://doi.org/10.3168/jds.2014-8087.

Li H. and Durbin R. (2009) Transform. Bioinformatics, 25:1754-60. https://doi.org/10.1093/bioinformatics/btp324

Liu G., Liao Y., Sun B., Guo Y., Deng M., *et al.* (2020) Gene 742:144550. doi.org/10.1016/j.gene.2020.144550

Love MI, Huber W, Anders S (2014). *Genome Biology*, 15, 550. https://doi.org/10.1186/s13059-014-0550-8.

Pertea M., Pertea G.M., Antonescu C.M. Chang T.C, Mendell J.T., Salzberg S.L. Nat Biotechnol. 2015;33(3):290-5. https://doi.org/10.1038/nbt.3122

Place R.F and Nooman E.J. (2014). Cell Stress and Chaperones 19(2):159-172. https://doi.org/10.1007/s12192-013-0456-5

678. Genomic regions exhibiting transmission distortions and copy number variants associated with stillbirth in cattle

A.A.A. Martin[1*], H.R. Oliveira[1,2], P.A.S. Fonseca[1], C.M. Rochus[1], F. Miglior[1,2], S. Id-Lahoucine[3], A. Cánovas[1], C.F. Baes[1,4] and F.S. Schenkel[1]

[1]Centre of Genetic Improvement of Livestock, University of Guelph, 50 Stone Rd E, N1G 2W1 Guelph, ON, Canada; [2]Lactanet, 660 Speedvale Ave W, N1K 1E5 Guelph, ON, Canada; [3]Department of Animal and Veterinary Sciences, Scotland's Rural College, Easter Bush, Midlothian EH25 9RG, United Kingdom; [4]Institute of Genetics, Vetsuisse Faculty, Länggassstrasse 120, 3012 Bern, Switzerland; amarti52@uoguelph.ca

Abstract

The incidence of stillbirth has increased in the past decades, negatively affecting the welfare and the economic profitability of dairy herds. Recently, high confidence copy number variants regions were found to be associated with stillbirth traits in Canadian Holstein cattle. Concomitantly, genomic regions showing transmission distortion associated with reproductive defects were identified in an independent study performed in the same population. The functional profile of the overlapped regions between those studies was investigated. Four genes (ENSBTAG00000047181, ENSBTAG00000023309, ENSBTAG00000046041, and *PIK3CD*) were shared between the three enriched pathways ('Antifolate resistance', 'ABC transporters' and 'cAMP signalling') hinting at abnormal uterine inflammation in late pregnancy, which could affect the viability of the placenta and/or foetus. With further studies validating these findings, breeding against stillbirth could potentially be improved by using CNV information.

Introduction

Stillbirth (SB) is defined by calf death after 260 days of gestation until 48 hours after parturition (Berglund *et al.*, 2003). The number of SB cases is a cause of significant economic and welfare losses for the dairy industry (Fonseca *et al.*, 2022). Causes of stillbirth are not well documented and are rarely assessed. For instance, about 50% of the stillborn calves can be attributed to dystocia or malformations, but more than 30% of the recorded stillbirths cannot be explained (Berglund *et al.*, 2003). A better understanding of the genetic background of this trait can help to explain some of the unknown causes of SB recorded on farms and decrease its incidence through genetic selection. Multiple approaches can be used to investigate and understand the biological mechanisms underlying SB. Recently, high-confidence copy number variants (CNV) regions were found to be associated with SB in Canadian Holstein cattle (Oliveira *et al.*, 2021). These CNV regions were found to overlap with genomic regions of a study investigating the mating success linked to the gametic compatibility of mated pairs (Martin *et al.*, 2021). This overlap suggests that CNV may be responsible for mechanisms of transmission ratio distortions (TRD) affecting stillbirth. Therefore, the functional profile of the overlapped regions was investigated in this study.

Materials & methods

Previously identified CNV and TRD regions. The CNV regions previously found associated with SB in Oliveira *et al.* (2021) were used for this study. In summary, the association analyses were performed using four alternate deregressed breeding values (heifers, cows, and bulls when mated to either heifers or cows) and high confidence CNV regions identified in 5,211 animals genotyped using a 95K SNP panel. The TRD regions were identified in an independent study performed using the same population (i.e. Canadian Holstein cattle; Martin *et al.*, 2021). Regions associated with gametic incompatibility were identified using a TRD model, which measures deviations from the Mendelian inheritance, i.e. one allele is preferentially transmitted to the next generation. An interaction effect related to the offspring genotype, called genotypic interaction, was added to the classic TRD model containing direct TRD effects (Id-Lahoucine *et al.*, 2019),

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_678, © A.A.A. Martin *et al.* 2022

to scan the genome for TRD signals potentially associated with gametic incompatibility, using different sizes of haplotypes. A set of 283,817 parent-offspring trios genotyped using 50K SNP genotypes was used in this analysis, and a total of 474 TRD regions were identified as potentially displaying gametic incompatibility.

Functional profile of the overlapped regions. Only overlapping CNV and TRD genomic regions were used for further functional analysis. The positional genes and quantitative trait loci (QTL) located in the overlapped regions were identified using the GALLO package (v1.3; Fonseca *et al.*, 2020) implemented in R (v4.1.1; R Core Team, 2019). Thereafter, a functional analysis was performed on the retrieved gene list using the AmiGO 2 software (v2.5.17; Carbon *et al.*, 2009). Additionally, a protein-protein interaction network was created using the String database (v11.0; Szklarczyk *et al.*, 2019), which was based on functional and physical protein associations. Furthermore, an over-representation analysis (ORA) was performed to determine if any Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways or Gene Ontology (GO) terms were more abundantly present in the investigated gene network compared to what would be expected from a random gene sample.

Results

Out of the 11 unique CNV regions identified by Oliveira *et al.* (2021) for the four SB trait EBVs, ten overlapped with at least one TRD region (Table 1). Interestingly, all TRD regions that overlapped with significant deletions (DEL) showed a strong genotypic interaction effect and moderate direct effect, indicating a strongly skewed allelic transmission. Only one CNV region containing duplication (DUP) was found to overlap with TRD regions. For the single DUP CNV region, the corresponding TRD region had a close-to-zero magnitude for both genotypic interaction and direct effects, showing a haplotype transmission close to the expected Mendelian proportions. A total of 67 positional genes were found, including 56 genes with NCBI annotation. A total of 572 unique QTL were identified, from which 9% were linked to reproduction. There were 44 QTL associated with calving ease and one with SB. The functional analysis of the whole gene list resulted in one enriched GO-Molecular Function term: 'ATPase-coupled transmembrane transporter activity' (*P*-value=0.02). There were four significant KEGG pathways (Table 2) but no enriched GO

Table 1. Overlapping of copy number variant (CNV) regions associated with stillbirth[1] and transmission ratio distortion (TRD) regions associated with gametic incompatibility[2].

CNV state[3]	BTA[4]	Start (bp)	End (bp)	Direct TRD mean (±SE)[5]	Genotypic interaction TRD mean (±SE)[5]
DEL	6	113,852,273	114,116,280	0.20 (0.05)	0.82 (0.07)
DUP	12	69,938,893	72,984,925	0.01 (0.01)	0.06 (0.04)
DEL	14	398,877	465,742	0.40 (0.27)	0.95 (0.01)
DEL	14	11,763,630	12,102,663	0.22 (0.10)	0.65 (0.07)
	14	14,168,275	14,196,101	0.10 (0.04)	0.76 (0.03)
DEL	16	43,676,591	44,527,833	0.17 (0.07)	0.61 (0.10)
	16	51,194,720	51,444,577	0.14 (0.04)	0.31 (0.08)
DEL	20	71,053,478	71,102,653	0.28 (0.14)	0.60 (0.29)
DEL	25	1,130,037	1,163,011	0.26 (0.12)	0.70 (0.22)
DEL	25	41,785,100	41,806,113	0.26 (0.11)	0.65 (0.20)
DEL	26	50,127,733	50,614,026	0.28 (0.05)	0.79 (0.05)

[1] CNV regions were reported in Oliveira *et al.* (2021).

[2] TRD regions were reported in Martin *et al.* (2021).

[3] CNV State: single deletion (DEL) and duplication (DUP).

[4] BTA: *Bos taurus* autosome.

[5] Mean (±standard error) of the absolute TRD magnitude values for the different haplotype combinations within the genomic region.

Table 2. Overrepresented KEGG[1] pathways for the positional candidate genes identified in this study[2].

KEGG pathways	Hits[3]	Total[4]	LogRatio[5]	FDR[6]
Antifolate resistance	4	38	1.62	0.002
ABC transporters	4	51	1.49	0.003
Bile secretion	4	84	1.27	0.011
cAMP signalling pathway	5	207	0.98	0.023

[1] KEGG: Kyoto Encyclopedia of Genes and Genome.

[2] Positional candidate genes found overlapping between copy number variant regions reported in Oliveira *et al.* (2021) and transmission ratio distortion regions reported in Martin *et al.* (2021).

[3] Hits: number of genes within pathway from our gene list.

[4] Total: number of genes within pathway from the whole genome.

[5] LogRatio: Log10(observed hits/expected hits) genes within pathway. [6]FDR: false discovery rate.

terms. The same four proteins coded by three genes (ENSBTAG00000047181, ENSBTAG00000023309, ENSBTAG00000046041), were shared between the four pathways, with the addition of *PIK3CD* for the cAMP signalling pathway.

Discussion

Both ATP-binding cassette (ABC) transporters and cAMP signalling pathways are essential to many cellular processes. ABC transporters family is one of the largest protein families that is conserved across vertebrates (Dean and Annilo, 2005), and cAMP is one of the most common and universal second messengers, internalizing external signals to the cell (Serezani *et al.*, 2008). Interestingly, there are several examples of recent duplications and deletions in the evolution of ABC transporters (Dean and Annilo, 2005). Remarkably, nine out of eleven CNV regions associated with SB were deletions. The sole overlapping DUP CNV region was associated with a low TRD magnitude, whereas the DEL CNV regions were associated with strong transmission distortion. DEL CNV could be linked to a loss of function or a decreased level of gene expression that would increase reproductive defects and potentially cause SB. The common four proteins shared between the four KEGG pathways were part of the ABC subfamily C member 4, also called multidrug resistance-associated protein, which are yet to be annotated in the *Bos taurus* genome but thoroughly studied in other mammals. The ABC transporters can be linked to a form of antifolate resistance as antifolate efflux transporters (Assaraf, 2007). Antifolates are the natural down regulators of DNA replication, controlling the proliferative activity of cells (Chabner *et al.*, 2006). In early pregnancy, the supplementation of folate reduces the chance of stillbirth (Imdad *et al.*, 2011), whereas excessive antifolates can induce abortion by blocking normal embryo development (Chabner *et al.*, 2006). The overexpression of ABC transporters could provoke an increased antifolate efflux, a known mechanism of antifolate resistance (Assaraf, 2007). The three genes, coding for parts of the ABC transporters, were positioned in the DUP CNV region, which could lead to an overexpression and induce resistance.

Fatal preterm birth can be associated with intrauterine inflammation, which affects gene expression in the placenta. Pathways including *PIK3CD* are altered in placenta exposed to pathogens markers (Lien *et al.*, 2020). Together, these results hinted at two potential hypotheses: (1) animals carrying the CNV could have a lower tolerance to infection leading to stillbirth; or (2) overly exported antifolates could trigger unexpected inflammation. For both, pathogen protection and parturition process, inflammation is essential but needs to be strictly controlled to avoid adverse effects on the foetus and/or the mother (Christiaens *et al.*, 2008). Any abnormal late intrauterine inflammation would be strongly selected against, shaping the genetic structure of the population. The four candidate genes (ENSBTAG00000047181, ENSBTAG00000023309,

ENSBTAG00000046041, and *PIK3CD)* identified in this study seem to play an important role in stillbirth in Holstein cattle. Breeding to reduce the frequency of stillbirth could likely be improved by using CNV information, however, further studies are required to validate our findings.

Acknowledgements

This research was financially supported by Agriculture and Agri-Food Canada, and by additional contributions from Dairy Farmers of Canada, Lactanet and the Canadian Dairy Commission under the Agri-Science Clusters Initiative. The authors also acknowledge the support from the Resilient Dairy Genome Project as administered through Genome Canada (Canada) and Genome Alberta (Canada).

References

Assaraf Y.G. (2007) Cancer Metastasis Rev. 26(1):153–181. https://doi.org/10.1007/s10555-007-9049-z

Berglund B., Steinbock L., and Elvander M. (2003) Acta Vet. Scand. 44(3-4):111–120. https://doi.org/10.1186/1751-0147-44-111

Carbon S., *et al.* (2009) Bioinformatics 25(2):288–289. https://doi.org/10.1093/bioinformatics/btn615

Chabner B.A., *et al.* (2006) Chemotherapy of Neoplastic Diseases. In: Brunton L.L., Lazo J.S., Parker K.L. (eds) Goodman & Gilman's the Pharmacological Basis of Therapeutics. McGraw-Hill, New York, USA, pp. 1341-1430

Christiaens I., *et al.* (2008) J. Reprod. Immunol. 79(1):50–57. https://doi.org/10.1016/j.jri.2008.04.002

Dean M. and Annilo T. (2005) Annu. Rev. Genomics Hum. Genet. 6(1):123–142. https://doi.org/10.1146/annurev.genom.6.080604.162122

Fonseca P., Schenkel F., and Cánovas A. (2022) J. Dairy Sci. 105(2):1–13. https://doi.org/10.3168/jds.2021-20936

Fonseca P.A.S., Suárez-Vega A., Marras G., and Cánovas Á. (2020) GigaScience 9(12):1–9. https://doi.org/10.1093/gigascience/giaa149

Id-Lahoucine S., *et al.* (2019) J. Dairy Sci. 102(4):3175–3188. https://doi.org/10.3168/jds.2018-15296

Imdad A., Yakoob M.Y., and Bhutta Z.A. (2011) BMC Public Health 11(S3):S4. https://doi.org/10.1186/1471-2458-11-S3-S4

Lien Y.C., *et al.* (2020) Front. Physiol. 11(11):1–18. https://doi.org/10.3389/fphys.2020.592689

Martin A., *et al.* (2021) Proc. of the 2021 ASAS-CSAS-SSASAS Annual Meeting and Trade show, Louisville, USA

Oliveira H.R., *et al.* (2021) Proc. of the XIV Simpósio Brasileiro de Melhoramento Animal, Santa Catarina, Brasil

R Core Team, R: A language and environment for statistical computing. Available at: https://www.r-project.org

Serezani C.H., Ballinger M.N., Aronoff D.M., and Peters-Golden M. (2008) Am. J. Respir. Cell Mol. Biol. 39(2):127–132. https://doi.org/10.1165/rcmb.2008-0091TR

Szklarczyk D., *et al.* (2019) Nucleic Acids Res. 47(D1):607–613. https://doi.org/10.1093/nar/gky1131

679. Mating allocations in Nordic Red Dairy Cattle using genomic information

C. Bengtsson[1,2*], H. Stålhammar[1], J.R. Thomasen[1], S. Eriksson[2], W.F. Fikse[3] and E. Strandberg[2]

[1]VikingGenetics Sweden AB, 53294 Skara, Sweden; [2]Dept. of Animal Breeding and Genetics, Swedish University of Agricultural Sciences, Box 7023, 75007 Uppsala, Sweden; [3]Växa Sverige, Box 288, 75105 Uppsala, Sweden; chben@vikinggenetics.com

Abstract

In this study, we compared mating allocations in Nordic Red Dairy Cattle using genomic or pedigree information. We used linear programming to optimise different economic scores within each herd, considering genetic level, semen cost, the economic impact of recessive genetic defects, and genetic relationships. We found high correlations (≥0.83) between the pedigree and genomic relationship measures. The mating results showed that it was possible to reduce the different genetic relationships between parents with minimal effect on genetic level. Including the cost of known recessive genetic defects eliminated expression of genetic defects. It was possible to reduce genomic relationships between parents with pedigree information, but it was best done with genomic measures. Linear programming maximised the economic score for all herds studied within seconds, which means that it is suitable for implementation in mating software to be used by advisors and farmers.

Introduction

Mating programs are an important support tool for livestock breeders, helping them to identify the best parental matings to maximise genetic level and avoid mating between closely related individuals, preventing excessive inbreeding (Carthy *et al.*, 2019; Bérodier *et al.*, 2021). New genetic insights at single nucleotide level can be used in mating programs. Single nucleotide polymorphism (SNP) markers can give information about major genes and genetic defects. Minimising the probability of obtaining offspring homozygous for a lethal recessive genetic defect is of economic importance for farmers (Pryce *et al.*, 2012). SNP markers also offer the possibility to reduce genomic relationships between parents when optimising mating plans. Various methods have been proposed for calculation of genomic relationships, including SNP-by-SNP relationships as described by e.g. VanRaden (2008). Further, there are methods using shared genomic segments, as described by e.g. Cara *et al.* (2013), which aim to reduce the number of runs of homozygosity in the offspring.

The objective of this study was to investigate the ability of different mating scenarios to maximise expected genetic level in the next generation, limiting parent relationship and the probability of expression of genetic defects. We investigated all scenarios at herd level using real data. We used linear programming to optimise economic scores within each herd, considering genetic level, semen cost, the economic impact of recessive genetic defects, and five different measures of relationships (two pedigree-based and three genomic-based).

Materials & methods

Animals. Breeding values, pedigree data, SNP data, and data on the carrier status of genetic defects were obtained from the Nordic Cattle Genetic Evaluation. Genotype information for a total of six genetic defects were available. We selected 9,841 genotyped Red Dairy Cattle (RDC) females born in Denmark, Finland, or Sweden in 2019 to simulate matings. All females included belonged to herds with 20 or more genotyped females. The females were potentially mated to 50 genotyped RDC bulls from the Nordic breeding cooperative VikingGenetics.

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_679, © C. Bengtsson *et al.* 2022

Relationship measures. We used two different pedigree relationship coefficients. The first relationship coefficient traced the pedigree three generations back from the parents of the potential mating (a_{3Gen}). The second pedigree relationship coefficient was based on all available pedigree information (a_{AllGen}). We used three different genomic relationship coefficients, one SNP-by-SNP genomic relationship coefficient (g_{SNP}) according to VanRaden (2008). The two segment-based genomic relationship coefficients were based on different minimum lengths of segments: 1 centimorgan (cM) (g_{SEG1}) and 4 cM (g_{SEG4}).

Economic score. For each potential mating between female i and bull j, we calculated an economic score similarly to Bérodier *et al.* (2021) and Pryce *et al.* (2012):

$$Score_{ij} = \left(\frac{NTM_i + NTM_j}{2} + \lambda F_{ij}\right) \times \text{prob}\,(♀) - \sum_{r=1}^{n_r} p(aa)_r \times v_r - \text{semen cost}$$

where NTM_i and NTM_j are the value of Nordic Total Merit units in Euros (€) for female *i* and bull *j*, λ is the economic consequence of an 1% increase in inbreeding similarly to Cole (2015), F_{ij} is the pedigree or genomic based co-ancestry (Relationship/2), prob(♀) is the probability of producing a female conceptus, n_r is the number of recessive genetic defects considered, $p(aa)_r$ is the probability of expression of a genetic defect *r*, v_r is the economic cost associated with the recessive genetic defect *r* similarly to Cole (2015) and Bérodier *et al.* (2021), and semen cost is the average amount (€) spent on semen for a pregnancy. In addition, an economic score based on only NTM and semen cost was calculated (MaxNTM). Mate allocation was programmed in R version 3.6.3 (R Core Team, 2020). Linear programming optimisation was performed with the 'Lp_solve' package in R (Berkelaar *et al.*, 2020). The objective of linear programming was to maximise the average economic score in the herd with a maximum of 5% of females per bull and herd. The mating R script was provided by Bérodier *et al.* (2021).

Results

The correlations between the genetic relationship coefficients were at least 0.83 or higher (Table 1). The mean value of the relationship coefficients between all possible combinations of females and males (Random) ranged from 0.009 to 0.188 (Table 2). The strongest correlation was between a_{AllGen} and a_{3Gen} (r=0.99), and the second strongest was between g_{SEG1} and g_{SEG4} (r=0.98). The strongest correlation between pedigree and genomic relationships was between a_{AllGen} and g_{SEG4} (r=0.88) (Table 1). The coefficients of regression of a_{3Gen} and g_{SNP} on a_{AllGen} were close to 1, and the coefficients of regression of g_{SEG1} and g_{SEG4} on a_{AllGen} were both 0.94.

It was possible to maximise economic score with limited impact on the average NTM level (Table 2). Including the cost of the known recessive genetic defects when optimising mating strategies eliminated the risk of loss from a genetic defect, regardless of which genetic relationship was used. When just maximising

Table 1. Correlations between the different relationship coefficients for all possible combinations of 9,841 females and 50 bulls.[1]

Relationship	a_{AllGen}	g_{SNP}	g_{SEG1}	g_{SEG4}
a_{3Gen}	0.99	0.88	0.83	0.87
a_{AllGen}		0.88	0.85	0.88
g_{SNP}			0.9	0.93
g_{SEG1}				0.98

[1] a_{3Gen} = pedigree relationships using three generations of ancestors; a_{AllGen} = pedigree relationships using all available pedigree information; g_{SNP} = genomic relationship calculated according to VanRaden (2008); g_{SEG1} (g_{SEG4}) = genomic segment-based relationship according to de Cara *et al.* (2013) with a minimum segment length of 1 (4) centimorgan.

Table 2. Comparison of outcomes of planned matings of 9,841 females for 7 mating scenarios in Nordic Red Dairy Cattle using various comparison criteria[1].

Comparison criterion[1]	Random	5% females/bull scenarios[2]					
		Max NTM	3Gen	AllGen	GSNP	GSEG1	GSEG4
NTM	19.5	20.8	20.8	20.8	20.8	20.7	20.8
a_{3Gen}	0.028	0.033	0.007	0.009	0.014	0.014	0.014
a_{AllGen}	0.066	0.070	0.046	0.043	0.050	0.050	0.050
g_{SNP}	0.009	0.014	-0.012	-0.016	-0.038	-0.034	-0.033
g_{SEG1}	0.188	0.191	0.167	0.163	0.148	0.143	0.146
g_{SEG4}	0.115	0.119	0.094	0.091	0.078	0.075	0.074
Expression of genetic defects (%)	0.4	0.4	0	0	0	0	0

[1] Average Nordic total merit (NTM) level, average five different genetic relationships between parents, the probability of expression of genetic defects.

[2] Maximum percent of females per bull and herd set to 5%. Random: All possible combinations of 9,841 females and 50 bulls. MaxNTM: mates were selected based on maximising an economic score including NTM, sexed semen, and semen cost.

[3] Gen, AllGen, GSNP, GSEG1, GSEG4: mates were selected based on maximising an economic score including NTM, sexed semen, semen cost, a penalty for genetic defects and a pedigree relationship include three generations (a_{3Gen}) or all available ancestors (a_{AllGen}), or a genomic relationship calculated according to VanRaden (2008) (g_{SNP}), according to Cara *et al.* (2013) with a minimum genomic segment length of 1 centimorgan (g_{SEG1}) or 4 centimorgan (g_{SEG4}).

NTM minus semen cost (MaxNTM), the NTM level improved compared with Random, but it resulted in higher average genetic relationship coefficients than Random and did not reduce the probability of loss from a genetic defect. Including a genomic relationship in the economic score also kept the other genomic relationship averages at a low level. For example, with the constraint of 5% females per bull and herd, including g_{SNP} in the objective function (scenario GSNP) resulted in a g_{SEG1} of 0.148, compared with 0.143 with GSEG1 (Table 2). Using the pedigree relationships also reduced the genomic relationships compared with Random and MaxNTM, but not as much as using genomic relationships in the objective function.

Discussion

The results show that it is possible to reduce genetic relationships between RDC parents in herds with minimal effect on the genetic level. Including the cost of known recessive genetic defects when optimising mating strategies eliminated expression of known genetic defects, regardless of the genetic relationship used.

The strong correlation estimated between pedigree and genomic relationships confirms that dairy pedigrees are well documented in the Nordic countries. The correlations between the pedigree relationship and genomic relationship estimates were high: ≥0.83 for a_{3Gen} and ≥0.85 for a_{AllGen} (Table 1). Carthy *et al.* (2019) reported a correlation of 0.57 between pedigree relationships and genomic relationship, which is lower than in other studies (0.67-0.88) (VanRaden *et al.*, 2011; Pryce *et al.*, 2012). Pryce *et al.* (2012) concluded that pedigree depth plays a major role for the strength of correlation between pedigree relationships and genomic relationships. They found that when the number of generations of recorded ancestry was 2, 4, 6, and 8, the correlations were 0.67, 0.73, 0.84, and 0.87, respectively (Pryce *et al.*, 2012).

In general, genomic relationships were better at reducing pedigree relationships than pedigree relationships were at reducing genomic relationships. Furthermore, using GSEG1 reduced g_{SEG1} compared with Random from 0.188 to 0.143, and 3Gen reduced g_{SEG1} to 0.163, i.e. the relative difference was 55%. Hence, using genomic relationships could be an overall better and safer option than using pedigree relationships in keeping all average relationships studied low.

Conclusions

We studied mating allocations in Nordic Red Dairy Cattle and found that it was possible to reduce genetic relationships between parents with minimal effect on genetic level. Including the cost of known recessive genetic defects entirely eliminated the risk of loss due to the six known genetic defects. It was possible to reduce genomic relationships between parents with pedigree measures, but it was best done with genomic measures. Linear programming maximised the economic score for all herds studied within seconds, which means that the method is suitable for implementing in mating software to be used by advisors and farmers.

References

Berkelaar and others. 2020. LpSolve: Interface to 'Lp_solve' v. 5.5 - Available at: https://CRAN.R-project.org/package=lpSolve

Bérodier M., Berg P., Meuwissen T., Boichard D., *et al.* 2021. Animal 15:100016. https://doi.org/10.1016/j.animal.2020.100016

Cole J.B. 2015. Genet. Sel. Evol. 47:94. https://doi.org/10.1186/s12711-015-0174-9

Carthy T.R., McCarthy J., and Berry D.P. 2019. J. Dairy Sci. 102:8210–8220. https://doi.org/10.3168/jds.2019-16283. https://doi.org/10.3168/jds.2019-16283

de Cara M.Á.R., Villanueva B., Toro M.Á., and Fernández J. 2013. Mol. Ecol. 22:6091–6099. https://doi.org/10.1111/mec.12560

Pryce J.E., Hayes B.J., and Goddard M.E. 2012. J. Dairy Sci. 95:377–388. https://doi.org/10.3168/jds.2011-4254

R Core Team. 2020. A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. - Available at: https://www.r-project.org/

VanRaden P.M. 2008. J. Dairy Sci. 91:4414–4423. https://doi.org/10.3168/jds.2007-0980

VanRaden P.M., Olson K.M., Null D.J., and Hutchison J.L. 2011. J. Dairy Sci. 94:6153–6161. https://doi.org/10.3168/jds.2011-4624

680. Genome wide association study of blood cell count measures in Holstein cows

C.J. Siberski-Cooper[1*], J. Chinchilla-Vargas[2], L.M. Kramer[1], K.S. Lim[1], M.S. Mayes[1] and J.E. Koltes[1]

[1]Department of Animal Science, Iowa State University, Ames, IA 50011, USA; [2]PIC, Hendersonville, TN 37075, USA; coris@iastate.edu

Abstract

Complete blood counts (CBCs) measure abundance of immune and red blood cells. CBCs can indicate immune status of an animal and may be related to disease resistance. The current study aimed to identify genomic regions associated with CBCs in Holstein cows. A genome wide association study was performed with the BLUPF90 suite of programs using 60K Affymetrix® genotypes from 258 cows, with CBCs measured at three timepoints. Estimated heritabilities ranged from 0.09±0.07 (leucocytes) to 0.78±0.09 (mean corpuscular volume), while repeatably of measurements ranged from 0.28±0.54 (mean corpuscular hemoglobin concentration) to 0.99±0.06E-05 (red cell distribution width). A total of 4 one megabase genomic windows explained ≥1% of genetic variance and all of these appear pleiotropic for CBC traits. This is the first study to evaluate the genomic heritabilities of CBCs in Holstein cattle.

Introduction

Animal health is important to the welfare and profitability of dairy cows. For example, animal health can impact feed efficiency and milk quality (Siberski *et al.*, 2019). Identifying novel health traits that are heritable and understanding genetic variants impacting susceptibility to illness may be beneficial for future use in selection. To date, genetic evaluation of health traits has been largely based on producer recorded data that can be subjective and inconsistent across farms, resulting in lower heritabilities (König and May, 2018). Due to these challenges, selection for the improvement of disease resilience has yet to be implemented (Berghof *et al.*, 2019). Identification of easier to measure indicator trait(s) for animal health and stress events could help identify more production resilient animals.

Blood cells, such as those measured in complete blood counts (CBCs) are indicative of an animal's innate and adaptive immunity. Two studies have evaluated the genetic architecture of CBCs and their relationship to growth or disease in beef cattle (Chinchilla-Vargas *et al.*, 2020; Leach *et al.*, 2013). Chinchilla-Vargas *et al.* (2020) reported moderate to high heritabilities and a genome-wide association study (GWAS) for CBCs in beef cattle. In pigs, Mpetile *et al.* (2015) reported differences in CBCs between high and low feed efficiency groups, indicating a potential genetic relationship. Few studies have examined the genetic architecture of CBCs in Holstein cattle, while none are known to have assessed their relationship with resilience. Denholm *et al.* (2017) evaluated phenotype based heritabilities of neutrophils, monocytes, lymphocytes, and eosinophils in Holstein-Friesians, which were low to moderately heritable (0.15-0.35). Given the relationship of CBCs with health and productivity traits in several species, CBCs may be indicators for production resilience in dairy cattle. Thus, the objective of the study was to 1) estimate preliminary heritability and repeatability of CBC traits in Holstein cattle, and 2) identify genomic regions and single nucleotide polymorphisms (SNP) associated with CBCs in Holstein dairy cows.

Materials & methods

Animal husbandry. This study was approved by the Iowa State University (ISU) Institutional Animal Care and Use Committee protocol 18-174. Blood samples were collected on 313 Holstein cows from 8 contemporary groups (CG). The CGs were based on pen and cohort group within year, and consisted of 12 to 48 cows based on pen size. Cows ranged from parity one to six with data collected between 24 to 313

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_680, © C.J. Siberski-Cooper *et al.* 2022

days in milk (DIM). Animals were housed in a free stall barn and received the ISU dairy farm standard total mixed ration. Animals were milked three times daily in CG 1-3 and twice a day in all other CGs. Animal health was monitored by ISU veterinary personnel.

Data collection. Blood samples were taken from the caudal vein in EDTA vacutainer tubes at three time points on all animals (2.5 weeks between samplings.) Blood samples were sent to the ISU Veterinary Diagnostic Laboratory where CBCs were measured using an Advia 2120i. Fifteen cellular measurements were reported (Table 1.) A total of 258 animals were genotyped with the 60k Affymetrix® Axiom® SNP chip. Genotypes were filtered to keep SNP markers with minor allele frequency ≥0.05 and samples with a genotyping call rate ≥0.90. Following filtering, 40,000 SNPs remained for use in the genomic analyses.

Statistical analyses. The distributions of CBC measures were assessed for normality with a quantile-quantile plot. The WBC, NEUT, LYMP, EOSI, BASO, and LUC were log transformed, and a Box-Cox transformation (lambda = -2) was used for RDW to obtain approximately normally distributed data. The remainder of the CBC measures did not require transformation for subsequent analyses.

Health events were classified into one of four categories: healthy, mastitis, lameness, and other events. The other category included multiple concurrent health events, respiratory issues, and injuries not included in lameness. The association of health status with CBC measures was assessed using the following model:

$$CBC_{ijkl} = Health_i + Lact_j + CG_k + DIM_l + e_{ijkl} \quad (1)$$

where CBC_{ijkl} is one of the 15 CBC measures (Table 1), $Health_i$ is the binary (yes/no) class effect of healthy vs each individual health event, $Lact_j$ and CG_k are the class effects of the j^{th} lactation (1, 2 or 3+) and k^{th} contemporary group, DIM_k is the fixed effect of the l^{th} day in milk, and e_{ijkl} is the random residual.

Table 1. Preliminary narrow sense heritability and repeatability estimates of complete blood count measures.

Measure	Abbreviation	Unit	Heritability ± SE[1]	Repeatability ± SE[1]
White blood cells[2]	WBC	K/ul	0.10±0.08	0.67±0.039
Red blood cells	RBC	M/ul	0.46±0.14	0.74±0.03
Hemoglobin	HGB	g/dl	0.19±0.13	0.65±0.038
Hematocrit	HCT	%	0.19±0.12	0.63±0.04
Mean corpuscular volume	MCV	fl	0.78±0.086	0.90±0.013
Mean corpuscular hemoglobin	MCH	Pg	0.76±0.09	0.89±0.014
MCH Concentration	MCHC	g/dl	0.18±0.069	0.28±0.054
Red cell distribution width[3]	RDW	%	0.51±0.029	0.99±0.06E-05
Platelets	PLT	K/ul	0.17±0.087	0.36±0.06
Neutrophils[2]	NEUT	K/ul	0.14±0.076	0.35±0.062
Lymphocytes[2]	LYMPH	K/ul	0.12±0.084	0.86±0.017
Monocytes	MONO	K/ul	0.18±0.071	0.33±0.058
Eosinophils[2]	EOSI	K/ul	0.26±0.095	0.39±0.062
Basophils[2]	BASO	K/ul	0.37±0.11	0.66±0.038
Leukocytes[2]	LUC	K/ul	0.09±0.07	0.34±0.058

[1] Standard error.

[2] Log transformed.

[3] Box-Cox transformed.

Estimation of narrow sense heritability and repeatability of CBC measures was conducted with THRGIBBSF90 and POSTGIBBSF90 from the BLUPF90 family of software using 10,000 sampling iterations and a burn-in of 2,500 samples (Aguilar *et al.*, 2018). The GWAS were completed with the single-step GBLUP approach for each of the CBC measures with THRGIBBSF90 and POSTGSF90 programs (Aguilar *et al.*, 2018.) The general linear mixed model used for the association analysis was:

$$Y = Xb + Za + e \quad (2)$$

where Y is the CBC measure, X is the incidence matrix for fixed effects including CG, parity (1, 2, 3+) and DIM, Z is the incidence matrix for random effects (i.e. animal), b is the vector of regression coefficients for fixed effects, a is the vector of additive genetic effects, and e is the random error vector. The GWAS utilized 50,000 sampling iterations and a burn-in of 10,000 samples to estimate the amount of variation explained by 1 megabase (Mb) windows across all bovine autosomes. Gene annotations within significant genomic windows (i.e. quantitative trait loci, QTL) were obtained from the NCBI Gene Database (https://www.ncbi.nlm.nih.gov/gene).

Results

Analyses from model 1 identified associations of mastitis and other health events with log LYMPH count ($P<0.05$). Significant associations were also identified for other health events and monocyte count, and mastitis with log WBCs ($P<0.05$). There was a tendency towards an association of log BASO with mastitis, and other health events with MCHC ($P<0.10$).

Preliminary heritability and repeatability estimates for CBC measures are presented in Table 1. The largest heritability was obtained for MCV (0.78±0.086), while most of the heritabilities were between 0.10 and 0.30 (n=8). The majority (n=9) of repeatabilities were >0.60. The most repeatable measure was Box-Cox transformed RDW (0.99±0.06E-05), while the next most repeatable was MCV (0.90±0.013). The least repeatable CBC measure was MCHC (0.28±0.054).

Between 3-4 QTL per CBC trait explained ≥1% of the genetic variation in CBC measures (Table 2). Several pleiotropic loci were identified. Two 1 Mb genomic regions on *Bos taurus* autosomes (BTA) 1 (1 Mb) and 6 (71 Mb) explained more than 1% of variation for all CBC measures, as well as two adjacent genomic windows on BTA 4 (113 and 114 Mb). Two neighbouring 1 Mb windows on BTA 21 (20 and 21 Mb) explained more than 1% of the genetic variation for 12 of the CBC measures. The QTL on BTA 1, 4 (two adjacent windows), 6, and 21 (two adjacent windows) contained 25, 47, 22, and 62 genes, respectively. The windows on BTA 4 and 6 accounted for the largest amount of genetic variance (4.67-6.88% and 5.37-8.79%, respectively). The largest amount of variance explained by any genomic window was on BTA 6 for log WBC (8.79%.)

Discussion

The heritability of CBC measures reported in the current study are of similar magnitudes as previous studies (Chinchilla-Vargas *et al.*, 2020; Denholm *et al.*, 2017), with some (i.e. MCV and MCHC) being much larger which may be due to the small study size. Additionally, the ranking of CBC heritabilities is similar to that in beef cattle (Chinchilla-Vargas *et al.*, 2020). The findings of the current study are preliminary, but they do indicate that CBC traits are heritable and repeatable in Holstein cows.

Four genomic regions individually accounted for ≥1% of variation for nearly all CBC measures. These regions were identified on BTA 1, 4, 6 and 21. Interestingly, the region on BTA 6 contains the *HOPX* gene, which has been associated with higher platelet counts and lower white blood cells (Lin *et al.*, 2017).

Table 2. Summary of 1 Mb windows explaining ≥1.0% of variance in complete blood count measures.

Measure	Number of regions explaining ≥1.0% GV[1]	Range in % GV[1] explained
Log white blood cells	4	1.15-8.79
Red blood cells	3	1.19-7.23
Hemoglobin	3	1.50-6.80
Hematocrit	4	1.01-6.37
Mean corpuscular volume	4	1.30-6.74
Mean corpuscular hemoglobin (MCH)	4	1.28-6.73
MCH Concentration	4	1.27-6.49
Box-Cox red cell distribution width	3	1.34-5.37
Platelets	4	1.07-7.32
Log neutrophils	4	1.07-7.81
Log lymphocytes	4	1.26-8.67
Monocytes	4	1.12-8.66
Log eosinophils	4	1.13-6.93
Log basophils	4	1.18-7.81
Log leukocytes	4	1.29-7.82

[1] GV = additive genetic variance.

Within the 1 Mb window on BTA 1 are the *IFNAR1* and *INFAR2*.genes. A study of *INFAR1* knockout mice indicated that neutrophil development and functionality were altered without a functional copy of the gene (Siakaeva *et al.*, 2019).

Though the sample size of the current study is small, it demonstrates CBCs are heritable in Holsteins and provides the first estimates of heritability and possible candidate genes for some of the CBCs. Several pleotropic QTL were identified, which indicates shared genetic control. Future research with larger populations is needed to determine genetic correlations of CBC traits with production traits to identify if they are useful in selection. In addition, validation of pleotropic QTL would be interesting to determine which mutations may have shared genetic control of multiple immune cell traits. This information may be important in understanding disease and production resilience (i.e. more robustly efficient dairy cattle.)

References

Aguilar, I., Tsuruta, S., Masuda, Y., Lourenco, D.A.L., Legarra, A., *et al.* (2018) Proc. of the 11th World Congress of Genetics Applied to Livestock Production, Auckland, New Zealand

Chinchilla-Vargas, J., Kramer, L.M., Tucker, J.D., Hubbell, D.S. III, Powell, J.G., *et al.* (2020) Front. Genet. 11:717. https://doi.org/10.3389/fgene.2020.00717

Denholm, S.J., McNeilly, T.N., Banos, G., Coffey, M., Russell, G.C., *et al.* (2017) J. Dairy Sci. 100:2850-2862. https://doi.org/10.3168/jds.2016-11679

König, S., and May, K. (2019) Animal 13(5):897-908. https://doi.org/10.1017/S1751731118003208

Leach, R.J., Chitko-McKown, C.G., Bennett, G.L., Jones, S.A., Kachman, S.D., *et al.* (2013) J. Anim. Sci. 91:3564-3573. https://doi.org/10.2527/jas.2012-5911

Mpetile, Z., Young, J.M., Gabler, N.K., Dekkers, J.C.M., and Tuggle, C.K. (2015) J. Anim. Sci. 93(3):892-899. https://doi.org/10.2527/jas.2014-8132

Siberski, C., Mayes, M.S., Gorden, P., Copeland, A., Goetz, B.M., *et al.* (2019) Proc. of the Annual American Dairy Science Association Meeting, Cincinnati, OH, USA.

681. Investigation of intestinal atresia in a Jersey sire family

T.R. Carthy[1*], O.M. Keane[1], J.P. Hanrahan[2], D. Matthews[3], J. McEwan[4], S. Rowe[4] and J. Mee[5]

[1]Animal & Bioscience Department, Teagasc, Grange, Dunsany, Co. Meath, Ireland; [2]University College Dublin, Belfield, 4, Dublin, Ireland; [3]Irish Cattle Breeding Federation, Bandon, Co. Cork Ireland; [4]AgResearch, Invermay Agricultural Centre, Private Bag 50034, Mosgiel, New Zealand; [5]Animal & Bioscience Department, Teagasc, Moorepark, Co. Cork, Ireland; tara.carthy@teagasc.ie

Abstract

Intestinal atresia is a congenital defect resulting in calf mortality within a few days of birth. This study focuses on two half-sibling Jersey sires who were identified as having a high proportion of their progeny exhibiting atresia from a longitudinal study of 39 dairy herds over an 8 year period. The objective of this study was to investigate the genetic cause of intestinal atresia within this family. Phased high-density genotypes of the two half-sibling sires were used to identify if common haplotypes were inherited by all affected animals. Whether or not the paternal haplotype that was inherited by the affected animal at a given locus deviated from the expectation was tested. A total of 16 SNPs were significantly overrepresented, 14 located in one region on chromosome 14 and the remaining 2 located in one region on chromosome 26.

Introduction

Intestinal atresia is a congenital defect, found in a wide variety of species, which results in complete occlusion of the intestinal lumen, occurring in the jejunum, ileum, colon and/or anus (Ducharme *et al.*, 1988). In calves, the blockage inhibits the normal movement of gut contents resulting in failure to pass meconium and faeces, inappetence, abdominal distension, and usually results in death due to autointoxication within a few days of birth (Syed and Shanks, 1992). The inheritance of atresia in cattle is still not completely understood (Constable and Morin, 2000). As part of a larger study, active surveillance of 39 dairy herds over 8 years was used to assess the incidence of intestinal atresia in dairy herds in Ireland. Of interest, two Jersey sires had a high proportion of their progeny exhibiting intestinal atresia; further investigation revealed that these two sires shared a common sire. The objective of this study was to investigate the genetic cause of intestinal atresia within this family.

Materials & methods

Data collection. A total of 39 dairy herds, including 3 research dairy herds from Teagasc, participated in a longitudinal study over 8 years to assess factors associated with intestinal atresia in dairy herds in Ireland. As part of the study any calf that died within the first 7 days of life was transported to Teagasc Moorepark and subjected to a full post-mortem by a veterinary surgeon. Following post-mortem, a total of 197 cases of intestinal atresia were identified. Additional data on all calves born on each farm during study participation was provided by the Irish Cattle Breeding Federation (ICBF) and consisted of calf ID, date of birth, sex, dam breed, dam ID, dam parity, sire breed, sire ID and birth type (single or twin).

Identification of sire family. During data exploration, generalized linear models were used in SAS version 9.4 (SAS Institute Inc., Cary, NC, USA) to quantify the association of risk factors associated with atresia. Initial results indicated sire was a significant factor in intestinal atresia among Jersey sired calves. Further investigation, limiting records to sires that had greater than 50 progeny born in at least 5 different herds, indicated an over-representation of affected progeny for a number of sires. The two most significant sires, which are paternal half-siblings, had 2.5 and 4.1% of their progeny exhibiting intestinal atresia.

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_681, © T.R. Carthy *et al.* 2022

DNA extraction and genotyping. Muscle tissue (>1 g) was collected post-mortem from 28 atresia cases from the two prominent sires, submerged in RNAlater solution (Ambion), left overnight at 4 °C and stored at -20 °C. For DNA extraction approximately 100 mg of tissue was minced in 1.5 ml of digestion buffer (1 M Tris-HCL pH 8.0, 0.5 M EDTA, 10% SDS, 500 µg/ml Proteinase K) and incubated for 4 h at 55 °C with shaking. The sample was then extracted twice with an equal volume of phenol:chloroform and once with an equal volume of chloroform. The DNA was precipitated with 0.1 volumes of 3 M sodium acetate pH 5.2 and 1 volume of isopropanol, washed with 70% ethanol, air dried and resuspended in 100 µl of H_2O. DNA was cleaned using the Genomic DNA Clean and Concentrator kit (Zymo Research) and quality assessed by agarose gel electrophoresis. DNA was genotyped with the Illumina 777K BovineHD beadchip (Illumina Inc, San Diego, CA, USA). Illumina Bovine50k beadchip genotypes from 640 unaffected progeny of either sire were also provided by ICBF as healthy controls and imputed to HD using FImpute (Sargolzaei *et al.*, 2014).

Linkage analysis with Jersey family. High-density genotypes were available on the two sires along with 28 affected progeny; the genotypes were phased using FImpute (Sargolzaei *et al.*, 2014). The phased genotypes of the two sires were used to determine which paternal haplotypes were inherited. Only haplotypes where both sires inherited the same paternal allele and which were heterozygous in both sires were retained. Following edits 46,177 SNPs remained. Whether or not a paternal haplotype that was inherited by the affected progeny at a given locus deviated from the expectation of equal representation was tested. The critical t-statistic was calculated using Equation 1.

$$\frac{P-0.5}{\sqrt{\frac{P(1-P)}{n}+0.001}} \quad (1)$$

In Equation 1, P is the proportion of progeny with the paternal haplotype inherited from the grandsire and n is the number of grand-progeny. The level of significance was determined using a one-tailed t-test with n-1 degrees of freedom; the mean *P*-value of a 10-SNP sliding window was used to avoid sporadic single SNP significance.

Bioinformatics. SNPs in close proximity (1 Mb) were combined into a single QTL. Genes within 50 kb span of the start and end of the QTL were identified using Ensembl 94 on the UMD3.1 bovine genome assembly implemented in BiomaRt in R.

Results

Of the 730,540 autosomal SNPs available from the phased genotypes, 46,177 SNPs remained after confirming the same paternal haplotype was inherited by both sires as well as ensuring both sires were heterozygous. A total of 16 SNPs were significant, 14 located in one region on chromosome 14 and the remaining 2 located in one region on chromosome 26 (Table 1 and Figure 1). The first region between 14: 24,769,617 and 14: 24,939,286 contained 41 SNPs, 14 of which were part of the binomial analysis; this region contained six genes (Table 1). A total of 27 out of 28 affected animals had the same haplotype inherited for 41 of these SNPs with the remaining animal differing at 8 SNPs. The second region on chromosome 26 between 18,028,791 and 18,029,994 contained only the two SNPs that were part of the binomial analysis; in total 26 of the 28 animals had the same haplotype inherited from the paternal-grandsire. There was one gene located in this region, PIK3AP1.

Discussion

There was a total of 197 animals identified through post-mortem as part of a longitudinal study across 39 herds over 8 years as having intestinal atresia; of these, 28 affected animals originated from two paternal half-sibling Jersey sires. All animals within the study originated from dairy herds with the majority

Table 1. Significant QTLs associated with intestinal atresia in 28 affected Jersey cross Holstein animals originated from a common paternal-grandsire.

Chromosome	Start position	End position	Number of SNPs	*P*-value	Genes
14	24,769,617	24,939,286	14	1.79E-15	TMEM68, TGS1, LYN, RPS20, U1, MOS
26	17,890,705	17,891,907	2	3.04E-08	PIK3AP1

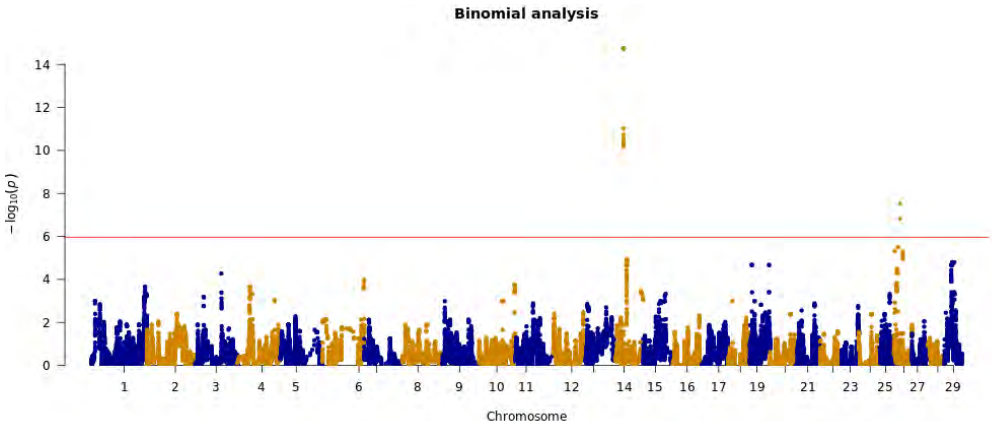


Figure 1. Binomial analysis of 28 affected Jersey cross Holstein animals originated from a common paternal-grandsire.

Holstein-Friesian or Holstein-Friesian crosses. Although Holstein-Friesian animals are reported to have a greater risk of atresia compared to other cattle breeds (Constable *et al.*, 1997), our data suggests that the genetic cause of intestinal atresia in these animals originated from the Jersey parent. The overall incidence of atresia in the current dataset was 0.35% whereas the two sires of interest in this study had an incidence among their progeny of 2.5 and 4.1%.

Unfortunately, the genotype of the paternal grandsire of the affected animals was unavailable; however, HD genotypes of both sires were available. Using both sire's genotypes, the haplotype of the paternal grandsire that was inherited by the affected animals was inferred using phased haplotypes. Due to the number of affected animals, traditional GWAS might not be able to provide strong evidence of regions associated with atresia. The alternative method used in the current study aimed to indicate any region that may require further investigation. One regions on chromosome 14 was shown to have the paternal grandsire's haplotype over-represented. In this region, all but one animal had inherited the same 169 kb region. Additionally, this region was also examined in the healthy controls; in the unaffected animals the region was also significantly over represented with 450 out of 680 animals also inheriting the same haplotype as the affected animals. None of the genes were obvious candidate genes, although mutations in these genes have been associated with multiple health and disease traits.

Conclusions

Intestinal atresia normally results in the death of the calf on-farm. Although it has been reported that bovine intestinal atresia has a genetic aetiology (Syed and Shanks, 1992) no causal mutation has been identified. In the current study we identified a sire-family that has an over-represented incidence of atresia. Although there is strong evidence of association with a region on chromosome 14, the genes within this

region have not been previously implicated in intestinal atresia. There are possible benefits to fine mapping this region in affected and unaffected animals to narrow the QTL containing any causal mutation.

References

Constable, P., R. Shanks, J. Huhn, and D. Morin. 1997. Evaluation of breed as a risk factor for atresia coli in cattle. Theriogenology 48:775-790.

Constable, P. D. and D. E. Morin. 2000. Atresia coli in Calves: Cause and Surgical Management.

Ducharme, N. G., M. Arighi, F. D. Horney, I. K. Barker, M. A. Livesey, M. H. Hurtig, and R. P. Johnson. 1988. Colonic atresia in cattle: a prospective study of 43 cases. The Canadian Veterinary Journal 29:818.

Sargolzaei, M., J. P. Chesnais, and F. S. Schenkel. 2014. A new approach for efficient genotype imputation using information from relatives. BMC genomics 15:478.

Syed, M. and R. Shanks. 1992. Incidence of atresia coli and relationships among the affected calves born in one herd of Holstein cattle. J. Dairy Sci. 75:1357-1364.

682. The genomic architecture of feed efficiency over the lactation of dairy cattle

K. Houlahan[1*], F.S. Schenkel[1], H.R. Oliveira[1,2], F. Miglior[1,2], G.A. Oliveira Jr.[1], M.J. VandeHaar[3], K. Weigel[4], R.J. Tempelman[3] and C.F. Baes[1,5]

[1]Centre for Genetic Improvement of Livestock, University of Guelph, 50 Stone Road E, Guelph, ON, N1G 2W1, Canada; [2]Lactanet, 660 Speedvale Ave., Guelph, ON, N1K 1E5, Canada; [3]Department of Animal Science, Michigan State University, 474 S Shaw Ln, East Lansing, MI 48824, USA; [4]Department of Animal and Dairy Science, University of Wisconsin-Madison, 1675 Observatory Dr., Madison, WI 53706, USA; [4]Institute of Genetics, Vetsuisse Faculty, University of Bern, Bremgartenstr.109a, 3012 Bern, Switzerland; khoulaha@uoguelph.ca

Abstract

To understand the potential genomic implications of selecting for feed efficiency, 38,294 phenotypic records from 4,024 first lactation North American Holsteins for dry matter intake, energy corrected milk, metabolic body weight, and residual feed intake were used to estimate genetic parameters using single trait random regression models. The lactation was then split into two time-points (pre-and post-peak lactation) and genome-wide association studies were performed on each stage of lactation, and the random regression coefficients using imputed whole-genome sequence genotypes. A total of seven and 24 functional candidate genes were identified for pre-and post-lactation, respectively. No overlap was observed between the functional candidate genes for pre-and post-lactation. Furthermore, single nucleotide polymorphism effects associated with functional candidate genes changed in magnitude and sign throughout the lactation. These results highlight the dynamic genomic behaviour of feed efficiency and the importance of collecting data across all stages of lactation.

Introduction

With costs of production increasing, improving farm and animal efficiency is paramount to assure sustainability in the dairy industry. The physiology of lactation must also be taken into consideration when improving feed efficiency (FE). It is well known that FE is heritable and consequently, that selection for improved FE is possible (Houlahan *et al.*, 2021), causing interest surrounding the genomic background of FE to grow (Li *et al.*, 2019). Residual feed intake (RFI) is a commonly used indicator trait of FE. Generally, RFI is calculated as the residual of the fixed linear regression of dry matter intake (DMI) on an energy sink related to milk production (e.g. energy corrected milk (ECM)), and maintenance metabolic function (e.g. metabolic body weight (MBW)). Previous studies have shown that FE (RFI) and its underlying traits, such as DMI, ECM, and MBW exhibit a dynamic behaviour over the lactation (Krattenmacher *et al.*, 2019), which suggests that the candidate genes associated with FE and their effects may also change throughout the lactation. The objective of this study was to identify genomic regions and candidate genes associated with DMI, ECM, MBW, and RFI considering two stages of lactation (pre-and post-peak), and to investigate the potential differential pattern of single nucleotide polymorphism (SNP) effects over the lactation.

Materials & methods

Phenotypic data. Data was provided by the Efficient Dairy Genome Project (https://genomedairy.ualberta.ca/) and the Resilient Dairy Genome Project (http://www.resilientdairy.ca/). The data contained 31,967 (DMI), 38,294 (ECM), 26,394 (MBW), and 23,083 (RFI) phenotypic records from 4,024 first lactation North American Holstein cows in 10 research herds and one commercial farm. All cows had their first calving between January 2007 and November 2020. Data editing procedures are detailed in Houlahan

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_682, © K. Houlahan *et al.* 2022

(2021). The RFI was calculated as the residual of the fixed linear regression of DMI on ECM and MBW (body weight$^{0.75}$), where ECM was calculated following the methodology outlined in Sjaunja *et al.* (1990).

Genotypes and quality control. Imputed (Minimac3 (Das *et al.*, 2016)) whole-genome sequence (WGS) data containing 42,158,739 SNP were available for 4,702 Holstein cows. Quality control was performed, which required that variants had an R^2 value >0.90 (calculated as the observed variance over the expected variance given the observed allele frequency), minor allele frequency >0.01, and no extreme deviation from Hardy-Weinberg equilibrium ($P>10^{-8}$). A total of 9,598,535 high confidence WGS SNP remained for the GWAS analysis.

Variance components. Genetic parameters were estimated using single-trait random regression animal models. Heterogeneous residual variances across four classes (5 to 70, 71 to 140, 141 to 210, and 211 to 305 days in milk (DIM)) were fitted for all traits. A quadratic Legendre polynomial was used for all analyses, as higher-order polynomials lead to convergence problems. The general model used for all traits is described in Houlahan (2021).

GWAS. All GWAS performed in this study used deregressed genomic estimated breeding values (dGEBV) as pseudo-phenotypes for pre-peak lactation (5-60 DIM) and post-peak lactation (61-305 DIM) and were derived from the random regression coefficients. The GWAS were performed using the *mlma* option in the program GCTA (Yang *et al.*, 2011), which fits a linear mixed model, including a polygenic effect. After the GWAS, SNP for pre-and post-peak lactation were ranked based on their *P*-values and clumped according to their linkage disequilibrium pattern ($r^2>0.9$) (Prive *et al.*, 2018). To correct for the potential false positives, the threshold of significance (α) per chromosome was calculated as Chen *et al.* (2020). This resulted in SNP being considered statistically significant if the -log10 *P*-value was higher than the chromosome-wide threshold, which ranged from 4.16 to 4.66 (average ± SD: 4.40±0.14). Significant SNP were analysed in downstream analysis for potential functional candidate genes (FCG). To assess the change in SNP effects over the lactation for each trait, the significant SNP associated with FCG in pre-and post-peak lactation were used. The SNP effects for the significant SNP were calculated as the change in kilograms per two-week segment of the lactation (21 segments), using the SNP effect solutions from the random regression coefficients.

Results

A total of 30 (DMI), 24 (ECM), 30 (MBW), and 37 (RFI) significant SNP (Figure 1a) were found in pre-peak lactation. These significant SNP were associated with two (DMI), two (ECM), one (MBW), and two (RFI) FCG in pre-peak lactation. In post-peak lactation, 151 (DMI), 78 (ECM), 35 (MBW) and 64 (RFI) significant SNP (Figure 1b) were observed to be associated with six (DMI), 10 (ECM), four (MBW) and four (RFI) FCG, respectively. The FCG in pre-peak lactation were previously associated with the digestive tract (salivary and gastric secretions; Bergstrom and Xia, 2013), mammary tissue development, milk, and component synthesis (Do *et al.*, 2017), and cortisol synthesis and secretion (Soares *et al.*, 2021). In post-peak lactation, the FCG were associated with the digestive tract (smooth muscle contraction; Fiscus, 2002), oxytocin signalling (Chatterjee *et al.*, 2016), and deposition of fat (Cantalapiedra-Hijar *et al.*, 2018).

The SNP effects associated with the FCG varied across the lactation (Figure 2). The amount of variation was dependent on the trait and SNP. For instance, the effects of a SNP associated with *KCNMA1* (a FCG associated with RFI) ranged from -3.38 to 4.43 kg from early to late lactation. Additionally, a significant SNP related to *PDE2A* (a FCG associated with DMI) changed up to 4 kg (-3.36 to 0.85 kg) across the lactation. Both ECM and MBW had lesser variation across the lactation compared to RFI and DMI, with a minimum variation of 1.50 kg (*GABRA5*, FCG associated with MBW) to a maximum variation of 3.33 kg (*CLDN23*, FCG associated with MBW) throughout the lactation.

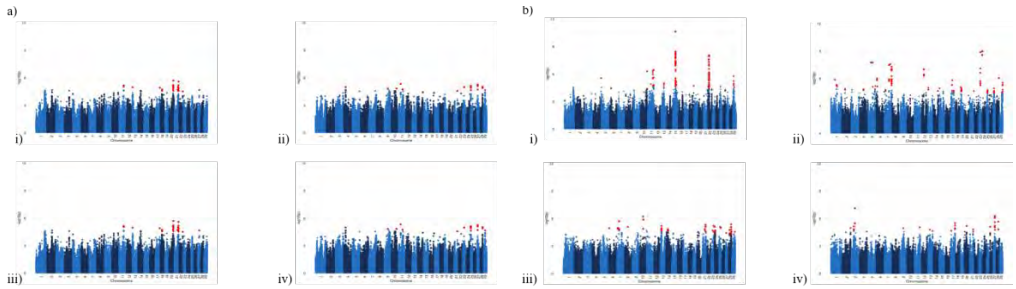


Figure 1. Manhattan plots for (a) pre-peak lactation and (b) post-peak lactation where: (1) dry matter intake (DMI); (2) energy corrected milk (ECM); (3) residual feed intake (RFI); and (4) metabolic body weight (MBW). Significant SNP are marked in red.

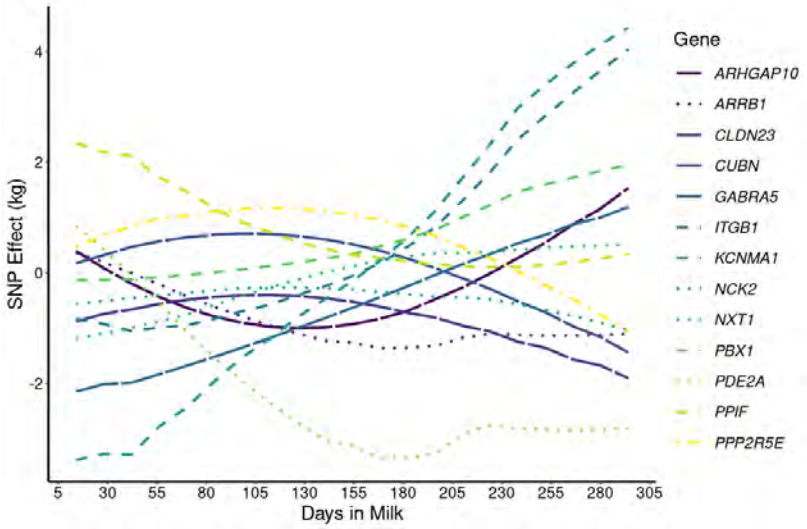


Figure 2. SNP effects over a 305-day lactation for the significant SNP associated with the functional candidate genes for dry matter intake (dotted line), energy corrected milk (large, dashed line), metabolic body weight (dot with dashed line), and residual feed intake (short, dashed line).

Discussion

Differences in the set of significant SNP and FCG between pre-and post-peak lactation suggest that the genetic architecture of FE changes over the lactation. A potential reason for the lack of overlap could be that there was a limited number of animals in this study, which limited power to detect SNP/FCG with small/moderate effects. The biological pathways associated with the FCG for all traits were different between pre-and post-peak lactation. Combined with the variability observed in the SNP effects over the lactation supports the idea that the genetic architecture of FE changes across the lactation. The changing genomic background suggests that multiple time points within the lactation should be considered for selection.

Conclusions

The different significant SNP and FCG between pre-and post-lactation, along with the dynamic SNP effects, support the idea that the genomic architecture of FE changes over the lactation. The results of this study

reinforce the complex nature of FE, which highlights the need for data collection across lactation to enable optimal selection for FE. Future studies should also consider the impact of including change in body weight or body condition score in the RFI equation.

Acknowledgements

The authors acknowledge partners and collaborators of the efficient dairy genome project (https://genomedairy.ualberta.ca/) and the resilient dairy genome project (http://www.resilientdairy.ca).

References

Bergstrom K.S.B., and L. Xia. (2013). Front. Genet.11:121. https://doi.org/10.3389/fgene.2020.00121.

Cantalapiedra-Hijar, G., M. Abo-Ismail, G.E. Carstens, L.L. Guan, R. Hegarty, *et al.* (2018). animal. 12(s2):S321–S335. https://doi.org/10.1017/S1751731118001489.

Chatterjee, O., K. Patil, A. Sahu, L. Gopalakrishnan, P. Mol. (2016). J. Cell Commun. Signal. 10(4):355–360. https://doi.org/10.1007/s12079-016-0353-7.

Chen, S.Y., H.R. Oliveira, F.S. Schenkel, V.B. Pedrosa, M.G. Melka, *et al.* (2020). J. Dairy Sci. 103(11):10383–10398. https://doi.org/10.3168/jds.2020-18897.

Das, S., L. Forer, S. Schönherr, C. Sidore, A.E. Locke, *et al.* (2016). Nat. Genet. 48(10):1284–1287. https://doi.org/10.1038/ng.3656.

Do, D.N., P.L. Dudemaine, R. Li, and E.M. Ibeagha-Awemu. (2017). Int. J. Mol. Sci. 18(17):1560. https://doi.org/10.3390/ijms18071560.

Fiscus, R.R. (2002). Neurosignals 11(4):175–190. https://doi.org/10.1159/000065431.

Houlahan K. (2021). Ph.D. Thesis. University of Guelph. https://atrium.lib.uoguelph.ca/xmlui/handle/10214/26420.

Houlahan, K., F.S. Schenkel, D. Hailemariam, J. Lassen, M. Kargo, *et al.* (2021). animals 11:1157. https://www.doi.org/10.3390/ani11041157.

Krattenmacher, N., G. Thaller, and J. Tetens. (2019). J. Dairy Sci. 102:3241–3253. https://www.doi.org/10.3168/jds.2018-15480.

Li, B., L. Fang, D.J. Null, J.L. Hutchison, E.E. Connor *et al.* (2019). J. Dairy Sci. 102(12):11067–11080. https://doi.org/10.3168/jds.2019-16645.

Prive, F., H. Aschard, A. Ziyatdinov, and M.G.B. Blum. (2018). Bioinformatics 34(16):2781–2787. https://doi.org/10.1093/bioinformatics/bty185. \

Soares, R.A.N., G. Vargas, T. Duffield, F. Schenkel, and E.J. Squires. (2021). J. Dairy Sci. 104:10076-10089. https://doi.org/10.3168/jds.2020-20101.

Sjaunja, L.O., L. Bævre, L. Junkkarinen, J. Pedersen, and J. Setälä. (1990). Page in International Committee for Recording the Productivity of Milk Animals.

Yang, J., S.H. Lee, M.E. Goddard, and P.M. Visscher. (2011). Am. J. Hum. Genet. 88(1):76–82. https://doi.org/10.1016/j.ajhg.2010.11.011.

683. Population structure and signatures of selection in Italian Holstein through genome-wide analysis of imputed SNP data

C. Persichilli[1*], M. Marusi[2], J.T. van Kaam[2], R. Finocchiaro[2], G. Senczuk[1], M. Di Civita[1], F. Pilla[1] and M. Cassandro[2,3]

[1]Department of Agriculture Environment and Food Science, University of Molise, Via F. De sanctis, 86100 CB Campobasso, Italy; [2]Italian Holstein, Brown and Jersey Breeders Association, Via Bergamo, 292, 26100 CR Cremona, Italy; [3]Department of Agronomy, Food, Natural resources, Animals and Environment (DAFNAE), University of Padova, Viale dell'Università 16, 35020 Legnaro (PD), Italy; c.persichilli@studenti.unimol.it

Abstract

Holstein Friesian is by far the most reared dairy breed. An important quota of the milk production is devoted to the production of Parmigiano Reggiano cheese. Consequently, the use of specific breeding values, and specific genotypes has become an essential selective tool for breeders leading to a possible genetic differentiation of the Italian population. In this work the genetic structure of the Italian Holstein has been investigated through an ADMIXTURE analysis performed on the SNP genotypes of 286 cows split into three groups: Italian, International and Parmigiano Reggiano. In addition, a selection signature analysis through a case-control F_{ST} outlier approach was performed to identify genes of interest. Results allowed to distinguish Italian Holstein from the International population and, to a lesser extent, from the Italian Holstein reared for Parmigiano Reggiano. The selection signature identified genes of interest for several productive traits such as reproductive traits, mammal gland dimension and resistance to infections.

Introduction

The Italian Holstein herd-book was established in 1956 by the National Association of Italian Friesian Cattle Breeders (ANAFI). In Italy, Italian Holstein is the most reared breed for milk for the production of Parmigiano Reggiano PDO cheese. Consequently, the use of specific breeding values (Marusi *et al.*, 2019), and specific genotypes such as the homozygosity of the B k-casein allele (Chessa *et al.*, 2020), has become an essential tool for breeders. This could have an effect on the genetic diversity of the Holstein breed as it is known that different selection purposes within the same breed may have an effect on its population structure (Ablondi *et al.*, 2021).

Aims of this study are to explore the genetic structure of the Italian Holstein assessing its distinctiveness from the American herd, and to identify whether the Italian Holstein used to produce Parmigiano Reggiano is also distinguishable from other Italian Holsteins. Moreover, a selection signature analysis was performed through a case-control F_{ST} outlier approach in order to identify possible genomic regions harbouring loci of production interest.

Materials & methods

Source of the genotypes and quality control. Based on pedigree and herd-book information, genotypic samples were selected according to their relationship level in order to avoid close kinship among samples. Three principal groups have been generated: one including only individuals from US and Canada, a second including Italian Holstein samples not used to produce Parmigiano Reggiano and a third group including only Italian Holstein cattle used to produce Parmigiano Reggiano (hereinafter called *PA-RE-MO*). For the latter group only individuals carrying homozygosity for the k-casein locus have been selected. In order to harmonize the genotypic data, as they come from different SNPs panels, an imputation approach using the software Pedimpute was performed (Nicolazzi *et al.*, 2013). 86,841 imputed SNPs were obtained and mapped against the ARS-UCD1.2 reference genome (Rosen *et al.*, 2020). The software PLINK (Purcell *et*

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_683, © C. Persichilli *et al.* 2022

al., 2007) was then used to filter for minor allele frequency (–maf 0.05) and missing genotype call rate (–geno 0.01) obtaining a final dataset of 79,464 SNPs.

Analysis of genetic relationships and population structure. A maximum likelihood clustering approach as implemented in the software ADMIXTURE (Alexander and Lange, 2011) with K values ranging from 2 to 5 has been used to assess the presence of genetic structure. The results have been plotted and ordered following the sampling grouping using the membercoef.circos function in the R package BITE (Milanesi *et al.*, 2017). Using the membercoef.cv function from the same package, the best values of K have been assessed.

FST-outlier analysis. In order to identify possible genomic regions under selection among the three Holstein groups, a genome-wide F_{ST} analysis was performed with a case–control model, (Mastrangelo *et al.*, 2019). Thus, F_{ST} pairwise comparisons of the three groups have been performed using the software PLINK. Significant SNPs have been detected and retained using the function *--assoc fisher* and the Bonferroni corrections have been calculated using the *--adjusted* function. To control the number of false positives, significant markers were defined by applying a *P*-value threshold of 1×10^{-20} which, corrected with the Bonferroni method resulted in a nominal value of 1.258431e-25 ($-\log^{10}(p) = 24.90$).

Results

Analysis of genetic relationships and population structure. The ADMIXTURE analysis showed a most reliable repartition at value of K=3 (Figure 1). Indeed, at this genetic structuring the *Foreign* group is clearly distinguishable from all the Italian Holstein showing a predominance of the red component. On the other hand, two main groups corresponding to the *Italian* and *PA-RE-MO* groups also showed a marked genetic distinctiveness.

F_{ST}-outlier analysis. The case-control F_{ST} analyses performed among groups, highlighted several potential genomic regions that might be under strong selection. As a consequence of our sampling strategy, therefore fully expected, a major selective sweep on BTA6 (which include the k-casein coding gene), is observable in all comparisons when including the *PA-RE-MO* group. However, additional 17 highly significant SNPs (Table 1) which are located on eight chromosomes (*1, 3, 7, 11, 12, 14, 27, 28)* are plainly detected in the *Italian* vs *Foreign* groups. Those significant SNPs fall within eight different coding genes as shown in Table 1.

Discussion

The major finding of the study is that Italian Holstein cattle have a different genetic architecture when compared to foreign cattle. To a lesser extent cattle reared for the production of Parmigiano Reggiano are also differentiated. Italian Holstein used for the production of Parmigiano Reggiano retains higher genomic diversity in terms of composition while both foreign and Italian groups seem to be characterized by less

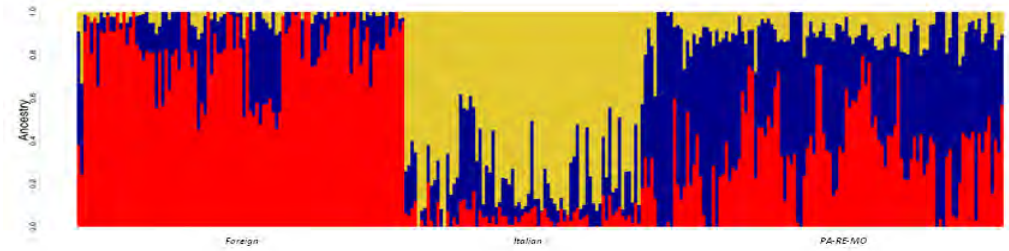


Figure 1. ADMIXTURE analysis depicted as a barplot at the most reliable number of K value (K=3).

Table 1. Results of case-control analysis between Italian and foreign Holstein.

SNP	CHR	POS	FST	Bonferroni corrected *P*-value	GENE
BOVINEHD0100023079	1	79639732	0.574183	2.948E-29	
BOVINEHD0100023533	1	81343507	0.563481	6.242E-28	
BOVINEHD0100023541	1	81380432	0.544815	1.942E-26	
BOVINEHD0100023776	1	82304016	0.563513	2.745E-29	*VPS8*
BTA-38502-NO-RS	1	82339579	0.581187	2.98E-32	*VPS8*
ARS-BFGL-NGS-14913	1	82360713	0.586364	9.23E-33	*VPS8*
BOVINEHD0300004169	3	12761961	0.544718	1.942E-26	*FCRL4*
BOVINEHD0700032048	7	107162220	0.632348	2.473E-33	*FBXL17*
BOVINEHD0700032242	7	107985586	0.600288	4.816E-30	*FER*
ARS-BFGL-NGS-77753	11	69177619	0.562703	1.633E-28	
BOVINEHD1200006184	12	20519397	0.551694	6.771E-26	*C12H13orf42*
BOVINEHD1200014439	12	51934707	0.556174	7.696E-26	
ARS-BFGL-NGS-29470	14	7917242	0.599573	4.115E-30	*ST3GAL1*
HAPMAP26061-BTC-063438	14	9485443	0.640775	3.292E-34	
BOVINEHD1400004018	14	12793296	0.544428	1.942E-26	
BOVINEHD2700009358	27	33494308	0.571073	4.888E-27	*NSD3*
BOVINEHD2800010243	28	37174238	0.614069	3.408E-31	*NRG3*

genetic diversity probably due to genetic drift (Senczuk *et al.*, 2020). Concerning the selection signature analysis, a strong signal was detected on chromosome 6, when comparing *PA-RE-MO* vs the other two groups. This is obviously due to the sampling criteria. On the other side, regions harbouring genes possibly responsible of the difference between Italian and Foreign population were identified. Several of these genes are also related to traits of interest in the zootechnical context: *VPS8* is involved in milk traits and reproductive traits (Taye *et al.*, 2017), *FER* has a putative role in the regulation of innate immune response (Flori *et al.*, 2009), *ST3GAL1* is related to female reproductive system in mammals (Sherblom *et al.*, 1985), *NSD3* seems to have a role in male fertility as semen quality traits (Capra *et al.*, 2019), while *NRG3* has been argued to be linked to bovine mammary gland development (Ogorevc *et al.*, 2009).

In conclusion a structure in the Holstein population reared in Italy has been identified according to the geographic origin of the individuals. Moreover, some of the loci involved in the selection for Parmigiano Reggiano production were highlighted and also loci putatively involved in the differentiation of the Italian population from the abroad.

References

Ablondi, M., Malacarne, M., Cipolat-Gotet, C., van Kaam, J.-T., *et al.* (2021) Genome-wide scan reveals genetic divergence in Italian Holstein cows bred within PDO cheese production chains. *Scientific Reports*, *11*(1): 12601. https://doi.org/10.1038/s41598-021-92168-1

Alexander, D. H., & Lange, K. (2011). Enhancements to the ADMIXTURE algorithm for individual ancestry estimation. *BMC bioinformatics*, *12*(1), 1–6.

Capra, E., Lazzari, B., Turri, F., Cremonesi, P., Portela, A. M. R., *et al.* (2019) Epigenetic analysis of high and low motile sperm populations reveals methylation variation in satellite regions within the pericentromeric position and in genes functionally related to sperm DNA organization and maintenance in *Bos taurus*. *BMC Genomics 20*(1):940. https://doi.org/10.1186/s12864-019-6317-6

Flori, L., Fritz, S., Jaffrézic, F., Boussaha, M., Gut, I., Heath, S., *et al.* (2009) The Genome Response to Artificial Selection: A Case Study in Dairy Cattle. *PLOS ONE*, *4*(8), e6595. https://doi.org/10.1371/journal.pone.0006595

Marusi, M., Visentin, G., Cipolat-Gotet, C., Summer, A., Formigon, A., *et al.* (2019) Cheesemaking and sustainability index – Parmigiano Reggiano: A new breeding index for the Italian Holstein. *Book of Abstracts*, *18, supplement 1*, 239. https://doi.org/10.1080/1828051X.2019.1622269

Mastrangelo, S., Criscione, A., Sottile, G., Portolano, B., *et al.*, (2019) Genome-wide analysis identifies potentially causative genes explaining the phenotypic variability in Pinzirita sheep. *Animal Genetics*, *50*(2):189–190. https://doi.org/10.1111/age.12763

Milanesi, M., Capomaccio, S., Vajana, E., Bomba, L., Garcia, J. F.,et al. (2017) BITE: An R package for biodiversity analyses. *BioRxiv*, 181610. https://doi.org/10.1101/181610

Nicolazzi, E. L., Biffani, S., & Jansen, G. (2013) Short communication: Imputing genotypes using PedImpute fast algorithm combining pedigree and population information. *Journal of Dairy Science*, *96*(4):2649–2653. https://doi.org/10.3168/jds.2012-6062

Ogorevc, J., Kunej, T., Razpet, A., & Dovc, P. (2009) Database of cattle candidate genes and genetic markers for milk production and mastitis. *Animal Genetics*, *40*(6):832–851. https://doi.org/10.1111/j.1365-2052.2009.01921.x

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A. R., *et al.* (2007) PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. *The American Journal of Human Genetics*, *81*(3), 559–575. https://doi.org/10.1086/519795

Rosen, B. D., Bickhart, D. M., Schnabel, R. D., Koren, S., Elsik, C. G., *et al.* (2020) De novo assembly of the cattle reference genome with single-molecule sequencing. *GigaScience. 9*(3):giaa021. https://doi.org/10.1093/gigascience/giaa021

Senczuk, G., Mastrangelo, S., Ciani, E., Battaglini, L., Cendron, F., *et al.* (2020). The genetic heritage of Alpine local cattle breeds using genomic SNP data. *Genetics Selection Evolution. 52*(1):40. https://doi.org/10.1186/s12711-020-00559-1

Sherblom, A. P., Smagula, R. M., Moody, C. E., & Anderson, G. W. (1985). Immunosuppression, sialic acid, and sialyltransferase of bovine serum as a function of progesterone concentration. *Journal of Reproduction and Fertility*, *74*(2), 509–517. https://doi.org/10.1530/jrf.0.0740509

Taye, M., Lee, W., Jeon, S., Yoon, J., Dessie, T., *et al.* (2017) Exploring evidence of positive selection signatures in cattle breeds selected for different traits. *Mammalian Genome 28*(11):528–541. https://doi.org/10.1007/s00335-017-9715-6

684. Genome-wide association studies for maturity rate in grazing dairy cows

N. Lopez-Villalobos*, H.B.P.C. Ariyarathne and R.E. Hickson

School of Agriculture and Environment, Massey University, Private Bag 11222, 4442, Palmerston North, New Zealand; n.lopez-villalobos@massey.ac.nz

Abstract

The objectives of this study were to estimate heritability and perform genome-wide association analysis for maturity rate (MR) in dairy cows from two dairy farms under New Zealand grazing conditions. MR was defined as the ratio between total lactation yields in the first and third lactations. Estimates of heritability of MR for milk (MY), fat (FY) and protein (PY) yields were 0.26, 0.22 and 0.24. The corresponding genomic heritability estimates were 0.15, 0.11 and 0.13. Genome-wide association analysis identified 1 gene in chromosome 10, 4 genes in chromosome 16 and 5 genes in chromosome 18 associated with MR_{FY}. These genes are involved in metabolic pathways related with lipid and carbohydrate metabolism. Eight genes in chromosome 1 and 2 genes in chromosome 7 were associated with MR_{PY}. Some of these genes are involved in metabolic pathways related with protein metabolism. This study provided some insights into the genetics and genomics of MR for milk production in grazing dairy cows.

Introduction

In dairy cows, milk yield increases towards the fourth lactation, but there is a great variation between herds and between individuals. Maturity rate (MR), defined as the ratio of total lactation yield of the trait at first and third lactations, has been given little consideration in dairy cattle research. Hickman and Henderson (1965) reported heritability (h^2) for MR of milk yield (MY) maturity was only one-fourth to one-third as large as h^2 of first lactation production. Hargrove (1974) reported estimates of h^2 for rate of maturity for MY and fat yield (FY) of 0.089 and 0.100, respectively. de Jong (2014) reported that rate of maturity is considered as one of the traits of the life-time production index that predicts the bull's genetic merit for daughter life-time production in Netherlands. The objectives of this study were to estimate heritability and perform genome-wide association analysis for MR in dairy cows from two dairy farms under New Zealand grazing conditions.

Materials & methods

Animals and phenotypes. The total lactation yield records from 1,113 cows measured during the 2013 to 2020 production seasons at the Massey University Dairy 1 and Dairy 4 farms in Palmerston North, New Zealand were used for this study. The number of cows per breed and farm were 84 Holstein-Friesian (F), 91 Jersey (J) and 157 crossbreed F×J at Dairy 1, and 316 F, 6 J and 453 F×J crossbred at Dairy 4. Dairy 1 is managed as a once-a-day (OAD) milking system with cows predominantly grazed on fresh ryegrass (*Lolium perenne*) and white clover (*Trifolium repens*) at a low stocking rate (2.1 cows/ha). Dairy 4 is managed as a twice a day (TAD) milking throughout the season, with higher stocking rate (2.8 cows/ha) grazing on ryegrass/white clover. Cows in Dairy 4 receive high levels of supplementary feed throughout the year. MR values for MY, FY and protein yield (PY) were estimated for each cow.

Genotypes and quality control. From the 1,113 cows with phenotypes, a subset of 947 cows had genotypes. The cows were genotyped using Bovine Illumina 50K SNP-chips. Quality control was carried out using SNP & Variation Suite (SVS 8.8) and loci with a call rate ≤80% or minor allele frequency ≤0.05, and animals with a call rate ≤80% were excluded from the data set. A total of 43,425 SNPs were available for association analysis after quality control.

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_684, © N. Lopez-Villalobos *et al.* 2022

Estimation of variance components and heritability. Estimates of prior values for animal additive genetic (σ^2_a) and residual (σ^2_e) variances were obtained by fitting a Bayesian linear mixed model in the JWAS package (Cheng *et al.*, 2018). The model included fixed effects of contemporary group defined as year of lactation started, milking frequency, breed group (F, J, or F×J), regression of deviation of cows from median caving date of the farm as a covariate and random effects of animal additive genetic and residual. A prior value for the genomic marker variance (σ^2_m) was estimated following the method proposed by Fernando and Garrick (2013). The variance explained by SNPs was estimated using a single-step Bayes C linear mixed model in the JWAS package using the estimated prior values for σ^2_a, σ^2_s, and σ^2_e. Genomic heritability (m^2) was estimated as the ratio between the posterior mean of variances of the genomic breeding value (σ^2_m) for each iteration across all cows and the phenotypic variance (σ^2_p) of the trait and h^2 was estimated as ratio between the sum of σ^2_m and additional polygenic variance ($\sigma^2_g = \sigma^2_a - \sigma^2_m$) and the σ^2_p. The chain included 50,000 Markov chain Monte Carlo (MCMC) iterations and marker effects from every 50th iteration were used to estimate genomic breeding values. Individual marker effects were estimated with a probability of 0.997 that any SNP would have a zero effect.

Genome-wide association analysis. The estimated prior values for σ^2_a, σ^2_m, and σ^2_e were used in this analysis. The marker effects were collectively used to predict the genomic breeding values of cows in non-overlapping 1-Mb windows based on their physical map order. The window genomic breeding values were computed by multiplying the number of minor alleles for each marker within each 1-Mb window by their corresponding posterior means of marker effects in that MCMC iteration. The proportion of genetic variance explained (PVE) by each window in any particular iteration was obtained by dividing the variance of the breeding values for that specific window by the variance of genomic breeding value for the whole genome in that iteration. There were 2,676 unique SNP windows across the whole genome. The expected proportion of variance explained by each 1-Mb window was 3.74×10^{-2}% (1/2,676 SNPs) The 1-Mb windows that explained 0.37% (10× expected PVE) of genetic variance were considered to reach genome-wide significance level. Estimated associations were represented in Manhattan plots where PVE by each 1-Mb window was plotted against genomic locations of the windows using the qqman package (Turner, 2014). The gene oncology tool (Ashburner *et al.*, 2000) integrated in Ensembl 2020 (Yates *et al.*, 2019) was used to report the functional classification of the identified genes.

Results

Means and standard errors of MR_{MY}, MR_{FY}, and MR_{PY} were 0.71 (±0.01), 0.74 (±0.01) and 0.72 (±0.01) in OAD milking, and 0.82 (±0.03), 0.81 (±0.03) and 0.82 (±0.03) in TAD milking. Estimates of variance components and heritability for MR_{MY}, MR_{FY}, and MR_{PY} are presented in Table 1. Genome-wide association graphs for maturity rate of the different traits are presented in Figure 1. No significant 1-Mb SNP windows were detected for MR_{MY} (Figure 1a). Three 1-Mb SNP windows from chromosomes 10, 16 and 18 were significantly (PVE≥0.37%) associated with MR_{FY} (Figure 1b). The proportion of genetic variance explained by the most significant 1-Mb window was 0.43% and the windows that were significantly associated with MR_{FY} at the genome-wide threshold collectively explained 1.26% of the total genetic variance. Two 1-Mb SNP windows from chromosomes 1 and 7 were significantly (PVE≥0.37%) associated with MR_{PY} (Figure 1c). The proportion of genetic variance explained by the most significant 1-Mb window was 0.44% and the windows that were significantly associated with MR_{PY} at the genome-wide threshold collectively explained 0.83% of the total genetic variance. Genome-wide association analysis identified 1 gene in chromosome 10 (SRP19), 4 genes in chromosome 16 (HLX, DUSP10, DEGS1, CNIH3), and 5 genes in chromosome 18 (N4BP1, CNEP1R1, NKD1, NOD2 and NREP) associated with MR_{FY}; and 8 genes in chromosome 1 (FOXL2, PIK3CB, ARMC8, A4GNT, DZIP1L, CLDN18, NCK1 and STAG1) and 2 genes in chromosome 7 (PPIP5K2 and C7H5orf30) associated with MR_{PY}.

Table 1. Estimates of variance components[1] for maturity rate of milk (MR_{MY}), fat (MR_{FY}) and protein (MR_{PY}) yields in grazing dairy cows at Massey University Dairy 1 and Dairy 4.

Trait	σ^2_g	σ^2_m	σ^2_e	σ^2_p	m^2	h^2
MR_{MY}	0.008	0.010	0.051	0.069	0.15	0.26
MR_{FY}	0.008	0.008	0.058	0.074	0.11	0.22
MR_{PY}	0.008	0.009	0.053	0.070	0.13	0.24

[1] σ^2_g = additional polygenic variance, σ^2_m = additive genetic variance explained by the markers, σ^2_e = residual variance, σ^2_p = phenotypic variance ($\sigma^2_g + \sigma^2_m + \sigma^2_e$), m^2 = genomic heritability calculated as σ^2_m/σ^2_p, h^2 = heritability calculated as $h^2 = (\sigma^2_g + \sigma^2_m)/\sigma^2_p$.

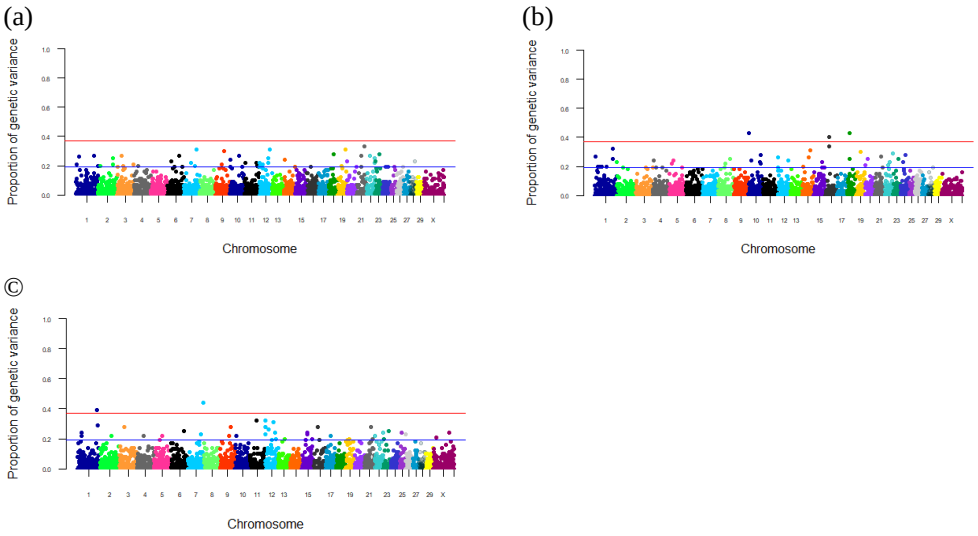


Figure 1. Manhattan plots showing genome-wide association for maturity rate of (a) milk, (b) fat and (c) protein yields. The red line shows the genome-wide significant level (PVE = 0.37%), and the blue line shows the suggestive association significant level (PVE = 0.19%).

Discussion

Cows in the OAD milking herd had lower MR than cows in TAD milking herd. This indicate that OAD milking is affecting the expression of milk production in the first lactation. The estimates of h^2 for MR of the three traits were greater than the 0.10 reported by Hickman and Henderson (1955) and Hargrove (1974) in USA Holstein dairy cows of the 1950's and 1960's. The candidate genes associated with MR_{FY} have previously been associated with fat deposition (N4BP1 gene, Martins *et al.*, 2020) and marbling score (NREP gene, Fonseca *et al.*, 2020) in Nellore cattle) and body weight and weaning weight (SRP19 gene, Grigoletto *et al.*, 2019) in Montana Tropical beef cattle. The CNEP1R1 and DEGS1 genes are associated with triglyceride biosynthesis (GO:0010867) and fat metabolism (GO:0006629), respectively. The HLX (GO:0030154) and DUSP10 (GO:0090335) genes have functions on cell differentiation whereas the former gene is also related to liver development (GO:0001889). Liver secretes bile which enables the metabolism of fat. Gene NKD1 is known to be related with FY (Cai *et al.*, 2018) while NOD2 (Wang *et al.*, 2015) and CNIH3 (Cai *et al.*, 2018) are associated with MY. Similarly, some of the candidate genes associated with MR_{PY} have been annotated in functions related to glycoprotein biosynthetic process (A4GNT, NCK1, FOXL2, STAG1 genes, GO:0009101, GO:0006417, GO:0000122, GO:0045944, respectively), protein catabolism (ARMC8,

GO:0043161) and transport of proteins (DZIP1L, CLDN18, C7H5orf30 genes, GO:0032880, GO:1900181, GO:0015031, respectively). The PPIP5K2 (GO:0016310) and PIK3CB (GO:0006468) are important genes in phosphorylation of synthesized proteins. The maturity rate is an important trait to acquire an insight into the life-time productivity of cows therefore it is important to understand the genetic and genomic control of this trait. Further studies based on a larger dataset would be beneficial to confirm these results.

References

Ashburner M., Ball C.A., Blake J.A., Botstein D., Butler H. *et al.* (2000). Nat. Genet. 25: 25-29. https://doi.org/10.1038/75556

Cai Z., Guldbrandtsen B., Lund M.S. and Sahana G. (2019). BMC Genet. 20, 15 https://doi.org/10.1186/s12863-019-0717-0

Cheng H., Fernando R.L., and Garrick D.J. (2018). Proc. of the 11th WCGALP, Auckland, New Zealand.

de Jong G. (2014). Interbull Bull. 48: 54-57. https://journal.interbull.org/index.php/ib/article/view/1345.

Fernando R.L., and Garrick D.J. (2013). Bayesian methods applied to GWAS. In: Gondro C., van der Werf J.H.J. and Hayes B. (eds). Genome-Wide Association Studies and Genomic Prediction. Springer, Berlin, Germany, pp. 237-274.

Fonseca L.F.S., dos Santos Silva D.B., Gimenez D.F.J., Baldi F., Ferro J.A. *et al.* (2020). Genomics 112:873-879. https://doi.org/10.1016/j.ygeno.2019.06.001.

Grigoletto L., Brito L.F., Mattos E.C., Eler J.P., Bussiman F.O. *et al.* (2019). Livest. Sci. 229:64-76. https://doi.org/10.1016/j.livsci.2019.09.013.

Hargrove G.L (1974). J. Dairy Sci. 57:328-331. https://doi.org/10.3168/jds.S0022-0302(74)84886-1.

Hickman C.G. and Henderson C.R. (1955). J. Dairy Sci. 38:883-890. https://doi.org/10.3168/jds.S0022-0302(55)95053-0.

Martins R., Machado P.C., Pinto L.F.B., Silva M.R., Schenkel F.S. *et al.* 2021. J. Anim. Breed. Genet. 138:360-378. https://doi.org/10.1111/jbg.12525.

SNP & Variation Suite ™ (Version 8.x) [Software]. Bozeman, MT: Golden Helix, Inc. http://www.goldenhelix.com.

Turner S.D. (2014). J. Open Source Softw. 3(25): 731. https://doi.org/10.21105/joss.00731

Wang W., Cheng L., Yi J., Gan J., Tang H., Fu, M.Z. *et al.* (2015). Genet. Mol. Res. 14:3570-3578. https://doi.org/10.4238/2015.April.17.6.

Yates, A.D., Achuthan, P., Akanni, W., Allen, J., Allen, J. *et al.* (2019). Nucleic Acids Res. 48: D682-D88. https://doi.org/10.1093/nar/gkz966.

685. Selection, validation, and utilization of mitogenome SNP array information in cattle breeding

V. Brajkovic[1*], L. Bradic[1], K. Turkalj[1], D. Novosel[2], S. Ristov[3], P. Ajmone-Marsan[4], L. Colli[4], V. Cubric-Curik[1], J. Sölkner[5] and I. Curik[1]

[1]University of Zagreb, Faculty of Agriculture, Svetosimunska 25, 10000 Zagreb, Croatia; [2]Croatian Veterinary Institute, Savska cesta 143, 10000 Zagreb, Croatia; [3]Ruđer Bošković Institute, Bijenička 54, 10000 Zagreb, Croatia; [4]Università Cattolica del Sacro Cuore, Faculty of Agriculture, food and environmental sciences, via Emilia Parmense 84, 29122 Piacenza, Italy; [5]University of Natural Resources & Life Sciences Vienna, Gregor-Mendel-Strasse 33, 1180 Vienna, Austria; vbrajkovic@agr.hr

Abstract

High-density SNP arrays are available and routinely used for genetic improvement of livestock populations. At the same time, mitogenome SNPs have rarely been used. In this study, we empirically evaluated 331 mitogenome SNPs included in the GGP Bovine 100K SNP array developed in collaboration with NEOGEN Genomics (Lincoln, NE, USA). Here we present a pipeline for the potential use of mitogenome SNPs with some examples: (1) classification of haplogroups with emphasis on rare ones (I, P, Q, R and T5); (2) pedigree verification; (3) pedigree and haplotype imputation; (4) deleterious mutation detection; (5) mitogenomic and nuclear diversity; and (6) standard mitogenome classification and phylogenetics. In summary, 317 mitogenome SNPs have been successfully validated and contribute as additional information to autosomal markers in population genomic analyses. We believe that our study will help to promote the use of mitogenome information in cattle breeding and biodiversity management.

Introduction

Mitochondria play a central role in cellular energy production. These organelles contain their own maternally inherited mitogenome, which in most vertebrates including mammalian livestock encodes 37 genes and has a non-coding region that has been widely used in studies of genetic diversity and evolution. While high-density SNP arrays are available and routinely used for genetic improvement of livestock populations, mitogenome SNPs have rarely been used. Therefore, this study aims to increase the use of mitogenome information in animal breeding by selecting and verifying 331 mitochondrial SNPs for the new NEOGEN GGP Bovine 100K SNP Chip and developing a pipeline for practical analysis.

Materials & methods

Reference dataset. The idea to include SNP mitogenomes in the new GGP100K SNP chip arose from a meta-analysis of sequence variability of the bovine mitogenome (Cubric-Curik *et al.*, 2021), which in the meantime, since publication, already includes a newly expanded set of 1068 sequences and over 150 breeds.

SNP selection. The discriminatory set of mitogenome SNPs was selected based on the pairwise FST values (equal to 1) between all haplogroups among them (I, P, Q, R, T1, T2, T3, T5) (Cubric-Curik *et al.*, 2021). F_{ST} values were calculated using DnaSP v6 software (Rozas *et al.*, 2017) and Arlequin v3.5 (Exoffier *et al.*, 2015), while SNP primers 150 bp in length were prepared for the GGP Bovine 100K SNP chip. Core SNP selection included 100 SNPs: (1) 70 SNPs for haplogroup classification; (2) 20 SNPs for more detailed haplogroup classification within the T3 haplogroup, as this is the most common one in commercial cattle breeds; and (3) 10 SNPs for LHON (leber hereditary optic neuropathy) disease. Additional SNP selection included 231 SNPs: (1) 199 SNPs for haplogroup classification; and (2) 32 SNPs for MELAS (mitochondrial encephalopathy, lactic acidosis, and stroke-like episodes) disease.

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_685, © V. Brajkovic *et al.* 2022

SNP array. From the study by Cubric-Curik *et al.* (2021), 129 samples (breeds: Austrian n=9, Greek n=5, Croatian n=4, Slovenian n=1) previously sequenced using NGS technology and an additional 135 new samples (breeds: Greek n=5, Croatian n=4, Austrian n=1) were selected for genotyping.

Results & discussion

Validation. 129 samples (sequenced, genotyped) were for 'full' validation, while the new 135 samples (genotyped) were for 'blind' functionality of mitogenome SNPs. 317 of 331 SNPs passed the sequence and genotype validation comparison. This is confirmed by the haplogroup classification of the new samples, which can be seen as grey-coloured haplotypes within the known haplogroups in the median-joining (MJ) network constructed with PopART (Leigh and Bryant) (Figure 1). It is important to note that all haplogroups except 'I' were successfully validated because we had no samples of that haplogroup. Due to the simplicity of the MJ network representation, SNP positions with gaps in ancient DNA (aDNA) samples were not considered in the construction of the haplotypes. It is worth noting that all aDNA samples were in or close to the core of the haplogroups.

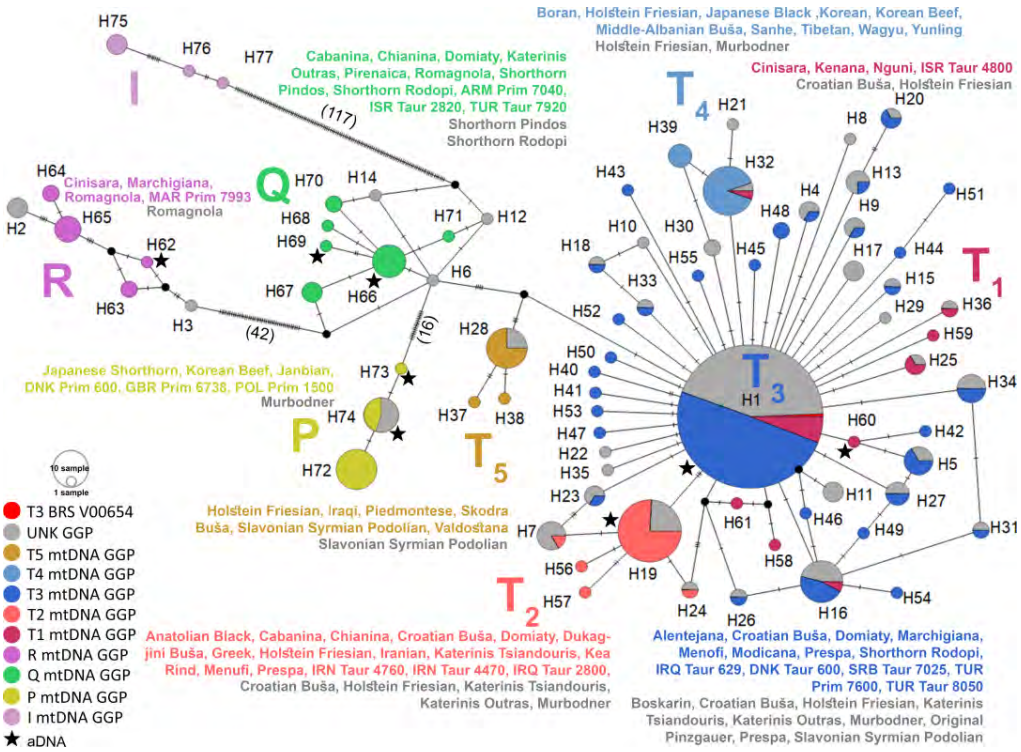


Figure 1. The Median-Joining network for the 181 mitogenome SNPs shows the phylogenetic positions of 77 haplotypes found in 383 samples of *Bos taurus*, *Bos primigenius*, and *Bos indicus* (129 samples sequenced and genotyped; 135 samples genotyped; additional 119 samples from the Cubric-Curik *et al.* (2021) study for better classification). For aDNA samples the number within the name indicates the year before present.

Mitogenome utilisation pipeline. Here we present a pipeline of potential use of mitogenome SNPs (Figure 2.)

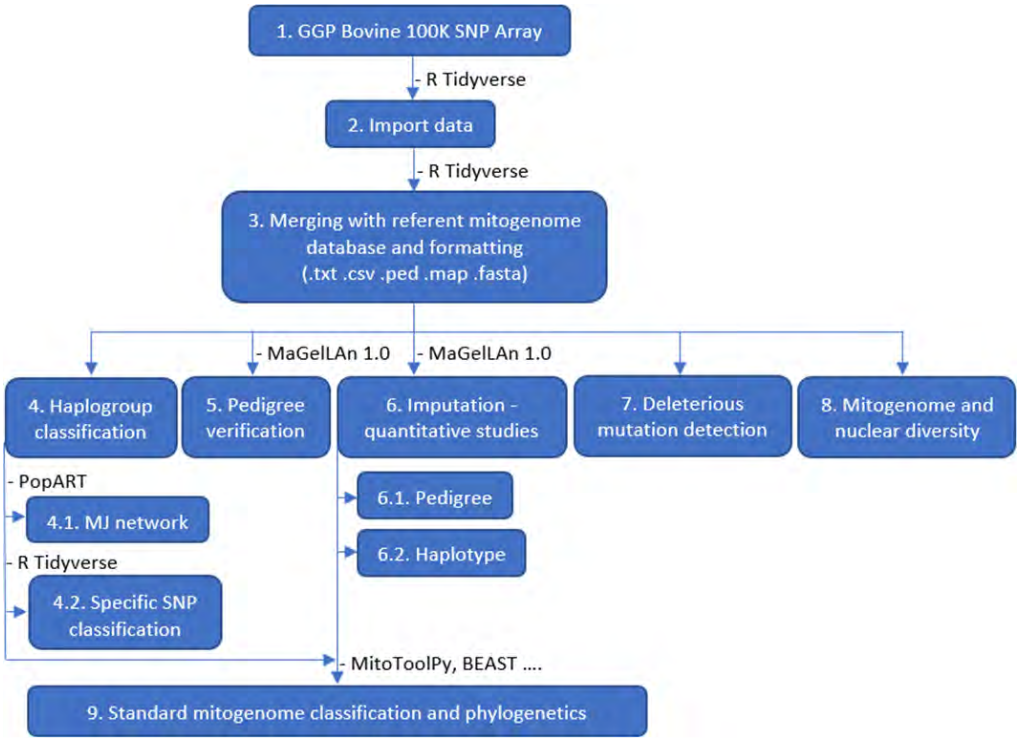


Figure 2. Utilization pipeline steps.

Step 1: Receipt of final report raw data of genotyped samples from NEOGEN Genomics.
Step 2: Import and visualise data with Rstudio (R Core Team, 2019; RStudio Team, 2020).
Step 3: Process the data using the tidyverse package collection (Wickham *et al.*, 2019), merge with the mitogenome reference database, and format according to the type of analysis.
Step 4: (1) Possible classification of haplogroups as shown in Figure 1. (2) specific SNP classification using the tidyverse package collection (Wickham *et al.*, 2019).
Step 5: Verification of complex pedigree errors using mitogenome haplotypes (maternally inherited), which has been well validated on a long and complex Lipizzaner horse pedigree using the computational approach of MaGelLAn 1.0 software (Čačić *et al.*, 2014; Ristov *et al.*, 2016).
Step 6: (1) Since the mitogenome is maternally inherited, MaGelLAn 1.0 makes it possible to quickly identify maternal lineages and assign the corresponding mitogenome sequences to all individuals in the pedigree, so that the phenotypic information can be used as input to any standard software for quantitative genetic (association) analyses (Ristov *et al.*, 2016). Such examples of estimating proportions of phenotypic variances of different traits explained by mitogenome variation can be found in rabbits (Nguyen *et al.*, 2018) and Holstein cattle (Brajkovic *et al.*, 2018). (2) Possible imputation of SNPs to construct the whole mitogenome consensus sequence considering SNPs that were not genotyped. For more information on mitogenome imputation in humans, see Ishiya *et al.* (2019) and McInerney *et al.* (2021).

Step 7: Ability to identify mitochondrial genetic disorders associated with LHON and MELAS disease by screening 42 potentially deleterious mtDNA mutations in cattle. While mitochondrial diseases are well documented in humans, to prove the concept, the first deleterious mutation, T10432C (T10663C in humans) in the ND4L subunit, was found in the Slovenian breed Cika with exophthalmos of the right eye (Novosel *et al.*, 2019), with aetiology and pathogenesis confirmed based on the tissue examined (unpublished data).
Step 8: Various diversity analyses can be performed using autosomal SNP array data. Here, we propose mitogenome-level analysis to enrich for maternal diversity. For example, the F_{ST} value between breeds can be calculated and potentially used to detect upgrading of breeds.
Step 9: Based on the imputed sequences, we can proceed with standard mitogenome classification and phylogenetic analysis, for example, using BEAST and Mitotoolpy.

In summary, 317 mitogenome SNPs have been successfully validated in the new GGP100K SNP chip and contribute as additional information to autosomal markers in population genomic analyses. We believe that our study will help to promote the use of mitogenome information in cattle breeding and biodiversity management.

Acknowledgements

This study was supported by the ANAGRAMS-IP-2018-01-8708, ALGSEQ18-IP-2018-01-7317 and Sir je IN-K.K.01.1.1.04.0058 projects. We also acknowledge financial support from NEOGEN Genomics (Lincoln, NE, USA) and H. Hofeneder-Barclay, J. R. Ferretti, R. K. Nelson, and R. G. Tait Jr. for their efforts related to the inclusion of mitochondrial SNPs in the GGP Bovine 100K.

References

Brajkovic V., Ferenčaković M., Špehar M., Novosel D. *et al.* (2018) Proc. of the 69th EAAP, Dubrovnik, Croatia.

Cubric-Curik V., Novosel D., Brajkovic V., Rota Stabelli O. *et al.* (2021) Evol. Appl. 00:1-16. https://doi.org/10.1111/eva.13315

Čačić M., Cubric-Curik V., Ristov S. and Curik I. (2014) Livest. Sci. 169: 42-47. https://doi.org/10.1016/j.livsci.2014.09.009

Excoffier L. and Lischer, H. E. (2005) Evol. Bioinform 1:117693430500100003. https://doi.org/10.1111/j.1755-0998.2010.02847.x

Ishiya K., Mizuno F., Wang L. and Ueda S. (2019) Bioinform. Biol. Insights. 13:1177932219873884. https://doi.org/10.1177/1177932219873884

Leigh, J. W., and Bryant, D. (2015) Methods Ecol. Evol. 6(9):1110-1116. https://doi.org/10.1111/2041-210X.12410

McInerney T. W., Fulton-Howard B., Patterson C., Paliwal D. *et al.* (2021) BMC Bioinform. 22(1): 1-15. https://doi.org/10.1186/s12859-021-04337-8

Novosel D., Cubric-Curik V., Simcic M., Dovc P. and Curik I. (2019) J. Cent. Eur. Agric. 20(1):19-24. https://doi.org/10.5513/JCEA01/20.1.2413

Nguyen N. T., Brajkovic V., Cubric-Curik V., Ristov S., Veir Z. *et al.* (2018) World Rabbit Sci. *26*(4): 287-298. https://doi.org/10.4995/wrs.2018.7644

Ristov S., Brajkovic V., Cubric-Curik V., Michieli I. and Curik I. (2016) Genet. Sel. Evol. 48(1): 1-10. https://doi.org/10.1186/s12711-016-0242-9

Rozas J., Ferrer-Mata A., Sánchez-DelBarrio J. C., Guirao-Rico S. *et al.* (2017) Mol. Biol. Evol. 34(12):3299-3302. https://doi.org/10.1093/molbev/msx248

R Core Team (2019) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

RStudio Team (2020) RStudio: Integrated Development for R. RStudio, PBC, Boston, MA URL http://www.rstudio.com/.

Wickham H., Averick M., Bryan J., Chang W. *et al.* (2019) J. Open Source Softw. 4(43):1686. https://doi.org/10.21105/joss.01686

686. Genomewide transcriptome profiling of milk derived primary bovine mammary epithelial cells after pathogen challenge

T. Iso-Touru[1*], D. Fischer[1], M. Kyläniemi[1,2], J. Tabell[1], A. Virta[1] and J. Vilkki[1]

[1]Natural Resources Institute Finland (Luke), Myllytie 1, 31600 Jokioinen, Finland; [2]current address: Turku Bioscience Centre, University of Turku and Åbo Akademi University, Turku, Finland; terhi.iso-touru@luke.fi

Abstract

A method was developed to collect and cultivate primary bovine mammary epithelial cells (pbMECs) from milk samples of healthy lactating cows. Cell cultures from ten Nordic Red cows were used in a pathogen challenge study with two common pathogens, *Escherichia coli* and *Staphylococcus aureus*. Transcriptome-wide changes in gene expression were monitored at four time points after challenge. We present here preliminary results from two cows, showing that pathogen specific differences in gene expression can be detected in the pbMECs, reflecting differences in the immune responses triggered by each pathogen.

Introduction

Mastitis, inflammation of the mammary gland, is the main cause of early weaning in humans, a major animal welfare concern and the costliest disease in dairy cattle. Mastitis can be expressed as mild sub-clinical, acute clinical or chronic inflammation of the mammary gland; caused by various bacteria, which may activate the mammary immune system in different ways, thereby influencing the severity and outcome of the disease (Taponen *et al.* 2017). Understanding of the molecular mechanisms that activate and regulate this response would be central to the development of effective prevention, prophylactics and treatments.

Previous studies show that the pathogen specific immune response of the mammary epithelial cells (MEC), but not of professional immune cells (e.g. macrophages) correlates best with the pathogen specific immune response of the udder (Petzl *et al.* 2018). For studying MEC function, usually tissue samples obtained by biopsies from live or slaughtered animals are used. We developed an ethical noninvasive method suitable for analysing many animals that would be available for follow-up studies. Therefore, we studied the possibility to extract and cultivate primary bovine mammary epithelial cells (pbMEC) from milk samples.

Our aim is to study the innate immune response by genomewide transcriptome profiling in pbMECs to two common mastitis pathogens known to induce different types of mastitis. The pathogens used include the Gram-negative *Escherichia coli* usually causing a strong acute inflammation and the Gram-positive *Staphylococcus aureus* causing a milder, often sub-clinical inflammation. In this paper we describe preliminary results from a subset of the full experiment, providing evidence that pbMECs extracted from milk samples can be used to identify differentially expressed genes.

Materials & methods

Primary cell culture establishment. Milk samples from two healthy lactating cows (Nordic Red Cattle) were collected by standard milking methodology to sterile collection bottles. A protocol modified from Danowski *et al.* (2012) and Hillreiner *et al.* (2017) was used to extract pbMECs. The epithelial character of the cells was confirmed by immunocytochemistry.

Pathogen challenges with *Escherichia coli* and *Staphylococcus aureus*. Cells (~1×10^5 cells/well, Greiner bio-one, collagen I coated, 392-0029) were challenged *in vitro* with two common heat inactivated mastitis pathogens, *E. coli* and *S. aureus* strains extracted from Nordic Red cows having mastitis (S. Taponen, personal communication) using 10 as multiplicity of infection (MOI). Cells were harvested before and at four different time points (3h, 6h, 12h and 24h) post challenge.

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_686, © T. Iso-Touru *et al.* 2022

RNA-Seq. Total RNA was extracted from the harvested cells using Rneasy Plus Micro kit (Qiagen) following manufacturer's protocol. RNA quantity and quality were measured with BioAnalyzer, and 50 ng of total RNA was used for ribosomal RNA depletion using RiboCop V1.3 (Lexogen). Ribosomal depletion allows detection of functionally relevant coding as well as non-coding transcripts through removal of ribosomal RNAs (Choe *et al.* 2021).

CORALL Total RNA-Seq Library Prep Kit (Lexogen) was used for library generation. The quality of pooled libraries was first checked on an Illumina iSeq100 platform at the Natural Resources Institute Finland and paired-end sequencing of three biological/technical replicates per timepoints and treatments was performed on an Illumina NovaSeq platform at the Finnish Functional Genomics Centre.

All samples from above-described challenge plan were sequenced with three technical replicates resulting in total 78 samples. An in-house Snakemake RNA-Seq pipeline following common RNA-Seq practices was applied for quality checks (FastQC, multiQC), trimming (fastp), alignment (STAR) and quantification (featureCounts). The pipeline, including additional features, is available at https://www.github.com/fischuu/Pipeline-RNA-seq.

Differentially expressed genes. The data was aligned against the ARS-UCD1.2 bovine reference genome and the quantification was done with reference to the Ensembl annotation 100. Technical replicates were combined by summing the corresponding read counts per gene. Following this practice, a data matrix with raw count data for 27,607 genes and 26 samples was generated and then split pathogen-wise, using the same control samples resulting in two matrices with 18 samples, each with 10 control and 8 pathogen related samples. Next, genes were filtered based on the sample read counts and only those genes were kept that had at least for eight samples 0.5 counts per million (cpm). This way, 11,891 genes were kept for *E. coli* and 11,674 for *S. aureus*. The differential gene expression (DE) analysis was performed with edgeR, using TMM normalized counts and only DE genes with a fold-change of 1.15 (*E. coli*) respective 1.1 (*S. aureus*) were kept.

Results

Primary cell culture. We were able to extract, cultivate and cryo-preserve pbMECs derived from freshly milked milk. Cells were successfully used in pathogen challenges. The pbMECs extracted from milk samples were shown to have epithelial characters by immunocytochemistry (Figure 1).

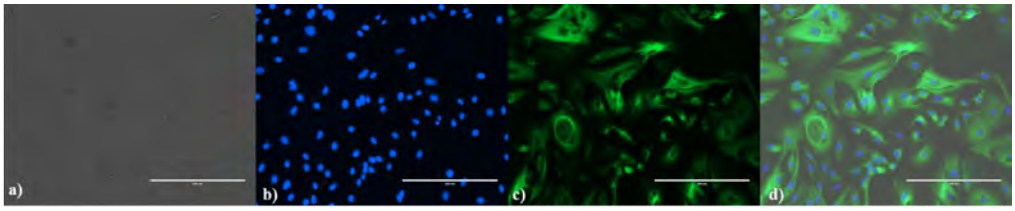


Figure 1. The (a) pbMECs under a light microscope, (b) cells stained with DAPI staining of pbMECs under fluorescence microscopy (c) cytokeratin staining using Monoclonal Anti-Cytokeratin, pan-FITC antibody (Sigma, F3418) d) merged image of light microscope, DAPI and cytokeratin staining images. The bars in the figures are 200 µm.

RNA-Seq analyses. The sequencing depth was between 41 and 87 mio 2×100 bp reads (mean: 62.8) per sample, after quality trimming and filtering the mean dropped only to 62.0, and the R1 (R2) read length to 86 bp (96 bp), indicating high-quality input data.

Prior to the DE analysis basic MDS plots of the raw data (data not shown) indicated a better separation between cases and controls for *E. coli* compared to *S. aureus*. Without fold-change filtering, 2,301 (1,849) downregulated and 2,332 (1,568) upregulated genes were detected after *E. coli* (*S. aureus*) challenge. However, after applying the fold-change filter 1,124 (868) up- and 1,219 (757) downregulated genes remained for *E. coli* (*S. aureus*).

For the top-30 differentially expressed genes, we applied a hierarchical clustering and visualised the results in Figure 2. The heatmaps show clear sample separations, as expected control samples are in a different branch compared to samples exposed to *E. coli*/*S. aureus*.

The specific top-30 up- or down-regulated DE genes differ between the two pathogens (Figure 2). The GO and KEGG analyses (data not shown) of the DE genes after each pathogen challenge show both same affected immune response related pathways (such as cytokine-cytokine receptor interaction, systemic lupus erythematosus) as well as specific affected pathways (such as oxidative phosphorylation and regulation of actin cytoskeleton for *E. coli* and ribosome, protein export and ECM-receptor interaction for *S. aureus*).

Discussion

We show here that pbMECs extracted from milk can be cultured and used for pathogen challenge experiments. The transcriptomic results indicate that we can detect pathogen specific changes in relevant pathways related to immune response, partly similar to those reported before by microarray-based analyses (Günther *et al.* 2017). The results including the full set of animals and timepoints will be published later.

Acknowledgements

This work is carried out within the Mastitis Challenge project funded by Academy of Finland (grant no. 317998). The CSC – IT Center for Science, Finland, is acknowledged for computational resources. We thank Finnish Functional Genomics Centre supported by University of Turku, Åbo Akademi University and Biocenter Finland for sequencing service.

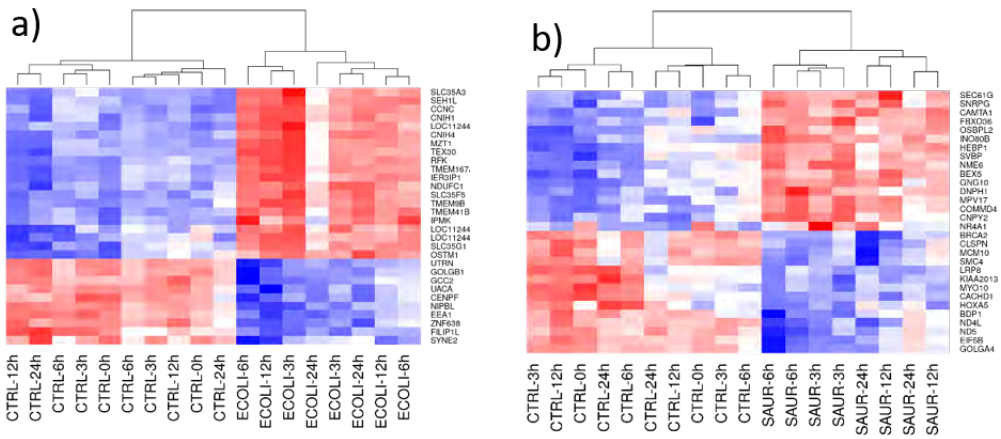


Figure 2. Heatmap visualisation of the clustering of (a) *Escherichia coli* and (b) *Staphylococcus aureus* exposed samples and corresponding control samples, based on the top 30 DE tested genes.

References

Choe D, Szubin R, Poudel S, Sastry A, Song Y, *et al.* (2021) PLoS Genet 17(9): e1009821. https://doi.org/10.1371/journal.pgen.1009821

Danowski K, Gross JJ, Meyer HH, Kliem H. (2013) J Anim Physiol Anim Nutr (Berl). 2013 Aug;97(4):647-55. https://doi.org/10.1111/j.1439-0396.2012.01305.x.

Hillreiner M, Müller NI, Koch HM, Schmautz C, Küster B, *et al.* (2017) *In vitro* Cell Dev Biol Anim. 2017 Sep;53(8):706-720. https://doi.org/10.1007/s11626-017-0169-7.

Günther J, Petzl W, Bauer I, Ponsuksili S, Zerbe H, *et al.* (2017) Sci Rep. 2017 Jul 6;7(1):4811. https://doi.org/10.1038/s41598-017-05107-4.

Petzl W, Zerbe H, Günther J, Seyfert HM, Hussen J, *et al.* Research in Veterinary Science,Volume 116, 2018,Pages 55-61,ISSN 0034-5288. https://doi.org/10.1016/j.rvsc.2017.12.012.

Taponen S, Liski E, Heikkilä AM, Pyörälä S. (2017) J Dairy Sci. 2017 Jan;100(1):493-503. https://doi.org/10.3168/jds.2016-11465.

687. Archetypal clustering reveals physiological mechanisms linking milk yield and fertility in dairy cattle

E. Ooi[1,2*], J.E. Pryce[2,3] and M.E. Goddard[1,2]

[1]Faculty of Veterinary and Agricultural Sciences, University of Melbourne, Parkville, VIC, 3010, Australia; [2]Agriculture Victoria Research, Agribio, Bundoora, VIC, 3083, Australia; [3]School of Applied Systems Biology, La Trobe University, Bundoora, VIC 3083, Australia; ecooi@student.unimelb.edu.au

Abstract

A negative genetic correlation between milk yield and fertility is now well-documented, however, the underlying biological causes are still uncertain. Our objective was to examine this problem from a genomic perspective by first using archetypal clustering to identify and group variants with similar patterns of effects across traits, and then by using pathway analysis to identify the biological processes underpinning variants with similar effects. Nine groups with distinct effects on production, fertility and conformation traits were identified. Cluster one had strong effects on fat, protein, and milk yields in contrasting directions with weak effects on fertility, and was characterized by genes affecting insulin, energy metabolism, and metabolic rate. Cluster three had strong effects on fertility but weak effects on production, characterized by genes associated with cell-cell adhesion and embryonic development. Initial results suggest that clusters are consistent with prior knowledge about the associated genes but may also suggest new areas for further research.

Introduction

Dairy cow fertility has declined significantly in dairy populations worldwide over the last 50 years. It has been well-established that there is a negative genetic correlation between milk production and fertility, and most dairy breeding programs now correct for this using genetic evaluations for this trait. However, although numerous hypotheses have been proposed, the exact physiological mechanisms linking the two traits are still unclear (Berry *et al.*, 2016). The correlation is also incomplete, indicating that some genetic variants affect milk production but have little effect on fertility and vice versa.

Our study aims to explore the physiological mechanisms linking milk production and fertility, using genes identified from a multi-trait genome wide association analysis (GWAS). We applied a combination of established and novel statistical methods to data collected from a large multi-breed population of Australian dairy cattle. Our pipeline proceeds from: (1) the identification of genomic variants in a multi-trait GWAS; to (2) the clustering of variants according to their pattern of effects on multiple traits of interest; and then (3) identification of genes through which these variants act; and (4) the prioritization of physiological mechanisms common to the genes within each cluster.

Materials & methods

Data preparation. Genotype and phenotype data for a total of 5,123 bulls and 29,081 cows from DataGene, Australia were used for this study. This data included a mix of Holstein-Friesians (4,061 bulls/22,899 cows) and Jerseys (1,062 bulls/6,174 cows).

Genotypes included a total of 46,771 sequence variants, which were selected from a total of 17,669,372 imputed variants selected by Xiang *et al.* (2021). These variants were selected based on evidence that they affect phenotype and that they are functional, as represented by their FAETH score. 247 additional variants were also included, as they were previously found to be informative for milk fat and protein percentage (Van Den Berg *et al.*, 2019).

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_687, © E. Ooi *et al.* 2022

Phenotype data included trait deviations and daughter trait deviations for cows and bulls, calculated using a model that corrects for fixed effects including herd, season, and year. Twelve traits were selected which are thought to have effects on production and/or fertility. These were protein yield, fat yield, protein percentage, fat percentage, milk yield, fertility, direct survival, stature, angularity, bone quality, udder texture and body condition score.

Single-trait GWAS. Each trait was analyzed one at a time in each sex with linear mixed models using GCTA (Yang *et al.*, 2011), with results combined using a weighted meta-analysis.

To remove variants in high LD, the snp_clumping function within R package was used. This function is analogous to the –clump function implemented in PLINK 1.9 but has been adapted for memory-efficient usage within the R environment. For our study, as we were most interested in the relationship between milk production and fertility traits, we used fertility t-values as our ranking statistic. This reduced the starting set of 46,771 variants to 15,220 variants.

Archetype-based clustering. We then clustered the sequence variants according to their pattern of effects on the 12 traits of interest. This was done by first ranking the variants in descending order according to the magnitude of their effect, where magnitude = sum of ^{t2} values across the 12 traits and t = t-value for the significance of each variant. The largest magnitude variant was declared the archetypal variant for cluster 1. Proceeding down the list, a new archetypal variant and thus cluster was declared whenever a variant had <0.8 cosine similarity with all previously declared archetypal variants. Using this method, we identified nine archetypal sequence variants with unique patterns of effects on the traits of interest. The remaining 15,211 variants were then assigned to the archetype with which they had the highest measure of cosine similarity, forming nine clusters. The direction of effects was then standardized across variants within a cluster.

Identification of genes. The identification of genes took account of the distance between the variant and the gene and whether or not the variant was an eQTL for that gene. We sampled one of the 10 closest genes within 500 kb either side of each variant, using a probability distribution that prioritized genes close to the variant, or multiplied the probability by 8.18 if the variant had been classified as an eQTL for that gene in expression studies (Chamberlain *et al.*, unpublished data), as they have been found to explain approximately 8.18 times the genetic variance compared to intergenic variants (Xiang *et al.*, 2019). The list of genes was then analyzed by g:Profiler, using GO:BP and KEGG as sources, and corrected for multiple testing with the g:SCS method. This was repeated 100 times for each cluster.

The top 10 GO:BP and KEGG terms for each iteration within a cluster were then aggregated by summing the inverse of the enrichment *P*-values obtained for each appearance of a biological term. A gene-weighted score was similarly calculated, with the added step of adjusting for annotation inequality using linear regression to determine the effect of the number of terms annotated for each gene.

Results

The pattern of effects for nine clusters of interest are shown in Figure 1. Solid blocks of color represent consistent effects in the same direction, with blue and red representing negative and positive effects respectively. For example, only clusters three and four have consistently negative effects on calving interval (representing better fertility).

For this brief paper, we chose to focus on clusters one and three as they have contrasting patterns of effects. Cluster one shows antagonistic effects on fat yield (blue), protein and milk yield (red), and weak effects

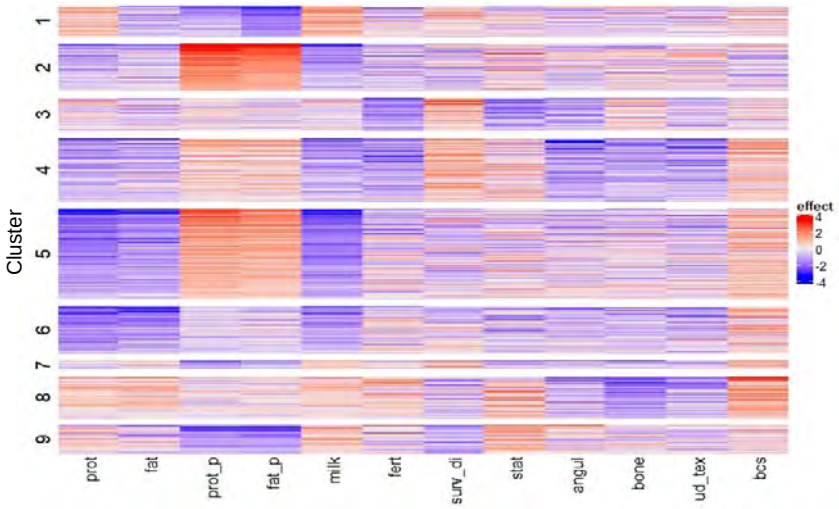


Figure 1. Heatmap showing the pattern of effects of 12,080 variants, sorted into 9 clusters across 12 traits of interest. Each horizontal line represents the effects of a single variant, with red representing positive effects and blue representing negative effects. As fertility is predominantly measured through calving interval, red represents poorer fertility.

on fertility, while cluster three shows strong effects on fertility (blue) but weak effects on production. The top genes identified for each cluster and degree of evidence for their involvement in common biological processes are shown in Figure 2.

Discussion

The archetypal variant for cluster one was associated with DGAT1, a gene which is primarily known to dairy scientists as responsible for lipid biosynthesis. However, our analysis shows that the dominant results demonstrated broader physiological mechanisms than fat synthesis, including lipid and glucose metabolism, obesity and insulin resistance/sensitivity, and differentiation of adipocytes. This is consistent with medical studies, which have focused on DGAT1 inhibition as a novel treatment for obesity and type 2 diabetes (Tsuda *et al.*, 2014). Our results suggest that this aspect of DGAT1 may be underexplored in dairy cattle.

Glucose tolerance tests in dairy cattle, have shown some evidence for adipose-specific insulin resistance (increased lipolysis) and mammary-specific insulin sensitivity (increased use of glucose as an energy substrate) in the form of phosphorylated Akt (Zachut *et al.*, 2013; Ma *et al.*, 2019). Tissue-specific insulin resistance may help to explain the high milk yields and poor body conditions which are often characteristic of the modern high-producing dairy cow. Interestingly, although our results support the hypothesis that high-yielding cattle have differences in energy partitioning, the cluster exhibits weak effects on fertility.

Cluster three, in contrast, exhibits strong effects on fertility and weak effects on production. The archetypal variant for cluster three was associated with ZNF613. Unlike DGAT1, ZNF613 is poorly studied, with little idea as to its function, although it has been implicated in dairy cattle gestation length, fertility, and conformation traits and is highly expressed in the developing embryo, placenta, and brain. However, our result may be helpful in suggesting areas of study for this gene, with the top five cluster three genes linked via their involvement in cell-cell adhesion. This area is less well-studied than diabetes and insulin, but the process may influence fertility through a variety of potential pathways, including GnRH migration in embryonic development and/or Wnt/β-catenin signal regulation.

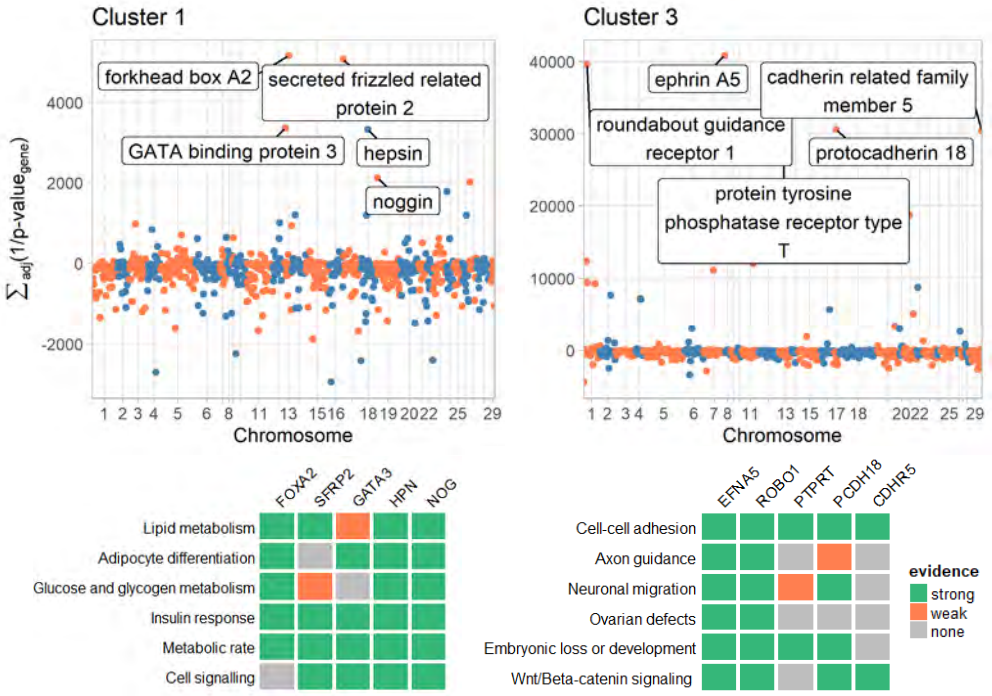


Figure 2. Top: The top five genes for cluster one were FOXA2, SFRP2, GATA3, HPN and NOG, while the top five genes for cluster three were EPHA5, ROBO1, PTPRT, PCDH18 and CDHR5. Bottom: Biological processes associated with clusters one and three. Green cells are supported by experimental evidence, orange have weak or incomplete evidence, grey have no supporting evidence.

References

Berry, D.P., N.C. Friggens, M. Lucy, and J.R. Roche. 2016. Annu Rev Anim Biosci 4:269–290. https://doi.org/10.1146/annurev-animal-021815-111406.

Fang, L., J. Jiang, B. Li, Y. Zhou, E. Freebern, P.M. Vanraden, *et al.* 2019. Commun Biol 2:100. https://doi.org/10.1038/s42003-019-0341-6.

Ma, Y.F., F. Batistel, T.L. Xu, L.Q. Han, R. Bucktrout, *et al.* 2019. J Dairy Sci 102:4264–4274. https://doi.org/10.3168/jds.2018-15451.

Tsuda, N., S. Kumadaki, C. Higashi, M. Ozawa, M. Shinozaki, *et al.* 2014. PLoS ONE 9:e112027. https://doi.org/10.1371/journal.pone.0112027.

Van Den Berg, I., B.J. Hayes, A.J. Chamberlain, and M.E. Goddard. 2019. BMC Genom 20:1–18. https://doi.org/10.1186/s12864-019-5656-7.

Xiang, R., I. van den Berg, I.M. MacLeod, B.J. Hayes, C.P. Prowse-Wilkins, *et al.* 2019. Proc Natl Acad Sci USA 116:19398–19408. https://doi.org/10.1073/pnas.1904159116.

Xiang, R., I.M. MacLeod, H.D. Daetwyler, G. de Jong, E. O'Connor, *et al.* 2021. Nat Commun 12:860. https://doi.org/10.1038/s41467-021-21001-0.

Yang, J., S.H. Lee, M.E. Goddard, and P.M. Visscher. 2011. Am J Hum Genet 88:76–82.

Zachut, M., H. Honig, S. Striem, Y. Zick, S. Boura-Halfon, *et al.* 2013. J Dairy Sci 96:5656–5669. https://doi.org/10.3168/jds.2012-6142.

Part 39

Bovine dairy – phenotype is king

688. Candidate genomic regions for forage dry matter intake in Swedish dairy cows using single-step genome-wide association

G.M. Tarekegn[1,3*], J. Karlsson[2], C. Kronqvist[2], K. Holtenius[2], S. Andonov[1], B. Berglund[1] and E. Strandberg[1]

[1]Swedish University of Agricultural Sciences, Department of Animal Breeding and Genetics, Uppsala, Sweden; [2]Swedish University of Agricultural Sciences, Department of Animal Nutrition and Management, Uppsala, Sweden; [3]Animal Breeding and Genomics, Scotland's Rural College, Roslin Institute Building, Easter Bush, Edinburgh EH125 9RG, United Kingdom; yafetgetinet@gmail.com

Abstract

The study was conducted to identify genomic regions associated with forage dry matter intake (DMI_{For}) in Swedish Red (SR) and Holstein (SH) cows using single-step genome-wide association methods. Phenotypic data were from 1,177 lactations of 356 SR and 219 SH cows, of which 255 were genotyped. A mixed linear animal model was fitted. The gene annotation revealed 84 variants associated with DMI_{For}, 76 of which were found for parity2+ in SH. The genomic variants were annotated and generated 33 candidate genes associated with DMI_{For} only in SH. The candidate genes that include MAPK10, SOX6, TCF7L2, CREB5, INPP4A, ATP13A3, SND1 and IRS1 were detected associated with DMI_{For}. The study revealed 10 QTL regions in 9 chromosomes. Our findings offer new insights into the genetic basis of DMI_{For} and provide a guide to breeding cows producing milk on forage.

Introduction

Productivity of dairy cows has increased due to enhanced health care, continuous selection and application of various breeding techniques and improved feeding strategies (Arefaine and Bertilsson, 2015). High-yielding cows are often fed high proportions of cereal grain and pulses that could be consumed directly by humans (FAO, 2014). Replacing these feeds with forage not suitable for human consumption would increase net food production and high-quality forage has the potential to sustain high milk production (Karlsson *et al.*, 2020). However, before introducing this system, we should be able to breed cows that can meet their nutritional requirements for high milk production on diets high in forage and low in concentrates. To do this, we need better knowledge about the genetic background of forage dry matter intake (DMI_{For}) though most studies on feed intake have been focused on total DMI. In this study, we implemented a single-step genome-wide association study (ssGWAS) approach to investigate candidate genomic regions associated with DMI_{For} in Swedish Red and Holstein dairy cows.

Materials & methods

Animals, forage intake and data editing. In total 1,177 lactations of 575 cows (748 lactations of 356 SR and 429 lactations of 219 SH cows) kept at Swedish Livestock Research Centre, Swedish University of Agricultural Sciences (SLU), Uppsala, Sweden, were used in the study. Observations were recorded between August 2013 and October 2018. The housing management and milking practices, the diet and feeding practices were described in previous studies (Karlsson *et al.*, 2020). Individual daily forage intake was recorded automatically and dry matter content was determined to calculate DMI_{For}. For the final analysis, the weekly average of daily observations of feed intake were used. The average DMI_{For} was 13.2 (SD 3.8) and 14.5 (4.3) for SR and SH, respectively.

Genotyping and quality control. A total of 255 cows (150 SR and 105 SH) were genotyped with low density panel (6,910 markers), and imputed to medium density (50K) by the Nordic Cattle Genetic Evaluation (NAV). A call rate of >95% was applied at both animal and SNP level, and MAF>5% was used as

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_688, © G.M. Tarekegn *et al.* 2022

the first quality control (QC) criterion for imputed data. After the QC assessment, the imputed data (50K) were reduced to 42,431 SNP markers. These markers were further reduced to 39,326 SNPs after checking for Mendelian conflicts and monomorphic and missing markers using blupf90 (Misztal *et al.*, 2002) in preparation for ssGWAS.

Variance components and ssGWAS. Variance components were estimated using the remlf90 package (Misztal *et al.*, 2002) separately for each breed, using a mixed linear animal model with season, parity (parity 1 and parity 2+), calving year and lactation week as fixed effects, and animal and permanent environmental effects as random effects, where the *H* matrix, combining pedigree and genomic relationships (Aguilar *et al.*, 2010) was used for the animal effect, resulting in genomic breeding values (GEBVs) for all animals. To obtain SNP effects, they were back-solved from the GEBVs as implemented by the program postGSf90 (version 1.46) following the procedure described in Wang *et al.* (2012). The SNP effects were evaluated using the SNP *P*-value method as suggested in Aguilar *et al.* (2019) applying the 20 adjacent SNPs as windows. postGSf90 software was employed to generate Manhattan plots. A more liberal threshold (i.e. 10^{-4}) was applied to test significant association of the variants (Teyssèdre *et al.* 2012).

Functional enrichment analysis of candidate regions and detection of QTL. The gene annotation was performed using the bovine genome assembly (ARS-UCD 1.2) as a reference genome embedded in Ensembl Genome Browser to identify candidate genes. To classify the genes according to their biological pathways using protein-protein interaction (PPI) networks and GO terms encoded by the candidate genes, the STRING Genomics 11.0 was employed to the annotated candidate genes. At least two SNPs found in the annotated genes in the 10 Mb window in a chromosome were used to define the QTL regions.

Results

A total of 7 and 77 differentiated markers were found associated with DMI_{For} in SR and SH, respectively, and most (76) were found in parity 2+ for SH (Figure 1). Only 5 and 2 markers were found for parity 1 and parity 2+, respectively for SR. The gene annotation analysis revealed 33 candidate genes, 5 of them were uncharacterized genes, that were generated from the 84 variants. The functional enrichment analysis in STRING demonstrated 30 KEGG pathways.

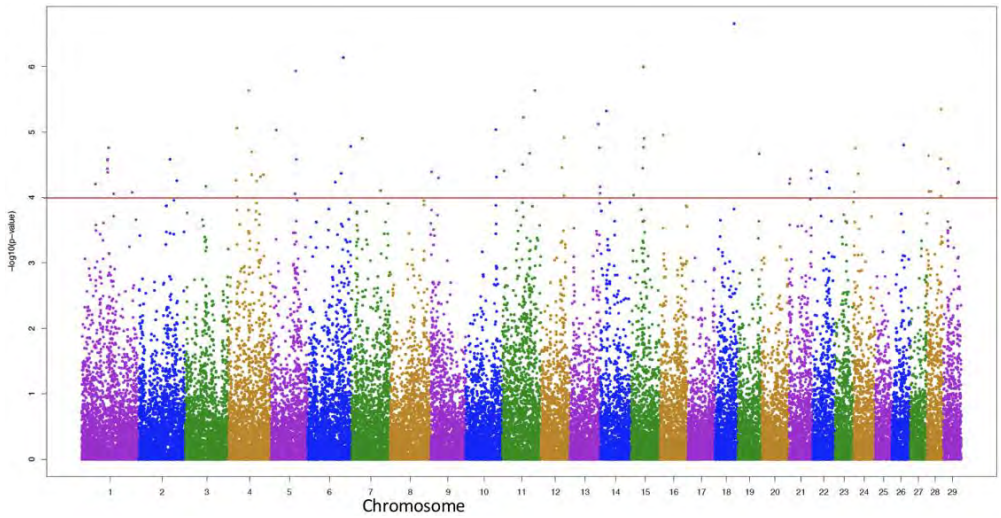


Figure 1. Manhattan plot generated for forage dry matter intake in parity 2+ in SH cows.

Detection of quantitative trait loci (QTL) regions. The identified candidate genes (18 genes, 3 of them are uncharacterized) were found embedded in 10 QTL regions in 9 BTAs, all in SH and in parity 2+. BTA4 had 2 QTL regions that contained 4 genes (AQAH, CREB5, AGMO, ENSBTAG00000051137). However, we detected only one QTL region per chromosome in the 8 BTAs. These include MAPK10 and CFAP299 in BTA6, INPP4A in BTA11, ENSBTAT00000050045 and ENSBTAG00000047383 in BTA12, KCNC1 and SOX6 in BTA15, WDR20 and MOK in BTA21, FHIT and CACNA1D in BTA22, GRID1 in BTA28, and MRPL49 and TCIRG1 in BTA29.

Discussion

In the current study, we identified variant associated with DMI_{For} in SR and SH, but the candidate genes obtained were detected only for SH indicating the breed difference on DMI_{For}. Those different genes detected were reported to be involved in various roles in metabolism, adipogenic conversion, lipogenesis, etc. For instance, the INPP4A gene associated with DMI_{For} for parity 1 and 2+ was reported associated to metabolism in cattle populations (Edea *et al.*, 2014). The CREB5 gene was reported associated to lipid and carbohydrate metabolism (Caldas, 2013). Over-expression of its family member, CREBL2, promotes adipogenesis and accelerates lipogenesis accompanied with increased GLUT (glucose transporter) 1 and GLUT4 via interaction with CREB (Ma *et al.*, 2011). The CREBL2 knockdown restrains adipogenic conversion and lipogenesis, whereas depletion of CREB could completely block the effects of overexpressed CREBL2 (Ma *et al.*, 2011). The SDN1 gene, detected in the present study, was reported in association with residual feed intake (RFI) in parity 2 in Holstein cows (Hardie *et al.*, 2017), whereas, the IRSI gene regulates milk fat metabolism and has roles in insulin activation, mediates and regulates carbohydrate and lipid metabolism (Yan *et al.*, 2019). According to Jiao *et al.* (2020), RS1 stimulated the level of Triglycerides (TAG) and increased the accumulation of lipid droplets in cow mammary epithelial cells (CMECs). The expression level of the IRS1 gene varies across lactation stages, it was highly expressed during the peak period of lactation, and was highly expressed in mammary tissue in dairy cows (Jiao *et al.*, 2020). The TCF7L2 gene was associated with fiber intake, is linked to pancreatic GLP-1 sensitivity (Loder *et al.*, 2008). The SOX6 gene has a role in carbohydrate metabolism pathway (Killeen *et al.*, 2014). It is a novel regulator of adipogenesis and modulate the transcriptional activity of carboxypeptidase M (CPM) to lower contents of C6:0 and C12:0, and increase contents of C18:1cis-9, C20:0 and total index as reported in Chinese Holstein cows, and is promising candidate gene for milk fatty acid (Shi *et al.*, 2020). Epidermal growth factor hydrolyzed by CPM enzyme (Skidgel *et al.*, 1996) involves in MAPK pathway to regulate fat metabolism (Taguchi *et al.*, 2018).

Overall, in the present study, we identified a range of genomic regions associated with DMI_{For}, but only in SH. We found only one gene, INPP4A, that showed pleiotropic effect, controlling DMI_{For} for parity 1 and parity 2+. Potential candidate genes to be further studied in SR and Holstein cows are CREB5, INPP4A, ATP13A3, SND1, TCF7L2, IRSI, SOX6 and MAPK10. Our findings offer new insights into the genetic basis of a previously little studied trait, DMI_{For}, and this new information may provide a guide to breeding strategies to forage efficient cows, however, validation of the SNPs effect with larger material could provide further evidence of the association of the candidate genes and their influence on the phenotypes.

References

Aguilar I., Misztal I., Tsuruta S., Legarra A., and Wang H. (2014). Pages 1-3 in Proc. 10th World Congr. Genet. Appl. Livest. Prod., Vancouver, BC, Canada.

Aguilar I., Legarra A., Cardoso F., Masuda Y., Lourenco D., *et al.* (2019). Genet Sel Evol. 51:28. https://doi.org/10.1186/s12711-019-0469-3

Arefaine H., Bertilsson J. (2015). Journal of Natural Sciences Research.5(17).

Caldas Y.R. (2013). PhD thesis, Universitat Autonoma de Barcelona.

Edea Z., Dadi H., Kim S.W., Park J.H., Shin G.H., *et al.* (2014). J. Anim. Breed. Genet. 131: 358-366. https://doi.org/10.1111/jbg.12083.

FAO. (2014). World mapping of animal feeding systems in the dairy sector. Rome, Italy.

Hardie L.C., VandeHaar M.J., Tempelman R.J., Weigel K.A., Armentano L.E., *et al.* (2017). J. Dairy Sci. 100:9061-9075. https://doi.org/10.3168/jds.2017-12604.

Jiao P., Yuan Y., Zhang M., Sun Y., Wei C., *et al.* (2020). Genes, 11:196. https://doi.org/10.3390/genes11020196.

Karlsson J., Lindberg M., Åkerlind M., and Holtenius K. (2020). J. Dairy Sci. 103:8922–8937. https://doi.org/10.3168/jds.2020-18204.

Killeen A.P., Morris D.G., Kenny D.A, Mullen M.P., Diskin M.G., *et al.* (2014). BMC Genomics, 15:234. https://doi.org/10.1186/1471-2164-15-234.

Loder M.K, Xavier S.G., McDonald A., Rutter G.A. (2008). Bioch.Soc.Trans 36:357-359. https://doi.org/10.2337/db08-1187.

Ma X., Zhang H., Yuan L, Jing H., Thacker P., *et al.* (2011). Biochem. J., 439: 27–38. https://doi.org/10.1042/BJ20101475.

Misztal I., Tsuruta S., Strabel T., Auvray B., Druet T., *et al.* (2002). Proc. 7th World Cong. Gen. Appl. Livest. Prod., Montpellier, France. Communication. 28:21-22.

Shu L., Matveyenko A.V., Kerr-Conte J., Cho J.H., McIntosh C.H., *et al.* (2009). Hum Mol Genet 18: 2388–2399. https://doi.org/10.1093/hmg/ddp178.

Skidgel R., McGwire G.B., and Li X.Y. (1996). Immunopharmacology. 32(1-3):48-52. https://doi.org/10.1016/0162-3109(96)00008-2.

Taguchi K., Bessho N., Hida M., Narimatsu H., Matsumoto T., *et al.* (2018) Naunyn-Schmiedeberg's Archives of Pharmacology 391:1075–83. https://doi.org/10.1007/s00210-018-1530-4.

Teyssèdre S., Dupuis M.C., Guérin G., Schibler L., Denoix J.M., *et al.* (2012). J. Anim. Sci. 90:45–53. https://doi.org/10.2527/jas.2011-4031.

Wang H., Misztal I., Aguilar I., Legarra A., Muir W.M. (2012). Genet Res (Camb). 94(2):73-83. https://doi.org/10.1017/S0016672312000274.

Yan X., Yang C., Lin G., Chen Y., Miao S., *et al.* (2019). J. Food Sci. 84:165–173. https://doi.org/10.1111/1750-3841.14415.

689. Stillbirth, dystocia and weaning rates in Australian dairy calves

M.M. Axford[1,2,3*], J.E. Pryce[1,2], M. Khansefid[1] and M Haile-Mariam[1]

[1]Agriculture Victoria, AgriBio, Bundoora, Victoria 3083, Australia; [2]School of Applied Systems Biology, La Trobe University, Bundoora, Victoria 3083, Australia; [3]DataGene Ltd, 5 Ring Road, Bundoora, Victoria 3083, Australia; maxford@datagene.com.au

Abstract

In a study spanning twenty years of calving records of Australian dairy cattle, stillbirths were observed in 7% of calvings, dystocia in 9% of calvings and 97% of heifers born alive survived to at least weaning age. Breed differences in stillbirth and weaning rate provide evidence that there may be a role for stillbirth and heifer livability EBVs to improve these metrics in Australia and complement the existing Calving Ease EBV for Holsteins. This study found that some Jersey animals experience dystocia which suggests a Calving Ease EBV for Jerseys may be helpful to reduce painful calving experiences and improve calf and cow survival and animal welfare.

Introduction

Successful calving and calf rearing systems deliver healthy cows to the milking herd, herd replacements as well as opportunities to increase income from livestock trading through generating surplus stock. Calves that are selected as herd replacements influence the rate of genetic gain and productive efficiency of the herd in subsequent generations. One way of defining successful calving and calf management is that cows deliver live calves in the absence of dystocia and calves survive to at least weaning age. This definition highlights three important metrics: stillbirths, dystocia and weaning rate. Stillborn calves are classically defined as calves that have died within 48h of birth (Cole *et al.*, 2007), but in practice, the time period is less defined due to recording practices. Dystocia is defined by Lombard *et al.* (2007) as delayed or difficult parturition. Weaning rate is the proportion of calves, excluding stillbirths, that survive at least to weaning. The aim of this paper was to quantify stillbirths, dystocia and weaning rate in Australian dairy calves born between 2000 and 2021.

Materials & methods

Calving records and termination dates are routinely collected from farmers that herd-record and were obtained from DataGene. Calvings recorded between January 2000 and April 2021 were matched with identities and termination dates of the resulting calves. Calving records include calf fate, calving ease, calf size, dam breed. Calf records include calf breed, date of birth, termination date. The presence of dystocia refers to calvings that were not recorded as 'ok', 'normal' or 'none'. 'Weaned' was calculated as a binary trait based on birth and termination dates. In Australia, weaning usually occurs by 12 weeks (Dairy Australia, 2020) so a threshold of 84 days (12 weeks) was selected. Calves with a termination date after 84 days were estimated to have survived to at least weaning. Calves were grouped according to the herd and year of birth to assess the weaning performance of herd/year groups. Data was summarised using RStudio (RStudio Team, 2021).

Results

In a dataset of more than 1.7M calvings, 7% of calves are reported as stillborn, with higher percentages reported in calves from Jersey dams (Table 1). Stillbirths have remained relatively constant in Holstein and Jersey females between 2016 and 2021. In males, stillbirth rates have risen 2 and 6% in Holstein and Jersey respectively, between 2006-2010 and 2016-2021.

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_689, © M.M. Axford *et al.* 2022

Table 1. Reported rates of stillbirth, dystocia and calf size in Australian dairy calves, 2000-2021 (n=1,738,300 calves).

Dam's breed	Sex	n calves	% stillborn	% dystocia	% big/huge	% small/ tiny	% stillborn calves with dystocia
Holstein	Female	836,400	5.1	7.5	5.5	8.1	34.1
	Male	696,779	5.0	12.9	11.8	7.7	39.6
Jersey	Female	107,690	6.9	2.7	2.8	7.5	19.8
	Male	97,831	9.7	3.6	3.9	7.3	18.8
All	All	1,738,300	6.6	9.1	7.8	7.9	34.6

Dystocia is reported in 9% of calvings (Table 1). Dystocia is more frequent in Holstein calvings compared to Jerseys. The highest level of dystocia occurs in calvings producing a male calf from a Holstein dam.

The next analysis shifts perspective from the cow's experience during calving to the calf's success to weaning. This subset of data contains only females that were born alive, given an identity at birth and had a recorded termination date which leads to a smaller dataset. The percentage of females that were weaned (i.e. survived to at least 84d in n=304,774 animals) averaged 97.4%. Jerseys were considerably lower than other breeds (94.5%), all other breeds were greater than the mean, with Red breed crosses being the highest and averaging 97.9% weaned, but this was only 0.3% more than Holsteins (97.6%).

Calves were grouped by herd and year with 5217 Holstein and 438 Jersey groups analysed (n=282,967 animals). Overall, Holstein herd/year groups had a weaning rate of 97.7% with a SD of 4.7, about 3% greater than Jersey groups (94.3%, SD 10.4). The percentage of Holstein and Jersey herds that achieved the 97% weaning target each year is plotted in Figure 1. Seventy-eight percent of Holstein groups achieved the target, 16% more groups than Jerseys. The most recent decade has generally poorer weaning rates than the previous across both breeds.

Discussion

Stillbirths are negatively impacting calf management on Australian dairy farms. The stillbirth rate of 7% is higher than 5.6% of Neupane *et al.* (2021) in US dairy cattle and 5.7% of Cuttance *et al.* (2017) in a group of New Zealand pastured based herds. The Holstein stillbirth rate (5%) is similar to 5.8% in reported by

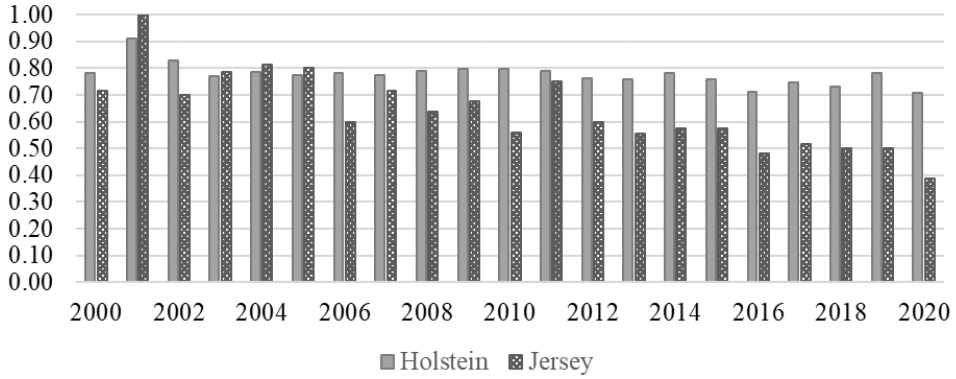


Figure 1. Proportion of Holstein and Jersey herd/year groups that achieve a target of 97% weaned 2000-2020 (n=282,967 calves, 5,655 herd/years).

Morek-kopec *et al.* (2021) Polish Holsteins. The higher stillbirth rate in Jerseys is consistent with results from Pryce *et al.* (2006). Differences in trait definition, time periods and breed proportions between studies could influence the small differences that are observed. Developing Stillbirth EBVs in Australia, such as those available in the USA (Cole *et al.*, 2007), will provide an opportunity to reduce the incidence of stillborn calves. In addition, strategies to reduce inbreeding could also contribute to lower stillbirths (Hinrichs and Thaller, 2011), especially at a time when inbreeding rates are increasing (Scott *et al.*, 2021).

Murray and Leslie (2013) suggest that dystocia is a highly painful experience for cow and calf and is known to contribute to poorer outcomes for both. The published rates of dystocia vary significantly depending on country, time-period, trait definition as well as animal factors such as parity and breed. At 9%, the level of dystocia reported in this study was higher than the USA estimate of 4.1% (USDA, 2018) and 1% in an early Australian study (McClintock *et al.*, 2003) but lower than an Irish estimate of 31% calving assistance (Mee *et al.*, 2011).

Stillborn calves are at greater risk of experiencing dystocia. More than one third of stillborn calves are dystocal, compared to just 5% of live Holstein heifer calves. Similarly, Haile-Mariam and Pryce (2018) observed a higher prevalence of stillborn calves amongst extremes i.e. tiny/small and big/huge Australian calves.

On average the weaning rate was more than 97% which is the target set for acceptable calf management by Dairy Australia's Rearing Health Calves resources (2020). Accounting for different time periods, the calf mortality to 12 weeks of 2.6% presented in this study is similar to the 4% to 18 months (Neupane *et al.*, 2021) in the USA. In a pasture-based New Zealand study which is more similar to Australian systems, post-natal mortality to 13 weeks was 4.1% (Cuttance *et al.*, 2017), slightly higher than this study.

There were some interesting breed differences in weaning rates. Compared to Jerseys, all other reported breed groups had a higher percentage of weaned calves. Norberg (2013) and Neupane *et al.* (2021) also reported higher losses in Jerseys compared to Holsteins in Denmark and USA respectively. Breed differences are also observed when weaning rate is assessed in herd/year groups. While the average weaning rate is more than the 97% target, there are herds and years that don't meet this target. About 16% more Holstein herd/year groups reach the target compared to Jerseys. While differences between breeds will provoke discussion, it is the within-breed variation that has the potential to be more useful to support farmers in making practice change that improves weaning rate through breeding. Genetic evaluations for heifer livability or survival have been released in the US (Neupane *et al.*, 2021) as well as other countries and are likely to improve weaning rates over time. This study shows that there is data on weaning rate that can be used for further study to estimate genetic variation and genetic evaluation of heifer livability in Australia.

Breed variation for stillbirth and weaning rate provide evidence that there may be a role for stillbirth and heifer livability EBVs, in Australia, to complement the existing Calving Ease EBV for Holsteins. This study found that some Jersey animals experience dystocia which suggests a Calving Ease EBV for Jerseys may be helpful to reduce painful calving experiences.

References

Cole J.B., Wiggans G.R., VanRaden P.M. 2007. J Dairy Sci 90, 2480-2488. https://doi.org/10.3168/jds.2006-435

Cuttance E.L., Mason W.A., McDermott J., Laven, R.A. McDougall, S., Phyn, C.V.C. 2017. J Dairy Sci 100, 8347-8357. https://doi.org/ 10.3168/jds.2017-12793

Dairy Australia, 2020. Rearing Healthy Calves. Dairy Australia, Melbourne, Vic. Available at: https://cdn-prod.dairyaustralia.com.au/-/media/project/dairy-australia-sites/national-home/resources/2020/10/07/rearing-healthy-calves-second-edition/rearing-healthy-calves-manual-october2020.pdf?rev=e86c53a7723e47c58dd481d1aa52ee62

Haile-Mariam M., Pryce J.E. 2018. Proc of 11th WCGALP, Auckland, NZ.

Hinrichs D., Thaller G. 2011. J Dairy Sci 94, 4726-4733. https://doi.org/10.3168/jds.2010-4100

Lombard J.E., Garry F.B., Tomlinson S.M., Garber L.P. 2007. J Dairy Sci 90, 1751-1760. https://doi.org/ 10.3168/jds.2006-295

McClintock S., Beard K.T., Gilmour A., Goddard, M.E. 2003. Proc of the 2003 Interbull meeting. Interbull, Rome, Italy.

Mee J.F., Berry D.P., Cromie A.R. 2011. Vet J 187, 189-194. https://doi.org/ 10.1016/j.tvjl.2009.11.018

Morek-kopeć M., Zarnecki A., Ptak E., Otwinowska-mindur, A. 2021. Animals (Basel) 11, 2792. https://doi.org/10.3390/ani11102792

Neupane M., Hutchison J.L., Van Tassell C.P., VanRaden P.M., 2021. J Dairy Sci 104, 8959-8965. https://doi.org/10.3168/jds.2020-19687

Norberg E., Pryce J.E., Pedersen J., 2013. J Dairy Sci 96, 4026-4030. https://doi.org/ 10.3168/jds.2012-5659

Pryce J.E., Harris B.L., Sim S., McPherson A.W., 2006. New Zealand Society of Animal Production, pp. 98-102.

RStudio Team, 2021. RStudio: Integrated Development Env. for R. PBC, Boston, MA.

Scott B.A., Haile-Mariam M., Cocks B.G., Pryce J.E., 2021. J Dairy Sci 104, 11832-11849. https://doi.org/ 10.3168/jds.2021-20326

USDA, 2018. Dairy 2014 Health and Management Practices on U.S. Dairy Operations, 2014. In: USEA (Ed.). USDA, Fort Collins, CO, p. 202. Available at: https://www.aphis.usda.gov/animal_health/nahms/dairy/downloads/dairy14/Dairy14_dr_PartIII.pdf

690. The relative importance of linear type traits in determining survival increases as Holstein-Friesian dairy cows age

M. Williams[1,2*], R.D. Sleator[2], C.P. Murphy[2], J. McCarthy[3] and D.P. Berry[1]

[1]Department of Animal Bioscience, Animal and Grassland Research and Innovation Centre, Teagasc, Moorepark, Fermoy, Co. Cork, Ireland; [2]Department of Biological Sciences, Munster Technological University, Bishopstown Campus, Co. Cork, Ireland; [3]Irish Cattle Breeding Federation, Bandon, Co. Cork, Ireland; maeve.williams@teagasc.ie

Abstract

Linear type traits were originally scored as a means of predicting functionality manifesting itself as greater cow longevity. Existing studies relating linear type traits to dairy cow longevity were generally restricted to relatively young cows, since dairy cows tended to be culled for infertility prior to reaching maturity. Improved reproductive performance, however, is contributing to an aging cow population. How linear type traits associate with survival in dairy cows should therefore be reassessed in light of this aging population. The objective of the present study was to use random regression models to quantify the genetic correlations between linear type traits and survival to the next parity, using survival data up to tenth parity. The genetic correlations between most of the linear type traits and survival were relatively consistent across parities; the exceptions being rear udder height, teat length, and udder depth, where the genetic correlations strengthened with advancing parity.

Introduction

Cow longevity has been a focus of dairy cow research since the 1950s (Rendel and Robertson, 1950) with some measure of longevity now included in most dairy cow breeding objectives (Cole and VanRaden, 2018; De Vries, 2020). Improved dairy cow longevity is associated with greater profitability, *via* a reduced requirement for replacement heifers, along with greater average herd yield, complemented with a greater opportunity for voluntary culling (van Pelt *et al.*, 2015). Nonetheless, selection for dairy cow longevity remains difficult, as a cow's true longevity is not known until she is actually culled (van Pelt *et al.*, 2015). Therefore, identifying early predictors of both phenotypic and genetic (merit for) survival are of potential importance. Although with the advent of genomic selection, having early indicators of longevity may become less critical in the future, as genotyped animals could receive accurate estimated breeding values for longevity at birth. Previous studies have genetically correlated linear type traits with measures of dairy cow longevity (Brotherstone *et al.*, 1998; Cruickshank *et al.*, 2002; Kern *et al.*, 2015), but these studies were generally limited to cows culled early on in their productive life. Nevertheless, linear type traits have frequently been included in multi-trait genetic evaluations to indirectly select for the genetic merit of dairy cow longevity (Miglior *et al.*, 2017). Despite this, no study to date has examined whether linear type traits are good (genetic) predictors of longevity in older cows. Therefore, the objective of the present study was to estimate the genetic correlations between linear type traits and survival from one parity to the next and, in doing so, evaluate if those genetic correlations change with advancing parity.

Materials & methods

Survival trait. Calving and pedigree information was available for 1,959,423 Irish Holstein-Friesian dairy cows calving in 40,437 herds. All cows were born between 1999 and 2008, inclusive, and calved for the first time in, or before, 2010. Only records from herds where <50% of the cows were registered with the breed society were retained; all calving records from the herdbook-registered cows in these herds were discarded, as were cows without a known sire. Records where first parity cows calved younger than 600 days of age were removed. The median age at calving was calculated for each parity, and cows calving more than 180

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_690, © M. Williams *et al.* 2022

days from the parity median was discarded. The entire records of an individual cow were discarded if any calving event (prior to culling) was set to missing after previous edits. Survival was defined for each parity (1st to 9th parity) based on whether or not a cow calved in the next lactation.

Linear type trait edits. Linear type traits data from 21 linear type traits from 203,136 primiparous Holstein-Friesian cows were available. Classification of cows was performed by a single professional classifier for a given herd, on a given date. Differences between classifiers were accounted for as described by Brotherstone (1994). Contemporary group for the linear type traits was defined as herd-date of inspection, and contemporary groups with <10 cows were discarded.

Subset selection for statistical analyses. Only survival records of cows with at least one linear classified paternal half-sib were retained. The retained survival records were assigned to contemporary groups based on proximity in calving dates within herd. A maximum duration of 60 days between calving events in the same contemporary group was allowed, and only contemporary groups with ≥10 records were retained. All survival data from herds where at least one cow had a linear type trait phenotype were retained. Survival information from an additional 852 herds were randomly selected. Similarly, all linear type trait information from herds where at least one cow had a survival phenotype were retained. Linear type trait data of cows in the linear type contemporary groups where at least one sire also had progeny with survival records were retained; all additionally retained linear scores, however, had to have been undertaken in or before 2010. After all edits, survival records were available for 52,447 Holstein-Friesian cows, with 152,894 lactations of information (1st to 9th parity). Linear type trait records were available from 52,122 Holstein-Friesian cows, of which 1,394 also had a phenotype for survival in the data for analysis.

Statistical analyses. The (co)variance components of survival were quantified using random regression animal models in ASReml (Gilmour *et al.*, 2009):

$$Surv_{ijklmno} = CG_j + Het_k + Rec_l + AGE_m + \sum_{n=1}^{3} b_n\, Parity^n + a_i * \sum_{o=1}^{3} b_o Parity^o + e_{ijklmno} \quad (1)$$

where $Surv_{ijklmno}$ was the survival observation; CG_j was the fixed effect of contemporary group j; Het_k was the fixed effect for heterosis class k (k=0 to 11); Rec_l was the fixed effect for recombination loss class l (l =0 to 6); Age_m was the fixed effect for age class relative to the parity median m (m=1 to 5); $\sum_{n=1}^{3} b_n\, Parity^n$ was the fixed effect nth order Legendre polynomial regression on parity number; $a_i * \sum_{o=1}^{3} b_o\, Parity^o$ was the random Legendre polynomial regression on parity number for animal$_i$ and $e_{ijklmno}$ was the residual term. No significant permanent environmental variance across lactation existed. Residual variances were estimated separately for each parity with no covariance assumed among parities. A series of bivariate animal random regression models were then used to estimate the covariances between survival at each parity and each linear type trait. The fixed and random effects included in the models were those used in the univariate analyses of survival and those used to estimate variance components of linear type traits described by Berry *et al.* (2004). The genetic (co)variances and genetic correlations between longevity in each parity, and each linear type trait, were estimated from the matrix of Legendre polynomial regression coefficients and the animal additive genetic covariance matrix estimated from the random regression models.

Results

A fixed cubic polynomial regression and a linear random regression were chosen as the most parsimonious orders for modelling longevity. Heritability estimates for survival per parity ranged from 0.02 (SE=0.004; 1st parity) to 0.05 (SE=0.01; 9th parity). Pairwise genetic correlations between survival among different parities varied from 0.42 (1st and 9th parity) to 1.00 (8th to 9th parity). The strength of the genetic correlations between survival in the different parities were inversely related to the interval between the compared

parities. Heritability estimates ranged from 0.22 to 0.39 for the body-related type traits, from 0.02 to 0.36 for the udder-related type traits, and from 0.08 to 0.26 for the feet and leg-related type traits.

The genetic correlations between survival and the individual linear type traits varied across parities for nine of the 21 linear type traits examined, but only three of the linear type traits became more strongly correlated with survival as cows aged; these three traits were rear udder height, teat length, and udder depth (Figure 1). With each successive parity post first-calving, shorter teats, a greater distance between the vulva and the udder (i.e. lower rear udder height scores), and a shallower udder depth, became increasingly more strongly genetically correlated with survival (Figure 1). While the genetic correlations between survival and locomotion were strongest in young parities, with each successive parity, after fourth parity, good locomotion scores became progressively more strongly associated with survival (Figure 1).

Discussion

The fact that the genetic correlations between survival and some linear type traits vary with age, suggest that different factors govern survival as the dairy cow ages; this was particularly true for some udder characteristics. In addition to the changes in genetic correlations between survival and individual linear type traits across ages, the range in the genetic correlations between survival in different parities further substantiates the hypothesis that survival may not be the same trait across the entire lifespan of a dairy cow. Previous studies have reported a large range in genetic correlations between individual linear type traits and different measures of longevity/survival in dairy cows; for example, the genetic correlations between longevity/survival traits and udder depth ranged from -0.07 for British Holstein-Friesian cows (Brotherstone *et al.*, 1998) to 0.31 for Brazilian Holstein cows (Kern *et al.*, 2015). The range in previously reported genetic correlations between longevity/survival and other individual linear type traits were similarly large (Brotherstone *et al.*, 1998; Cruickshank *et al.*, 2002; Kern *et al.*, 2015), potentially reflecting differences in scoring systems between countries/breed organisations, and the different definitions of survival/longevity.

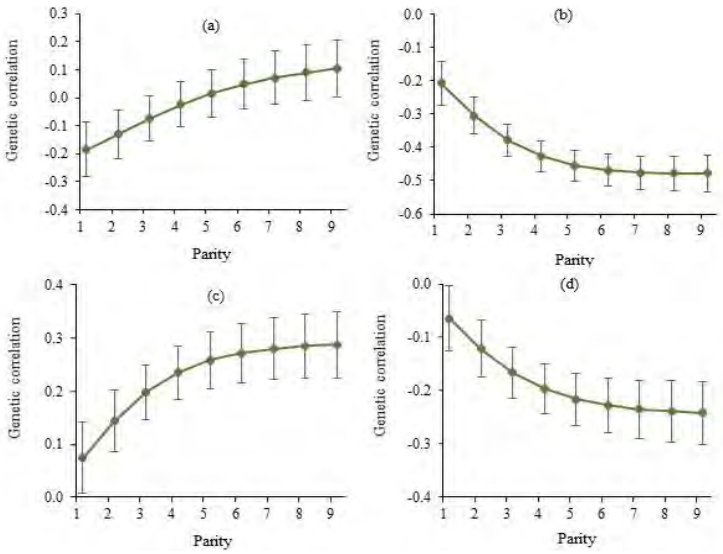


Figure 1. Genetic correlations between survival and (a) locomotion, (b) rear udder height, (c) udder depth, and (d) teat length in each parity. Standard error bars represent above and below the genetic correlation. Locomotion is scored from lame (1) to excellent (9); rear udder height is scored from low (1) to high (9); udder depth is scored from below hocks (1) to above hocks (9) and teat length is scored from short (1) to long (9).

Nonetheless, many of the genetic correlations reported in the present study were similar to those previously reported in other dairy cow populations, even though many of these studies were limited to younger cows.

As expected, better locomotion scores and shallower udders with shorter teats were favourably associated with survival, especially in older parities. Given linear type traits are frequently scored in first parity (Berry *et al.*, 2004), and are genetically correlated with survival in older parities, they may be suitable early predictors of survival. Additionally, the direction of the genetic correlations between survival and rear udder height, teat length, and udder depth did not change between parities; hence, selection for survival in older parities using these linear type traits should not hinder survival in younger parities.

References

Berry D.P., Buckley F., Dillon P., Evans R.D. and Veerkamp R.F. 2004. Irish Journal of Agricultural and Food Research. 43(2):161-176. https://www.jstor.org/stable/25562515.

Brotherstone S. 1994. Anim. Sci. 59(2):183-187. https://doi.org/10.1017/S0003356100007662.

Brotherstone, S., Veerkamp R.F. and Hill W.G. 1998. Anim. Sci. 67(3):405-411. https://doi.org/10.1017/S135772980003280X.

Cole J.B. and VanRaden P.M. 2018. J. Dairy Sci. 101(4):3686-3701. https://doi.org/10.3168/jds.2017-13335.

Cruickshank J., Weigel K.A., Dentine M.R. and Kirkpatrick B.W. 2002. J. Dairy Sci. 85(5):1307-1313. https://doi.org/10.3168/jds.S0022-0302(02)74195-7.

De Vries, A. 2020. J. Dairy Sci. 103(4):3838-3845. https://doi.org/10.3168/jds.2019-17361.

Gilmour, A. R., B. J. Gogel, B. R. Cuillis, and R. Thompson. 2009. VSN International Ltd, Hemel Hempstead.

Kern E.L., Cobuci J.A., Costa C.N., McManus C.M. and Braccini J. 2015. Scientia Agricola. 72(3):203-209. https://doi.org/10.1590/0103-9016-2014-0007.

Miglior F., Fleming A., Malchiodo F., Brito L.F., Martin P., Baes C.F. 2017. J. Dairy Sci. 100(12):10251:10271. https://doi.org/10.3168/jds.2017-12968.

Rendel J.M. and Robertson A. 1950. Empire Journal of Experimental Agriculture. 18:49-56.

Van Pelt M.L., Meuwissen T.H.E., De Jong G. and Veerkamp R.F. 2015. J. Dairy Sci. 98(6):4117-4130. https://doi.org/10.3168/jds.2014-9090.

691. Genetic parameters of milk composition and fertility traits in cows milked once- or twice-daily in New Zealand

J.M.D.R. Jayawardana[1*], N. Lopez-Villalobos[1], R.E. Hickson[1] and L.R. McNaughton[2]

[1]School of Agriculture and Environment, Massey University, Palmerston North 4410, New Zealand; [2]Livestock Improvement Corporation, Private Bag 3016, Hamilton 3240, New Zealand; d.jayawardana@massey.ac.nz

Abstract

The objective of this study was to estimate genetic and phenotypic correlations and heritabilities for milk composition and fertility traits in cows milked either once-a-day (OAD) or twice-a-day (TAD) across 21 spring-calving dairy herds in New Zealand. Fertility traits evaluated were submission in the first 3 weeks of mating (SR21), the interval from the start of mating to conception (SMCO), in-calf in the first 3 weeks of mating (PR21), and conception to the first service (PRFS). Heritability estimates of fertility traits were low and varied from 0.02 to 0.07 across the population. Phenotypic correlations between milk production and fertility traits were close to zero and most of the genetic correlations were antagonistic in both milking systems. Protein to fat ratio had positive genetic correlations with SR21, PR21 and PRFS in both milking frequencies. Estimates of genetic correlations of P:F ratio with PR21 and PRFS were relatively greater in OAD than TAD indicating that cows milked OAD had a higher P:F ratio and were more likely to get in-calf early in the breeding season than cows milked TAD.

Introduction

Genetic parameters of milk production and fertility traits have been estimated in seasonal dairy-producing systems (Grosshans *et al.*, 1997, Berry *et al.*, 2013). The main conclusion from cited studies is that genetic correlations between milk production and fertility traits are antagonistic. Since the late 1990s, some farmers have adopted OAD milking for the entire lactation in New Zealand (Davis *et al.* 1999), and for the production season of 2015-16, about 9% of the tested herds were milked OAD for the whole lactation (Edwards, 2018). Hemming *et al.* (2018) reported herds milked OAD for the entire lactation had better reproductive performance than those milked TAD. Genetic correlations between milk production and fertility traits have not been reported for the OAD population. Therefore, this study aimed to estimate genetic and phenotypic correlations between milk composition and fertility traits of cows milked OAD in New Zealand.

Materials & methods

Data. Herd test milk yields, calving, mating, and pregnancy diagnosis information of spring-calved dairy cows in production season 2017-18 were extracted from the animal database of Livestock Improvement Corporation (LIC). Selected herds had at least 50 cows per herd, herd tested four or more times per lactation and at least 80% of cows in a herd had early, aged pregnancy testing information to allow conception dates to be calculated. Herds were classified as OAD or TAD for entire lactation if more than 90% of the tested cows had the same milking frequency at each herd test-day and all the herd tests were classified as the same milking regimen. After the OAD herds had been identified, the geographically closest TAD milking herd was selected within 20 km from OAD using GPS visualizer (Schneider, 2012). The final dataset contained 4,629 milking cows from 10 OAD (2,008 cows) and 11 TAD (2,621 cows) milking herds with Holstein Friesian 816 (F), Jersey 1566 (J), and 2247 their crossbred (F×J).

Production traits. Milk production data included milk yield (MY), fat percentage (FP), protein percentage (PP), and lactose percentage (LP). Protein to fat (P:F) and protein to lactose (P:L) was calculated as the ratio between protein and fat percentages and protein and lactose percentages, respectively. Only animals with lactation lengths from 150-305 days were analysed.

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_691, © J.M.D.R. Jayawardana *et al.* 2022

Fertility traits. Four fertility traits were defined: submission for artificial insemination in the first 3 weeks of mating (SR21), interval from the start of mating to conception (SMCO), in-calf in the first 3 weeks of mating (PR21), and conception to the first service (PRFS). Pregnancy testing results where the foetal age was estimated and where cows were tested between 35-122 days (inclusive) from conception were used. Conception dates were calculated as the date of pregnancy diagnosis minus the estimated foetal age for cows with a pregnancy status of 'pregnant'. Non-pregnant cows (11.8%) were retained to estimate genetic parameters and penalized conception dates were calculated by adding 105 days (a 12-week mating period plus 21 days) to the herd start of mating date. The mating start date of each herd was sourced from the LIC database. The variables SR21, PR21 and PRFS were coded as 1 if the cow was mated or conceived in the first 3 weeks of mating, or conceived to the first service, and 0 otherwise. Cows in their first four parities were considered separately and cows in parity five and above were amalgamated into one group.

Statistical analysis. Heritability and genetic and phenotypic correlations were estimated using ASReml 4.1 (Gilmour *et al.*, 2015). A single-trait animal model was used for the estimation of heritability for each trait and bivariate animal models were used separately to estimate the genetic and phenotypic correlations. Models included the fixed effects of herd, parity, proportion of F, heterosis, and linear and quadratic effects of deviation of calving date from the herd median calving date and the random effect of animal.

Results

Cows milked TAD produced greater MY and higher LP than those cows milked in OAD (Table 1), but had poorer fertility outcomes. Fat and protein percentages were higher in cows milked OAD (Table 1). Compared with cows milked TAD, a higher proportion of cows milked OAD were mated (>2.5%) and conceived (>9.9%) within 3 weeks of the mating season and were pregnant (>7.4%) to their first AB.

Heritability of production and fertility traits varied from 0.22 to 0.72 and 0.02 to 0.05 in OAD, from 0.13 to 0.67 and 0.02 to 0.07 in TAD respectively (Table 2). In both milking systems, the genetic correlations of MY with FP, PP, and P:L were moderate and negative. Milk yield was moderately positively correlated with P:F in TAD. In both milking systems there were strong negative genetic and phenotypic correlations between FP and P:F, while correlations of FP with PP and P:L were positive and moderate to strong. Genetic and phenotypic correlations between PP and P:L were strongly positive but LP and P:L were moderately negatively correlated in both OAD and TAD herds.

Table 1. Summary statistics for milk production and fertility performance of cows milked once daily (OAD) and twice daily (TAD).

Traits[1]	OAD		TAD	
	Mean	SD	Mean	SD
MY (kg)	2,696.6	686.5	4,437.7	1,112.0
FP (%)	5.6	0.9	5.0	0.8
PP (%)	4.1	0.3	3.9	0.3
LP (%)	4.9	0.2	5.0	0.2
SMCO (days)	21.1	17.8	24.6	19.4
SR21 (%)	85.6	35.2	83.1	37.5
PR21 (%)	58.6	49.3	48.7	50.0
PRFS (%)	68.3	46.5	60.9	48.8

[1] MY = milk yield; FP = fat percentage; PP = protein percentage; LP = lactose percentage; SMCO = start of mating to conception; SR21 = proportion of cows that presented for mating within 21 days of the start of mating; PR21 = proportion of cows that conceived within 21 days of the start of mating; PRFS = proportion of cows that conceived to their first service.

Table 2. Heritabilities, genetic and phenotypic correlations of milk production and fertility traits for cows milked in once daily (OAD) and twice daily (TAD).[1]

	Traits[2]	MY	FP	PP	LP	P:F	P:L	SMCO	SR21	PR21	PRFS
OAD	MY	**0.36**	-0.30	-0.27	0.11	0.13	-0.28	-0.02	0.01	-0.04	0.03
	FP	-0.53	**0.64**	0.54	-0.11	-0.79	0.50	-0.05	0.03	0.03	0.01
	PP	-0.65	0.64	**0.72**	-0.12	0.03	0.89	-0.12	0.09	0.11	0.07
	LP	0.01	0.01	-0.13	**0.40**	0.01	-0.56	-0.04	0.01	0.01	0.04
	P:F	0.11	-0.77	-0.04	-0.11	**0.67**	0.03	-0.02	0.02	0.03	0.04
	P:L	-0.57	0.55	0.93	-0.49	0.01	**0.65**	-0.08	0.07	0.08	0.03
	SMCO	-0.03	0.18	-0.36	0.62	-0.45	-0.49	**0.04**	-0.25	-0.74	-0.55
	SR21	0.06	-0.51	-0.29	-0.19	0.46	-0.13	-0.36	**0.02**	0.38	0.21
	PR21	-0.39	-0.20	0.16	-0.59	0.38	0.33	-0.69	0.44	**0.05**	0.54
	PRFS	-0.16	-0.28	0.09	-0.44	0.44	0.22	-0.97	0.60	0.97	**0.04**
TAD	MY	**0.30**	-0.30	-0.29	0.02	0.18	-0.24	0.01	0.02	-0.03	0.01
	FP	-0.72	**0.65**	0.61	-0.02	-0.82	0.48	-0.03	0.03	0.04	0.01
	PP	-0.63	0.61	**0.67**	-0.05	-0.09	0.81	-0.09	0.07	0.08	0.05
	LP	-0.14	0.10	0.25	**0.21**	-0.03	-0.59	0.01	0.03	0.01	0.03
	P:F	0.59	-0.85	-0.33	-0.03	**0.40**	-0.05	-0.02	0.02	0.01	0.02
	P:L	-0.59	0.78	0.98	-0.36	-0.34	**0.29**	-0.07	0.04	0.06	0.02
	SMCO	0.24	-0.05	-0.30	-0.04	-0.23	-0.28	**0.07**	-0.24	-0.72	-0.58
	SR21	-0.04	-0.34	0.11	-0.27	0.70	0.29	-0.98	**0.04**	0.33	0.11
	PR21	-0.48	0.08	0.33	0.38	0.25	0.16	-0.80	0.96	**0.05**	0.59
	PRFS	-0.46	0.15	0.38	0.57	0.10	0.06	-0.96	0.29	0.99	**0.02**

[1] Genetic correlations are shown below the diagonal, phenotypic above, and heritabilities along in the diagonal. Standard errors for all phenotypic and genetic correlations were ≤0.03 and ≤0.53 in both milking frequencies respectively.

[1] MY = milk yield; FP = fat percentage; PP = protein percentage; LP = lactose percentage; P:F = protein to fat ratio; P:L = protein to lactose ratio; SMCO = start of mating to conception (days); SR21 = proportion of cows that presented for mating within 21 days of the start of mating; PR21 = proportion of cows that conceived within 21 days of the start of mating; PRFS = proportion of cows that conceived to their first service.

In general, phenotypic correlations between milk production and fertility traits were low and close to zero for both milking frequencies. Most genetic correlations between milk production and fertility traits were negative and low to moderate. Genetic correlations of MY with PR21 and PRFS were moderate and negative in TAD, which was relatively greater compared to OAD. In OAD milking, FP was moderate and negatively correlated with SR21, PR21 and PRFS. In both milking frequencies, genetic correlations of PP, P:F and P:L with PR21 and PRFS were positive and low to moderate whereas PP, P:F and P:L were moderate negatively correlated with SMCO. Genetic correlations with P:F and SR21 were positive in OAD and TAD, but relatively greater in TAD. Genetic correlations of LP with SMCO were positive and moderate in OAD, but close to zero in TAD. Genetic correlations of LP with PR21 and PRFS were positive and moderate in TAD, but this was moderately negatively correlated in OAD.

Discussion

Heritability estimates and phenotypic and genetic correlations between milk yield and compositional traits in cows milked OAD and TAD for this study were consistent with Lembeye *et al.* (2016) in New Zealand. Estimated heritabilities for fertility traits were low and similar to previous findings (Grosshans *et al.*, 1997; Berry *et al.*, 2013). The low heritability of fertility traits are a consequence of small genetic variation and relatively large residual variation of cows in both milking frequencies. Our results indicate that increases in

MY were associated with a decrease in submission and in-calf and lengthening SMCO. This antagonistic genetic relationship between production and fertility of cows implies that selecting only for cows with higher milk production leads to a genetic decline in cow fertility. For this reason, New Zealand and most other national genetic evaluation systems now include fertility in their selection indices.

Our findings show that cows with a higher PP, P:F and P:L ratio, are more likely to get pregnant early in the breeding season and to conceive to their first service in both milking populations. A positive genetic association of PP with fertility has been reported by Morton *et al.* (2018), highlighting better fertility in cows with higher milk PP. Though the genetic correlations of PP with PR21 and PRFS are higher in TAD compared to OAD milking, genetic correlations of P:F and P:L with fertility are comparatively greater in OAD than TAD milking cows. This indicates that herds milked OAD have better fertility and a higher milk solids percentage, so each kilogram of milk provides a greater return to the farmer. The reason for the negative genetic correlation between FP and fertility outcomes with OAD milking in this study has not yet been identified. Davis *et al.* (1999) reported that LP decreased in OAD milking cows due to changes in mammary permeability which enhanced efflux of lactose through mammary tight junctions. The negative association between LP and fertility in cows milked OAD should be further investigated. The findings of this study indicate that there is potential to increase the fertility performance of dairy cattle by identifying the causes of the association between milk protein concentration and fertility of cows.

References

Berry D.P., Kearney J.F., Twomey K. and Evans R.D. (2013) Irish J. Agric. Food Res. 52(1):1-16. http://www.jstor.org/stable/23631014

Davis S.R., Farr V.C., and Stelwagen K. (1999) Livest. Prod. Sci. 59(1):77-94. https://doi.org/10.1016/S0301-6226(98)00204-8

Edwards J. P. (2018) Anim. Prod. Sci. 59(3): 570-580. https://doi.org/10.1071/AN17484

Gilmour A.R., Gogel B.J., Cullis B.R. and Thompson R. (2015) ASReml User Guide Release 4.1.

Grosshans T., Xu Z.Z., Burton L.J., Johnson D.L., and Macmillan K.L. (1997) Livest. Prod. Sci 51:41-51. https://doi.org/10.1016/S0301-6226(97)00104-8

Hemming N., McNaughton L., and Couldrey C. (2018) Proc. N. Z. Soc. Anim. Prod., Lincoln, New Zealand.

Lembeye F., Lopez-Villalobos N., Burke J.L., and Davis S.R. (2016) Livest. Sci. 185: 142-147. https://doi.org/10.1016/j.livsci.2016.01.022

Morton J.M., Pryce J.E., and Haile-Mariam M. (2018) J. Dairy Sci. 101(6):5227-5239. https://doi.org/10.3168/jds.2017-13268

Schneider A. (2012) About GPS visualizer. Available at: https://www.gpsvisualizer.com

692. Identifying phenotypes to enhance stress resilience of dairy cows

T.M. Sullivan[1], A. Sharma[1], K. Lamers[1], C. White[1], B.A. Mallard[2], A. Cánovas[1] and N.A. Karrow[1*]

[1]Department of Animal Biosciences, Centre for Genetic Improvement of Livestock, University of Guelph, Guelph, ON, N1G 2W1, Canada; [2]Department of Pathobiology, Ontario Veterinary College, Guelph, ON, N1G 2W1, Canada; nkarrow@uoguelph.ca

Abstract

Dairy cattle routinely face a variety of different stressors such as intense heatwave events that can lead to more frequent bacterial infections. In the present study, Holstein heifers (n=20) were allocated to different treatment groups receiving saline, 100, 200 or 400 ng/kg of bacterial lipopolysaccharide (LPS) administered intramuscularly, and serum was collected to profile changes in circulatory stress biomarkers. All LPS-challenged animals demonstrated distinct cortisol responses as compared to the saline controls 2 hr post-LPS challenge. Cytokines TNFα and IL-10, were found to be induced between 2 and 4 hr. Finally, regulatory miR-1246, miR-223, miR-29 and miR-31 were also significantly induced by LPS challenge. The present study demonstrates that the bovine stress response to LPS is dynamic, and variability in the response warrants further investigation to better understand the genetic contribution to stress resilience.

Introduction

Dairy cattle routinely face a variety of different stressors; for instance, cattle are being subjected to intense heatwave events as a result of climate change and subsequently to more frequent bacterial infections (Lacetera, 2019). Higher bacterial load leads to reduced milk production and reproductive ability, thereby contributing to economic losses to dairy producers (Abdelhay Kaoud, 2019). To combat this, farmers have become reliant on antimicrobials that have contributed to the development of antimicrobial resistant pathogens, which are now a global health threat (Robinson *et al.*, 2016). It is imperative to have a thorough understanding of the bovine stress response triggered by bacterial stressors, like bacterial lipopolysaccharide (LPS), to identify genetic and epigenetic factors that regulate these physiological responses so that enhanced stress resilience of animals can be pursued through selective breeding. The aim of this study was to optimize an immune stress challenge using LPS and identify biomarkers that may be used to stress phenotype dairy cattle.

Materials & methods

Twenty Holstein heifers (7.4±1.9 months of age) were used to determine an intramuscular dose of LPS sufficient for eliciting a systemic stress response. Animals were randomly divided into four different treatment groups (n=5/treatment); each receiving 1 ml of either 100, 200 or 400 ng/kg of *E. coli* LPS (O111: B4, Sigma-Aldrich), or equal volume of sterile phosphate-buffered saline (control) injected into the trapezius muscle. Blood was collected hourly for 8 hr and again at 24 hr. Measured serum parameters included total cortisol (IMMULITE/ IMMULITE 1000 Cortisol chemiluminescent enzyme immunoassay, Siemens Medical Solutions Diagnostics, Los Angeles, CA, USA) and cytokine (MILLIPLEX® MAP Custom 13-Plex Bovine Magnetic Bead Panel, EMD Millipore Corp., Burlington, MA, USA) concentrations. Serum miRNA was extracted (Qiagen miRNeasy® Serum/Plasma Advanced Kit, Hilden, Germany) and analysed by PCR using a StepOne Plus (ABI) qPCR machine. The qPCR reaction was subjected to denaturation at 95 °C for 10 min, 40 PCR cycles of 95 °C for 15 sec, then 58 °C for 1 min of primer annealing and amplification. Primer sequences were derived from Sharma *et al.* (2020). The statistical analyses were performed using repeated measures ANOVA using PROC GLIMMIX procedures available in SAS (SAS/STAT®, SAS Institute Inc, NC, USA).

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_692, © T.M. Sullivan *et al.* 2022

Results

Dynamic changes in serum cortisol concentration occurred in response to LPS challenge (Figure 1). At 2 hr, serum cortisol increased significantly across all treatment groups, and remained significantly induced until 4 hr in the 200 and 400 ng/kg LPS treatments compared to the control (P<0.01).

Dynamic changes in IL-10 and TNF-α also occurred following LPS challenge. IL-10 was most significantly induced in heifers 2 hr post challenge with 400 ng/kg of LPS (P<0.0001) as compared to the saline control group (P<0.0001) and returned towards basal levels by 4 hr (Figure 2A). TNFα was most significantly induced at 2 hr for animals challenged with 200 and 400 ng/kg of LPS (P<0.01) as compared to the saline control animals (Figure 3A).

LPS-induced changes in serum miRNA are shown in Table 1. Of the six miRNA analysed, miR-223, miR-31, miR-29b and miR-1246 were found to be significantly expressed during the study period as compared to their baseline values (P<0.05).

Discussion

Activation of the innate immune system by LPS is triggered primarily by TLR4, which results in a signalling cascade that triggers the release of pro- and anti-inflammatory mediators, and activation of the HPA axis leading to the release of cortisol (Burdick Sanchez). As expected, cortisol was significantly induced in all of the treatment groups in this study with concentrations peaking between 2 and 4 hr. These findings are

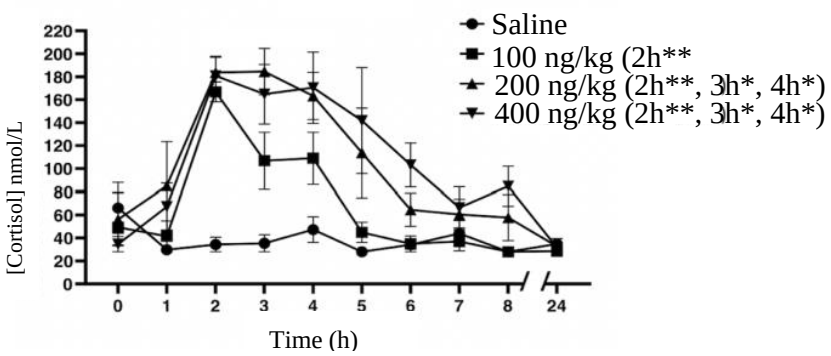


Figure 1. Bovine serum total cortisol response following intramuscular LPS immune challenge. ** indicates P<0.01, * indicates P<0.5.

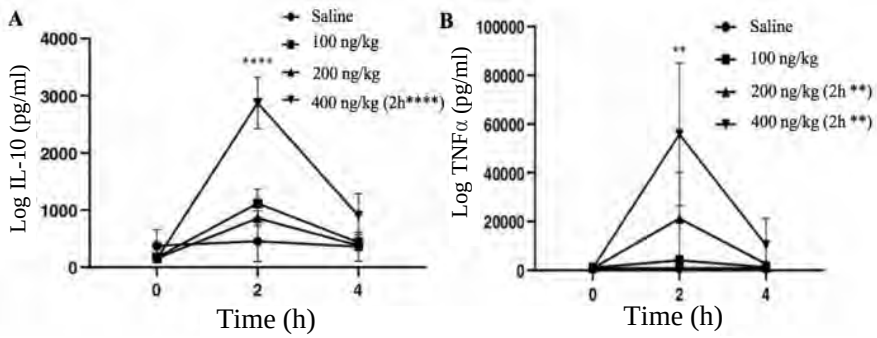


Figure 2. Bovine serum IL-10 (A) and TNF-α (B) response following intramuscular LPS challenge. **** indicates P<0.0001, and ** indicates P<0.01.

Table 1. Bovine serum miRNAs following intramuscular LPS challenge.

Fixed effects	miRNA-223		miRNA-29b		miRNA-31		miRNA-1246	
	Std. error	*P*-value	Std. error	*P*-value	Std. error	*P*-value	Std. error	*P*-value
100:2	1.32	0.458	8.48	0.284	1.18	0.922	42.29	0.755
100:4	1.37	<0.001***	8.48	0.848	1.43	<0.001***	43.83	0.937
100:6	1.35	<0.001***	8.01	0.981	1.15	0.822	42.29	0.910
200:2	1.32	0.978	8.19	0.840	1.86	0.470	42.29	0.835
200:4	1.32	0.441	7.71	0.879	1.24	<0.001***	42.29	0.005**
200:6	1.32	0.696	7.89	0.517	1.15	0.963	43.22	0.904
400:2	1.37	0.034*	8.48	0.035*	1.28	0.737	42.28	0.780
400:4	1.46	0.937	9.02	0.774	1.50	0.368	54.67	0.882
400:6	1.40	0.046*	8.48	0.734	1.33	0.664	44.73	0.736

[1] * indicates $P<0.05$, ** indicates $P<0.01$, *** indicates $P<0.001$, Fixed effects = interaction of LPS dose (ng/kg): hr post LPS challenge.

consistent with the findings of Boam *et al.* (2019), who assessed the effect of Holstein genotype on innate immunity and metabolism when heifers were challenged intravenously with 500 ng/kg of LPS; they report a peak cortisol response of approximately 220 nmol/l 3 hr following the *iv* LPS challenge. Our study was capable of eliciting a comparable cortisol response when heifers were challenged intramuscularly with 200 and 400 ng/kg of LPS, with averages 183 and 184 nmol/l, respectively. It has been demonstrated using sheep that the cortisol response to bacterial LPS is moderately heritable ($h^2 \cong 0.13$), and 16 single nucleotide polymorphisms (SNPs) on 13 different chromosomal regions have been implicated in the ovine stress response (Pant *et al.*, 2010). Future studies should aim to assess the heritability of the cortisol response in dairy cattle.

In the present study, LPS also induced a cytokine response as noted in the significant increase in serum TNF-α and IL-10. TNF-α has a diverse range of functions; it elicits inflammation, the acute-phase response, and triggers programmed cell death as a means of resisting microbial infections (Idriss and Naismith, 2000). In a beef steer calf study, researchers demonstrated a peak in TNF-α production 1 hr after intravenous administration of 250 ng/kg of LPS, which remained significantly induced until 2 hr (Burdick Sanchez *et al.*, 2020) these findings are in agreement with our study. IL-10 is a potent anti-inflammatory cytokine that is known to protect against LPS lethality by suppressing the production of inflammatory cytokines. In mice, IL-10 exhibited a biphasic response to intravenous LPS challenge, with an early peak at 1.5 hr and decrease to baseline at 6 hr (Barsig *et al.*, 1995). These authors found the IL-10 response was synchronous with the TNF-α response, and we also observed this. Peak IL-10 and TNF-α expression also coincided with peak cortisol concentrations.

MiRNAs play a key role in immune cell differentiation and modulation in response to pathogenic infections. Specifically, Hsieh *et al.* (2012) demonstrated a significant increase of 15 miRNAs in murine whole blood 2 hr following intraperitoneal LPS challenge that persisted for at least 6 hr. Despite their potential to be used as stress biomarkers, there is limited literature supporting the expression of bovine circulatory miRNA. Bovine miR-223 demonstrated the most prominent response to LPS as it was strongly upregulated by 100 and 400 ng/kg of LPS. MiR-223 helps to induce the proliferation and differentiation of neutrophils and is a proposed diagnostic biomarker of human sepsis (Liu *et al.*, 2020). MiR-29b has been reported to upregulate production of the inflammatory chemokine CXCL8 (IL-8) and may contribute to intestinal fibrosis by up-regulating fibroblast collagen production in Crohn's patients (Nijhuis *et al.*, 2017). Since LPS is known to target the lungs in ruminants (Warner *et al.*, 1988) the role of miR-29b in lung fibrosis in response to

bacterial and viral challenges could be more thoroughly investigated. LPS doses 100 and 200 ng/kg also upregulated bovine miR-31, which is in contrast to our previous ovine study (Naylor *et al.*, 2020). MiR-31 plays a role in maintaining homeostasis by acting as a negative regulator of the FOXP3 gene, which is a master regulator for the expansion of T regulatory (Treg) cells (Magni *et al.*, 2014). The upregulation of miR-31 in the present study implicates a potential reduction of Treg cell activity. Finally, animals challenged with 200 ng/kg of LPS demonstrated significant upregulation of miR-1246 and these findings are consistent in heat-stressed Holstein cows (Zheng *et al.*, 2014). The present study shows the sensitivity of cortisol, inflammatory cytokines and miRNA to low-dose intramuscular LPS challenge and indicates their potential to be used as biomarkers of bovine stress in response to a bacterial challenge.

References

Abdelhay Kaoud, H. (2019) Introductory Chapter: Bacterial Cattle Diseases - Economic Impact and Their Control. IntechOpen, https://doi.org/10.5772/intechopen.83635.

Barsig, J., S. Küsters, K. Vogt, H. -D Volk, G. Tiegs, *et al.* (1995). Lipopolysaccharide-induced interleukin-10 in mice: role of endogenous tumor necrosis factor-α. Eur J Immunol. https://doi.org/10.1002/eji.1830251027.

Boam, G.C., W.J. Weber, A. Benjamin, S. Kahl, G. Allen Bridges. (2019) Effect of bovine genotype on innate immune response of heifers to repeated lipopolysaccharide (LPS) administration. Vet Immunol Immunop. https://doi.org/10.1016/j.vetimm.2019.109914.

Burdick Sanchez, N.C., J.A. Carroll, P. Rand Broadway, T.S. Edrington, I. Yoon, *et al.* (2020). Some aspects of the acute phase immune response to a lipopolysaccharide (LPS) challenge are mitigated by supplementation with a Saccharomyces cerevisiae fermentation product in weaned beef calves. Transl Anim Sci. https://doi.org/10.1093/tas/txaa156.

Hsieh, C.H., C.S. Rau, J.C. Jeng, Y.C. Chen, T.H. Lu, *et al.*(2012). Whole blood-derived microRNA signatures in mice exposed to lipopolysaccharides. J Biomed Sci. https://doi.org/10.1186/1423-0127-19-69.

Idriss, H., and J. Naismith. (2000). TNF alpha and the TNF receptor superfamily: structure-function relationship(s). Microscop Res Techniq. 50:184–195.

Lacetera, N. (2019). Impact of climate change on animal health and welfare. Anim Front. https://doi.org/10.1093/af/vfy030.

Liu, D., Z. Wang, H. Wang, F. Ren, Y. Li, *et al.* (2020) The protective role of miR-223 in sepsis-induced mortality. Sci Rep-UK. https://doi.org/10.1038/s41598-020-74965-2.

Magni, S., G.B. Comani, L. Elli, S. Vanessi, E. Ballarini, *et al.* (2014) MIRNAs affect the expression of innate and adaptive immunity proteins in celiac disease. Am J Gastroenterol. https://doi.org/10.1038/ajg.2014.203.

Naylor, D., A. Sharma, Z. Li, G. Monteith, T. Sullivan, (2020) Short communication: Characterizing ovine serum stress biomarkers during endotoxemia. J Dairy Sci. https://doi.org/10.3168/jds.2019-17718.

Nijhuis, A., R. Curciarello, S. Mehta, R. Feakins, C.L. Bishop, *et al.* (2017). MCL-1 is modulated in Crohn's disease fibrosis by miR-29b via IL-6 and IL-8. Cell Tissue Res. https://doi.org/10.1007/s00441-017-2576-1.

Pant, S.D., F.S. Schenkel, C.P. Verschoor, Q. You, D.F. Kelton *et al.* (2010) A principal component regression based genome wide analysis approach reveals the presence of a novel QTL on BTA7 for MAP resistance in holstein cattle. Genomics. https://doi.org/10.1016/j.ygeno.2010.01.001.

Robinson, T.P., D.P. Bu, J. Carrique-Mas, E.M. Fèvre, M. Gilbert, *et al.* (2016). Antibiotic resistance is the quintessential One Health issue.T Roy Trop Med H. https://doi.org/10.1093/trstmh/trw048.

Sharma, A., U.K. Shandilya, T. Sullivan, D. Naylor, A. Canovas, *et al.* (2020). Identification of ovine serum mirnas following bacterial lipopolysaccharide challenge. Int J Mol Sci. https://doi.org/10.3390/ijms21217920.

Warner, A.E., M.M. DeCamp, R.M. Molina, and J.D. Brain. (1988) Pulmonary removal of circulating endotoxin results in acute lung injury in sheep. Lab Invest.

Zheng, Y., K.L. Chen, X.M. Zheng, H.X. Li, and G.L. Wang (2014) Identification and bioinformatics analysis of microRNAs associated with stress and immune response in serum of heat-stressed and normal Holstein cows. Cell Stress Chaperon. https://doi.org/10.1007/s12192-014-0521-8.

693. Estimation of genetic correlations between predicted energy balance and milk fatty acids of Holsteins in Japan

A. Nishiura[1*], O. Sasaki[1], M. Aihara[2], Y. Saito[1], R. Tatebayashi[1] and H. Takeda[1]

[1]Institute of Livestock and Grassland Science, National Agriculture and Food Research Organization, 2 Ikenodai Tsukuba, 3050901, Ibaraki, Japan; [2]Livestock Improvement Association of Japan, 11-17 Fuyuki Koto, 1350041, Tokyo, Japan; akinishi@affrc.go.jp

Abstract

It is important to optimize the energy balance of dairy cows in lactation to prevent any deterioration of health or fertility. We estimated the genetic correlations between predicted energy balance (PEB) and milk β-hydroxybutyrate (BHB) and three types of milk fatty acid (FA) composition (*de novo* FA on a milk basis (DNM), *de novo* FA on a fat basis (DNF) and preformed FA on a fat basis (PRF)) from Holsteins using datasets for test-day milk records. PEB was calculated using multiple regression equations. The genetic parameters of PEB, BHB, DNM, DNF and PRF were estimated using a random regression model. Genetic correlations in the early stage of lactation were estimated to be negative for PEB-BHB and PEB-PRF, and positive for PEB-DNF. Sufficient data was acquired to reveal the relationship between each milk components and energy balance, suggesting PEB to have potential for use as an index for improving energy balance.

Introduction

Increases in milk yields over the last few decades have resulted in postpartum energy deficits (Hüttman *et al.*, 2009). The energy balance (EB) of dairy cows tends to be negative, especially in early lactation, because energy intake via feed cannot match the energy demands of milk production (de Vries *et al.*, 1999). A negative energy balance causes a deterioration in fertility, metabolic disorders such as ketosis, and mastitis (Goff and Horst 1997; Collard *et al.*, 2000; Ingvartsen *et al.*, 2003). It is important to improve EB, but EB is difficult to measure, making it a challenge to obtain enough data to carry out a genetic evaluation (Coffey *et al.*, 2001). This prompted us to formulate an equation to predict EB from readily-acquired records of milk yield and its components. Records of milk β-hydroxybutyrate (BHB) and milk fatty acid (FA) composition are increasingly obtainable from routine herd tests in Japan. BHB and milk FA composition are thought to be related to energy status (Koeck *et al.*, 2014, Stoop *et al.*, 2009). It is necessary to elucidate the relationship between predicted EB (PEB) and BHB and milk FA composition. Our objective was to investigate the relationships between PEB and BHB and milk FA composition using herd test data.

Materials & methods

Data. The datasets we used consisted of test-day milk records for the first three lactations of Holstein cows in two regions: Hokkaido and the rest of Japan. The records were collected by the Livestock Improvement Association of Japan. All the cows had records of milk β-hydroxybutyrate (BHB), three types of milk FA composition (*de novo* FA (C4 to C14) composition on a milk basis (DNM), *de novo* FA composition on a fat basis (DNF) and preformed FA (≥C18) composition on a fat basis (PRF)). Milk components were predicted using mid-infrared spectrometry and BHB was $\log_e(x+1)$ transformed. All the lactation records of the cows covered days in milk (DIM) 6-305. All the cows were aged 18-50 months, 30-67 months and 42-84 months at the time of calving for their first, second and third lactation respectively. There were more than three cows for each herd test day and number of milking times class and more than four records for each lactation. The pedigree data used for the analysis were traceable back through three generations of cows that had records. There were 1,100,399 records from 175,154 cows in Hokkaido and 1,553,475 records from 142,887 cows in the rest of Japan.

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_693, © A. Nishiura *et al.* 2022

Prediction of energy balance. EB was predicted using Equation 1:

$$\begin{aligned}PEB = {} & 401.015 - 1.640sDIM + 7.925sDIM^2 + 27.302sDIM^3 - 30.097sDIM^4 - 34.243FPR - \\ & 5.566MY + 0.178MY^2 - 0.002MY^3 + 4.853P - 130.278L^2 + 45.568L^3 - 4.430L^4 + 3.932dMY \\ & - 2.690dMY^2 - 0.778dMY^3\end{aligned} \quad (1)$$

where DIM is days in milk, sDIM is standardized DIM ((2×(DIM-6))/(305-6)-1), FPR is fat-to-protein ratio, MY is milk yield, P is protein %, L is lactose %, and dMY is the current MY minus the previous MY value.

Estimation of genetic parameters. Genetic parameters were estimated using a random regression model, Equation 2:

$$y_{hijk} = HTDT_h + \sum_{m=0}^{5} CM_{im} x_m(t) + \sum_{m=0}^{5} PCA_{jm} x_m(t) + \sum_{m=0}^{3} a_{km} x_m(t) + \sum_{m=0}^{3} pe_{km} x_m(t) + e_{hijk} \quad (2)$$

where y_{hijk} is the value of a trait (BHB, DNM, DNF, PRF, EB); $HTDT_h$ is the fixed effect of the h^{th} herd test day and number of milking times; CM_{im} is the fixed regression coefficient of the i^{th} calving month corresponding to the m^{th}-order polynomial; PCA_{jm} is the fixed regression coefficient of the j^{th} parity calving age (in months), corresponding to the m^{th}-order polynomial; $x_m(t)$ is the Legendre polynomial of order m (m=0-4) and Wilmink's exponential function (exp(–0.05t); m=5) for t DIM; a_{km} is the m^{th}-order additive genetic random regression coefficient of the k^{th} animal; pe_{km} is the m^{th}-order permanent environmental random regression coefficient of the k^{th} animal; and e_{hijk} is the random residual effect. The Gibbs2f90 program (Misztal *et al.*, 2002) was used to estimate variance-covariance components. A single chain of 100,000 samples was generated, and 50,000 samples after a burn-in period of 50,000 cycles were used to estimate posterior means for the model parameters.

Results

Daily average values of BHB, DNM, DNF and PRF fluctuated widely in the early stages of lactation but stabilized in the mid to late stages (Figure 1). There were few differences among parities and regions except for BHB. A different pattern was shown in the early stages of lactation of BHB average changes for different parities and regions.

The heritability estimates of BHB were mostly stable at 0.1-0.2 throughout lactation (Figure 2). The heritability estimates of DNM were around 0.2 soon after parturition and rose to around 0.5 in the mid to late stages of lactation, similar to milk fat %. The heritability estimates of DNF were 0.1-0.3 and heritability estimates of PRF were 0.2-0.3 throughout lactation.

The genetic correlation between PEB and BHB was estimated to be -0.2 to -0.3 soon after parturition, decreasing in the mid to late stages of lactation (Figure 2). The genetic correlation between PEB and DNM was estimated to be -0.1 to -0.3 and remained negative throughout lactation. The genetic correlation between PEB and DNF was estimated to be about 0.4 soon after parturition and became negative in the late stage of lactation. In contrast, the genetic correlations between PEB and PRF were estimated to be about –0.25 soon after parturition but became positive in the mid to late stages of lactation. For both DNF and PRF, there was little difference between Hokkaido and the rest of Japan.

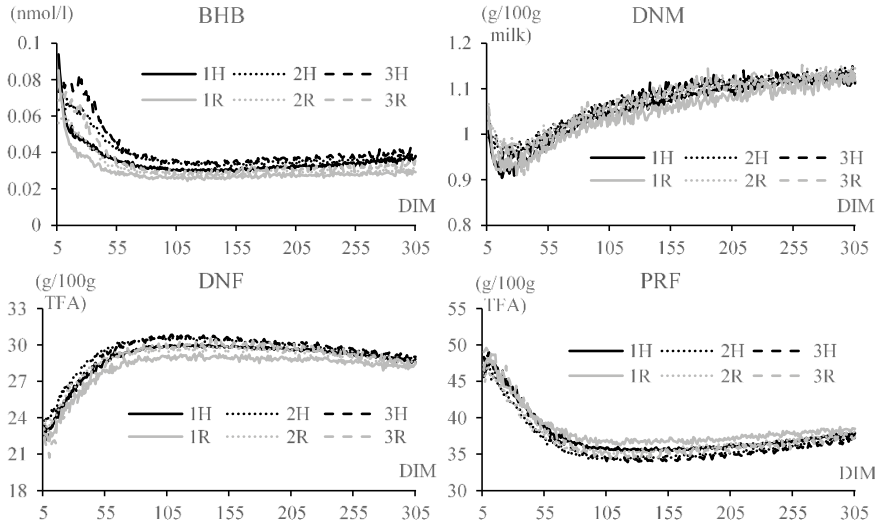


Figure 1. Changes in average BHB, DNM, DNF and PRF across days in milk (DIM) (1: 1st parity, 2: 2nd parity, 3: 3rd parity, H: Hokkaido, R: Rest of Japan, TFA: total fatty acids).

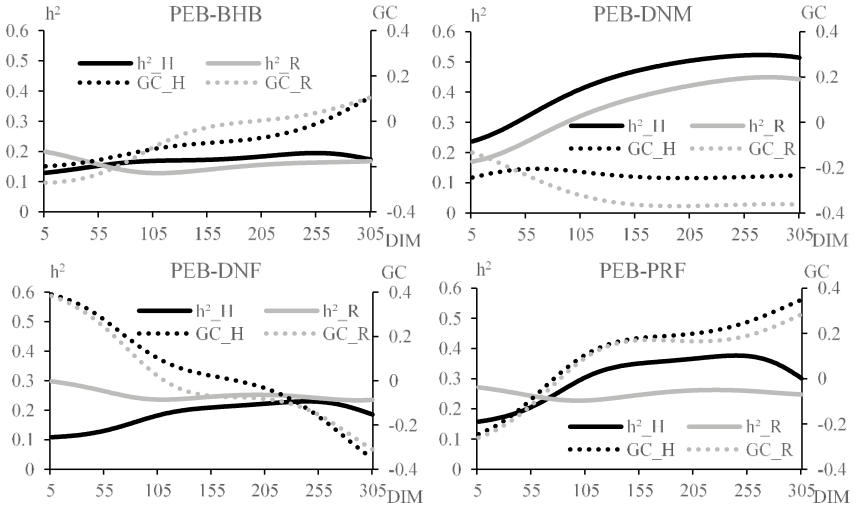


Figure 2. Changes in heritabilities (h^2) for BHB, DNM, DNF and PRF and genetic correlations (GC) between PEB and BHB, DNM, DNF and PRF across days in milk (DIM) (H: Hokkaido, R: rest of Japan).

Discussion

The heritability estimates for BHB were similar to those of Koeck *et al.* (2014) and Jamrozik *et al.* (2016) and can therefore be considered reliable. The genetic correlations of PEB and BHB were estimated to be negative in the early stages of lactation, consistent with the phenotypic relationship. However, the magnitude of these values were smaller than those of the genetic correlations between FPR and BHB (0.49: Koeck *et al.*, 2014). This may be because that FPR is a trait based on milk fat and protein, and genetic correlations between BHB and fat or protein were very high (Klein *et al.*, 2020), whereas EB is a trait estimated using

a more wide-ranging equation. The heritability estimates for *de novo* fatty acids were higher for DNM than for DNF: this was in good agreement with previous reports suggesting milk fat content to be a highly heritable trait (Bastin *et al.*, 2011). Some reports have indicated that shorter chain fatty acids tend to have higher heritability (Bastin *et al.*, 2011; Fleming *et al.*, 2018), but this was not seen in our results. The estimated heritability of PRF was in good agreement with that reported in previous studies (Stoop *et al.*, 2008; Mele *et al.*, 2008), so it appears that the heritability of DNF is slightly lower. This may be due to the multiple types and quantities of fatty acids present. For example, saturated fatty acids had a higher heritability than unsaturated fatty acids (Bastin *et al.*, 2011; Fleming *et al.*, 2018). A detailed analysis of fatty acid composition is needed to clarify this.

The genetic correlations between PEB and DNF and PRF appeared to be in good agreement with the phenotypic relationship in the early stages of lactation. The reason for the slight deviation from the phenotypic relationship between PEB and DNF and PRF in the late stages of lactation may be due to different mechanisms of energy metabolism and origin of milk components in the early and late stages of lactation.

In conclusion, PEB was estimated to have a desirable genetic correlation with BHB and fatty acid composition in the early stage of lactation. It therefore has potential for use as an index for improving the energy balance of Holsteins.

References

Bastin C., Gengler N., and Soyeurt H. (2011) J. Dairy Sci. 94(8):4152-4163. https://doi.org/10.3168/jds.2010-4108

Collard B.L., Boettcher P.J., Dekkers J.C.M., Peticlerc D., and Schaeffer L.R. (2000) J. Dairy Sci. 83(11):2683-2690. https://doi.org/10.3168/jds.S0022–0302(00)75162-9

De Vries M.J., van der Beek S., Kaal-Lansbergen L.M.T.E., Ouweltjes W., and Wilmink J.B.M. (1999) J. Dairy Sci. 82(9):1927-1934. https://doi.org/10.3168/jds.S0022–0302(99)75428-7

Fleming A., Schenkel F.S., Malchiodi F., Ali R.A., Mallard B., *et al.* (2018) J. Dairy Sci. 101(5):4295-4306. https://doi.org/10.3168/jds.2017-14089

Goff J.P. and Horst R.L. (1997) J. Dairy Sci. 80(7):1260-1268. https://doi.org/10.3168/jds.S0022–0302(97)76055-7

Hüttmann H., Stamer E., Junge W., Thaller G., and Kalm E. (2009) Animal 3(2):181-188. https://doi.org/10.1017/S175173110800325X

Ingvartsen, K.L., Dewhurst R.J., and Friggens N.C. (2003) Livest. Prod. Sci. 83(2-3):277-308. https://doi.org/10.1016/S0301-6226(03)00110-6

Jamrozik J., Koeck A., Kistemaker G.J., and Miglior F. (2016) J. Dairy Sci. 99(3):1990-1998. https://doi.org/10.3168/jds.2015-10505

Klein S.-L., Scheper C., May K., and König S. (2020) J. Dairy Sci. 103(11):10332-10346. https://doi.org/10.3168/jds.2020-18339

Koeck A., Jamrozik J., Schenkel F.S., Moore R.K., Lefebvre D.M., *et al.* (2014) J. Dairy Sci. 97(11):7286-7292. https://doi.org/10.3168/jds.2014-8405

Mele M., Dal Zotto R., Cassandro M., Conte G., Serra A., *et al.* (2009) J. Dairy Sci. 92(1):392-400. https://doi.org/10.3168/jds.2008-1445

Misztal I., Tsuruta S., Strabel T., Auvray B., Druet T., *et al.* (2002) Proc. of the 7th World Congress on Genetics Applied to Livestock Production, Montpellier, France.

Stoop W.M., van Arendonk J.A.M., Heck J.M.L., van Valenberg H.J.F., and Bovenhuis H. (2008) J. Dairy Sci. 91(1):385-394. https://doi.org/10.3168/jds.2007–0181

Stoop W.M., Bovenhuis H., Heck J.M.L., and van Arendonk J.A.M. (2009) J. Dairy Sci. 92(4):1469-1478. https://doi.org/10.3168/jds.2008-1468

694. Associations between milk protein polymorphisms and milk production in Holstein-Friesian cows

D.N. Kelly[1*], F.L. Dunne[1], R.D. Evans[2] and D.P. Berry[1]

[1]Teagasc, Moorepark, P61 C996, Cork, Ireland; [2]Irish Cattle Breeding Federation, Bandon, P72 X050, Cork, Ireland; david.kelly@teagasc.ie

Abstract

Milk protein variants for beta casein (β-CN), kappa casein (κ-CN), and beta lactoglobulin (β-LG) affect different properties in milk influencing milk product manufacturing. The impact of selection on such variants on milk production in Irish Holstein-Friesian cows is, however, unknown. Mixed models were used to determine the associations between milk production and milk protein genotypes for β-CN, κ-CN, and β-LG in 13,095 lactation records from 5,431 cows. Presence of the β-CN A_1A_3 and A_2A_3 genotypes were associated with a greater milk yield, while the β-CN II and A_2A_3 genotypes were associated with a greater protein yield and concentration. The β-LG AA and BB genotypes were associated with a lower and higher fat concentration, respectively, relative to the AB genotype. The κ-CN AE and β-LG BB genotypes were both associated with a lower protein yield and a lower protein concentration.

Introduction

The impact of the major milk proteins namely beta casein (β-CN), kappa casein (κ-CN), and beta lactoglobulin (β-LG) on bovine milk composition and cheese making properties has been extensively studied (Gai *et al.*, 2021). For example, the presence of the β-CN B and A_1 variants or the κ-CN B variant is favourable for rennet coagulation (Gai *et al.*, 2021). Each genetic polymorphism for these milk proteins are mainly governed by a single base pair change on chromosome six for β-CN and for κ-CN and on chromosome 11 for β-LG; thus direct selection (or editing) for a specific milk variant could be relatively easily achieved. Nonetheless, before selection (or editing) for a specific milk protein variant takes place, the effect of the milk variants on other economically important traits must first be established. To that end, the impact of such variants on milk production in Irish dairy cows is unknown. Therefore, the objective of the present study was to quantify the association between milk genetic variants for three major milk proteins and milk production traits in a population of Irish Holstein-Friesian cows.

Materials & methods

Genotype data. Genotype data were available on animals genotyped with the custom International Beef and Dairy (IDB) v3 chip (Mullen *et al.*, 2013). Animals with a genotype call rate <95% were removed as well as single nucleotide polymorphisms (SNP) that were not called in at least 50% of the population. Genotypes were phased and imputed using FImpute (Sargolzaei *et al.*, 2014) in order to obtain parental haplotypes for the milk variants of β-CN, κ-CN, and β-LG. Only the milk protein variant data for β-CN, κ-CN, and β-LG for 9,069 dairy cows with a known sire and dam, of ≥87.5% Holstein-Friesian, and born from 2000 onwards were available.

Performance data. Only cows that calved for the first time between 540 and 1,240 d of age were retained (Ring *et al.*, 2021). Lactation records for milk yield, protein yield, fat yield, protein concentration, and fat concentration all adjusted to a 305-d mature equivalent were available for 33,257 lactations from 8,508 genotyped cows across 976 herds. Lactation records were removed if any milk trait deviated greater than three standard deviations from the parity mean. Lactation records following calvings that resulted in multiple births were excluded as were records where a calving event occurred <545 d or >545 d of age from the parity median. Contemporary groups were defined as per the Irish national dairy cow genetic

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_694, © D.N. Kelly *et al.* 2022

evaluations using an algorithm described in detail by Berry *et al.* (2013). Cows that calved within 10 days of each other in the same herd were initially clustered together. Contemporary groups with <10 records were clustered together if the interval between the first and last calving event per contemporary group did not exceed 100 days. Lactation records from contemporary groups with <15 records were removed from all analyses. Subsequently, 13,095 lactation records from 5,431 cows across 748 contemporary groups remained for analysis.

Statistical analyses. Associations between each milk variant genotype and performance were determined using a series of univariate animal models in DMU (Madsen and Jensen, 2008) as follows:

$$y = \text{parity}\times\text{age} + cg + \beta\text{-CN} + \kappa\text{-CN} + \beta\text{-LG} + cow_{pe} + cow_{a} + e \quad (1)$$

where y = milk yield (kg), fat yield (kg), protein yield (kg), fat concentration (%) or protein concentration (%); parity×age = fixed effect of the interaction between cow parity (1, 2, 3, 4, 5+) and age at calving (in months) relative to the parity median; cg = fixed effect of the contemporary group; β-CN = the fixed effect of β-CN genotypes; κ-CN = the fixed effect of κ-CN genotypes; β-LG = the fixed effect of β-LG genotypes; cow_{pe} = the random permanent environmental effect of the cow; cow_{a} = the random additive genetic effect; e = random residual effect. The pedigree of each cow was traced back at least four generations and contained 27,905 animals. The significance of the fixed effects for the milk protein genotypes were determined using likelihood ratio tests.

Results

The mean (± SD) milk yield, fat yield, protein yield, fat concentration, and protein concentration was 7,220.24 kg (±1,485.96 kg), 302.36 kg (±64.71 kg), 254.87 kg (±49.19 kg), 4.22% (±0.53%), and 3.55% (±0.25%), respectively. The β-CN genotypes of A_1A_2, A_2A_2, A_2I, and A_1A_1 were the most common in the sample population with genotype frequencies of 30.2, 23.2, 13.2 and 10.1%, respectively. The most common κ-CN genotypes were AB, AA, and BB with genotype frequencies of 45.5, 33.1, and 14.5% respectively. Similarly, the genotype frequencies of β-LG AB, AA, and BB were 49.9, 23.8 and 26.2%, respectively.

The estimated effects of the β-CN, κ-CN, and β-LG genotypes on the milk performance traits are in Table 1. A_2A_3 cows produced 623.83 kg (i.e. 8.55%) more than A_1A_2 cows. Cows with the β-LG BB genotype produced 64.67 kg (i.e. 0.89%) less milk than AB cows. Less milk was associated with the κ-CN BB genotype relative to the AA genotype. All of the β-CN FI, κ-CN BB, and β-LG AA genotypes were associated with a less fat yield compared to the respective reference genotypes. A_2A_3 cows produced milk with a 5.38% lower fat concentration than their A_1A_2 genotype contemporaries. β-LG AA cows had a 0.62% lower fat concentration, and BB cows had 0.72% higher fat concentration, than AB cows. κ-CN was not associated with fat concentration.

The β-CN genotypes A_2A_3, II, and A_2I had respective protein yields 8.02, 3.22 and 1.66% higher than the A_1A_2 genotype. Similarly, compared to the A_1A_2 genotype, the β-CN genotypes BF, II, A_1I, A_2I, and A_1A_1 were associated with a 2.47, 2.02, 1.07, 0.99 and 0.60% higher protein concentration, respectively. In contrast, the κ-CN AE and β-LG BB genotypes were both associated with a less protein yield and a lower protein concentration than their respective reference genotypes of κ-CN AB and β-LG AB.

Discussion

Modern large-scale genotyping programs use SNP-chips that also contain probes to call major genetic variants. These data are being relayed back to the producer who, in turn, are rightly questioning the value of such data and the impact of selection on individual variants. Evidence is clear that some of the casein

Table 1. Milk variant genotype solutions (SE in parentheses)[1], as well *P* -values for the protein effect, for milk yield, fat yield, protein yield, fat concentration, and protein concentration.

	Milk (kg)	Fat (kg)	Protein (kg)	Fat (% × 10)	Protein (% × 10)
Beta casein					
P-value	<0.001	<0.001	<0.001	<0.001	<0.001
A_1A_1	-41.26 (39.30)	0.39 (1.64)	0.00 (1.26)	0.31 (0.19)	0.21 (0.08)
A_1A_3	463.55 (346.41)	16.08 (14.62)	13.75 (11.12)	-0.83 (1.66)	-0.31 (0.72)
A_1B	-5.88 (72.81)	0.73 (3.02)	-2.67 (2.33)	0.09 (0.36)	-0.28 (0.16)
A_1F	-93.08 (83.33)	-1.44 (3.46)	-2.49 (2.67)	0.46 (0.41)	0.18 (0.18)
A_1I	6.43 (44.88)	0.52 (1.87)	3.31 (1.24)	-0.02 (0.22)	0.38 (0.10)
A_2A_2	35.30 (29.21)	-1.4 (1.22)	1.29 (0.93)	-0.34 (0.14)	0.03 (0.06)
A_2A_3	623.83 (223.35)	9.32 (9.20)	20.41 (7.10)	-2.39 (1.11)	-0.16 (0.49)
A_2B	43.67 (60.36)	0.42 (2.52)	0.78 (1.93)	-0.26 (0.30)	-0.16 (0.13)
A_2F	-32.81 (63.93)	-1.57 (2.67)	-0.91 (2.05)	-0.04 (0.31)	-0.02 (0.14)
A_2I	43.46 (38.72)	1.07 (1.61)	4.22 (1.24)	-0.18 (0.19)	0.35 (0.08)
BB	7.31 (280.65)	7.22 (11.71)	-2.26 (9.00)	0.93 (1.36)	-0.50 (0.59)
BF	-297.23 (198.43)	-5.89 (8.37)	-4.68 (6.37)	1.48 (0.96)	0.87 (0.42)
BI	-102.4 (108.58)	-2.37 (4.55)	-1.98 (3.48)	0.10 (0.53)	0.19 (0.23)
FF	48.75 (285.24)	2.02 (12.08)	1.16 (9.15)	0.09 (1.38)	-0.22 (0.60)
FI	-209.25 (120.85)	-10.49 (5.06)	-6.83 (3.87)	-0.29 (0.59)	0.11 (0.26)
II	100.35 (87.78)	1.66 (3.66)	8.20 (2.81)	-0.33 (0.43)	0.71 (0.19)
Kappa casein					
P -value	<0.001	<0.001	0.023	0.061	0.026
AA	18.30 (28.07)	0.86 (1.17)	-0.96 (0.90)	-0.06 (0.14)	-0.25 (0.06)
AE	-2.21 (60.26)	-0.89 (2.50)	-3.25 (1.62)	-0.04 (0.30)	-0.41 (0.13)
BB	-71.64 (34.92)	-2.98 (1.45)	-0.80 (1.12)	0.04 (0.17)	0.28 (0.07)
BE	-8.27 (68.52)	-1.35 (2.85)	0.68 (2.19)	-0.15 (0.33)	0.12 (0.15)
EE	189.71 (199.17)	-1.13 (8.27)	3.50 (6.37)	-0.71 (0.97)	-0.40 (0.43)
Beta lactoglobulin					
P-value	<0.001	0.0135	<0.001	<0.001	<0.001
AA	-18.04 (25.96)	-2.61 (1.08)	-0.34 (0.83)	-0.26 (0.13)	0.07 (0.06)
BB	-64.67 (25.70)	-0.45 (1.07)	-2.97 (0.82)	0.30 (0.13)	-0.11 (0.05)

[1] Solutions were relative to the genotypes of A_1A_2 for beta casein, AB for kappa casein, and AB for beta lactoglobulin.

variants are favourable for cheese manufacturing, including rennet and acid coagulation properties (Gai *et al.*, 2021). This could prompt milk processors to actively promote the use of dairy sires carrying those alleles or the creation of herds with favourable allele combinations aligned to the product portfolio of the processor. Prior to embarking on such an aggressive strategy, the impact on other performance traits should firstly be quantified prompting this study.

The A_2-type β-CN variants include A_2, A_3, and I and allele combinations of these variants (i.e. A_2A_3, A_2I, II) generally had more favourable relationships with protein yield and concentration than the A_1-type variants; these A_2-type variants are also associated with higher casein yields (Gai *et al.*, 2021). Milk with A_2-type β-CN variants, however, is not always desirable for cheese manufacturing due to poor rennet coagulation but may be more suitable for yoghurt products (Gai *et al.*, 2021). The greater protein concentration associated with the κ-CN BB genotype, compared to the AA and AB genotypes, is advantageous as a desirable shorter

rennet coagulation time and higher cheese yield is associated with this genotype (Gai *et al.*, 2021). On the other hand, there must be cognisance of the trade-off between the higher protein content but lower milk and fat yield associated with the κ-CN BB genotype; the genotype frequency of κ-CN BB genotype is actually higher within the sample population used in the present study compared to other Holstein/Friesian populations (Gustavsson *et al.*, 2014; Comin *et al.*, 2008). The present study demonstrates that variation exists in the milk production of Holstein-Friesian cows of different milk variant genotypes. Producers or processors that wish commence selection (or editing) for individual variants or allele combinations should remain mindful of the potentially adverse, as well as favourable, impacts of such an approach on economically important milk production traits.

References

Berry, D.P., Kearney, J., Twomey, K. and Evans, R. (2013). Irish J. Agric. Food Res. 52(1): 1-16. http://hdl.handle.net/11019/445.

Comin, A., Cassandro, M., Chessa, S., Ojala, M., Dal Zotto, R., *et al.* (2008). J. Dairy Sci. 91(10): 4022-4027. https://doi.org/10.3168/jds.2007-0546.

Gai, N., Uniacke-Lowe, T., O'Regan, J., Faulkner, H. and Kelly, A.L. (2021). Foods. 10(10): 2409. https://doi.org/10.3390/foods10102409.

Gustavsson, F., Buitenhuis, A.J., Johansson, M., Bertelsen, H.P., *et al.* (2014). J. Dairy. Sci. 97(6): 3866-3877. https://doi.org/10.3168/jds.2013-7312.

Madsen, P. and Jensen, J. (2008). A User's Guide to DMU: A Package for Analysing Multivariate Mixed Models Version 6. Release 5.2. Available at: https://dmu.ghpc.au.dk/dmu/index.html.

Mullen, M., McClure, M., Kearney, J.F., Waters, S., *et al.* (2013). Proc. of INTERBULL. Nantes, France.

Ring, S.C., Evans, R.D., Cromie, A.R. and Berry, D.P. (2021). J. Dairy Sci. 104(7): 8076-8093. https://doi.org/10.3168/jds.2020-19957.

Sargolzaei, M., Chesnais, J.P. and Schenkel, F.S. (2014). BMC Genom. 15(1): 478. https://doi.org/10.1186/1471-2164-15-478.

695. Genetic analysis of milking traits in Holstein cows based on automatic milking system data

J. Kawakami[1*], Y. Gotoh[1], T. Baba[1], S. Yamaguchi[2], S. Nakagawa[2], S. Yamaguchi[3], T. Oka[4] and T. Kawahara[1]

[1]Holstein Cattle Association of Japan, Hokkaido Branch, Sapporo-shi, 001-0015, Hokkaido, Japan; [2]Hokkaido Dairy Milk Recording and Testing Association, Sapporo-shi, 060-0004, Hokkaido, Japan; [3]Livestock Improvement Association of Japan, Kouto-ku, 135-0041, Tokyo, Japan; [4]Holstein Cattle Association of Japan, Nakano-ku, 164-0012, Tokyo, Japan; j.kawakami@holstein.jp

Abstract

The aim of this study was to estimate heritabilities and breeding values for milk flow (MF), dead milk time (DT) and milking time (MT) using the records of 3,069 primiparous Holstein cows collected by an automatic milking system. Estimated heritabilities using the three-trait animal model was 0.37 for MF, 0.22 for DT and 0.29 for MT. Genetic correlations among the traits ranged from 0.76 to 0.88 on an absolute scale. While the genetic trends of MF and MT were slightly positive, a negative trend was observed for DT. The results of this study suggest that genetic improvement of these milking traits is feasible. The breeding value of milking traits was correlated with the somatic cell score, udder traits and milking speed. These traits can be used as indicators for indirect selection of milking traits.

Introduction

The automatic milking system (AMS) is an effective tool to reduce labour cost and increase milk production; thus, the number of dairy herds using AMS has dramatically increased in Japan. AMS can also collect detailed milking information for milking cows (Aerts *et al.* 2021). Santos *et al.* (2018) suggested that compared to conventional milking systems, AMS data could provide a more accurate estimation of genetic parameters for milk production and milking traits. In addition, AMS offers the ability to assess traits where genetic evaluation has not been implemented. Therefore, further analysis of these data may help improve the efficiency of milking cows (Wethal and Heringstad 2019).

To increase the efficiency of AMS operation, cows with faster milk flow and milking time are preferable. In this study, we aimed to estimate genetic parameters and breeding values for milk flow (MF), dead milk time (DT) and milking time (MT) in Japanese Holstein cows. In addition, genetic relationships between milking traits and milk production, type traits, milking speed which is scored subjectively on a 5-point scale and Nippon total profit index were investigated.

Materials & methods

Data. We used the records of primiparous Holstein cows, which were collected by the Livestock Improvement Association of Japan and Hokkaido Dairy Milk Recording and Testing Association, from 2007 to 2021 at 249 dairy herds with Lely AMS. The following were excluded: (1) individuals moved to the other herd between 5 and 305 days in milk; (2) milking records with extreme intervals from last milking (i.e. <1 hour or >24 hours); and (3) cows with unknown parents. A total of 967,531 AMS records consisting of 3,069 animals were available for genetic analysis.

Three milking traits were considered in this study: MF (kg/min), which is the average milk flow rate during milking; DT (sec), which is the time gap between cluster attachment and start of milking; and MT (min), which is the time difference between milking start and cluster removal time. Traits were averaged for every milking. Descriptive statistics for the milking traits are presented in Table 1. The averages of the traits were 2.7 kg/min for MS, 17.4 sec for DT and 3.3 min for MT.

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_695, © J. Kawakami *et al.* 2022

Table 1. Means, standard deviations (SD), minimum, and maximum values for milking traits.

Trait	N	Mean	SD	Minimum	Maximum
Milk flow (kg/min)	967,531	2.7	0.90	0.2	9.0
Dead milk time (sec)	967,531	17.4	8.46	2.0	60.0
Milking time (min)	967,531	3.3	1.28	1.0	10.2

Model. The three-trait animal model was used for genetic analysis of milking traits. We considered the same effects in the model, as shown in the following equation:

$$y_{ijklmn} = HYS_i + CA_j + CM_k + SOL_l + u_m + pe_m + e_{ijklmn}$$

where y_{ijklmn} is the observation; HYS_i is the fixed effect of the i-th herd-year-season (1,542 levels); CA_j is the fixed effect of the j-th calving age (14 levels); CM_k is the fixed effect of the k-th calving month (12 levels); SOL_l is the fixed effect of the l-th stage of lactation (10 levels); u_m and pe_m are random effects for the m-th cow for additive genetic and permanent environmental effects, respectively; and e_{ijklmn} is the random residual associated with each record. The season at HYS_i was grouped into four categories of March to May, June to August, September to November and December to February. We used 13,950 pedigree animals that were traced back up to three generations from cows with records.

Genetic parameters and breeding values were estimated using the AIREMLF90 and BLUPF90 programs, respectively (Misztal *et al.*, 2002). To investigate genetic relationships with other traits, Pearson correlations between breeding values of the milking traits and milk production, type traits, milking speed, and Nippon total profit index were calculated. The breeding values of their traits except milking traits, were obtained from the national genetic evaluation in August 2021 by the National Livestock Breeding Centre (Fukushima, Japan).

Results

Estimates of heritabilities for milking traits and their genetic and phenotypic correlations are presented in Table 2. The heritabilities ranged from 0.22 to 0.37. While strong genetic correlations were estimated on an absolute scale with a range of 0.76 to 0.88, the phenotype correlations were moderate.

The genetic trends of milking traits for cows with records and domestic sires (n=347) with at least one daughter are shown in Figure 1. Genetic bases were based on cows born in 2015. While slight positive trends were observed for MF and MT, DT showed a slightly negative trend.

Pearson correlations between the breeding values of milking and other traits are presented in Table 3. In the cows, all milking traits were significantly correlated with the protein yield, somatic cell score, mammary system, front teat length, milking speed and Nippon total profit index. On the other hand, in domestic sires, significant correlations with the somatic cell score, udder depth and milking speed were found.

Table 2. Heritability (diagonal) and genetic (upper triangle) and phenotypic (lower triangle) correlation of milking traits.

Trait		1	2	3
1	Milk flow	0.37	-0.88	-0.76
2	Dead milk time	-0.55	0.22	0.83
3	Milking time	-0.59	0.44	0.29

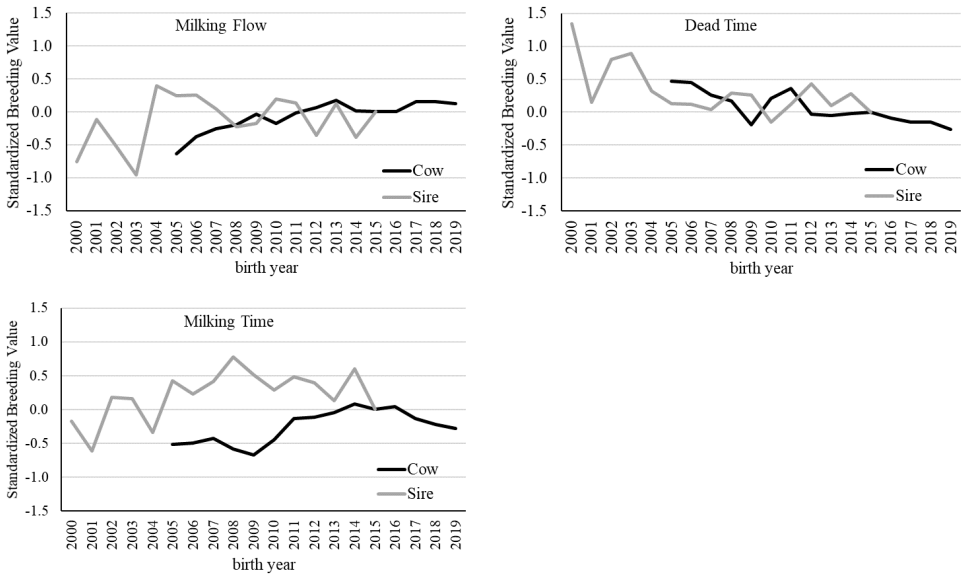


Figure 1. Genetic trends of milking traits in cows with records and sires with at least one daughter.

Table 3. Estimated correlation between breeding values of milking traits and milk production, linear type traits, milking speed and Nippon total profit index.[1]

Trait	N	Milking flow		Dead time		Milking time	
Cow							
MY	3,069	0.12	**	-0.04	n.s.	0.12	**
FY	3,069	0.13	**	-0.08	**	0.01	n.s.
PY	3,069	0.12	**	-0.06	*	0.07	**
SCS	2,533	0.25	**	-0.24	**	-0.24	**
MSM	989	0.17	**	-0.15	**	-0.09	*
UA	989	0.14	**	-0.14	**	-0.03	n.s.
UD	989	0.13	**	-0.14	**	-0.01	n.s.
FTL	989	-0.22	**	0.20	**	0.09	*
MSP	591	0.44	**	-0.36	**	-0.47	**
NTP	989	0.16	**	-0.11	**	0.11	**
Sire							
MY	347	0.00	n.s.	0.02	n.s.	0.27	**
FY	347	-0.01	n.s.	-0.01	n.s.	0.09	n.s.
PY	347	0.04	n.s.	0.00	n.s.	0.17	*
SCS	347	0.14	**	-0.20	**	-0.16	*
MSM	347	0.18	**	-0.15	*	-0.07	n.s.
UA	347	0.17	*	-0.13	*	-0.11	n.s.
UD	347	0.18	**	-0.15	*	-0.10	*
FTL	347	-0.13	*	0.15	*	0.05	n.s.
MSP	347	0.38	**	-0.38	**	-0.45	**
NTP	347	0.06	n.s.	-0.04	n.s.	0.12	n.s.

[1] MY : milk yield; FY : fat yield; PY : protein yield; SCS : somatic cell score; MSM : mammary system; UA : fore udder attachment; UD : udder depth; FTL : front teat length, MSP : milking speed; NTP : Nippon total profit index. **: $P<0.001$; *: $P<0.01$; n.s.: not significant.

Discussion

Heritabilities of milking traits was moderate, and the values for MF and MT were similar to those in previous reports, which ranged from 0.25 to 0.55 for MF and 0.19 to 0.42 for MT (Aerts *et al.*, 2021; Carlstrom *et al.*, 2013; Carlstrom *et al.*, 2014; Dechow *et al.*, 2020; Santos *et al.*, 2018). To the best of our knowledge, no study has reported the heritability of DT. The results of this study suggest that the improvement of these traits can be achieve through genetic selection.

There were strong genetic correlations among milking traits, as previously reported by Dechow *et al.* (2020) and Santos *et al.* (2018). We observed negative genetic correlations between MF and the other milking traits, MT and DT. This suggests that genetic improvement for faster MF can lead to a shorter DT and MT. Since the records of milking traits were collected only from dairy herds with AMS, indirect selection by traits that which have been routinely evaluated in national genetic evaluation can be one way to improve milking traits. Therefore, the somatic cell score, udder traits and milking speed, which were genetically related to milking traits in this study, are expected to be valuable indicators of genetic improvement.

The results of the genetic trend for MF and DT indicated that milk flow and the time from teat cup attachment to milking improve annually. A slight positive trend of MT may be caused by genetic improvement in milk production.

References

Aerts J., Piwczynski D., Ghiasi H., Sitkowska B. *et al.* (2021) Animals 11(7):1943. https://doi.org/10.3390/ani11071943

Carlström C., Pettersson G., Johansson K., Strandberg E. *et al.* (2013) J. Dairy Sci. 96(8):5324-5332. https://doi.org/10.3168/jds.2012-6221

Carlström C., Strandberg E., Johansson K., Pettersso G. *et al.* (2014) J. Dairy Sci. 97(1):497-506. https://doi.org/10.3168/jds.2013-6948

Dechow C.D., Sondericker K.S., Enab A.A., and Hardie L.C. (2020) J. Dairy Sci. 103(12):11503-11514. https://doi.org/10.3168/jds.2020-18786

Misztal l., Tsuruta S., Strabel T., Auray B. *et al.* (2002) Proc. Of the 7th WCGALP, Montpellier, France.

Santos L.V., Brügemann K., Ebinghaus A., and König S. (2018) Arch. Anim. Breed. 61(2):161-171. https://doi.org/10.5194/aab-61-161-2018

Wethal K.B., and Heringstad B. (2019) J. Dairy Sci. 102(9):8221-8233. https://doi.org/10.3168/jds.2019-16625

696. Fertility traits of Mexican Holstein cattle: genetic parameters, correlations and changes in genetic diversity over time

C. Durán-Alvarez[1*], A. García-Ruiz[2], R.A. Alonso Morales[3], L.E. Eguiarte [4], D. Piñero[4] and F.J. Ruiz-López[2]

[1]Posgrado en Ciencias Biológicas-UNAM, Circuito de Posgrados, Edificio D, 1° Piso, Ciudad Universitaria, Alcaldía Coyoacán. C. P. 04510, Ciudad de México, México; [2]CENID FyMA-INIFAP; Km. 1, Carretera a Colón, Col. Ajuchitlán Colón, C.P. 76280, Querétaro, México; [3]Facultad de Medicina Veterinaria y Zootecnia-UNAM, Circuito escolar s/n, Ciudad Universitaria, Alcaldía Coyoacán, C.P. 04510, Ciudad de México, México; [4]Instituto de Ecología-UNAM, Circuito exterior s/n anexo al Jardín Botánico, Ciudad Universitaria, Alcaldía Coyoacán, C.P. 04500, Ciudad de México, México; caro_du@ciencias.unam.mx

Abstract

The present research aimed to describe the genetic parameters and correlations among fertility traits of Holstein cattle from Mexico and their changes over time. The reproductive parameters (RP) included were: calving interval to first service, number of inseminations per conception, days open, calving interval, and age at first calving. The study included female records provided by the Mexican Holstein association (2006-2019). Variance components for the mentioned RP were estimated using restricted maximum likelihood and an animal model using the BLUPF90 software. Heritability estimates ranged from 4 to 9%, and repeatability estimates between 8 and 9%, similar to previously reported for fertility traits. The phenotypic and genetic correlations showed positive values among almost all RP. Estimates of additive genetic variance in the analyzed sample showed changes over time for the fertility characteristics of Mexican Holsteins probably due to the implementation of genomic selection in the origin populations of imported germplasm.

Introduction

Fertility is one of the factors of utmost importance in productivity, because its decrease can be the cause of economic losses. Genetic fertility studies are important to support the future profitability of production systems and reduce waste generated by poor fertility. Genetic parameters (heritability and repeatability) are specific to the populations or subpopulations evaluated and can vary through the years, so their estimation is justified in the Holstein population of Mexico. Considering that the fertility characteristics have been improved in Holstein cattle in the United States and that Mexico has imported germplasm from the US, it is possible that the latter shows changes over time in these traits even though no direct selection has been applied for these traits. The objectives of this study were to describe the parameters, correlations and genetic trends of reproductive traits in Holstein cattle from Mexico.

Materials & methods

Reproductive data and parameters analyzed. This study included 415,859 reproductive records of Holstein cows (*Bos taurus taurus*) from January 2006 to December 2019, information provided by the Mexican Holstein association. The reproductive parameters calculated were: calving to first insemination (CFI), number of services per conception (NSC), days open (DO), calving interval (CI) and age at first calving (AFC).

Variance components. Variance components were estimated with an animal model, using restricted maximum likelihood (Patterson and Thompson, 1971), and running the BLUF90 program (Misztal *et al.*, 2014). The general model used was:

$$y = Xb + Zu + e \quad (1)$$

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_696, © C. Durán-Alvarez *et al.* 2022

where y = vector of the records of: CFI, NSC, DO, CI or AFC; b = vector of fixed effects (herd-year-station (hys), production level (np), calving number (nop), calving age (ep) and herd-year-season of birth for AFC (hysn)); u = vector of random effects (effect of the animal and permanent environment when applicable); X and Z = incidence matrices that relate the observations to the fixed or random effects, respectively; e = vector of the effects of the error or residuals.

Genetic parameters and correlations. Heritability, repeatability and genetic correlations were obtained from the estimates of the variance components. The standard deviation of the estimated parameters was approximated from their expectations using the REMLF90 program (Misztal, 2002).

Analysis by periods. Considering the exchange of germplasm between the United States and Mexico, as well as the changes in the values of the genetic parameters of the former, the data was grouped into three periods (P1: 2006-2009, P2: 2010-13 and P3: 2014-2017).

Results

Genetic parameters and correlations. The heritability of the reproductive parameters ranged from 4 to 9%, while the repeatability estimates were around 8% (Table 1).

The estimated phenotypic and genetic correlations are shown in Table 2.

Table 1. Number of records and individuals with phenotypic information included in the study, as well as descriptive statistics and genetic parameters.[1]

RP	Record number	Individuals number	$\bar{x} \pm SD$	$h^2 \pm SD$	$r \pm SD$
CFI	202,545	107,179	71.5±22.8	0.04±0.003	0.09±0.003
NSC	202,545	107,179	2.7±1.8	0.05±0.004	0.08±0.003
DO	194,816	101,763	131.6±74	0.05±0.004	0.08±0.003
CI	139,901	77,819	413.6±84.6	0.05±0.004	0.09±0.004
AFC	103,467	103,467	24.1±2.7	0.09±0.007	-

[1] CFI = calving to first insemination in days; NSC = number of services per conception; DO = days open; CI = calving interval in days; AFC = age at first calving in months; $\bar{x}$ = phenotypic average for the characteristic; SD = standard deviation; h^2 = heritability; r = repeatability.

Table 2. Phenotypic and genetic correlations, as well as number of records for each pair of characteristics.[1]

Traits	Record number	Genetic covariance	Phenotypic correlation	Genetic correlation
CFI-NSC	202,545	0.20	-0.05*	0.13 (0.06)
CFI-DO	192,527	32.02	0.26*	0.55 (0.05)
CFI-CI	138,277	28.73	0.23*	0.51 (0.06)
CFI-AFC	103,467	0.28	0.08*	0.09 (0.07)
NSC-DO	194,816	4.60	0.77*	0.84 (0.02)
NSC-CI	139,901	5.93	0.78*	0.88 (0.02)
NSC-AFC	84,675	0.05	0.01*	0.22 (0.08)
DO-CI	137,158	224.60	0.98*	0.99 (0.002)
DO-AFC	80,467	0.82	0.06*	0.09 (0.08)
CI-AFC	62,999	1.04	0.07*	0.12 (0.11)

[1] CFI = interval to first service in days; NSC = number of services per conception; DO = days open; CI = calving interval in days; AFC = age at first delivery in months; * $P<0.001$, in parentheses the standard deviation.

Analysis by periods, additive genetic variance. The results can be divided into two groups of comparisons, group I: period 1 and 2, and group II: period 2 and 3 (Table 3). The additive genetic variances for group I were similar for all traits except for AFC which showed a significant decrease. In group II there was an increase in CFI, a decrease in NSC, DO and CI while the AFC was similar.

Genetic values. Figure 1a and 1b show how, in general, average BV was similar between P1 and P2, decreasing in P3.

Table 3. Additive genetic variance for five reproductive parameters in three time periods (P1: 2006-2009, P2: 2010-13 y P3 2014-2017).[1]

RP	Period	$\sigma^2_a \pm$ SD
CFI	P1:06-09	12.48±2.2
	P2:10-13	11.17±2.0
	P3:14-17	19.88±3.0
NSC	P1:06-09	0.15±0.02
	P2:10-13	0.16±0.02
	P3:14-17	0.10±0.02
DO	P1:06-09	229.8±35.32
	P2:10-13	214.6±33.87
	P3:14-17	194.8±25.66
CI	P1:06-09	478±51.1
	P2:10-13	466.7±52.5
	P3:14-17	266.8±53.9
AFC	P1:06-09	1.23±0.12
	P2:10-13	0.19±0.04
	P3:14-17	0.15±0.03

[1] CFI = interval to first service in days; NSC = number of services per conception; DO = days open; CI = calving interval in days; AFC = age at first delivery in months; σ^2_a = additive genetic variation; SD = standard deviation.

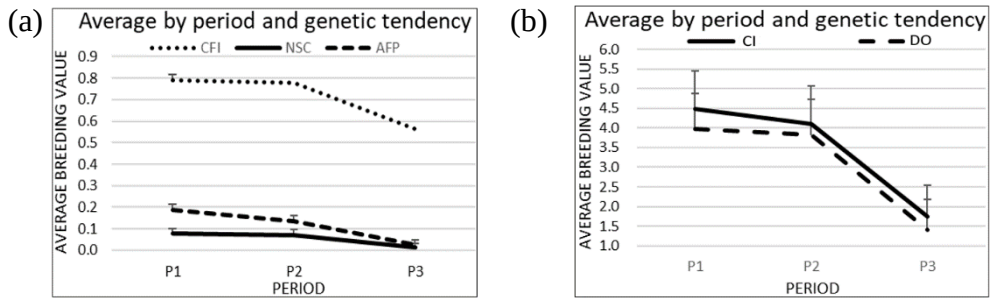


Figure 1. Trends of the average genetic values for CFI, NSC, AFP (a) and CI, DO (b) for the Holstein population of Mexico in three time periods.

Discussion

Phenotypic averages of the five reproductive parameters were similar to those previously reported in other dairy cattle populations (Weigel and Rekaya 2000; Ríos-Ultrera *et al.*, 2010). The variations could be due to the fact that there is no consensus on the biologically acceptable limits of the measurements of these characteristics, which generates possible biases to the estimates and, on the other hand, the fertility characteristics vary between individuals largely due to the environment (Ghiashi *et al.*, 2011).

The heritability and repeatability estimates were found within previously reported limits (González-Recio and Alenda 2005). The estimated values in the Holstein population of Mexico are low, less than 10%, so that in addition to genetics, the environment and management issues as a source of variation have an important effect (Weigel and Rekaya, 2000). Making genetic improvement of reproductive traits will allow permanent benefits in the populations, unlike improvements in management (environments), which offer benefits temporarily.

The phenotypic and genetic correlations showed positive values between the reproductive parameters (Table 2), similar to those reported (González-Recio and Alenda, 2005, Ríos-Ultrera *et al.*, 2010) The high genetic correlation between DO-CI of 0.99, as reported by González-Recio and Alenda (2005) and by Ghiashi *et al.* (2011), suggests that these two traits, are almost equivalent.

When comparing additive genetic variance changes across time, in P1 vs P2 we observe a decrease for age at first calving, possibly due to the selection. In the second period (P2 vs P3), there was an increase in CFI, which tells us about genetic diversity and the greater the genetic diversity, the greater the opportunity for the response to selection (Oldenbroek, 2017); while decreases were observed for NSC, DO and CI. Average genetic values were similar in the first two periods but showed a decrease for four of the five reproductive parameters (CFI, DO, CI and AFC) from P2 to P3, this reduction in genetic values being desirable. These changes in fertility traits could be attributed to the incorporation of genomic information in the selection processes in the United States as of 2009, as shown by Wiggans *et al.* (2017), and García-Ruiz *et al.* (2016), because of the use of germplasm imported mainly from the United States. For the NSC, no changes were observed in the genetic values.

Our result suggest that it would be convenient to select only for one parameter, since it could be considered that its effects are controlled by the same genetic regions (i.e. they are the same trait).The estimates of additive genetic variances and genetic values in different periods showed the existence of changes over time in our population, possibly derived from the influence of the use of germplasm from imported bulls from countries where genomic selection has been implemented.

References

García-Ruiz A., Cole J.B., VanRaden P.M., Wiggans G.R., Ruiz-López, F. J., *et al.* (2016) PNAS 113(28):E3995-E4004.

Ghiasi H., Pakdel A., Nejati-Javaremi A., Mehrabani-Yeganeh H., Honarvar M., González-Recio O., *et al.* (2011) Livestock Science 139(3):277-280.

González-Recio O., and Alenda R. (2005) J Dairy Sci 88:3282-3289.

Misztal I., Tsuruta S., Lourenco D., Aguilar I., Legarra A. *et al.* (2014) Manual for BLUPF90 family of programs – Available at: ttp://nce.ads.uga.edu/wiki/lib/exe/fetch.php?media=blupf90_all7.pdf

Oldenbroek J.K. (2017) Wageningen Academic Publishers. Wageningen, the Netherlands pp. 13-45.

Patterson H. D., and Thompson R. (1971) Biometrika 58:545-554.

Ríos-Ultrera A., Calderón-Robles R., Rosete-Fernández J., and Lagunes-Lagunes J. (2010) Agron. Mesoam 21:245-253.

Weigel K. A., and Rekaya R. (2000) J Dairy Sci 83(5):1072-1080.

Wiggans G.R., Cole J. B., Hubbard S. M., and Sonstegard, T. S. (2017) Annu. Rev. Anim. Biosci. 5, 309-327.

697. Genetic parameters for residual feed efficiency in three dairy cattle breeds in commercial farms using 3D cameras

C.I.V. Manzanilla-Pech[1*], R.B. Stephansen[1], T. Andersen[2] and J. Lassen[1,3]

[1]Center for Quantitative Genetics and Genomics, Aarhus University, Blichers Alle 20, 8830 Tjele, Denmark; [2]SEGES Innovation P/S, Agro Food Park 15, 8200 Aarhus N, Denmark; [3]Viking Genetics, Ebeltoftvej 16, Assentoft, 8960 Randers, Denmark; coralia.manzanilla@qgg.au.dk

Abstract

This is a preliminary study on estimating genetic parameters for dry matter intake (DMI), BW and residual feed intake (RFI) using 3D cameras in 4,069 cows from three dairy breeds. Weekly records of DMI and BW were recorded during 2019-2021 for 1,643 Holstein, 1,211 Jersey and 1,215 Red Dairy Cattle (RDC) cows at Danish commercial farms. Estimated heritabilities for DMI ranged from 0.23 to 0.34, whereas for BW ranged from 0.38 to 0.49. Estimated heritabilities for RFI were 0.19 for Holstein, 0.15 for Jersey and 0.16 for RDC. Genetic correlations between DMI and BW were moderate positive (0.23-0.57), whereas, the genetic correlations between RFI and DMI were highly positive (0.76-0.98) and between RFI and BW were low positive (0.28) to close to zero. With these results we conclude that feed efficiency (RFI) calculated using DMI and BW measured by 3D cameras (and ECM) is heritable for the three dairy breeds.

Introduction

In the last decade, several countries have included feed efficiency in their breeding goal (Veerkamp *et al.* 2014; Pryce *et al.* 2015, Holstein USA, 2021). The Saved feed index in the Nordic Total Merit Index (NTM; NAV, 2020) now includes the breeding value for metabolic efficiency and maintenance (Lidauer *et al.*, 2019). Residual feed intake (RFI) has been proposed as proxy trait for feed efficiency in several species including cattle, pig and poultry (Martin *et al.* 2021). RFI is usually calculated as the partial regression of dry matter intake on energy sinks (energy-corrected milk; ECM, metabolic body weight; MBW, and BW change; Tempelman *et al.* 2015). As dry matter intake (DMI) and BW records are required to calculate RFI, new technologies are being developed to increase the number of individual records in an easier and more efficient way. In the last years, a 3D camera's technology to measure feed intake and body weight (BW) on individual cows at large scale has been developed. Using artificial intelligence and 3D cameras, the Cattle Feed InTake System (CFIT, Viking Genetics, 2020; Lassen *et al.* 2018) is an alternative tool to record individual feed intake and BW compared to scale based systems. The CFIT 3D cameras, located in barns, can identify individual cows and by using artificial intelligence algorithms, record individual DMI and BW for the entire herd. The squared correlation between traditional feed bins and 3D cameras is 0.9 (Lassen *et al.* manuscript submitted to WCGALP 2022). However, as with every new phenotype, the traits (DMI and BW) need to be quantified genetically. In this study, we estimated the genetic parameters including genetic correlations between DMI, BW (obtained by 3D cameras) and RFI from 4,069 Danish Holstein, Jersey and RDC commercial cows.

Materials & methods

Data description. The data included weekly records of DMI and BW from 4,069 Danish cows. Only data from 1st to 6th parity from the first 330 days in milk was utilized. The cows were recorded for DMI and BW using 3D cameras technology on 12 (4 per breed) commercial farms in Denmark during 2019-2021. The cows were fed with a total mixed ration diet that mainly consisted of maize silage, grass silage and concentrates. The cameras were located above the feeding area floor, and the cows were recorded when eating. An algorithm based on artificial intelligence identify the cows and translate their 3D images into phenotypes (DMI and BW). Lassen *et al.* (2018) have presented a description of the 3D cameras

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_697, © C.I.V. Manzanilla-Pech *et al.* 2022

methodology to measure feed intake. Body weight is also predicted using 3D images of the back of the cow, from these images, contours of the back were obtained. Using a PLS method, a prediction model was developed based on scale measures of the cows. Posteriorly DMI and BW weekly averages were calculated. Residual feed intake (RFI; Tempelman *et al.*, 2015) was the residual of the partial regression of DMI on metabolic BW (MBW), energy corrected milk (ECM) and body weight change (ΔBW). The model used to calculate RFI is the one used to calculate the breeding values for RFI by the Nordic Cattle Genetic Evaluation (NAV; Stephansen *et al.* 2021). Milk yield and content were available through the national milk recording system. Metabolic BW (MBW) was defined as $BW^{0.75}$. Body weight change (ΔBW) is described as change in kg per day. As the recording cameras were installed at different times at the different farms, there is a great variation in the number of records available per animal. These records are also at different times of the lactation. For this reason, a repeatability model was not considered at this time. Instead, averages per animal per lactation for DMI, BW and RFI were calculated. Given that only quarter of the animals have more than one lactation, only one lactation per animal was considered.

Statistical analyses. A univariate animal model was performed to estimate the variance and covariance components using DMU software (Madsen and Jensen 2014) for each trait. To estimate the genetic correlations, pairwise bivariate models between the three traits were fitted.

The general model used to estimate the variance components for DMI, BW, and RFI was:

$$y = Xb + Z_1a + e \quad (1)$$

Where y is the vector of phenotypes (DMI, BW, RFI); b represent the vector of fixed effects (herd-year; parity and age of cow); X is the incidence matrix relating observations with fixed effects; a is the vector of direct additive genetic effects; Z_1 is the incidence matrix relating observations with random genetic effects; and e is the vector of residual effects. Distributions of the random effects are $var(a) = A\sigma^2a$ where A is the pedigree relationship matrix, and $var(e) = I\sigma^2e$, where I is an identity matrix of an order equal to the number of observations and σ^2e is the residual variance. The pedigree for Holstein included 8,349 cows, whereas, for Jersey 5,405 cows and for RDC 7,963 cows.

Results

Descriptive statistics for DMI, BW, and RFI in Danish Holstein, Jersey and RDC cows are presented in Table 1. DM content was assumed to be 40%. The mean DMI was 27.9 kg/d in Holstein, 22.1 kg/d in Jersey and 27.6 kg/d in RDC. While, the mean BW was 680.6 kg for Holstein, 454.4 for Jersey and 658.2 for RDC. Genetic variances, heritabilities, genetic correlations, and phenotypic correlations of DMI, BW, ECM and RFI in Danish Holstein cows are shown in Table 2. Estimated heritabilities for DMI were 0.34, 0.23 and 0.30, for Holstein, Jersey and RDC. Whereas, for BW were 0.42, 0.38 and 0.49; and for RFI were 0.19, 0.15 and 0.16 for Holstein, Jersey and RDC respectively. Genetic correlations between DMI and BW ranged from 0.23 to 0.57 for the three breeds. RFI and BW showed low correlations (0.06 to 0.26) with large standard errors. Additionally, genetic correlations between DMI and RFI were highly positive for the three breeds (0.76 to 0.98).

Discussion

The average DMI for Holstein were slightly higher than previously reported (19.5 kg/d and 607 kg) for Danish Holstein cows by Difford *et al.* (2020). Additionally, Li *et al.* (2018) reported lower DMI and BW averages for the three breeds in primiparous Nordic cows measured within 32 weeks of lactation (Holstein= 19.4 kg/d and 601.2 kg; Jersey= 15.8 kg/d and 433.0 kg; RDC=18.5 kg/d and 575.1 kg, respectively).

Table 1. Descriptive statistics for dry matter intake (DMI), body weight (BW), energy corrected milk (ECM), residual feed intake (RFI) in Danish Holstein, Jersey and Red Danish (RDC) cows with number of animals in parenthesis.

	Holstein (n=1,643)				Jersey (n=1,211)				RDC (n=1,215)			
Trait	Mean	SD	Min	Max	Mean	SD[1]	Min	Max	Mean	SD	Min	Max
DMI	27.9	4.0	17.9	39.8	22.1	3.5	12.3	33.9	27.6	3.4	14.1	36.3
BW	680.6	76.3	449.9	907.3	454.4	40.6	351.3	568.2	658.2	76.0	428.0	889.2
RFI	0.0	1.2	-4.4	4.8	-0.1	2.0	-6.5	6.5	0.0	1.4	-5.3	4.8

Table 2. Genetic and phenotypic variances, heritabilities (diagonal), genetic correlations (lower diagonal) and phenotypic correlations (upper diagonal) of dry matter intake (DMI, g/d), body weight (BW, kg), residual feed intake (RFI, g/d) in Danish Holstein, Jersey and Red Danish (RDC) cows with standard errors in parenthesis.

Breed	Trait	σ^2_a	σ^2_e	DMI	BW	RFI
Holstein	DMI	1.6	3.0	**0.34 (0.06)**	0.49 (0.01)	0.72 (0.01)
	BW	1,352.7	1,857.8	0.57 (0.09)	**0.42 (0.06)**	0.07 (0.01)
	RFI	0.3	1.2	0.76 (0.07)	0.12 (0.15)	**0.19 (0.05)**
Jersey	DMI	1.8	5.9	**0.23 (0.07)**	0.24(0.02)	0.84 (0.01)
	BW	304.6	488.5	0.23 (0.17)	**0.38 (0.07)**	0.24 (0.02)
	RFI	0.6	3.5	0.98 (0.02)	0.06 (0.15)	**0.15 (0.05)**
RDC	DMI	1.8	4.1	**0.30 (0.07)**	0.42 (0.03)	0.67 (0.02)
	BW	1,732.4	1,782.1	0.50 (0.12)	**0.49 (0.07)**	0.06 (0.03)
	RFI	0.3	1.6	0.80 (0.09)	0.26 (0.18)	**0.16 (0.06)**

Genetic parameters and correlations. Estimated heritabilities for DMI (Table 2) using 3D cameras were within the range (depending on the lactation week) reported by Li *et al.* (2016, 2018) for the 3 breeds (0.30-0.55 for Holstein; 0.17-0.0.52 for Jersey; and 0.20-0.48 for RDC) in primiparous Nordic cows. Likewise, estimated heritabilities for BW were within the range reported by Li *et al.* (2018) for the 3 breeds, ranging from 0.49 to 0.63 in Holsteins, and from 0.46 to 0.61 in Jersey cows, whereas in RDC was from 0.32 to 0.53. Additionally, NAV (2021) reported heritabilities between 0.51 and 0.53 for second and third lactation Holsteins and Jersey cows. Furthermore, Difford *et al.* (2020) reported higher heritabilities for DMI (0.43), BW (0.52) and RFI (0.21) for Danish Holstein cows measured with conventional methods. Finally, heritabilities for RFI reported by Li *et al.* (2017) were similar (0.10-0.23) than the reported in this study (0.19) for Holstein. Moderate to high genetic correlations were estimated between DMI and BW, these values were within the range of values previously reported (0.3-0.7) for Danish Holstein and RDC cows across lactation stages with traditional methods (Li *et al.* 2018). Furthermore, Manzanilla-Pech *et al.* (2017) reported lower genetic correlations (0.43) between DMI and BW in Dutch Holstein cows. Genetic and phenotypic correlations between RFI and DMI were large and positive as expected given that RFI is the residual of DMI after been corrected by energy sinks. Thus, genetic and phenotypic correlations between RFI and BW were between low positive and close to zero (with large standard errors) as expected, as RFI has been corrected by BW. This study has showed that 3D phenotypes (DMI and BW) are heritable and their heritability is comparable to the heritabilities reported with conventional measuring methods. Furthermore, feed efficiency (RFI) calculated using DMI and BW measured by 3D cameras is also heritable. The results of this study are promising, endorsing a new technique of recording feed intake and body weight that has be implemented in commercial farms. We conclude that feed intake, efficiency and body weight measured by 3D cameras are heritable traits, endorsing a new technology of recording feed intake

and weight that has be implemented in commercial farms allowing to measure feed intake in a larger and efficient way.

References

Difford, G.F., P.Løvendahl, R.F.Veerkamp, H.Bovenhuis, M.H.P.W.Visker, *et al.* (2020). 103(3):2442-2459. https://doi.org/10.3168/jds.2019-16966

Holstein USA, 2021. TPI Formula. https://www.holsteinusa.com/genetic_evaluations/ss_tpi_formula.html

Lassen, J., J. R. Thomasen, R. H. Hansen, G. G. B. Nielsen, E. Olsen, P. R. B. Stentebjerg, *et al.* (2018). Proc. of the 11th WCGALP. Auckland, New Zeland.

Lassen J., Thomasen JR and S Borchersen (2022) WCGALP 2022

Li, B., W.F.Fikse, P.Løvendahl, J.Lassen, M.H.Lidauer, *et al.* (2016). J. Dairy Sci. 99(9): 7232-77239. https://doi.org/10.3168/jds.2015-10669

Li, B., W.F.Fikse, P.Løvendahl, J.Lassen, M.H.Lidauer, P.Mäntysaari, B.Berglund. (2018). J. Dairy Sci. 101(11): 10011-10021. https://doi.org/10.3168/jds.2018-14611

Li, B., B.Berglund, W.F.Fikse, J.Lassen, M.H.Lidauer, P.Mäntysaari, P.Løvendahl. (2017). J. Dairy Sci. 100(11): 9076-9084. https://doi.org/10.3168/jds.2017-12775

Lidauer, M.H., A. M. Leino, R. S. Stephansen, J. Poso, J, U. S. Nielsen, *et al.* 2019. Proc. of the 2019 Interbull meeting, Cincinnati, Ohio, USA June 22-24.

Madsen, P., and Jensen. (2014). Aarhus Res Cent. Foulum Box 50, 8830 Tjele Denmark.

Manzanilla-Pech, C.I.V., R. Veerkamp, Y. de Haas, M. Calus, and J. tenNapel. (2017). J. Dairy Sci. 100(11):9103-91014. https://doi.org/10.3168/jds.2017-12741

Martin, P., V. Ducrocq, P. Faverdin, N. C. Friggens. 2021. J. Dairy Sci. 106: 6329-6342.

NAV. (2020). https://www.nordicebv.info/wp-content/uploads/2020/11/Metabolic-efficiency-included-in-the-Saved-feed-index.pdf

NAV. (2021). https://nordicebv.info/wp-content/uploads/2021/10/NAV-routine-genetic-evaluation_EDITYSS-08102021.pdf

Pryce, J., O. Gonzalez-Recio, G. Nieuwhof, W.J. Wales, M.P. Coffey, *et al.* (2015) J. Dairy Sci. 98(10): 7340-73450. https://doi.org/10.3168/jds.2015-9621

Stephansen, R.V., M. H. Lidauer, U.S. Nielsen, J. Pösö, F. Fikse, *et al.* Proc. of the 2021 Interbull meeting.

Tempelman, R., D. Spurlock, M. P. Coffey, R. F. Veerkamp, L. Armentano, *et al.* (2015). J. Dairy Sci. 98(3):2013-2026. https://doi.org/10.3168/jds.2014.8510

Viking Genetics. (2020). https://www.vikinggenetics.com/about-us/innovative-breeding/ntm/saved-feed

Veerkamp, R.F., M.P.L. Calus, G. de Jong, R. van der Linde, Y. De Haas. (2014). Proc. of the 10th WCGALP, Vancouver, Canada.

698. Insemination values to support mating decisions under dairy heifer calf herd size constraints

A. De Vries[1*], N. Bliznyuk[1], P. Sharma[1], Y. Han[1] and P. Pinedo[2]

[1]University of Florida, P.O. Box 110910, Gainesville, FL 32611, USA; [2]Colorado State University, 307 University Ave, Fort Collins, CO 80523, USA; devries@ufl.edu

Abstract

Dairy producers have many factors to consider regarding insemination of dairy cattle, including semen type (sexed, conventional), semen breed (dairy, beef), price, and probability of conception based on sire and dam factors. Usually, the decisions are constrained by the need to produce a certain number of dairy heifer calves. To aid decision making, we calculated insemination values for each potential sire mated to every non-pregnant cow based on dynamic programming. We also calculated the probability of a live heifer calf born for each possible mating. Herd profitability is maximized by choosing per cow the sire with the greatest insemination value, subject to generating a target number of dairy heifer calves per batch of inseminations. We used a mixed integer linear program to make individual mating decisions. An illustration with 50 open cows showed opportunity cost from $3 to $10 per mating if matings are based on calf values instead of insemination values.

Introduction

Dairy producers have many options when it comes to mating their heifers and cows. In the last decade, the options have expanded to the availability of sexed dairy and beef semen from multiple sires. Combined with much better reproductive performance on dairy farms, dairy producers now also need to plan for the right size of the heifer inventory to replace cows to be culled and herd expansions.

The ability to accurately predict the probability that an insemination results in a pregnancy and subsequent calving is also increasing due to activity sensors, information contained in past health events, and genetic merit of reproduction traits. The mating decision further depends on the expected genetic merit of the calf, the level of inbreeding, and the price of semen. The best decision could be to delay insemination, do-not-breed, or replace the cow. Collectively, these many factors call for better guided mating decisions with the goal to increase profitability while meeting herd constraints such as a required number of dairy heifer calves.

Research conducted decades ago on optimization of replacement and insemination decisions (Van Arendonk, 1988; Mourits *et al.*, 1999) provides direction for the further development of insemination values to support mating decisions. The objective of our project is to develop and deploy insemination values for individual mating pairs subject to producing a target number of dairy heifer calves. If successful, this outcome is an important step towards solving a practical problem on many US dairy farms.

Materials & methods

Dynamic programming models. Economically optimal mating decisions result in the highest profit subject to constraints. Conceptually, this involves a comparison of future cash flows resulting from each decision and making the decision that results in the greatest net present value (NPV) of future cash flows. Cash flows can be optimized by also taking optimal decisions in the future, for example, by calculating when to optimally stop inseminating cows that have failed to get pregnant. Such decisions that span several points in time often break apart recursively. Dynamic programming (DP) is a mathematical optimization

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_698, © A. De Vries *et al.* 2022

method that can be used to simplify a decision by breaking it into simpler sub-problems in a recursive manner. See De Vries (2006a) and Nielsen and Kristensen (2015) for an overview in livestock.

For this project, we developed two related DP models: a herd DP model and an individual cow DP model. We developed a new herd DP model that used value-iteration to optimize replacement and insemination decisions for cows. The DP model had weekly steps, 10 lactations, 140 weeks in milk, 3 oestrus cycle states, 40 weeks of pregnancy, and 3 types of sires for matings (sexed dairy, conventional beef, conventional dairy), and 9 levels of economy-corrected milk. A bio-economic module included biological functions such as lactation curves for milk, fat and protein, dry matter intake, body weight, and probabilities of involuntary culling, oestrus detection, and conception. Prices were added and weekly revenues and costs were calculated per state. Transition probabilities between states were calculated based on biological functions and model decisions, for example when the cow was culled and replaced with a calving heifer. Starting from arbitrary starting conditions, the DP model ran until convergence was reached and steady-state results were obtained. Convergence criterion was the weekly change in NPV of the cash flow from a calving heifer. This change equals the expected weekly profit of the herd in steady state conditions.

Although this herd DP model contains optimal mating decisions for the states in the model that represent cows eligible for mating, these decisions were not used further. Instead, the model then was run for another 100 weeks while the NPV of cash flows of the first week in milk for each lactation and level of economy-corrected milk were saved. This 3-dimentional matrix of NPV of cash flows (100 weeks × 10 lactations × 9 levels of economy-corrected milk) was stored for use by the individual cow DP model, as follows.

Although herd DP models may contain millions of states, in practice some cows may not be well represented. Even in large herds, cows are unique and vary in many attributes relevant for future cash flow predictions, such as expected milk production, health risks, fertility, genetic merits of many traits, body weight, dry matter intake, etc.

The accuracy of cash flow predictions of individual cows can be improved by separating calculations for the current lactation from those for future lactations (De Vries, 2006b). The idea is to model the biological performance in the remainder of the current lactation for individual cows as accurately as possible and apply DP to only the current lactation for mating and replacement decisions. The NPV of cash flows after the cow calves again, or if she is replaced before calving again, are read from the stored 3-dimentional matrix using the most appropriate lactation number, economy corrected milk levels, and weeks into the future. This individual cow DP model is run separately for every cow in the herd for which mating decisions need to be made. This solution allows for every cow to have her own lactation curves, probabilities of conception and abortion, body weight, dry matter intake, health risks, genetic merit, etc. in the current lactation. It also allows for accurate valuing of the calf that follows from a mating.

Insemination values. An insemination value is the NPV of the cash flow following an insemination, minus the NPV of the cash flow when the decision is to delay the insemination (Mourits *et al.*, 1999). The decision to delay implies that the cow may be inseminated at a future insemination opportunity, for example in the next oestrus cycle, or be replaced with a calving heifer, as determined by the individual cow DP model. The decision to inseminate includes the cost of semen, the probabilities of conception, abortion and involuntary culling, and the probability a dairy heifer calf will be born whose value depends on the genetic merit of the dam and service sire.

An insemination value is calculated for every sire considered for a cow's mating. Sire attributes include semen price, relative probability of conception and abortion, genetic merit, and sex and breed of the calf. The DP models make the mating decision that results in the highest cash flow, which is equivalent to the decision associated with the highest insemination value. However, for actual decision support for cows that need a mating decision today, all insemination values are presented to the decision maker. In addition, the probability that a mating results in the birth of a live dairy heifer calf is also calculated.

Mixed integer linear program. Herd profitability is in theory maximized when the mating with the highest insemination value is chosen for each cow. However, this policy may result in more or fewer dairy heifer calves than are needed. The problem now is one of maximizing the sum of insemination values in a batch of cows, subject to the required number of dairy heifer calves. A simple mixed integer linear program solves this problem.

Illustration. We obtained cow records from a herd of 1,007 cows. Using default prices, we first ran the herd DP model and stored the 3-dimentional matrix of NPV of future cash flows. Next, we calculated 3 insemination values and three probabilities of a live dairy heifer calf for every non-pregnant cow, considering sexed, conventional, and beef semen. We also calculated the value of the calf for each mating pair. Finally, we determined optimal mating decisions for a batch of 50 cows eligible for insemination, subject to bounds on the number of dairy heifer calves.

Figure 1 shows the 150 insemination values for the 50 cows. The values ranged from -$34 to +$96. Unbound by the number of dairy heifer calves, the profit maximizing mating decisions resulted in a sum of insemination values of $2,430, equivalent to $48.50 per cow. This policy resulted in 4.57 expected dairy heifer calves. Table 1 shows the average insemination values if the policy is bound by a minimum and maximum number of dairy heifer calves. Bounds on the number of dairy heifer calves reduce the average insemination values. The opportunity cost shows the loss per mating if mating decisions were based on the highest calf values instead of the highest insemination values.

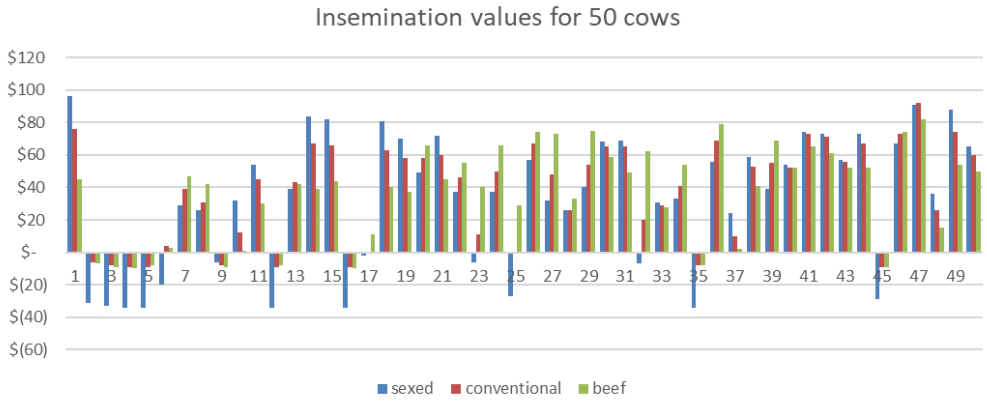


Figure 1. Insemination values for 50 cows. Each cow has 3 insemination values reflecting mating options with sexed dairy, conventional dairy, and conventional beef semen.

Table 1. Average insemination values, calf values and opportunity cost for matings of 50 cows bound by a window of the number of dairy heifer calves.

Mating	# dairy heifer calves			Inseminations (50 cows)				Average	Opportunity
criterion	min	max	actual	#sexed	#convent	#beef	#delay	ins. value[1] ($)	cost/insem. ($)
Ins. value	0	50	4.571	21	3	17	9	48.60	-
Ins. value	2	3	2.927	12	4	25	9	47.18	-
Ins. value	5	6	5.036	24	2	16	8	48.36	-
Ins. value	8	50	8.002	32	13	5	0	43.76	-
Calf value	0	50	8.959	45	0	5	0	38.14	10.46
Calf value	2	3	2.971	15	0	35	0	44.36	2.82
Calf value	5	6	5.987	30	0	20	0	43.92	4.44
Calf value	8	50	8.959	45	0	5	0	38.14	5.62

[1] Insemination value.

Discussion

The framework of insemination values combined with a mixed integer linear program offers a solution to the problem of determining optimal mating decisions given multiple sire choices and requirements on the number of dairy heifer calves. The application is fast and can in principle be applied on dairy farms. Many extensions and improvements are possible. For example, the probability of conception may be better predicted from the strength of the oestrus display (Cerri *et al.*, 2021) and health events (Pinedo *et al.*, 2020). A DP model for replacement and insemination decisions for heifers will be added soon.

Acknowledgements

This work is supported by Food and Agriculture Cyberinformatics and Tools grant no. 2019-67021-28823 from the USDA National Institute of Food and Agriculture.

References

Cerri R.L.A., Burnett T.A., Madureira A.M.L., Silper B.F. *et al.* (2021) J. Dairy Sci. 104:1220–1231. https://doi.org/10.3168/jds.2019-17893

De Vries A. (2006A) J. Dairy Sci. 89:3876-3885. https://doi.org/10.3168/jds.S0022-0302(06)72430-4

De Vries A. (2006B) Proc. of the 4th World Congress on Comp. in Agric. and Natural Resources. Orlando, FL, USA

Mourits M.C.M., Huirne R.B.M., Dijkhuizen A.A., Kristensen A.R., *et al.* (1999) Agric.Sys. 61: 17-31. https://doi.org/10.1016/S0308-521X(99)00029-3

Nielsen L.R. and Kristensen A.R. (2015) Markov decision processes to model livestock systems. In: L.M. Pla (ed.), Handbook of Operations Research in Agriculture and the Agri-Food Industry, International Series in Operations Research & Management Sci. 224

Pinedo P., Santos J.E.P., Chebel R.C., Galvão K.N. *et al.* (2020) J. Dairy Sci. 103:10560-10576. https://doi.org/10.3168/jds.2019-17951

Van Arendonk J.A.M. (1988). J. Dairy Sci. 71:1050-1057. https://doi.org/10.3168/jds.S0022-0302(88)79651-4

699. Onset of heat stress and development of genomic predictions for heat tolerance in US Holsteins and Jerseys

T.M. McWhorter[1,2*], M. Sargolzaei[2], C.G. Sattler[2], M.D. Utt[2], S. Tsuruta[1], I. Misztal[1] and D. Lourenco[1]

[1]*University of Georgia, Department of Animal and Dairy Science, 425 River Road, Athens, GA 30602, USA;* [2]*Select Sires, Inc.,11740 US-42, Plain City, OH 43064, USA; taylor.mcwhorter@uga.edu*

Abstract

The impact of heat stress on dairy production traits was investigated and genomic predictions for heat tolerance in US Holsteins and Jerseys were developed. Data included 12.8 and 2.1 million test-day records for milk, fat, and protein yield (kg) for the first five parities of 923,311 Holstein and 153,714 Jersey cows in 27 US states. Genotypes were available for 76,481 and 46,046 Holsteins and Jerseys, respectively. Production was affected by heat stress when the temperature-humidity index (THI) exceeded 69 for Holsteins and 72 for Jerseys. The additive genetic component of heat tolerance for production traits was estimated using a random regression repeatability model on a function of THI. Negative correlations between general and heat tolerance additive effects were observed for Holsteins (Jerseys): -0.38 (-0.37) milk, -0.51 (-0.49) fat, and -0.43 (-0.52) protein. Estimates for the heat tolerance additive genetic values were obtained with single-step GBLUP, potentially identifying heat-tolerant dairy cattle.

Introduction

High temperatures and relative humidity have negative impacts on production (Fuquay, 1981) and reproduction (Hansen and Aréchiga, 1999) of dairy cattle thus affecting overall economic return (Kadzere *et al.*, 2002; St-Pierre *et al.*, 2003). Cattle experience heat from both environmental factors including radiation, conduction, and convection as well as through metabolic processes such as exercise, growth, lactation, gestation, and feeding (Fuquay, 1981). Excess environmental heat unable to be managed through thermoregulation can result in heat stress. The stress can be mediated by provisions of cooling technologies such as sprinkler systems, fans, and shade. Additionally, developing a genetic evaluation for heat tolerance of dairy cattle will allow producers to select animals whose production is least affected by environmental heat stress. Different approaches have been previously explored with test-day yields in US Holsteins (Ravagnolo *et al.*, 2000; Bohmanova *et al.*, 2005, Aguilar *et al.*, 2009), non-return rate in US Holsteins (Ravagnolo *et al.*, 2002) and test-day yields in Australian Holsteins and Jerseys (Nguyen *et al.*, 2016). The objectives of this study were to investigate the impact of heat stress on test-day yields and develop single-step genomic predictions to identify heat tolerant US Holsteins and Jerseys.

Materials & methods

Test-day data. Select Sires, Inc. (SSI; Plain City, OH, USA) supplied 12.8 and 2.1 million test-day records from the first five parities containing milk, fat, and protein yield (kg) for 923,311 Holstein and 153,714 Jersey cows. Each cow was required to have at least 5 test-day records per lactation. Test-day records were obtained from 331 herds in 27 different US states from 2015 to 2021. Genotypes from different density arrays were available for 76,481 and 46,046 Holstein and Jersey animals, respectively. Genotypes were imputed to approximately 50k SNP density using FImpute3 software (Sargolzaei *et al.*, 2014).

Weather data. Weather data was obtained from Automated Surface Observing System (ASOS) stations at airports closest to each herd. Iowa State University's Iowa Environmental Mesonet (IEM) provides this data and can be accessed using R package 'riem' (Salmon, 2016). Geographical coordinates for each herd were obtained using R package 'ggmap' (Kahle and Wickham, 2013) and the distances between each herd

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_699, © T.M. McWhorter *et al.* 2022

and each airport were calculated using the Haversine formula in R package 'geosphere' (Hijmans, 2019). The Haversine formula determines the distance between two geographical coordinates on Earth using an average radius of 6,371 km. This measure was used to determine the airport with the minimum distance to each herd. Daily mean temperature and daily mean relative humidity for the 4-days prior to and the day of test (Nguyen *et al.*, 2016) were obtained for each airport and the daily temperature-humidity index (THI; NOAA, 1976) were calculated as proposed by Ravagnolo *et al.* (2000) as:

$$THI = (1.8 \times t + 32) - (0.55 - 0.0055 \times rh) \times (1.8 \times t - 26) \quad (1)$$

where t is the mean air temperature in degrees Celsius at approximately 2 meters, and rh is the mean relative humidity as a percentage. A final THI for each herd and test-day was calculated as the average THI over the 5 days.

Threshold. Statistical analyses were conducted separately based on the breed of the cow providing the test-day records. To determine the THI threshold where heat stress affects production traits, the least square estimates (LSE) of production were obtained from a repeatability multi-trait animal model as:

$$y_{hijklmno} = HTD_{ih} + (DIM \times s)_{jh} + lact_{kh} + state_{lh} + THI_{mh} + age_{nh} + a_{oh} + pe_{oh} + e_{hijklmno} \quad (2)$$

where $y_{hijklmno}$ is the yield of milk, fat, and protein in kg for trait h, categorical fixed effects included herd test-day i (HTD_{ih}), days-in-milk class by calving season class j $(DIM \times s)_{jh}$, lactation k ($lact_{kh}$), state l ($state_{lh}$), 5-day average THI of test-day m (THI_{mh}), and age of calving class n (age_{nh}). Random effects included general additive genetic effect for animal o (a_{oh}), general permanent environmental effect (pe_{oh}) for animal o, and residual ($e_{hijklmno}$). The LSE of production were regressed on THI. The discovered thresholds ($THI_{threshold}$) for each breed were used in a function, f(THI), and implemented in a random regression repeatability model to identify animals with superior genetic merit for heat tolerance:

$$f(THI) = maximum(0, THI_{TD} - THI_{threshold}) \quad (3)$$

where THI_{TD} is the calculated average THI of a given test-day.

Random regression. The multi-trait random regression repeatability model used for variance component estimation and genetic evaluation of test-day milk, fat, and protein yield (y) was:

$$y_{hijklno} = HTD_{ih} + (DIM \times s)_{jh} + lact_{kh} + state_{lh} + age_{nh} + a_{oh} + \alpha_{oh}[f(THI)] + pe_{oh} + \pi_{oh}[f(THI)] + e_{hijklno} \quad (4)$$

that included the additive genetic random regression effect of heat tolerance for animal o (α_{oh}) and the permanent environmental random regression effect of heat tolerance for animal o (π_{oh}). Additive and permanent environmental variances were assumed only to change with heat stress. Variance components were calculated using GIBBS2F90 and estimates for heat tolerance additive genetic values were calculated using BLUP90IOD2OMP1 as implemented in the BLUPF90 family of programs (Misztal *et al.*, 2014). The Algorithm for Proven and Young (APY; Misztal, 2016) was implemented with 11,000 Jersey and 15,000 Holstein core explaining >99% of variation.

Results

The average distance between herds and matched airports was 20 (12 SD) km with a minimum of 2 and maximum distance of 75 km. The 331 US herds averaged 44,913 (65,892 SD) test-day records but ranged from contributing 3 to 459,326 records. A substantial decrease in milk, fat, and protein yield for Holsteins

occurred after a THI of 69, whereas for Jerseys this value was 72. Approximately 14.6% Holstein and 9.9% Jersey test-day records were recorded at THI above the respective thresholds and therefore under heat stress. The maximum THI observed was 82. Heritability for general genetic merit of Holsteins was 20.8% for milk, 15.7% for fat, and 17% for protein, whereas Jerseys was found to be 20.0, 10.5 and 13%, in the same order. The heritability for production at each heat stress level is presented in Figure 1. A negative correlation between general and heat tolerance additive effects was observed for Holsteins (Jerseys): -0.38 (-0.37) milk, -0.51 (-0.49) fat, and -0.43 (-0.52) protein for all THI. Statistics of heat tolerance additive genetic values for milk, fat, and protein yield are in Table 1; these are the random regression coefficients and represent the expected change in yield (kg) from average change for each unit above THI threshold.

Discussion

Production yields decreased due to heat stress at different THI thresholds for each breed. It would be beneficial to have a greater number of test-days for cows experiencing heat stress. Heritability at each THI was greater than previously found in Ravagnolo *et al.* (2000) where first-parity Holstein milk yield was 16% at 72 THI and increased to 21% at 85 THI and little to no increase in fat and protein. In this study, a greater increase in heritability was observed from 20% at 70 THI to 35% at 82 THI when all five lactations were included. Heritability of protein increased from 16 to 31% from THI 70 to 82; thus, selection for heat tolerance for production traits is possible. Small increases in heritability at extreme THI could be due to artifacts of the random regression model and a need for more records under heat stress. The negative correlations between general and heat tolerance additive effects for production traits indicate an antagonistic relationship. High-producing cows are expected to have a lower heat tolerance. However, the

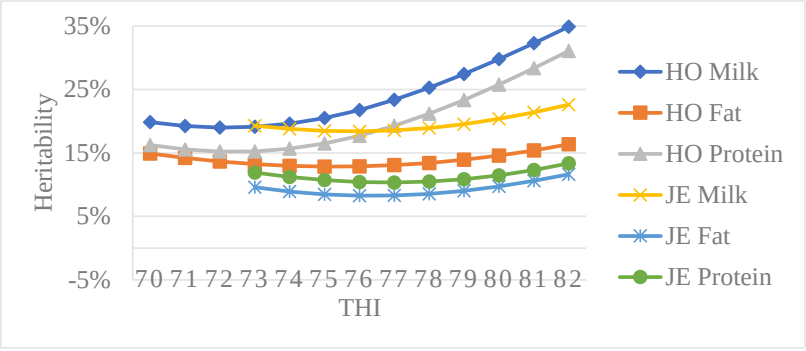


Figure 1. Heritability for production under heat stress for Holsteins (HO) and Jerseys (JE).

Table 1. Heat tolerance additive genetic values[1] for milk, fat, and protein yield for 1,189,386 Holstein and 240,813 Jersey animals.

Breed	Trait	Mean	Min	Max	SD
Holstein	Milk	0.033	-1.132	1.384	0.169
	Fat	0.001	-0.028	0.034	0.005
	Protein	0.001	-0.030	0.041	0.005
Jersey	Milk	-0.009	-0.516	0.475	0.086
	Fat	-0.001	-0.023	0.023	0.004
	Protein	0.000	-0.016	0.017	0.003

[1] Interpreted as the expected change in yield (kg) from average change for each unit above THI threshold.

low to moderate correlations and increasing heritability for heat tolerance over THI show potential to improve both production traits and heat tolerance simultaneously using a selection index. Estimates of heat tolerance additive genetic values for milk, fat, and protein yield were obtained with single-step GBLUP, potentially allowing the identification of heat-tolerant Holsteins and Jerseys. A validation on the genetic predictions will be explored. Production for first lactation cows has been observed to be least affected by heat stress. Further analysis of models containing first lactation only, and later lactations (lactations 2-5) will be explored.

References

Aguilar I., Misztal I., and Tsuruta S. (2009) J Dairy Sci 92(11):5702-5711. https://doi.org/10.3168/jds.2008-1928.

Fuquay J.W. (1978) J Anim Sci 52(1):164-174. https://doi.org/10.2527/jas1981.521164x.

Hansen P.J., and Aréchiga C.F. (1999) J Anim Sci 77(2):36-50. https://doi.org/10.2527/1997.77suppl_236x.

Hijmans R.J. (2019) geosphere: Spherical Trigonometry. R package version 1.5-10. Available at: https://CRAN.R-project.org/package=geosphere.

Iowa Environmental Mesonet (IEM), Iowa State University 2021. Available at: https://mesonet.agron.iastate.edu/.

Kadzere C.T., Murphy M.R., Silanikove, N., and Maltz E. (2002) Liv Prod Sci 77(1):59-91. https://doi.org/10.1016/S0301-6226(01)00330-X.

Kahle D., and Wickham H. (2013) R Journal 5(1):144-161. https://doi.org/10.32614/RJ-2013-014.

Misztal I. (2016) Genetics 202(2):401-409. https://doi.org/10.1534/genetics.115.182089.

Misztal I., Tsuruta S., Lourenco D., Aguilar I., and Legarra A., *et al.* (2014) Manual for BLUPF90 family of programs. Athens: University of Georgia, USA.

NOAA (1976) Livestock hot weather stress. Regional operations manual letter C-31-76. US Dept Commerce, Natl Oceanic and Atmospheric Admin.

Nguyen T.T.T., Bowman P.J., Haile-Mariam M., Pryce J., and Hayes B.J. (2016) J Dairy Sci 99(4):2849-2862. https://doi.org/10.3168/jds.2015-9685.

Ravagnolo O., and Misztal I. (2002) J Dairy Sci 85(11):3092-3100. https://doi.org/10.3168/jds.S0022-0302(02)74396-8.

Ravagnolo O., Misztal I., and Hoogenboom G. (2000) J Dairy Sci 83(9):2120-2125. https://doi.org/10.3168/jds.S0022-0302(00)75094-6.

Ravangolo O., and Misztal I. (2000) J Dairy Sci 83(9):2126-2130. https://doi.org/10.3168/jds.S0022-0302(00)75095-8.

Salmon M. (2016) riem: Accesses Weather Data from the Iowa Environmental Mesonet. R package version 0.1.1. Available at: https://CRAN.R-project.org/package=riem.

Sargolzaei, M., J., Chesnais J.P., and Schenkel, F.S. (2014). BMC Genomics 15:478. https://doi.org/10.1186/1471-2164-15-478.

St-Pierre N.R., Cobanov B., and Schnitkey G. (2003) J Dairy Sci 86:E52-E77. https://doi.org/10.3168/jds.S0022-0302(03)74040-5.

700. Detection of genetic variability in cattle infectivity for bovine tuberculosis (bTB)

E. Sánchez-Molano[1*], D. Madenci[1], S. Tsairidou[1], M. Winters[2], A.P. Mitchell[3], G. Banos[1,4] and A. Doeschl-Wilson[1]

[1]The Roslin Institute and R(D)SVS, Easter Bush Campus, Midlothian EH25 9RG, Edinburgh, United Kingdom; [2]Agriculture and Horticulture Development Board, Kenilworth, Warwickshire CV8 2TL, United Kingdom; [3]Animal and Plant Health Agency, Surrey KT15 3NB, United Kingdom; [4]Scotland's Rural College, Easter Bush Campus, Midlothian EH25 9RG, Edinburgh, United Kingdom; enrique.sanchez-molano@roslin.ed.ac.uk

Abstract

Bovine tuberculosis (bTB) is a zoonotic disease primarily affecting cattle and leading to important economic losses. Eradication programmes in various countries focus on surveillance, movement restrictions and culling, with some considering also selective breeding. These breeding strategies focus on reducing animal susceptibility to the disease and thus, the potential of reducing disease transmission by selecting also animals with low infectivity remains unknown. The first step in assessing this potential is to detect underlying genetic variation in infectivity. In the present study, the number of secondary cases attributed to a given index case in a breakdown was used as an infectivity phenotype, and a range of linear and generalised linear mixed sire models were used to estimate its genetic variance and heritability. Substantial genetic variance was detected for this trait, suggesting that infectivity can be reduced via selective breeding, hence contributing to more efficient eradication strategies.

Introduction

Bovine tuberculosis (bTB) is a chronic disease caused by *Mycobacterium bovis*, a Gram-positive bacterium of the *Mycobacteriaceae* family. Although mainly found in cattle, it also affects many other mammals, including humans (Michel *et al.* 2010; Payeur 2014). In the UK, bTB is one of the major causes of economic losses in cattle industry, due to the costs of the testing programme and the compensation payments to farmers when animals are culled (Abernethy *et al.* 2013).

Given the costs associated with this disease, several programmes across different countries were set up to eradicate rather than to control bTB. These are mainly based on intensive test and culling regimes, and stringent biosecurity measures to limit the spread of the disease. Although in many cases these strategies reported successful results, in others, such as the UK, the disease has proven difficult to eradicate with these measures alone (Allen *et al.* 2018). Given this, the TB Advantage genetic index was introduced as a complementary strategy in the UK in 2016 (Banos *et al.* 2017). This index indicates the degree of resistance to bTB that a bull is expected to pass on to his offspring and thus, allowing to reduce bTB susceptibility via genetic selection.

There is evidence for the presence of bTB super-spreaders, i.e. highly infectious animals responsible for transmitting the disease to many herd-members (O'Hare *et al.* 2014). Hence, a promising additional strategy to eradicate the disease may also be to consider infectivity as part of the selection index, allowing for the reduction of the propensity of infected animals to transmit the infection. Previous studies have demonstrated that including infectivity in the selection index could drastically accelerate reduction in risk and prevalence of disease (Tsairidou *et al.* 2018), however the genetic basis of bTB infectivity is still unknown. Therefore, the purpose of this preliminary study was to assess the existence of underlying genetic variation for infectivity using real data from UK bTB breakdowns.

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_700, © E. Sánchez-Molano *et al.* 2022

Materials & methods

Breakdown data. Repeated bTB skin test records collected in 60-day intervals, from all animals in a herd, for 32,745 breakdowns, were available for the period between 1995 to 2017. From these data, only closed breakdowns were considered, and the first animal with a positive test result (reactor) or slaughtered after an inconclusive test (inconclusive reactor) was identified as a potential index case. An index case is defined as the first infected animal in a herd that transmits infection to its herd members. In this study, breakdowns with more than one index cases (reactors detected on the same first test date) were discarded. After quality control, the total number of breakdowns retained for analysis was 4,598, with each breakdown corresponding to one given index case.

Index cases with high infectivity would be expected to generate, on average, more secondary cases than index cases with low infectivity. Hence, for each breakdown, the response variable (numR) was calculated as the total number of secondary cases (non-index case reactors or slaughtered due to inconclusive test results) identified in a given period since the start of the breakdown (either 120 or 365 days). Based on previous estimates of the disease stage durations (O'Hare *et al.* 2021), the chosen periods were assumed to be sufficiently long for secondary cases to be detectable and sufficiently short to minimise bTB cases resulting from transmission between secondary cases.

Pedigree data. A full pedigree consisting of 910,170 records was available. From this data, two shorter pedigrees corresponding to the 4,598 index cases described above were extracted, comprising the last seven generations (58,753 records). These index cases corresponded to 2,544 different sires, of which 80.7% (2,054) were sires of at least two index cases.

Data analysis. The above data was analysed with sire models to estimate the genetic effect of the index cases on the total number of secondary cases per breakdown (numR) identified in either the first 120 or 365 days after the breakdown start date. A sire would be associated to index cases in different breakdowns, thus expected to account for potential across-herd genetic variation in susceptibility of the infected animals. Different distributions of the response variable were examined, thus estimating genetic variances and heritabilities in the latent scale.

Initial exploratory analyses were performed only for the 365 days period using ASReml (Gilmour *et al.* 2015), without including any fixed effects. However, given the skewed distribution of the number of secondary cases and the high proportion of breakdowns with zero secondary cases, the response variable was transformed to log(numR+1), which was assumed to be normally distributed.

A further analysis was performed through a generalised linear mixed model accounting for an over-dispersed Poisson distribution (de Villemereuil *et al.* 2016) of the response variable using ASReml (Gilmour *et al.* 2015). In this case, the response variable was transformed to numR+1, and the following covariates and fixed effects were included in the analysis: index case age in months at test, maximum herd size during breakdown, risk code (type of risk area), season at onset of breakdown and breakdown duration (in days).

One potential issue of using Poisson distributions (even if accounting for over-dispersion) is the fact that they account poorly for zero inflation. Therefore, additional distributions for the response variable were considered in the generalised linear mixed models.

A negative binomial distribution was firstly assumed, with the number of successes (φ) being 1 and, therefore, effectively resembling a geometric distribution. This analysis was done in ASReml, with the

response variable being transformed to numR+1, given the logarithmic link function and using all the previously described fixed effects and covariates.

An additional two-step model was also applied to the data, following the rationale of a Hurdle model (Mullahy 1986). In this analysis, after accounting for the above fixed effects and covariates, the data was assumed to follow a mixture of a Binomial and Poisson distributions. The response variable (numR) was firstly analysed in ASReml as a binary variable (presence or absence of secondary cases) and the probability of non-zero values (breakdowns with numR≥1) was analysed following an over-dispersed Poisson model in ASReml.

A final model was tested using Markov Chain Monte Carlo techniques via the MCMCglmm package in R (Hadfield 2010). In this model, the response variable (numR) was assumed to follow a zero-inflated Poisson distribution (ZIP), and all previous fixed effects and covariates were included. The model was run for 200,000 iterations, plus 10,000 burn-in, with a thinning interval of 100. The prior assumed independent residual and random effect variances for the count and binary terms.

Results

The distribution of the number of secondary cases per breakdown in the first 365 days (Figure 1) was skewed, presenting an excess of zeros, and with a mean of 2.23 and variance of 34.37. About 0.8% of the breakdowns presented more than 30 secondary cases during this interval, with the maximum being one breakdown with 108 secondary cases.

Results for the estimates of genetic variance and heritabilities on the latent scale are presented in Table 1 for each of the two intervals considered. Heritability and genetic variances for the binomial part of the two-step models were not estimable.

Discussion

Under the assumption of a normal distribution for the transformed response variable, none of the analyses resulted in a significant genetic variance estimate. This was clearly due to the lack of normality in the response variable and the high frequency of zero secondary cases. Similarly, the use of generalised linear mixed models assuming over-dispersed Poisson distributions resulted in significant but overestimated heritabilities, likely due to the lack of concordance of these distributions with the observed in the data.

Both the negative binomial or the zero-inflated Poisson distribution models derived significant and relatively reliable heritability estimates; these estimates are in line with those observed for resistance to bTB in previous studies (Banos *et al.* 2017) as well as with those expected for other fitness-related traits (Gibson and Dechow 2018).

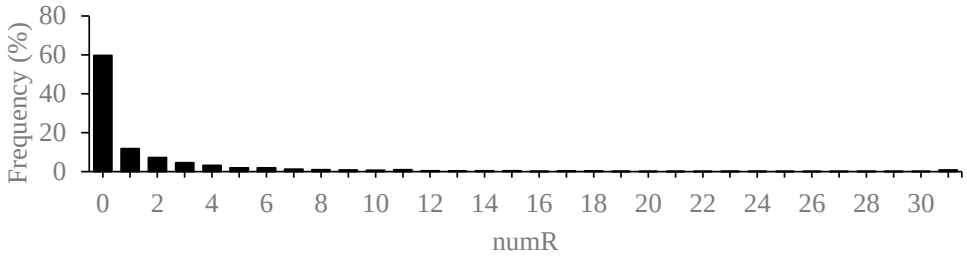


Figure 1. Distribution of the average count of secondary cases per sire.

Table 1. Heritabilities (h^2) and genetic variances (σ_a^2) in the latent scale for the different models and periods.

Distrib.[1]	Response variable	365 days		120 days	
		σ_a^2	h^2	σ_a^2	h^2
Normal	log(numR+1)	0.004±0.017	0.006±0.021	-	-
OP	log(numR+1)	1.657±0.107	0.599±0.038	1.169±0.085	0.674±0.047
NB	log(numR+1)	0.094±0.048	0.144±0.072	0.151±0.060	0.214±0.083
TS-P	numR≥1	1.582±0.154	0.408±0.044	1.475±0.161	0.591±0.068
ZIP	numR	0.157±0.004	0.066±0.002	0.188±0.004	0.086±0.002

[1] Studied distributions were: Normal, Over-dispersed Poisson (OP), Negative Binomial (NB), Poisson part of the two-step model (TS-P) and Zero-inflated Poisson (ZIP).

These results showed consistently higher heritabilities for secondary cases in the first 120 days than for the first 365 days. This was probably the consequence of the effects of the increased strength of the genetic component (the smaller the timeframe, the more likely that the infection of secondary case was transmitted by the index case) and the lower total phenotypic variance.

This study presents a simple approach to estimate the genetic component and heritability of bTB infectivity, resulting in low but substantial significant heritabilities for the studied trait. However, further research is required to validate these results. Given the length of the different disease stages reported in recent studies (O'Hare *et al.* 2021), a more accurate definition of secondary cases attributed to the index case is needed. Furthermore, additional models must be tested and used to take advantage of the data available from breakdowns with more than one index cases and to incorporate transmission from all infected cows.

References

Abernethy D.A., Upton P., Higgins I.M., McGrath G., Goodchild A.V., *et al.* (2013) Vet Record 172: 312-312. https://doi.org/10.1136/vr.100969.

Allen A.R., Skuce R.A., and Byrne A.W. (2018) Front Vet Sci 5:109. https://doi.org/10.3389/fvets.2018.00109.

Banos G., Winters M., Mrode R., Mitchell A.P., Bishop S.C., *et al.* (2017) J Dairy Sci 100:1272–1281. https://doi.org/10.3168/jds.2016-11897.

de Villemereuil P., Schielzeth H., Nakagawa S., and Morrissey M. (2016). Genetics, 204(3):1281–1294. https://doi.org/10.1534/genetics.115.186536.

Gibson K.D. and Dechow C.D. (2018) J Dairy Sci 101:1251–1257. https://doi.org/10.3168/jds.2017-13041.

Gilmour A.R., Gogel B.J., Cullis B.R., Welham S.J., and Thompson R. (2015). ASReml User Guide Release 4.1 Functional Specification. VSN International Ltd, Hemel Hempstead, UK.

Hadfield J.D. (2010) J Stat Softw 33(2):1-22. https://doi.org/10.18637/jss.v033.i02

Michel A.L., Muller B., and van Helden P.D. (2010) Vet Microbiol 140:371–81. https://doi.org/10.1016/j.vetmic.2009.08.029.

Mullahy J. (1986) J Econom 33:341–365. https://doi.org/10.1016/0304-4076(86)90002-3.

O'Hare A., Orton R.J., Bessell P.R., and Kao R.R. (2014) Proc Biol Sci 281(1783): 20140248. https://doi.org/10.1098/rspb.2014.0248

O'Hare A., Balaz D., Wright D.M., McCormick C., McDowell S., *et al.* (2021) PLoS Comput Biol 17(6): e1009005. https://doi.org/10.1371/journal.pcbi.1009005.

Payeur J.B. (2014). Mycobacterium. In: Batt C.A. and Tortorello M.L. (eds) Encyclopedia of Food Microbiology. Academic Press, Cambridge, US, pp. 841-853.

Tsairidou S., Allen A., Banos G., Coffey M., Anacleto O., *et al.* (2018) Front Vet Sci 5:310. https://doi.org/10.3389/fvets.2018.00310.

701. Dairy female fertility traits in seasonal and non-seasonal herds in Ireland

K. Stachowicz[1*], P.R. Amer[1], M.M. Kelleher[2], S.C. Ring[2] and R.D. Evans[2]

[1]AbacusBio Limited, P.O. Box 5585, Dunedin 9058, New Zealand; [2]The Irish Cattle Breeding Federation, Highfield House, Bandon, Co. Cork, Ireland; kstachowicz@abacusbio.co.nz

Abstract

In Ireland, the genetic trait representing dairy female fertility, namely calving interval, has high emphasis in the selection index (EBI), but recent validation testing has raised concerns about its predictive effectiveness deteriorating in recent years (Ring *et al.*, 2021). The goal of this work was to evaluate a wide range of possible fertility trait definitions in order to find ones that best fit seasonal and non-seasonal herds in Ireland. Phenotypes for over 40 traits were evaluated. Heritabilities were estimated as well as genetic correlations for the same traits between seasonal and non-seasonal systems. Based on these results a sample of the most promising traits were selected and genetic correlations were estimated between those traits within and between calving systems. Finally, recommendations were made for prototype genetic evaluation systems to be tested and validated against the current genetic evaluation system.

Introduction

In Ireland, the proportion of herds that operate a seasonal calving pattern has been systematically increasing in the last 20 years from around 50% in the early 2000s to over 70% in 2021. Seasonal calving reflects a high proportion of calvings occurring within the first four months from the beginning of the calving season. Despite this fact the genetic evaluation system for fertility and survival traits continues to be the one developed when less compact calving patterns were prevalent. Fertility and survival traits are evaluated jointly in a multiple trait animal model which includes 23 traits (Berry *et al.*, 2013). The main fertility trait is calving interval (CINT), with calving to first service, number of services (NS) and age at first calving (AFC) as predictor traits. The use of CINT is consistent with the fertility trait definition used in many countries with predominantly all year-round calving systems. It is also a trait widely accepted by the scientific community as providing an accurate reflection of genetic merit for fertility. However, CINT is not the optimal selection criterion for fertility in seasonal systems, as the goal in seasonal herds is not to reduce the CINT, but rather to maintain a compact calving pattern. For example, early calvers which are likely the more fertile cows often have longer CINTs compared to their later calving contemporaries because they are withheld from mating, to avoid them calving too early in the next calving season.

The objective of this work was to investigate a wide range of possible fertility traits definitions to find ones that best fit seasonal and non-seasonal herds in Ireland. Phenotypes for over 40 traits were created and assessed for differences between seasonal and non-seasonal herds. Heritabilities were estimated as well as genetic correlations for the same traits between seasonal and non-seasonal systems. Based on these results a sample of the most promising traits were selected and genetic correlations were estimated between those traits within and between calving systems. Finally, recommendations were made for prototype genetic evaluations systems to be tested and validated against the current genetic evaluation system.

Materials & methods

Phenotypic data were obtained from the Irish Cattle Breeding Federation database (www.icbf.com). The criteria used to differentiate seasonal from non-seasonal herds has been previously described by Stachowicz *et al.* (2018). Data was filtered at the individual cow level to include only dairy animals with both parents known, and only cows which did not move herd in their lifetime. Data was filtered at the herd level to

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_701, © K. Stachowicz *et al.* 2022

ensure reasonable size, consistent seasonal or non-seasonal data across the 2010 to 2018 seasons, and herds that record matings systematically.

Definitions of fertility and survival trait phenotypes are presented in Table 1. Traits more suited for non-seasonal calving system were derived based on mating and calving dates, and intervals between them. Traits more suited to seasonal calving were derived based on intervals between mating or calving dates and planned start of mating (PSM) or calving (PSC) dates for the contemporary group. Definitions of PSM and PSC were as described by Stachowicz *et al.* (2018). Using PSM and PSC dates as the reference points for seasonal fertility traits accounts for the involuntary withholding period. CINT preadjusted for involuntary withholding period (CINT*) was introduced, which is the standard CINT trait after pre-adjustment for the time between the first calving within the herd-season and PSM date (i.e. the withholding period). In addition to cow survival as in the current evaluation (Berry *et al.*, 2013) three heifer survival traits were tested.

Variance components were estimated using animal models in ASReml (Gilmour *et al.*, 2015) Fixed effects and covariates included in the models were: herd-year for seasonal herds and herd-year-calving season (2 seasons per year) for non-seasonal herds; month of birth for heifer traits, month of calving for cow traits; linear and quadratic effects of age at previous calving; mating type for mating traits; breed as a categorical fixed effect with 4 levels: Friesian, Holstein, Cross between Friesian and Holstein and Other Dairy; breed specific heterosis; recombination loss. Cow traits were modelled using repeated records up to 6 parities.

Results & discussion

Heritabilities and genetic correlations estimated for and between all the traits are presented in Table 1. Inflated heritabilities were found for birth to mating and birth to calving survival traits in non-seasonal herds, possibly due to farmers being less likely to keep heifers from lower index sires through to mating. The binary trait, of whether or not the heifer calved following a recorded mating showed reasonable and consistent heritability across herd types and should be pursued further. However, the estimate of the genetic correlation for this trait between seasonal and non-seasonal herds is low. Age at first service and AFC had lower heritabilities in seasonal herds. Heifer traits based on reference to mating or calving planned start dates were more promising in seasonal herds (TCCH, PR21H, PM21H and CSDH) with h^2 ranging as expected from 0.02 to 0.03. CINT had almost zero heritability in seasonal herds, which confirms that it is not a trait suitable for seasonal calving systems. CINT* (adjusted for withholding period) recovers some heritability in seasonal herds and might be a viable alternative to CINT whereby a single trait is evaluated, with pre-adjustment in herds defined as seasonal, and no adjustment in herds defined as non-seasonal. Pregnant to first mating, Pregnant to any mating, and the interval traits have low h^2 across both herd types. Calving to first mating (and CINT) is an exception in non-seasonal herds. The conception-based traits look comparable to calving based traits in heritability. They have the advantage of being independent from gestation length, and potential to get data into a genetic evaluation earlier in the season. Three/six week in calf rate could be a good early predictor of fertility, as NS, which is currently used is censored for cows still available to be mated, whereas not getting in-calf on the first or second service is the key economic driver. The continuous timing of conception trait (TCC) has higher h^2 than binary pregnancy rates, but binary traits are more practical to use as they avoid censoring.

Based on these results the traits that are included in Table 2 were chosen for further analysis. Genetic correlations between chosen traits within and between seasonal and non-seasonal herds are presented in Table 2. The genetic correlation between CINT in non-seasonal herds and CINT* (adjusted for voluntary withholding period) in seasonal herds was 0.99. The genetic correlation between 6-week in calf rate for cows (pregnant in 1st 21 days from 1st herd-year mating) in non-seasonal herds with 6-week in calf rate

Table 1. Fertility trait definitions, heritabilities and genetic correlations between seasonal and non-seasonal herds (H=heifer, C=cow, B=binary, T=continuous, X=categorical).

Trait	Trait definition[1]	h^2 in non-seasonal herds	h^2 in seasonal herds	r_G between seasonal and non-seasonal herds
Mated	Survival from birth to mating (H, B)	0.048	0.235	
Calved	Survival from birth to calving (H, B)	0.062	0.162	
MatCalv	Survival from mating to calving (H, B)	0.032	0.021	0.29
AFS	Age at first service (T)	0.097	0.040	-0.19
AFC	Age at first calving (T)	0.097	0.012	0.40
PR1MH	Pregnant to first mating (H, B)	0.000	0.013	
PR1MC	Pregnant to first mating (C, B)	0.015	0.013	
PRAMH	Pregnant to any mating (H, B)	0.014	0.019	
PRAMC	Pregnant to any mating (C, B)	0.040	0.016	
FMTCH	First mating to calving (H, T)	0.006	0.000	
FMTCC	First mating to calving (C, T)	0.020	0.005	0.52
FSTCH	First service to conception (H, T)	0.006	0.000	
FSTCC	First service to conception (C, T)	0.026	0.003	0.66
FSTC21H	Pregnant in 1st 21 days from 1st mating (H, B)	0.003	0.014	0.30
FSTC21C	Pregnant in 1st 21 days from 1st mating (C, B)	0.007	0.005	0.72
FSTC42H	Pregnant in 1st 42 days from 1st mating (H, B)	0.003	0.000	-0.13
FSTC42C	Pregnant in 1st 42 days from 1st mating (C, B)	0.020	0.003	0.54
CTFM	Calving to first mating (T)	0.044	0.007	0.54
NSH	Number of services (H, X)	0.034	0.032	-0.01
NSC	Number of services (C, X)	0.024	0.016	0.98
CINT	Calving interval (T)	0.033	0.002	0.52
PMCH	Days from PSM to first mating (H, T)	0.070	0.038	0.04
PMCC	Days from PSM to first mating (C, T)	0.040	0.019	0.95
PM21H	Presented for mating in 1st 21 days from PSM (H, B)	0.019	0.018	0.74
PM21C	Presented for mating in 1st 21 days from PSM (C, B)	0.027	0.031	0.50
TCCH	Days from PSM to conception (H, T)	0.071	0.033	0.33
TCCC	Days from PSM to conception (C, T)	0.057	0.023	0.88
PR21H	Pregnant in 1st 21 days from PSM[1] (H, B)	0.026	0.021	-0.54
PR21C	Pregnant in 1st 21 days from PSM (C, B)	0.015	0.017	0.64
PR42H	Pregnant in 1st 42 days from PSM (H, B)	0.024	0.021	-0.62
PR42C	Pregnant in 1st 42 days from PSM (C, B)	0.023	0.026	0.65
CSDH	Days from PSC to calving (H, T)	0.108	0.026	0.14
CSDC	Days from PSC to calving (C, T)	0.032	0.026	0.81
CR21H	Calved in the 1st 21 days from PSC (H, B)	0.093	0.018	-0.01
CR21C	Calved in the 1st 21 days from PSC (C, B)	0.019	0.027	0.64
CR42H	Calved in the 1st 42 days from PSC (H, B)	0.065	0.010	0.30
CR42C	Calved in the 1st 42 days from PSC (C, B)	0.027	0.020	0.30
CINT*	Calving interval adjusted for involuntary withholding (T)	0.038	0.020	0.93
SURV	Survival to the next calving (B)	0.045	0.021	0.66

[1] PSC = planned start of calving date for contemporary group; PSM = planned start of mating date for contemporary group.

Table 2. Genetic correlations between chosen traits within and between seasonal and non-seasonal herds.

		Seasonal				Non-seasonal				
		CSDH	**CINT***	**PR42C**	**SURV**	**MatCalv**	**AFC**	**CINT**	**FSTC42C**	**SURV**
Seasonal	MatCalv	-0.82	-0.24	0.37	0.59	0.29	-0.55	-0.31	0.18	0.20
	CSDH		0.62	-0.46	-0.43	0.15	0.58	0.15	-0.01	0.27
	CINT*			-0.90	-0.86	0.29	-0.02	0.99	-0.78	-0.96
	PR42C				0.95	-0.24	-0.14	-0.84	0.92	0.77
	SURV					0.09	-0.22	-0.74	0.55	0.66
Non-seasonal	MatCalv						-0.03	-0.05	0.62	0.42
	AFC							0.25	0.09	-0.15
	CINT								-0.90	-0.74
	FSTC42C									0.44

for cows (pregnant in 1st 21 days from PSM) in seasonal herds was also found to be very high (0.92). This indicates that both CINT and 6-week in calf rate, although defined differently in the two herd types, could be evaluated as one trait in a genetic evaluation system. A similar recommendation of combining traits across systems was made for survival despite the lower genetic correlation between seasonal and non-seasonal herds of 0.66. It is anticipated that further enhancement of contemporary group definitions for survival in both herd types as well as heterogenous residual variances could improve this correlation. Heifer survival from mating to calving, if included in a genetic evaluation, would need to be a separate trait for seasonal and non-seasonal herds due to the low genetic correlation and differences in variances. Additionally, AFC in non-seasonal herds and calving season day (CSDH) for heifers in seasonal herds were recommended for consideration when building a new genetic evaluation. AFC is currently utilised as an early predictor of fertility. CSDH would be a novel trait for Ireland but was recently introduced into fertility evaluation in New Zealand seasonal system (Stachowicz *et al.*, 2021).

References

Berry, D.P., Kearney, J.F., Twomey, K. and Evans, R.D. (2013). Irish J. Agric. Food Res. 52: 1-16.

Gilmour, A.R., Gogel, B.J., Cullis, B.R., Welham, S.J. and Thompson, R. (2015). ASReml User Guide Release 4.1 Structural Specification, VSN International Ltd, Hemel Hempstead, UK.

Ring, S.C., Evans, R.D., Cromie, A.R. and Berry, D.P. (2021) J. Dairy Sci. 104: 8076-8093.

Stachowicz, K., Jenkins, G.M., Amer, P.R., Berry, D.P., Kelleher, M.M., *et al.* (2018). Interbull Bull. 53:57-63.

Stachowicz, K., Amer, P.R., Falconer, B. and D.P. Garrick. (2021). Proc. Assoc. Advmt. Anim. Breed. Genet. 24:370-373.

702. Genetic aspects of age at first calving in Italian Holstein dairy cows

V. Ferrari[1,2*], F. Galluzzo[1], R. Finocchiaro[1], J.B.C.H.M. van Kaam[1], M. Marusi[1], M. Penasa[2] and M. Cassandro[1,2]

[1]Associazione Nazionale Allevatori della Razza Frisona, Bruna e Jersey Italiana (ANAFIBJ), Via Bergamo 292, 26100 Cremona, Italy; [2]Department of Agronomy, Food, Natural resources, Animals and Environment (DAFNAE), University of Padova, Viale dell'Università 16, 35020 Legnaro (PD), Italy; valentinaferrari@anafi.it

Abstract

Age at first calving (AFC) is around a 2-years non-productive period (on average in Holstein cows) which plays a relevant role in the cost of rearing replacements in dairy herds. In the present study we aimed at evaluating the heritability of AFC and estimating breeding value of the trait in the Italian Holstein population. A linear animal model was used and heritability was 0.03 (Posterior Standard Deviation of 0.005). Further analyses will be performed to obtain a genomic evaluation of AFC and to evaluate the economic importance of this trait.

Introduction

Age at first calving (AFC) is the period from birth to first calving that a female calf needs to reach puberty and to reproduce for the first time. The AFC plays a relevant role in the cost of rearing replacements in dairy herds. Boulton *et al.* (2017), considering an average AFC of 25.8 months, estimated that the mean cost of rearing increased by £2.87 for each additional day of AFC. An earlier AFC allows animals to generate income earlier and the effects of AFC on production traits in dairy cattle have been estimated (Hutchison *et al.*, 2017). Hence, the aim of this study is to evaluate the genetic aspects of AFC (heritability, genetic trend and correlation among estimated breeding values) in Italian Holstein population.

Materials & methods

Data editing. Data stored in the national database at ANAFIBJ including AFC of 4,510,986 Holsteins born from 1993 and with first calving from January 1996 to 2020 were used. AFC was limited from 18 to 36 months and gestation length was restricted to 240 to 305 d. Furthermore, only animals with both parents in the pedigree were selected. In order to have an effective comparison between contemporaries for herd-year-season of birth (HYS), groups with less than 4 observations were removed. Cows' pedigree was traced back to at least 4 generations. Table 1 reports the description of the dataset.

Table 1. Dataset description.

Item	N	Parameter	Mean	SD
Animals	4,510,986	Mean AFC 1996[2], months	28.14	2.94
Herds	21,080	Mean AFC 2020[3], months	26.09	3.07
HYS[1] groups	428,650	Animals/HYS	10.52	

[1] HYS = Herd/Year/Season of birth class.

[2] Mean age at first calving of cows that calved in 1996.

[3] Mean age at first calving of cows that calved in 2020.

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_702, © V. Ferrari *et al.* 2022

Statistical model. In order to evaluate sources of variation of AFC, a forward stepwise approach based on Akaike Information Criterion was used. Fixed effects were selected using the general linear model procedure of R version 4.0.3. The linear animal model used to evaluate AFC was:

$$AFC_{ijkl} = HYS_i + GL_j + A_k + e_l \quad (1)$$

where, HYS_i is the herd-year-season of birth class, with season defined according to 4 classes (1 = Nov-Dec-Jan; 2 = Feb-Mar-Apr; 3 = May-Jun-Jul; 4 = Aug-Sep-Oct), GL_j is the gestation length in classes (240 to 260, 261 to 280, 281 to 290 and 291 to 305 days), A_k is the random additive genetic effect, and e_l is the random residual error.

Genetic parameters estimation. Variance components estimation was carried out with the Gibbs sampler THRGIBBS1F90 (Misztal *et al.*, 2002; Misztal, 2008) on the entire final dataset, with 160,000 iterations, a burn-in of 10,000 and a thinning rate of 10. The post-Gibbs analysis was carried out using the software POSTGIBBSF90 (Misztal *et al.*, 2002) using the last 150,000 samples. From the total final dataset, 3 randomized sampling were created selecting heifers from herds with at least 200 observations. The heritability (h^2) was calculated as:

$$h^2 = \frac{\sigma_a^2}{\sigma_a^2+\sigma_e^2}$$

where σ_a^2 is the additive genetic variance, and σ_e^2 is the residual variance.

Breeding values estimation (EBV). For pedigree analyses Relax2 program version 1.65 was used. Animals' EBVs were estimated with the linear animal model described above with MiX99 software (Lidauer *et al.*, 2019) and standardized to mean 100 and standard deviation 5, where high values mean animal with lower AFC within the population.

Results & discussion

From 1996 to 2020, the phenotypic mean of AFC decreased by 2.05 months, probably due to the increased selection for milk yield and for better management made by farmers resulting in earlier AFC in recent years (Figure 1). Indeed, this is observable when comparing the frequency of average AFC in 1996 and 2020; in 1996 the greater percentage of animals calved at 27 months, whereas it decreased at 24 months in 2020

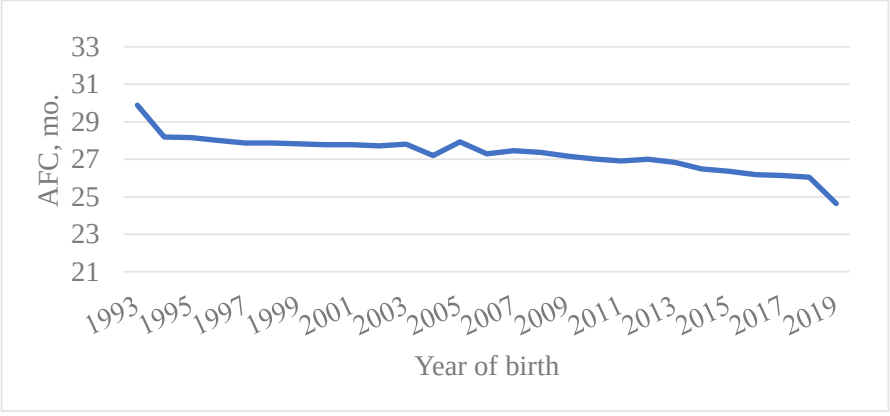


Figure 1. AFC trend by year of birth.

(Table 2). These results are in accordance with findings of Hutchison *et al.* (2017) who reported a decrease of AFC by 2.4 months for Holstein from 1997 to 2015.

The R^2 of the selected model was 47%. The analysis of the fixed effects revealed that 45.6% of the total variance was explained by HYS, whereas GL explained only 1.3% of the total variance.

Results of genetic parameters estimation are reported in Table 3. Heritability of AFC ranged from 0.03 to 0.04, showing low values. Low heritability values were also found by VanRaden and Klaaskate (1993), by Hutchison *et al.* (2019), and Heise *et al.* (2018) in Holstein cattle (0.02, 0.027, and 0.10, respectively), whereas Brzáková *et al.* (2020) found different results of heritability (from 0.18 to 0.23 in Aberdeen Angus and Charolais, respectively). Differences can be due to differences in genetic variation between breeds or due to management choices applied within breeds population.

Figure 2 depicts the genetic trend of AFC by year of birth. The tendency line displays a positive trend across years resulting in higher value in more recently bulls. The strong positive tendence across years is due to the increased selection for milk yield. This can be observed in Figure 3 also, where correlations of AFC with other traits are presented. The positive relationship of AFC with both production and fertility traits is displayed in Figure 3; only body condition score presents a negative correlation with AFC.

Table 2. Frequency of age at first calving (AFC) by month.

AFC, months	1996	2020
18-20	0.10	1.09
21	0.30	3.02
22	1.04	8.05
23	3.49	14.30
24	7.53	15.19
25	12.24	14.39
26	13.86	11.19
27	14.35	9.07
28	12.12	6.61
29	9.87	5.02
30	7.35	3.57
31	5.74	2.80
32	4.16	2.00
33-35	7.86	3.72

Table 3. Results of genetic parameters estimation.[1]

	h^2	PSD	Burn-in
Sample 1	0.04	0.005	10,000
Sample 2	0.03	0.004	10,000
Sample 3	0.03	0.005	10,000
μ	0.03	0.005	

[1] h^2 = heritability; PSD = posterior standard deviation.

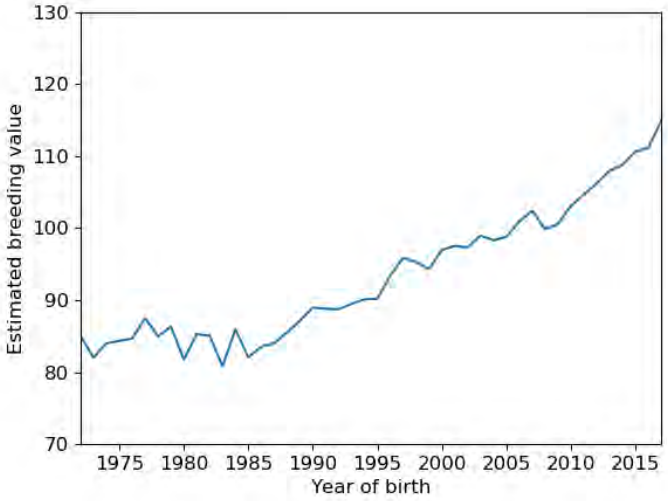


Figure 2. Genetic trend of AFC by year of birth.

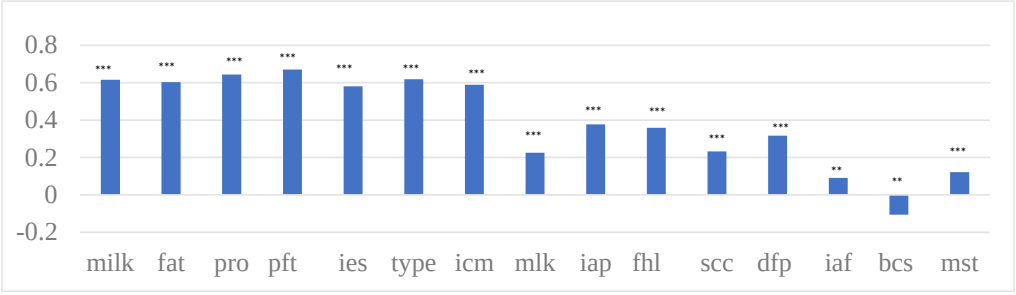


Figure 3. Correlation of indexes with AFC – Milk: Milk yield; fat: fat yield (kg); pro: protein yield (kg); pft: productivity, functionality and type index; ies: economic and functional index; type: type index; icm: udder composite index; mlk: milkability; iap: feet and legs composite index; fhl: functional longevity; scc: somatic cell score; dfp: calving ease; iaf: female fertility; bcs: body condition score; mst: udder healt index (*P<0.05, **P<0.01, and ***P<0.001).

Conclusions

Based on the results, the estimated heritability confirms the possibility for AFC genetic selection. Positive correlation of AFC with both productive and fertility traits reveals that selecting for an earlier AFC can improve herd profitability. Further studies will be conducted to evaluate its optimum value and to define the economic importance of this trait and to investigate the possibility to include AFC in the others selection index for Holstein.

References

Boulton, A. C., Rushton, J., and Wathes, D. C. (2017). An empirical analysis of the cost of rearing dairy heifers from birth to first calving and the time taken to repay these costs. *Animal*, *11*(8), 1372–1380. https://doi.org/10.1017/S1751731117000064

Brzáková, M., Čítek, J., Svitáková, A., Veselá, Z., and Vostrý, L. (2020). Genetic parameters for age at first calving and first calving interval of beef cattle. *Animals*, *10*(11), 1–11. https://doi.org/10.3390/ani10112122

Heise, J., Stock, K. F., Reinhardt, F., Ha, N. T., and Simianer, H. (2018). Phenotypic and genetic relationships between age at first calving, its component traits, and survival of heifers up to second calving. *Journal of Dairy Science*, *101*(1), 425–432. https://doi.org/10.3168/jds.2017-12957

Hutchison, J. L., VanRaden, P. M., Null, D. J., Cole, J. B., and Bickhart, D. M. (2017). Genomic evaluation of age at first calving. *Journal of Dairy Science*, *100*(8), 6853–6861. https://doi.org/10.3168/jds.2016-12060

Hutchison, J. L., VanRaden, P. M., Null, D. J., Cole, J. B., and Bickhart, D. M. (2019). Available at: https://www.uscdcb.com/cdcb-changes-to-evaluation-system-april-2019/

Lidauer, M., Matilainen, K., Strandèn, I., Mäntysaari, E., Pitkänen, T., and Taskinen, M. (2019). *MiX99 : Technical reference guide for MiX99 pre-processor*.

Misztal, I. (2008). Reliable computing in estimation of variance components. In *Journal of Animal Breeding and Genetics* (pp. 363–370). https://doi.org/10.1111/j.1439-0388.2008.00774.x

Misztal, I., Tsuruta, S., Strabel, T., Auvray, B., Druet, T., and Lee, D. H. (2002). BLUPF90 and related programs (BGF90). Commun. No. 28-07. *7th World Congress on Genetics Applied to Livestick Production, Montpellier, France*.

VanRaden, P. M., and Klaaskate, E. J. H. (1993). Genetic Evaluation of Length of Productive Life Including Predicted Longevity of Live Cows. *Journal of Dairy Science*, *76*(9), 2758–2764. https://doi.org/10.3168/jds.S0022-0302(93)77613-4

Part 40

Caprine/ovine

703. Genetic and genomic approaches to ewe's artificial insemination success in the Spanish Assaf and Manchega sheep breeds

M. Serrano[1], M.A. Jiménez[1], M. Ramón[2*] and O. González-Recio[1]

[1]Instituto Nacional de Investigación y Tecnología Agraria y Alimentaria CSIC-INIA, Ctra. de la Coruña km 7.5, 28040 Madrid, Spain; [2]Instituto Regional de Investigación y Desarrollo Agroalimentario y Forestal de Castilla-La Mancha (IRIAF), Avda. del Vino s/n, 13300 Valdepeñas, Spain; m.ramon.fernandez@gmail.com

Abstract

Artificial insemination (AI) is an essential tool in genetic programs of dairy livestock. In sheep AI has low success rates for several reasons, among which, the anatomy of the reproductive tract of the sheep, the non-detection of the oestrous cycle and the use of fresh semen, can be highlighted. In this work classical genetic and genomic approaches to the evaluation of this trait in the Spanish Assaf and Manchega breeds have been performed. Several threshold models have been tested considering the male and female contribution in different ways. Heritability estimates for AI success ranged from 0.07 to 0.10 for Assaf and from 0.01 to 0.05 for Manchega. The inclusion of genetic markers in the evaluation model yields more accurate breeding values predictions, but highly inflated. Results pointed out the possibility to exert some selective pressure over this trait to improve AI results.

Introduction

In the current dairy sheep sector, reproductive rates achieved by artificial insemination (AI) are one of the major concerns, due to the importance of this tool for the efficiency of the breeding programs. AI allows the progeny testing of rams, herds' connection, and an efficient dissemination of genetic improvement. Fertility, defined as the ability to produce offspring, is due to effects contributed by the male and the female, most likely in a non-purely additive fashion. The phenotype observed for the fertility of a mating, is a binary trait (0/1) and a continuous underlying variable is assumed under a threshold model assumption (Gianola, 1982). Two models can be fitted to the binary results of AI, considering the female and the male components, the additive and the product threshold models (David *et al.*, 2011). Both models yield low heritability estimates for the female component (0.10) and very low for the male one, (0.01) in the additive model, and (0.10) in the product model (David *et al.* 2011). Modern genomic tools, as medium and high SNPs platforms, can help through the incorporation of genotypic information to the classical genetic models, recovering a proportion of the genetic component responsible of fertility variation and achieving more accurate estimates of the animals' genetic values. In this work, we studied the AI fertility in the Assaf Spanish and the Manchega sheep breeds, by fitting different threshold animal models. Traditional genetic and genomic approaches, including 11,269 and 4,874 genotyped animals for the Assaf and Manchega, respectively, with the Affimetrix 50K platform, were compared in terms of genetic parameters estimates and accuracy of breeding values predictions.

Materials & methods

Data. Assaf breed data consisted of 203,621 AI records provided by the Spanish Assaf breeders association (ASSAFE) between 1998 and 2021 and 11,269 genotyped animals. Manchega breed data consisted of 168,154 AI records provided by the Manchega breeders association (AGRAMA) recorded between 2006 and 2017, and 4,874 genotyped animals (Table 1).

Methods. A threshold model was used to estimate variance components and genetic parameters, as well as to predict breeding values for fertility using a binary trait 0/1 as an indicator of the absence/presence of an ewe AI outcome. Some genetic models were fitted according to how the ram effect was considered,

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_703, © M. Serrano *et al.* 2022

Table 1. Summary of the data, pedigree and genotypes bases.

	Assaf	Manchega
No. AIs / no. herds	203,621 / 210	168,154 / 104
No. ewes / no. rams	146,982 / 848	104,673 / 994
Average (sd) AI fertility	0.35 (0.48)	0.44 (0.39)
Average (sd) inseminations per herd	969.2 (1,205.07)	1,617 (1,107.0)
Average (sd) inseminations by ram	240.4 (315.5)	169.2 (274.5)
Average (sd) inseminations per ewe	1.38 (0.76)	1.61 (0.86)
Average (sd) age of ewes in years at AI	2.6 (1.7)	3.0 (1.0)
Average (sd) age of rams in years at AI	2.7 (1.8)	3.3 (1.9)
No. animals in pedigree	211,908	147,896
No. animals with sire or dam known	166,081	111,965
No. animals with sire and dam known	126,818	84,307
No. animals genotyped	11,269	4,874
No. ewes genotyped (with data)	4,144 (2,016)	1,591 (615)
No. rams genotyped (with data)	7,125 (559)	3,378 (184)

that is, if the ram was included as a fixed or a random permanent effect and the inclusion of repeated measures of the ewes as a random effect. All models were implemented either in a genetic or a genomic approach. For the Assaf breed, the comparison group herd-AI date (2,857 levels), the ewe age and the ram age (8 levels in both cases) were included as fixed effects in the model. For the Manchega breed, fixed effects considered were herd-AI date (2,474 levels), number of parity-age interaction (16 levels), age of the ram (9 levels), days from lambing to AI service (5 levels) and level of milk production of the ewe at the service (10 levels). Technicians making inseminations were not included in the model since each of them inseminates always the same herds and therefore effects are confounded. Model 1 included the ram as a fixed effect and the additive effect of the ewe; Model 2 fitted the ram as a random permanent effect and the additive effect of the ewe; Model 3 included the ram and ewe permanent effects in addition to the additive effect of the ewe; Model 4 included both permanent and additive effects for rams and ewes. Genetic groups were included defined on a basis of a generation interval of 3 years. The animal additive genetic and the residual were assumed to be randomly distributed. Genomic information file consisted of 11,269 animals genotyped for 42,479 SNPs in the Assaf breed, and of 4,874 animals genotyped for 38,433 SNPs in the Manchega breed, after a quality control on animal call_rate (≥90%), SNPs call-rate (≥90%) and Minimum allele frequency (MAF≥3%). The software used for both the genetic and genomic approaches was the THRGIBBSF90 (BLUPF90 family programs. University of Georgia. http://nce.ads.uga.edu/software/). Gibbs sampling consisted of a single chain of 100,000 iterations, with a burning period of 10,000 samples and saving samples every 10 iterations, leaving a total of 9,000 samples for final inferences. Models were compared based on DIC (deviance information criterion) values, and genetic and genomic predictions, based on heritability and accuracy estimates.

Results

Table 2 and 3 shows DIC values and variance components estimates obtained in each of the breeds and genetic models fitted under both, the classic genetic and genomic approaches. For the Assaf breed, models comparison revealed under both, genetic and genomic approaches, that Model 3, which includes the ram and ewe environmental permanent effects and the additive effect of the ewe, had the best performance in terms of DIC values (the lower the better). For the Manchega sheep, model 1 with the ram as a fixed effect and the permanent and additive effects of the ewe resulted in a better fitting. In general, variances estimates and heritabilities were of lower magnitude for the Manchega sheep. Additive variances estimates were very similar for all models, and something higher under the genomic approach with respect to the classic genetic

Table 2. Deviance information criterion (DIC), variance components and standard deviations, and heritability, estimated for all the models tested under the genetic and genomic approaches in the Assaf sheep breed.

Approach	Genetic			Genomic		
model ID[1]	1	2	3	1	2	3
DIC[2]	-41,890	-51,261.6	-66,223.3	-43,795.2	-53,224.6	-65,938.7
σ^2a_{ewe}(sd)	0.200 (0.008)	0.197 (0.008)	0.133 (0.008)	0.210 (0.008)	0.206 (0.008)	0.151 (0.008)
σ^2p_{ewe} (sd)	-	-	0.146 (0.001)	-	-	0.125 (0.001)
σ^2p_{ram} (sd)	-	0.014 (0.001)	0.015 (0.001)	-	0.014 (0.001	0.015 (0.001)
σ^2e (sd)	1.012 (0.004)	1.009 (0.004)	1.008 (0.004)	1.012 (0.004)	1.009 (0.004)	1.008 (0.004)
h^2s	0.165	0.161	0.102	0.172	0.168	0.116
h^2o	0.099	0.097	0.062	0.104	0.101	0.070

[1] 1 Fixed ram; 2 Permanent ram; 3 Permanent ram and Permanent ewe.

[2] DIC deviance information criterion; σ^2a (sd) additive variance and its standard deviation; σ^2p_{ewe} (sd) permanent variance of the ewe and its standard deviation; σ^2p_{ram} (sd) permanent variance of the ram and its standard deviation; σ^2e (sd) residual variance and its standard deviation; h^2s heritability on the underlying scale; h^2r heritability under the observed scale.

Table 3. Deviance information criterion (DIC), variance components and standard deviations, and heritability, estimated for all the models tested under the genetic and genomic approaches in the Manchega sheep breed.

Approach	Genetic			Genomic		
model ID[1]	1	3	4	1	3	4
DIC[2]	167,100.8	168,803.74	168,476.8	167,050.2	168,783.2	168,472.5
σ^2a_{ewe} (sd)	0.096 (0.009)	0.098 (0.006)	0.099 (0.007)	0.093 (0.005)	0.097 (0.006)	0.100 (0.008)
σ^2p_{ewe} (sd)	0.174 (0.009)	0.174 (0.009)	0.171 (0.009)	0.183 (0.009)	0.174 (0.009)	0.170 (0.010)
σ^2a_{ram} (sd)	-	-	0.014 (0.003)	-	-	0.009 (0.001)
σ^2p_{ram} (sd)	-	0.033 (0.003)	0.019 (0.003)	-	0.033 (0.003)	0.024 (0.002)
σ^2e (sd)	1.017 (0.005)	1.012 (0.004)	1.012 (0.005)	1.017 (0.005)	1.012 (0.005)	1.012 (0.004)
h^2s ($_{ewe/ram}$)	0.075/-	0.076/-	0.077/0.013	0.072/-	0.076/-	0.078/0.008
h^2o ($_{ewe/ram}$)	0.048/-	0.049/-	0.050/0.010	0.047/-	0.049/-	0.050/0.006

[1] 1 Fixed ram; 3 Permanent ram; 4 Permanent plus additive ram and ewe.

[2] DIC deviance information criterion; σ^2a_{ewe} (sd) additive variance of the ewe and its standard deviation; σ^2p_{ewe} (sd) permanent variance of the ewe and its standard deviation; σ^2a_{ram} (sd) additive variance of the ram and its standard deviation; σ^2p_{ram} (sd) permanent variance of the ram and its standard deviation; σ^2e (sd) residual variance and its standard deviation; h^2s heritability on the underlying scale; h^2r heritability under the observed scale.

one. Rams permanent environmental variance estimates were very low, about 10 times less than that of ewes. For the Assaf breed, heritability on the underlying scale ranged from 0.10 to 0.17, being slightly higher those estimated under the genomic approach. Under the observed scale, heritability estimates ranged from 0.06 to 0.10. Estimates were lower for the Manchega breed, ranging between 0.01 (male trait) to 0.08 (female trait) in the underlying scale and between 0.01 (male trait) and 0.05 (female trait) when back-transforming to the observable scale.

In the Assaf breed and considering results from the best model (Model 3), lower DIC, correlations among the estimated breeding values for AI fertility under the genetic (EBV) and genomic (GEBV) approaches was 0.93 for all evaluated animals and 0.80 for the genotyped ones. Average accuracies of fertility EBVs under Model 3, was 0.37. For the genomic evaluation with Model 3, average accuracies of GEBVs were 0.44 and 0.56, for all animals and genotyped animals, respectively. For Manchega, Model 1 showed the best goodness of fit. For this model, correlations between EBV and GEBV were 0.99 for all animals, and 0.89 for

the genotyped ones. Average accuracies were 0.40 for model 3 not including genotypes, 0.41 when genomic information was included and 0.45 for the genotyped animals. In this breed, when the additive effect of the ram was included (model 4), correlations between EBV and GEBV were 0.99 for all animals and 0.82 for the animals with genomic information. Average accuracies were 0.37 for model 4 with no genotypes, 0.38 when including genomic information and 0.51 for the animals genotyped.

Discussion

Improving artificial insemination efficiency is critical to dairy genetic programs. Most of this improvement will come through innovations in reproductive techniques, however, genetic selection is also possible despite the low heritability of the trait. Precisely due to the low genetic determinism of AI fertility, genomic information could contribute to improve the accuracy in the choice of future breeders. A single-step approach that uses the combined relationships, allows for a simple evaluation that takes all available phenotypic, pedigree, and genomic information (Chen *et al.*, 2011). In this work a heritability of 7 to 10% has been estimated for the AI success in Assaf and of 5% for Manchega breeds. Despite the fact that no direct selection pressure has been exerted on this trait, a positive trend of breeding values over the last decade has been observed in both breeds. Even though only a few animals with phenotypic data have genotype information, the genomic approach provides higher accuracy of breeding values.

References

Chen C.Y., Misztal I., Aguilar I., Legarra A. and Muir W.M. (2011) J. Anim. Sci. 89: 2673-2679. https://doi.org/10.2527/jas.2010-3555

David I., Carabaño M.J., Tusell L., Diaz C., González-Recio O. *et al.* (2011). J Anim Sci 89: 321-328. https://doi.org/10.2527/jas.2010-3167

Gianola D. (1982). J. Anim. Sci. 54: 1079–1096. https://doi.org/10.2527/jas1982.5451079x

704. The impact of sire on variation in methane production and dry matter intake in sheep

E. O'Connor[1,2*], F.M. McGovern[1], E. Dunne[1], S.J. Morrison[3], T.M. Boland[2] and N. McHugh[4]

[1]Teagasc, Animal and Grassland Research and Innovation Centre, Athenry, Co. Galway, H65 R718, Ireland; [2]School of Agriculture and Food Science, University College Dublin, Belfield, D04 V1W8, Dublin 4, Ireland; [3]Agri-Food and Bioscience, Hillsborough, BT26 6DR, Co. Down, United Kingdom; [4]Teagasc, Animal and Grassland Research and Innovation Centre, Fermoy, P61 P302, Co. Cork, Ireland; edel.oconnor@teagasc.ie

Abstract

The objective of this study was to investigate the impact of sire on variation in methane (CH_4) production and dry matter intake (DMI). Methane emissions were measured on 60 Suffolk and Texel ewe lambs from 20 sires using respiration chambers (RC) and portable accumulation chambers (PAC). Daily DMI was measured on lambs fed an *ad libitum* grass pellet diet. The mean CH_4 output was 13.65±2.63 g/day (PAC Pre-RC), 28.07±4.96 g/day (RC) and 13.53±3.83 g/day (PAC Post-RC). Methane production differed by sire ($P<0.05$) and varied from 20.70 to 40.25 g/day in the RC. A strong correlation (0.68) was observed between CH_4 output measured in PAC Pre-RC and RC. Dry matter intake differed by sire ($P<0.05$). Methane production did not differ by breed ($P>0.05$), however, DMI differed by breed ($P<0.05$), with greater intakes associated with Suffolk (2.14 kg DM/day) compared to Texel lambs (2.01 kg DM/day).

Introduction

Respiration chambers are considered the 'gold standard' method of measuring CH_4 emissions in ruminants (Robinson *et al.*, 2014). To date RC have been used to quantify an animals CH_4 output, compare emissions on different diets (Zhao *et al.*, 2017) and to validate new CH_4 measurement techniques (Jonker *et al.*, 2018). In recent years there has been a movement away from using the RC due to the expense involved in construction and maintenance, and also the restriction on the number of animals that can be measured at any one time (Jonker *et al.*, 2018). One such alternative technique for measuring CH_4 emissions are the PAC, a short-term spot sampling method, which measures the accumulation of gas over a 50-minute period (Jonker *et al.*, 2018). This method allows for a high throughput of animals over a short period of time (O' Connor *et al.*, 2021), thereby providing CH_4 emission values on grazing sheep (Jonker *et al.*, 2018). Sire variation in CH_4 emissions measured using the RC and subsequent DMI has been estimated previously (Robinson *et al.*, 2014). However to date the simultaneous measurement of CH_4 and DMI from Irish sheep is relatively unknown. Feed intake has been shown to be the main driver of CH_4 production (Oddy *et al.*, 2019). Dry matter intake is easily measured during the RC period (Jonker *et al.*, 2018; Zhao *et al.*, 2017) while fewer studies have attempted to measure DMI around PAC measurements due to the challenges with feeding protocols prior to measurement (Hegarty, 2013). Therefore, the objectives of this study were to investigate the effect of sire on the variation in: (1) CH_4 emissions measured using both RC and PAC, and (2) DMI.

Materials & methods

Methane and DMI measurements were taken on a total of 60 nulliparous purebred Suffolk (30) and Texel (30) ewe lambs (c. 6 to 7 months of age) from October to December 2019. Each lamb was assigned to one of ten groups; each group consisted of six ewe lambs. These 60 ewe lambs represented the progeny from 20 sires (11 Suffolk and 9 Texel), with the number of ewe lambs per sire ranging from 1 to 6, and the number of CH_4 production and DMI records per sire ranging from 6 to 36.

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_704, © E. O'Connor *et al.* 2022

Dry matter intake. At the commencement of the experiment each group, consisting of six ewe lambs balanced for weight and breed, were housed in individual pens for 20 days. Following a 5 day diet adaption period, all animals had *ad libitum* access to a grass pellet (91.31% dry matter) diet. Individual DMI was measured from days 6 to 14 in individual pens, days 15 to 17 in the RC and days 18 to 20 in individual pens. Daily DMI of each animal was determined based on the quantity of feed offered and refused on a daily basis.

Methane measurements. Portable accumulation chamber measurements as described by Jonker *et al.* (2018) were taken on days 13 and 14, the two days prior to entry to the RC (hereon referred to as PAC Pre-RC) and on days 18 and 19, two days after the RC (hereon referred to as PAC Post-RC). One hour prior to CH_4 measurement, animals were removed from feed in their individual pens and were weighed. On day 18, immediately after a 72 hour RC measurement animals were returned to the individual pens and offered *ad libitum* access to feed for 2 hours before being removed from feed for a minimum of 1 hour and a CH_4 measurement was taken using the PAC. Methane (ppm), oxygen (%) and carbon dioxide (%) measurements were obtained using individual chambers over a 50 minute measurement period at three specific time points (0, 25 and 50 minutes) using the RKI Eagle 2 monitor (Weatherall Equipment and Instruments Ltd, UK).

Open-circuit respiration chamber measurements, as described by Zhao *et al.* (2017), were carried out over a 72 hour period from day 15 to 17. Within each group, consisting of six animals, individual lambs were assigned randomly to one of six chambers, with the first 24 hour period used as an acclimatisation period and 2×24 hour test periods thereafter. Animals had *ad libitum* access to the grass pellet diet throughout the 72 hour period.

Statistical analysis. To investigate the effect of sire on CH_4 (g/day, CH_4 per kg live-weight and CH_4 yield) and DMI (kg DM/day), linear mixed models in PROC MIXED (SAS, Inst. Inc., Cary, NC.) were used with animal live-weight, measurement group (n=1 to 10), sire (n=20), age of dam, birth and rearing type of the lamb (single, twin, triplet or quadruplet) included as fixed effects. In all models animal was included as a repeated effect. When the dependent variable was CH_4 production, DMI, chamber number, chamber number of measurement and measurement period (PAC Pre-RC, RC, PAC Post-RC) were also included as fixed effects. Biological interactions of interest were also investigated. In a secondary analysis sire was substituted with breed (Suffolk or Texel) to determine the effect of breed on CH_4 production and DMI. To quantify the relationship between gaseous measurements undertaken in the PAC and RC, the correlations between both CH_4 output (g/day) measured in PAC and RC were calculated.

Results

The mean CH_4 output measured during each period was 13.65±2.63 g/day (PAC Pre-RC), 28.07±4.96 g/day (RC) and 13.53±3.83 g/day (PAC Post-RC). Methane production differed by both sire and measurement period (i.e. PAC Pre-RC, RC and PAC Post-RC; $P<0.001$). Although a significant interaction between sire and measurement period was observed for CH_4 ($P<0.05$), further investigation showed that this was due to the re-scaling rather than the re-ranking of sires across measurement periods. Sire 1 (Texel) ranked as the highest emitter for CH_4 output measured in both the PAC and RC, with two other sires remaining in the top six CH_4 emitters across both measurement techniques. Two of the lowest emitting sires remained in the bottom six when measured in both the PAC and RC. The coefficient of variation for CH_4 production across sires ranged from 4.40 to 9.04%. Moderate to strong correlations (0.51 and 0.68) were estimated between CH_4 output measured in PAC, either Pre or Post-RC, and RC. The correlation between the two PAC measurements Pre- and Post-RC was 0.56. The ranking of sire for CH_4 output measured using the RC is shown in Figure 1 and ranged from 20.70 to 40.25 g/day. Methane output did not differ by breed ($P>0.05$).

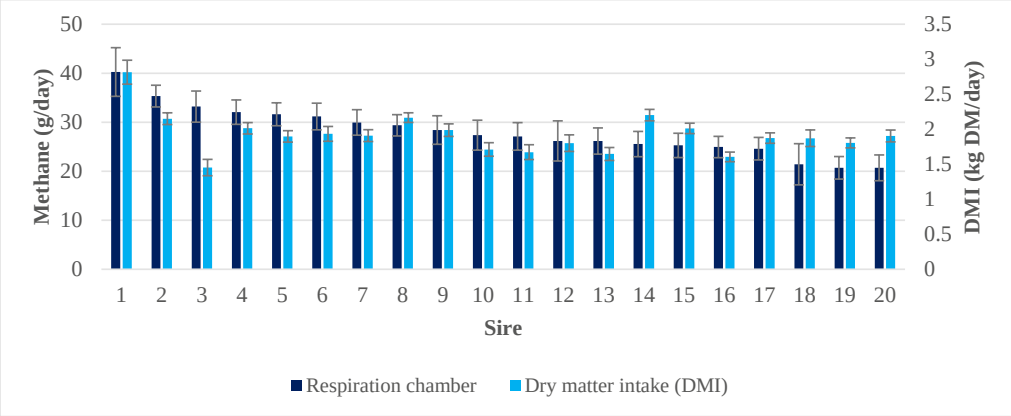


Figure 1. Ranking of sires from highest to lowest for methane output (g/day) measured in the respiration chamber and their respective dry matter intake (kg DM/day).

When methane emissions were expressed as CH_4 per kg live-weight, emissions were found to differ by sire ($P<0.05$). Strong to moderate correlations ranging from 0.62 to 0.76 were observed between CH_4 per kg live-weight measured during the three measurement periods (i.e. PAC Pre- and Post-RC, and RC). Methane per kg live-weight ranged from 0.14 to 0.63 g CH_4/kg live-weight when measured in either PAC measurement period with values ranging from 0.36 to 0.88 g CH_4/kg live-weight when measured in the RC.

The average DMI measured for all animals was 1.92 kg DM/day and ranged from 1.45 to 2.81 kg DM/day. Sire had a significant effect on DMI ($P<0.05$). The DMI of each sire is shown in Figure 1. The same sire (Sire 1) had the highest DMI along with CH_4 output (g/day) across the measurement period. Dry matter intake differed by sire breed ($P<0.05$), with Suffolk ewe lambs having a greater DMI (2.14 kg DM/day) compared to Texel ewe lambs (2.01 kg DM/day). Suffolk lambs also had a higher DMI per kg metabolic live-weight (0.12 kg DMI/kg metabolic live-weight) compared to Texel lambs (0.11 kg DMI/kg metabolic live-weight; $P<0.05$) Methane yield differed by sire ($P<0.05$) and ranged from 10.62 to 18.58 g/kg DMI when measured in the RC, while a smaller range of methane yield was observed during both PAC measurement periods (range 6.41 to 11.77 g/kg DMI). The correlations between CH_4 yield across the three measurement periods ranged from 0.40 to 0.50.

Discussion

Variation across sire is common among a plethora of traits such as lamb birth weight (Speijers *et al.*, 2010), however, the ability to identify sires with low CH_4 emissions is paramount in order to identify and breed sheep for reduced CH_4 emissions. Similar to results observed by Robinson *et al.* (2014) using RC data, sire was found to have a significant effect on CH_4 production (g/day) irrespective of the measurement technique used. As there was a 14.81 g/day of CH_4 difference between the top and bottom emitting sires this suggests there is ample variation between sires to be able to select for reduced CH_4 production. The large variation between sires was observed in the current study irrespective of how CH_4 was expressed (i.e. CH_4 production (g/day) or expressed in terms of live-weight or DMI), which suggests that variation in CH_4 between sires could not be solely accounted for due to live-weight or DMI differences. Dry matter intake differed by sire in the current study, which is in agreement with Robinson *et al.* (2014) and Oddy *et al.* (2019). A 1.36 kg DM/day difference was seen between the top and bottom sire, highlighting the large variation between sires for DMI. The phenotypic correlations estimated between the measurement techniques investigated in

the present study (i.e. PAC or RC) were similar to the genetic correlation (0.67) calculated by Jonker *et al.* (2018) between the same measurement techniques; this aids the validation of the PAC for more widespread use in larger genetic studies. The lack of differences between the two breeds investigated in the present study for CH_4 production was not unexpected, as both Texel and Suffolk are terminal breeds with relatively similar live-weights, however, interestingly DMI did differ by breed. Results from this study highlight large sire variation in both CH_4 production and DMI, indicating that selecting sires with low CH_4 emissions, while taking cognisance of DMI will aid in breeding sheep with lower environmental footprints.

Acknowledgements

Funding from the Irish Department of Agriculture, Food and Marine GreenBreed (17/S/2135) and GrasstoGas (2019EN202) is gratefully acknowledged. The authors also wish to acknowledge the extensive input and direction from AgResearch, New Zealand and especially Dr John Mc Ewan and Dr Suzanne Rowe on this manuscript.

References

Hegarty, R. S. (2013). Animal, 7, 401-408. https://doi.org/10.1017/S1751731113000839

Jonker A., Hickey S.M., Rowe S.J., Janssen P.H., Shackell G.H. *et al.* (2018) J. Anim. Sci. 96(8):3031-3042. https://doi.org/10.1093/jas/sky187

O'Connor E., McHugh N., Boland T.M., Dunne E., McGovern F.M. (2021) J. Anim. Sci. 99(8). https://doi.org/10.1093/jas/skab132

Oddy, V. H., Donaldson, A. J., Cameron, M., Bond, J., Dominik, S., (2018). Ani. Prod. Sci., 59(3), 441-448. https://doi.org/10.1071/AN17447

Robinson, D. L., Goopy J. P., Donaldson A. J., Woodgate R. T., Oddy V. H., *et al.* (2014) Animal, 8(12):1935-1944. https://doi.org/10.1017/S1751731114001773

Speijers, M. H. M., Carson, A. F., Dawson, L. E. R., Irwin, D., & Gordon, A. W. (2010). Animal, 4(3), 486-496. https://doi.org/10.1017/S1751731109991236

Zhao Y.G., Annett R., Yan T. (2017) J. Anim. Sci. 95(8):3762-3771. https://doi.org/10.2527/jas.2017.1598

705. Milk transcriptome analysis to elucidate the impact of prepubertal nutrition in dairy ewes residual feed intake

A. Suárez-Vega[1*], B. Gutiérrez-Gil[1], G. Hervás[2], H. Marina[1], R. Pelayo[1], P.G. Toral[2], C. Esteban-Blanco[1], P. Frutos[2] and J.J. Arranz[1]

[1]University of León, Campus de Vegazana s/n, 24071 León, Spain; [2]Instituto de Ganadería de Montaña (CSIC-University of León), Grulleros, 24346 León, Spain; asuav@unileon.es

Abstract

Prepubertal nutrition has been proposed as an important factor for feed efficiency and milk production in later life. Moreover, nutritional restrictions during prepubertal mammary gland development may influence the milk production potential. This study aims to evaluate the potential impact of a nutritional challenge during the allometric growth of the mammary gland on the milk transcriptome of dairy ewes with different feed efficiency phenotypes. The differential expression analyses identified 29 genes whose expression was significantly different for feed efficiency and relevant regarding the nutritional challenge. Eight differentially expressed genes were clustered under the biological process 'cell adhesion'. Molecules involved in cell adhesion are crucial in the proper development and function of the mammary gland. In conclusion, the main genes impacted by the nutritional challenge and showing a different expression pattern related to feed efficiency could be involved in the development of the proper architecture of the mammary gland parenchyma.

Introduction

In the last years, selecting animals for feed efficiency (FE) has become a major research objective in the livestock industry. Feed provision accounts for around 60-70% of costs in dairy sheep farming. Moreover, breeding high feed efficient sheep reduces carbon fingerprint, thus, decreasing farming environmental impact. The residual feed intake (RFI) is one of the most used indexes to measure feed efficiency. In order to elucidate the genetic determinism of this complex trait in livestock, several RNA-Seq approaches have been applied in different tissues such as the liver, muscle, rumen, and duodenum to identify polymorphism and genes associated with FE (Brito *et al.*, 2020). However, till now, RNA-Seq has not been applied to the milk somatic cell (MSC) transcriptome in relation to this trait.

Early live nutrition has been proposed as an important factor for feed efficiency and milk production in later life. It has been demonstrated that a restricted diet, compared to an *ad libitum* one during the prepubertal stage, improves the ewes' feed conversion potential (Villeneuve *et al.*, 2010). Moreover, several studies in heifers, goats, and sheep have focused on elucidating the impact of nutrition during this period on the mammary gland parenchyma, showing that a high plane of nutrition reduces mammary gland development while nutrient restriction improves its development (Bowden *et al.*, 1995; Capuco *et al.*, 1995; Villeneuve *et al.*, 2010).

In dairy sheep, postnatal mammary gland development implies a succession of allometric and isometric growth to accomplish a full development by the onset of lactation. Although most mammary gland growth is related to pregnancy, duct network ramifies during the prepubertal allometric growth, which may influence the milk production potential (Anderson, 1975).

This study aims to evaluate the potential impact of a nutritional challenge during the allometric growth of the mammary gland on the MSC transcriptome of dairy ewes with different feed efficiency.

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_705, © A. Suárez-Vega *et al.* 2022

Materials & methods

Experimental design. An experimental population of 40 new-born female lambs of the Assaf breed was constituted. Once these animals passed the veterinary control and got adapted to the experimental facilities, they were separated into Control group (C, n=20) and Nutritional Challenge group (NC, n=20). The NC lasted approximately two months, between 3.5 and 5.5 months of age of the animals. In this period, the NC group received a 42% protein restriction diet compared with the C group. Ewes were mated at around ten months of age.

Management at feed efficiency sampling period. After lambing, the sheep were housed in individual tie stalls and fed *ad libitum* a total mixed ration formulated from alfalfa hay (particle size>4 cm) and concentrate (50:50). Feed intake was individual and daily measured over four weeks (week 6^{th}-9^{th} of lactation) by weighting the diet dry matter (DM) offered and refused by each animal. At the morning and evening milkings (08:30 and 18:30 h), total milk produced by each animal was also collected and weighed individually and daily to calculate the milk yield. The body weight (BW) of each sheep was recorded at the beginning and the end of the 4-week sampling period.

Estimation of feed efficiency. RFI was calculated as the residuals of the following regression model using the GLM procedure of the SAS software package (version 9.4; SAS Institute Inc., Cary, NC, USA):

$$DMI = \mu + a \times ECM + b \times MBW + c \times \Delta BW + RFI \quad (1)$$

Where, DMI represents the mean DM intake over the period (kg/day); μ is the intercept; ECM is the mean energy-corrected milk (kg/day); MBW is the mean metabolic body weight (BW0.75; kg); ΔBW is body weight change over the period adjusted for mean body weight (ΔBW × BW); RFI is the residuals; and a, b, and c are the regression coefficients. All variables included in the model were statistically significant.

Sampling, RNA extraction and sequencing. Milk samples were collected on week 7^{th}-9^{th} of lactation. RNA was extracted following the protocol by Suarez-Vega *et al.* (2016). From the 40 ewes with records for milk production and individual feed intake, 24 animals were selected for mammary gland transcriptome analysis by RNA-seq (12 NC vs 12 C). For the analysis of the present study, we choose the most extreme animals for the RFI value: 6 HighFE (3 NC, 3C), with an average RFI of -0.17 (SD=0.05), and 5 LowFE (3 NC, 2C), with an average RFI of 0.16 (SD=0.06). The RNA-Seq was conducted at NOVOGENE in Cambridge (UK). cDNA libraries were sequenced to a minimum depth of 30 million paired-end reads on a Novaseq 6000 (Illumina), generating stranded paired-end reads of 150 bp.

RNA-Seq data analyses. The alignment against the reference genome (Oar_rambouillet_v2.0) was carried out using the STAR software (Dobin *et al.*, 2013), and the transcript quantification step was performed with the RSEM software (Li and Dewey, 2011). Differential expression analyses between the groups or conditions were performed with the DESeq2 package (Love *et al.*, 2014). The model for DESeq2 was fitted assuming an interaction between the nutritional challenge and the RFI:

$$Y = \text{nutritional challenge} + RFI + \text{nutritional challenge:RFI} \quad (2)$$

Differentially expressed genes (DEGs) were considered at a false discovery rate (FDR)<0.05. Functional enrichment analyses of DEGs were performed with ToppGene (Chen *et al.*, 2009) and terms were considered significantly enriched when there were more than two genes clustered in Gene Ontology (GO) term and FDR<0.05.

Results

After massive parallel RNA sequencing, an average of 34 million paired-end reads were obtained for each animal. On average, 95% of the reads aligned to unique locations of the Oar_rambouillet_v2.0 reference genome.

When compared between HighFE and LowFE groups, 213 DEGs were found for the control group, and 86 DEGs were found for the NC group; from these two subsets, 12 DEGs were found in common in both analyses (Figure 1a). When comparing C vs NC groups, 37 and three DEGs genes were found for the HighFE and LowFE ewes, respectively. Among the DEGs identified between C vs NC, 17 genes were also DEGs between the feed efficiency comparisons (Figure 1a). The functional enrichment analysis used the 29 genes identified as differentially expressed by at least two comparisons. For the GO database, we found 5 terms enriched for the GO: Molecular function category, 2 terms for the GO: Biological Process, and 1 term for the GO: Cellular component categories. The top enriched term in each category was 'calcium ion binding', 'cell adhesion', and 'cell surface', respectively.

Discussion

This study represents the first evaluation of the transcriptomic changes in the mammary gland of dairy ewes with extreme feed efficiency phenotypes. It is remarkable that a different response in the MSCs transcriptome is observed between HighFE and LowFE animals when subjected to a nutritional challenge. Fewer differences in gene expression were observed for the RNA-Seq data in the NC group, suggesting that changes in the milk transcriptome for FE phenotype in the C group are more relevant than for NC animals.

Regarding the enriched GO terms identified, the 'calcium-binding' function has also been found enriched in previous studies on feed efficiency in pigs (Xu *et al.*, 2018), which supports that this function plays an important role in feed efficiency.

Eight DEGs were involved in the GO:biological process term 'cell adhesion'. Cell adhesion is important for the mammary gland as it is required for the survival and differentiation of luminal cells during lactation. In this sense, Eph- and ephrin-signalling pathways play a role in the morphogenic processes of various tissues (Andres and Ziemiecki, 2003). In our study, the expression of the *EPHA4* gene, included in the last cited pathway, was downregulated in the HighFE C group while it was higher expressed in both HighFE and LowFE NC ewes (Figure 1b). Several molecules from the ephrin-Eph system have been shown to be involved in branching morphogenesis and the establishment of the mammary gland functional structure (Kaenel *et al.*, 2014). Moreover, genes from this family have been shown to act as positive factors for the increase and

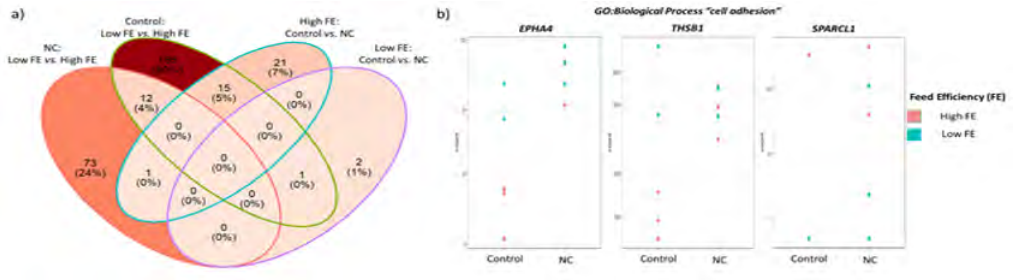


Figure 1. Differentially expressed genes between the conditions evaluated: Feed Efficiency (FE) and Nutritional Challenge (NC). (a) Venn Diagram integrating the results from the differential expression analyses performed. (b) Expression of three genes clustered under the GO: Biological Process 'cell adhesion'.

maintenance of the epithelial cells in the mammary gland (Kang *et al.*, 2018). Other genes included under the 'cell adhesion' GO term, such as *SPARCL* or *THBS1*, were previously identified as important genes in relation to feed efficiency in livestock (Tizioto *et al.*, 2016; de Lima *et al.*, 2020), highlighting their relevancy for this complex trait.

Taking into consideration our results, it can be hypothesised that the main genes affected by the nutritional challenge and showing a different expression pattern between the HighFE and LowFE groups could be involved in the proper development and function of the mammary gland parenchyma.

Acknowledgements

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 772787 (SMARTER).

References

Anderson, R. R. (1975). J. Anim. Sci. 41, 118–123. https://doi.org/10.2527/jas1975.411118x.

Andres, A.-C., and Ziemiecki, A. (2003). J. Mammary Gland Biol. Neoplasia 8. https://doi.org/10.7892/boris.117638.

Bowden, C. E., Plaut, K., Maple, U. L., and Caler, W. (1995). J. Dairy Sci. 78, 1728–1733. https://doi.org/10.3168/jds.s0022-0302(95)76798-4.

Brito, L. F., Oliveira, H. R., Houlahan, K., Fonseca, P. A. S., Lam, S., *et al.* (2020). Can. J. Anim. Sci. 100, 587–604. https://doi.org/10.1139/CJAS-2019-0193/ASSET/IMAGES/CJAS-2019-0193TAB5.GIF.

Capuco, A. V., Smith, J. J., Waldo, D. R., and Rexroad, C. E. (1995). J. Dairy Sci. 78, 2709–2725. https://doi.org/10.3168/jds.s0022-0302(95)76902-8.

Chen, J., Bardes, E. E., Aronow, B. J., and Jegga, A. G. (2009). Nucleic Acids Res. 37, W305-11. https://doi.org/10.1093/nar/gkp427.

de Lima, A.O., Koltes, J.E., Diniz, W.J.S., de Oliveira, P.S.N., Cesar, A.S.M., *et al.* (2020). Front. Genet. 11, 189. https://doi.org/10.3389/FGENE.2020.00189/BIBTEX

Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., *et al.* (2013). Bioinformatics 29, 15–21. https://doi.org/10.1093/bioinformatics/bts635.

Kaenel, P., Hahnewald, S., Wotzkow, C., Strange, R., and Andres, A. C. (2014). Dev. Growth Differ. 56, 255–275. https://doi.org/10.1111/DGD.12126.

Kang, M., Jeong, W., Bae, H., Lim, W., Bazer, F. W., *et al.* (2018). J. Cell. Physiol. 233, 2560–2571. https://doi.org/10.1002/JCP.26131.

Li, B., and Dewey, C. N. (2011). BMC Bioinformatics 12, 323. https://doi.org/10.1186/1471-2105-12-323.

Love, M. I., Huber, W., and Anders, S. (2014). Genome Biol. 15, 550. https://doi.org/s13059-014-0550-8.

Suarez-Vega, A., Gutierrez-Gil, B., Klopp, C., Tosser-Klopp, G., and Arranz, J.-J. (2016). Sci. Data 3. https://doi.org/10.1038/sdata.2016.51.

Tizioto, P.C., Coutinho, L.L., Oliveira, P.S.N., Cesar, A.S.M., Diniz, W.J.S., *et al.* (2016). Sci. Reports 2016 61 6, 1–12. https://doi.org/10.1038/srep39493

Villeneuve, L., Cinq-Mars, D., and Lacasse, P. (2010). Animal 4, 944–950. https://doi.org/10.1017/S1751731110000170.

Xu, Y., Qi, X., Hu, M., Lin, R., Hou, Y., *et al.* (2018). Genes 2018, Vol. 9, Page 336 9, 336. https://doi.org/10.3390/GENES9070336.

706. Inbreeding and effective population size of United States Katahdin sheep

G.M. Becker[1], J.M. Burke[2], R.M. Lewis[3], J.E. Miller[4], J.L.M. Morgan[5], B.D. Rosen[6], C.P. Van Tassell[6], D.R. Notter[7] and B.M. Murdoch[1*]

[1]Department of Animal, Veterinary and Food Sciences, University of Idaho, 875 Perimeter Dr, Moscow, ID 83844, USA; [2]USDA, ARS, Dale Bumpers Small Farms Research Center, 6883 AR-23, Booneville, AR 72927, USA; [3]Department of Animal Science, University of Nebraska-Lincoln, 1400 R St, Lincoln, NE 68588, USA; [4]Department of Pathobiological Sciences, School of Veterinary Medicine, Louisiana State University, Skip Bertman Dr, Baton Rouge, LA 70803, USA; [5]Round Mountain Consulting, Fayetteville, AR 72701, USA; [6]USDA, ARS, Animal Genomics and Improvement Laboratory, 10300 Baltimore Avenue, Beltsville, MD 20705, USA; [7]Department of Animal and Poultry Sciences, Virginia Tech, 925 Prices Fork Road, Blacksburg, VA 24060, USA; bmurdoch@uidaho.edu

Abstract

Katahdin sheep are a composite hair breed reared for meat production and known for parasite resistance traits. The aim of this study was to calculate observed heterozygosity (Het_O), inbreeding coefficients and effective population size (Ne) in a study population representing 23 U.S. flocks (n=4,884 sheep) with 50K genotype data. In the U.S. Katahdin, mean Het_O was 0.37 and mean inbreeding coefficient (Fhat2) was 0.01. Inbreeding coefficient estimates were statistically different (Kruskal-Wallis $P<2.2e-16$) between flocks. The Ne of the study population examined was n=172 at 13 generations ago, with average linkage disequilibrium r^2 of 0.04 for the generation. Maintaining genetic diversity relies upon maximizing Ne and minimizing inbreeding within a population. The current study found that mean inbreeding of Katahdin sheep was lower than published for other sheep breeds, mean Het_O was greater and Ne was intermediate to estimates for other breeds.

Introduction

Genetic diversity is important for the preservation of livestock species. Inbreeding can result in increased homozygosity, reduction of genetic diversity and the accumulation of detrimental alleles (Vostry *et al.*, 2018). Negative outcomes related to increased homozygosity can also occur through inbreeding depression, which can affect economically important traits such as birth weight, weaning weight, average daily gain and litter size (Gholizadeh and Ghafouri-Kesbi, 2016; Selvaggi *et al.*, 2010). The Katahdin breed of sheep was developed in the later half of the twentieth century (Wildeus 1997) and has since grown to become economically important to the United States sheep industry (Thorne *et al.*, 2021). The aim of this study was to characterize observed heterozygosity (Het_O), inbreeding and effective population (Ne) size in a large population of 4,884 Katahdin sheep with 50K genotype data.

Materials & methods

Twenty-three U.S. Katahdin flocks enrolled with the National Sheep Improvement Program (NSIP) participated in the current study. Sheep producers collected blood cards and DNA samples were extracted and genotyped with the Neogen GGP Ovine 50K genotype array. All autosomal markers were filtered by the following exclusionary parameters: sample call rate <0.90, marker call rate <0.90, marker minor allele frequency <0.01. Sex chromosome markers were not included in analyses. In total, 37,856 markers were used for principal component analysis (PCA) and effective population (Ne) size estimation. Data for heterozygosity and inbreeding analyses were pruned to remove markers in high linkage-disequilibrium (LD) using the --indep-pairwise function in plinkv1.9 with the parameters window size of 50 kb, variant count moving five markers per window and r^2 threshold of 0.8. Following LD-pruning, 34,772 autosomal

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_706, © G.M. Becker *et al.* 2022

markers were used in inbreeding and mean heterozygosity analyses. Observed heterozygosity, inbreeding and PCA analyses were conducted using plink v1.9 with the functions --het, --ibc and –pca, respectively (Chang *et al.*, 2015). Ne analysis using LD-based methods were conducted with SNeP1.11 with default parameters (Barbato *et al.*, 2015). R version 4.1.1 was used for data visualization with packages 'ggpubr' and 'ggplot2' (Kassambara 2020; Wickham 2016). Statistical significance of (Fhat2 ~ Flock) was evaluated with the Kruskal-Wallis test. The r package 'rcompanion' was used to calculate 95% confidence intervals for inbreeding coefficients (Mangiafico 2020).

Results

The structure of the study population was investigated by PCA (Figure 1). The flocks in this study were located across the five geographical regions of the U.S.: the East Coast (EC), Gulf Coast (GC), Midwest (MW) and the West Coast/Rocky Mountain (WR) regions. There was greater diversity of sheep from the GC region in comparison to sheep from other regions.

The mean heterozygosity observed (Het_O) in the study population was 37%. This was found to be higher than mean Het_O for previously reported breeds (Table 1). The range in mean Het_O by flock was 0.36 to 0.39 with a standard deviation of 0.01. The range in Het_O for individual sheep of all flocks was 0.25 to 0.45 with a standard deviation of 0.02.

Population and flock inbreeding coefficients were calculated using Fhat2, which is calculated (observed homozygous – expected) / (1 – expected) (Rovelli *et al.*, 2021). Individual sheep inbreeding coefficients for all flocks ranged from -0.42 to 0.34 with a standard deviation of 0.07, and the mean inbreeding coefficients by flock ranged from -0.1 to 0.07 with a standard deviation 0.04 (Figure 2). Inbreeding coefficients differed significantly across flocks (Kruskal-Wallis $P<2.2e-16$). The mean inbreeding coefficient over the entire study population was 0.01 and all inbreeding coefficient mean calculations were supported by 95% confidence intervals.

Fhat2 inbreeding coefficients have been previously estimated in other sheep breeds (Davenport *et al.*, 2020; Rochus *et al.*, 2020). The mean inbreeding coefficient calculated for this study population was smaller than

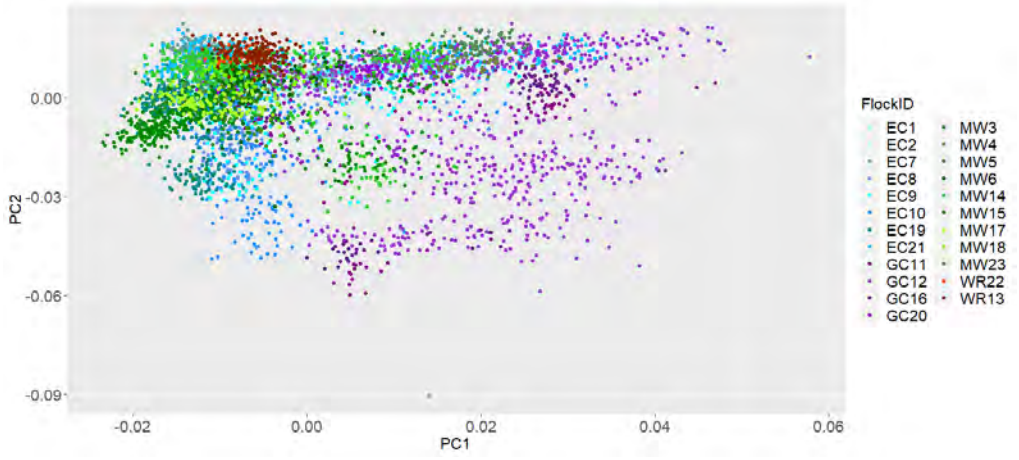


Figure 1. Principal component analysis (PCA) for U.S. Katahdin sheep. Animals are color-coded by region (East Coast in blue shades, n=2,167; Gulf Coast in purple shades, n=1,088; Midwest in green shades, n=1,414; West Coast/Rocky Mountain in orange shades, n=215) and labelled by flock from 1 to 23, with a total of 4,884 sheep.

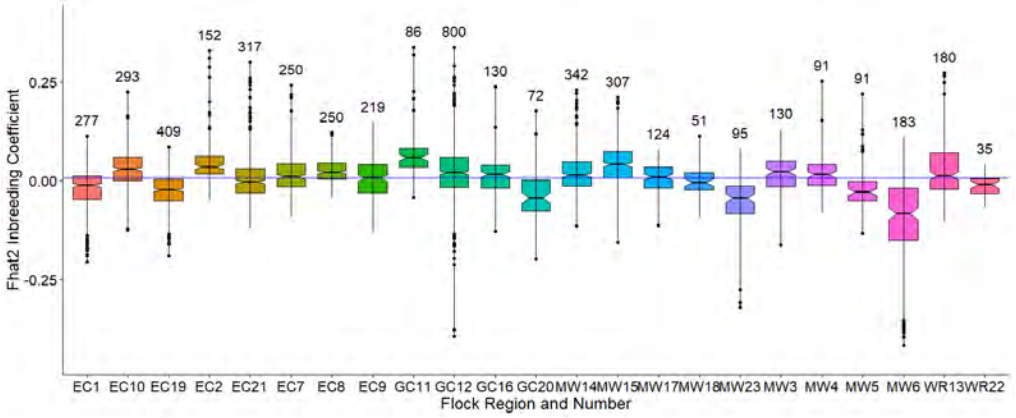


Figure 2. Fhat2 inbreeding coefficients by Katahdin flock. Overall population mean is given by the blue line and individual flock size (n) are given above each boxplot.

mean estimates published for other breed populations (Table 1). Inbreeding coefficients may be difficult to directly compare across studies due to differences in sample structure and size.

For this population, historic Ne was estimated to be 901 (as of 234 generations ago) while the most recent Ne estimate was 172 (as of 13 generations ago). The average LD per bin decreased from r^2 of 0.12 (standard deviation 0.16) at 234 generations ago to r^2 of 0.04 (standard deviation 0.06) at 13 generations ago.

Discussion

This study encompasses nearly 5,000 sheep from 23 Katahdin flocks located across the United States. Some of the sheep in this study exhibited inbreeding coefficients of 6.25% or greater, indicating cousin mating (Li *et al.*, 2011); however, the mean inbreeding coefficient was smaller than estimates for other breeds and the mean observed heterozygosity was higher. Previous estimations of Ne for other breeds range from 25 to 1,317 (Kijas *et al.*, 2012; Pasandideh *et al.*, 2020). Two of the founding breeds of the Katahdin, the Wiltshire and Suffolk, have previously reported Ne estimates of 100 and 300/569 (Irish/Australian Suffolk), respectively (Kijas *et al.*, 2012). In the current study, 4,884 Katahdin sheep were evaluated and found to have an estimated Ne that falls intermediate of other breed reports. These statistics have allowed for characterization of genetic diversity within a robust sample population of U.S. Katahdin sheep.

Table 1. Inbreeding coefficient and observed heterozygosity compared to other published estimates.[1]

	KT[2]	Rochus *et al.*, 2020					Davenport *et al.*, 2020					
		D	F	G	Gute	K	H	S	WS	O	SH	R
Number of animals	4,884	21	10	19	22	21	45	68	37	11	44	43
Inbreeding coefficient	0.01	0.3	0.36	0.03	0.25	0.26	0.14	0.13	0.14	0.05	0.09	0.16
Mean observed heterozygosity	0.37	0.30	0.30	0.33	0.31	0.34	0.33	0.33	0.34	0.35	0.34	0.30

[1] KT-Katahdin; F-Fjällnäs; G-Gotland; K-Klövsjö; H-Hampshire; S-Suffolk; WS-Western Suffolk; O-Oxford; SH-Shropshire; R-Rambouillet.

[2] Data from this study.

References

Barbato, M., Orozco-terWengel, P., Tapio, M., & Bruford, M. W. (2015) Front. Genet. 6:109. https://doi.org/10.3389/fgene.2015.00109

Chang, CC., Chow, CC., Tellier, LCAM., Vattikuti, S., *et al.* (2015) GigaScience 4(1). https://doi.org/10.1186/s13742-015-0047-8

Davenport, K. M., Hiemke, C., McKay, S. D., Thorne, J. W., *et al.* (2020) An. Genet. 51(2): 284-291. https://doi.org/10.1111/age.12905

Gholizadeh, M., & Ghafouri-Kesbi, F. (2016) Sm. Rum. Research 144: 184-190. https://doi.org/10.1016/j.smallrumres.2016.09.012

Kassambara, Alboukadel. ggpubr: 'ggplot2' Based Publication Ready Plots. R package version 0.4.0. Available at: https://CRAN.R-project.org/package=ggpubr

Kijas, J. W., Lenstra, J. A., Hayes, B., Boitard, S., *et al.* & International Sheep Genomics Consortium. (2012) PLoS Bio., 10(2). https://doi.org/10.1371/journal.pbio.1001258

Li, M. H., Strandén, I., Tiirikka, T., Sevón-Aimonen, M. L., *et al.* (2011) PLoS One, 6(11): e26256. https://doi.org/10.1371/journal.pone.0026256

Mangiafico, Salvatore. rcompanion: Functions to Support Extension Education Program Evaluation. R package version 2.4.13. Available at: https://CRAN.R-project.org/package=rcompanion

Pasandideh, M., Gholizadeh, M., & Rahimi Mianji, G. (2020). Iranian J. App. An. Sci., 10(2): 287-295.

Rochus, C. M., Jonas, E., & Johansson, A. M. (2020) BMC Genet. 21(1): 1-9. https://doi.org/10.1186/s12863-020-0827-8

Rovelli, G., Luigi-Sierra, M. G., Guan, D., Sbarra, F., *et al.* (2021) PeerJ, 9:e12049. https://doi.org/10.7717/peerj.12049

Selvaggi, M., Dario, C., Peretti, V., Ciotola, F., *et al.* (2010) Sm. Rum. Research, 89(1): 42-46. https://doi.org/10.1016/j.smallrumres.2009.12.005

Thorne, J. W., Murdoch, B. M., Freking, B. A., Redden, R. R., *et al.* (2021) An. Genet., 52(4): 395-408. https://doi.org/10.1111/age.13067

Vostry, L., Milerski, M., Schmidova, J., & Vostra-Vydrova, H. (2018) Sm. Rum. Research 168: 25-31. https://doi.org/10.1016/j.smallrumres.2018.09.004

Wickham H. *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. ISBN 978-3-319-24277-4, Available at: https://ggplot2.tidyverse.org.

Wildeus, S. (1997) Jour. An. Sci., 75(3): 630-640. https://doi.org/10.2527/1997.753630x

707. Genetic diversity and population structure of dairy goat populations in Kenya

C.M. Kamidi[1*], R.W. Waineina[1,2], C.B. Wasike[3], D.K. Ngeno[2] and E.D. Ilatsia[1]

[1]Dairy Research Institute, Kenya Agricultural and Livestock Organization, P.O. Box 25-20117 Naivasha, Kenya; [2]Animal Breeding and Genomics Group, Department of Animal Sciences, Egerton University, P.O. Box 536-20115, Egerton, Kenya; [3]Livestock Efficiency Enhancement Group (LEEG), Department of Animal and Fisheries Sciences, Maseno University, P.O Box Private Bag 40105, Maseno, Kenya; tmuhonja@gmail.com

Abstract

This study aimed to decipher the genetic diversity and structure of different dairy goat populations kept at KALRO-research station Naivasha to facilitate national genetic improvement program. We used 49,773 genome-wide SNP data to quantify levels of and patterns of genetic diversity among 243 individuals from five goat populations: Toggenburg (n=69), Alpine (n=54), Saneen (n=24), Toggenburg Kenya Dual Purpose goat (TKDPG) (n=4), Kenya Dual Purpose Goat (KDPG)(n=11) and crosses (n=81). Inbreeding coefficient, observed and expected heterozygosity ranged from-0.01-0.07, 0.56-0.60 and 0.57 respectively. Principal component analysis clustered the populations into three distinct clusters namely; cluster one-Saanen, Alpine and their crosses, cluster two-Toggenburg, Alpine, TKDPG and their crosses, and cluster three-KDPG, Toggenburg, TKDPG, Alpine and their crosses. Genetically admixed genotypes were identified in each cluster, indicating intermixing within the populations. High number of ROH proofed there is some level of homozygosity in the goat genotypes in Kenya. Results provided understandings into genomic variations and relationships which is vital in furthering genetic improvement and conservation of goat kept at Kenya Agricultural and Livestock Research Organization (KALRO) research station and Kenya as a whole.

Introduction

Goats are an integral part of the whole farming system with multiple biological (meat, milk, manure and skin) and non-biological (asset, security, and socio-cultural) functions (Ajmone-marsan *et al.*, 2014). Recent studies have also shown that goat farming could be one of the alternative agricultural practices due to the diminishing land sizes in the medium to high potential areas as a result of human population pressure and climate change-related challenges. These challenges are a trigger for alternative farming practices such as intensive dairy goat production which offers more multi-functionality, flexibility and adaptability to varied production conditions (Scarpa *et al.*, 2003). In Kenya, dairy goats were introduced through community-based breed improvement programs by various NGOs. Common dairy goat breeds include Saanen, Toggenburg and Alpines, and their crosses. Crossbreeding between the exotic breeds (Saanen, Toggenburg, Alpine and Anglo-Nubian) and local goat (Galla and Small East African) populations has resulted in different dairy goat genotypes kept by farmers (Peacock *et al.*, 2011). Since introduction of different goat genotypes, farmers have been practicing unsystematic crossing due to lack of breeding programs to guide in genetic improvement of dairy goats in the country (Kosgey and Okeyo, 2007). Currently, admixtures of exotic and local goats reared as dairy goats are present in different parts of the country under different production systems. The genetic diversity and relationships of these genotypes remain undocumented. It is therefore important to decipher their genetic diversity and relationships of dairy goat breeds in Kenya for conservation strategy as well as to further genetic improvement.

Materials & methods

Sample collection. A total of 243 blood samples were collected from goat population maintained on-station at the Dairy Research Institute of the Kenya Agricultural and Livestock Organization (KALRO) in

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_707, © C.M. Kamidi *et al.* 2022

Naivasha. The breeds sampled included Toggenburg (n=69), Saneen (n=24), Alpine (n=54), KDPG (n=11), TKDPG (n=4) and crosses (n=81). Blood samples were collected via jugular venipuncture from each animal into EDTA vacutainer tubes. The collected blood samples were kept in iceboxes until refrigeration at 4 °C before DNA extraction. The study was conducted in strict accordance with the recommendations for the Institute of Primate Research (IPR) ethical guidelines on animal care and the use of laboratory animals (ISERC/03/2020).

DNA extraction and genotyping. Genomic DNA was extracted from the whole blood using the DNeasy Blood and tissue kit 149 (Qiagen®, Hilden, Germany). Genomic DNA was quantified using Nanodrop Spectrophotometer 150 (Nanodrop ND-1000) and genotyped using the GoatSNP50 Bead Chip (Illumina, Inc. San Diego, 151 CA 92122 USA) developed by the International Goat Genome Consortium (IGGC).

Data analysis. Genetic diversity indices; inbreeding coefficient, observed and expected heterozygosity were calculated within genotypes using PLINK (Purcell *et al.*, 2007). Minor allele frequency (MAF) was determined for each goat genotype using PLINK. The MAF were categorized into 0.0- 0.1, 0.1-0.2, 0.2-0.3 and 0.4-0.5. Population structuring and genetic relatedness were estimated by principal components analysis (PCA) in TASSEL and visualized in R-package ggplot2 (Zheng *et al.*, 2012).

Results

Breed genetic diversity. Inbreeding coefficient and observed heterozygosity (Ho) ranged from -0.01 to 0.07 and 0.56 to 0.60 respectively (Table 1). A negative F value was recorded in the KDPG breed indicating heterozygote excess contrary to the deficit (positive F-values) in other goats. Mean Ho were generally close to the expected heterozygosity (He) than He levels were 0.43.

Population genetic structure. Principal component analysis (PCA) was used to investigate the genetic relationships among the dairy goat genotypes. The PCA showed three distinct clusters (Figure 1) with the principal components one (PC1) and two (PC2) accounting equally (50%) to the genetic variation. The three PCA clusters are; cluster 1 (Saanen, Alpine and their crosses), cluster 2 (Toggenburg, Alpine, TKDPG and their crosses, and cluster 3 (KDPG, Toggenburg, TKDPG, Alpine and their crosses).

Discussion

Our study revealed some levels of diversity within the 243 goat genotypes. Toggenburg Kenya Dual Purpose Goat (TKDPG) genotype was the most inbred genotype with an inbreeding coefficient of 0.07. KDPG as non-inbred (-0.01). The studied goat genotypes reserved high levels of genomic diversity as shown by mean Ho and He above 0.39 and 0.42, respectively (Table 1). While these heterozygozities are high, they are in the close range to those described in other goats (Onzima *et al.*, 2018).

Table 1. Inbreeding coefficient, observed and expected heterozygosities for different goat genotypes.

Genotype	Inbreeding coefficient (F)	Observed heterozygosity (Ho)	Expected heterozygosity (He)
Alpine	0.0377	0.411	0.427
KDPG	-0.0165	0.434	0.427
Saanen	0.0292	0.415	0.427
TKDPG	0.0799	0.393	0.427
Toggenburg	0.0298	0.414	0.427
Crosses	0.0406	0.410	0.427

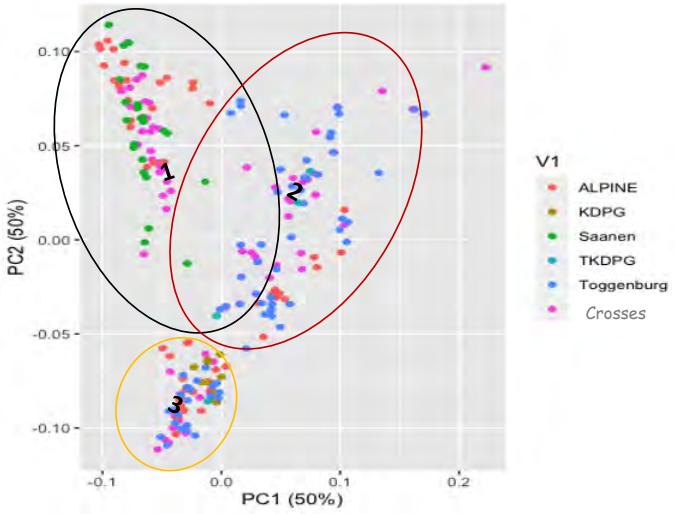


Figure 1. PCA plots for the first principal components (PC1) and second principal component (PC2) and their respective variations.

Principal component of population structure confirmed distinctiveness among the goat genotypes (Saanen, KDPG, Crosses, Toggenburg, TKDPG and Alpine) in Kenya. PCA showed 3 distinct clusters (Figure 1) corresponding to their relatedness between populations. Principal component analysis clustered the populations into three distinct clusters namely; cluster one-Saanen, Alpine and their crosses, cluster two-Toggenburg, Alpine, TKDPG and their crosses, and cluster three-KDPG, Toggenburg, TKDPG, Alpine and their crosses. This clustering revealed that admixture has occurred and genetic links exist between the genotypes due to crossbreeding. Genomic analysis of goat genotypes revealed three distinct genetic pools, and these genetic relationships is essential for selection and breeding in the ongoing breeding program at KALRO, Naivasha.

References

Ajmone-marsan, P. *et al.* (2014) 'The characterization of goat genetic diversity: Towards a genomic approach', 121, pp. 58–72. https://doi.org/10.1016/j.smallrumres.2014.06.010.

Alexander, D. H., Novembre, J. and Lange, K. (2009) 'Fast model-based estimation of ancestry in unrelated individuals', pp. 1655–1664. https://doi.org/10.1101/gr.094052.109.vidual.

Kosgey, I. S. and Okeyo, A. M. (2007) 'Genetic improvement of small ruminants in low-input, smallholder production systems: Technical and infrastructural issues', *Small Ruminant Research*, 70(1). https://doi.org/10.1016/j.smallrumres.2007.01.007.

Letunic, I. and Bork, P. (2019) 'Interactive Tree Of Life (iTOL) v4 : recent updates and'. Oxford University Press, 47(April), pp. 256–259. https://doi.org/10.1093/nar/gkz239.

Ngeno, K. *et al.* (2021) 'Performance of Dairy Goat Genotypes in different Production Systems in Kenya', 20(1), pp. 111–117.

Onzima, R. B. *et al.* (2018) 'Genome-wide characterization of selection signatures and runs of homozygosity in Ugandan goat breeds', *Frontiers in Genetics*, 9(AUG), pp. 1–13. https://doi.org/10.3389/fgene.2018.00318.

Peacock, C. *et al.* (2011) 'Meru dairy goat and animal healthcare project Meru dairy goat and animal healthcare project', 5903. https://doi.org/10.3763/ijas.2010.0571.

Purcell, S. *et al.* (2007) 'REPORT PLINK : A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses', 81(September), pp. 559–575. https://doi.org/10.1086/519795.

708. Resolving the genetic structure of 18 Greek and neighbouring sheep populations

A. Kominakis[1*], E. Tarsani[1], A. Hager[1], G. Antonakos[2], I. Mastranestasis[3], D. Gkelia[4] and I. Hadjigeorgiou[1]

[1]Department of Animal Science, Agricultural University of Athens, Iera Odos 75, 11855 Athens, Greece; [2]Agricultural and Livestock Union of Western Greece, 13th km N.R. Agrinio-Ioannina, 30100 Lepenou, Greece; [3]Breeder's Association of the Lesvos sheep, 81109 Anaxos Lesvos, Greece; [4]Association of Pastoral Farmers of Epirus, H. Trikoupi 36, 45332 Ioannina, Greece; acom@aua.gr

Abstract

Aim of the present study was to explore the genetic relationships of 18 sheep breeds originating from Greece (n=9), Albania (n=4), North Macedonia (n=2), Turkey (n=1), Romania (n=1), and Cyprus (n=1). To this end, admixture, population structure and phylogenetic analyses were performed. Results of admixture analysis showed the fat-tailed sheep (Chios, Sakiz, and Cyprus fat tailed) to be genetically differentiated from the rest breeds. Both admixture and population structure analyses revealed Katsika (Greece) and Karakachanska (North Macedonia) sheep to be clearly distinct from the rest sheep breeds. Phylogenetic analysis distinguished the Greek sheep breeds from Albanian (Shkodrane, Ruda, Lara and Recka), North Macedonian (Ovchepolean and Karakachanska), and the Romanian Tsigaia sheep. Current findings disclose the genetic relationships between the studied sheep breeds and are relevant in terms of priorities in conservation actions for rare indigenous sheep breeds.

Introduction

Interrogation of 50k genetic loci from 74 world sheep breeds revealed high levels of admixture during the development of modern sheep breeds, as well as a major migration route along an axis running from South-West Asia to the Mediterranean region, and then via central Europe to Britain and the Nordic regions (Kijas *et al.* 2012). The Balkan region has served as a hub along this migration route, with sheep populations exhibiting significant genetic overlap and distinctness from breeds in neighbouring regions (Ciani *et al.* 2020). Nevertheless, due to the limited number of breeds used, the genetic relationships among southern Balkan sheep populations remain elusive. Driven from the scarcity of relevant reports, the current study aimed to provide an insight into this part of the world's sheep breeds map.

Materials & methods

Data and quality. control Seven Greek sheep populations (Lesvos, Lemnos, Skyros, Oreino, Katsika, Agrinio and Peloponnesus) were sampled and genotyped using the Illumina SNP 50k array. Application of typical quality control criteria at the marker level (call rate>0.95, MAF>0.05, autosomal markers) resulted in a total number of 42,234 SNPs available. Quality control was also performed at the sample level and samples with a call rate<0.95 and an identity by descent degree of recent shared ancestry ($\hat{\pi}$) threshold of 0.1875 (halfway between third- and second-degree relatives) were removed. We also mined repository genotypic data on two more Greek sheep breeds (Kymi and Chios) and nine neighbouring sheep breeds from Albania (Shkodrane, Ruda, Lara and Recka), North Macedonia (Ovchepolean and Karakachanska), Turkey (Sakiz), Romania (Tsigaia) and Cyprus (Fat-tailed) made available by Ciani *et al.* (2020). A total number of 223 genotypes and 19,816 SNPs on 18 sheep breeds (see Figure 1 for details) were finally remained for further genetic analyses.

Genetic analyses. Admixture analysis was first conducted, with the *ADMIXTURE* software, assuming K (ancestral populations) in the range from 2 to 7 and a five-fold CV procedure to determine the optimal

R.F. Veerkamp and Y. de Haas (eds)
CC BY

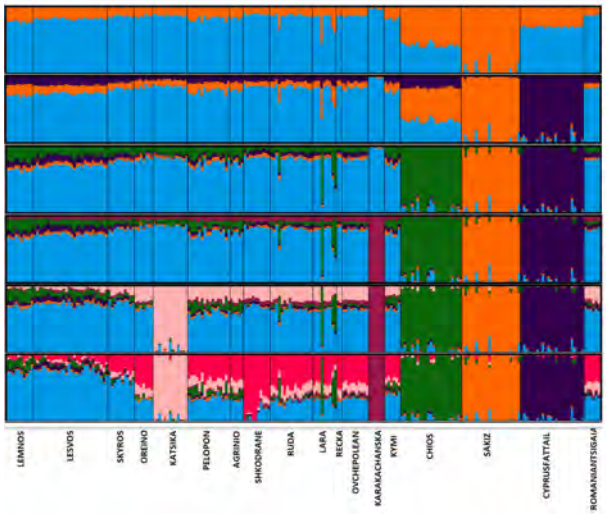
DOI: 10.3920/978-90-8686-940-4_708, © A. Kominakis *et al.* 2022

K-value presenting the lowest CV error. Admixture results were graphically visualized using Distruct from *CLUMPAK*. Population structure was then assessed via the IPCAPS (Chaichoompu *et al.* 2019) R package. *IPCAPS* assigns individuals to subpopulations and infers the total number of subpopulations present via utilizing an unsupervised clustering algorithm and an iterative process with multiple splits. Finally, pairwise Reynolds' genetic distances between breeds were calculated using the stamppNeisD' function in the *StAMPP R* package and were graphically visualized using NeighborNet graphs constructed in *SplitsTree5*.

Results

Results from admixture analysis (for K=2 to K=7) are shown in Figure 1. With the number of clusters (K) set at 2, only one distinct cluster was generated with animals of the Sakiz breed (orange color). This cluster was persistent from K=2 to K=7. At K=3, the Cyprus fat-tailed sheep (dark blue color) first emerged and remained as distinct group from K=3 to K=7. At the optimal step (K=4), another cluster was generated with the Chios breed genotypes (green color). Finally, the Karakachanska sheep (purple color) and the Katsika sheep (light pink color) were revealed as distinct genetic clusters at higher K values (K=5 and K=6, respectively).

Clustering analysis by IPCAPS resulted in a total of 13 groups (Figure 2). Here, the Karakachanska and Katsika breeds formed distinct clusters (1 and 8, respectively). Furthermore, the Cyprus fat tail genotypes were distributed across 3 clusters (2-4) not overlapping with any other breed and this was also the case for the Sakiz sheep (clusters 5-7). Furthermore, genotypes from east Aegean islands (Lemnos, Lesvos) along with those from Skyros and Kymi clustered together with Agrinio, Oreino and some Peloponnesus genotypes (cluster_11), whereas cluster_12 included genotypes from Lara, Oreino, Ovchepolean, Peloponnesus, Tsigaia, and Ruda sheep. Finally, cluster_13 consisted of the Chios genotypes and one Lara genotype (Figure 2).



Breed/population	Country	n
Agrinio	Greece	5
Chios	Greece	23
Fat tailed	Cyprus	24
Karakachanska	North Macedonia	6
Katsika	Greece	13
Kymi	Greece	6
Lara	Albania	9
Lemnos	Greece	10
Lesvos	Greece	28
Oreino	Greece	7
Ovchepolean	North Macedonia	10
Peloponnesus	Greece	16
Recka	Albania	2
Ruda	Albania	16
Sakiz	Turkey	22
Shkodrane	Albania	10
Skyros	Greece	10
Tsigaia	Romania	6

Figure 1. Ancestry proportions (left) and number of samples (right) of 18 Greek and neighbouring sheep breeds. The x-axis represents individuals grouped according to breed. Each colour represents a population and each individual is represented by a vertical line partitioned into colored segments whose lengths represent the admixture proportions from K=2-7 ancestral populations.

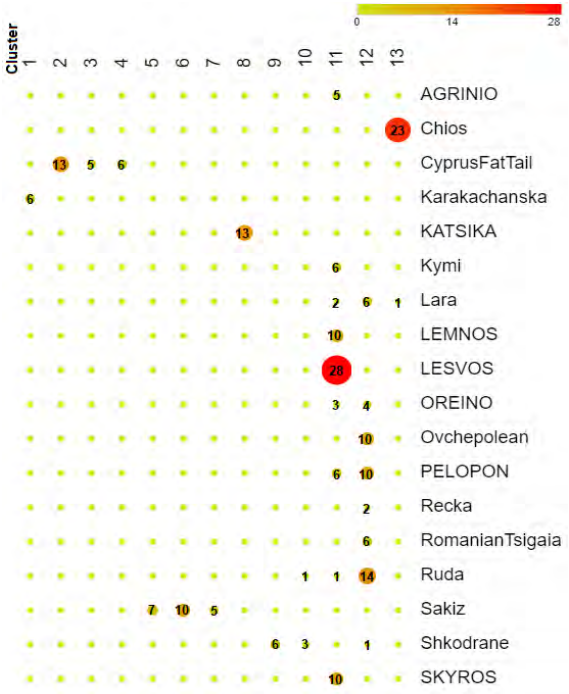


Figure 2. Distribution of sample genotypes across the inferred clusters obtained by IPCAPS (numbers indicate the number of animals per breed assigned to clusters).

Phylogenetic network (Figure 3) revealed two distinct geographically corresponding branches: one with the Karakachanska, Shkodrane, Ruda, Ovchepolean, Recka and Tsigaia breeds (west-north) and the other (east) with the rest of the breeds. One sub-branch emerged at the base of the latter branch, consisting of Greek sheep from Peloponnesus, Epirus (Katsika and Oreino), Kymi and Agrinio. Finally, a major east

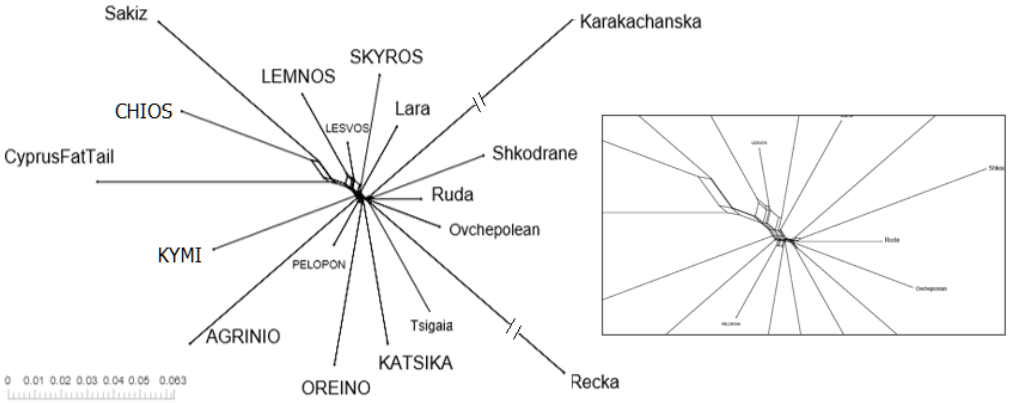


Figure 3. Phylogenetic network based on the estimated pairwise Reynolds' genetic distances between 18 Greek and neighboring sheep herds. Long terminal distances reflect small samples. Breed names in capital letters are of Greek origin.

sub-branch emerged, consisting of one group of Aegean islands populations (Skyros, Lesvos, and Lemnos) along with the Lara breed (Albania) and another outgroup formed by the fat-tailed breeds (Sakiz, Chios, and Cyprus fat tail). For the Greek breeds, an east-west genetic cline was noticed.

Discussion

In the present study, we attempted to resolve the population structure and phylogenetic relationships between nine Greek and nine neighbouring sheep populations. Our results exhibited a genetic differentiation of the Albanian and North Macedonian sheep from their Greek counterparts, and this was obvious even for geographically neighbouring breeds, such as those from Epirus (Greece) (Katsika and Oreino) and Albania. Our results also confirmed previous findings, as they showed: (1) a geography-related genetic differentiation for sheep from Albania and Greece (Pariset *et al.* 2011); (2) an east-west genetic cline for the Greek breeds (Ciani *et al.* 2020, Kominakis *et al.* 2021); and (3) close genetic relationships between the Albanian sheep breeds (Shkodrance, Recka and Ruda) (Hoda *et al.* 2011). Finally, since some of the sheep breeds evaluated here are endangered (e.g. Agrinio, Oreino, Katsika, Shkodrane, Lara, Karakachanska), the current findings could be used to design and prioritize future conservation efforts.

References

Chaichoompu K., Abegaz F., Tongsima S. *et al.* (2019). Source Code Biol Med. https://doi.org/10.1186/s13029-019-0072-6.

Ciani, E., Mastrangelo, S., Da Silva, A. *et al.* (2020). Genet. Sel. Evol. 52(1) https://doi.org/10.1186/s12711-020-00545-7.

A. Hoda, G. Hykaj, L. Sena *et al.* (2011). Acta Agriculturae Scand. Section A, 61:12-20, https://doi.org/10.1080/09064702.2010.542250.

Kijas J.W., Lenstra J.A., Hayes B., Boitard S., Porto Neto L.R. *et al.* (2012). PLoS Biol. 10(2):e1001258. https://doi.org/10.1371/journal.pbio.1001258.

Kominakis A, Tarsani E, Hager-Theodorides AL *et al.* (2021). PLoS One. 2021 Sep 16;16(9):e0257461. https://doi.org/10.1371/journal.pone.0257461.

Pariset L, Mariotti M. Gargani M., Joost S., Negrini R. *et al.* 2011. TheScientificWorldJournal, 11, 1641–1659. https://doi.org/10.1100/2011/186342.

709. Genetic determinism of quality of lamb leather in Lacaune dairy sheep breed

H. Larroque[1], E. Dugas[1], F. Woloszyn[1], F. Plisson-Petit[1], J. Gayraud[2], J.-C. Duchêne[3], J. Frayssignes[4], A. Gille[4], S. Brenot[5], D. Allain[1] and L. Drouilhet[1*]

[1]GenPhySE, Université de Toulouse, INRAE, ENVT, 31326 Castanet Tolosan, France; [2]OVITEST, 12780 Saint-Léon, France; [3]ALRIC mégisserie, 12100 Millau, France; [4]UNICOR, 12000 Rodez, France; [5]GenESI, INRAE, 17700 Surgères, France; laurence.drouilhet@inrae.fr

Abstract

Lacaune lamb skins are well known for their high quality and their use in the leather industry and luxury ready-to-wear clothing. In France, for several decades, the quality of this leather has deteriorated, mainly due to two skin defects: pinhole and straw-like bushiness. To explore the genetic determinism of these defects, around 1,400 Lacaune lambs, from 72 rams, were phenotyped for the 2 skin defects and 4 wool traits. The pinhole defect showed a high heritability (0.59±0.07) and was highly genetically correlated to the wool predicted core bulk (r_g=0.54±0.08). The straw-like bushiness defect had a lower heritability (0.25±0.06), and was correlated to the mean fiber diameter (r_g=0.41±0.15). A genome wide association study revealed for the pinhole defect one major signal on chromosome 3 close to a cluster of keratin genes, known to play an important role in skin development.

Introduction

Lamb skins are generally considered to be a by-product of sheep production and genetic studies of traits directly related to sheep skin are infrequent in the literature. The Lacaune sheep milk breed is bred in south of France, in the Roquefort cheese production area. Lacaune lamb skins are well known for their high quality and their use in the leather industry and luxury ready-to-wear clothing. Therefore, historically, a large majority of French tanneries was located in the same geographical region. However, for several decades, a degradation of this leather quality was observed, mainly due to the presence of two skin defects: pinhole and straw-like bushiness. Pinhole consists of small holes in the grain surface of the dewooled pelt, which decreases the leather quality. The straw-like bushiness defect is defined as perforations due to the presence of forage debris planted in the thickness of the dermis. In the present study, we explored the genetic determinism for these two main skin defects and their possible relationship with wool traits.

Materials & methods

Animals and phenotypes. Around 1,400 Lacaune lambs (from 72 rams) were born around February 2020 and 2021 (in 16 different farms) and identified at birth with electronic ear tags. At around 35 days-old lambs were brought together in a unique fattening farm. Around 60 days-old, a blood and a wool samples of each animal were collected (they were carried out in the fattening herd within the framework of the usual practices of breeding by authorized persons from OVI-test AI organization). After an average fattening period of 79 days, animals were slaughtered and their skins were sent to the ALRIC tannery to proceed to the tanning. Each skin was identified in order to be assigned to one lamb. The two defects were scored at the end of the tanning procedure: skin pinhole score (from 0=absent to 4=abundant corresponding to several heavily marked regions), and straw-like bushiness score (from 0=absent to 2=abundant). Wool samples were analysed by Optical Fibre Diameter Analyser (OFDA) for determining mean fibre diameter (MFD, µ), coefficient of variation of fibre diameter (CVD, %) and mean fibre curvature (MFC, °/mm). Predicted core bulk (PCB), expressed in cm3/g, represents the occupied volume by 1 gram of wool under a pressure of 10^{-3}

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_709, © H. Larroque *et al.* 2022

N (Sumner *et al.*, 2005). At the end, 1379 lambs were phenotyped for wool traits, 1,166 for leather defects, and 1,158 for both (Table 1).

SNP genotypes. DNA of a subset of 1,248 animals was extracted from blood sample and genotyped using the Illumina OvineSNP50 BeadChip. Genomic analyses were performed using 1,248 phenotyped and genotyped lambs, plus their 72 rams' genotypes (available from a previous project). SNP quality was checked (i.e. SNP call rates >95%, minor allele frequency >5%, inconsistent for Hardy-Weinberg disequilibrium). Finally, 38,439 autosomal SNPs were retained and 14 lambs failed QC and were removed.

Statistical analysis. Variance-covariance components were estimated with the airemlf90 software (Misztal *et al.*, 2002) using pedigree information (8,575 animals over 6 generations) combined with SNPs information. Heritabilities were estimated with linear single trait animal models, and genetic correlations with linear bivariate animal models. For leather defects, the models included environmental fixed effects of birth herd (16 levels), year×housing condition during the fattening period (3 levels), litter size (2 levels: 1, 2 or more), duration of fattening period (as co-variable, 79±15.71 d) and a random additive genetic animal effect. For wool traits, the models were the same excepting for duration of fattening period replaced by the age at sampling (61±12.07 d). Genome wide association analysis (GWAS) was performed using GEMMA software (Zhou and Stephens, 2012) using phenotypes adjusted for environmental effects defined above with the procedure GLM of SAS software, the threshold of significance was at the genome level (probability of 0.05 divided by the number of SNP used).

Results

Phenotypes characteristics of leather and wool traits are presented in Table 1. The mean of the pinhole score was 1.47, and 25% of skins did not express the defect whereas the percentages of skins classified of 1 to 4 were of 28, 27, 17 and 4% respectively. For Straw-like bushiness, the mean was 0.86, and 45% of skins showing no defect, whereas 27% presented this defect with a low density (score 1) and 28% with a high density (score 2). Straw-like bushiness presented a moderate estimated heritability of 0.25, whereas the other traits had a high estimated heritability ranging from 0.48 for MFD to 0.61 for MFC (Table 1).

Phenotypic and genetic correlations estimated between each leather defect and wool traits are shown in Figure 1. The two leather defects showed low to moderate phenotypic correlations with wool traits and reach 0.33 and 0.17 with PCB (for pinhole and straw-like bushiness defects respectively). Although the two defects were genetically highly correlated (0.66±0.11), their correlations with wool traits were somewhat different. Pinhole defect presented a strong positive genetic correlation with MFC (0.38±0.09) and PCB (0.54±0.08). Lambs with high MFC or high PCB wool had genetically a skin with more pinhole defects. Straw-like bushiness defect showed a moderate positive genetic correlation with PCB (0.31±0.14) and a higher genetic correlation with MFD (0.41±0.15). The two defects were related to PCB.

Table 1. Summary of phenotypes used for estimation of genetic parameters and heritability estimations.

	Records	Mean	Std	Heritability	SE[1]
Pinhole	1,166	1.47	1.14	0.59	0.07
Straw-like bushiness	1,166	0.86	0.84	0.25	0.06
MFD	1,379	19.31	2.02	0.48	0.07
CVD	1,379	25.75	5.85	0.51	0.06
MFC	1,379	93.45	10.38	0.61	0.06
PCB	1,379	28.02	1.43	0.57	0.06

[1] Approximated standard error of heritability.

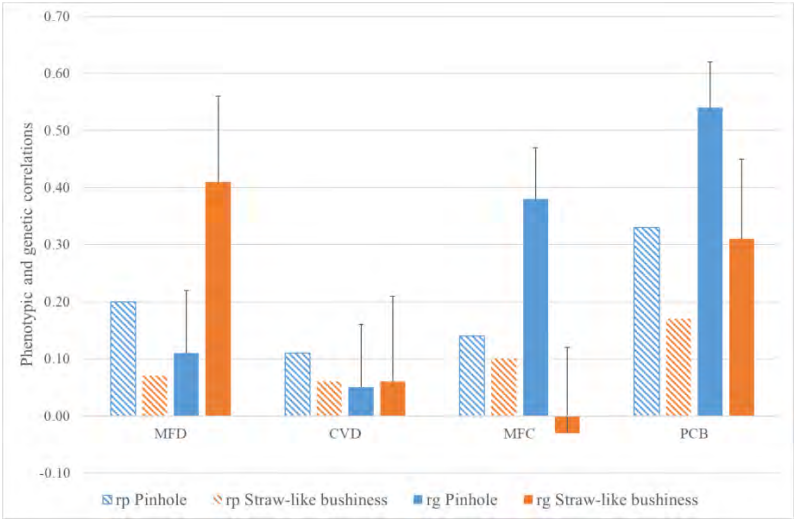


Figure 1. Phenotypic and genetic correlations between leather defects and wool traits (rp: phenotypic correlation, rg: genetic correlation, vertical line: approximated standard error).

Genome wide association studies were performed on 1,147 animals for the pinhole defect and the straw-like bushiness defect. A significant signal was detected at the genome-wide threshold for the pinhole defect, localized on OAR3 (Figure 2).

We identified the most significant SNP at a position of 146.2 Mb on the ovine reference genome (Oar_rambouillet_v1.0). Considering the significant region at the chromosome-wide threshold, it encompasses 4 Mb and the cluster of Keratin genes localised between 143-144 Mb. In contrast, no significant signal was found for the straw-like bushiness defect.

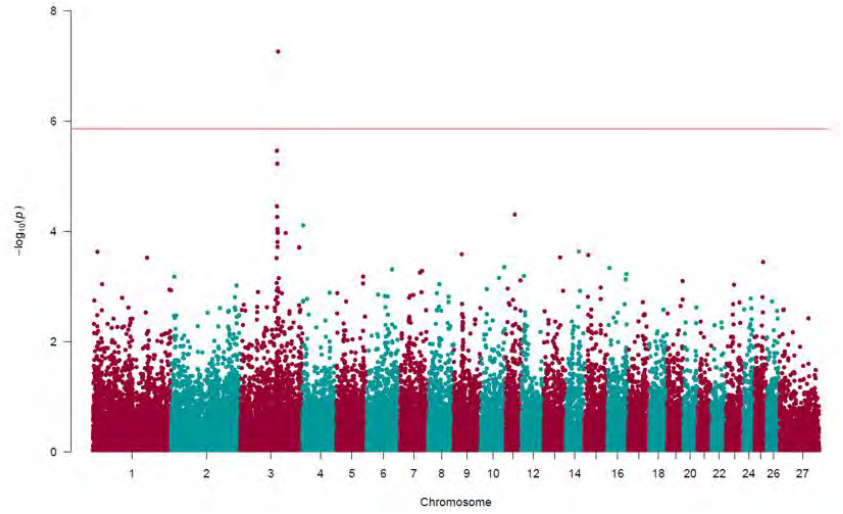


Figure 2. Manhattan plot for GWAS of pinhole defect. This plot is based on −log 10 (*P*-value) from GWAS against chromosome position, each colour represents different chromosome. Red line indicates genome-wide significant threshold.

Discussion

A defect similar to the pinhole defect was studied in New-Zealand and Australia. We found an estimation of pinhole heritability of 0.59, which was close to the estimation of 0.6 found by Campbell *et al.* (1997) on fewer animals. We have found a high genetic correlation (0.54) between pinhole defect and PCB, a higher level of wool PCB being related to a higher presence of pinhole defect. This can be related to previous studies carried out in New-Zealand, where a phenotypic positive relationship was found between the presence of the pinhole defect and PCB (Scobie *et al.*, 1997; Cooper, 2002). In these previous studies, high PCB was related to a more pronounced curvature of fibers. These authors proposed that when removing the fleece, fibers that have a high curvature remove part of the collagen matrix at the same time, affecting the skin (Scobie *et al.*, 1997).

No data on genetic parameters for the straw-like bushiness defect are available in literature. Its heritability estimation was 0.25, suggesting a large influence of the environment on this skin defect. Length of the fattening period (highly correlated with age at slaughter) seemed to influence the presence of this defect. Indeed, lambs with a longer fattening period had more straw-like bushiness defects, which could be linked to a longer exposition of their skin to the plant debris.

Our genetic study led to the identification of a region nearby the cluster of Keratin genes localized on OAR3. Keratin genes is a family of structural fibrous proteins also known as scleroproteins. It is the key structural material making up scales, hair, nails, feathers, horns, claws, hooves, and the outer layer of skin among vertebrates. Keratin also protects epithelial cells from damage or stress. Sequencing of finely chosen animals is currently on-going to identify one or several candidate mutations in our region of interest.

For pinhole defect, its high heritability, its strong genetic correlation with a wool trait measurable on live animals and the detection of a QTL affecting it, indicate that a genetic selection could be considered in order to decrease its frequency in Lacaune breed. However, additional studies must be conducted in order to better understand its genetic determinism and to evaluate its relationship with traits currently selected in this dairy sheep breed.

Acknowledgements

This project has received funding from Occitanie region's Readynov program 'L'A-PROPOS'.

References

Campbell I. P., Hopkins A. F., and Kenney P. A. (1997). Proc. of AAABG, Dubbo, Australia.

Cooper S. M. (2002). Proc. of NZSAP, Palmerston North, NZ.

Misztal I., Tsuruta S., Strabel T., Auvray B., Druet T. *et al.* (2002). Proc. of the 7th WCGALP, Montpellier, France.

Scobie D. R., Cooper S. M., Passman A., and Young S. R. (1997). Proc. of NZSAP, NZ.

Sumner R. M. W., Baxter B. P., and Upsdell M. (2005). Int J of Sheep and Wool Sci., 57(1).

Zhou X., and Stephens M. (2012). Nat Genet 44, 821–824. https://doi.org/10.1038/ng.2310

710. Heritability and genomic analysis of coagulation event in sheep milk

G. Gaspa[1*], A. Cesarani[2], F. Correddu[2], M. Congiu[2], C. Dimauro[2], A. Pauciullo[1] and N.P.P. Macciotta[2]

[1]*Department of Agricultural Forest and Food Science, University of Turin, Largo Braccini 2, 10095 Grugliasco, Italy;* [2]*Dipartimento di Agraria, University of Sassari, Viale Italia 39, 07100 Sassari, Italy;* *giustino.gaspa@unito.it*

Abstract

Sheep farming plays an important role in the rural economy and about 250 million sheep are reared for dairy purpose worldwide. The whole amount of sheep milk is used for cheese making and the presence of non-coagulating milk affects negatively economics of dairy industry. Beside the effects exerted by the environmental factors, additive genetic individual differences in coagulation exist and allow genetic improvement. The aim of this work was to study differences between coagulating and non-coagulating milk samples, to estimate h^2 and conducting a genomic association study on milk coagulation phenotypes. A total of 8.7% samples did not coagulate within 30 min and a difference in milk composition between normal and non-coagulating samples was observed. Genetic analysis allows both to estimate h^2 for the coagulation event of 0.23 (0.04) and to suggest 45 genes involved in mammary gland metabolism and udder health status.

Introduction

Sheep farming plays an important role in the rural economy, in particular for Mediterranean area, Asian, and developing countries. About 250 million of dairy purpose sheep farmed worldwide produce about 10 million t of ovine milk (Pulina *et al.*, 2018), which is nearly all processed into cheese. Thus, the ability of milk to coagulate properly is crucial. Milk coagulation is assessed by individual laboratory rennet coagulation time (RCT) and curd firmness (A_{30}) (i.e. the time between the addition of rennet to milk and the beginning of the clot formation and the curd consistence, respectively). The combination of low RCT and high A_{30} are usually associated with higher cheese yield (De Marchi *et al.*, 2008). However, a variable proportion of individual milk samples of different species did not coagulate within this time range. The latter aspect is strongly associated with milk composition (e.g. protein, lactose, pH, and somatic cells) (Bittante *et al.*, 2012). In small ruminants, beside the effects exerted by different environment and management, individual differences exist due to the additive animal variability (Puledda *et al.*, 2017). Thus, the coagulation ability could be improved though breeding schemes. Moreover, genome-wide analysis (GWA) of non-coagulating milk allowed to map QTL in cattle (Duchemin *et al.*, 2016) but a few reports are available in dairy ewes (Marina *et al.*, 2021). The aim of this work was to study differences between coagulating and non-coagulating milk samples, to estimate the variance components and heritability and to find regions associated with this trait.

Materials & methods

Data. Individual milk samples of 1,018 ewes from 47 flocks located in Sardinia (Italy) were analyzed. Data came from a project aimed to investigate milk coagulation properties (MCP) in Sarda dairy sheep (Manca *et al.*, 2016; Puledda *et al.*, 2017). RCT was determined by Formagraph instrument (Foss Electric A/S, Hillerød, Denmark) for all samples, that were divided in two classes: (1) COAG, samples coagulating within 30 minutes; (2) NON-COAG, samples that did not coagulate in this time range. Milk composition traits (fat, protein, lactose, SCS, chloride and urea) were recorded for all samples to compare the two groups.

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_710, © G. Gaspa *et al.* 2022

Moreover, 769 ewes were genotyped with Infinium Ovine SNP50 v1 BeadChip (Illumina Inc., San Diego, CA). After quality control, 44,619 SNPs on 27 OAR chromosomes remained for GWA study.

Heritability estimation. The following threshold animal model was fitted to estimate variance components and heritability (h^2)

$$y = Xb + Z_1f + Z_2a + e \quad (1)$$

where: y is the vector of coagulation binary trait; b is the vector of fixed effects including the month of lambing (4 levels), days in milk in classes of 30d (5 levels) and parity (3 levels); f is the vector of the random effect of flock-test days of sampling combination (69 levels) ~$N(0, I\sigma^2_{ftd})$; a is the vector of the random additive genetic effects ~$N(0, A\sigma^2_a)$ where A is the numerator relationship (5,031 animals) and e is the vector of random residuals ~$N(0, I\sigma^2_e)$ The σ^2_{ftd} σ^2_a σ^2_e are the flock-test day, additive genetic and residual variances. The X, Z_1 and Z_2 are the incidence matrices relating records to effects. Variance components and heritability were estimated using thrgibbs1f90 software (Misztal *et al.*, 2014), which implements a Gibbs sampling method. The following parameters were used: 50,000 samples were generated, with the first initial 5,000 rounds discarded as burn-in, and all were saved. The heritability was estimated as intra-herd h^2. The value on the liability scale, which is a function of h^2 and incidence of coagulation classes, was transformed on the observed scale.

GWA study and gene discovery. The model (1) was modified under the ssGBLUP framework to accommodate H^{-1} matrix in place of A^{-1} and later used for GWA study according to Cesarani *et al.* (2021). Briefly, the GEBV were back-solved into SNP effects and the additive genetic variance explained by each SNP was computed. The SNP exceeding the 99.9th percentile were retained as suggestive of trait association. Using genome browser (https://genome.ucsc.edu/) genes included in the neighbour of suggestive SNP were retrieved, and a gene-by-gene literature review was performed in order to assess previously association with MCP and milk traits.

Results

A total of 89 milk samples (8.7% of the total) did not coagulate within 30 min from the rennet addition. Average RCT for the COAG group was 15.13±6.5 minutes. Samples of the NON-COAG group showed larger values for protein (5.72 vs 5.38%, P<0.001), chloride (194.5 vs 141 mg/100 ml, P<0.001), and SCS (7.50 vs 4.37, P<0.001), whereas they exhibited lower values for lactose (4.44 vs 4.85%, P<0.001) and urea (33.2 vs 38.4 mg/dl, P<0.001). The heritability of the coagulation ability was moderate (0.23±0.04).

As far as GWA is concerned, a total of 458 SNP overcome the threshold defining 40 QTL regions over 18 chromosomes (autosome and sexual); OAR12, 13, 15, 16, 17, 18, 19, 24 and 25 did not show any associated SNP with failure in coagulation phenotype. The distribution of suggestive SNPs and the proportion of additive genetic variance are presented in Figure 1.

Among the genomic regions identified by tag SNPs, nearby 17 of them, 260 genes were retrieved, but only 45 (Table 1) were involved in cellular function associated to dairy traits in ruminant species (i.e. milk coagulation, milk production and composition, udder health, heat stress resistance).

Discussion

Literature reports quite high percentage of non-coagulating milk samples, both in cows and sheep. In Swedish Red cattle breed, about 18% of milk samples did not coagulate within 30-40 minutes (Gustavsson *et al.*, 2014); lower percentages were reported for Holstein (9.7%; Cecchinato *et al.,* 2011) and Brown Swiss

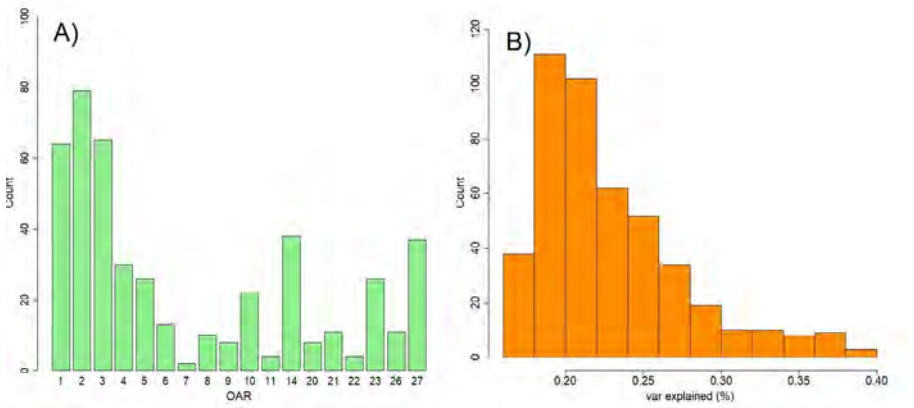


Figure 1. (A) Number of top 0.1% SNP per chromosome and (B) distribution of variance explained by top SNP.

Table 1. QTL regions identified and genes retrieved.

QTL	OAR[1]	Start	End	Peak SNP (max% var)	Mbp	Gene
1	1	19.01	22.59	OAR1_20007545.1	20.03	*AKR1A1,CMPK1,TMEM69, UROD,LRRC41,PIK3R3,TESK2*
2	2	107.07	114.70	OAR2_122611468.1	114.65	*MFAP3L, ARHGEF4, OCA2*
3	2	201.82	202.26	s35200.1	201.86	*BZW1,PPIL3*
4	2	234.72	235.51	s46218.1	235.10	*FABP3, NKAIN, ZCCHC17*
5	3	195.80	198.52	OAR3_211332869.1	196.05	*SLC15A5*
6	3	94.92	95.16	s61740.1	95.16	*CCT7*
7	4	14.35	15.14	OAR4_14557628.1	14.35	*ASNS*
8	5	79.30	79.82	OAR5_87409839_X.1	79.47	*RPS23*
9	5	41.59	42.57	s44617.1	42.16	*CSNK1G2*
10	11	46.99	47.30	OAR11_50094068.1	47.05	*ACE*
11	14	45.38	46.07	OAR14_48226980.1	45.81	*APLP1,CAPNS1,ZNF529, ALKBH6,SYNE4*
12	14	35.07	35.58	s14680.1	35.29	*ZBTB7C, CDH1*
13	20	45.96	46.30	OAR20_50378146.1	46.22	*SLC35B3*
14	21	44.31	45.22	s23338.1	44.96	*CHKA, TCIRG1, CABP2, AIP, CLCF1, NUDT8*
15	22	22.33	22.62	OAR22_26729825.1	22.62	*BORCS7, TRIM8, WBP1L,ARL3*
16	26	6.15	6.75	s41368.1	6.45	*SPCS3, ASB5*
17	X	51.09	52.98	s05480.1	52.39	*TFE3, SLC35A2, AKAP4*

[1] OAR *Ovis aries* chromosomes.

(3.5%; Cecchinato *et al.*, 2011). For sheep milk, about 10% of samples were found to be NON-COAG in studies involving Sarda sheep milk (Manca *et al.*, 2016).

A higher proportion (17.7-19.4%) of NON-COAG samples was reported by Caballeros-Villalbos *et al.* (2017) and Garzón *et al.* (2021) in Manchega sheep breed. In our study, the main differences between the two investigated groups of milk samples seem to be associated with health indicator or involving permeability of mammary gland epithelium, which facilitates the exchanges of molecules from bloodstream to the alveolus of the mammary gland. As far as the heritability was concerned, we did not find reports on coagulation trait as binary trait in sheep. In dairy cattle, threshold model applied on binary outcome rather than linear model on RCT (with the exclusion NON-COAG) halved the h^2 estimates (Cecchinato and

Carnier, 2011). In our case, the h^2 estimate was in agreement with previous report on RCT excluding the NON-COAG samples in different sheep breed (Puledda *et al.*, 2017; Sanchez Mayor *et al.* 2019).

The gene discovery on GWA results allow us to prioritize a set of genes previously associated to milk or cheese related traits in dairy ruminants, such as those affecting somatic cell count (*CHKA*, *TCIRG1*, *PPIL3*), mastitis (*ALKBH6, SPCS3, LRRC41*), either overexpressed in mammary gland in different lactation stage (*CMPK1*) or in different milk ability group (*ZCCHC17*) and involved in lactose synthesis (*SLC35A2*) (Bonnefont *et al.* 2011; Dhorne-Pollet *et al.*, 2012; Ghahramani *et al.*, 2021; Michailidou *et al.*, 2021; Sadovnikova *et al.*, 2021). Many other retrieved genes have been previously associated to milk fat and/or composition in sheep milk. In conclusion, the genes of interest are mostly linked to mammary gland metabolism, udder health status and milk compound known to affect the ability of milk to coagulate. These findings are consistent in explaining the differences in the milk composition observed between COAG and NON-COAG individual milks.

References

Bittante G., Penasa M., and Cecchinato A. (2012). J. Dairy Sci. 95(12):6843-70.

Bonnefont, C. M., Toufeer, M., Caubet, C., Foulon, E., Tasca, *et al.,* (2011). BMC genomics. 12

Caballero-Villalobos J., Garzón Sigler A.I., Oliete B., Arias Sánchez R., Jiménez L., *et al.* (2015) Mljekarstvo 65:138-143.

Cecchinato A, Carnier. (2011). J. Dairy Sci. 94():4214-9.

Cecchinato A., Penasa M., De Marchi M., Gallo L., Bittante G., *et al.* (2011). J. Dairy Sci. 94:4205-4213.

Cesarani, A., Garcia, A., Hidalgo, J., Degano, L., Vicario, D. *et al.* (2021). J. Dairy Sci. 104(5):5719-27

Duchemin, S.I., Glantz, M., de Koning, D.J., Paulsson, M., and Fikse, W.F. (2016). Front. Genet. 2016:7.

Dhorne-Pollet, S., Robert-Granié, C., Aurel, M. R., and Marie-Etancelin, C. (2012). Anim. Genet. 43(2):199-209.

Ghahramani, N., Shodja, J., Rafat, S. A., Panahi, B., and Hasanpur, K. (2021). Front. Genet. 12.

Gustavsson, F., Buitenhuis, A.J., Glantz, M., Stålhammar, H., Lindmark-Månsson, H., *et al.* (2014). Int. Dairy J. 39(1):102-107.

Manca, M.G., Serdino, J., Gaspa, G., Urgeghe, P., Ibba, I., *et al.* (2016). J. Dairy Sci. 99(6):4547-4557.

Marina, H., Pelayo, R., Suarez-Vega, A., Gutierrez-Gil, B., Esteban-Blanco, C. and Arranz, J.J. (2021). J. Dairy Sci. 104:11850-66.

Michailidou, S., Gelasakis, A., Banos, G., Arsenos, G., & Argiriou, A. (2021). Front. Genet. 1293.

Puledda, A., Gaspa, G., Manca, M.G., Serdino, J., Urgeghe, P.P., *et al.* (2017). Animal 11(6):920-928.

Pulina, G., Milan, M.J., Lavin, M.P., Theodoridis, A., Morin, E. *et al.* (2018). J. Dairy Sci. 101(8):6715-29.

Sánchez-Mayor, M., Pong-Wong, R., Gutiérrez-Gil, B., Garzón, A., de la Fuente, L.F., *et al.* (2019). Livest. Sci. 228:76-83.

Sadovnikova, A., Garcia, S. C., and Hovey, R. C. (2021). J. Mammary Gland Biol. Neoplasia 1-16.

Tyrisevä, A.M., Ikonen, T., and Ojala, M. (2003). J. Dairy Res. 70(01):91-98.

711. Genetic parameters across European and North American Alpine goats for two milk production and one udder type traits

M. Teissier[1], L.F. Brito[2], F. Schenkel[3], B. Bapst[4], G. Bruni[5], P. Fresi[6], C. Robert-Granié[1*] and H. Larroque[1]

[1]GenPhySE, Université de Toulouse, INRAE, ENVT, 31326 Castanet-Tolosan, France; [2]Department of Animal Sciences, Purdue University, West Lafayette, IN 47907, USA; [3]Centre for Genetic Improvement of Livestock, Department of Animal Biosciences, University of Guelph, Guelph, ON, Canada; [4]Qualitas AG, 6300 Zug, Switzerland; [5]ARAL, 26013 Crema, Italy; [6]AssoNaPa, 00187 Roma, Italy; christele.robert-granie@inrae.fr

Abstract

The development of an international genomic evaluation may represent an interesting opportunity for small ruminant populations, especially to increase the size of their reference population. In this perspective, 4 countries have decided to evaluate the feasibility of an international caprine genomic evaluation of the Alpine breed. For this purpose, variance-covariance components analysis was performed in order to estimate heritabilities and across countries genomic correlations for 3 traits, milk yield, fat content and rear udder attachment. The heritabilities were accurately estimated and coherent with previous estimations. The genomic correlations were high for rear udder attachment ranging from 0.76 (0.037) to 0.92 (0.004) and for fat content with correlation of 0.78 (0.013). However, they were only estimable in one case for milk yield with a moderate value of 0.45 (0.095). The results for the first 2 traits are encouraging but the results for milk yield need to be further investigated.

Introduction

International genomic evaluations may represent an interesting opportunity for small ruminant populations to pool their efforts to access genomic selection or to expand their reference population (Brøndum *et al.*, 2011), especially for traits that are difficult or costly to measure. In goats, 4 countries, Canada (CAN), France (FRA), Italy (ITA) and Switzerland (CHE) have decided to evaluate the feasibility of an international caprine genomic evaluation in Alpine breed, within the H2020 Smarter project. For that, they shared their phenotypes, pedigrees and genotypes. A preliminary study evaluated their genetic exchanges, as well as the genomic distance between their populations (Teissier *et al.*, 2022). The results of the PCA and Admixture analyses demonstrated that European populations, in particular French and Italian, are very close, while the Canadian populations are further apart. Italy is the country that imports the most animals from France, which may explain the genetic proximity between these two populations. In contrast, the Canadian and European populations are more genetically distant. This might be explained by the little exchange of animals and the geographical distance that separates Canada and the European countries. To develop an international genomic evaluation a prerequisite is that the genetic correlations across countries for common traits under selection are high. The objective of our study was therefore to estimate in Alpine breed the variance components and genetic correlations between the four countries for common traits. We present here the results for 3 traits representative of situations encountered according to the type of trait, the number of countries involved and the level of heritability: milk yield (MYI), milk fat content (FCO), and rear udder attachment (RUA).

Materials & methods

Phenotypes and pedigrees. The phenotypes considered are those currently used by each country for its genetic evaluations. MYI is phenotyped by the four countries, RUA by Canada Italy and France, and FCO is only common to Italy and France. Table 1 present all records shared by each country and used in these analyses. Given the large amount of phenotypes recorded in France and Italy for milk production,

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_711, © M. Teissier *et al.* 2022

we limited the phenotypes for animals born from 2000 to 2010 in France and born from 2002 to 2015 in Italy. For Canada and Switzerland, all MYI phenotypes shared were used. For udder type trait (RUA), all phenotypes from the four countries were used. Even so, France had the highest number of phenotypes, followed by Italy, Switzerland and Canada (Table 1).

The RUA trait is recorded in the three concerned countries according to a scale from 1 to 9. For FCO, France expresses it in g/kg of milk and Italy in dg/kg of milk. MYI in kg is not similar across the 4 countries. Indeed, France calculates MYI for a reference lactation duration of 250 days, Canada for a 305-day lactation, Switzerland evaluates MYI from 1 to 100 DIM and from 101 to 200 DIM so the two quantities were summed up, and Italy uses a multiplicative pre-adjustment for parity, age and number of kids.

In total for MYI analysis, 807,534, 71,276, 25,497 and 8,007 animals were kept in the French, Italian, Canada and Swiss pedigree file respectively. For FCO, 806,877 French and 63,617 Italian animals were in the pedigree file. And for RUA, it contained less animals (474,963 French, 29,721 Italian and 9,434 Canadian animals respectively) (Table 1). The pedigree consists of 91% females and these animals were born between 1936 and 2020 for milk production traits and between 1960 to 2020 RUA.

Genotypes. The four countries also shared their genotypes on the Goat SNP50 BeadChip (Illumina Inc., San Diego, CA, USA). SNP quality was checked with PLINK 1.9 software (Purcell *et al.*, 2007) within breed and country and then by merging the genotypes of the 4 countries (i.e. SNP call rates >90%, minor allele frequency >1%, inconsistent for Hardy-Weinberg disequilibrium, genotypes call rate >90%). Finally, 50,578 SNPs and 6,102 animals were retained, with around 50% of the genotypes coming from France (Table 1).

Statistical analysis. Variance-covariance components were computed via Gibbs sampling, with the gibbs1f90 software (Misztal *et al.*, 2002), using pedigree information combined with SNPs information. For each trait, a multi-trait animal model was used where each country was considered as a different trait and for which the model was the one used in routine national genetic or genomic evaluation. For milk yield traits, random animal and permanent environment effect were added in addition of national fixed effect. For France, fixed effects were herd×year×parity, age_at_kidding×parity×year×area, month_kidding×parity×year×area, length_of_dry_period×parity×year×area; for Italy herd-year-season of calving; for Switzerland lactation, year×season, herd (random); and for Canada herd, lactation and birth. For RUA, models contained one random genetic animal effect and fixed effects: herd×year×parity, age at scoring×year,

Table 1. Summary of data used for the analysis and average heritabilities estimates within each country.

Country[1]	Trait[2]	# pedigree	# genotype	# phenotype	Phenotype mean	Phenotype std	Official h^2	Estimated h^2 (SE)
FRA	FCO	806,877	2,968	1,878,616	37.05	5.17	0.48	0.50 (0.002)
ITA	FCO	63,617	1,061	150,162	339.78	54.31	0.18	0.23 (0.007)
CAN	MYI	8,007	793	7,414	949.47	284.43	0.29	0.32 (0.027)
CHE	MYI	25,497	1,280	66,794	565.74	222.62	0.15	0.11 (0.012)
FRA	MYI	807,534	2,968	1,882,056	872.53	262.72	0.31	0.22 (0.002)
ITA	MYI	71,276	1,061	170,932	524.02	203.95	0.24	0.13 (0.006)
CAN	RUA	9,434	793	5,235	5.49	1.43	0.08	0.25 (0.029)
FRA	RUA	474,963	2,968	326,423	4.50	1.50	0.47	0.38 (0.005)
ITA	RUA	29,721	1,061	22,757	5.59	1.56	0.22	0.19 (0.015)

[1] FRA: France; ITA: Italy; CAN: Canada; CHA: Switzerland.

[2] FCO: fat content; MYI: milk yield; RUA: rear udder attachment.

lactation stage at scoring×year (France); DIM×parity, classifier, herd-season of classification, age(linear + quadratic) (Italy); herd×year×classifier(random), breed, parity, age, DIM (Canada). Covariances across country was estimated only between countries for animal effect. For each trait, heritability estimates obtain in 2-by-2 multi-country model were averaged. Genomic correlations were computed with the countries involved for each trait. However, for MYI it was not possible to estimate correlations across the 4 countries. Only, correlation between France and Italy has proven to be sufficiently accurate.

Results

Phenotypes characteristics are presented in Table 1. For RUA, the mean score is close between Canada and Italy and a little weaker in France in Alpine breed. For FCO, despite the difference of unit, we can observe that the means for the two countries (France, Italy) are close but a little higher in France. For MYI, due to the differences in calculation between countries mentioned above, the means of this phenotype are very different between countries. Italy and Switzerland have lower lactation MYI levels than the other two countries: 524.02 and 565.74 kg respectively in the Alpine breed (Table 1).

All estimates of heritabilities had small to moderate standard error. Averages of estimates with 2-by-2 multi-country models are presented in Table 1. For each country, these estimates are relatively close to the ones used in the national genetic evaluations. MYI had low (0.11 in Switzerland) to moderate (0.32 in Canada) heritabilities. FCO had high a heritability in France (0.50 in ALP) and a moderate one in Italy (0.23). RUA shown comparable estimated heritabilities in Canada and Italy (0.25 and 0.19, respectively) and almost twice as high in France (0.38). Highest differences were observed for Canada between official h^2 (0.08) and estimated h^2 (0.25).

Genomic correlations between countries in the Alpine breed are presented in Table 2. For MYI, only estimates between France and Italy are presented. For RUA, correlations were high ranging from 0.76 between France and Italy to 0.92 between France and Canada. For FCO, the correlation between France and Italy was also high (0.78). Whereas for MYI, the correlation between France and Italy of 0.45 was only moderate.

Discussion

In our study heritabilities estimates were relatively accurate and consistent with the within-countries estimations (Official h^2, Table 1) (Carillier *et al*., 2013; Massender *et al*., 2022). However, the estimations of genomic correlations showed different patterns according to traits. Our results shown that for RUA, despite a low level of genetic connection between the Canadian and European populations and very unbalanced data amounts, the correlations across the 3 countries are high. The definition of the phenotype for this trait seems close between countries. The lower correlations for FCO than for RUA between France and Italy could be due to little difference in calculation of this phenotype. The most surprising result was the

Table 2. Across countries genomic correlations (and SE) in Alpine breed.

Trait[1]	Country	Italy	Canada
MYI	France	0.45 (0.095)	
FCO	France	0.78 (0.013)	
RUA	France	0.76 (0.037)	0.92 (0.004)
	Italy		0.78 (0.002)

[1] FCO: fat content; MYI: milk yield; RUA: rear udder attachment.

failure to accurately estimate genomic correlations for the most prevalent trait (common to all 4 countries) MYI. For this trait, convergence was not reach when using data from Canada and Switzerland. It can be supposed that the estimation of covariances for this trait is the victim of a combination of circumstances: a weak genetic connection except between France and Italy, a very unbalanced amount of phenotypes (more than 80% from France), different calculations of the phenotypes together with different heritabilities and analysis models. Only correlation between France and Italy could be estimated but is relatively low.

These analyses were performed in Saanen breed too and similar results regarding heritabilities as well as genetic correlations have been found on the same traits (results not shown). In the same manner, difficulties in estimating genetic correlations for MYI across countries were encountered in this breed.

The strong genomic correlations estimated for RUA and FCO across countries, especially between France and Italy, are promising for future international genomic evaluations in goats. However, given the importance of MYI in breeding objectives, work will have to be done to better understand the difficulty to calculate the across countries correlations for this trait.

Acknowledgements

The authors thank the partners of the Smarter project and the national organizations that made the data sharing possible. This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 772787 (SMARTER).

References

Brøndum, R.F., Rius-Vilarrasa E., Strandén I., Su G., Guldbrandtsen B., *et al.* (2011) J. Dairy Sci. 94(9):4700–4707. https://doi.org/10.3168/jds.2010-3765

Carillier, C., Larroque H., Palhière I., Clément V., Rupp R., *et al.* (2013) J. Dairy Sci. 96(11):7294–7305. https://doi.org/10.3168/jds.2013-6789

Massender E., Brito L. F., Maignel L., Oliveira H. R., Jafarikia M., *et al.* (2022) J. Dairy Sci. In Press. https://doi.org/10.3168/jds.2021-20558

Misztal, I., Tsuruta S., Strabel T., Auvray B., Druet T., *et al.* (2002) Proc. of the 7th WCGALP, Montpellier, France.

Purcell S., Neale B., Todd-Brown K., Thomas L., Ferreira M. A.R., *et al.* (2007) AJHG 81(3):559–575. https://doi.org/10.1086/519795

Teissier, M., Brito L. F., Schenkel., F, Bapst.,B., *et al.* (2022) 12th WCGALP, Rotterdam, the Netherlands

712. Wool colour traits in Gotland sheep and association with a deletion in ASIP

A.M. Johansson

Department of Animal Breeding and Genetics, Swedish University of Agricultural Sciences, Box 7023, 750 07 Uppsala, Sweden; anna.johansson@slu.se

Abstract

In Gotland sheep several different traits including wool colour traits are recorded in most breeding rams. Here, the distributions of three wool colour traits are described and an association analysis is performed with a five base pair deletion (D_5) in the gene A*SIP*. Phenotypes were available for 96 rams of which 43 also had a D_5 genotype. There was a significant association between presence of D_5 and the trait 'cleanness of colour'. This investigation also confirmed earlier indications of that all Gotland sheep seem to have multiple copies of *ASIP*.

Introduction

Gotland sheep is a Swedish sheep breed famous for their grey curly and shiny wool. Gotland sheep can be of different shades of grey and some have the same grey colour all over the fleece whereas others have different grey shades in different parts or even some white strings in the grey fleece. The history of the breeding and the traits are described in (Näsholm 2005). The heritability for the nuance of fleece colour have been estimated to 0.48 (Näsholm 2005). The undesired phenotype of white strings in the grey fleece is called 'pepper and salt' and has been observed to increase but have a low heritability (Näsholm 2009).

Adalsteinson (1978) suggested an allele at the *agouti* locus together with modifiers to explain the variation in grey colour in Gotland sheep. The *agouti* locus has later been identified as the gene *ASIP*. In a previous study of Gotland sheep and some other Swedish breeds it was shown that the nine base pair deletion (D_9) of the gene *ASIP* was present in all investigated Gotland sheep and in most grey sheep of other Swedish breeds (Rochus *et al*., 2019). Another deletion of five base pairs (D_5) of *ASIP* have been shown to be associated with black colour in homozygotes for D_5 (Norris and Whan 2008, Rochus *et al.* 2019). Norris and Whan (2008) also identified a duplication of *ASIP* as the dominant white allele at the *agouti* locus. In the study of Rochus *et al.* (2019) it was shown that Gotland sheep most likely have multiple copies of *ASIP* since they all were heterozygous for most of the polymorphic sites. The only site with more than one genotype in *ASIP* in Gotland sheep was regarding D_5, where some sheep were heterozygous for the deletion and the non-deleted allele (D_5N_5) and some were homozygous for the non-deleted allele (N_5N_5). In the study by Rochus *et al.* (2019) different grey shades were not considered but all grey sheep were grouped into the same phenotype. In this study I use a more detailed description of the phenotype and investigate the association of the D_5 mutation with three colour traits in Gotland sheep.

Materials & methods

Records from 96 rams of Gotland sheep from the Gotland sheep association were used and nasal swab samples were taken using Performagene™ from DNA Genotek, Canada. The data were derived from the routine ram examination that is usually done for potential breeding rams of this breed. Most of the rams (64) were young rams born in 2020, but 32 of them were older rams (born 2015-2019). The rams were from 10 different farms. There were three colour phenotypes: (1) colour (categories: white, extra light grey, light grey, grey, dark grey, black grey); (2) cleanness of the colour (categories: medium, clean, blue); and (3) the undesired trait 'pepper and salt' with white strings in the grey fleeces (categories: no, some, yes, a lot). The distributions of the three colour phenotypes in the 96 rams are shown in Table 1.

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_712, © A.M. Johansson 2022

Table 1. Distribution of colour phenotypes in the 96 rams.

Main colour		Cleanness of colour		Pepper and salt	
Colour	**Number**	**Cleanness**	**Number**	**Presence**	**Number**
white	1	medium	82	no	46
extra light grey	1	clean	13	some	33
light grey	13	blue	1	yes	15
grey	60			a lot	2
dark grey	17				
black grey	4				

DNA extraction and Sanger sequencing was done by a research technician at Department of Animal Breeding and Genetics, Swedish University of Agricultural Sciences. DNA was extracted from the nasal swabs according to the instructions from the manufacturer. The first part of the gene *ASIP*, including the sites for the nine base pair deletion (D_9) and the five base pair deletion (D_5), was sequenced with the same primer pair and the same conditions as in Rochus *et al.* (2019). The DNA quality from the nasal swabs was not good enough for getting high quality sequences in all samples. Good enough sequences in the region encompassing the D_5 mutation were obtained from 45 samples and for D_9 from 56 samples. The D_5 allele is easiest to detect in Gotland sheep in the reverse sequence due to the proceeding heterozygosity for the D_9 allele in all Gotland sheep in the forward sequence. For the D_5 allele it is the sequence AGGAA that is deleted in the forward sequence and thus it is the sequence TTCCT that is deleted in the reverse sequence. The non-deleted alleles for the sites of D_9 and D_5 are denoted N_9 and N_5, respectively.

The sequences were examined with the software 4peaks (from Nucleobytes, the Netherlands). The Fisher's exact test for testing of association between D_5 genotype and the three colour phenotypes was done in R (R Core Team 2021). For the Fisher's exact test, phenotype groups with small numbers were combined with one or two other groups, i.e. for the main colour dark grey plus black grey were in one group and the lighter colours in one group, for the pepper and salt phenotype 'no' was in one group and the phenotypes with some degree of pepper and salt was in the other group, for cleanness of colour the medium was in one group and clean and blue was combined into the other group (see footnotes in Table 2). In this way was there was 2×2 contingency tables for Fisher's exact test for all three tests.

Results

In agreement with the results from Rochus *et al.* (2019) the D_9 mutation was observed in heterozygous form in all samples where that region was possible to analyse (56 samples). This supports that there is a duplication of *ASIP* in Gotland sheep. The D_5 mutation was more difficult to accurately determine the

Table 2. Distribution of colour phenotypes among the 43 samples with good *ASIP* sequence in the D_5 region and phenotype.

Main colour			Cleanness of colour			Pepper and salt		
Colour	D_5N_5	N_5N_5	**Cleanness**	D_5N_5	N_5N_5	**presence**	D_5N_5	N_5N_5
light grey[1]	1	2	medium	14	21	no	12	10
grey[1]	17	14	clean[1]	6	1	some[1]	6	9
dark grey[1]	2	4	blue[1]	1	0	yes[1]	3	2
black grey[1]	1	2				a lot[1]	0	1

[1] Combined in Fisher's exact test.

genotype, since the peaks from the haplotype with D_5 were always lower than the haplotype without a deletion (see an example in Figure 1). Thus, in some samples it was difficult to distinguish a possible D_5 haplotype from background noise. The presence or absence of D_5 was clearly determined in 45 samples but two of these had some unclarities in the linking of IDs to phenotypes. Only the 43 samples that both had been successfully genotyped for D_5 and had undoubtful phenotypes were used for association analysis with the colour phenotypes. Out of these 43 sheep 21 were D_5N_5 and 22 were N_5N_5. The phenotype distribution among these two groups can be seen in Table 2. The Fisher's exact test showed that the distribution of the phenotype cleanness was significantly different (P=0.021) between the sheep with D_5 and those without D_5. For the other two phenotypes there were no significant differences. The six rams with phenotype 'clean' and genotype D_5N_5 are born in four different farms (3 of them from one farm) and all six are thus not closely related. The ram with phenotype 'clean' and genotype N_5N_5 is born in the same farm as three with phenotype 'clean' and genotype D_5N_5.

Discussion

The nasal swabs are less likely to give good enough DNA quality than blood samples or tissue samples, but the advantages are that it is a non-invasive testing method that can be done by the owner and that samples can be stored in room temperature which facilitates the sampling.

The significant association of presence of D_5 with cleanness of colour is interesting and genotyping of this mutation could be useful for breeding where the colour phenotypes are important. The number of sheep in this study with the phenotype clean is however low and confirmation of the results from a larger number of sheep would be necessary before using the genotype data in breeding. In this study, data on relatedness between the rams is not available, but the rams with phenotype 'clean' and genotype D_5N_5 comes from four different herds and is thus not closely related. The lower peaks from the haplotype harbouring the D_5 mutation in combination with the indications of a duplication of *ASIP* in all Gotland sheep is interpreted as a presence of the D_5 allele in a single copy and the N_5 allele on multiple copies of *ASIP* in the sheep heterozygous D_5N_5. Quantification of the different alleles of *ASIP* in Gotland sheep would be valuable in future research projects and can hopefully give further insights into the background of different colour traits. Also, investigation of other colour genes would be interesting.

Figure 1. A screenshot from the software 4peaks showing the region with the D_5 allele in *ASIP* in one of the Gotland sheep with the D_5N_5 genotype where it was clearly visible. The highlighted region is the region that is deleted in in D_5 haplotype (TTCCT here in reverse, AGGAA in forward). The deletion is detected by the occurrence of double peaks after the deletion. The lower peaks of the haplotype harbouring D_5 was observed in all Gotland sheep with the D_5 allele.

The category blue in the cleanness of colour phenotype is a desired trait and it would be of importance to investigate the genetic background to this trait, but in this data there was unfortunately only one ram with this phenotype. A larger number of samples with this phenotype would be necessary for further research.

References

Adalsteinsson S., Lauvergne J.J., Boyazoglu, J.G., Ryder M.L. (1978) Ann Genet Sel Anim 10: 329-342

Norris B.J. & Whan V.A. (2008) Genome Research 18, 1282–93. https://doi.org/10.1101/gr.072090.107

Näsholm A. (2005) Acta Agric Scand Sect A, 55:57-65. https://doi.org/10.1080/09064700500335996

Näsholm A. (2009) Fårskötsel 6:16–18 (in Swedish).

R Core Team (2021). R: A language and environment for statistical computing. R

Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/

Rochus C.M., Westberg Sunesson K., Jonas E., Mikko S., Johansson A.M. (2019) Anim Genet 50(6):712-717. https://doi.org/10.1111/age.12837

713. Breeding initiatives around reduction of methane emissions in meat sheep – an international review

N.R. Lambe

Scotland's Rural College, Hill and Mountain Research Centre, Crianlarich, West Perthshire, FK20 8RU, Scotland, United Kingdom; nicola.lambe@sruc.ac.uk

Abstract

This review considers breeding initiatives around reduction of methane emissions in meat sheep. It discusses genetic strategies that are being tested, considers research results to date, and summarises the current state of implementation across different countries for the potential to reduce methane emissions from sheep production systems through breeding.

Introduction

Globally, ~1.2 billion sheep produce 6.87 million tonnes of methane (i.e. 186.9 million tonnes of CO_2 equivalent) per annum due to enteric fermentation (FAOSTAT, 2019). A long-term strategy for cumulative reductions in greenhouse gas (GHG) emissions is to breed more efficient animals with lower methane emissions, which fits the Organisation for Economic Co-operation and Development's approach for mitigating GHG emissions (Macleod *et al.*, 2015), as selective breeding is permanent, cumulative and highly cost-effective.

Various genetic solutions have been proposed to mitigate methane emissions from sheep production systems. As more productive animals have lower environmental impact per unit of product (Macleod *et al.*, 2015), wider use of existing breeding goals, designed to reduce waste and increase meat yield per animal, may go a long way towards reducing GHG emissions. Another long-term genetic strategy is to breed more feed-efficient animals. Residual feed intake (RFI) gives a measure of feed efficiency independent of growth and body size, and selection for RFI should produce more biologically and economically efficient animals that consume less feed for the same output, reducing methane production per day or per unit of dry matter intake. Alternatively, livestock can be bred directly for reduced methane, by measuring or predicting emissions under relevant conditions (Robinson *et al.*, 2020). As methane measurement technology is expensive relative to the value of sheep, proven, low-cost predictors are needed, benchmarked against methane measurements. Proxy measures related to rumen morphology and function have been investigated. Understanding their relationships with methane emissions, and their genetic control may be key to identifying more practical ways to accurately predict individual methane emissions.

The aim of this paper is to review breeding initiatives around reduction of methane emissions in meat sheep – genetic strategies being tested, research results to date, and the current state of implementation across different countries to reduce methane through breeding.

Materials & methods

Several studies have modelled the effects of current genetic selection goals on GHG emissions over time and the impact of different traits on GHG emissions (e.g. Nakielny and Roden, 2011; Wall *et al.*, 2010). The impacts of changing index weightings and/or including feed efficiency and methane measures in indices, with different economic values, have been modelled (Cottle and Conington, 2012a,b; Wall *et al.*, 2010). Other studies have used biological or life cycle analysis models to predict methane emissions based on data from sheep of high or low genetic merit for a production index (Bohan, 2021; Lambe *et al.*, 2014).

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_713, © N.R. Lambe 2022

Internationally, sheep breeding research is at different stages in developing resource populations to generate data on feed intake and methane emissions. Although feed intake measurement has been more widely studied in cattle, various methods and proxy measures are now under investigation in sheep. Animal-mounted sensors and biological markers for predicting feed intake at pasture have been tested by several research groups. However, for large scale recording of feed intake for breeding programmes, automated feed intake recording equipment, suitable for recording individual animal intakes from groups of sheep in indoor or outdoor pens, is closest to wide-scale implementation.

Methods for direct measurement of methane include respiration chambers (considered to be the gold standard) and shorter-term techniques, such as portable accumulation chambers (PAC), extraction hoods (e.g. GreenFeed), sniffers, or laser detectors (Jonker *et al.*, 2020). The sulphur hexafluoride (SF6) method allows methane measurements from grazing animals (Johnson *et al.*, 1994), but is unsuitable for large-scale use. Proxies relating to rumen function or methanogenesis pathways include: rumen morphology measured by computed tomography (CT) or post-mortem (Goopy *et al.*, 2014); characterisation or genotyping of rumen microbial communities (Rowe *et al.*, 2019); and analysis of rumen metabolites (Jonker *et al.*, 2019).

Results

Current UK breeding goals are predicted (Wall *et al.*, 2010) to reduce GHG emissions per breeding female or kg product for upland and lowland sheep systems, but to *increase* GHG emissions for hill sheep, largely due to an emphasis on maternal traits. Using selection indices for Welsh sheep, Nakielny and Roden (2011) predicted small annual methane reductions for hill and lowland sheep, assuming no ewe weight change. Concurring research in Ireland found that high genetic merit (5-star) ewes produced less methane compared to low genetic merit (1-star) ewes (Bohan, 2021). Changing index weightings (Lambe *et al.*, 2014; Wall *et al.*, 2010), or including methane emissions and/or feed intake as breeding objectives and selection criteria could further reduce emissions (Cottle and Conington, 2012a,b).

Understanding of the genetic control of feed efficiency and methane emission traits in sheep is increasing rapidly, laying the foundations for breeding strategies to reduce methane emissions. Low to moderate heritabilities are estimated for feed intake, RFI, feed efficiency and feeding behaviour traits, across sheep studies where environmental and experimental conditions vary (Pulina *et al.*, 2013). High genetic correlations imply that selection to decrease feed intake or RFI will decrease methane and carbon dioxide (Paganoni *et al.*, 2017; Robinson *et al.*, 2020). Additionally, selecting lambs post-weaning to decrease feed intake and methane would also decrease these traits in adult sheep (Paganoni *et al.*, 2017) and promising repeatabilities have been found for methane production across ages and physiological states (Oddy *et al.*, 2019). An earlier review suggested that methane emissions measured in respiration chambers are moderately repeatable across diets and feeding levels, and that shorter-term methane measurement techniques show promise for use in genetic evaluations under optimal protocols (Pickering *et al.* 2015; Jonker *et al.* 2020). Heritabilities for methane production are low to moderate (h^2 0.11-0.29) when measured by respiration chambers or PAC (Jonker *et al.*, 2018; Paganoni *et al.*, 2017; Robinson *et al.*, 2020). Genetic parameters have also been estimated for methane proxy traits. In Texel sheep, rumen volume measured during routine CT scanning was moderately heritable (h^2 0.45; Lambe *et al.*, 2022), although genetic correlations with methane emissions are not yet available. Rumen microbial profiles are heritable and correlated with methane emissions and feed intake (Rowe *et al.*, 2019). Volatile fatty acid concentrations are also moderately heritable and repeatable, with moderate genetic correlations to methane yield, but depend on the diet (Jonker *et al.*, 2019).

Discussion

Selection indices designed to reduce methane emissions. Although studies suggest potential to reduce methane emissions further by changing current index weightings, or including breeding goals of feed intake and/or methane emissions, such indices are not yet being widely used in commercial sheep breeding. However, with increasing research knowledge, alongside potential financial incentives to reduce methane emissions expected in many countries, this may be closer to realisation. Implementation of carbon credits (or penalties), and their value, could influence selection pressure to reduce methane emissions (Black *et al.*, 2021).

Selection for improved feed efficiency. Advances in electronic animal identification and recording of feed consumption by computer-controlled machinery present opportunities to measure individual feed intake during performance trials. Best methodologies to measure feed efficiency in sheep flocks in different systems are under investigation and selection for feed efficiency has proven successful in research flocks (e.g. Tortereau *et al.*, 2020), resulting in breeding stock selected for RFI becoming available in some countries and breeds.

Selection for reduced methane emissions. Viable proxy traits for methane prediction are emerging. Rumen sampling to predict methane emissions from the microbiome, when coupled with high-throughput, rumen profiling (Hess *et al.*, 2020), could provide a quick and accurate method to determine genetic merit in breeding programmes. If CT-measured rumen volumes prove to accurately predict methane at the genetic level, then mechanisms are in place in some countries (e.g. UK, NZ) to incorporate them into breeding programmes with other CT traits. Direct PAC measurements of methane are a promising route to large-scale measurement of emissions across sheep production systems and are in use within research flocks (e.g. Australia, NZ, Uruguay, Ireland, Norway), often with strong genetic links to national breeding programmes. In some cases (e.g. NZ, Ireland, Norway), PACs are measuring methane from industry flocks participating in these breeding programmes. The next steps are to develop environmental selection indices, including methane emissions.

Work underway to address knowledge gaps. Numerous national and international projects have been funded to progress understanding and implementation of strategies to reduce GHG emissions from livestock, including sheep. The Global Research Alliance on Agricultural Greenhouse Gases (GRA) oversees projects to improve international research in this area, including the production of good practice guidelines for more consistent measurement methodology (Jonker *et al.*, 2020). Another pertinent example is the 'GrassToGas' project (Strategies to mitigate GHG emissions from pasture-based sheep systems; ERA-NET), with seven partner countries combining expertise to address knowledge gaps, considering various technologies, protocols and approaches to breed sheep for reduced GHG emissions.

Future potential. Although some lower-cost, shorter-term methods show potential for widespread estimation of methane emissions from individual sheep, most methods still require expensive equipment, or laboratory processes, and intensive animal handling. The most feasible breeding strategy is likely to be through genomic selection (Black *et al.*, 2021). Datasets of genotypes and phenotypes are developing in several countries, with some already of sufficient size to estimate genomic breeding values for methane emissions. International collaboration, to standardise protocols and methodology and combine data and resources, can fast-track genetic and genomic strategies to mitigate methane emissions from sheep production systems. A trans-national approach is necessary to address this global issue.

References

Black J.L., Davison T.M., and Box I. (2021) Animals. 11:951. https://doi.org/10.3390/ani11040951

Bohan A. (2021) Five star animals producing more profit and less emissions Irish Farmers' Journal. Available at: https://www.farmersjournal.ie/five-star-animals-producing-more-profit-and-less-emissions-593070

Cottle D.J. and Conington J. (2012a) J.Agric.Sci. 150:570-583. https://doi.org/10.1017/S0021859611000967

Cottle D.J. and Conington J. (2012b) J.Agric.Sci. 151:872-888. https://doi.org/10.1017/S0021859612001037

FAOSTAT (2019). Available at: https://www.fao.org/faostat/en/#data/

Goopy J.P., Donaldson A.J., Hegarty R.S., Vercoe P.E., Haynes F. *et al.* (2014) Br.J.Nutr. 111:578-585. https://doi.org/10.1017/S0007114513002936

Hess M.K., Rowe S.J., Van Stijn T.C., Henry H.M., Hickey S.M. *et al.* (2020) PLoS ONE. 15:e0219882. https://doi.org/10.1371/journal.pone.0219882

Johnson K., Huyler M., Westberg H., Lamb B., and Zimmerman P. (1994) Environ.Sci.Technol. 28:359-362. https://doi.org/10.1021/es00051a025

Jonker, A., S. Hickey, J. McEwan and G. Waghorn. (2020) Guidelines for estimating methane emissions from individual ruminants. Ministry for Primary Industries, Wellington, NZ, pp. 55-62. https://www.mpi.govt.nz/resources-and-forms/publications/

Jonker A., Hickey S., Rowe S., Janssen P., Shackell G. *et al.* (2018) J.Anim.Sci. 96:3031-3042. https://doi.org/10.1093/jas/sky187

Jonker A., Hickey S.M., McEwan J.C., Rowe S.J., Janssen P.H. *et al.* (2019) J.Anim.Sci. 97:2711-2724. https://doi.org/10.1093/jas/skz162

Lambe N.R., McLaren A., McLean K.A., Gordon J., and Conington J. (2022) Anim.Prod.Sci. in press. https://doi.org/10.1071/AN21423

Lambe N.R., Wall E., Ludemann C.I., Bunger L., and Conington J. (2014) Small Rumin.Res. 120:27-34. https://doi.org/10.1016/j.smallrumres.2014.04.011

Macleod, M., V. Eory, G. Gruere and J. Lankoski. (2015) Cost-effectiveness of greenhouse gas mitigation measures for agriculture: A literature review. OECD Publishing, Paris

Nakielny C. and Roden J.A. (2011) Modelling the Effect of Genetic Improvement Programmes on Greenhouse Gas Emissions in the Welsh Sheep Industry. Available at: https://meatpromotion.wales/images/resources/HCC_Methane_A5_low_res.pdf

Oddy V.H., Donaldson A.J., Cameron M., Bond J., Dominik S. *et al.* (2019) Anim.Prod.Sci. 59:441-448. https://doi.org/10.1071/AN17447

Paganoni B., Rose G., Macleay C., Jones C., Brown D.J. *et al.* (2017) J.Anim.Sci. 95:3839-3850. https://doi.org/10.2527/jas.2017.1499

Pickering N.K., Oddy V.H., Basarab J., Cammack K., Hayes B. *et al.* (2015) Animal. 9:1431-1440. https://doi.org/10.1017/S1751731115000968

Pulina G., Avondo M., Molle G., Francesconi A.H.D., Atzori A.S. *et al.* (2013) Rev.Bras.de Zootec. 42:675-690. https://doi.org/10.1590/S1516-35982013000900010

Robinson D.L., Dominik S., Donaldson A.J., and Oddy V.H. (2020) Anim.Prod.Sci. 60:880-892. https://doi.org/10.1071/AN18383

Rowe S.J., Hess M., Zetouni L., Hickey S., Brauning R. *et al.* (2019) Proceedings. 36:177. https://doi.org/10.3390/proceedings2019036177

Tortereau F., Marie-Etancelin C., Weisbecker J.L., Marcon D., Bouvier F. *et al.* (2020) Animal. 14:681-687. https://doi.org/10.1017/S1751731119002544

Wall E., Ludemann C.I., Jones H.E., Audsley E., Moran D. *et al.* (2010) The potential for reducing GHG emissions for sheep and cattle in the UK using genetic selection. Available at: http://randd.defra.gov.uk/Document.aspx?Document=FinalReportIF0182.doc

714. Measuring methane in sheep production systems – phenotypic factors affecting output

F.M. McGovern[1*], E. O'Connor[1], L. Farrell[1], E. Dunne[1], E. Wall[3] and N. McHugh[2]

[1]Teagasc, Animal and Grassland Research and Innovation Centre, Athenry, H65 R718, Co. Galway, Ireland; [2]Teagasc, Animal and Grassland Research and Innovation Centre, Mooreaprk, Fermoy, P61 C996, Co. Cork, Ireland; [3]Sheep Ireland, Highfield House, Shinagh, Bandon, P72 X050, Co. Cork, Ireland; fiona.mcgovern@teagasc.ie

Abstract

Quantifying methane (CH_4) output from ruminant animal production systems is crucial in the development of mitigation strategies. This study aimed to develop an understanding of CH_4 output across life stage and the effects of key performance indicators on CH_4 output in a multi breed sheep population. A total of 7,027 CH_4 records from 2,404 animals collected from four flocks between 2019 and 2021 were available. Factors associated with CH_4 output were determined using linear mixed models. Highest CH_4 output was associated with ewes, with lambs being the lowest (P<0.001). When measuring CH_4 using portable accumulation chambers the atmospheric pressure, relative humidity and time since feed removal must all be recorded (P<0.001). Singleton reared litters had lower CH_4 output than multiple rearing litters when measured in both the ewe and the lamb (P<0.001). Results highlight the impact of multiple environmental and animal factors on CH_4 output in sheep systems.

Introduction

There is intensifying pressure on the global agricultural community to reduce greenhouse gas emissions from animal agriculture. Worldwide agricultural methane (CH_4) emissions account for over 38% of total CH_4 emissions (FAO, 2020). As it is one of the primary outputs of digestion in ruminant animals, there is a major focus on identifying more efficient animals with lower lifetime levels of CH_4 output per unit of production. Quantifying CH_4 output of individual ruminant animals is labour and resource intensive, particularly when using the gold standard CH_4 output measurement method, respiration chambers. Subsequently, alternative apparatus have been developed allowing for data of CH_4 output to be collected for large cohorts of animals in a resource efficient manner. Scientific evidence shows that there is a possibility to breed animals for lower methane emissions (Rowe *et al.*, 2019). However in order to incorporate this trait into a national breeding programme investigation must firstly be carried out into the variation surrounding the trait and the factors affecting it. Understanding the impact of phenotypic factors associated with correlated traits and primary traits of interest enables appropriate adjustments within genetic evaluation models. While striving for lower methane emitting animals is merited, the importance of maintaining production levels and overall farm gate efficiency and profitability cannot be forgotten. Therefore, the objective of the present study was to determine the phenotypic factors affecting methane output in sheep systems.

Materials & methods

Phenotypic data collection. Methane measurements were obtained using portable accumulation chambers (PAC) as described by Jonker *et al.* (2018). Briefly, animals were removed from feed a minimum of 1 hour prior to entering the PAC for a 50 minute period. Methane, oxygen (O_2) and carbon dioxide (CO_2) levels were recorded at three separate intervals; 0 minute (animal entry), 25 minute and 50 minute post entry to the PAC. A total of 7,027 methane measurements were collected from 2,404 animals across four flocks between the years 2019 and 2021, inclusive. Animals were classified as lambs (<12 months; n=2,151 records), nulliparous hoggets (12 to 24 months; n=763 records) and ewes (primiparous or greater; 24 to 174 months; n=3,892 records) which were further classified as lactating or dry (non-pregnant and non-

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_714, © F.M. McGovern *et al.* 2022

lactating). The exact time of each measurement, the time since feed was removed, ambient temperature, atmospheric pressure and relative humidity were recorded for each PAC measurement run. Live-weight was recorded on all animals on the day of measurement. Data was also available on date of birth, age, breed composition, stage of production (referred to hereafter as life-stage) and diet type for all animals. A contemporary group of the date of measurement, flock and PAC run the animal was measured in was fitted to the data. For all ewes the lambing date, parity, age at first lambing, number of lambs born and number of lambs reared, combined litter weight at birth, 40d, weaning and ewe body condition score (BCS) were also available. For all lambs data on the sex, birth, pre-weaning and weaning weight, litter size at birth and rearing litter size and daily live-weight gain between each time point were available. Ewe parity was categorised as 1, 2, 3, 4, or ≥5. Age of the dam was categorised as lambing either: (1) between 8 and 18 months of age; or (2) between ≥18 and 28 months of age. Number of lambs born (litter size at birth) was the total number of lambs (dead and alive) per lambing event; only litter sizes between one and four were retained for final analysis. Number of lambs reared (rearing litter size) was defined as the number of lambs reared in the litter (including the lamb itself); only rearing sizes between one and three were retained. Following all edits there were a total of 6,771 records from 2,392 animals, including 2,140 individual lamb records and 3,868 individual ewe records.

Statistical data analysis. All PAC and live-weight data were checked for normality using PROC UNIVARIATE (SAS Inst. Inc., Cary, NC, USA); records with data outside three standard deviations from the mean were removed for the final analysis. Factors associated with animal production performance and methane output were estimated separately using linear mixed models in PROC HPMIXED; contemporary group (date-flock-PAC run) was included as the random effect in all models. Potential fixed effects considered in the models included: life-stage, time of measurement, time off feed prior to measurement, ambient temperature, atmospheric pressure, relative humidity, age and live-weight on the day of measurement and breed. For ewe analysis parity, age at first lambing, number of lambs born, number of lambs reared, BCS and combined litter weight were included while for lambs sex, birth, pre-weaning and weaning weight, birth and rearing litter size were included. For all dependent variables and all analysis, a multiple regression model was also built using stepwise forward – backward regression; the significance threshold for entry and exit of variables into/from the model was set at 5%.

Results

The average live-weight of lambs, hoggets and ewes was 43.7, 66.1 and 74.5 kg, respectively, with lactating ewes weighing on average 3 kg heavier than non-lactating ewes. Methane output increased with stage of production whereby lambs ranked the lowest for CH_4 output, 12.62 g CH_4/day while lactating ewes ranked highest, 27.12 g CH_4/day (P<0.001). The phenotypic standard deviation in CH_4 output of lambs, hoggets and ewes was 6.14, 6.26 and 7.48 g, representing 49, 45 and 35% of the mean methane output, respectively. Factors associated with CH_4 output included animal life stage, diet type, breed, time from feed removal, atmospheric pressure, relative humidity, PAC chamber number, age and live-weight (P<0.001). Overall animal life-stage and diet type accounted for 35% of the variation in methane output. Regardless of life-stage the time from feed removal affected methane output as shown in Figure 1.

The main breed proportion of the animal affected CH_4 output, with Charollais and Texel animals having higher CH_4 output in comparison to Belclare and Suffolk animals (P<0.01). Across the full dataset for every 1 kg increase in live-weight CH_4 output increased by 0.12 g/day, however for lambs a 1 kg increase in live-weight resulted in a CH_4 output 0.26 g/day higher (P<0.001). In addition, ewe CH_4 output increased up to a BCS of 3 after which it declined (P<0.001). Animals consuming perennial ryegrass or a combination of perennial ryegrass plus chicory ranked higher for CH_4 output when compared to those grazing a combination of perennial ryegrass plus clover (red or white; P<0. 001). Animals measured indoors on a

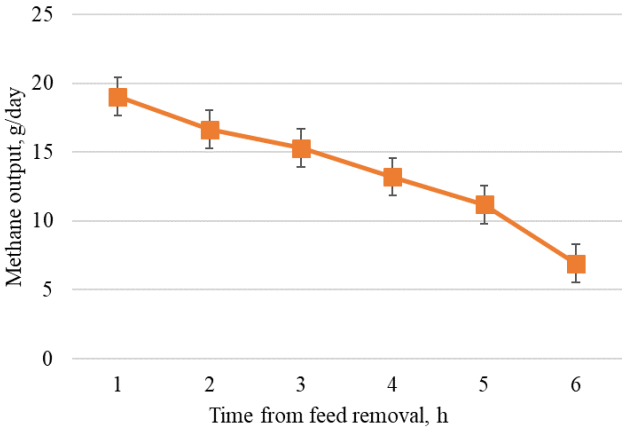


Figure 1. The effect of time from feed removal on methane output as measure in the PAC.

diet of perennial ryegrass silage ranked lower for CH_4 output than animals grazing outdoors ($P<0.001$). For both lambs and ewes rearing litter size affected CH_4 output. When ewes were rearing multiple lambs (>1 lamb per litter) CH_4 output was on average 6.7% higher and similarly when lambs were reared as multiples their CH_4 output was 8.6% higher than those reared as singletons. Parity 1 ewes ranked higher for CH_4 output which declined as parity increased, whereby parity 4 and 5 ewes ranked lowest ($P<0.001$).

Discussion

One of the main challenges for ruminant scientists is developing mitigation strategies for methane output without comprising animal production. Understanding the effect of various phenotypic and environmental factors on methane output in sheep production systems is therefore crucial to facilitating the development of mitigation strategies and comprehending relationships with measures of animal efficiency. Results from this study signify that a plethora of key performance indicators in conjunction with environmental conditions impact methane output when measured in small ruminants using PAC. Methane is produced in the rumen whereby methanogenic bacteria convert excess hydrogen to methane as a metabolic end product of digestion. Multiple factors have previously been shown to impact methane production including the type of carbohydrate in the diet, feed intake and feed passage rate (McAllister *et al.*, 1996). Recording the time of feeding relative to when the methane output of animals was measured is important as the rate of digestion declines over time as seen in the results of this study. However, further work from our research group has shown the ranking of animals for methane output remains unaffected by the length of time since feeding (O'Connor *et al.*, 2021). There was an effect of diet type whereby animals consuming feeds with higher organic matter digestibility had lower methane output. Due to the large population of animals measured in the current study, it was impractical to measure individual feed intake. Previous research has shown that dry matter intake increases in line with animal age and stage of production. Therefore, the increase in methane output across each of the life-stages investigated was expected and aligns with Jonker *et al.* (2018) who reported methane output from lambs and ewes measured using the PAC. As BCS increased CH_4 output declined, while this merits further biological investigation, it aligns with results previously reported in sheep (Reintke *et al.*, 2020) and dairy (Zetouni *et al.*, 2018) whereby negative genetic correlations were identified between BCS and CH_4 production. The relationships between CH_4 and other, more widely recorded, production traits outlined in this study will deepen the understanding of their interaction which is required prior to inclusion in genetic evaluations.

Acknowledgements

Funding from the Irish Department of Agriculture, Food and Marine Research Fund GREENBREED (17/S/235) and Grass2Gas (2019/EN/202) is gratefully acknowledged.

References

Food and Agriculture Organization of the United Nations (FAO). 2020. Emissions shares. Available at: http://www.fao.org/faostat/en/#data/EM/visualize

Jonker, A., Hickey, S.M., Rowe, S.J., Janssen, P.H., Shackell, G.H. *et al.* (2018) J. Anim. Sci., 96(8):3031-3042 https://doi.org/10.1093/jas/sky187

McAllister T.A., Cheng K.J., Okine E.K. and Mathison, G.W. (1996) Canadian J. Anim. Sci., 76(2):231-243 https://doi.org/10.4141/cjas96-035

O' Connor E., McHugh N., Boland T.M., Dunne E., McGovern F.M. (2021) J. Anim. Sci. 99(8). https://doi.org/10.1093/jas/skab132

Ramin, M. and Huhtanen, P. (2013) J. Dairy Sci, 96(4): 2476-2493 https://doi.org/10.3168/jds.2012-6095

Rowe, S.J., Hickey, S.M., Jonker, A., Hess, M.K., Janssen *et al.* (2019). Proc. of the 23rd Conference of the Association for the Advancement of Animal Breeding and Genetics (AAABG), New South Wales, Australia 23:306-309

Reintke, J., Brügemann, K., Yin, T., Engel, P., Wagner, H. *et al.* (2020) Archives anim. Breeding, 63(1):113-123 https://doi.org/10.5194/aab-63-113-2020

Zetouni, L., Kargo, M., Norberg, E. and Lassen, J. (2018) J. Dairy Sci., 101(3):2273-2280 https://doi.org/10.3168/jds.2017-13402

715. Genetics of methane emissions in a multi-breed sheep population

N. McHugh[1*], E. O'Connor[2], D.P. Berry[1], E. Dunne[2], L. Farrell[2], P. Rafter[1], K. McDermott[3] and F.M. McGovern[2]

[1]*Teagasc, Animal and Grassland Research and Innovation Centre, Fermoy, P61 P302, Co. Cork, Ireland;*
[2]*Teagasc, Animal and Grassland Research and Innovation Centre, Athenry, H65 R718, Co. Galway, Ireland;*
[3]*Sheep Ireland, Highfield House, Shinagh, Bandon, P72 X050, Co. Cork, Ireland;* noirin.mchugh@teagasc.ie

Abstract

Mitigation strategies are required to reduce methane emissions without impacting animal productivity; genetic improvement is one potential strategy. The objective of this study was to estimate genetic variance components for methane emissions in an Irish sheep population using a spot-sampling technique. Spot-sampling methane measures using portable accumulation chambers were available from 1,213 animals in 4 flocks. Variance components were estimated using an animal linear mixed model. Mean methane (CH_4) emissions of the ewes was 19.29 (SD=7.50) g CH_4/day, with a coefficient of genetic variation of 9.67%. The direct heritability and animal repeatability were 0.25±0.12 and 0.39±0.05, respectively. The direct heritability estimate for methane emissions expressed in daily g CH_4/LW was 0.13±0.07. The mean estimated breeding value of methane emission per breed varied from -0.13 (Belclare breed) to 0.14 (Suffolk breed) g CH_4/day.

Introduction

Global agriculture sector accounts for an estimated 38.62% of the potent greenhouse gas methane (FAO, 2020), with mounting international pressure to identify mitigation strategies. Reducing ruminant livestock populations is one drastic option which could impact food security. Nutritional strategies using complementary feeds have also been proposed. Breeding programs, by comparison, have a long, well-proven history in tackling grand challenges. However, successful breeding programs are conditional on the availability of phenotypes from which to undertake genetic evaluations. The development of spot-sampling, portable, high-throughput measurement techniques for sheep (Jonker *et al.*, 2018) has enabled the collection of vast quantities of methane phenotypes. Previous studies have clearly demonstrated genetic variability in methane emissions from sheep (Robinson *et al.*, 2014; Jonker *et al.*, 2018), but estimates of the genetic variation in methane emissions from the Irish national sheep population at pasture are lacking. The objective of this study was to estimate genetic variance components for spot-sampling measures of methane emissions in a multi-breed population. Results from this study will be useful in determining the suitability of methane for inclusion in sheep breeding programs.

Materials & methods

Data. A total of 4,007 spot measure methane records from 2,075 animals collected from 4 flocks between the years 2019 to 2021, inclusive were available. Methane records were collected using the portable accumulation chambers (PAC) as described by Jonker *et al.* (2018). Briefly animals were removed from feed one hour prior to measurement and the live weight of each animal was recorded; animals were then placed in 1 of 12 chambers and methane was measured at the point of entry and immediately prior to removal from the chamber (approximately 50 minutes after entry). A range of PAC-specific measurements such as duration of each individual animal measurement, the time since last available feed, ambient temperature, atmospheric pressure and relative humidity were available. Only records from females who had at least one recorded lambing event were considered in the present study; furthermore, only data from ewes weighing

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_715, © N. McHugh *et al.* 2022

between 45 and 120 kg were retained. Methane or carbon dioxide (CO_2) records that were ±3 SD from the mean were removed (n=45 records removed).

Data were also available on date of lambing, ewe parity, breed composition, and the ancestry of each ewe. Ewe parity was categorised as 1, 2, 3, 4, or ≥5. Age of the dam at first lambing was categorised as lambing either: (1) between 8 and 18 month of age; or (2) between ≥18 and 28 month of age. Birth type (single to quadruplets) and rearing size (single to triplets) of each ewe for the same production year as methane measurement was also available. Only ewes with a known sire and dam were retained for analysis. Each ewe was allocated to a contemporary group of flock-date of measurement-measurement lot (i.e. group of 12 animals measured at the same time). Following all edits, 2,114 methane measurements on 1,213 individual ewes remained. Methane emissions were expressed both on a total daily emissions (i.e. g CH_4/day) but also relative to live weight (i.e. daily g CH_4/LW).

Data analysis. Genetic and residual components were estimated in ASReml (Gilmour *et al.*, 2015). Fixed effects included in the model were contemporary group (flock-date-run), ewe parity, ewe age at first lambing, the breed proportion of the most common breeds (Belclare, Charollais, Suffolk and Texel), and the heterosis and recombination loss coefficient of the ewe. Approximately 15% of the methane data was recorded on lactating ewes, therefore days since lambing and number of lambs born and reared by the ewe within the same production year were also included as fixed effects. A direct ewe additive genetic effect and a ewe permanent environmental effect were included as random effects; 42% of ewes had >1 record. With genotypes available on 54% of ewes, relationships among animals was accounted for using a H matrix which combines the genomic and pedigree-relationship matrix.

Due to large temporal variation in grass quality between paddocks or farms and time off feed (Jonker *et al.*, 2018), standardisation of methane emissions was undertaken to account for the heterogeneity of residual variances across time. Before the estimation of variance components for methane emissions, residuals from a fixed effects model with methane emissions as the dependent variable and the aforementioned fixed effects included in the model (excluding contemporary groups) were standardised to the mean residual SD within the contemporary group of flock-date-run. Standardised residuals were added back to the respective fixed effects solutions for each animal to generate a standardised methane emission trait. Variance components of the standardised methane emissions are reported in the present study.

Results

The mean methane emissions of the ewes was 19.29 (SD=7.50) g CH_4/day and varied from 18.61 g CH_4/day for ewes during the dry period to 26.75 g CH_4/day for lactating ewes. When expressed on a per live weight basis, the mean methane emissions was 0.26 (SD=0.10) daily g CH_4/LW. The genetic SD for the two methane emissions traits were 1.87 g CH_4/day and 0.02 daily g CH_4/LW. The coefficient of genetic variation for methane emissions expressed as total daily emissions was 9.67%. The direct heritability estimate for methane emissions ranged from 0.13±0.07 (expressed in daily g CH_4/LW) to 0.25±0.12 (expressed in g CH_4/day). The animal repeatability was 0.39±0.05 (expressed in g CH_4/day) and 0.49±0.03 (expressed in daily g CH_4/LW).

The mean carbon dioxide produced per ewe was 1,255 (SD=484) g CO_2/day. The genetic SD and coefficient of variation was 142 g CO_2/day and 11.29%, respectively. The heritability for carbon dioxide was 0.14±0.11 while the animal repeatability was 0.60±0.04. The phenotypic and genetic correlation between the methane emissions and carbon dioxide (both expressed as g/day) was 0.24±0.05 and 0.41±0.10, respectively.

The distribution of methane emission (g CH_4/day) estimated breeding values (EBVs) for animals of the four main breeds is in Figure 1. The mean methane emissions EBV for the Belclare (n=399), Charollais (n=77), Texel (n=723), Suffolk (n=723) breeds was -0.13, 0.18, -0.03 and 0.14 CH_4/day, respectively.

Discussion

Animal genetic improvement can play a pivotal role in the mitigation strategy to reduce greenhouse gases from the ruminant livestock sector. Parameters estimated in this study will be useful in firstly generating genetic evaluations (i.e. the lambda in the mixed model equations), secondly determining the optimal phenotypic strategy for a breeding program (i.e. number of phenotypes per family versus number of families phenotyped) and thirdly populating models to determine the possible rate of genetic gain and mitigation potential (i.e. selection index theory).

The moderate heritabilities estimated for methane emissions in sheep in the present study (0.13 to 0.29) corroborate previous estimates from the southern hemisphere using both PAC and respiration chambers (Robinson *et al.*, 2014; Jonker *et al.*, 2018). Assuming a heritability of 0.13, 29 phenotyped progeny per sire would be required to achieve an accuracy of selection of 0.70 using traditional pedigree-based genetic evaluations. The high animal repeatability (0.39) for ewes in the present study, coupled with the strong correlation (0.85 to 0.99) calculated by Jonker *et al.* (2018) between methane emissions in lambs and ewes suggests that methane remains genetically similar across an animal's lifetime.

Dry matter intake (DMI) was not available in the current study but has been shown to explain 76% of the variation in methane production between animals (Swainson *et al.*, 2018). In cattle a strong correlation has been estimated between methane yield (i.e. CH_4 per unit of DMI) and methane emissions expressed as CH_4 per unit of live weight, indicating the suitability of methane emissions expressed as daily g CH_4/LW as a proxy for methane yield where DMI data are not available, especially in growing animals. Nonetheless, the impact of selecting animals for lower methane emissions on DMI warrants further investigation in an Irish grass based context, to ensure that indirect selection for lower DMI does not negatively impact individual

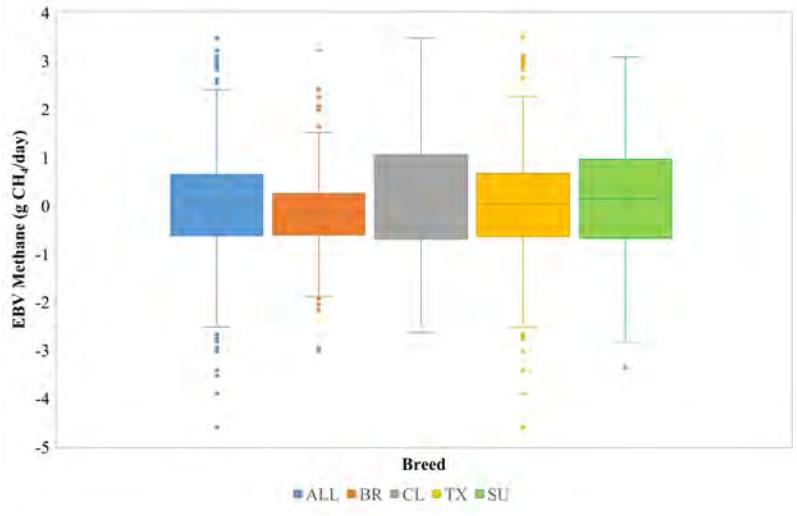


Figure 1. Box and whisker plot of estimated breeding values (EBVs) for methane emissions (g CH_4/day) for all animals (ALL), and the Belclare (BR), Charollais (CL), Texel (TX) and Suffolk (SU) breeds.

animal performance, such as growth rate, which could actually contribute to greater total flock emissions per unit of output.

Data used in the present study were generated on commercial crossbred ewes, but the future phenotyping strategies should also consider the phenotyping of seedstock young rams for methane emissions thereby ensuring that breeding values of high accuracy are generated on the parents of the next generation. Using a repeatability of 0.39, the accuracy of selection for an individual ram with 1, 2, or 3 measures is 0.36, 0.43 and 0.47, respectively.

With the advent of spot-sampling techniques such as PACs, methane emissions can now be routinely generated on large numbers of sheep. This coupled with the existence of ample genetic variation for methane emissions, facilitates their inclusion in Irish sheep breeding objectives to aid in the reduction of the carbon footprint for the Irish sheep industry.

Acknowledgements

Funding from the Irish Department of Agriculture, Food and Marine GreenBreed (17/S/2135) and GrasstoGas (2019EN202) is gratefully acknowledged.

References

Donoghue K.A., Herd R.M., Bird S.H., Arthur P.F., and Hegarty R.F. (2013). Proc. of Assoc. Advmt. Anim. Breed. Genet. Napier, New Zealand.

Food and Agriculture Organization of the United Nations (FAO), Emissions shares. Available at: https://www.fao.org/faostat/en/#data/EM.

Gilmour A.R., Gogel B.J., Cullis B.R., and Thompson R. (2015). ASReml User Guide Release 4.1 Functional Specifications. VSN International Ltd, Hemel Hempstead, HP1 1ES, UK. Available at: https://www.vsni.co.uk.

Jonker A., Hickey S.M., Rowe S.J., Janssen P.H., Shackell G.H. *et al.* (2018) J. Anim. Sci. 96(8):3031-3042. https://doi.org/10.1093/jas/sky187.

Robinson D.L., Goopy J.P., Hegarty R.S., Oddy V.H., Thompson A.N. *et al.* (2014) J. Anim. Sci. 92(10):4349-4363. https://doi.org/10.2527/jas.2014-8042

Swainson N., Muetzel S., and Clark H. (2018). Anim. Prod. Sci. 58(6):973–979. https://doi.org/10.1071/AN15766

716. Modelling of productivity, profitability, and greenhouse gas emissions of Irish sheep flocks divergent in genetic merit

L. Farrell[1*], N. McHugh[2], T. Pabiou[3], A. Bohan[4], K. McDermott[3], L. Shalloo[2] and J. Herron[2]

[1]Animal & Grassland Research and Innovation Centre, Teagasc, Athenry, H65 R718, Co. Galway, Ireland; [2]Animal & Grassland Research and Innovation Centre, Teagasc, Moorepark, Fermoy, P61 C996, Co. Cork, Ireland; [3]Sheep Ireland, Highfield House, Shinagh, Bandon P72 X050, Co. Cork, Ireland; [4]Ballymote Local Advisory Office, Teagasc, Carrownanty, Ballymote, F56 A585 Co. Sligo, Ireland; lydia.farrell64@gmail.com

Abstract

Genetic selection using multi-trait breeding objectives can be used to improve flock productivity, profitability, and greenhouse gas (GHG) emissions intensities (kg CO_2-eq /kg of product), however validation is required. Phenotypic data from commercially farmed animals divergent in genetic merit for maternal traits were inputted into a bio-economic model to compare two contrasting flocks, ranked as either High (top 20%) or Low (bottom 20%) on the Irish replacement sheep breeding objective. The two flocks were then simulated using life cycle assessment (LCA) to estimate GHG emissions. The flock of High genetic merit ewes sold 3.29 kg more lamb carcass per ewe and €18/ewe higher net profit than the flock of Low genetic merit. Although total flock GHG emissions were higher for the High genetic merit flock, GHG emissions intensities were 22.1 and 23.7 kg CO_2-eq /kg lamb carcass sold for the High and Low genetic merit flocks, respectively.

Introduction

Improvements in flock productivity and profitability can be achieved through genetic selection using multi-trait breeding objectives. Ireland's sheep breeding objectives have not yet been validated for effects on flock productivity and profitability. Validation will confirm the sheep industry is breeding suitable animals and ensure farmer confidence in the breeding objectives. Previous studies have validated herd level improvements in production and profit for commercial Irish dairy farms with cows ranked higher on the Irish economic breeding index (Ramsbottom *et al.*, 2012). Previous modelling studies of pastoral sheep and dairy production systems have predicted reductions in GHG emissions intensity (GHG emissions / kg of product) of up to 10% through improvements in genetic merit without compromising profitability (Harrison *et al.*, 2014; Jones *et al.*, 2014). To date no study has yet modelled the potential economic and environmental benefits of improved sheep genetic merit using data from a national database of commercial farms. The objectives of this study, therefore, were to model production data differences from sheep divergent in genetic merit and to quantify their impact on farm profit and GHG emissions.

Materials & methods

Two flocks were modelled in the current study: a *High* and *Low* maternal genetic merit flock. The *High* genetic merit flock consisted of ewes ranked in the top 20% of the Irish Replacement breeding objective, and a proportion were bred with rams ranked in the top 20% of the Replacement breeding objective to produce replacement ewe lambs of *High* genetic merit. The remaining ewes were bred with terminal rams ranked in the top 20% of the Terminal breeding objective to produce lambs destined for slaughter. The *Low* genetic merit flock consisted of ewes ranked in the bottom 20% of the Replacement breeding objective, with a proportion of ewes bred with replacement rams ranked in the bottom 20% of the Replacement breeding objective and a proportion of ewes bred with rams ranked in the bottom 20% of the Terminal breeding objective. Both modelled flocks consisted of 168 ewes stocked at 7.91 ewes/ha, lambing annually in spring, grazing on pasture the majority of the year, and housed over winter until 48 hours post-lambing. Ewes were mated for the first time at 19 months of age to lamb at two years old and both flocks had a flock replacement

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_716, © L. Farrell *et al.* 2022

rate of 22%. Production values assumed for animals divergent in genetic merit in this study were derived from the Irish national sheep production database, using data from 4,200 animals born between 2018 and 2020 in commercial flocks in the Irish Central Progeny Test scheme (McHugh *et al.*, 2022). Traits investigated in the study included lambing (birth weights, lambing difficulty and survival), lamb growth rates (weight pre and at weaning, as well as, days to slaughter) and ewe traits (ewe litter size, barren rates and mature weight).

An established bio-economic model of an Irish lowland sheep flock, the Teagasc Lamb Production Model (Bohan *et al.*, 2016), was utilised to predict the production and profit of the *High* and *Low* genetic merit flocks. Key model inputs included animal production data, market prices and wool sales, variable and fixed costs, stocking rate, and fertiliser use. Key model outputs included stock inventories, total income and expenses, farm gross margin, and net profit expressed on a whole farm, per ha, per ewe, or per kg of carcass sold basis. Net profit was estimated as gross margin (total income less variable costs) less fixed costs and overdraft loan interest costs. Farm income was generated from sheep and wool sales, with average market prices from 2017 to 2020 utilised, with variable and fixed costs were informed by market values from the same period and from sources described in Bohan *et al.* (2016). Lambs were sold after reaching the target 20 kg carcass weight. After 1 October all lambs remaining on-farm were offered concentrate supplementation. A monthly feed budget was estimated, with feed surpluses conserved as silage and feed deficits requiring diet supplementation with silage and/or concentrates (Bohan *et al.*, 2016). Life cycle assessment (LCA) was conducted in accordance with the International Organisation for Standardization 14040 and 14044 standards to calculate greenhouse gas emissions generated by the two flocks. A cradle-to-farm gate system boundary was adopted, encompassing emissions up to the point in which the animals leaves the farm. Environmental impact embodied in capital goods (i.e. farm machinery and infrastructure) and medicines were not included in the LCA. For prediction of emission intensity, economic allocation was used to disaggregate the total environmental impact between co-products based on their economic value at farm level. Greenhouse gas emissions were scaled to per kg carcass weight and per hectare. A full description of the LCA model used can be found in O'Brien *et al.* (2016).

Results

Ewes in the *High* genetic merit flock had a higher scanning rate and their lambs had a higher survival rate, resulting in a higher weaning rate (1.70; lambs weaned per ewe joined) than the *Low* genetic merit flock (1.54; Table 1). With higher weaning and post-weaning weights, the *High* genetic merit flock sold more lambs before 1 October (from an all-pasture diet) with an average age of slaughter of 190 days, 13 days sooner than lambs in the *Low* genetic merit flock. The *High* genetic merit flock also sold 28 more lambs than the *Low* genetic merit flock, equating to 3.29 kg more lamb carcass per ewe. Lamb income was € 27,693 for the *High* genetic merit flock, € 3,083 higher than for the *Low* genetic merit flock. Total farm costs were similar for the two modelled flocks, and net profit was € 45/ewe for the *High* genetic merit flock, € 18/ewe higher than for the *Low* genetic merit flock. Total GHG emissions per hectare were 5,750 and 5,470 kg CO_2-eq for the High and Low genetic merit flocks, respectively. However, the GHG emissions intensity for the High genetic merit flock was 22.1 kg CO_2-eq/kg of lamb carcass sold, 6.7% lower than for the Low genetic merit flock.

Discussion

Through use of both replacement and terminal animals of *High* genetic merit, the *High* flock was effectively exploiting the potential production benefits of selecting for higher genetic merit on both of the Irish national breeding objectives. The current analysis has shown that the higher level of productivity at a trait level for animals of superior genetic merit validated by McHugh *et al.* (2022) resulted in higher production and profit at a flock level for the modelled *High* genetic merit flock compared to the *Low* genetic merit

Table 1. Production, profit, and greenhouse gas emissions of modelled flocks ranked *High* (top 20%) and *Low* (bottom 20%) of the Irish Replacement breeding objective.

	Flock	*High* genetic merit	*Low* genetic merit
Production	Ewes bred with terminal sires (%)	61	57
	Scanning rate (%)	1.87	1.70
	Lamb survival to weaning (%)	93.48	92.87
	Weaning rate (%)[1]	170	154
	Lamb weaning weight (kg)	31.52	30.65
	Lamb post-weaning weight (kg)[2]	39.24	38.29
	Days to slaughter (days)	190	203
	Lambs sold (head)	248	220
	Carcass sold per ewe joined (kg/head)	30.43	27.14
	Lambs drafted before 1 October (%)	66	57
Profitability	Lamb income (€)	27,693	24,613
	Total income (€)	30,045	26,962
	Total variable costs (€)	16,370	16,264
	Gross margin (€)	13,675	0,699
	Total fixed costs (€)	5,040	5,039
	Total farm costs (€)	21,410	21,303
	Debt servicing (€)	1,104	1,101
	Net profit (€)	7,585	4,607
	Net profit per ewe (€/ewe)	45	27
GHG emissions	Emission intensity (kg CO_2-eq/ha)	5,750	5,470
	Emissions intensity (kg CO_2-eq/kg carcass)	22.1	23.7

[1] Lambs weaned per ewe joined.

[2] Lamb weight at 120 days of age.

flock. The difference in weaning rate between the two modelled flocks appeared to be a main driver of the greater production and profitability of the *High* genetic merit flock, with more lambs weaned and sold than the *Low* genetic merit flock. Traits related to lambing and weaning rate are heavily weighted in the replacement breeding index, thus differences in lamb production for flocks with higher ranking on the index were expected.

In a secondary analysis, both flocks were modelled with *Average* terminal sires, i.e. rams ranked in the top 40 to 60% of the Terminal breeding index, so that the effects of choosing replacement vs terminal rams of *High* genetic merit could be explored. The results suggested increasing the ewe genetic merit into the next strata (quintile) ranking on the replacement breeding index to increase farm net profit by € 4/ewe. For example, a farm with a flock of *Average* genetic merit, ranked in the top 40 to 60% of the replacement index, might increase net profit by € 4/ewe through increasing flock genetic merit into the top 60 to 80% of the replacement index. Conversely, increasing terminal sire genetic merit into the next strata (quintile) ranking on the Terminal breeding index was predicted to increase farm net profit by € 1/ewe.

Without published data on differences in the feed efficiency and GHG emissions of animals divergent in genetic merit, the higher production of the modelled *High* genetic merit flock was assumed to increase flock feed intake and enteric methane emissions, GHG emissions were higher at a flock level for the *High* genetic merit flock. However, the higher weaning rate (lamb production) and lower age of lamb sale (reduced feed intake per lamb sold) contributed towards the lower emissions intensity of the *High* genetic merit flock, i.e.

the increase in production of the *High* genetic merit flock compared with the *Low* genetic merit flock was proportionally greater than the increase in GHG emissions.

Overall, the study results validated the greater productivity and profitability of flocks of higher genetic merit according to their ranking on the Irish breeding indices. Key maternal traits which did not differ between both the High and Low genetic merit flocks in the present analyses, but would be beneficial for the profitability and environmental impact of sheep production systems, were ewe longevity (determining flock replacement rate), barren rate and dam live weight. These traits should be considered in future adjustments to the Replacement breeding objective. When selecting animals of *High* genetic merit according to the national Irish replacement and terminal breeding objectives, net profit per ewe was predicted to be 67% higher and GHG emissions intensity 6.7% lower than the modelled flock with *Low* genetic merit animals. The results suggest Irish sheep producers can make substantial profit gains through selection according to the national breeding indices while also reducing their environmental impact, and farmers should consider genetic merit when purchasing their rams, particularly replacement rams.

Acknowledgements

Funding from the Irish Department of Agriculture, Food and Marine Research Fund GREENBREED (17/S/235) and Grass2Gas (2019/EN/202) is gratefully acknowledged.

References

Bohan, A., Shalloo, L., Malcolm, B., Ho, C., Creighton, P., *et al.* (2016) Agric. Sys. 148: 124-134. http://dx.doi.org/10.1016/j.agsy.2016.07.008

Harrison, M.T., Jackson, T., Cullen, B.R., Rawnsley, R.P., Ho, C., *et al.* (2014) Agric. Sys. 131: 23-33. https://doi.org/10.1016/j.agsy.2014.07.008

Jones, A.K., Jones, D., Cross, P. (2014). Agric. Sys. 123: 97-107. https://doi.org/10.1017/S0021859613000245

McHugh, N., Farrell1, L., Pabiou, T., Bohan, A., McDermott, K., *et al.* (2022). Unpublished.

O'Brien, D., Bohan, A., McHugh, N., Shalloo, L. (2016) Agric. Sys. 148: 95-104. https://doi.org/10.1016/j.agsy.2016.07.004

Ramsbottom, G., Cromie, A., Horan, B., Berry, D.P. (2012) Anim. 6(7): 1031-1039. https://doi.org/10.1017/S1751731111002503

717. Using milk fatty acids as biomarkers to improve feed conversion ratio in dairy sheep

H. Marina[1*], G. Hervás[2], R. Pelayo[1], P.G. Toral[2], A. Suárez-Vega[1], B. Gutiérrez-Gil[1], C. Esteban-Blanco[1], P. Frutos[2] and J.J. Arranz[1]

[1]University of León, 24007 León, Spain; [2]Instituto de Ganadería de Montaña (CSIC-University of León), Grulleros, 24346 León, Spain; hmarg@unileon.es

Abstract

Feed costs constitute a substantial portion of livestock production costs; therefore, improving feed efficiency (FE) is a major research topic nowadays. However, obtaining individual FE records in commercial populations requires specialized infrastructure. This research aimed to study the association between milk fatty acids (MFA) and FE in Assaf dairy sheep to determine potential MFA biomarkers for FE. Individual feed conversion ratio (FCR) of 39 ewes were estimated. Spearman correlation, regression, and orthogonal partial least squares were then used to analyze the relationship between FE and a profile of 101 MFA. We identified 33 MFA significantly associated with the FCR. These MFAs achieved a predictive performance of 0.82 for the FCR, and were successfully used to identify the most and least feed efficient animals. Thus, we propose a list of potential biomarkers to validate as an easy-to-measure phenotype for FE trait in a larger population.

Introduction

Nowadays, livestock feed efficiency (FE) is a major research topic, and it constitutes a desirable target for animal breeding. Selecting feed-efficient animals would reduce feed usage and farming costs while maintaining animal performance (Ellison *et al.* 2017). However, individual FE measurement in commercial populations is costly, time-consuming, and requires specialized infrastructure. In this sense, obtaining molecular biomarkers correlated to FE could be useful for the genetic improvement of the trait. Previous research in our group has suggested that more efficient ewes tend to have higher milk concentrations of de novo fatty acids and lower proportions of cis-9 18:1 and C20 to 22n-6 polyunsaturated fatty acids than less efficient ewes (Toral *et al.* 2021). The study presented herein uses an independent population of Assaf sheep and three analytical approaches (Spearman correlation, regression, and orthogonal partial least squares analyses) to identify potential biomarkers for FE in milk and validate those proposed in the previous study.

Materials & methods

Feed efficiency indicators. To estimate the individual FE of 39 Spanish Assaf ewes housed in individual tie stalls, we monitored the dry matter intake (DMI) by weighing the feed offered and refused by each animal over four weeks. Moreover, the milk yield and milk composition data were used to calculate the energy-corrected milk (ECM) parameter according to INRA (2018). The feed conversion ratio (FCR) was calculated as the relationship between the average DMI and ECM.

Milk composition. Lipids in 1 ml of milk sampled from the 39 Assaf ewes were extracted and converted to fatty acid methyl esters by base-catalyzed transesterification (Shingfield *et al.* 2003). The fatty acid methyl esters profile was determined following the procedure detailed by Hervás *et al.* (2021).

Centered-log ratio. The constant sum of the 101 milk fatty acids (MFAs) examined in this study suggests compositional data (Gloor *et al.* 2016). Therefore, MFA measurements were transformed via centered log-ratio (CLR) for correcting sparse compositionality (Aitchison, 1982).

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_717, © H. Marina *et al.* 2022

Association analyses. To assess whether MFA composition discriminates animals in high and low FCR, three quantitative analyses were applied. In the first analysis, we estimated the Spearman correlation and regression coefficient for each MFA analysed with the FCR parameter to characterise the association level of the CLR-transformed MFA measurements ith the FCR. Secondly, we carried out an orthogonal partial least squares (OPLS) analysis on the CLR-transformed MFA data and an orthogonal partial least squares discriminant analysis (OPLS-DA) considering only the ten extreme animals (most and least efficient) ranked using the FCR index. These last analyses (OPLS and OPLS-DA) aimed to identify those MFAs useful in predicting the FCR parameter and were performed using the R package ROPLS (Thévenot *et al.* 2015).

Results

Basic statistics. The MFA phenotypes were normally distributed. The average of the CLR-transformed MFA values was close to zero, and the minimum standard deviation corresponded to the *'iso 17:0'* MFA (0.07) and the maximum to the *'trans-11 cis-15 18:2'* MFA (1.06). The distribution of the values was not affected by the CLR transformation, as expected. Furthermore, the FCR parameter was on average 1.36 with a standard deviation of 0.26.

Association analysis. Firstly, the Spearman correlation analyses between the MFA matrix and FCR revealed a total of 50 significant correlations (P-value<0.05) ranging from 0.32 to 0.70 (absolute values). Furthermore, the regression analysis revealed 38 significant associations (P-value <0.05) between the MFAs and the FCR parameters studied. The R^2 Adjusted values identified ranged from 0.08 to 0.31. Table 1 depicts the 36 common MFAs associated with the FCR parameter identified through both the Spearman correlation and the regression analyses.

Orthogonal partial least squares analysis (OPLS). MFAs achieved a predictive performance of 0.82 for the FCR index. A total of 40 MFAs achieved variable importance in the projection (VIP) score higher than one (P<0.05) (Figure 1A). Out of the 40 MFAs highlighted by the OPLS analysis, 33 matched the 36 MFAs highlighted by Spearman correlation and regression analysis (Table 1). The OPLS-DA revealed a predictive performance of 0.90 for the FCR index. As can be appreciated in Figure 1B, the results of the OPLS-DA confirmed the ability of the values of the 101 MFAs considered in this study to distinguish between the most efficient and least efficient group of animals in terms of FCR.

Discussion

The concordances in the results obtained by the three independent analyses (Spearman correlation, regression, and OPLS analyses) performed in this study emphasise the importance of this list of MFA as potential biomarkers for the FCR index. Moreover, we also identified 17 out of 33 MFA previously described as associated with FE in dairy sheep by Toral *et al.* (2021) (Table 1). Ten long-chain MFAs out of the 33 MFAs highlighted in this study showed negative correlations with FCR, suggesting that selecting for higher FE would reduce their concentration in the milk. The negative correlation between cis-9 18:1 and FE was also previously estimated by Toral *et al.* (2021), which influences the physical properties of milk fat by increasing its melting point (Chilliard *et al.* 2000). Particularly, the FCR index showed a positive and negative correlation with the lauric acid (C12:0) and conjugated linoleic acid (CLA), respectively. Both fatty acids have been extensively studied in relation to human health (Markiewicz-kęszycka *et al.* 2013; Djordjevic *et al.* 2019). The results from OPLS analyses indicate that the MFA profile can be used to predict the FCR, which could potentially reduce the costs associated with its measurement and enhance its implementation in breeding programs. In addition, the discriminant analysis (OPLS-DA) revealed that the MFA profile could be used to classify the most and least FE animals. The robust associations revealed in this study between certain MFAs and the FCR suggest that selecting the most efficient animals in terms of milk production might impact the MFA profile of dairy sheep. Regarding the importance of the milk

Table 1. Significant Spearman correlation and regression coefficients estimated between the MFAs and the FCR parameter.[1]

Milk fatty acids	Spearman coef.[2] (*P*-value)	R adjusted[3] (*P*-value)	Milk fatty acids	Spearman coef.[2] (*P*-value)	R adjusted[3] (*P*-value)
7:0[§]	0.403 (**)	0.081 (*)	*9:0*	0.465 (**)	0.119 (*)
11:0	0.539 (***)	0.182 (**)	*12:0*[†]	0.553 (***)	0.153 (**)
15:0[†]	0.494 (***)	0.172 (**)	*17:0*[†]	-0.585 (***)	0.194 (**)
22:0[†]	0.687 (***)	0.277 (***)	*24:0*	0.704 (***)	0.294 (***)
16-oxo-18:0	-0.577 (***)	0.303 (***)	*18:3n-3*[†]	0.458 (**)	0.15 (**)
18:3n-6	0.345 (*)	0.091 (*)	*20:5n-3*[†]	0.393 (*)	0.127 (*)
cis-11 18:1[†]	-0.434 (**)	0.101 (*)	*cis-11 20:1*[†]	-0.492 (***)	0.155 (**)
cis-12 14:1[†]	0.418 (**)	0.113 (*)	*cis-12 18:1*[†]	0.521 (***)	0.203 (**)
cis-13 18:1	-0.428 (**)	0.117 (*)	*cis-16 18:1*	0.63 (***)	0.313 (***)
cis-7 14:1[†]	0.376 (*)	0.084 (*)	*cis-9 12:1*	0.429 (**)	0.128 (*)
cis-9 17:1	-0.61 (***)	0.264 (***)	*cis-9 18:1*[†]	-0.544 (***)	0.159 (**)
cis-9 trans-11 trans-15 CLnA	0.365 (**)	0.076 (*)	*cis-9 trans-13 18:2*[§]	0.364 (*)	0.124 (*)
iso 15:0[§]	0.399 (**)	0.105 (*)	*iso 17:0*[†]	0.505 (***)	0.267 (***)
iso 18:0[†]	-0.384 (**)	0.104 (*)	*trans-10 18:1*[†]	-0.595 (***)	0.261 (***)
trans-10,trans-14 18:2	0.366 (**)	0.116 (*)	*trans-11 cis-15 18:2*	-0.57 (***)	0.154 (**)
trans-11 trans-13 CLA	0.479 (**)	0.219 (**)	*trans-12 18:1*[†]	0.375 (*)	0.129 (*)
trans-15 18:1	0.466 (**)	0.258 (***)	*trans-4 18:1*	-0.435 (**)	0.183 (**)
trans-9 cis-11 CLA[†]	-0.478 (**)	0.143 (**)	*trans-9 trans-12 cis-15 + cis-9 cis-12 trans-15 18:3*	0.527 (***)	0.157 (**)

[1] [§] = MFA not associated with the FCR parameter according to the OPLS analysis; [†] = MFA in agreement with the results published by Toral *et al.* (2021); * *P*-value≤0.05; ** *P*-value≤0.01; *** *P*-value≤0.001.

[2] Spearman Coefficients and its correspondents *P*-values between brackets.

[3] R^2 Adjusted values and their corresponding *P*-values between brackets.

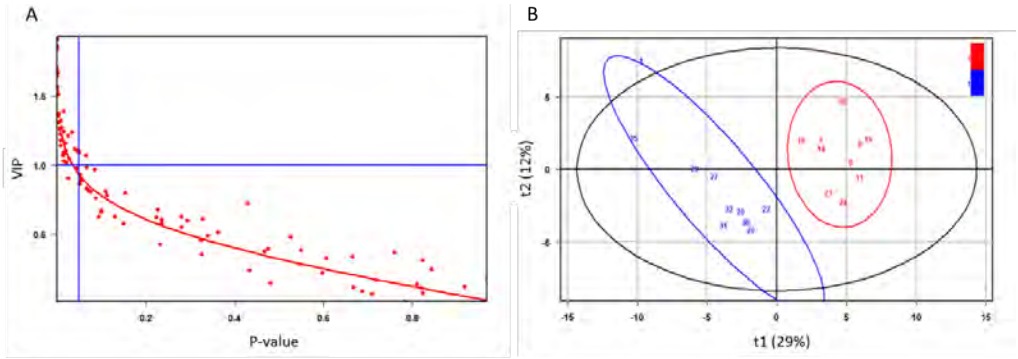


Figure 1. Results from the OPLS and OPLS-DA analyses. (A) Representation of VIP scores (Y-axis) and *P*-values (X-axis) reported by the OPLS analysis for the 101 MFAs analysed in relation to the FCR parameter. (B) Representation of the most (coloured in blue: 1) and least (coloured in red: 2) efficient animals included in the OPLS-DA concerning the FCR parameter studied in this work.

composition for dairy sheep products (Moioli *et al.* 2007), further studies should focus on the advantages or consequences of implementing FCR as a breeding target on the technological and organoleptic characteristics of dairy sheep products.

Acknowledgements

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 772787 (SMARTER).

References

Aitchison, J. (1982). J. R. Stat. Soc. 44:139–177. https://www.jstor.org/stable/2345821

Chilliard, Y., Ferlay, A., Mansbridge, R.M., and Doreau, M. (2000). Ann. Zootech. 49:181–205. https://doi.org/10.1051/ANIMRES:2000117

Djordjevic, J., Ledina, T., Baltic, M.Z., Trbovic, D., Babic, M., *et al.* (2019). IOP Conf. Ser. Earth Environ. Sci. 333. https://doi.org/10.1088/1755-1315/333/1/012057

Ellison, M.J., Conant, G.C., Lamberson, W.R., Cockrum, R.R., Austin, K.J., *et al.* (2017). Small Rumin. Res. 156: 12–19. https://doi.org/10.1016/J.SMALLRUMRES.2017.08.009

Gloor, G.B., Wu, J.R., Pawlowsky-Glahn, V., and Egozcue, J.J. (2016). Ann. Epidemiol. 26:322–329. https://doi.org/10.1016/J.ANNEPIDEM.2016.03.003

Hervás, G., Toral, P.G., Fernández-Díez, C., Badia, A., and Frutos, P. (2021). Anim. 11: 2476. https://doi.org/10.3390/ANI11082476

INRA (2018). Alimentation des ruminants; Editions Quae: Versailles (France), 2018.

Markiewicz-kęszycka, M., Czyżak-runowska, G., Lipińska, P., and Wójtowski, J. (2013). Bull Vet Inst Pulawy. 57:135-139. https://doi.org/10.2478/bvip-2013-0026

Moioli, B., D'Andrea, M., and Pilla, F. (2007). Small Rumin. Res. 68:179–192. https://doi.org/10.1016/J.SMALLRUMRES.2006.09.008

Shingfield, K.J., Ahvenjärvi, S., Toivonen, V., Ärölä, A., Nurmela, K.V.V., *et al.* (2003). Anim. Sci. 77:165–179. https://doi.org/10.1017/S1357729800053765

Thévenot, E., Roux, A., Xu, Y., Ezan, E., and Junot, C. (2015). J. Proteome Res. 14:3322–3335. https://doi.org/10.1021/ACS.JPROTEOME.5B00354

Toral, P.G., Hervás, G., Fernández-Díez, C., Belenguer, A., and Frutos, P. (2021). J. Dairy Sci. 104:5569–5582. https://doi.org/10.3168/JDS.2020-19061

718. Six years of genomic selection have increased the genetic gain in French dairy sheep

J.M. Astruc[1*], G. Lagriffoul[1,2], A. Legarra[3] and D. Buisson[1]

[1]Institut de l'Elevage, 31321 Castanet-Tolosan, France; [2]Comité National Brebis Laitières, 31321 Castanet-Tolosan, France; [3]INRAE, GenPhySE, 31326 Castanet-Tolosan, France; jean-michel.astruc@idele.fr

Abstract

Genomic selection (GS) started in French dairy sheep in 2015, first in the Lacaune breed, then in the Pyrenean breeds, and finally in the Corsican breed. This was favoured by a strong genetic organisation and enabled thanks to an intensive phase of research and development to undertake all the prerequisites: constitution of reference populations, implementation of a pipeline for genotype editing and of a genomic evaluation, design of a GS scheme suitable to each breed. This paper first presents an overview on how GS was organised in dairy sheep in France and in which way the breeding scheme differs from the conventional program based on progeny tested. Second, the results of 6 years of GS in the Lacaune breed and 4 years in the Pyrenean breeds are studied in terms of genetic gain, both through the evolution of the breeding values over time and through the analysis, year after year, of the different pathways of selection.

Introduction

Selection programs are implemented in French dairy sheep in 5 breeds: the Lacaune breed (LAC), the Pyrenean breeds (Manech Tête Rousse – MTR; Manech Tête Noire – MTN; Basco-Béarnaise – BB), the Corse breed (COR). The programs are designed on a pyramidal organisation of the population with selection flocks limited to 15 to 25% of the total ewes, targeting at generating the genetic gain and producing the rams for the whole population. Before genomic era, conventional breeding schemes were based on progeny-test of artificial insemination (AI) rams. The AI rate in the selection flocks ranges from 45% (COR) to 85% (LAC). The number of yearly progeny-tested rams varied from 20 rams in COR breed to 450 in LAC breed. Accordingly, the annual genetic gain (AGG) obtained was estimated between 0.09 (COR) to 0.22 (LAC) genetic standard deviation. Consequently, the reference populations are constituted by progeny-tested rams with a depth getting back to the early 2000's. After half a decade of R&D programs aiming at genotyping the rams, building pipelines for genotype handling and genomic prediction, assessing the feasibility and the cost-benefit of genomic selection (GS) in dairy sheep situation, the LAC breed decided in 2015 to shift to GS (replace the progeny-test process by using the genomically pre-selected rams). The Pyrenean breeds followed in 2020 and the COR breed in 2020. Initiation and conception of French dairy sheep GS was the last undertaking and accomplishment before retirement of Francis Barillet, who spent his successful whole career in dairy sheep genetics. We dedicate this paper to him.

Materials & methods

Genomic predictions. Genomic evaluation has been implemented since 2015 with the shift to GS of the LAC breed. We compute a single-step GBLUP from the package BLUPf90 (Misztal *et al.* 2002, Aguilar *et al.* 2010 and 2011), using all performances, pedigree and genotyped rams. The predictions were validated, showing a significant gain of precision compared to the pedigree predictions (Baloche *et al.*, 2014). In 2021, the reference populations vary across breeds (Table 1), from 326 rams in the COR breed up to 6,397 in the LAC breed.

Genomic breeding schemes. GS schemes were designed following different simulations aiming at optimising the genetic gain and the cost of the program (Buisson *et al.*, 2014, 2018). The key points are the use of fresh semen in sheep and the seasonality of the mating period (the majority of the AI are concentrated

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_718, © J.M. Astruc *et al.* 2022

Table 1. Importance of genomic breeding schemes in French dairy sheep in 2021.

Breed	First year of GS	Females in selection	# new AI rams/ year	# adult AI rams	# sires of sires	# rams genotyped/ year	Reference population[1]
LAC	2015	192,928	297	674	75	2,974	6,397
BB	2017	28,061	60	110	26	240	880
MTN	2017	10,547	28	68	22	102	585
MTR	2017	84,286	174	302	49	714	2,842
COR	2020	18,863	21	28	14	374	326

[1] Rams genotyped with daughters.

in 2-3 months). This leads to the presence in the AI centre of a great number of rams. The young rams are genotyped at 3-months old and a strong pre-selection on genomic prediction is applied (with a selection intensity equivalent to an elimination rate of 75% in the LAC breed). At 8 months (LAC breed) or 1.5 year (other breeds), the rams are used in AI for the first time in the selection flocks with the goal to maintain the reference population. Very few of them are considered as potential sire of breeding rams because, as young rams, they have a too low semen production to be mated to enough dams of rams. There is no longer lay-off, and at 1.5 year-old (LAC breed) or 2.5 year-old (other breeds), the rams are fully used as sire of males and females, both in the selection flocks and in dissemination towards the base population. When daughters are included in the predictions (2.5 or 3.5 year-old depending on the breeds), a second selection occurs, lower than in conventional progeny-test programs, with a 20% elimination. The rams are then used based on prediction with daughters, during 1 to 3 years. The dairy sheep breeding schemes are therefore hybrid, with a part of the rams with genomic prediction without daughters performances (genomic rams), and a part of rams with daughters performances. In 2021, in the selection flocks, the rate of AI from genomic rams varies from 62% (LAC) to 88% (MTN).

Breeding goal. The breeding goals differ across the breeds. As the COR breed selects for milk yield only, the other breeds select for a total merit index (TMI), named ISOL, including fat and protein yield, fat and protein content, somatic cell score (SCS), udder morphology. The respective part of production traits (yield, content) and functional traits (SCS, udder) in the TMI is 50/50 in the LAC breed, 56/44 in the MTR breed, 59/41 in the BB breed, 91/9 in the MTN breed. Within the functional trait, SCS and udder account for the same part. ISOL is expressed in equivalent liters, like milk yield, since the breeders got used to its dimension and scale.

Estimation of genetic gain through evolution of EBVs. Genetic gain over a given period may be estimated by the slope of a plot of the average TMI (or the elementary traits) into the birth years. The method is used both for males and females. We estimated the slope for a 15-year period before GS and for the period since GS has started. This comparison could actually be done only in the LAC breed, for which GS started in 2015, allowing to get a reliable genetic gain across a 5-year period.

Estimation of expected genetic gain through the pathways of selection. The selection pathways methodology permits a vision of the effort of selection each year. Thus we studied how the replacement males and females of a given birth year (e.g. 2021) are generated from their sires and dams. We calculated the AGG in all 4 selection paths (MM, MF, FM and FF for males to males, males to females, females to males, and females to females, respectively), as follows (Rendel and Robertson, 1950; Lindhe, 1968):

$$AGG = \frac{\Delta GMM + \Delta GMF + \Delta GFM + \Delta GFF}{LMM + LMF + LFM + LFF}$$

where ΔG is the estimated genetic superiority of the selected animals over their contemporaries born in the same year, L is the average age of the selected animals when their offspring were born, and AGG is the annual genetic gain. ΔG was calculated by multiplying the selection differential by the root square of the reliabilities. We verified that the standard deviation of ISOL (expressed in milk yield scale) was very similar to the genetic standard deviation of milk yield. The sires of rams (MM) and dams of rams (FM) are respectively the sires and dams of the genotyped rams. The candidate males are the progeny tested rams (for birth year within the conventional selection period) or the genotyped rams (for birth year within the GS period. The selection differentials are calculated within each birth year of sire and dam, then averaged across year, weighted by the part of the replacement animals produced by each sire's and dam's birth year. The unknown sires and dams are not considered. All the rams have known dams and sires. The percentage of ewes born from unknown dam is 0.7, 3.7, 4.6, 1.9, 19.2, respectively in LAC, BB, MTN, MTR and COR breeds. The percentage of ewes born from unknown sire is 4.1, 29.9, 30.7, 19.7, 62.7, respectively in LAC, BB, MTN, MTR and COR breeds.

Results & discussion

The evolution of TMI over the time in the LAC breed (Figure 1) shows a clear inflexion after 2015, with the implementation of GS. Between 2000 and 2014, the estimated yearly AGG (expressed in equivalent litres) is 5.74 (0.19 σ_g) and 5.23 (0.17 σ_g) in rams and ewes. After 2015, the AGG is 9.19 (0.31 σ_g) and 7.96 (0.27 σ_g) in rams and ewes. The extra genetic gain due to GS is 60% in rams and 52% in ewes. Such an inflexion is not yet clear in the other breeds since the GS is more recent. Moreover the shift in selection criteria at the same time of GS in the Pyrenean breeds, makes a reliable analysis on a long period difficult.

The AGG observed using the methodology of the selection pathways are presented in Table 2. When comparing the AGG calculated in 2021 vs the AGG calculated as an average of the years 2011-2015, we can derive the increase in AGG in the LAC and Pyrenean breeds. This increase is 57% in the LAC breed and ranges from 16 to 29% in the Pyrenean breeds. This difference may be attributed to how long the GS has been implemented (6 years in LAC vs 3 years in the Pyrenean breeds).

The results obtained with both methods are similar in the LAC breeds (0.35 σ_g with the pathways method; 0.27-0.31 with the TMI evolution method). The slight differences comes from the fact that the slope in the TMI evolution reflects a mean increase over 6 years. Yet the effort of selection since 2015 has increased steadily, as the rams used were more and more generated by a process of genomic pre-selection before use in AI.

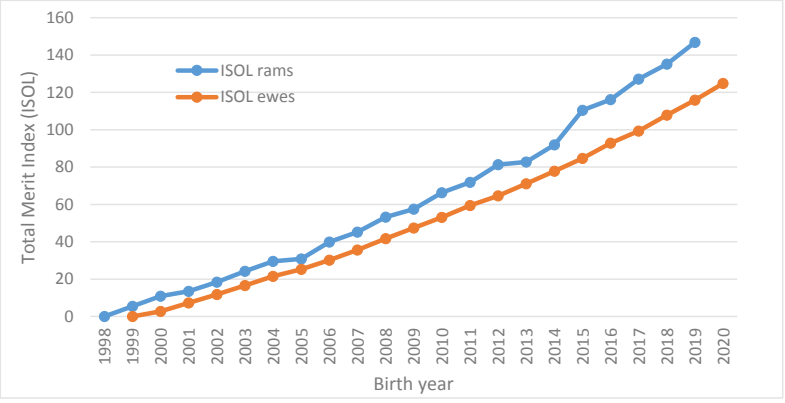


Figure 1. Evolution of rams and ewes TMI over the birth years in the Lacaune breed.

Table 2. Genetic trends (in genetic standard deviation or σ_g) obtained through the selection pathways in the French dairy sheep breeds in genomic situation and comparison with classical program.

Breed	Sum of progress in litres	Sum of generation interval in years	σ_g of milk yield (litres)	Genetic trend 2011-2015 in σ_g	Genetic trend 2021 in σ_g	Extra genetic gain genomic vs classical program
Lacaune	4.13	11.93	30	0.22	0.35	+57%
BB	2.07	13.66	26	0.13	0.15	+16%
MTN	1.74	14.92	25	0.09	0.12	+29%
MTR	2.15	13.24	28	0.13	0.16	+25%
Corse	2.23	19.91	15	0.11	0.11	

Conclusions

Extra genetic gain permitted by GS in the LAC breed appears to be above the predictions obtained in the simulations (Buisson *et al.*, 2014), thanks to a strong selection intensity of the young rams. Close attention is paid by the breeding programs to manage the inbreeding, which is made easier by the high number of rams used quite evenly. The highest genetic gain in GS is intended to be used for including new resilience (functional longevity, resistance to parasites) and efficiency (proxies of feed efficiency) traits in the next years. These good results obtained from GS and the motivating perspectives towards more balanced breeding goals are strong drivers to reinforce the genetic organisation that has existed for decades in French dairy sheep.

References

Aguilar I., Misztal I., Johnson DL., Legarra A., Tsuruta S. *et al.* (2010) J. Dairy Sci. 93:743-752.

Aguilar I., Misztal I., Legarra A., Tsuruta S. (2011) J. Anim. Breed. Genet., 128, 422-428.

Baloche, G., Legarra, A., Sallé, G., Barillet F. *et al.* (2014) J. Dairy Sci. 97:1107-1116

Buisson D., Lagriffoul G., Baloche G., Astruc J.M., Barillet F. *et al.* (2014) Proc. of the 10th WCGALP, Vancouver, Canada.

Buisson D., Lagriffoul G., Legarra A., Astruc J.M., *et al.* (2018) Proc. of the 11th WCGALP, Auckland, New Zealand.

Lindhe, B. (1968). Acta Agric. Scand., 18, 33-41.

Misztal I., Tsuruta S., Strabel T., Auvray B., Druet T., *et al.* (2002) Proc. of the 7th WCGALP, Montpellier, France.

Rendel J.M., Robertson A. (1950) J Genet 50(1):1–8

719. Cross-validation of single-step genetic evaluation in U.S. Katahdin sheep

A.J. McMillan[1*], D.J. Brown[1], J.M. Burke[2], J.L. Morgan[3] and R.M. Lewis[4]

[1]AGBU, a joint venture of NSW Department of Primary Industries and University of New England, 2351, Armidale, Australia; [2]USDA, ARS, Dale Bumpers Small Farms Research Center, Booneville, AR 72927, USA; [3]Round Mountain Consulting, Fayetteville, AR 72701, USA; [4]Department of Animal Science, University of Nebraska – Lincoln, Lincoln, NE 68583, USA; amcmill4@une.edu.au

Abstract

Genomic information is used in genetic evaluation to improve prediction accuracy in most livestock species. Such is less so in sheep, particularly in the U.S. In this study, the impact of implementing single-step genomic BLUP as compared to pedigree BLUP for weight and faecal egg count traits was evaluated in U.S. Katahdin sheep. Two methods of cross validation were utilised to compare the predictive ability and bias of estimated breeding values of these methods. Genomic information improved predictive ability for both traits, and reduced bias in the evaluation of faecal egg counts. Accuracies of estimated breeding values improved appreciably in genotyped animals. The benefit from including genomic information based on cross validation was minimal but is expected to improve as the reference population grows. Single-step genomic BLUP procedures developed will be used to update those applied in the routine genetic evaluation offered through the U.S. National Sheep Improvement Program.

Introduction

Many global genetic evaluation systems now utilise genomic information routinely. To date, this has not been possible for the National Sheep Improvement Program (NSIP) evaluation system, which provides genetic evaluation services to the U.S. sheep industry (Notter, 1998). Through a U.S. Department of Agriculture grant, and in collaboration with NSIP Katahdin breeders, a large number of genotypes and phenotypes were collected on key traits. Katahdin are an increasingly popular hair sheep breed in the U.S., recognized for tolerance to parasite resistance and twinning. These data provided a foundation to test and implement genomically assisted genetic evaluation. As a result of this research and development, and as a first for the U.S. sheep industry, NSIP recently implemented single step genomic BLUP (SS-GBLUP) (Legarra *et al*., 2014) within their genetic evaluation for Katahdin sheep. SS-GBLUP has been implemented and is successfully running in Australian evaluations since 2016 (Brown *et al.,* 2018), each using the same evaluation platform (OVIS; Brown *et al.,* 2007*)*. The aim of this study was to use cross validation methods with forward prediction to compare pedigree based BLUP (PBLUP) with SS-GBLUP in the Katahdin evaluation for weaning weight (wwt), post-weaning weight (pwt), weaning faecal egg count (wfec), and post-weaning faecal egg count (pfec).

Materials & methods

Data for this paper was based on the June 2021 Katahdin NSIP genetic evaluation. The analysis has close to 100,000 sheep and is multi-trait with 12 traits across age stages including body weights, carcase scans, faecal egg counts and reproductive traits. Approximately 5,000 animals have been genotyped using the GeneSeek Genomic Profiler Ovine 50k array (Neogen Corp., Lincoln, NE, USA) with most genotyped animals being born within the last 5 years.

Forward cross validation was undertaken using both PBLUP and SS-GBLUP procedures. These were full multi-trait analyses. The SS-GBLUP evaluation used the equations of Aguilar *et al.* (2010). A lambda value

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_719, © A.J. McMillan *et al.* 2022

of 0.5 was used (McMillan and Swan, 2017), which placed equal emphasis on the pedigree and genomic relationships for animals when both recorded.

Phenotypic data were pre-corrected for standard fixed effects such as age of dam, birth-rearing type, and age of measurement; a fixed effect for contemporary group was included directly. Phenotypes for animals born after 2018 were removed from the data and the BLUP analyses used to estimate breeding values (EBV) from this partial dataset. In Table 1, phenotypic data for wwt, pwt, wfec and pfec are shown. Accuracy of prediction was assessed by examining the ability for EBV to predict phenotypes pre-corrected using the OVIS evaluation procedures (Brown *et al.*, 2007). The expectation of these regressions is a slope of 1.0. The accuracy of prediction was also calculated using the correlation of adjusted phenotypes with EBV and dividing by the square root of the heritability. As an additional method of validation, accuracy and bias metrics were calculated using the LR method (Legarra and Reverter, 2018) as the correlation and regression slopes between the EBV from the full analysis and the partial analysis for the animals in the validation group with phenotypes (Legarra *et al.,* 2008). The EBV accuracy was estimated for each model using OVIS and compared as a further test of the improvements in predictive ability from each model.

Results

Changes in phenotypic predictive ability and bias are shown in Table 2. For all traits the SS-GBLUP approach increased predictive ability based on the variance explained by the model and accuracy of prediction. The improvements ranged from 2% for wwt up to 10% for pfec across both metrics. For the weight traits, there was a tendency for genomic information to overestimate the phenotypic differences compared to PBLUP. Although, the fec traits showed less bias with the SS-GBLUP model, the phenotypes still were under-predicted.

Result for the EBV regression are shown in Table 3. Based on the correlation of EBV, over all for animals in the validation set, the predictive ability slightly increased for all traits with the incorporation of genomic information. For animals with a genotype, the improvement was more substantial and ranged from 2-3% across all traits.

The bias was either not significantly different or improved when comparing SS-GBLUP to PBLUP. In genotyped animals the bias was closer to the expectation of 1 across traits with SS-GBLUP. The trait showing the largest bias under SS-GBLUP evaluation for genotyped animals was wfec although the bias was much improved compared to PBLUP evaluation.

Table 1. Summary of pre-2018 (training) and post-2018 (validation) datasets.

Trait[1]	Pre-2018			Post-2018		
	Records	Mean	Std	Records	Mean	Std
wwt	57,791	20.52	3.99	14,038	20.52	4.22
pwt	44,525	31.50	4.82	12,254	31.70	5.08
wfec	6,910	9.41	5.96	3,522	8.89	5.53
pfec	9,779	8.78	5.63	4,058	9.63	5.37

[1] wwt = weaning weights (kg); pwt = post-weaning weight (kg); wfec = weaning faecal egg count (egg/g, cube root transformed); pfec = post-weaning faecal egg count (egg/g, cube root transformed).

Table 2. Results from regression of adjusted phenotypes and accuracy of prediction and EBV using each evaluation method.

Analysis[1]	Trait[2]	Slope (s.e.)	R^2	Accuracy
PBLUP	wwt	1.08±0.08	0.39	0.64
PBLUP	pwt	0.97±0.06	0.38	0.57
PBLUP	wfec	0.57±0.08	0.27	0.28
PBLUP	pfec	0.58±0.07	0.24	0.13
SS-GBLUP	wwt	1.20±0.07	0.41	0.70
SS-GBLUP	pwt	1.12±0.06	0.44	0.63
SS-GBLUP	wfec	0.63±0.08	0.33	0.35
SS-GBLUP	pfec	0.70±0.07	0.34	0.19

[1] PBLUP = pedigree BLUP; SS-GBLUP = single-step genomic BLUP.

[2] Traits defined as in Table 1.

Table 3. Results from LR method for regression of full EBV with part EBV.

Analysis[1]	Trait[2]	All animals			Genotyped animals		
		Records	Slope (s.e.)	Corr	Records	Slope (s.e.)	Corr
PBLUP	wwt	13,768	1.00±0.01	0.77	2,076	0.94±0.02	0.79
PBLUP	pwt	11,784	0.92±0.01	0.71	2,529	0.95±0.02	0.76
PBLUP	wfec	3,394	0.84±0.02	0.62	1,691	0.85±0.03	0.60
PBLUP	pfec	3,906	0.88±0.02	0.65	1,761	0.97±0.02	0.68
SS-GBLUP	wwt	13,768	0.98±0.01	0.77	2,076	1.00±0.02	0.81
SS-GBLUP	pwt	11,784	0.93±0.01	0.72	2,529	1.04±0.02	0.79
SS-GBLUP	wfec	3,394	0.87±0.02	0.65	1,691	0.90±0.03	0.64
SS-GBLUP	pfec	3,906	0.94±0.02	0.67	1,761	1.02±0.02	0.71

[1] PBLUP = pedigree BLUP; SS-GBLUP = single-step genomic BLUP.

[2] Traits defined as in Table 1.

The comparison of EBV accuracy from each model is shown in Figure 1. Clear improvements in EBV accuracy are seen for genotyped animals from the SS-GBLUP model compared to PBLUP. Such was particularly the case where the accuracies based on PBLUP were low.

Discussion

Accuracy of prediction of phenotypes was significantly improved with the use of SS-GBLUP for all traits. The benefit was more obvious for fec traits. This is likely a function of the difference in heritabilities, and the number of phenotypes available on fec as compared to weight traits on the validation animals as weights are more commonly recorded. Thus, the weight traits had a larger amount of data supporting them compared to fec, and therefore a higher starting accuracy. The benefit from genomic information therefore becomes smaller. These differences were more obvious for genotyped animals.

The prediction of phenotypes from EBV were more biased for weight traits when comparing SS-GBLUP and PBLUP. In contrast, the predictions of phenotypes for fec traits were much less under predicted with genomic information. The benefit of the SS-GBLUP model can also be observed by comparing the improvement in estimated EBV accuracy from this model over PBLUP. This benefit is greater for animals with lower PBLUP accuracy.

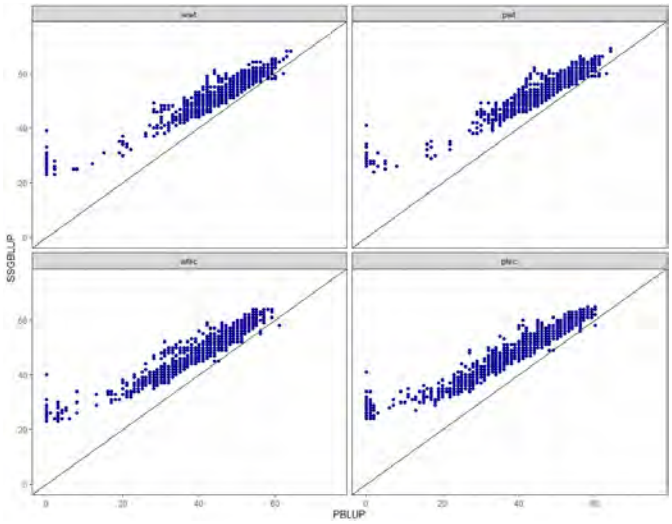


Figure 1. Comparison of EBV accuracy for genotyped animals from pedigree BLUP (PBLUP) and single-step genomic BLUP (SS-GBLUP) models for each trait.

In most cases, though, a small improvement in accuracy of prediction was observed. The benefit from including genomic information based on cross validation is minimal currently but is expected to improve as the size of the reference population grows. The new genetic evaluation procedures and models developed are now being used in the routine genetic evaluation of U.S. Katahdin sheep through NSIP. They will be the foundation for applying SS-GBLUP to other U.S. breeds once sufficient genomic information becomes available.

Acknowledgements

This project is supported by the Agriculture and Food Research Initiative Grant number 2016-51300-25723 from the USDA National Institute of Food and Agriculture, Organic Agriculture Research and Extension Initiative program.

References

Aguilar I., Misztal I., Johnson D.L., Legarra A., Tsuruta S. *et al.* (2010) J. Dairy Sci. 93:743-752. https://doi.org/10.3168/jds.2009-2730

Brown D.J., Huismann A.E., Swan A.A., Graser H.-U, Woolaston R.J. *et al.* (2007) Proc. Of the 18th AAABG, Armidale, Australia.

Brown D.J., Swan A.A., Boerner V., Li l., Gurman P.M., *et al.* (2018) Proc. of the 11th WCGALP, Auckland, New Zealand.

Legarra A., Christensen O.F., Aguilar I., and Misztal I. (2014) Livestock Science 166:54-65. https://doi.org/10.1016/j.livsci.2014.04.029

Legarra A, Robert-Granié C, Manfredi E, and Elsen J-M. (2008) Genetics 180(1):611-618. https://doi.org/10.1534/genetics.108.088575

Legarra A., and Reverter A. (2018) Genet. Sel. Evol. 50:53. https://doi.org/10.1186/s12711-018-0426-6

McMillan A.J., and Swan A.A (2017) Proc. of the 22nd AAABG, Townsville, Australia.

Notter, D.R. (1998) J. Anim. Sci. 76(9):2324–2330. https://doi.org/10.2527/1998.7692324x

720. Partitioning of genetic trends by flock and gender in Istrian sheep breed

M. Spehar[1*], J. Ramljak[2], V. Poslon[2] and A. Kasap[2]

[1]Croatian Agency for Agriculture and Food, Svetošimunska 25, 10000 Zagreb, Croatia; [2]University of Zagreb Faculty of Agriculture, Svetošimunska 25, 10000 Zagreb, Croatia; marija.spehar@hapih.hr

Abstract

Istrian sheep is Croatian native breed traditionally selected for milk yield, but the success of past selection efforts has never been studied. Genetic trends play an important role in this regard, and their decomposition is effective in identifying the main contributors to trait related genetic change. The objective of this study was to estimate recent genetic trends (2014-2018) obtained under framework of single step genomic BLUP (test-day animal model) and to partition them by gender and flock. The overall genetic trend was positive, but negligible and inconsistent. Males contributed slightly more to the overall genetic trend than females, while the partitioning by flock was inconclusive. The results implicate absence of systematic selection in the recent period and impose a need to conduct selection based on the estimated breeding values. The ongoing transition to single step genomic BLUP in this population will provide sufficiently accurate breeding values to provide faster and consistent selection gain.

Introduction

Istrian sheep is Croatian native multi-purpose breed (milk-meat). The milk recording in this population started in early 2000 following ICAR guidelines, while the BLUP genetic evaluation for dairy traits has been carried out for about a decade. Recently, genotyping of a large part of the population with 50K SNP chip has been conducted. The aim is to upgrade the existing genetic evaluation system to genomic selection in order to increase the accuracy of estimated breeding values (EBVs) (Duchemin *et al.*, 2012; Legarra *et al.*, 2014). There are many uncertainties about success of historical selection work in this population since up to date there has not been a systematic analysis of genetic trends. Monitoring of selection success (genetic trends) can be simply obtained by averaging EBVs per birth year (Blair and Pollak, 1984) or by regressing BVs on birth year. Following the idea of Garcia-Cortes *et al.* (2008), these trends can be partitioned into contributions by the predefined different 'paths' (country, gender, flock, line, etc.). This can provide additional information on the main contributors to selection gain. The method has already been used to assess the contribution of different countries in Brown Swiss cattle (Gorjanc *et al.*, 2011), Croatian Simmental cattle (Špehar *et al.*, 2011), and two pig breeds in Croatia (Škorput *et al.*, 2015). To our knowledge, there are no reports of using this methodology in sheep. In order to analyse the success of previous selection work in the Istrian sheep, we estimated genetic trend for daily milk yield and decomposed it by flock and gender.

Materials & methods

Data. Phenotypic records for daily milk yield (DMY) and pedigree were provided by the Ministry of Agriculture. Phenotypic records were collected following the ICAR guidelines (ICAR, 2018). Before the final inferential statistical analysis, the original data set was 'pruned' leaving a total of 37,703 records obtained from 5,071 ewes. The pedigree file was created using all available relationships to phenotyped ewes and included a total of 7,208 animals. A total of 719 animals were genotyped using the Illumina OvineSNP50K BeadChip® (52,152). Genotyped animals and SNPs were included in the analysis after quality control parameters for call rate per animal, call rate per SNP, and MAF set to 0.9, 0.9, and 0.05, respectively. A total of 26 animals and 5,150 did not pass the above quality control. Monomorphic markers (n=898) and markers with unknown genome position or located on the sex chromosome (n=1,176) were also removed. The final number of genotyped animals and markers was 693 and 45,349, respectively.

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_720, © M. Spehar *et al.* 2022

Genetic evaluation. The BLUPF90 family of programs (Misztal *et al.*, 2018) was used in genetic evaluation of the reference population. RENUMF90 was used to renumber data sets and prepare parameter files for estimation of variance components (AIREMLF90) and estimation of BVs (BLUPF90). Covariance components were estimated using the AI REML algorithm within single trait repeatability test–day model. The model included the following fixed class effects: parity, litter size, season of lambing, and flock. Days in milk (DIM) and age at lambing were fitted as covariates. DIM was modelled using the Ali–Schaeffer lactation curve (Ali and Schaeffer, 1987) nested within parity and litter size for DMY. Age at lambing was modelled as linear regression nested within parity. The random part of the model included flock–test–day, permanent environmental effect within lactations, direct additive genetic effect (modelled by combining pedigree and genomic relationships), and residual. The same model was used to predict BVs. The matrix notation of the model was:

$$y = Xb + W_c c + W_p p + W_a a + e \quad (1)$$

where: y is a vector of the phenotypic observations for DMY, X is an incidence matrix for the fixed effects; W_c, W_p and W_a are incidence matrices for the flock-test-day, permanent environmental, and additive genetic effects, respectively; b is a vector of unknown fixed effects; c, p and a are vectors of unknown random effects; e is a vector of residuals. Pedigree additive relationship was combined with genomic information following the theory of ssGBLUP (Aguilar *et al.*, 2010). The inverse of numerator relationship matrix A^{-1} in the mixed model equation was replaced by matrix H^{-1} as follows:

$$H^{-1} = A^{-1} + \begin{bmatrix} 0 & 0 \\ 0 & G^{-1} - A_{22}^{-1} \end{bmatrix} \quad (2)$$

Partitioning of genetic trends. Genetic trends obtained as the average BV by year of birth were estimated for animals born from 2014 to 2018. The BV of an individual can be partitioned as $a_i = 1/2a_s + 1/2a_d + w_i$ where a_i is the BV of the individual, a_s and a_d are BVs of sire and dam, and w_i is the Mendelian sampling term. These partitions of BVs can be allocated into group-specific contributions to the change in genetic mean (Garcia-Cortes *et al.*, 2008). The â was partitioned by the gender and flock using the package AlphaPart 0.8.1. (Obšteter *et al.*, 2021) in R (R Core Team, 2020) following the method proposed by Garcia-Cortes *et al.* (2008) as follows:

$$\hat{a} = \boldsymbol{TP_1T^{-1}}\hat{a} + \boldsymbol{TP_2T^{-1}}\hat{a} + \cdots + \boldsymbol{TP_kT^{-1}}\hat{a} \quad (3)$$

where matrix T traces the flow of genes from one generation to the other, i.e. it describes the expected genetic contribution of genes identical by descent between each individual and all its ancestors, while P_i matrix is the 'path' matrix for the i^{th} origin that selects a partition of animals' BV attributed to selection work performed by the i^{th} origin.

Results

Regression of BVs on birth year revealed a negligible positive (0.005 kg per year) overall genetic trend over the period studied. However, the gain was inconsistent throughout the period with significant decreases in 2016 for both genders and in 2018 for ewes (Figure 1a). Decomposition of genetic trend by gender revealed that rams had greater contribution in selection gain than ewes throughout the period, especially in the last examined year where discrepancy was the most pronounced. When BVs were regressed on birth year within gender, the estimated slopes for rams and ewes were 0.007 and 0.005 kg, respectively. Because of the high oscillations in estimated BVs between adjacent years, these regressions were estimated with a relatively high standard errors and were statistically insignificant (P>0.05). However, due to the similar number of animals throughout the years, these regressions are still to some extent informative in general

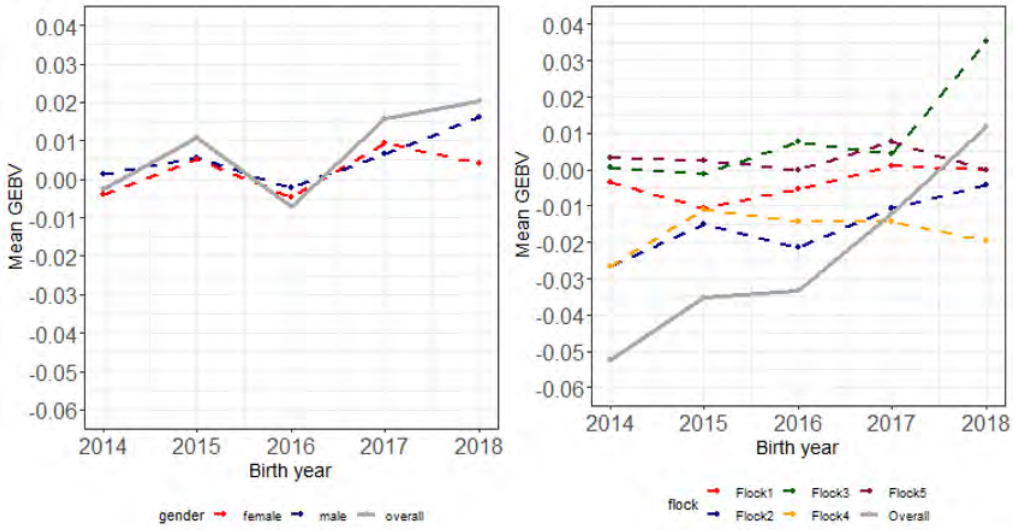


Figure 1. (A) Decomposition of overall genetic trends by gender and (B) flock for DMY.

description of genetic trends. The Figure 1b. shows decomposition of genetic trend by flock for the most informative and most reliable part of this population (flocks with >50 animals). The estimated annual rate for this subpopulation was 0.01435 kg with inconclusive contribution of the flocks' genetics to the overall selection gain.

Discussion

Analysis of genetic trends is important to monitor the success of past breeding activities and to carefully plan future breeding actions. The main intention of this study was to examine recent genetic trends in the Istrian sheep with special attention to determining the main contributors to the genetic trend. We used all the available phenotypic, pedigree, and genomic information to estimate genetic trends. The obtained results suggest that selection had a negligible positive overall trend, but oscillations in average BVs between years suggest that selection was not systematically conducted. The BVs used in this analysis were post-hoc obtained (after selection had been done), so it is ungrateful to derive firm conclusions on this issue. By not having the same information on genetic merits of the animals (potential disproportion of current and past BVs), as well as information which breeders selected based on EBVs, we can only hypothesise that selection on many of the examined flocks was 'happening by chance' rather than following the estimated BVs. In addition, partition of genetic trend by gender indicates that breeders in this population paid more attention to the selection of rams. There are many indications that rams are selected (based on different criteria among the breeders), while ewes are often 'recruited' by chance (often the ones not slaughtered for meat). According to the national breeding programme (Mioč *et al.*, 2011), animals should be selected based on the BVs, but since this is not obligatory, it is obviously sporadically applied. Decomposition of genetic trend by flock was inconclusive because none of the flocks studied had positive trend throughout the study period. Inconsistent genetic trends among flocks can be partially explained by differences in breeders selection criteria, but also as a consequence of limited genetic connectedness between the flocks. It is well known that many sheep populations, including this one, suffer from poor connectedness which leads to inability of the BLUP procedure to disentangle average genetic level from flock effects. Implementation of genomic information in genetic evaluation can alleviate this issue to some extent (Yu *et al.*, 2017) and therefore the results obtained probably nicely reflect true genetic trends in this population. A great precaution is needed

in ranking the flocks according to their estimated genetic merit, thus the obtained results should serve the best on within-flock level. The initial steps towards genomic selection in this population have been taken and continuous genotyping should additionally enlarge the reference population, likely leading to even more accurate predictions. To conclude, the results obtained indicate a very slow and inconsistent genetic gain for DMY and additional efforts need to be made to make it faster and more consistent. Breeders should be more decidedly advised to select replacements based on the estimated BVs. High accuracy of genomic BVs should serve as a strong argument in this regard. The connection between advisors and breeders should be more firm to adopt and carry out the proposed selection methods in order to bring this breed much closer to the world-famous dairy sheep breeds.

Funding

This research has been fully supported by the Croatian Science Foundation under the project Genomic characterization, preservation, and optimum contribution selection of Croatian dairy sheep (OPTI-SHEEP), grant number IP-2019-04-3559.

References

Aguilar I., Misztal I., Johnson D.L., Legarra A., Tsuruta S. *et al.* (2010) J. Dairy Sci. 93(2):743-752. https://doi.org/10.3168/jds.2009-2730.

Ali T.E., and Schaeffer L.R. (1987) Can. J. Anim. Sci. 67(3):637-644. https://doi.org/10.4141/cjas87-067.

Blair H.T., and Pollak E.J. (1984) Anim. Sci. J. 58(4):878-886. https://doi.org/10.2527/jas1984.584878x.

Duchemin S.I., Colombani C., Legarra A., Baloche G., Larroque H. *et al.* (2012) J. Dairy Sci. 95(5): 2723-2733. https://doi.org/10.3168/jds.2011-4980.

García-Cortés L.A., Martínez-Ávila J.C., and Toro M.A. (2008) Animal 2(6):821-824. https://doi.org/10.1017/S175173110800205X.

Gorjanc G., Potočnik K., García-Cortés L.A., Jakobsen J., and Durr J. (2011) Proc. of the 2011 Interbull Meeting, Stavanger, Norway.

International Committee for Animal Recording (ICAR) (2018) Section 16. Guidelines for Performance Recording in Dairy Sheep and Dairy Goats. Available at: https://www.icar.org/index.php/icar-recording-guidelines/.

Legarra A., Baloche G., Barillet F., Astruc J.M., Soulas C. *et al.* (2014) J. Dairy Sci. 97(5):3200-3212. http://doi.org/10.3168/jds.2013-7745.

Mioč B., Pavić V., Barać Z., Vnučec I., Prpić Z. *et al.* (2011) Program uzgoja ovaca u Republici Hrvatskoj. Hrvatski savez uzgajivača ovaca i koza, Zagreb.

Misztal I., Tsuruta S., Lourenco D.A.L., Masuda Y., Aguilar I. *et al.* (2018) Manual for BLUPF90 family of programs. Athens: University of Georgia.

Obšteter J., Holl J., Hickey J.M., and Gorjanc G. (2021) Genet. Sel. Evol. 53:30. https://doi.org/10.1186/s12711-021-00600-x.

R Core Team (2020) R: A Language and Environment for Statistical Computing; R Foundation for Statistical Computing: Vienna, Austria. Available at: https://www.R-project.org.

Škorput D., Gorjanc G., Kasap A., and Luković Z. (2015) Animal 9(10):1605-1609. http://doi.org/10.1017/S1751731115001056.

Špehar M., Ivkić Z., Bulić V., Barać Z., and Gorjanc G. (2011) Agric. Conspec. Sci. 76(4):1-4.

Yu H., Spangler M.L., Lewis R.M., Morota, and Gorjanc G. (2017) G3: Genes, Genomes, Genetics 7(10):3543-3556. http://doi.org/10.1534/g3.117.300151.

721. Comparison of GEBV accuracies predicted using the pedigree matrix of female and male potential reference populations

A. Carta*, M.G. Usai, S. Casu and S. Salaris

Agris Sardegna, Genetics and Biotecnology, Località Bonassai, 07100 Sassari, Italy; acarta@agrisricerca.it

Abstract

We propose a procedure to predict the accuracies of GEBV of selection candidate rams which uses the pedigree matrix to estimate the amount of information on relatives in a female (4,486 ewes) and a male (3,074 Herd Book sires) potential reference populations. Two predictive functions were obtained fitting the estimates of the amount of information on relatives to the GEBV accuracies of a sample of genotyped males. The two functions provide good predictions (R^2: 0.94 and 0.73). The predictive functions were applied to 677 ungenotyped selection candidate rams. The predicted accuracies for the male reference population show higher average (0.55 vs 0.46) and higher variability (sd: 0.044 vs 0.020). We showed that the traditional approach for genomic selection in Sarda breed is less effective than in dairy cattle and other dairy sheep breeds and that the female nucleus provides accurate GEBV for a portion of well genetically connected selection candidates.

Introduction

Over the past decade, genomic selection has become the reference approach for the genetic improvement of dairy livestock (Wiggans *et al.*, 2017). In dairy cattle, relative high accuracies of GEBV of young bulls are due to the consistency of male reference populations that include progeny tested bulls with large numbers of daughters. Moreover, GEBV of selection candidates can be calculated for several productive and functional traits that have been recorded on a large number of genotyped animals or their progeny. Although there are some exceptions (Baloche *et al.*, 2014), in dairy sheep the situation is generally worse (Carta *et al.*, 2009). First, the artificial insemination (AI) is less efficient and not widely used which makes the male reference populations less powerful compared to dairy cattle. Second, the overall consistency of genetic evaluations is negatively affected by the relative high costs of performance recording which led to very simplified recording schemes. Third, the range of traits for which GEBV may be calculated is limited by the lack of records for genotyped animals or their progeny for most functional and health traits. In these conditions, the applications of genomic selection may take advantage of experimental female nuclei which can integrate the information content of male reference populations for traits routinely recorded in the Herd Book (HB) and serve as reference populations for traits costly to record on a large scale (Pryce *et al.*, 2018). The potential contribution of experimental female nuclei depends on the number of animals bred, the depth of historical records and the strength of the genetic links with the HB population. For the Sarda breed, a female nucleus, consisting of ewes generated by HB rams, has been pedigree recorded, genotyped and phenotyped for most functional, health and production traits for twenty years. However, the design of a breeding program with genomic selection that includes a female nucleus as reference population requires to appreciate its effectiveness by predicting the accuracy of GEBV (Dekker *et al.*, 2021). In this study, we propose a procedure to predict the expected accuracies of GEBV of selection candidate rams which uses the pedigree matrix to estimate the amount of information on relatives in the reference populations. The final aim was to compare the potential effectiveness as reference population for genomic selection of the Sarda female nucleus and a hypothetical population of HB sires before starting their genotyping.

Materials & methods

The female nucleus proposed as reference population (FRP) was created in 2000 for a QTL detection experiment. It consists of approximately 800 milking ewes with a replacement rate of about 25% generated

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_721, © A. Carta *et al.* 2022

by mating adult ewes with rams coming from the HB of the Sarda breed. Ewes of FRP are raised in an experimental farm that has a typical Sardinian dairy sheep farming system. Ewes are routinely recorded with intense scheme for milk yield and 14,999 lactation records from 4,486 ewes born from 1999 to 2019 were available. All the ewes, 272 male ancestors and further 1,311 HB rams were genotyped with the Illumina Inc. OvineSNP50 Beadchip. SNP editing was performed on the 26 OAR autosomes using call rate and MAF thresholds of 95 and 1% respectively. After quality control, 43,390 SNP were retained.

First, GEBV of 1,561 genotyped HB rams were predicted using the FRP records with the model $y = \mu + Xb + Zg + Wp + e$; where y is the vector of the lactation records; μ is the overall mean; b is the vector of fixed effects; g is the vector of GEBV; p is the vector of random permanent environmental effects; e is the vector of random residuals and X, Z and W are incidence matrices. Vectors p, g and e were assumed distributed as $p \sim N(0, I\sigma_p^2)$, $g \sim N(0, G\sigma_g^2)$ and $e \sim N(0, I\sigma_e^2)$, where G is a genomic relationship matrix including FRP ewes and HB rams (Van Raden, 2008) and σ_p^2, σ_g^2 and σ_e^2 are the corresponding variances.

An alternative GEBV prediction was performed by using as reference population 995 genotyped rams with at least two milk-recorded daughters in HB and using their DYD calculated basing on 173,147 lactations from 70,981 ewes (on average 71 daughters per sire with a maximum of 929). DYD were weighted by $\{1 - [SE_i^2/(\sigma_s^2)]\}^{1/2}$; where SE and σ^2_s are the approximated standard errors and the sire variance estimated by a model including the ram as random effect. Then, GEBV of the 566 genotyped rams with no daughters in HB were predicted with the model $y = \mu+Zg+e$; where y is the vector of DYD and μ, e and Z are as above and g is the vector of GEBV distributed as $\sim N(0, G\sigma_g^2)$ where G is a genomic relationships matrix including the 1,561 genotyped HB rams.

Both the analyses were carried out using the airemlf90 software (Misztal *et al.*, 2002). For both the models the expected GEBV accuracies (r_{GEBV}) were calculated as $\{1 - [SE_i^2 / (G_{ii}\times\sigma_g^2)]\}^{1/2}$ where SE is the standard error, G_{ii} is the diagonal element of the G matrix and σ_g^2 is the model-specific genetic variance. The amount of information on relatives (IR) of a selection candidate on the two reference populations was estimated using the pedigree relationships matrix A including all the available pedigree records in HB of the 6,047 individuals (4,486 ewes and 1,561 rams) as $IR_{i=}[(a_{ir}'A_{rr}^{-1} a_{ir})/ a_{ii}]^{1/2}$, where A_{rr} are the sub-matrices including the relationships between the animals of FRP or the 995 sires of daughters in HB, a_{ir} are the vectors of relationships between the selection candidate i and FRP or the 995 sires of daughters in HB, a_{ii} is the diagonal element of the relationships matrix corresponding to the selection candidate i. Then, for each reference population a specific function of IR fitting to r_{GEBV} was identified. The next step was to compare the expected accuracies predicted by the two functions (Pr_{GEBV}) of the 677 ungenotyped selection candidates that will start breeding in the next reproduction season. Finally, a hypothetical reference population of males (MRP) was formed by all the sires born from 2010 to 2019 with at least two recorded daughters in HB. Finally, 3,074 rams were included in MRP. The potential impact of the two genomic selection approaches was discussed focusing on the distributions of the 677 Pr_{GEBV} obtained by the specific predictive functions in both FRP and MRP.

Results

Figure 1 shows the pattern of r_{GEBV} with IR for the 1,561 validation rams evaluated using FRP as reference population. The values of r_{GEBV} ranged from 0.28 to 0.94 (0.53 on average). The values of IR ranged from 0.00 to 0.99 (0.43 on average). The large variability of r_{GEBV} and IR is related to the different origins of the genotyped rams. Indeed, 272 out of them are ancestors of the FRP ewes and showed the highest r_{GEBV}. On the other hand, most of the other selection candidates recently genotyped come from flocks not genetically linked with FRP due to the scarce aptitude of farmers to use external rams or participate to the AI programs.

Figure 2 shows the pattern of r_{GEBV} with IR for the 556 validation rams evaluated using the 995 sires of daughters in HB as reference population. The values of r_{GEBV} ranged from 0.25 to 0.76 (0.49 on average). The values of IR ranged from 0.19 to 0.83 (0.56 on average). Polynomial functions of IR were fitted to r_{GEBV} for FRP and the male reference data set. Both the functions provide good predictions of r_{GEBV} with R^2 of 0.94 and 0.73 respectively. The lower fitting ability of the function applied to the 556 rams may be due to the high percentage of false pedigree records in HB where the ascertainment of paternity through DNA is not routinely realized. Nevertheless, the predictive function accomplished the aim of comparing the distributions of Pr_{GEBV} of the HB selection candidates rather than predicting accurately the single r_{GEBV}.

In Figure 3 the distributions of IR for the 677 selection candidates are reported for both FRP and MRP. As expected, IR was on average higher in MRP than FRP (0.66 vs 0.31) but with lower variability (sd: 0.054 vs 0.064). Indeed, most selection candidates have male ancestors in MRP whereas, at present, only those coming from the AI center or from flocks that have continuously participated to the AI programs have relatives in FRP. Figure 4 shows the distributions of Pr_{GEBV} of the 677 selection candidates for both FRP and MRP. The values for MRP showed a higher average (0.55 vs 0.46) and higher variability (sd: 0.044 vs 0.020).

Discussion

Our results confirm that IR in potential reference populations can be used to predict GEBV accuracies and to compare different scenarios of breeding programs with genomic selection. It has to be noted that the predictive functions applied in this study are population-specific and should be re-estimated in different populations. Although we hypothesized 10 years of routine genotyping of the HB male

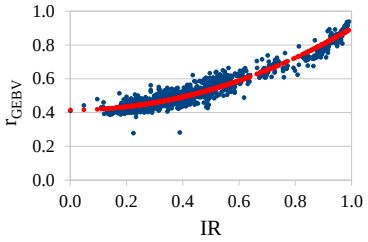


Figure 1. Relationship between expected accuracies (r_{GEBV}) and amount of information on relatives (IR) in FRP. In red the pattern of the predictive function.

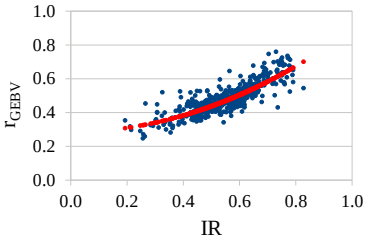


Figure 2. Relationship between expected accuracies (r_{GEBV}) and amount of information on relatives (IR) in 995 sires of daughters in HB as reference population. In red the pattern of the predictive function.

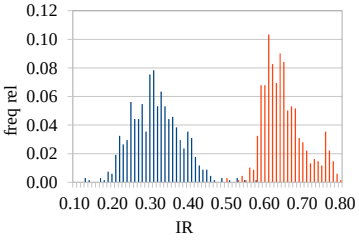


Figure 3. Distribution of amount of information on relatives (IR) of the first service rams candidates estimated from FRP (in blue) and MRP (in red) reference population.

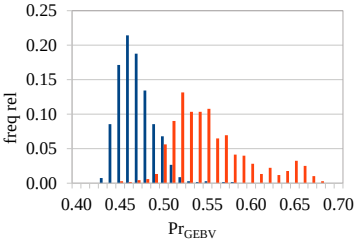


Figure 4. Distribution of expected accuracies (Pr_{GEBV}) predicted from the amount of information on relatives of the first service rams candidates in FRP (in blue) and MRP (in red) reference population.

population, the average Pr_{GEBV} obtained for MRP is not particularly promising for the application of a traditional genomic selection approach based on genotyping the HB rams. The average Pr_{GEBV} obtained for FRP shows that the current genetic structure of FRP can realize sufficient accuracies of GEBV only for few selection candidates well connected with the nucleus. However, it should be considered that the effects of pedigree errors in HB and the higher accuracies of performance records in FRP were not considered in our study. From a practical point of view, we showed that the traditional approach for genomic selection in Sarda breed is less effective than in dairy cattle and other dairy sheep breeds with more structured selection schemes. For traits routinely measured in HB, FRP may be useful to start genomic selection while planning and realizing the genotyping of HB rams. Successively, it could integrate the information content of the male reference population along with the progressive piling up of genotypings of the HB rams. In this perspective, the availability of single-step Genomic BLUP methodology is fundamental to develop models able to jointly exploit information from male and female reference populations (Lourenco *et al.*, 2020). For traits too costly to measure in the HB, FRP can currently provide accurate GEBV at least for the breeding rams well genetically connected. However, the number of breeding rams with sufficiently accurate GEBV can be increased by using the predictions of the expected accuracies to set up the size of the nucleus and the flow of animals between HB and FRP.

References

Baloche G., Legarra A, Sallé G., Larroque H., Astruc J.M. *et al.* (2014). J. Dairy Sci.,97(2), 1107-1116. http://dx.doi.org/10.3168/jds.2013-7135

Carta A., Casu S., Salaris S. (2009). J. Dairy Sci., 92, 5814-5833. https://doi.org/10.3168/jds.2009-2479

Dekkers J.C.M, Su H., and. Cheng, J. (2021). Genet Sel Evol 53, 55. https://doi.org/10.1186/s12711-021-00647-w

Lourenco D., Legarra A., Tsuruta S., Masuda Y., Aguilar I. *et al.* (2020) Genes, 11, 790; https://doi.org/10.3390/genes11070790

Misztal, I., Tsuruta, S., Strabel, T., Auvray, B., Druet, T., *et al.* (2002) Proc. of the 7th WCGALP, Montpellier, France

Pryce J.E., Nguyen T.T.T., Axford M., Nieuwhof G., and Shaffer M. (2018). J. Dairy Sci. 101:3702–3713. https://doi.org/10.3168/jds.2017-13377

Wiggans G.R., Cole J.B., Hubbard S.M., and Sonstegard T.S.. (2017). Annu. Rev. Anim. Biosci. 5:309-327. https://doi.org/10.1146/annurev-animal-021815-111422

722. Quantification of the effect of incomplete reproduction data on the accuracy of selection

W.J. Olivier

Grootfontein Agricultural Development Institute, Private Bag X529, Middelburg (EC), 5900, South Africa; willemo@dalrrd.gov.za

Abstract

The aim of this study was to quantify the effect of incomplete reproduction data on the accuracy of estimated breeding values. The trait included in the analysis was total weight of lamb weaned per lambing opportunity (TWW). The complete dataset for reproduction traits of the Grootfontein Merino Stud was used in this study and subsets of the data were created by excluding or including the reproduction records of ewes. This was done to simulate different levels of data completeness. It is clear from the results that the omission of mating and lambing data will result in an increase of the mean of TWW, as well as a decrease in the variation available for selection. It is therefore of the utmost importance that all relevant mating, lambing and weaning data must be recorded for the evaluation of reproduction traits and the improvement of flock productivity.

Introduction

It is a well-established fact that incomplete phenotypic or pedigree data will reduce the accuracy of BLUP models (Henderson, 1973), which will subsequently lead to bias in the estimated breeding values (EBV; Stock and Distl, 2010). Using these EBV in selection programs will have a negative impact on the genetic gain of the population which will subsequently result in the decrease of the profitability of the enterprise, making all the effort and time worthless (Id-Lahoucine and Casellas, 2017).

Reproduction is the most important trait for small stock enterprises in South Africa to be considered during the selection of replacement rams and ewes. Accurate selection for reproduction is challenging because of the low heritability of these traits, as well as the fact that it is a sex limited trait that is expressed after selection was performed (Olivier *et al.*, 1998). On top of this, genetic progress could even further be hampered by incomplete and inaccurate reproduction data.

The lack of measurements for production traits are seen as a missing value in the estimation of EBV and will probably not have a big influence on the EBV of related animals (Olivier, 2018). However, for reproduction traits, it is important to note that a ewe that did not lambed or wean a lamb should not be regarded as missing data and be excluded from the analysis. The completeness of reproduction data can be defined as all data with regard to ewes that weaned a lamb, ewes that lambed that did not wean a lamb and ewes that were mated but did not lambed. Omittance of the mating and lambing information can therefore be a major constraint to the usefulness of reproduction data. The aim of this study was therefore to quantify the effect of incomplete reproduction data on the accuracy of estimated breeding values.

Materials & methods

Data collected on the Grootfontein Merino stud (GMS) from 1968 to 2017 were used for this analysis. A detailed description of the management and selection of the stud was done by Nemutandani (2017).

The trait included in the analysis was total weight of lamb weaned per lambing opportunity (TWW). The complete dataset for reproduction traits of the GMS was used in this study, which included data from all the ewes that were born since 1968 in the stud. Subsets of the data were created by excluding or including

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_722, © W.J. Olivier 2022

the reproduction records of ewes. This was done to simulate different levels of data completeness. The first subset (TR – all mating, lambing and weaning data) included data from all the ewes that were mated, lambed or did not lamb and weaned or did not weaned a lamb/s for each lambing opportunity. The second subset (TL – lambing and weaning data) included the data of the ewes that weaned at least one lamb of the litter, as well as the data of ewes that lambed but did not wean a lamb for each lambing opportunity. The third subset (TW – only weaning data) include only the data of the ewes that weaned at least one lamb of a litter of each lambing opportunity.

The Proc MEANS-procedure of SAS (SAS, 2016) was used to obtain the descriptive statistics of the datasets. The estimation of the genetic parameters for TWW of the three subsets, as well as the EBVs was done with ASReml 4.0 program (Gilmour *et al.*, 2014) by fitting a repeatability model. In order to quantify the effect of data completeness, the Spearman rank correlation under the PROC CORR-procedure of SAS (SAS, 2016) was used to estimate the correlations between the ranks of these EBVs for TWW of all three subsets.

Furthermore, lists were compiled of the Top 10% and Top 20% of the animals having the highest EBVs for TWW, estimated in the different data subsets. These lists were compiled for all animals in the pedigree file. The number and percentage of animals common to the different top lists were evaluated among the subsets.

Results

The description of the data for the different subsets is summarised in Table 1. It is clear that there was a clear increase in the number of lambs born (NLB) and weaned (NLW) and total weight of lamb weaned (TWW) with a decrease in completeness of the data. The increases in the means of NLB, NLW and TWW are clearly linked to the exclusion of information of ewes that did not lamb or wean a lamb. An important part in the success of selection for reproduction is the available variation in these traits. It is evident from Table 1 that there is a substantial decrease in the coefficient of variation (CV) of the traits as the completeness of the data decreases.

The heritabilities of the TWW were low and ranged from 0.03±0.01 to 0.06±0.01 and increased with an increase in the completeness of the data. These values are in correspondence with values summarised in the

Table 1. Data description of different data subsets (TR – All mating, lambing and weaning data; TL – Lambing and weaning data; TW – Weaning data) used for this analysis.

Reproduction data	Traits	Number of records	Mean	CV[1] (%)
Subset 1 – TR include: all mating, lambing and weaning data	Lambing opportunities	3,866	3.34	48.09
	Number of lambs born	3,866	1.2	45.27
	Number of lambs weaned	3,866	0.92	58.03
	Total weight of lambed weaned	3,866	24.52	58.09
Subset 2 – TL include: all lambing and weaning data	Lambing opportunities	3,565	2.93	49.91
	Number of lambs born	3,565	1.55	25.43
	Number of lambs weaned	3,565	1.18	42.45
	Total weight of lambed weaned	3,565	31.28	41.83
Subset 3 – TW include: all weaning data	Lambing opportunities	3,340	2.72	51.27
	Number of lambs born	-	-	-
	Number of lambs weaned	3,340	1.39	26.53
	Total weight of lambed weaned	3,340	36.97	25.35

[1] Coefficient of variation.

literature (Olivier, 2014). The estimates for c^2_{pe} were very low and ranged from 0.02±0.01 to 0.04±0.01 and are at the lower end of the range of 0.00 to 0.13 summarised by Olivier (2014).

The average and range of repeatability model EBV and accuracies of the EBVs for total weight of lamb weaned (TWW) for the respective subsets are summarised in Table 2. It is evident from Table 2 that the top end of the EBVs decreased with 2.15 kg as the variation in TWW decreased because of the unavailability of mating and lambing data. However, the lower boundary of the TWW EBVs increased substantially with 4.73 kg as the available variation decreased.

These changes in the upper and lower boundaries resulted in a decrease in the EBV range between TR and TL of 24%, between TR and TW of 59% and between TL and TW of 28%. These changes suggest that EBVs of the better reproductive ewes will be effected negatively. However, the biggest concern is the fact that the EBVs of the ewes with poorer reproduction will increase. It is also evident that there were a decrease of in the mean accuracies between the TWW EBVs of TR and TL of 8% and between TR and TW of 15%.

The Spearman rank correlations between TR and TL was 0.84, between TR and TW 0.63 and TL and TW 0.75. According to these values it is evident that the unavailability of mating and lambing data will result in a major re-ranking of the ewes, especially when only weaning data is included.

The number and proportion of common animals in the Top 10% and Top 20% lists when all animals in the pedigree were included are presented in Table 3 for TWW EBV estimated for TR, TL and TW. The highest percentage common animals in Table 3 were in the Top 20% lists between TR and TL, as well as between TL and TW, with just more than 70% common animals. The lowest percentage of common animals in both the Top 10% and Top 20% lists were 54.26 and 57.24% between TR and TW respectively, which could be expected due to difference in the level of completeness of the data.

Table 2. The average of repeatability model estimated breeding values (EBVs; Min-Max) and accuracies of the EBVs (Accuracy; Min-Max) for total weight of lamb weaned for the respective subsets.[1]

Traits	EBV	Accuracies
TR (Total reproduction)	0.54 (-8.92 to 9.53)	51.79 (23.33-67.23)
TL (Lambing data)	0.42 (-5.80 to 9.06)	47.96 (0.18-65.48)
TW (Weaning data)	0.23 (-4.19 to 7.38)	45.19 (0.13-64.14)

[1] TR = all mating, lambing and weaning data; TL = lambing and weaning data; TW = weaning data.

Table 3. The number and proportion of common animals in the Top 20% and Top 10% lists for total weight of lamb weaned (TWW) for the three subsets.[1]

Subsets	Top 10% list (n=387)		Top 20% list (n=774)	
	Number	%	Number	%
TR vs TL	259	66.93	556	71.83
TR vs TW	210	54.26	443	57.24
TL vs TW	256	66.15	549	70.93

[1] TR = all mating, lambing and weaning data; TL = lambing and weaning data; TW = weaning data.

Discussion

It is evident from the results of this study that the completeness and accuracy in the collection of reproduction data will have a substantial impact on the improvement in total weight of lamb weaned per generation through selection. It is clear from the results that the omission of mating and lambing data will result in an increase of the mean of TWW, as well as a decrease in the variation available for selection. This decrease in the phenotypic variation was also reflected in the decrease of h^2_a, as well as the decrease in the EBV ranges. It is therefore of the utmost importance that all relevant mating, lambing and weaning data must be recorded for the evaluation of reproduction traits and the improvement of flock productivity.

References

Gilmour, A.R., Gogel, B.J., Cullis, B.R., Welham, S.J. and Thompson, R. (2014) VSN International Ltd, Hemel Hempstead, HP1 1ES, UK. www.vsni.co.uk.

Henderson, C.R. (1973). Proc. Anim. Breeding Genet. Symp. In honor of Dr Jay L. Lush. Am. Soc. Anim. Sci. Champaign, IL.

Id-Lahoucine, S. and Casellas, J. (2017) Livest. Sci., 198:76-81. https://doi.org/10.1016/j.livsci.2017.02.011

Nemutandani, K.R., 2017. MScAgric Thesis. University of Pretoria. http://hdl.handle.net/2263/64181

Olivier, W.J. (2014) PhD thesis, University of Stellenbosch. http://hdl.handle.net/10019.1/86675

Olivier, W.J. (2018) Grootfontein Agric, 18(1):33-39. http://gadi.agric.za/Agric/Vol18No1_2018/Olivier%202018%20Effect%20of%20culling.pdf

Olivier, W.J., Snyman, M.A., Van Wyk, J.B. and Erasmus, G.J. (1998) Livest. Prod. Sci. 56:71-77. https://doi.org/10.1016/S0301-6226(98)00131-6

SAS, 2016. SAS Online Doc, Version 9.2. SAS Institute Inc. Cary, N.C., USA.

Stock, K.F. and Distl, O. (2010) J. Anim. Breed. Genet. 127:42-52. https://doi.org/10.1111/j.1439-0388.2009.00809.x

723. Using random regression models to model growth in sheep with irregular data

E. Mutch[1*], M. Coffey[1], G. Simm[2] and R. Mrode[1]

[1]*Scotland's Rural College, Roslin Building Institute Bush Estate, Midlothian, EH25 9RG, United Kingdom;* [2]*Global Academy of Agriculture & Food Security, University of Edinburgh, The Royal (Dick) School of Veterinary Studies and The Roslin Institute, Easter Bush Campus, Midlothian, EH25 9RG, United Kingdom;* *emutch@sruc.ac.uk*

Abstract

Random Regression is increasingly popular as a method to model longitudinal traits, such as growth, in animal populations. However, these models perform optimally when some specific criteria are met such as the availability of sufficient records at the beginning and end of the age distribution to avoid 'edge' effects. Also, the performance of these models with sparse data is not well known. This research investigates the use of random regression models with data for Scottish Blackface (SBF) and mixed breed sheep, where data is rather sparse. Polynomial regressions and Splines presented a good fit at the phenotypic level. Estimates of heritabilities and genetic correlations for SBF varied from medium to high. This study will help expand the functions used to model growth in sheep and allow for more flexible phenotype recording strategies.

Introduction

Breeding animals for desirable traits is a practice as old as domestication. One of current animal breeding's top priorities is creating sustainable and efficient food supply for the growing human population. Thus, animal breeding uses genetic analysis to identify and breed animals that have the genetic potential to grow quickly. Accurately modelling growth curves is an important way to understand how quickly and efficiently animals grow. With the advent of modern statistical genetics, genetic models can be used to precisely identify animals with superior genetic effects and then select those animals for breeding. Over several generations of breeding this benefit can be disseminated population wide (Hermesch *et al.*, 2003, Kirkpatrick and Heckman 1989). There are several ways to model growth and select for superior genetic potential. The random regression model has been used to accurately represent the entire dynamic process of growth across an animal's life. The basic structure of the random regression model remains generally constant across various applications (Schaeffer 2004).

The structure is: $Y_{ijkn:t} = F_i + g(t)_j + r(a,x,m1)_k + r(pe,x,m2)_k + e_{ijkn:t}$ where Y is the n^{th} observation of the k^{th} animal at time t belonging to the i^{th} fixed factors and the j^{th} group. *Fi* is a fixed effect. The next variable, $g(t)$ j is a function that accounts for the phenotypic trajectory of the average observations across all animals belonging to the j^{th} group. The next term $r(a,x,m1)k$ is the notation adopted for the random regression function. The next variable is another regression coefficient for the permanent environmental effect for the k^{th} animal. Lastly e is the random residual effect (Schaeffer 2004).

Random regression models represent growth as a function rather than a finite set of measurements, which allows it to account for variability in the rate of growth and variability in environmental effects throughout an individual's life. However, random regression models can be difficult to fit to real world data which may be unbalanced or sparse. Firstly, it cannot accommodate non-linear functions to represent the data. Although generally traits such as growth curves can be modelled with a linear function, a specific dataset can have variable rates of change or be highly correlated. It is difficult to validate the assumption that the data would necessarily follow a linear relationship. The second key issue is the number of records required for the model to converge. Random regression models are often used on lactation data in dairy cattle where

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_723, © E. Mutch *et al.* 2022

test day records follow a well-defined pattern compared to growth data; especially in sheep where data may be sparse. Thus, this research examines the application of random regression models when the data is sparse, highly correlated, or there are large gaps in age between weight recordings for an individual animal.

Materials & methods

The project was conducted with two datasets. The first was provided by the Agriculture and Horticulture Development Board. It consisted of mixed breed animals included in the RamCompare (RC) project which sought to identify the genetic merit of key sires currently used in UK sheep breeding. The dataset consisted of approximately 74,124 individuals with a total of 174,815 weight records. This dataset included three possible weight recordings with an accompanying age for each individual sheep. The second dataset was records from an SRUC SBF research flock and included 23,256. The datasets were also split into two versions. The first version of the dataset included weight records across the entire age distribution while the second included weight records that were taken from 1 to 200 days of age. The first version of the RC data included 174,815 records and the second version included 120,897 records. The first version of the SBF data included 23,256 records and the second version included 18,595 records. Three generation pedigrees were used.

Phenotypic curves. Initially, several curves were fitted at the phenotypic level to determine how several functions fit the data before including the genetic component. The datasets were analyzed with polynomial regression models of several orders (1 to 6) with weight fitted as a function of age using excel. Also spline functions were fitted in R using the 'stats' package and the 'rm. smooth' function in R studio. These methods were applied to the cut version of the dataset and the full version of the dataset when the latter was possible. The quantity and size of the full datasets were, at times, impossible to use when they exceeded the computation capacities of excel or R studio.

Genetic analysis. A random regression model consisting of the fixed effects: birth year, sex, genetic line, birth rearing rank, dam's age, grazing code, a fixed regression on age, month of recording, a combination of year month and day of recording, season of recording, and a combination of year and season of weight recording. A random animal effect with Legendre polynomial (LP) of order 1 and a permanent environmental (PE) effect were fitted to the full data set for SBF using ASREML (Gilmour *et al*., 2006). A random regression model including the fixed effects: breed, flock id, sex, birth type, and a fixed regression on age. A random animal effect with Legendre polynomial of order 1 and a PE effect were fitted for the cut data set for RC data using ASREML. Heritabilities and genetic correlations among several weights at different ages were generated from the analysis. It was observed that correlation between the intercept and slope for the LP parameters was close to the parameter boundary at 0.99 and this yielded genetic parameters that were not biologically reasonable. Therefore this correlation was fixed at three different levels and the correlation with the most biologically reasonable genetic parameters was chosen. These correlations were 0.2, 0.5 and 0.7 for SBF and 0.1, 0.5 and 0.9 for RC.

Results

Phenotypic curves. A summary of the results for the various phenotypic curves fitted is presented in Table 1. The polynomial regression had R squared values above 0.80 for both the RC and SBF datasets. The R squared values for the SBF were slightly higher indicating a slightly better goodness of fit. Spline functions were selected based on R squared values with the best for RC containing 80 knots and the best for the SBF containing 100 knots. The Gompertz functions for both datasets converged and resulted in equations for both models. Results are in Table 1.

Table 1. Phenotypic model summary values.

	Polynomial regression R^2 values	Spline	Gompertz:
RamCompare	Order 2: 0.826	R^2: 0.615	Residual standard error:
	Order 3: 0.835	$P<0.001$	5.344 on 15,053 df
	Order 4: 0.837		
	Order 5: 0.837		Equation: $Y=43.12e^{-1.73*0.98\wedge t}$
	Order 6: 0.838		
Scottish Blackface	Order 2: 0.870	R^2: 0.942	Residual standard error:
	Order 3: 0.872	$P<0.001$	3.561 on 23,253 df
	Order 4: 0.877		
	Order 5: 0.877		Equation: $Y=45.52^{e(2.19*0.98\wedge t)}$
	Order 6: 0.879		

Genetic analysis. The model with fixed correlation of 0.5 between both components of LP was chosen for the SBF and a corresponding value of 0.7 for the RC based on the biological plausibility of the genetic parameters obtained. The estimates of heritabilities for weight at different ages, as well as genetic and phenotypic correlations among weights at several ages are shown in Table 2 for SBF Sheep and Table 3 for RC. Heritabilities increased with age to reasonably high estimates at 376 days for SBF and 197 days for RC. In general, genetic correlations were high between close ages and much lower between ages that were further apart. Phenotypic correlations also followed an expected pattern, being higher between adjacent ages and decreasing at weights which are far apart in age.

Discussion

The curves from the phenotypic analyses, indicates the feasibility of fitting polynomial regressions of several orders. The polynomial regressions in excel increased in R squared as the order of the polynomial regression increased which is a trend that must be balanced with the increasing degrees of freedom in the model. The spline functions had reasonably high R squared values with SBF having a higher R squared value indicating a better goodness of fit. However, at the genetic level such high number of knots will not be feasible. The Gompertz models had similar resulting equations suggesting that the models for both data would be similar, which may indicate a poor goodness of fit as other curves have shown differences between the data sets. In addition, Gompertz functions may not be easy to fit in random regression setting.

Table 2. Genetic (below diagonal) and Phenotypic (above diagonal) correlations among several ages in days with heritabilities (bold) on the diagonal from the random regression model for SBF.

Records	**5,153**	**11**	**150**	**191**	**10**	**20**	**31**	**2**	**3**
	1	30	60	120	180	200	280	330	376
1	**0.37**	0.36	0.35	0.32	0.28	0.27	0.21	0.18	0.14
30	1.00	**0.36**	0.35	0.32	0.29	0.28	0.23	0.21	0.18
60	0.98	0.99	**0.34**	0.33	0.30	0.30	0.41	0.24	0.21
120	0.91	0.94	0.97	**0.33**	0.32	0.32	0.31	0.29	0.28
180	0.79	0.84	0.90	0.97	**0.34**	0.34	0.35	0.35	0.34
200	0.74	0.80	0.86	0.95	1.00	**0.35**	0.36	0.36	0.36
280	0.54	0.62	0.70	0.85	0.95	0.97	**0.40**	0.42	0.43
330	0.43	0.51	0.60	0.77	0.90	0.93	0.99	**0.45**	0.47
376	0.34	0.43	0.52	0.70	0.85	0.88	0.97	1.00	**0.49**

Table 3: Genetic (below diagonal) and Phenotypic (above diagonal) correlations among several ages in days with heritabilities (bold) on the diagonal from the random regression model for RC Data.

Records	**32,397**	**35**	**1,832**	**250**	**25**	**1**	**1**
	1	30	60	120	170	191	197
1	**0.30**	0.28	0.25	0.18	0.11	0.09	0.08
30	0.98	**0.24**	0.25	0.2	0.16	0.14	0.14
60	0.91	0.97	**0.24**	0.23	0.2	0.19	0.19
120	0.62	0.76	0.89	**0.27**	0.29	0.29	0.29
170	0.35	0.53	0.71	0.95	**0.34**	0.36	0.36
191	0.26	0.44	0.64	0.92	1	**0.38**	0.39
197	0.24	0.42	0.62	0.91	0.99	1	**0.39**

The correlation between the intercept and slope of the LP being at the parameter boundary space accounted for the high heritabilities initially obtained for both breeds. However, examining this correlation at different levels helped to obtained reasonable genetic parameters. The impact of this on the rate of genetic progress will be examined. These results provide an initial basis to investigate further the impact of sparse data on the performance of random regression models. Although the birth weight heritability did not follow the general pattern of increasing over time, this is reasonable if birth weight is under different genetic control than growth at other points during a lamb's life. Further investigations on the use of covariance functions and more flexible functions such as the Spline are planned.

References

Gilmour, A.R., Gogel, B.J., Cullis, B.R., and Thompson, R. (2006) ASReml User Guide Release 2.0 VSN International Ltd, Hemel Hempstead, HP1 1ES, UK.

Hermesch S, Kanis E and Eissen JJ,(2003) *J. Animal Science* 81: 895–903. https://doi.org/10.2527/2003.814895x

Kirkpatrick M and Heckman N, (1989) *J. Math Biol* 27(4): 429–450. https://doi.org/10.1007/BF00290638

Schaeffer,LR,(2004) *Livest* 86:35-34. https://doi.org/10.1016/S0301-6226(03)00151-9.

724. Impact of the order of polynomials in random regression model on the accuracy of genetic evaluation for weight traits

U. Paneru[1*], D.J. Brown[2], N. Moghaddar[1] and J.H.J. van der Werf[1]

[1]School of Environment and Rural Science, University of New England, 2351, Armidale, Australia; [2]AGBU, a joint venture between NSW Department of Primary Industries and University of New England, 2351, Armidale, Australia; upaneru@myune.edu.au

Abstract

Random regression is suggested as a better model for genetic evaluation for a trait with multiple measurements. A random regression model with a higher order of polynomials is often selected as a more parsimonious model based on log-likelihood and information criteria. However, limited empirical studies assessed whether higher order of polynomials corresponds to better predictive ability. This study investigated the predictive ability of random regression models with different polynomials using forward prediction in Australian maternal sheep, using 141,056 repeated weight records from 33,564 animals. Results showed a random regression model with quadratic polynomials for all random effects was the best model, producing an average accuracy estimate of 0.72 and a slope of 0.96 in regressing phenotypes on estimated breeding values. A random regression model with a quadratic polynomial is suggested as the best alternative to model multiple weight records for genetic evaluation of sheep.

Introduction

Random regression (RR) is advocated as a method of choice to estimate genetic parameters and estimated breeding values (EBVs) for a trait with multiple records because it allows variance and covariance to vary across a trajectory of time (Fischer *et al.*, 2004; Schaeffer, 2004; Van der Werf *et al.*, 1998). Further, the RR model does not limit the number of records that need to be considered per animal, removes the need for age correction, and allows the genetic merit of animals to be estimated at any point in the trajectory (Meyer, 2001). Most RR analyses tend to select an optimal order of fit in the estimation of variance components, based on higher log-likelihood, lower Akaike Information Criteria (AIC), and Bayesian Information Criteria (BIC) (Fischer *et al.*, 2004; Meyer, 2001). However, the predictive ability of EBVs was seldom assessed to investigate whether a higher order of fit also corresponds to a better prediction of progeny performance. This study aimed to examine the accuracy of genetic evaluation from various RR models using weight records of Australian sheep.

Materials & methods

Data were taken from the LAMBPLAN dataset consisting of weight records of Australian and New Zealand sheep from maternal breeds (Brown *et al.*, 2007). Weights recorded from 61 to 525 days of age were analysed for animals born between 2009 and 2019, resulting in a data set of 141,056 records from 33,564 animals. A plot of the number of records and mean weights of animals across ages are presented in Figure 1.

Fixed effects. Mean weight was modelled as fixed cubic regression of body weight on a Legendre polynomial of age in days. Contemporary group included breed, flock, management group, sex, and day of measurement and was fitted with Legendre polynomial of order 1. Additional fixed effects included birth type and rear type fitted with a Legendre polynomial of order one and two, respectively and dam age with order zero based on a partial F test (Gilmour *et al.*, 2015). The same fixed effects were fitted for all models, to allow analysis to be directly comparable.

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_724, © U. Paneru *et al.* 2022

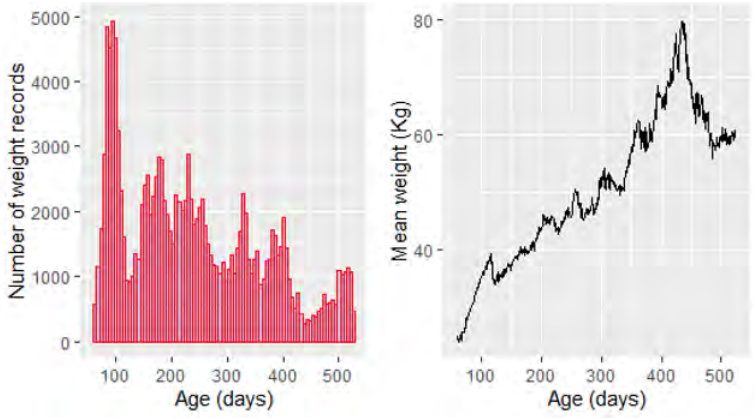


Figure 1. Histogram of number of weight records across age on the left, and mean weight of animals across age on the right.

Random effects fitted for RR. The RR models fitted random effects as Legendre Polynomials of age at recording (in days). A simple model was fitted first, with a first-order polynomial for all random effects, i.e. with effects being constant across ages. Higher-order polynomials for each random effect were added successively, and log-likelihood, lower AIC, and BIC values were compared. All analysis was done using ASReml 4.1 (Gilmour *et al.*, 2015).

The model in matrix notation is:

$$y = Xb + Z_4\alpha + Z_5\gamma + W_3\delta + W_4\rho + Z_4\eta + e$$

The structure of the expected values and covariance matrix was:

$$\mathbf{E}\begin{bmatrix} y \\ \alpha \\ \gamma \\ \delta \\ \rho \\ \eta \\ e \end{bmatrix} = \begin{bmatrix} X\beta \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}; \mathbf{V}\begin{bmatrix} \alpha \\ \gamma \\ \delta \\ \rho \\ \eta \\ e \end{bmatrix} = \begin{bmatrix} K_P \otimes A & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & K_Q \otimes I & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & K_R \otimes A & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & K_S \otimes I & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & K_T \otimes AQ & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & R \otimes I \end{bmatrix}$$

Where y is a vector of observations, β is a vector of fixed effects, α is a vector of additive genetic (breeding value) coefficients, γ is a vector of animal permanent environmental coefficients, δ is a vector of maternal genetic coefficients, ρ is a vector of maternal permanent environmental coefficients, η is a vector of genetic group coefficients, and e is a vector of random residuals. X, Z_4, Z_5, W_3, W_4, Z_4 are incidence matrices; K_P, K_Q, K_R, K_s and K_T are variance and covariance matrices between respective random regression coefficients; A is a relationship matrix, Q is breed proportion matrix, I is an identity matrix containing a number of levels for each effect and $\otimes$ is a Kronecker product, and R is a matrix of heterogeneous residual variance and covariance with eight classes from 61 to 525 days at 58 days interval.

Forward prediction. Forward cross-validation was performed to investigate the empirical predictive ability of the models and to observe how well the breeding values predict future animal's performance. The full data set was divided into two subsets: animals born from 2009 to 2016 were used as a training set, and animals born after 2016 were used as a validation set. EBVs of all animals were estimated using the training set. EBVs of the validation animals were compared to their phenotypic performance using two parameters: accuracy and slope. Performance of the validation animals was corrected for fixed effects at fixed ages for

each weight stage. The Pearson correlation coefficient between EBVs and adjusted phenotype was scaled by dividing it by the square root of the heritability to estimate the accuracy of EBVs. Slope was estimated by regression of adjusted phenotype on EBVs, with an expectation of slope one, i.e. no bias. A lower value indicates the under-dispersion of EBVs, while a higher value indicates over-dispersion.

Accuracy and slope estimates obtained from the above method are sensitive to the correction of the fixed effect phenotype. Therefore, accuracies and slopes estimates were also derived from the method of Legarra and Reverter (2018). With this method, accuracy was defined as the correlation between EBVs of the validation animals obtained from the partial set and the full set (true EBVs). Similarly, slope was estimated by regressing EBVs estimated from the full set (true EBVs) on EBVs estimated from the partial set.

Results

Log-likelihood and Aikaike's and Bayesian information criteria for all models are presented in Table 1. Under the simplest model (Model 22222); 1st order polynomial was fitted for each additive genetic, maternal genetic, maternal permanent environmental and genetic group effect, with a total of 22 parameters. Model 44334, with a total of 44 parameters, was observed as a more optimum model, assessed based on a significantly higher log-likelihood of -232121, lower AIC 464331, and lower BIC value 464763.

Accuracy and slope estimates of EBVs for a different order of fit calculated using adjusted phenotypes were presented in Table 2. Accuracy estimates from model 22222 were 0.75,0.74, 0.68, 0.64, and slope were 0.87, 0.94, 0.91, 0.90, respectively, for Wwt, Pwt, Ywt and Hwt. Accuracy increased to 0.77, 0.75, 0.69, and 0.66 for Wwt, Pwt, Ywt and Hwt using model 33333 was significant (P<0.05). Similarly, regression slope improved to 0.95, 0.93, 0.97, 0.99 and was significant improvement (P<0.05) over 22222 model. Higher-order of fit 44334 produced no improvement over the 33333 model.

The accuracy estimate observed from the Legarra-Reverter method for Model 22222 was 0.89, 0.86, 0.83 and 0.81 and slope was 0.96, 0.96, 0.96 and 0.92, respectively, for Wwt, Pwt, Ywt and Hwt (Table 3). There was no significant improvement from the higher order of fit over the 22222 and 33333 models.

Table 1. Log-likelihood and corresponding information criteria.

Model[1]	Parameters estimated	Log likelihood	AIC	BIC
22222	22	-234,831	469,706	469,922
33333	35	-233,134	466,338	466,681
44331	42	-232,133	464,351	464,764
44334	44	-232,121	464,331	464,763

[1] Order of fit for additive genetic effect, animal permanent environmental effect, maternal genetic effect, maternal permanent environmental effect, and genetic groups effect, respectively.

Table 2. Accuracy and slope estimates of EBVs through forward prediction.

Model	Accuracy				Slope			
	Wwt	Pwt	Ywt	Hwt	Wwt	Pwt	Ywt	Hwt
22222	0.75	0.74	0.68	0.64	0.87±0.03	0.94±0.03	0.91±0.03	0.90±0.04
33333	0.77	0.75	0.69	0.66	0.95±0.04	0.93±0.03	0.97±0.03	0.99±0.04
44331	0.75	0.73	0.69	0.66	0.88±0.03	0.89±0.03	0.92±0.03	1.02±0.04
44334	0.76	0.75	0.69	0.66	0.95±0.04	0.86±0.03	1.00±0.03	1.02±0.04

Table 3. Validation result using the method of Legerra and Reverter.

Model	Accuracy				Slope			
	Wwt	Pwt	Ywt	Hwt	Wwt	Pwt	Ywt	Hwt
22222	0.89	0.86	0.83	0.81	0.96±0.004	0.96±0.004	0.96±0.007	0.92±0.010
33333	0.87	0.86	0.83	0.77	0.94±0.005	0.96±0.004	0.98±0.007	0.96±0.012
44331	0.88	0.84	0.84	0.83	0.98±0.005	0.94±0.004	0.97±0.007	0.95±0.009
44334	0.87	0.86	0.81	0.80	0.94±0.005	0.94±0.004	0.97±0.007	0.94±0.011

Discussion

We observed only minor differences in accuracy between RR models with various orders of fit that differed significantly in log-likelihood, AIC, and BIC. The ability of a model to predict progeny performance was improved for a model with a quadratic fit for all random effects, which showed that high order of polynomial beyond quadratic fit is not required for accurate genetic evaluation. Many large-scale genetic evaluations could not easily implement the RR model because of higher computation requirements, and the present study showed that a reduced order of polynomial fit performs not worse compared to a higher order of fit. All animals had at least four measurements which were sufficient to model a cubic fit. Further, we validated our findings by checking how well EBVs estimated from the partial set can predict true EBVs; accuracy observed from this method was higher than regressing phenotypes on EBVs, but there was a non-significant difference between different models. Li *et al.* (2020) also observed RR with quadratic fit better in dairy cattle when validating EBVs. Therefore, the RR model with quadratic order of fit is suggested to be used in genetic evaluation for a trait with repeated measurements.

Conclusions

Random regression with quadratic order of Legendre polynomials performed better than RR models with other polynomials to model repeated weight records in Australian maternal sheep. Therefore, RR models with Legendre polynomial orders higher than three are not required to model multiple weight records in Australian sheep.

References

Brown, D., Huisman, A., Swan, A., Graser, H., Woolaston, R., Ball, A., Atkins, K., & Banks, R. (2007) Proc. Assoc. Advmt. Anim. Breed. Genet. Armidale, Australia.

Fischer, T. M., van der Werf, J. H. J., Banks, R. G., & Ball, A. J. (2004) Livest. Prod. Sci. 89(2), 175-185. https://doi.org/10.1016/j.livprodsci.2004.02.004

Gilmour, A. R., Gogel, B. J., Cullis, B. R., Welham, S. J., & Thompson, R. (2015) VSN International Ltd, Hemel Hempstead, HP1 1ES, UK.

Legarra, A., and Reverter, A. (2018) Genet. Sel. Evol. 50(1), 53. https://doi.org/10.1186/s12711-018-0426-6

Li, J., Gao, H., Madsen, P., Li, R., Liu, W., Bao, P., Xue, G., Gao, Y., Di, X., & Su, G. (2020) Front. Genet. 11(1269). https://doi.org/10.3389/fgene.2020.586155

Meyer, K. (2001) Genet. Sel. Evol. 33(5), 487. https://doi.org/10.1186/1297-9686-33-5-487

Schaeffer, L. R. (2004) Livest. Prod. Sci. 86(1), 35-45. https://doi.org/10.1016/S0301-6226(03)00151-9

Van der Werf, J.H.J., Goddard, M., & Meyer, K. (1998) J. Dairy. Sci. 81(12), 3300-3308. https://doi.org/10.3168/jds.S0022-0302(98)75895-3

725. Estimation of linkage disequilibrium in Pag and Istrian sheep breed: towards genomic optimum contribution selection

A. Kasap[1*], J. Ramljak[1] and M. Špehar[2]

[1]University of Zagreb Faculty of Agriculture, Svetošimunska 25, 10000 Zagreb, Croatia; [2]Croatian Agency for Agriculture and Food, Svetošimunska 25, 10000 Zagreb, Croatia; akasap@agr.hr

Abstract

Knowledge of the genetic architecture of the population is important to predict the potential of genomic selection. The strength of association between genome wide single nucleotide polymorphisms (SNPs) and unknown quantitative trait loci (QTL), measured by linkage disequilibrium (LD), represents the corner stone of genomic selection. The aim of this study was to estimate LD in two Croatian native dual purpose (milk-meat) sheep breeds (Pag sheep, n=719 and Istrian sheep, n=749) using the Illumina Ovine 50K SNP chip. The main goal in these populations is to improve dairy yield traits, preferably via genomic optimum contribution selection. The estimated LD was low (r^2_{Pag}=0.04, D'_{Pag}=0.28, $r^2_{Istrian}$=0.06, $D'_{Istrian}$=0.39,) with very steep LD decay in both breeds. The results obtained are promising for conservation of these breeds, but less promising for the success of genomic selection. In order to maximize benefits of genomic information, single step genomic BLUP should be applied.

Introduction

Istrian and Pag sheep are Croatian native dairy orientated breeds with great economical, traditional and cultural importance. In order to preserve them '*in situ*', it is necessary to permanently increase their productivity. Small breeds under selection are under risk of genetic erosion, but low productivity leads them towards extinction due to non-competitiveness with highly selected breeds (FAO, 1998). Balancing between selection gain and loss of genetic variability is a challenging but feasible task, best achieved by following the principles of optimum contribution selection (OCS) (Wellmann *et al.*, 2021; Wooliams *et al.*, 2015). High-density SNP arrays have been used as an important source of information for selection (Van Raden, 2020) and conservation (Windig and Engelsma, 2010) in many livestock populations. However, benefits of genomic information in selection vary between populations, and depend heavily on population genetic parameters such as linkage disequilibrium (LD). Association between marker alleles (SNP) and causative alleles (QTL) directly affects the accuracy of genomic selection (Meuwissen *et al.*, 2001) and estimation of genetic diversity parameters (Boettcher *et al.*, 2010, Li *et al.*, 2011). Wishing to upgrade the existing selection strategy to genomic optimum contribution selection, we aimed to estimate the genome-wide levels of LD. The obtained results will serve as a base to ascertain the benefits of potentially novel selection strategy in these breeds.

Materials & methods

Samples for genotyping were collected with the permission of the Bioethical committee for protection and welfare of animals at University of Zagreb Faculty of Agriculture. Ear tissue from 719 Istrian sheep (13 flocks) and 749 Pag sheep (11 flocks) were taken with Allflex tissue sample unit applicator. Animals were originally genotyped for 52,152 SNPs, using the Illumina OvineSNP50K BeadChip. Only autosomal SNPs were included in the analysis after quality control parameters for missingness per animal, missingness per SNP, and minor allele frequency set to 0.9, 0.95, and 0.01, respectively. SNPs were also removed if they deviated from the Hardy-Weinberg equilibrium (*P*-value cut-off set to 1×10^{-6}). A total of 693 Istrian (44,743 SNPs) and 713 Pag sheep (46,685 SNPs) passed the above control. The linkage disequilibrium (LD) between pairs of SNPs was calculated with the PLINK (version 1.9) (Purcell *et al.*, 2007; Chang *et al.*, 2015) using the two most commonly used statistics r^2 and D'. Post hoc statistical analysis of genome wide LD was

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_725, © A. Kasap *et al.* 2022

conducted with dplyr (Wickham *et al.*, 2021) and ggplot (Wickham 2016) packages in R (R Core Team, 2020).

Results

The minimum and maximum chromosome-wide average LDs were: r^2_{min}=0.046, r^2_{max}=0.078, D'_{min}=0.314, D'_{max}=0.423 for Istrian; and r^2_{min}=0.027, r^2_{max}=0.049, D'_{min}=0.245, D'_{max}=0.307 for Pag sheep breed. The overall genome wide LDs were: $r^2_{Istrian}$=0.06, $D'_{Istrian}$=0.39, r^2_{Pag}=0.04, D'_{Pag}=0.28. The tabular (Table 1) and graphical representations (Figure 1) of LD change accompanied by between-marker distance revealed very similar pattern of LD decay between these populations. Graphical representation of LD decay obtained by plotting the average LD vs average between-marker distance of bins spanning 20 Kbp revealed a very steep decay of LD within first 150 Kbp, an intermediate decay from 150 to 250 Kbp, and negligible decay after 250 Kbp.

Table 1. Genome wide linkage disequilibrium (LD) classified by between-marker distance.

	Istrian sheep				Pag sheep			
Kbp	**mean (r^2)**	**sd (r^2)**	**mean (D')**	**sd (D')**	**mean (r^2)**	**sd (r^2)**	**mean (D')**	**sd (D')**
[0,100]	0.12	0.19	0.51	0.33	0.10	0.17	0.44	0.33
(100,200]	0.06	0.10	0.40	0.29	0.04	0.08	0.30	0.26
(200,300]	0.05	0.07	0.37	0.27	0.02	0.05	0.25	0.23
(300,400]	0.05	0.07	0.35	0.27	0.02	0.04	0.24	0.21
(400,500]	0.04	0.06	0.35	0.26	0.02	0.03	0.23	0.21
(500,600]	0.04	0.06	0.34	0.26	0.02	0.03	0.22	0.20
(600,700]	0.04	0.06	0.34	0.26	0.02	0.03	0.22	0.20
(700,800]	0.04	0.06	0.35	0.26	0.02	0.03	0.22	0.20
(800,900]	0.04	0.07	0.34	0.26	0.02	0.03	0.21	0.20
(900,1000]	0.04	0.05	0.33	0.26	0.02	0.03	0.21	0.20

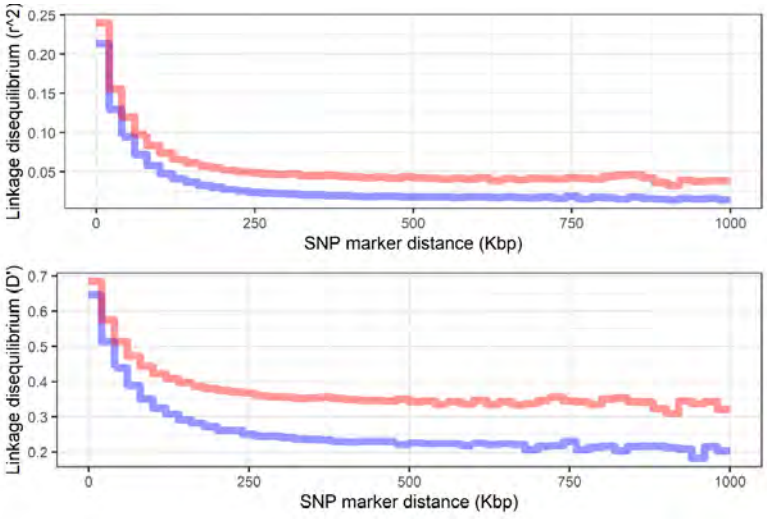


Figure 1. The linkage disequilibrium (LD) decay in Istrian (red) and Pag (blue) sheep.

Discussion

The extent of genome-wide LD and pattern of LD decay are important for GWAS studies, genomic selection, conservation genetics and many other applications. Therefore, our intention was to ascertain this parameter in Istrian and Pag sheep breeds and compare them with some other populations. In this study, we used the two most commonly used statistics, r^2 and D', to allow easy comparison of these results with some previous reports. However, in the between-study comparisons, we focused on r^2 due to its' greater robustness to allele frequency variation than D' (Ardlie *et al.*, 2002). Estimates of genome-wide LD and LD decay in both breeds were very low and steep, respectively. The estimates were very similar to those obtained for French Lacaune (Baloche *et al.*, 2014), Australian Merino, crosses Merino × Border Leicester, crosses Merino × Border Leicester × Poll Dorset (Al-Mamun *et al.*, 2015) and Chinese Merino (Liu *et al.*, 2017). The LD was considerably lower and LD decay less pronounced compared to pure Border Leicester and Poll Dorset breeds (Al-Mamun *et al.*, 2015). The results of the study by Alvarenga *et al.* (2018), conducted in Brazilian Santa Inês sheep, are also comparable to the estimates from this study, especially their reports for chromosome wide LD that were considerably lower (LD r^2_{min}=0.010, r^2_{max}=0.023, D'_{min}=0.176, D'_{max}=0.248). The low levels of LD and high rates of LD decay suggest high level of genetic variability in both breeds which is perspective from the conservation point of view, and less perspective for success of genomic selection. The required level of r^2 for genomic selection to achieve an accuracy of 0.85 should be 0.2 (Meuwissen *et al.*, 2001). The estimates of LD and LD decay have been reported in some other studies in sheep (e.g. Prieur *et al.*, 2017), but because of the different methodology and metric used in presentation of results it would be inappropriate to compare them with results from this research. The low LD in this study confirm many previous conclusions that sheep populations exhibit substantially higher genetic variability than cattle (e.g. De Roos *et al.*, 2008, Sargolzaei *et al.*, 2008). However, regardless of being less favorable than in dairy cattle, the estimated LD should be sufficient for accurate genomic selection (Goddard, 2009). To conclude LD profiles of the Istrian and Pag sheep indicate substantial genetic variability of the breeds and limited benefits of applying genomic selection in term of increased accuracy of estimated breeding values. Some objectively applicable strategies to bridge this gap in practice include genotyping with high density chips in combination with imputation techniques and implementation single step genomic approach in routine genetic evaluation system. Financial investments in novel selection strategies in these populations should be carefully examined to optimize selection under reasonable costs.

Funding

This research has been fully supported by the Croatian Science Foundation under the project Genomic characterization, preservation, and optimum contribution selection of Croatian dairy sheep (OPTI-SHEEP), grant number IP-2019-04-3559.

References

Al-Mamun H.A., Clark S.A., Kwan P., and Gondro C. (2015) Genet. Sel. Evol. 47(90). https://doi.org/10.1186/s12711-015-0169-6.

Alvarenga A.B., Rovadoscki G.A., Petrini J., Coutinho L.L., Morota G. *et al.* (2018) Sci. Rep. 8(8851). https://doi.org/10.1038/s41598-018-27259-7.

Ardlie K.G., Kruglyak L., and Seielstad M. (2002) Nat. Rev. Genet. 3:299-309. https://doi.org/10.1038/nrg777.

Baloche G., Legarra A., Sallé G., Larroque H., Astruc J.M. *et al.* (2014) J Dairy Sci, 97(2):1107-1116. https://doi.org/10.3168/jds.2013-7135.

Boettcher P.J., Tixier-Boichard M., Toro M.A., Simianer H., Eding H. *et al.* (2010) Anim. Genet. 41(1): 64-77. https://doi.org/10.1111/j.1365-2052.2010.02050.x.

Chang C.C., Chow C.C., Tellier L.C., Vattikuti S., Purcell S.M. *et al.* (2015) Gigascience 4(1): s13742-015-0047-8. https://doi.org/10.1186/s13742-015-0047-8.

De Roos A.P.W., Hayes B.J., Spelman R.J., and Goddard M.E. (2008) Genetics 179:1503-1512. https://doi.org/10.1534/genetics.107.084301.

Food and Agriculture Organization of the United Nations (FAO) (1998) Secondary guidelines for development of National FAGR. Management of small populations at risk. Available at: https://www.portalces.org/sites/default/files/migrated/docs/smlpopn.pdf.

Goddard, M. (2009) Genetica 136(2):245-257. https://doi.org/10.1007/s10709-008-9308-0.

Li M.H., Strandén I., Tiirikka T., Sevón-Aimonen M.L., and Kantanen J. (2011) PLoS ONE 6(11): e26256. https://10.1371/journal.pone.0026256.

Liu S., He S., Chen l., Li W., Di J. *et al.* (2017) Genes Genomics 39(7):733-745. https://doi.org/10.1007/s13258-017-0539-2.

Meuwissen T.H.E., Hayes B.J., and Goddard M.E. (2001) Genetics 157(4):1819-1829. https://doi.org/10.1093/genetics/157.4.1819.

Prieur V., Clarke S.M., Brito L.F., and McEwan J.C. (2017) BMC Genetics 18(68). https://doi.org/10.1186/s12863-017-0534-2.

Purcell S.N.B., Neale B., Todd-Brown K., Thomas L., Ferreira M.A.R. *et al.* (2007) Am J Hum Genet 81(3):559-575. https://doi.org/10.1086/519795.

R Core Team (2020) R: A Language and Environment for Statistical Computing; R Foundation for Statistical Computing: Vienna, Austria. Available at: https://www.R-project.org.

Sargolzaei M., Schenkel F.S., Jansen G.B., and Schaeffer L.R. (2008) J. Dairy Sci. 91:2106-2117. https://doi.org/10.3168/jds.2007-0553.

Van Raden P.M. (2020) J. Dairy Sci. 103:5291-5301. https://doi.org/10.3168/jds.2019-17684.

Wellmann R. (2021) optiSel: Optimum Contribution Selection and Population Genetics. R package version 2.0.5. 2021. Available online: https://CRAN.R-project.org/package=optiSel.

Wickham H. (2016) ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York.

Wickham H., François R., Henry L., and Müller K. (2021) dplyr: A Grammar of Data Manipulation. R package version 1.0.7. Available online: https://CRAN.R-project.org/package=dplyr.

Windig J.J., and Engelsma K.A. (2010) Conserv. Genet. 11:635-641. https://doi.org/10.1007/s10592-009-0007-x.

Wooliams J.A., Berg P., Dagnachew B.S., and Meuwissen T.H.E. (2015) J. Anim. Breed. Genet. 132(2):89-99. https://doi.org/10.1111/jbg.12148.

726. Genome-wide association study for milk composition traits of Canadian Alpine and Saanen goats

E. Massender[1*], H.R. Oliveira[1,2], L.F. Brito[1,2], L. Maignel[3], M. Jafarikia[1,3], C.F. Baes[1,4], B. Sullivan[3] and F.S. Schenkel[1]

[1]Centre for Genetic Improvement of Livestock, Department of Animal Biosciences, University of Guelph, 50 Stone Road East, N1G 2W1, Guelph, Canada; [2]Department of Animal Sciences, Purdue University, 270 Russell Street South, West Lafayette, IN 47907, USA; [3]Canadian Centre for Swine Improvement Inc., 960 Carling Avenue, K1A 0C6, Ottawa, Canada; [4]Institute of Genetics, Vetsuisse Faculty, University of Bern, Bremgartenstr.109a, 3012 Bern, Switzerland; emassend@uoguelph.ca

Abstract

The aim of this study was to investigate the genetic architecture of milk composition traits (i.e. protein and fat percentages), in first and later lactations, for the Canadian Alpine and Saanen dairy goat breeds. A total of 24 unique single nucleotide polymorphisms (SNP) were significantly associated with the traits at the genome-wise level. These SNP were located within 50 kb of 35 unique positional candidate genes on *Capra hircus* autosomes (CHI) 6 and CHI14, including the casein genes and *DGAT1*, which have been previously associated with milk composition traits in various dairy species.

Introduction

In the last decade, the Canadian dairy goat sector rapidly expanded. The province of Ontario produces most of the goat milk in Canada, with an estimated 55 million litres produced (86.8% of national production) in 2019 (Canadian Dairy Information Centre, 2021). The number of goats on Ontario dairy goat operations increased by 23.1% between 2011 and 2016, as the sector grew to meet market demand (McGonegal, 2017). Canadian dairy goat producers are paid based on milk components. However, unlike Canadian dairy cattle, goat milk production is not supply managed and prices continuously fluctuate based on supply and demand. Most Canadian goat milk is processed into cheese. Thus, milk composition contributes to both producer revenue and affects the cheese-making properties of the milk produced.

Although milk production traits in dairy goats are highly polygenic, there are several genes, such as *CSN1S1* and *DGAT1,* that have been reported to have a major influence on the composition of goat milk (Martin *et al.*, 2017). In French dairy goats, genotypes for *CSN1S1* have been reported to explain 24 to 38% of the phenotypic variation for protein percentage (PP) and 9 to 18% of the phenotypic variation for fat percentage (FP) (Carillier-Jacquin *et al.*, 2016), while *DGAT1* genotypes explained up to 46% of the phenotypic variation in FP (Martin *et al.*, 2017). Genetic testing for *CSN1S1* has been used as a selection tool in dairy goat breeding programs in several countries, including Canada, and it has also been shown that accounting for the effects of major genes may increase the accuracy of genomic predictions (Carillier-Jacquin *et al.*, 2016). However, little is known about the genetic architecture of milk composition traits in the Canadian dairy goat populations.

In this context, the objective of this study was to perform a genome-wide association study (GWAS) for milk composition traits (PP and FP) in the two main Canadian dairy goat breeds.

Materials & methods

Phenotypes and pedigree. The data used in this study were provided by the Canadian Centre for Swine Improvement (www.ccsi.ca) through the Canadian Dairy Goat Genetic Improvement Program (www.goatgenetics.ca). Phenotypes were 305-day lactation records for PP and FP recorded for does of the Alpine

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_726, © E. Massender *et al.* 2022

and Saanen breeds. The first and later lactation (2-11) records were considered separate traits and are noted with the number 1 for first lactation records (i.e. PP1) and 2+ for later lactation records (i.e. PP2+). The final datasets included first lactation records for 6,409 Alpine and 3,434 Saanen does, whereas the later lactation datasets included 12,236 Alpine records on 5,827 does and 5,008 Saanen records on 2,632 does. Pedigree information was obtained from the Canadian Livestock Records Corporation (www.clrc.ca) and trimmed to only include ancestors of animals with phenotypes or genotypes. The pedigrees for Alpine and Saanen had 13,437 and 7,379 animals, respectively, with an average pedigree depth of 15.8-18.4 generations for animals with records. Additional description of the studied populations is presented in Massender *et al.* (2022).

Genotypes. A total of 1,707 genotypes were used in this study, of which 833 were from the Alpine breed (78 bucks, 755 does) and 874 were from the Saanen breed (97 bucks, 777 does). All animals were genotyped with the first version of the GoatSNP50 Bead Chip (Tosser-Klopp *et al.*, 2014) from Illumina Inc. (San Diego, CA, USA), which contains 53,347 single nucleotide polymorphisms (SNP). Genotype quality control was performed within breed, as described in Massender *et al.* (2022), retaining 44,598 (83.6%) and 43,598 (81.7%) SNP for further analyses for the Alpine and Saanen breeds, respectively.

GWAS. The single-step GWAS was performed using the blupf90 family of programs (Aguilar *et al.*, 2014; Aguilar *et al.*, 2019) and the models described in Massender *et al.* (2022). Briefly, the single-trait animal models included fixed effects of the trait mean, doe age (in months, 8 to 108+), as well as parity (2 to 7+) for later lactation traits, and a linear covariate of days in milk at the final milk test for the lactation. The random effects included contemporary group (herd and year of test) and animal additive genetic effect, as well as permanent environmental effect in the models for later lactation traits. The SNP were considered to be significantly associated with the trait if the $-\log_{10}$ *P*-value was greater than the genome-wise significance threshold, after a Bonferroni correction for multiple comparisons at a 5% family-wise error rate (5.95). The GALLO package (Fonseca *et al.*, 2020), available in the R software version 4.0.4 (R Core Team, 2021), was used to retrieve positional candidate genes located within 50 kb of the significant SNP using the gene annotation information for *Capra hircus* assembly ARS1 in NCBI Datasets release 102 (www.ncbi.nlm.nih.gov/assembly/GCF_001704415.1).

Results

Significant SNP. There were 24 unique SNP significantly associated with the milk composition traits at the genome-wise level. The significant SNP were located on CHI 6 for PP1 and PP2+ in both breeds, CHI6 and CHI14 for FP1 in Saanen and CHI14 for FP2+ in Saanen (Figure 1). No SNP were significantly associated with FP1 or FP2+ in Alpine.

Positional candidate genes. A total of 35 unique positional candidate genes were annotated across both breeds and lactations, of which 30 (85.7%) were protein coding. Four positional candidate genes were common in both breeds (*CSN2, TRNAC-ACA, MOB1B, DCK*), three were unique to Alpine (*CSN1S1, SLC4A4, LOC102179276*) and 28 unique to Saanen (Table 1).

Discussion

The genome-wise correction for multiple comparisons was strict, thus, only major genes were expected to be detected. Genes in the casein cluster on CHI6 were identified as positional candidate genes for PP in both breeds, which was expected as *CSN1S1* has a major effect on goat milk composition (Carillier-Jacquin *et al.*, 2016; Martin *et al.*, 2017). Significant SNP for FP were only observed for Saanen, suggesting possible differences in the genetic architecture for this trait between the two breeds, differing from results reported in other populations (Martin *et al.*, 2017). Positional candidate genes identified on CHI14 have

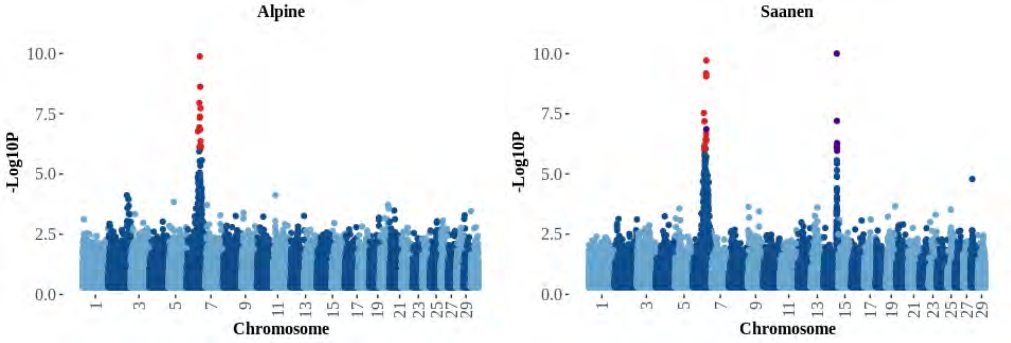


Figure 1. Manhattan plots for milk composition traits in the Alpine and Saanen breeds across first and later lactations. Genome-wise significant SNP are red for protein percentage and purple for fat percentage.

Table 1. Positional candidate genes within 50 kb of genome-wise significant SNP.

Breed	Trait[1]	SNP Position[2]	-Log$_{10}$P	Positional candidate genes
Alpine	PP1	6:85,979,655	6.03	*CSN1S1, CSN2, TRNAC-ACA*
	PP2+	6:85,979,655	8.62	*CSN1S1, CSN2, TRNAC-ACA*
	PP2+	6:85,996,534	6.85	*CSN1S1, CSN2, TRNAC-ACA*
	PP2+	6:86,858,026	6.36	*DCK, MOB1B*
	PP2+	6:87,026,431	7.73	*SLC4A4*
	PP2+	6:89,709,498	6.13	*LOC102179276*
Saanen	PP1	6:86,858,026	9.05	*DCK, MOB1B*
	PP1	6:86,050,088	6.66	*CSN1S2, CSN2, LOC102178810, TRNAC-ACA*
	PP1	6:86,155,374	6.44	*CSN3, LOC102178810, LOC106502214, ODAM*
	PP2+	6:86,858,026	9.71	*DCK, MOB1B*
	PP2+	6:72,867,612	6.01	*LOC102180384, LOC102181107, LOC102183584*
	PP2+	6:73,013,014	6.18	*IGFBP7, POLR2B*
	PP2+	6:86,050,088	9.18	*CSN1S2, CSN2, LOC102178810, TRNAC-ACA*
	PP2+	6:86,155,374	6.36	*CSN3, LOC102178810, LOC106502214, ODAM*
	FP1	6:86,858,026	6.86	*DCK, MOB1B*
	FP1	14:80,143,561	6.14	*ADGRB1, ARC*
	FP2+	14:80,143,561	10.00	*ADGRB1, ARC*
	FP2+	14:81,347,395	7.20	*ADCK5, BOP1, CPSF1, DGAT1, FBXL6, HSF1, SCRT1, SCX, SLC52A2, TMEM249*
	FP2+	14:81,658,383	6.09	*C14H8orf33, LOC106502817, RPL8, ZNF16, ZNF34, ZNF250*

[1] Traits: protein percentage (PP) and fat percentage (FP) in first (1) and later (2+) lactations.
[2] Position of genome-wise significant SNP recorded as *Capra hircus* autosome: base pair.

been previously reported for milk production traits of goats, including: *DGAT1* for FP in the French Alpine and Saanen breeds (Martin *et al.*, 2017) and *ZNF16* for fat yield in multiple New Zealand breeds (Scholtens *et al.*, 2020). Other genes identified (e.g. *DCK*, *MOB1B*, *SLC4A4*, *ODAM*, *RPL8*, *ADCK5*, *CPSF1*, *HSF1*, *SCX*, *BOP1*) have been reported to influence milk composition in dairy cattle (Oliveira *et al.*, 2019; Pedrosa *et al.*, 2021). Overall, this study provided insight into the genetic architecture of milk composition traits in Canadian goats.

Acknowledgements

Funding for this research was provided by the Ontario Ministry of Agriculture, Food and Rural Affairs (OMAFRA), through the Ontario Agri-Food Innovation Alliance. This research is part of the European Union's Horizon 2020 research and innovation SMARTER project.

References

Aguilar, I., I. Misztal, S. Tsuruta, A. Legarra, and H. Wang (2014) Proc. of the 10th WCGALP, Vancouver, Canada.

Aguilar, I., A. Legarra, F. Cardoso, Y. Masuda, D. Lourenco *et al.* (2019) Genet. Sel. Evol. 51:28. https://doi.org/10.1186/s12711-019-0469-3

Canadian Dairy Information Centre (2021) Goat milk production by province. Available at: https://agriculture.canada.ca/en/canadas-agriculture-sectors/animal-industry/canadian-dairy-information-centre/

Carillier-Jacquin, C., H. Larroque, and C. Robert-Granié (2016) Genet. Sel. Evol. 48:54. https://doi.org/10.1186/s12711-016-0233-x

Fonseca, P.A.S., A. Suárez-Vega, G. Marras, and Á. Cánovas (2020) Gigascience 9(12):1-12. https://doi.org/10.1093/gigascience/giaa149

Martin, P., I. Palhière, C. Maroteau, P. Bardou, K. Canale-Tabet, *et al.* (2017) Sci. Rep. 7:1872. https://doi.org/10.1038/s41598-017-02052-0

Massender, E., L.F. Brito, L. Maignel, H.R. Oliveira, M. Jafarikia, *et al.* (2022) J. Dairy Sci. 105:2393-2407. https://doi.org/10.3168/jds.2021-20558

McGonegal, M. (2017) Dairy goats in Ontario: a growing industry. Statistics Canada. Available at: https://www150.statcan.gc.ca/n1/pub/96-325-x/2017001/article/54873-eng.htm

Oliveira, H.R., J.P. Cant, L.F. Brito, F.L.B. Feitosa, T.C.S. Chud, *et al.* (2019) J. Dairy Sci. 102:8159-8174. https://doi.org/10.3168/jds.2019-16451

Pedrosa, V.B., F.S. Schenkel, S-Y. Chen, H.R. Oliveira, T.M. Casey, *et al.* (2021) Genes. 12:1830. https://doi.org/10.3390/genes12111830

R Core Team. 2021. R: A language and environment for statistical computing. Available at: https://www.r-project.org/

Scholtens, M., A. Jiang, A. Smith, M. Littlejohn, K. Lehnert, *et al.* (2020) J. Anim. Sci. Biotechnol. 11:55. https://doi.org/10.1186/s40104-020-00453-2

Tosser-Klopp, G., P. Bardou, O. Bouchez, C. Cabau, R. Crooijmans, *et al.* (2014) PLoS ONE 9(1): e86227. https://doi.org/10.1371/journal.pone.0086227

727. Evaluating performance of Suffolk, Columbia, and crossbred lambs at birth and weaning

R.M. Lewis[1*], N. Vargas Jurado[1,5], D.J. Brown[2], D.R. Notter[3] and J.B. Taylor[4]

[1]University of Nebraska – Lincoln, Lincoln, NE 68583, USA; [2]AGBU, a joint venture of NSW Department of Primary Industries and University of New England, 2351, Armidale, Australia; [3]Virginia Tech, Blacksburg, VA 24060, USA; [4]USDA, ARS, Range Sheep Production Efficiency Research Unit, U.S. Sheep Experiment Station, Dubois, ID 83423, USA; [5]Current address: Natural Resources Institute Finland (Luke), 31600 Jokioinen, Finland; rlewis5@unl.edu

Abstract

Crossbreeding is an important strategy in the sheep industry. The evaluation of purebred individuals alongside crossbred progeny requires considering heterosis in the crossbreds and the potentially differing impacts of systematic environmental effects on performance. As such, the aims of this study were to: (1) compare effects of dam age and birth (or birth and rearing) type; and (2) estimate direct and maternal heterosis effects on body weights. An animal model with maternal and genetic group effects was fitted to birth and weaning weight records from Columbia, Suffolk, and crossbred lambs. Purebred and crossbred records were treated as different traits. Dam age effects differed among breeds at birth and weaning, while effects of birth (or birth and rearing) type did not. Heterosis estimates ranged from 5.2±1.2 to 12.2±4.1%. Direct and maternal heritability estimates were moderately low. The importance of accounting for heterosis and systematic effects was highlighted.

Introduction

Crossbreeding is a common practice in the sheep industry (Alvarez *et al.*, 2013; Notter and Taylor, 2019). Despite the beneficial effects of crossbreeding (Gregory and Cundiff, 1980), and the potential gain in accuracy of sire and dam evaluations due to the increase in the number of records available, crossbred performance data are not often included in routine genetic evaluation.

Due to the diversity in selection and breeding strategies among producers, variability within a given breed is often observed (Brown *et al.*, 2010). To account for this additional source of variability when combining different breeds into a single evaluation, it is important to define founder breed or line genetic group. Furthermore, retained heterosis (direct and maternal) may influence the performance of crossbred progeny (Van Raden, 1992).

When evaluating individuals from different breeds and their crossbred progeny, it is also necessary to consider potential differences in the magnitude of systematic environmental effects. As such, the aims of this study were to: (1) compare the effect of dam age and birth (or birth and rearing) type; and (2) estimate direct and maternal heterosis effects on crossbred lambs, on birth and weaning weights. In addition, variance components for direct and maternal additive effects, maternal uncorrelated environmental effects, and genetic groups effects were estimated.

Materials & methods

Data used were from various sheep genetics projects conducted at the USDA, Agricultural Research Service, Range Sheep Production Efficiency Research Unit, US Sheep Experiment Station (Idaho, USA). All husbandry practices and related animal-care procedures were reviewed and approved by the Unit's Institutional Animal Care and Use Committee.

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_727, © R.M. Lewis *et al.* 2022

Traits. Performance records of purebred Suffolk and Columbia lambs, and crossbred lambs produced by mating purebred and crossbred Suffolk, Columbia, and(or) Texel sires and dams, were used. A total of 4,135, 2,481, and 5,075 birth weight (BN), and 2,486, 1,049, and 3,673 weaning weight (WW) records were available for Columbia, Suffolk, and crossbred lambs, respectively. Mean (SD) BN was 5.2 (1.1), 5.4 (1.1), and 5.5 (1.2) kg. Mean (SD) WW was 40.1 (6.8), 40.7 (7.7), 39.5 (7.7) kg for Columbia, Suffolk, and crossbred lambs, respectively. Correspondingly, these lambs were, on average, 123 (11.2), 124 (11.7), and 121 (12.7) d of age at weaning.

Fixed effects, heterosis, and variance component estimation. Tri-variate models were fitted separately to BN and WW in ASREML 4.0 (Gilmour *et al.*, 2009), with records of Columbia, Suffolk, and crossbred lambs considered different traits The model fitted was:

$$y = Xb + Z_a a + Z_m m + Z_a Qg + Wc + e \quad (1)$$

where $y = [y_1, y_2, y_3]$ was the vector of phenotypic records for the trait of interest (BN or WW), $b = [b_1, b_2, b_3]$ was the vector of fixed effects, $a = [a_1, a_2, a_3]$ was the vector of direct additive effects, $m = [m_1, m_2, m_3]$ was the vector of maternal additive effects, $g = [g_1, g_2, g_3]$ was the vector of genetic group effects, $c = [c_1, c_2, c_3]$ was the vector of maternal uncorrelated environmental effects, and $e = [e_1, e_2, e_3]$ was a vector of residuals. Similarly, $X = \text{diag}(X_1, X_2, X_3)$, $Z_a = \text{diag}(X_{1a}, X_{2a}, X_{3a})$, $Z_m = \text{diag}(Z_{1m}, X_{2m}, X_{3m})$, and $W = \text{diag}(W_1, W_2, W_3)$ were the design matrices relating to fixed, direct additive, maternal additive, and maternal uncorrelated environmental effects, respectively. The diag(·) refers to a block diagonal matrix. The matrix Q contained covariates for genetic group fractions. Three genetic groups were defined for each of the Columbia and Suffolk breeds, while for crossbreds there was an additional genetic group for Texel sires for a total of 7 genetic groups. Fixed effects included dam age (1, …,7+ years) for BN and WW, birth type (1, 2, 3+) for BN, or birth and rearing type (for single-single, multiple-single, multiple-double, and multiple-multiple) for WW, the interaction between dam age and birth type (or birth and rearing type), contemporary group, and lambing day. For the crossbred trait, direct and maternal heterosis, calculated from breed composition were included as continuous covariates. Direct and maternal additive effects for purebreds and crossbreds at BN and WW were assumed to be correlated while additive effects between purebreds were assumed to be uncorrelated. Furthermore, maternal environmental, genetic group, and residual effects were assumed to be uncorrelated among breeds.

Results

Using a likelihood ratio test, the final model for BN included all effects described above, while for WW the model did not include maternal additive effects. Based on an approximate F-test ($P<0.001$) and their confidence intervals, dam age had a smaller effect on Columbia lambs as compared to Suffolk and crossbred lambs at birth and weaning (Table 1). The effect of birth type as well as birth and rearing type on the other hand, differed little among purebreds and crossbred lambs.

In crossbreds at birth, direct and maternal estimates of heterosis were 7.9 (3.9) and 5.9 (1.8) percent, respectively. Similarly at weaning, direct and maternal heterosis estimates were 12.2 (4.1) and 5.2 (1.7), respectively.

Variance estimates are provided in Table 2. At birth, direct and maternal heritability estimates were, respectively, 0.13 (0.04) and 0.31 (0.05) in Columbia, 0.29 (0.06) and 0.30 (0.06) in Suffolk and 0.18 (0.03) and 0.28 (0.04) in crossbreds. Correspondingly, at weaning direct heritability estimates were 0.26 (0.06), 0.18 (0.07), and 0.23 (0.04). At birth, the direct additive correlation between Columbia and crossbreds was 0.55 (0.34) and between Suffolk and crossbreds was 0.87 (0.21). The maternal additive correlation between

Table 1. Contrasts between levels of dam age (years), birth type (single, double, multiple) and birth and rearing type (single-single, multiple-single, multiple-double, multiple-multiple) for purebred and crossbred lambs at birth and weaning.

Trait	Effect	Breed	Contrast		
			1-2	1-3	1-4+[1]
Birth	Dam age	Columbia	0.86 (0.09)	1.23 (0.11)	1.19 (0.06)
		Suffolk	1.29 (0.10)	1.57 (0.13)	1.78 (0.10)
		Crossbred	1.51 (0.06)	1.80 (0.08)	1.90 (0.07)
	Birth type	Columbia	-1.18 (0.08)	-1.75 (0.11)	-
		Suffolk	-1.29 (0.10)	-2.02 (0.18)	-
		Crossbred	-1.20 (0.07)	-1.60 (0.33)	-
Weaning	Dam age	Columbia	7.80 (0.65)	9.48 (0.85)	9.65 (0.48)
		Suffolk	8.62 (0.92)	12.44 (1.15)	11.72 (0.90)
		Crossbred	9.06 (0.43)	11.98 (0.52)	11.93 (0.51)
	Birth-rearing type	Columbia	-3.55 (1.09)	-7.65 (0.95)	-8.36 (0.92)
		Suffolk	-4.49 (1.79)	-7.89 (1.27)	-10.62 (3.30)
		Crossbred	-2.23 (1.35)	-8.70 (0.69)	-6.96 (2.74)

[1] For dam age the contrast is the difference between age 1 and the average of ages 4 to 7+.

Table 2. Estimates of phenotypic (σ_p^2), direct and maternal additive ($\sigma_{A_D}^2$, $\sigma_{A_M}^2$), and maternal uncorrelated environmental variance (σ_c^2) for Columbia, Suffolk, and crossbred lambs at birth and weaning. Standard errors of estimates in parentheses.

Trait	Breed	σ_p^2	$\sigma_{A_D}^2$	$\sigma_{A_M}^2$	σ_c^2
Birth	Columbia	0.88 (0.03)	0.12 (0.03)	0.27 (0.05)	0.06 (0.03)
	Suffolk	1.02 (0.05)	0.29 (0.06)	0.31 (0.07)	0.01 (0.03)
	Crossbred	0.97 (0.03)	0.18 (0.04)	0.27 (0.04)	0.04 (0.02)
Weaning	Columbia	29.41 (1.06)	7.73 (1.82)	-	1.92 (0.71)
	Suffolk	32.46 (1.69)	5.94 (2.52)	-	4.34 (1.41)
	Crossbred	32.01 (0.93)	7.36 (1.42)	-	4.24 (0.67)

Columbia and crossbreds was 0.69 (0.12) and between Suffolk and crossbreds was 0.77 (0.15). At weaning, direct additive correlations between Columbia and crossbreds, and between Suffolk and crossbreds, were 0.97 (0.22) and 0.37 (0.34), respectively. The proportions of phenotypic variance explained by the maternal uncorrelated environmental variance were low for both traits.

Discussion

Anett *et al.* (2011) reported different effect sizes for dam age at birth for Scottish Blackface and crossbred hill ewes, especially for three- and four-years old ewes. In our study, differences between age categories at birth, were larger for the crossbreds and Suffolk while smaller for Columbia.

Estimates of direct and maternal heterosis were considerably larger than those reported in Brown *et al.* (2016). The discrepancy could be due to fewer observations on crossbreds in this study and different breeds considered. Also, in the current study, records from the purebreds and crossbred were considered different traits, which was not the case in Brown *et al.* (2016). Still, the estimates of heterosis highlight the

benefit of using crossbreeding in sheep. Estimates of the correlation for direct additive effects between Suffolk and crossbreds at birth, and between Columbia and crossbreds at weaning, were consistent with those of Brown and van der Werf (2015) for Dorset sheep, and their crosses. Heritability estimates were, in general consistent with the literature (Borg *et al.*, 2009; Fitzmaurice *et al.*, 2020). Due to the moderate size of maternal heritability estimates it seems feasible to enhance the influence that dams have on the performance of their progeny.

Due to the different impact of systematic effects among the breeds evaluated, and the considerable effect of direct and maternal heterosis, adjustment factors can be developed to help reduce bias when evaluating purebred and crossbred individuals.

References

Álvarez J.M., Rodríguez Iglesias R.M., García Vinenta J., Giorgetti H., Rodríguez G., *et al.* (2013) Small Rumin Res. 109(1):9-14. http://dx.doi.org/10.1016/j.smallrumres.2012.08.001

Annett R.W., Carson A.F., Dawson L.E.R, Irwin D., and Kilpatrick D.J. (2011) Animal 5(3):356-366. https://doi.org/10.1017/S1751731110002090

Borg R.C., Notter D.R., and Kott R.W. (2009) J Anim Sci 87(11):3506–3514. https://doi.org/10.2527/jas.2008-1622

Brown D.J., Swan A.A., and Gill J.S. (2010) J Anim Prod Sci. 50(12):1060-1068. https://doi.org/10.1071/AN10102

Brown D.J., and van der Werf J. (2015) Proc. Assoc. Advmt. Breed. Genet. 21st. Conf. Victoria, Australia.

Brown, D. J., Swan A.A., Gill J.S., Ball A.J., and Banks R.G. (2016) Anim. Prod. Sci. 56(9):1442-1448. https://doi.org/10.1071/AN14787

Fitzmaurice S., Conington J., Fetherstone N., Pabiou T, McDermott K., *et al.* (2020) Animal 14(5):899–909. https://doi.org/10.1017/S1751731119002908

Gilmour A.R., Gogel B.J., Cullis B.R., Welham S.J., and Thompson R. (2021). VSN International Ltd, Hemel Hempstead, UK.

Gregory K.E., and Cundiff L.V. (1980) J Anim Sci 51(5):1224-1242. https://doi.org/10.2527/jas1980.5151224x

Notter D.R., and Taylor J.B. (2019) Transl Anim Sci 3(1):1627–1631. https://doi.org/10.1093/tas/txz043

VanRaden, P.M. (1992) J Dairy Sci 75(11):3136-3144. https://doi.org/10.3168/jds.S0022-0302(92)78077-1

728. Genetic selection for sensory eating quality of lamb using consumer assessments

S.Z.Y. Guy[1*], S.I. Mortimer[2], L. Pannier[3], P. McGilchrist[4], D.J. Brown[1], D. Pethick[3] and A.A. Swan[1]

[1]AGBU, a joint venture of NSW Department of Primary Industries and University of New England, 2351, Armidale, Australia; [2]NSW Department of Primary Industries, Armidale Livestock Industries Centre, 2351, Armidale, Australia; [3]College of Science, Health, Engineering and Education, Murdoch University, 6150, Perth, Australia; [4]School of Environmental and Rural Science, University of New England, 2351, Armidale, Australia; sarita.guy@une.edu.au

Abstract

This paper reports genetic parameter estimates for correlations between consumer eating quality and carcase traits for Terminal sire sheep breeds. Four consumer sensory-assessment traits (tenderness, flavour, juiciness, liking) were collected on loin and topside cuts (n~2,700). Heritability estimates (h^2) were low to moderate for these traits ($0.12<h^2<0.33$), with the topside cut having slightly more genetic variation than the loin. Genetic correlations ($\hat{r}_g$) between sensory traits were high within cuts ($\hat{r}_g>0.94$) and across cuts ($\hat{r}_g>0.61$). There were moderate to high correlations of sensory with the objective eating quality traits (intramuscular fat and shear force) and low to moderately negative correlations with carcase lean meat yield. These estimates can be used to revise selection indexes, and potentially develop breeding values for consumer eating quality in multi-trait models with on-farm and carcase traits.

Introduction

Australian terminal sire breeders have achieved substantial genetic gains in various traits (Swan *et al.* 2017). Balanced selection for carcase yield and eating quality was further enhanced from 2017, with new genetic tools underpinned by substantial investment in genomic reference populations. Genomically enhanced breeding values are now available for carcase yield traits, as well as the objectively measured eating quality traits of intramuscular fat and shear force. Breeding programs can now balance the antagonistic relationship between yield and eating quality using modified selection indexes. These indexes are based on breeding objectives where an economic value is defined for consumer-assessed eating quality (Swan *et al.* 2015).

Genetic correlations between consumer eating quality and selection traits (carcase yield and objective measures of eating quality) are required to develop indexes. Current eating quality indexes are based on correlations estimated from consumer eating quality trials on approximately 1,500 animals from multiple breeds. The continued collection of this information has resulted in almost double the amount of both consumer and carcase data available. This paper reports updated genetic parameter estimates between consumer eating quality and carcase traits for Terminal sire sheep breeds. These estimates can be used to revise selection indexes, and potentially develop new multi-trait breeding values for consumer eating quality.

Materials & methods

Data were available for crossbred lambs from terminal sire breeds (mainly Poll Dorset, White Suffolk and Dorper) recorded in Australian sheep reference populations across multiple years and data sources; mainly the Sheep CRC Information Nucleus Flock 2009- and 2010-born (Van der Werf *et al.* 2010), and the 2017- and 2018-born lambs from the MLA Resource Flock. Lambs were processed at an average age of 266±54 days (mean ± SD), hot carcase weight of 24.2±3.9 kg and carcase fat of 4.5±2.5 mm.

Carcase composition and objective eating quality traits examined were carcase eye muscle depth (mm), lean meat yield assessed from computed tomography scanning, and carcase fat (mm). The objective eating

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_728, © S.Z.Y. Guy *et al.* 2022

quality traits, measured in the loin, were intramuscular fat (%) and shear force 5 days of aging (N) (collected according Pearce (2009).

Consumer eating quality traits were collected by obtaining a representative subset of animals from each project. There were 335 sires represented by an average of 8 progeny per sire. Samples were collected, prepared, cooked and tasted by consumers according to the protocol described by Pannier *et al.* (2014). Briefly, five slices were collected from the topside *(Musculus semimembranosus*) and loin (*M. longissimus lumborum*) muscles from each carcase. Samples were prepared and grilled under standardised conditions during each consumer tasting session, and randomly allocated to untrained consumers according to a 6×6 Latin square design. Consumers scored each sample for tenderness, juiciness, flavour and overall liking, scored between zero and 100 (with 100 being most preferred). The scores for each animal were averaged across the 10 consumer responses for each muscle without filtering.

The above traits were analysed in a series of bivariate models ASReml (Gilmour *et al.* 2021):

$$y = \mu + contemporary\ group + animal + genetic\ group + e \quad (1)$$

where y is the phenotype (average consumer eating quality scores per animal, or carcase composition and objective eating quality traits, preadjusted for the fixed effects of birth type, rearing type, age of dam and age at measurement); μ is the mean; *contemporary group* is a fixed effect defined as flock, management group, year of measurement, sex, and kill group sub-classes (as well as laboratory for shear force and intramuscular fat); *animal* is the random additive genetic effect based on the pedigree; *genetic group* is the random effect to account for different base animal breeds and sub-populations within breeds (not fitted for topside traits due to lack of significance); and e is the residual random effect.

Results

On average, the loin samples scored higher for eating quality compared to the topside samples, though there was generally more variation in the topside cuts (Table 1).

Table 1. Summary statistics for eating quality traits (consumer-assessed and objectively measured) and carcase composition traits.

	Trait	*n*	Mean	Min	Max	Coefficient of variation (%)
Loin (0 to 100)	Tenderness	2,707	69.3	28.2	94.0	14.7
	Flavour	2,707	67.9	40.5	91.2	12.4
	Juiciness	2,707	64.4	28.1	93.0	15.1
	Overall liking	2,707	68.6	37.8	92.0	13.1
Topside (0 to 100)	Tenderness	2,756	46.0	5.0	83.9	25.4
	Flavour	2,756	54.2	6.0	81.7	16.4
	Juiciness	2,756	49.7	7.0	84.4	21.1
	Overall liking	2,756	51.4	6.0	84.2	19.5
Intramuscular fat (%)	16,182	4.4	1.2	11.2	25.0	
Shear force (N)	16,123	32.2	10.8	110.2	37.3	
Lean meat yield (%)	1,642	57.5	46.4	68.4	6.3	
Carcase eye muscle depth (mm)	17,944	32.5	16.0	50.8	14.2	

Genetic parameter estimates are presented in Table 2. Heritability estimates were moderate for the consumer eating quality traits (0.12 to 0.19 for the loin, 0.16 to 0.33 for the topside), with the topside generally having higher estimates. Tenderness in the topside had the highest estimate (0.33±0.06), with flavour traits having the lowest estimates for both cuts. Heritability estimates were high for intramuscular fat and lean meat yield, and moderate for shear force and carcase eye muscle depth.

Figure 1 presents estimates of genetic correlation ($\hat{r}_g$). The consumer eating quality traits were highly correlated with each other within cuts (ranging from 0.95 to 1; SEs~0.03), as well as across cuts (0.62 to 0.80; Ses~0.10). There were also moderate to strong genetic correlations between consumer eating quality traits and objective eating quality traits.

Table 2. Additive genetic ($\hat{\sigma}_a^2$), phenotypic ($\hat{\sigma}_p^2$), genetic group ($\hat{\sigma}_{GG}^2$) and heritability estimates ($\hat{h}^2$)(± standard error), averaged from a series of bivariate analyses of eating quality traits (consumer-assessed and objectively-measured) and carcase traits.

	Trait	$\hat{\sigma}_a^2$	$\hat{\sigma}_p^2$	$\hat{\sigma}_{GG}^2$	$\hat{h}^2$
Loin (0 to 100)	Tenderness	20.92±4.70	89.54±2.52	13.59±8.29	0.19±0.05
	Flavour	7.68±2.74	61.81±1.73	5.90±3.80	0.12±0.04
	Juiciness	13.03±3.90	83.48±2.33	6.10±4.31	0.17±0.04
	Overall liking	9.92±3.15	69.30±1.94	7.27±4.56	0.14±0.04
Topside (0 to 100)	Tenderness	39.31±7.04	118.19±3.38	-	0.33±0.06
	Flavour	11.78±3.33	71.48±1.98	-	0.16±0.05
	Juiciness	19.58±4.65	88.70±2.49	-	0.22±0.05
	Overall liking	21.99±4.75	87.96±2.48	-	0.25±0.05
Intramuscular fat (%)	0.47±0.02	0.87±0.02	0.12±0.05	0.57±0.03	
Shear force (N)	24.58±2.07	78.97±1.13	3.65±2.02	0.28±0.02	
Lean meat yield (%)	2.83±0.59	5.96±0.23	6.48±2.85	0.47±0.09	
Carcase eye muscle depth (mm)	3.43±0.28	12.30±0.14	1.69±0.64	0.28±0.02	

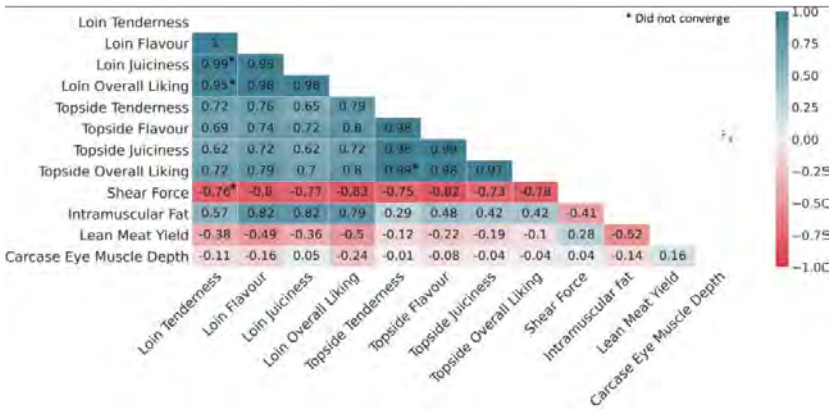


Figure 1. Genetic correlation estimates ($\hat{r}_g$) between eating quality traits (consumer-assessed and objectively-measured) and carcase composition traits (all Ses~0.10).

Shear force had strong negative (favourable) correlations with loin and topside traits (range -0.83 to -0.73). Meanwhile, intramuscular fat (sampled from the loin) had a slightly stronger positive relationship with the eating quality traits from the loin (0.57 to 0.82) compared to the topside (0.29 to 0.48).

The relationships between consumer eating quality traits and lean meat yield were moderately strong in the loin (-0.50 to -0.36), but weak to negligible in the topside (-0.22 to -0.10). Meanwhile, there were weak to no significant correlations of eating quality traits with carcase eye muscle depth (-0.24 to 0.05 for the loin; -0.08 to 0.01 for the topside).

Discussion

Consumer-assessed eating quality of lamb exhibited genetic variation, with the highest heritability estimated for tenderness in the topside cut, which reflects the multi-breed estimates reported in Mortimer *et al.* (2015). Although consumer eating quality traits of the different muscles may be genetically different traits, selection for better eating quality for one muscle will still result in an improvement in eating quality in another muscle.

Consumer eating quality traits were generally more highly correlated with shear force than intramuscular fat, particularly in the topside. However, both traits will be useful as selection criteria in breeding programs. With technological advances in objective carcase measurements, price signals are now emerging for intramuscular fat measured at processing. Therefore, future indexes may directly include intramuscular fat in the breeding objective as an economic trait.

The re-estimated genetic parameters between topside overall liking and lean meat yield align with the previous estimate reported in Swan *et al.* (2015). However, notably stronger genetic correlations were evident between topside overall liking and shear force (-0.78 vs -0.31) in the current study compared to previous estimates (Swan *et al.* 2015). This study also found a negligible relationship between topside overall liking and carcase eye muscle depth (previously estimated at -0.17). Since this study used more data, there is greater confidence in the precision of estimates of genetic correlation between these traits. However, due to changes in correlations with the updated data, it will be necessary to review indexes that include eating quality, along with predictions of response to selection. Future work will explore the development of breeding values for consumer eating quality in multi-trait models with on-farm and carcase traits.

Acknowledgements

This work was partly funded by MLA Project L.GEN.1704. The authors are grateful for the staff involved in collecting this data, funded through MLA L.GEN.1814 and B.SGN.0144.

References

Gilmour A.R., Gogel B.J., Cullis B.R., Welham S.J. and Thompson R. (2021) 'ASReml SA User Guide Release 4.2' VSN International Ltd, Hemel Hempstead, HP1 1ES, UK.

Pannier L., Gardner G.E., Pearce K.L., McDonagh M., Ball A.J. *et al.* (2014) Meat Sci. 96(2):1076-1087. https://doi.org/10.1016/j.meatsci.2013.07.037

Pearce K.L. (2009). Sheep CRC program 3: Next generation meat quality project 3.1.1

phenotyping the information nucleus flocks: Operational protocol series. (1st ed.), Murdoch University, Perth, Western Australia.

Mortimer S.I., Swan A.A, Pannier L., Ball A.J., Jacob R.H *et al.* (2015) Proc. of the 21st AAABG, Lorne, Australia.

Swan A.A., Pleasants T. and Pethick D. (2015) Proc. of the 21st AAABG, Lorne, Australia.

Swan A.A., Banks R.G., Brown D.J. and Chandler H.R. (2017) Proc. of the 22nd AAABG, Townsville, Australia.

Van der Werf J.H.J, Kinghorn B.P. and Banks R.G. (2010). Anim. Prod. Sci. 50 (12):998-1003. https://doi.org/10.1071/AN10151

729. Genetic characterization and connectedness of dairy goats in Canada, France, Italy and Switzerland

M. Teissier[1*], L.F. Brito[2], F. Schenkel[3], G. Bruni[5], P. Fresi[6], B. Bapst[4], C. Robert-Granié[1] and H. Larroque[1]

[1]GenPhySE, Université de Toulouse, INRAE, ENVT, 31326 Castanet-Tolosan, France; [2]Department of Animal Sciences, Purdue University, West Lafayette, IN 47907, USA; [3]Centre for Genetic Improvement of Livestock, Department of Animal Biosciences, University of Guelph, Guelph, ON, Canada; [4]Qualitas AG, 6300 Zug, Switzerland; [5]ARAL, 26013 Crema, Italy; [6]AssoNaPa, 00187 Roma, Italy; marc.teissier@inrae.fr

Abstract

Size of reference population for genomic prediction can limit the increase of accuracy for small populations. Multi-country evaluation can overcome this limit but requires prerequisites such as pedigree connectedness and small genetic distances between populations. Within the SMARTER project, pedigree information and genotypes on SNP50 BeadChip of Alpine and Saanen dairy goats from Canada, France, Italy and Switzerland were shared. Pedigrees analyses have shown unilateral exchanges of animals with flows from France to other countries. Analyses of genotypes revealed a genetically closely related population between France and Italy for both Saanen and Alpine breed. On the opposite, Canadian populations showed the largest differences with European populations. There is connectedness between the different populations included in the project and the genetic distance between populations seems suitable for testing a multi-country genomic evaluation.

Introduction

Genomic prediction of breeding values are routinely performed in several livestock species (Baloche *et al.*, 2014; Boichard *et al.*, 2012; Carillier *et al.*, 2013, Massender *et al.*, 2022). However, the success of these genomic predictions depends on parameters related to the studied populations, including the level of linkage disequilibrium (LD) between single nucleotide polymorphism (SNPs), effective population size, genetic relationship between the training and target populations, pedigree connectedness and trait heritability (Lund *et al.*, 2011; VanRaden, 2020). This condition is not always met for small breeds, in particular in dairy goats. Combining data from genetically-related populations can be an efficient strategy for enlarging training populations for genomic predictions. The objective of this study was to evaluate the genetic connectedness and population structure of Alpine and Saanen dairy goats from four countries involved in the European project SMARTER (Small Ruminants Breeding for Efficiency and Resilience), including Canada, France, Italy, and Switzerland. These analyses are paramount for assessing the potential feasibility of an across-country genomic evaluation in dairy goats.

Materials & methods

Pedigrees and genotypes available. Pedigrees were recorded from 4 countries (Canada, France, Italy and Switzerland) and two goat breeds (Alpine and Saanen). A quality control was performed independently in each country to check the format of animal's identification. We checked that all animals recorded as parent where also identified as animal in the pedigree. 6,435,189 animals were registered in the total pedigree files (58% Alpine and 42% Saanen). The pedigree consists of 91% females and these animals were born between 1936 and 2020. French, Italian, Canadian and Swiss pedigree represented respectively 94.2, 3.4, 1.4 and 1.0%.

Genotypes on the Goat SNP50 BeadChip ((Illumina Inc., San Diego, CA, USA) were filtered out with PLINK 1.9 software. The quality control, performed on all genotypes, consisted of removing duplicate genotypes

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_729, © M. Teissier *et al.* 2022

based on call rate, SNPs with MAF<0.01 and call rate <0.90. Besides, animals with a call rate <0.90 were also removed. After quality control, 9,855 genotypes for a total of 50,578 SNPs were available (Table 1). France have the largest number of genotyped animals (2,968 in Alpine and 2,009 in Sannen) born between 1997 to 2019, whereas Italy started to genotype animals later (2009 to 2019) with less animals (Table 1).

Connections between pedigrees. One of the most complicated stage of the project was the standardization of animal identifiers between countries to find the pedigree connections. Imported animals may have multiple identifiers (one from the country of origin and another one in the importing country). Therefore, up to three aliases were provided by the partners in addition to the id of the animal. This identification, which required a lot of computer and manual work, was important to enable tracing the origin of the curated data but also useful for finding the connections between the different pedigrees.

Genetic structure of populations. We used several metrics to characterize the genetic structure of the four populations. Firstly, linkage disequilibrium (LD) with r^2 was estimated within-country and within-breed with PLINK 1.9 software (Purcell *et al.*, 2007). Then, the consistency of gametic phase was estimated as the correlation coefficient of the signed square roots of the LD between different country or breed. The consistency was calculated for nine categories according to the distance between SNPs: (0, 1 kb], (1, 10 kb], (10, 20 kb], (20, 40 kb], (40, 60 kb], (60, 100 kb], (100, 200 kb], (200, 500 kb], and (500, 1000 kb]. Lastly, we ran Principal Component Analysis (PCA) with PLINK 1.9 and genetic admixture analysis to investigate population structure. In order to comply with the data independence assumption for performing a PCA, the genotypes were pruned using PLINK software. A total of 31,951 SNPs was then retained for the analyses. For Admixture (Alexander *et al.*, 2009), we tested the number of clusters k values ranging from 2 to 8.

Results

Pedigree connectedness. Pedigree analysis indicates that animal flows between partner countries are rather limited. French animals were found in all pedigrees, indicating that France exported animals to all country partners of the project, but France did not import any animals from these countries. In contrast, Italian and Canadian animals were not exported to any other countries based on available recording. Italy was the only country that imported animals from France (9,037 animals) and Switzerland (1,905 animals). Only 322 (Canada) and 180 (Switzerland) French animals are shared with France (Table 2). Exports/imports are well balanced between Alpine and Saanen with 54% Alpine and 46% Saanen.

Table 1. Number of genotypes per breed and per country.

	Alpine	Saanen	Range birth year	Total
Canada	793	903	2001-2019	1,696
France	2,968	2,009	1997-2019	4,977
Italy	1,061	338	2009-2019	1,399
Switzerland	1,280	503	2000-2013	1,783
Total	9,102	3,753	1997-2019	9,855

Table 2. Summary statistics of the impact of French animals (Alpine and Saanen) in Canadian, Italian and Swiss pedigree.

	# French animals shared with France	# animals with French sire	Size pedigree
Canada	180	542	93,342
Italy	9,037	17,137	238,942
Switzerland	322	879	56,989

Consistency of gametic phase. Consistencies of gametic phase remained high between France and Italy even for long distances between SNPs (from 1 for SNP distant of (0, 1 kb] to 0.67 for SNP distant of (500, 1000 kb] in Alpine and from 0.99 for SNP distant of (0, 1 kb] to 0.60 for SNP distant of (500, 1000 kb] in Saanen. We observed lower consistency when comparing all other countries especially for long distance. For Alpine, we observed a consistency on average of 0.97 (std 0.01) for SNP distant of (0, 1 kb] to 0.0.17 (std 0.08) for SNP distant of (500, 1000 kb]. The same trend was observed for Saanen breed.

Principal component analysis. Figure 1 presents the two first components of the PCA analyses. They explained 3.2 and 2.3% of the variability respectively. The French and Italian populations largely overlap and are indistinguishable for both breeds. The Canadian Saanen population is the most differentiated and does not group with the other Saanen populations. The few individuals present between the Canadian Saanen and the European populations are in fact animals with at least one French parent. The Swiss Saanen population is also more differentiated from the other European Saanen populations than the Alpine. Indeed, for the Alpine, there is an overlap of individuals for France, Italy, and Switzerland which is not the case in Saanen with a more homogeneous cluster. We observed one of the lowest cross validation errors with k=4 for the Admixture analyses. As observed with PCA analyses, French and Italian animals have close profiles in both breeds. On the contrary, the Canadian Saanen had distinct profile compared to other groups even when comparing to Canadian Alpine.

Discussion

The first difficulties we encounter is the traceability of animals across countries. For some animals, the original ID was not retained after export, so we were not able to recover their pedigree. The importance and difficulty of exhaustive research of pedigrees has been discussed in previous studies, such as in beef cattle for Interbeef (Venot *et al.*, 2007). International genomic evaluation requires exchange between countries to be performant (VanRaden, 2020). In this case, dairy goat seems to be far from optimal with mainly exchange from France to others. Exception would be between France and Italy where connectedness is strong enough to consider an international genomic evaluation.

Our estimation of LD and consistency of gametic phase is consistent with previous studies performed in each country independently, at least for France and Canada (Brito *et al.*, 2015; Carillier *et al.*, 2013). For Italy and Switzerland, it is the first one with genomic data one in dairy goats, showing lower LD profiles than France. When comparing the consistency between Alpine and Saanen within or between each country, we observed low values. It could make a multi-breed genomic evaluation ineffective. The consistency of the gametic phases between France and Italy for both breeds is high enough to make an international genomic evaluation interesting.

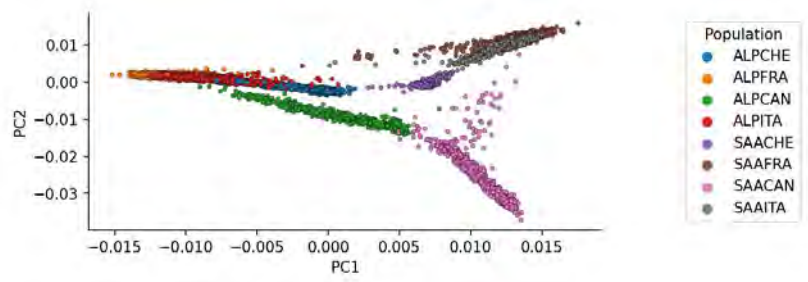


Figure 1. Principal Component Analysis (PCA) with all genotypes for each breed (ALP: Alpine, SAA: Saanen) and each country (CAN: Canada, FRA: France, ITA: Italy, CHE: Switzerland).

Genetic analyses (PCA and Admixture) revealed a distinction between Alpine and Saanen breed. This results is consistent with previous studies performed in France (Oget *et al.*, 2019) or in Canada (Brito *et al.*, 2015). No comparison was available for Italy and Switzerland but we showed that differences between Alpine and Saanen were also present. Canadian populations are quite far from European populations, especially for the Saanen. It may be an obstacle to implementation of an international evaluation. Nevertheless, French and Italian populations are both the most connected and the most genetically similar. An international genomic evaluation for these two countries seems meaningful to implement.

Funding

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 772787 (SMARTER).

Acknowledgements

The authors thank the partners of the Smarter project and the national organizations that made the data sharing possible.

References

Alexander, D. H., Novembre, J., & Lange, K. (2009). Genome Research, 19(9), 1655–1664. https://doi.org/10.1101/gr.094052.109

Baloche, G., *et al.* (2014). Journal of Dairy Science, 97(2), 1107–1116. https://doi.org/10.3168/jds.2013-7135

Boichard, D., *et al.* (2012). Animal Production Science, 52(3), 115–120. https://doi.org/10.1071/AN11119

Brito, L. F., *et al.* (2015). BMC Genetics, 16(1), 67. https://doi.org/10.1186/s12863-015-0220-1

Carillier, C., *et al.* (2013). Journal of Dairy Science, 96(11), 7294–7305. https://doi.org/10.3168/jds.2013-6789

Lund, M. S., (2011). Genetics Selection Evolution, 43, 43. https://doi.org/10.1186/1297-9686-43-43

Massender, E., *et al.* (2022). Journal of Dairy Science. https://doi.org/10.3168/jds.2021-20558

Oget, C., Servin, B., & Palhière, I. (2019). Animal Genetics, 50(1), 54–63. https://doi.org/10.1111/age.12752

Purcell, S., *et al.* (2007). American Journal of Human Genetics, 81(3), 559–575. https://doi.org/10.1086/519795

VanRaden, P. M. (2020). Journal of Dairy Science, 103(6), 5291–5301. https://doi.org/10.3168/jds.2019-17684

Venot, E., *et al.* (2007). InterBull Bulletin, 36, 41–41.

730. Value of data from ram breeding flocks as an industry reference population for Australian sheep

P. Alexandri[1*], S.F. Walkom[1], A.A. Swan[1], J.H.J. van der Werf[2] and D.J. Brown[1]

[1]AGBU, a joint venture of NSW Department of Primary Industries and University of New England, 2351, Armidale, Australia; [2]School of Environmental and Rural Science, University of New England, Armidale, NSW 2351, Australia; palexan8@une.edu.au

Abstract

This study examined the value of using carcass data from seedstock ram breeding flocks to build upon an industry sheep reference population in Australia. Data from 995 lambs managed in 15 commercial ram breeder flocks were collected between 2017 and 2020 for carcass and meat quality measurements: hot carcass weight, tissue depth on the 12th rib (GR site), eye muscle depth, fat between the 12th and the 13th rib (C – site), intramuscular fat and shear force. Industry data were cross-validated with and without reference data from the MLA Resource Flock. Industry data did not bias the estimation of breeding values when used in combination with the reference population (MLA Resource Flock). Therefore, industry data can be used to expand an industry reference population if data collection is accurate and consistent with industry standards.

Introduction

The MLA Resource Flock (RF) is the current Australian sheep reference population and a succession of the Sheep CRC Information Nucleus Flock (Van der Werf *et al.* 2010). The flock provides a well-designed and ongoing sheep reference population and is mostly funded with national levy funds. This industry investment helps subsidise the high cost for progeny testing and measuring expensive traits such as shear force and intramuscular fat of lamb meat.

Historically most RF sites were well-resourced research stations and RF data has been used for research as well as in routine genetic evaluation to produce Australian Sheep Breeding Values (ASBVs). Large reference populations lead to increased accuracy of genomic predictions (Habier *et al.*, 2010) which can be further enhanced by collecting phenotypic and genotypic information from additional animals. As part of expanding the reference population, a series of projects were established where breeders funded data collection from their own flocks with co-investment support from industry funds. These projects allowed for more animals to be genotyped and measured for carcass and meat quality traits to help build an industry reference population with reduced investment per animal measured from industry funds. The quality of data collected on these projects, however, varied and procedures were not always fully comparable to data from the research stations used in the RF. Therefore, we investigated the value of the records collected on industry animals via the co-investment funding model as part of the reference population for sheep in Australia.

Materials & methods

Data. The seedstock data came from 995 lambs born between 2017 and 2020. The lambs came from 15 seedstock ram breeding flocks and were the progeny of 281 sires across 3 breeds (Poll Dorset, White Suffolk and Southdown). A second dataset with 4,027 RF animals was used for validation. The RF animals were born between 2015 and 2020 and were the progeny of 445 sires across the same three breeds (Poll Dorset, White Suffolk and Southdown). Carcass traits for industry and RF animals were measured after slaughter in commercial abattoirs. Weight (CWT) and tissue depth at the GR site (GRFAT, total tissue depth measured with a GR knife on the 12th rib) were measured on the hot carcass. After overnight chilling (3-4 °C), at a cut

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_730, © P. Alexandri *et al.* 2022

between the 12th and 13th ribs of each carcass, eye muscle (*M. longissimus thoracis et lumborum*, LL) depth (EMD) and fat depth at the C site (CFAT) were measured. The percentage of intramuscular fat (IMF) at the LL was determined using a near infrared procedure (NIR) as described by (Perry *et al.* 2001). Shear force (SF5) at 5 days after slaughter was measured at a section of the LL as described by Hopkins *et al.* (2010).

Statistical analyses. Estimated Breeding values (EBVs) were estimated using the LAMBPLAN genetic evaluation software OVIS (Brown *et al.* 2018). All data was pre-adjusted for birth and rear type (single, twins, triplets and quadruplets or more lambs), lamb age and age of dam. Contemporary group, defined by breed, flock, management group, sex, date of measurement and kill group, was used as a fixed effect. Hot carcass weight was included as a covariate to adjust other carcass traits to a weight constant. All models included the random effects of animal, genetic group (Swan *et al.* 2016) and sire × flock interaction. To estimate the differences in accuracy of prediction an internal cross-validation procedure was used for each data set as described by Legarra and Reverter (2018). RF data were separated into four data sets with approximately the same size as the seedstock data. RF animals were randomly assigned to groups one to four based on their sires so that half-sib families were not represented in multiple groups. All seedstock animals were assigned to group five. Numbers of records, sires and unadjusted trait means for each group are shown in Table 1.

Following analysis of the full data set, three different validation scenarios were investigated. First, EBVs were calculated for each RF validation group, using RF data from the other three groups as the training population (RF – RF analysis, performed four times – one for each group). Second, prediction of industry animals was carried out using only RF data (groups one to four) (F – seedstock analysis, replicated four times). Third, prediction of RF animals was performed using four different replicated combinations of RF (groups one to four) and industry animals (group five) (combined analysis).

For each trait, four validation metrics were calculated and averaged across replicates. Accuracy and dispersion metrics were calculated using the LR method (Legarra and Reverter 2018) as the correlation and regression slopes between the EBVs from each of the three analyses (RF – RF, RF – industry, combined) with EBVs from the full analysis. The regression slopes between EBVs are expected to have a value close to one if there is no over or under dispersion. Accuracy and dispersion were also calculated as the correlation and regressions for EBVs on phenotypes adjusted for fixed effects, with regressions performed in ASReml (Gilmour *et al.* 2015).

Table 1. Number of animals (N), sires (Sires), sires per contemporary group (Sires/CG) and unadjusted mean record values (standard deviation) for all traits and validation groups used in the validation analysis. CWT: hot carcass weight (kg), EMD: carcase eye muscle depth (cm), CFAT: fat depth at the C-site (cm), GRFAT: tissue depth at GR-site (cm), IMF: intramuscular fat (%), SF5: shear force 5 days after slaughter (N).

Group[1]	N	Sires	Sires/CG	CWT	EMD	CFAT	GRFAT	IMF	SF5
1	1,100	112	16.7	23.6 (3.8)	32.9 (4.8)	4.3 (2.4)	14.5 (6.2)	4.4 (1.0)	39.3 (12.9)
2	1,162	110	17.7	26.3 (3.7)	35.1 (4.2)	4.8 (2.6)	18.0 (5.1)	4.4 (1.0)	35.3 (10.1)
3	953	111	15.4	26.7 (3.1)	35.4 (4.2)	5.2 (2.4)	20.7 (5.8)	4.7 (1.2)	32.9 (10.3)
4	812	112	26.4	27.5 (3.5)	34.9 (4.1)	5.7 (2.9)	22.3 (6.0)	4.7 (1.1)	29.0 (8.9)
5	995	140	9.6	26.9 (5.2)	37.0 (4.5)	4.4 (2.8)	16.4 (6.7)	3.9 (1.0)	38.3 (15.0)

[1] Data for groups 1 to 4 came from RF animals, group 5 included seedstock data.

Results

Validation results followed similar patterns for two of the three different validation scenarios. For the RF – RF and combined scenarios LR EBV dispersion metrics had values close to one for CWT and CFAT and greater than one for EMD, GRFAT, IMF and SF5 (Table 2). LR EBV correlations ranged from 0.36 (EMD, combined) to 0.52 (GRFAT for RF – RF and combined, SF5 for RF – RF). Phenotypic dispersions for the same validation scenarios (RF – RF and combined) were close to one for GRFAT (0.95 and 0.94 respectively) and CFAT (0.87 for both). However, estimates greater than one were observed for EMD, IMF and SF5 for both analyses. Correlations between EBVs and phenotypes were consistent between the two analyses ranging from 0.19 (EMD) to 0.45 (CWT).

Differences were observed in the metrics when RF data was used to predict into industry animals (scenario RF – seedstock). LR EBV dispersions were closer to one for GRFAT (0.94), CWT (0.85) and CFAT (0.73) although for the last two traits the values observed were lower than the other two validation scenarios (Table 2). LR EBV accuracies were higher for all traits ranging from 0.49 (CFAT) to 0.64 (EMD). Higher phenotypic dispersions were observed for all traits ranging from 0.67 (CWT) to 1.67 (SF5). Phenotypic accuracies were similar to other validation analyses for EMD and CFAT but higher for the rest of the traits. In general, validation patterns were very similar when using RF data to predict RF progeny (RF – RF) and combined RF and seedstock data (combined) were used to validate RF phenotypes but over-dispersion was observed when RF data was used to predict industry phenotypes (RF – seedstock).

Discussion

Accuracy of genomic predictions can benefit from larger reference populations (Habier *et al.* 2010). In this study the value of expanding the Australian sheep reference population by using data from seedstock ram breeding flocks for carcass and meat quality traits was explored. Cross validation results showed that it was possible to use data from seedstock ram breeding flocks to expand the industry reference population for traits recorded using common protocols. Comparison between observed and predicted performances in a

Table 2. Validation metrics for each validation scenario averaged across replicates. CWT: hot carcass weight, EMD: carcase eye muscle depth, CFAT: fat depth at the C-site, GRFAT: tissue depth at GR-site, IMF: intramuscular fat, SF5: shear force 5 days after slaughter.

	Trait						
Metric	CWT	EMD	CFAT	GRFAT	IMF	SF5	Scenario[1]
LR EBV dispersion	0.93	1.16	0.94	1.07	1.28	1.14	RF – RF
	0.85	1.56	0.73	0.94	1.13	1.07	RF – seedstock
	0.92	1.15	0.94	1.08	1.27	1.13	Combined
LR EBV correlations	0.46	0.37	0.44	0.52	0.49	0.52	RF – RF
	0.50	0.64	0.49	0.57	0.56	0.60	RF – seedstock
	0.45	0.36	0.43	0.52	0.48	0.50	Combined
Phenotypic dispersion	0.61	1.21	0.87	0.95	1.33	1.12	RF – RF
	0.67	1.61	1.25	1.01	1.06	1.67	RF – seedstock
	0.58	1.19	0.87	0.94	1.31	1.10	Combined
Phenotypic – EBV correlations	0.45	0.19	0.24	0.40	0.30	0.40	RF – RF
	0.49	0.19	0.24	0.43	0.59	0.65	RF – seedstock
	0.45	0.19	0.24	0.40	0.30	0.40	Combined

[1] RF-RF: prediction using different datasets of RF animals, RF – seedstock: using RF animals to predict into seedstock animals, Combined: using different combinations of RF animals.

cross-validation analysis is important to measure the efficiency of the application of the analysis to specific data sets (Legarra *et al.* 2008). Our results show that for reference and seedstock data, phenotypic and EBV dispersions can be similar when seedstock data are used in combination with a well recorded reference population (Analyses RF – RF and combined, Table 2). Using seedstock and reference data together does not introduce further bias to breeding values estimation. When reference data was used to predict seedstock phenotypes results depended upon the recorded traits. For example, EMD presents higher biases than other carcass traits both for EBV and phenotypic dispersions (Table 2) but also exhibits higher mean values in the seedstock animals compared to the reference data sets (Table 1). Moreover, data structure is different between RF and seedstock animals. The number of sires per contemporary group is typically lower for seedstock data (Table 1) and RF data also normally represents a bigger range of breeds (van der Werf *et al.* 2010). This highlights the importance of recorded data quality; the establishment of an industry reference population can benefit from accurate and consistent data recording.

In conclusion, data collected from seedstock ram breeding flocks can be used to complement managed progeny test sites to create an industry reference population. The effectiveness of commercial data depends on the trait measured (completeness of data and good representation of the flock's diversity) and the influence of fixed effects recorded on the flock. Co-investment into industry recorded carcass traits using levy funds was found to be beneficial to the growth of the reference population. However, these projects should be carried out with caution as there is a risk of reduced sire diversity, less consistent data collection, and reduced data quality if they are not managed properly.

Acknowledgements

This work was conducted as part of the Advanced Measurement Technologies for globally competitive Australian meat project, funded by the Australian Government Department of Agriculture and Water Resources as part of its Rural R&D for Profit programme. The authors thank the teams behind the Sheep CRC Information Nucleus Flock and MLA resource flock.

References

Brown D.J., Swan A.A., Boerner V., Li L., Gurman P.M., *et. al.* (2018) Proc. of the 11th WCGALP, Aukland, New Zealand.

Gilmour A.R., Gogel B.J., Cullis B.R., Welham S.J. & Thompson R. (2015) ASReml User Guide Release 4.1. In: *VSN International Ltd, Hemel Hempstead, UK*

Habier, D., Tetens J., Seefried F.R., Lichtner P. & Thaller G. (2010) Genet. Sel. Evol. 42:5. https://doi.org/10.1186/1297-9686-42-5.

Hopkins D.L., Toohey E.S., Warner R.D., Kerr M.J. & van de Ven R. (2010) *Anim. Prod. Sci.* 50(6):382-385. https://doi.org/10.1071/AN09162.

Legarra A. & Reverter A. (2018) *Genet. Sel. Evol.* 50:53. https://doi.org/10.1186/s12711-018-0426-6.

Legarra A., Robert-Granie C., Manfredi E. & Elsen J.M. (2008) *Genetics* 180 (1):611-618. https://doi.org/10.1534/genetics.108.088575.

Perry D., Shorthose W.R., Ferguson D.M. & Thompson J.M. (2001) *Aust. J. Exp. Agric.* 41(7):953-957. https://doi.org/10.1071/EA00092.

Swan A.A., Brown D.J. & van der Werf J.H.J. (2016) *Ani. Prod. Sci.* 56(1):87-94. https://doi.org/10.1071/AN14560.

Van der Werf J.H.J., Kinghorn B.P. & Banks R.G. (2010) *Ani. Prod. Sci.* 50(12):998-1003. https://doi.org/10.1071/AN10151.

731. Pairwise principal component analysis of African goat breeds identifies loci involved in meat and/or milk production

S.A. Rahmatalla[1,2*#], D. Arends[1#], G.M. Tarekegn[3,4], R. Onzima[5] and G.A. Brockmann[1]

[1]Humboldt-Universität zu Berlin, Invalidenstraße 42, 10115 Berlin, Germany; [2]University of Khartoum, Shambat, P.O. Box 32, Khartoum North, Sudan; [3]Bahir Dar University, P.O. Box 79, Bahir Dar, Ethiopia; [4]Scotland's Rural College, Roslin Institute Building, Easter Bush, EH125 9RG Edinburgh, United Kingdom; [5]Muni University, P.O. Box 725, Arua; Uganda; siham.rahmatalla@hu-berlin.de; [#]these authors contributed equally to this work

Abstract

The selection process in Africa, both environmentally and human-driven, is expected to have led to distinctive variations on the genome between milk and meat goat breeds. We used genotypic data from the caprine 50k Illumina BeadChip to assess the genetic differentiation between nine milk-type and 12 meat-type breeds from Africa by applying a novel pairwise principal component analysis (PCA) method. Using the PCA approach, 95 comparisons contribute to the optimal dissipation on PC1 for milk-type versus meat-type breeds. We identified nine genomic regions that highly differed between milk-type and meat-type breeds in over 25% of comparisons. These SNPs are located on chromosomes 2, 3, 4, 5, 7, 12, 18, and 25. In particular, near these SNPs, candidate genes that might potentially control milk and meat production traits were located. The novel PCA approach helps the identification of genomic regions that contributed to the genomic variation of milk-type versus meat-type breeds.

Introduction

Growing human populations generate an increased demand for food, including protein-rich products from animals (Henchion *et al.* 2017). Goats can play an important role in meeting these demands (Aziz 2010). In Africa, goats are profoundly entrenched in almost every African culture, especially within communities that cannot afford large livestock (Peacock 2005). Goats offer an advantage in animal production and food security as they can produce milk and meat under harsh environmental conditions by using inexpensive food resources. Therefore, costs of goat production are relatively low (Peacock 2005; Mataveia *et al.* 2021) and milk and meat production are essential niches in most African countries to overcome poverty and malnutrition.

Research and developments to improve the relatively low level of productivity of African goats do not match their importance. As a result, many goat breeds are not sufficiently phenotypically explored. To improve the productivity, a better understanding of the animal's genomic data and its relation to phenotypes is necessary (Rexroad *et al.* 2019). Getting, genotypic data of African goat breeds is a challenge. Nevertheless, over the last decade, researchers were successful at collecting a significant amount of genotypic data.

This study aims to identify genomic regions that differentiate between milk and meat goats.

Materials & methods

Animals. Illumina Goat SNP 50K genotypes from African goat breeds were obtained from the Adaptmap project (Bertolini *et al.* 2018) collaborators, and previous own genotyping. The first step is quality control (QC) of the data: Animals where the breed designation is missing, animals that do not cluster inside their designated breed, animals classified as cross-breed animals, and animals that show high relatedness to other animals were removed from the data set. Furthermore, per breed we require a minimum of 20 animals to be genotyped after QC. This resulted in 2,044 animals belonging to 45 African breeds suitable for further

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_731, © S.A. Rahmatalla *et al.* 2022

analysis. The next step was to classify each breed for its most common usage type, being primarily 'milk' or 'meat'-type goats. This was done by literature research and talking to experts on African goat breeds. In total, we reliably classify nine breeds as being milk-type breeds (Abergelle, Burundi, Guera, Gumez, Keffa, Nubian, Peulh, Saanen, and Sofia) and 12 breeds as being meat-type breeds (Ambo, Angora, Boer, Cameroon, Central Highland, Long-eared Somali, Mubenda, Nilotic, Sahel, Sebei, Small East African, Taggar). This leads to the final data set, which contained 1,098 animals belonging to 21 African goat breeds.

Genotype quality control. Since the method used here is based on PCA, no missing genotype data is allowed. We remove animals with more than 5% of SNP genotypes missing. After this step, we only kept fully genotyped markers (call rate 100%). This resulted in 1,090 animals genotyped at 21,889 markers.

Pairwise principal component analysis. To detect SNPs which have a high likelihood of being involved in meat or milk production, we adapted our principal component analysis method (Rahmatalla *et al.* 2017) into a repeated pair-wise comparison method. The method works as follows for a single comparison: Take one milk-type and one meat-type breed, perform standard PCA on the genotypes of the animals belonging to these two breeds. Find if there is a threshold in PC1 that perfectly separates the milk-type animals from the meat-type animals. If a perfect separation is found in PC1, we continue by computing the contribution of each SNP to PC1 as described previously (Rahmatalla *et al.* 2017). In short, variable correlations with PC1 were calculated by multiplying the loading factors with the principal component standard deviation. The quality of representation for variables on the factor map (cos^2) was calculated. All cos^2 values for PC1 are summed up, and the contribution of each SNP was expressed as the percentage of the total variation. After this, we estimate the significance of the SNP contribution, which is then adjusted for multiple testing using Benjamini-Hochberg correction. SNPs with an adjusted contribution <0.05 were deemed significantly contributing to the separation of the two breeds.

We exhaustively test all possible combinations of milk-type and meat-type African goat breeds, each time recording which marker is significantly contributing to PC1. We then compute for each SNP the percentage of comparisons in which it is found to differ between two breeds. Many phenotypes are different when comparing any two breeds, but only the milk/meat-type classification is consistently different in our approach. SNPs contributing to milk/meat production will be detected 'consistently', while SNPs contributing to other phenotypic differences will only be found spuriously.

Results

We performed pairwise PCA of autosomal SNPs to assess the difference between the milk type and meat type African goat breeds. In total, 95 comparisons were performed, which showed a perfect separation on PC1 that split the milk breed from the meat breed. Figure 1 illustrates the percentage of times a SNP was found to be significantly contributing to the PC1 of the pairwise PCA comparison. PCA was done using 21,889 SNPs to detect which SNPs contributed highly to PC1. After 95 comparisons we found nine SNPs that contributed highly in more than 25% of comparisons. These markers are found in chromosomes 2, 3, 4, 5, 7, 12, 18, and 25 (Table 1). Although these markers consistently differentiate between milk and meat type goat breeds we do not know the direction of the effect. The genes, *SCN1A*, *SCN9A*, *SCN7A*, *ST6GALNAC5*, *FERD3L*, *SLC41A2*, *APPL2*, *AK5*, *BCKDK*, *ITGAD*, *COX6A2*; *AHSP*, *SEPTIN14*, and *PHKG1* are located in a region ±500 kb around the identified SNPs (Table 1) and could potentially be functionally associated with the milk or/and meat traits.

Discussion

The novel approach based on pairwise PCA identified SNPs contributing consistently to PC1. These SNPs are biologically crucial in the distinction between milk and meat-type goat breeds. Both milk and meat

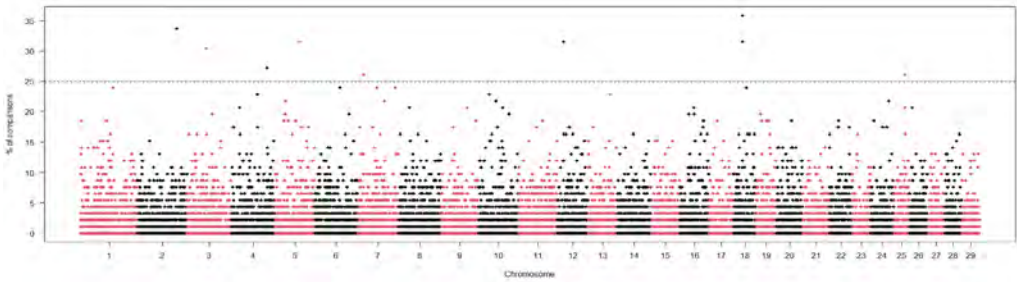


Figure 1. Percentage of pairwise comparisons in which SNP is found to significantly contribute to the observed difference between a milk-type and meat-type breed.

Table 1. SNPs contributing consistently (>25%) to PC1 when comparing milk versus meat-type goats.

SNP ID	Chr[1]	Position	Perc[2]	Genes	QTL[3]
snp31725-scaffold3533-2077	2	106,410,342	33.7	*SCN1A, SCN9A, SCN7A*	Rahmatalla *et al.* 2018; Talouarn *et al.* 2020
snp1663-scaffold1047-65932	3	53,787,794	30.4	*ST6GALNAC5, AK5*	Rahmatalla *et al.* 2018; Luigi-Sierra *et al.* 2020
snp32092-scaffold362-92730	4	92,298,257	27.2	*FERD3L*	
snp28835-scaffold310-10750	5	67,569,878	31.5	*SLC41A2, APPL2*	Rahmatalla *et al.* 2018; Martin *et al.* 2016; Kang *et al.* 2020; Talouarn *et al.* 2020
snp29287-scaffold316-61510	7	18,816,632	26.1		Luigi-Sierra *et al.* 2020
snp24150-scaffold246-06744	12	14,987,932	31.5		Martin *et al.* 2016
snp14564-scaffold1589-8962	18	24,578,286	31.5		Luigi-Sierra *et al.* 2020
snp14565-scaffold1589-8303	18	24,617,627	35.9		Luigi-Sierra *et al.* 2020
snp44137-scaffold6-559066	25	27,423,327	26.1	*BCKDK, ITGAD, COX6A2; AHSP, SEPTIN14, PHKG1*	Luigi-Sierra *et al.* 2020

[1] Chromosome.
[2] Percentage.
[3] Known quantitative trait loci.

production are complex and polygenic traits. Therefore, the distribution of the nine markers on eight chromosomes might be supporting the polygenic nature of these traits. Near to the SNP regions, different studies had QTL for conformation traits that related to meat production, such as bicoastal diameter, body length and rump length (Rahmatalla *et al.* 2018), chest width, and rump width (Luigi-Sierra *et al.* 2020) on chromosomes 2, 3, 5, 7, and 18, respectively. Recently, Talouarn *et al.* (2020) detected new signals for milk yield on chromosomes 2 and 5 in Alpine and Saanen goats, respectively, when using a multi-breed panel. Additional QTLs for milk protein yield, milk fat yield and teat number were identified in different goat breeds on chromosome 5 (Martin *et al.* 2016; Kang *et al.* 2020). Furthermore, Martin *et al.* (2016) and Luigi-Sierra *et al.* (2020) found genomic regions associated with teat number in Alpine and Saanen dairy goats on chromosome 12 and teat placement in Murciano-Granadina goats on chromosome 25. Potential candidate genes that might be associated directly or indirectly with milk or meat production traits are suggested from the PCS results of this study. Further research is required to validate the direction of the allele effect and to prioritize the positional candidate genes.

References

Aziz M.A. (2010). Lohmann Information 45(2): 42-52.

Bertolini F., Servin B., Talenti A., Rochat E., Kim E.S., *et al.* (2018). Genet Sel Evol 50(1): 57. https://doi.org/10.1186/s12711-018-0421-y.

Henchion M., Hayes M., Mullen A.M., Fenelon M. and Tiwari B. (2017). Foods 6(7): 53. https://doi.org/10.3390/foods6070053.

Kang X., Li M., Liu M., Liu S., Pan M.G., *et al.* (2020). Genomics 112(6): 4934-4937. https://doi.org/10.1016/j.ygeno.2020.09.007.

Luigi-Sierra M.G., Landi V., Guan D., Delgado J.V., Castello A., *et al.* (2020). J Dairy Sci 103(12): 11605-11617. https://doi.org/10.3168/jds.2020-18461.

Martin P., Palhiere I., Tosser-Klopp G. and Rupp R. (2016). J Dairy Sci 99(11): 8891-8900. https://doi.org/10.3168/jds.2016-11210.

Mataveia G.A., Visser C. and Sitoe A. (2021). Intechopen. https://doi.org/10.5772/intechopen.97792.

Peacock C. (2005). Small Ruminant Research 60(1): 179-186. https://doi.org/10.1016/j.smallrumres.2005.06.011.

Rahmatalla S.A., Arends D., Reissmann M., Said Ahmed A., Wimmers K., *et al.* (2017). BMC Genet 18(1): 92. https://doi.org/10.1186/s12863-017-0553-z.

Rahmatalla S.A., Arends D., Reissmann M., Wimmers K., Reyer H., *et al.* (2018). Anim Genet 49(5): 478-482. https://doi.org/10.1111/age.12686.

Rexroad C., Vallet J., Matukumalli L.K., Reecy J., Bickhart D., *et al.* (2019). Frontiers in Genetics 10(327). https://doi.org/10.3389/fgene.2019.00327.

Talouarn E., Bardou P., Palhiere I., Oget C., Clement V., *et al.* (2020). BMC Genet 21(1): 19. https://doi.org/10.1186/s12863-020-0826-9.

732. Linking genetic merit to phenotypes: a validation of the terminal and replacement Irish sheep breeding objectives

T. Pabiou[1*], K. McDermott[1], A. Bohan[1], L.J. Farrell[2] and N. McHugh[3]

[1]Sheep Ireland, Highfield House, Shinagh, Bandon P72 X050, Co. Cork, Ireland; [2]Teagasc, Animal and Grassland Research and Innovation Centre, Athenry, H65 R718, Co. Galway, Ireland; [3]Teagasc, Animal and Grassland Research and Innovation Centre, Moorepark, Fermoy, P61 P302, Co. Cork, Ireland; tpabiou@icbf.com

Abstract

Genomic evaluations provide producers with a tool to highlight the increase in performance achievable at farm level through genetic gain. The objective of the present study was to validate phenotypic differences for a range of economically important traits for animals divergent in Terminal and Replacement index. Validation animals were ranked into 5 strata of Terminal and Replacement index separately. Least squares regression analyses were constructed to quantify the effect of index strata on each phenotype adjusted for all fixed effects. Growth traits responded favourably to selection: +0.390 to 0.990 kg depending on the trait and index considered. Very high strata animals were drafted for slaughter 8 days earlier than the very low strata for Terminal index. Number of lambs born increased by +0.17 lamb in the Very high ewe Replacement index strata. Differences in other traits where recording may be less accurate were mostly not significantly different from zero.

Introduction

Sheep production systems are traditionally low input with low profit margins, therefore an increased focused is required identifying strategies to improve flock sustainability and profitability. Genetic indexes provide producers with a tool to aid in breeding decisions and highlight the increase in performance achievable at farm level through genetic gain. The dual sheep breeding objectives in Ireland have been defined previously (Santos *et al.*, 2015; Bohan *et al.*, 2019) and the overall breeding indexes in sheep are represented as Terminal or Replacement indexes. The objective of the present study was to validate the phenotypic differences for a range of economically important traits for animals divergent in genetic merit for the Irish national Terminal and Replacement indexes.

Materials & methods

The data used in the present study were obtained from a pre-existing database managed by Sheep Ireland.

Validation population. The validation population used in this analysis were animals born in the Central Progeny Test (CPT) flocks between 2018 and 2021. All phenotypes recorded in the validation population animals were set to missing prior to the one step genomic evaluation (MiX99 development team); their Terminal and Replacement indexes (Bohan *et al.*, 2019) were therefore computed as parent average using the economic weightings currently employed in the Irish national evaluations. Validation animals were then classified into five separate strata of equal size based on their Terminal and Replacement index values, respectively: Very high (top 20%), High (top 60 to 80%), Average (40 to 60%), Low (20 to 40%) and Very low (0 to 20%).

Validation phenotypes. Performance data from the CPT flocks across a range of animal-specific events including lambing, lamb performance, ewe performance and animal health records between October 2018 and June 2021 were extracted from the Sheep Ireland database.

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_732, © T. Pabiou *et al.* 2022

Lambing performance data were available on three farmer recorded lambing traits, birth weight, lambing difficulty score and lamb survival as described by McHugh *et al.* (2016). Lamb live weights were recorded at three time points: pre-weaning, weaning and post-weaning (McGovern *et al.*, 2020) by farmers using weigh-scales. Lamb carcass data including age at slaughter, carcass conformation and carcass fat were also available on a subset of lambs slaughtered between 95 and 365 days of age (McHugh *et al.*, 2021). Ewe live-weight was defined as the weight of a female who had at least one recorded lambing event; only ewe live-weights recorded between 45 and 120 kg were retained (McHugh *et al.*, 2019). Ewe litter size was defined as the number of lambs born per litter (1 to 4 lambs per litter) for each lambing event. Barren rate was derived from the farmer recorded binary barren trait. Dag score was measured by trained technicians on a five-point scale as: 1 (no faecal soiling) to 5 (faecal soiling) and dags covering the breech area and extending down the hind legs towards the pasterns (O'Brien *et al.*, 2016). Lameness was measured by trained technicians assessing each animal individually as it walked and was recorded for each animal as a binary trait: lame or not (O'Brien *et al.*, 2016) and was treated as a separate trait for ewes and lambs.

Data analysis. The association of the Terminal and Replacement index strata and the corresponding phenotypic performance validation phenotypes was quantified separately for Terminal and Replacement indexes using a fixed effects model (Table 1) with SAS Procedure LSMEANS (SAS Institute, Cary, NC).

Results

This study will only report differences between the two most divergent startum (i.e. Very high and Very low). The average index for the Very low and Very high strata were -€0.70 and +€1.27, respectively for the Terminal index, and -€0.26 and +€2.02, respectively for the Replacement index (Table 2). For lambing traits, differences between the Very high and Very low strata were only significantly different from zero for lamb survival and the Replacement index of the ewe (+0.98%). Relative to lambs born to parents in the Very low strata, live weight traits increased by +0.390 kg to 0.990 kg for lambs in the Very high strata, depending on the traits and the index under investigation ($P<0.05$). Similar to live weight traits, the Very high strata of index showed a reduction of -8.34 and -4.85 days in age at slaughter for Terminal and Replacement, respectively relative to lambs born to parents of the Very low strata ($P<0.05$). Health traits did not differ by the index strata of either the Replacement or Terminal index ($P>0.05$). Relative to ewes in the Very low strata for Replacement index, the number of lambs born increased by +0.17 lamb for ewes of Very high strata ($P<0.05$).

Table 1. Model composition by trait types.

Trait	Fixed class effects (number/range of levels)	Fixed regression effects
Lambing traits	Ewe parity (5), contemporary group (541-628), birth type (4), sex of lamb (2), Index strata[1](5)	Breed composition[2], Index[3]
Lamb live performances	Ewe parity (5), contemporary group (190-236), birth type (4), sex of lamb (2), rearing type (3), Index strata[1](5)	Breed composition[2], Index[3]
Lamb carcass performances	Ewe parity (5), contemporary group (194), birth type (4), sex of lamb (2), rearing type (3), lamb month of birth (5), Index strata[1](5)	Breed composition[2], Index[3], carcass weight
Ewe traits	Ewe parity (5), contemporary group (76-823), age at first lambing (2), Index strata[1](5)	Breed composition[2], Index[3]
Health traits	Ewe parity (5), contemporary group (281-357), birth type (4), sex of lamb (2), Index strata[1](5)	Breed composition[2], Index[3]

[1] Terminal index strata and Replacement index strata were analysed separately.

[2] 10 breeds: Belclare, Beltex, Bluefaced Leicester, Charollais, Llyen, Suffolk, Rouge de l'Ouest, Texel, Vendeen and other.

[3] Terminal index of sire when Replacement strata was analysed; Replacement index of dam when the Terminal strata was analysed.

Table 2. Differences in phenotypes between the Very high and Very low strata for Terminal and Replacement indexes.[1]

Trait (unit)	Desired selection direction	Replacement		Terminal	
		Average phenotypes Very high strata	Phenotypic differences (Very high – Very low) strata	Average phenotypes Very high strata	Phenotypic differences (Very high – Very low) strata
Lamb survival (%)	+	93.55	0.98*	93.27	-0.80
Lambing ease singleton (%)	+	84.29	-0.03	85.59	-0.35
Lambing ease twin+ (%)	+	80.97	0.44	80.37	0.89
Pre-weaning weight (kg)	+	18.27	0.39***	18.15	0.30**
Weaning weight (kg)	+	31.87	0.80***	31.71	0.51**
Post-weaning weight (kg)	+	39.56	0.99***	39.35	0.70**
Age at slaughter (day)	-	206.76	-4.85***	206.75	-8.34***
Carcass conformation (1-5)	+	3.45	0.00	3.43	-0.05***
Carcass fat (1-5)	-	2.96	-0.01	2.96	-0.02
Lamb dagginess (1-5)	-	3.82	0.02	3.82	0.00
Lamb lameness (%)	-	8.01	-0.34	8.07	-0.19
Number of lambs born	+	1.91	0.17***	na	na
Ewe barren rate (%)	-	8.55	-0.12	na	na
Ewe mature weight (kg)	-	73.29	0.53	na	na
Ewe lameness (%)	-	12.25	0.12	na	na

[1] * $P<0.05$; ** $P<0.01$; *** $P<0.001$; na: not applicable.

Discussion

Since its inception in 2008, Sheep Ireland has promoted more profitable pedigree breeding stock for use in commercial sheep enterprises. Today Irish sheep farmers can use the Terminal and Replacement indexes as overall selection tools to increase farm profit (Bohan *et al.*, 2019). The accuracy of genetic and genomic evaluations is fundamental to realising the benefits of using index information in breeding decisions. Nonetheless, scepticism among farmers sometimes exists on the accuracy of evaluations in commercial production systems. Validation studies that report the predictability of indexes have been published in beef cattle in Ireland (e.g. McHugh *et al.*, 2014), no such study has been published on Irish sheep. Results from the present validation study show that, on average, selection for desired genetic attributes will result in favourable improvements in performances. This is particularly true for traits that show the largest differences between Very high and Very low strata: in the present validation analysis for traits that were recorded directly (live weight traits, number of lambs born) and/or with higher heritability (age at slaughter). Lower heritability traits such as lambing traits would potentially require more validation records to show pertinence. Health trait results did suffer from later recording compared to other more routinely collected traits; the recording of health traits started in the last decade only.

In conclusion, results from this study clearly show that genetic selection for improved performance will materialise in improvements in phenotypic performance. The capacity for improvement is linked to the accuracy and easiness of recording.

References

Bohan, A., L. Shalloo, P. Creighton, D. P. Berry, T. M. Boland, A. C. O'Brien, T. Pabiou, E. Wall, K. McDermott, and N. McHugh. 2019. Deriving economic values for national sheep breeding objectives using a bio-economic model. Livest. Sci, 227:44-54. https://doi.org/10.1016/j.livsci.2019.05.018

McGovern, F.M., N. McHugh, S. Fitzmaurice, T. Pabiou, K. McDermott, E. Wall and N. Fetherstone. 2020. Phenotypic factors associated with lamb live weight and carcass composition measurements in an Irish multi-breed sheep population. Transl. Anim. Sci. 4:1-9. https://doi.org/10.1093/tas/txaa206.

McHugh, N., Cromie A. R., Evans R. D., and Berry D. P. 2014. Validation of national genetic evaluations for maternal beef cattle traits using Irish field data. J. Anim. Sci. 92:1423–1432. https://doi.org/10.2527/jas.2013-6658

McHugh, N., D.P. Berry, and T. Pabiou. 2016. Risk factors associated with lambing traits. Anim. 10:89-95. https://doi.org/10.1017/S1751731115001664.

McHugh, N., F. McGovern, P. Creighton, T. Pabiou, K. McDermott, E. Wall and D.P. Berry. 2019. Mean difference in live-weight per incremental difference in body condition score estimated in multiple sheep breeds and crossbreds, Anim. 13:549-553. https://doi.org/10.1017/S1751731118002148.

McHugh, N., A.C. O'Brien, T. Pabiou, K. McDermott, E. Wall and D.P. Berry. 2021. Genetic parameters for carcass traits in a multibreed sheep population. Anim. Submitted.

MiX99 Development Team, Biometrical Genetics, Natural Resources Institute Finland (Luke), FI-31600 Jokioinen, Finland.

O'Brien, A.C, N. McHugh, E. Wall, T. Pabiou, K. McDermott, S. Randles, S. Fair and D.P. Berry. 2016. Genetic parameters for lameness, mastitis and dagginess in a multi-breed sheep population. Anim. 11:911-919. https://doi.org/10.1017/S1751731116002445.

Santos, B. F. S., N. McHugh, T. J. Byrne, D. P. Berry, and P. R. Amer. 2015. Comparison of breeding objectives across countries with application to sheep indexes in New Zealand and Ireland. J. Anim. Breed. Genet. 132 144-154. https://doi.org/10.1111/jbg.12146.

733. Influence of a nutritional restriction in dairy ewe lambs on the response to a later inflammatory intramammary challenge

R. Pelayo[1], H. Marina[1], A. Suarez-Vega[1], C. Esteban-Blanco[1], G. Foucras[2], J.J. Arranz[1] and B. Gutierrez-Gil[1*]

[1]Department of Animal Production, Faculty of Veterinary Sciences, University of León, Campus de Vegazana s/n, 24071 León, Spain; [2]Université de Toulouse, École Nationale Vétérinaire de Toulouse, INRAE, Interactions Hôtes – Agents Pathogènes, 31076, Toulouse, France; beatriz.gutierrez@unileon.es

Abstract

The objective of this work was to evaluate the potential influence of a nutritional challenge performed in ewe lambs on the inflammatory response triggered against an experimental intramammary lipopolysaccharide challenge carried out at the end of the first lactation. Sixty Assaf ewes at 2 months of age were divided into a control and a nutritional challenge group, the second of which was subjected to a severe protein restriction during the prepubertal growth stage. The ewes were inseminated, and after lambing, 24 of them were subjected to an experimental inflammatory challenge of the mammary gland to induce a local inflammation. Measurements of somatic cell counts, rectal temperature and plasma concentrations of 14 cytokines were collected at different time points from the lipopolysaccharide infusion. The results showed that the effect of the nutritional challenge did not show a significant influence on the local indicator somatic cell count (SCC) traits, whereas it had a significant influence on systemic indicator traits. This research opens the way to further studies related to the potential long-time effects of the nutritional plane at early life on immune response.

Introduction

In dairy sheep, mastitis causes economic losses due to decreased milk production and cheese yield (Leitner *et al.* 2008). In dairy cattle, it has been proven that in order to moderate the mastitis related inflammatory response as well as to improve the immune status of dairy herds it is very important that the animals' nutritive needs are well balanced (Turk *et al.* 2017). Different studies in dairy cattle have shown that the plane of nutrition at the prepubertal stage, when the allometric growth of the mammary tissue starts, influences the mammary gland development and the production potential of dairy heifers (Ford and Park, 2001; Sejrsen *et al.* 1982). The majority of these studies have shown that high feeding levels of heifers during that period is detrimental to milk production as cows (Sejrsen *et al.* 2000). In sheep, Villeneuve *et al.* (2010) reported that the restricted feeding before puberty improves mammary gland development without compromising growth performance in ewe lambs. However, to our knowledge, no previous studies in dairy sheep have assessed how the nutritional plane at the time of the mammary gland development may influence the response of this organ to an inflammatory response later during the productive life of the animals. Based on this, the main objective of this work was to evaluate the potential influence that a nutritional challenge performed in ewe lambs at the prepubertal age may have on the inflammatory response triggered over time against an experimental intramammary lipopolysaccharide (LPS) inoculation carried out at the end of the first lactation.

Materials & methods

A total of 60 Assaf ewes were acquired at 2 months of age and randomly divided in two groups: a control group (C=30) and a nutritional challenge group (NC=30), the second of which was subjected to a severe protein restriction during the prepuberal stage (from 3 to 5 months of age). At 10 months of age, the ewes were inseminated, and about 5 months after lambing, 24 of the ewes (11 from the C and 13 from the NC group), were subjected to an experimental inflammatory challenge of the mammary gland through infusion

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_733, © R. Pelayo *et al.* 2022

of an *E. coli* lipopolysaccharide (LPS) solution to induce a local inflammation. For all the considered ewes, rectal temperature measurements (T^{ra}) and plasma samples to measure the concentration of 14 cytokine/chemokine biomarkers were collected one day before de LPS injection (-24h) and at different time points after the LPS injection (2h, 4h, 6h, 8h, 12h, 24h, 48h, 72h and 144h). Likewise, for all the animal's measurements of the milk somatic cell count in the control udder (SCC_c) and for the SCC in the LPS-treated udder (SCC_i) were obtained at seven of the total considered time points (-24h, 0h, 6h, 24h, 48h, 72h and 144h after injection). This inflammatory challenge was approved by the Junta de Castilla y León regional government (Resolution 23/01/2020, Junta de Castilla y León).

Data processing. To consider the evolution of the different trait values across the different time points, the raw value of each measurement was transformed into a ratio trait.

Hence, for each considered systemic phenotype (T^{ra} and 14 biomarkers) and time point combinations, the -24h time point measurement of the same trait, was considered as the basal point, as indicated in the following formulae:

$$Ratio\ Trait\ 1 = [raw\ data\ Trait\ 1\ (0h/2h/4h/\ \ldots 144h)\ /\ raw\ data\ Trait\ 1\ (-24h)] \quad (1)$$

By applying the same transformation to the available measurements of the two phenotypes related to the local inflammatory response, two SCC ratio traits were also defined (r_SCC_c, r_SCC_i).

Statistical analysis. In order to analyze the relative importance of the studied traits throughout the inflammatory challenge, we performed 17 Generalized Linear Model (GzLM) analyses with the SPSS software (Statistics, 2019). The considered model included as fixed effects the 11 considered sample collection time points (-24h, 0h, 2h, 4h, 6h, 8h, 12h, 24h, 48h, 72h and 144h), and the nutritional challenge group (NC vs C). Statistical significance of the GzLM analysis was defined as $P \leq 0.05$.

Results

Phenotypic characterization. After the LPS injection, the treated half-udders of all the animals showed clinical signs of local inflammation, such as redness and pain on palpation, whereas no signs of local inflammation were observed in the control half-udder. A local response to inflammation was observed in all the animals based on a remarkable increase of the SCC in the LPS-treated udder with respect to the control udder with its maximum at the 6h time point after the inoculation. Also, a systemic response to the LPS injection was observed based on the peak of the values of some of the systemic phenotypes analysed, such as the rectal T^{ra} and some plasma biomarkers, especially IL-6, which showed the maximum concentrations between the 4h and 6h time points, with a profile very similar to those observed for the SCC and T^{ra} traits.

GLM-analysis results. The results from the GLM-analysis performed for the defined ratio traits of both groups of phenotypes, local and systemic, in relation to the different time points considered across the experiment and in relation to the early life nutritional challenge group (NC vs C) are shown in Table 1. For both r_SCC_i and r_SCC_c, the 6h and 24h time points were the only highly significant time points of the experiment. Regarding the systemic ratio traits, the r_T^{ra} trait was significant for most time points, and, r_IL-6, r_IL-8, r_IL-10, r_IL- 36RA and r_VEGF-A were the plasma biomarkers that showed the most significant values for the different time points analysed. However, this analysis did not evidence a significant effect of the nutritional challenge groups (NC vs C) on the two local traits studied (r_SCC_i and r_SCC_c) whereas significant effects of the NC group for all the systemic ratio traits considered except for r_IL-6 and r_IL-8 were identified (Table 1).

Table 1. The effect of changes in inflammatory indicators according to different time points and early life nutrition group based on the results of the Generalized linear model (GLMs) analyses.[1]

Traits	0h	2h	4h	6h	8h	12h	24h	48h	72h	144h	NC
r SCC i				***			***	*			
r_SCC_c				***			***				
r_T[a]			**	***	***	*			**	*	***
r_IFN-γ											**
r_IL-1α									*		***
r_IL-1β											***
r_IL-4											*
r_IL-6			**	***	***						
r_IL-8							***	**			
r_IL-10			***	***					*		***
r_IL-17A											*
r_MIP-1α											**
r_MIP-1β											*
r_IL-36RA		**	**		*						*
r_IP-10											***
r_TNF-α								*	*		***
r_VEGF-A			**	**		*	*				***

[1] Statistically significance is marked ***<0.001, **<0.01 and *<0.05, accordingly.

Discussion

Given the importance of the mastitis disease in dairy sheep and the increased knowledge about how the plane of nutrition at growth stages can influence phenotypes of economic interest in adult dairy animals (Villeneuve *et al.* 2010), the present study has assessed, for the first time to our knowledge, how a feeding restriction in ewe lambs may influence the response of these animals to an experimental inflammatory challenge of the mammary gland performed at their first lactation. Locally, the intramammary inoculation of LPS described in this report induced an inflammatory response within the inoculated mammary glands, characterized by a remarkable increase in SCC. A very light increase of SCC was also observed in the control udder. After 72 hours, the SCC reached its basal levels in both udders. Our results agree with those presented by Rainard *et al.* (2013) in cows, who found that the infusion of inflammatory compounds such as LPS and ovalbumin into the mammary gland induced an inflammatory response that lasted no longer than 72 h. Likewise, the systemic responses here described, based on the T^{ra} and some plasma biomarkers (especially r_IL-6, r_IL-36RA and r_IL-10), also agrees with that reported by others researchers (Bochniarz *et al.* 2017; Naylor *et al.* 2020). The results of the GLM-analysis here presented (Table 1) showed that most time points were significant for the T^{ra} trait as previously reported by Naylor *et al.* (2020) in a group of female lambs after a systemic challenge with an intravenous bolus of LPS. The increase of body temperature or fever is considered a cardinal response to infection and is thought to be driven by some proinflammatory cytokines, particularly IL-1, IL-6, and TNF-α (Naylor *et al.* 2020). In our work, the r_IL6 ratio trait was the biomarker that showed the highest relative concentration increase relative to the basal point (-24h), showing the most significant results between the 4h to 8h time points post-LPS- injection. Among the tested anti-inflammatory cytokines, the IL-10 was highly significant at 4h and 6h time points post LPS-injection and significant at the 72h time point where the levels of this protein decreased. Other studies have suggested that IL-10 is produced at a later time than the proinflammatory cytokines and regulates the course of inflammation (Bochniarz *et al.* 2017). IL-36RA, another anti-inflammatory cytokine, is known

to inhibit the effects of Interleukin-36 cytokines and thereby inhibit their proinflammatory effects (Towne *et al.* 2004). The concentration of the r_IL-36RA oscillated throughout the experiment, but only the time points 2h, 4h, and 8h were significant for this cytokine. Regarding the concentration of the VEGF-A, which is a master regulator of vascular homeostasis and angiogenesis (Pandey *et al.* 2018), it varied over the experiment, showing significant time points at 4h and 6h (P-value<0.01), but also within the 12-24h time interval (P-value<0.05). Lastly, of the chemokines studied here, r_IL-8 was the most remarkable showing a significant increase in their concentration at 24 and 48 hours after LPS injection, coinciding with the review of Kaplanski *et al.* (2003), who reported that when neutrophils or other cells are stimulated by inflammatory cytokines, IL-8 is usually produced over 24h, recruiting and activating more neutrophils locally. Regarding the results of the GLM analysis in relation to the NC/C groups effect on the local inflammation response (Table 1), our results suggest that energy restriction at the prepuberal age did not modify the local inflammatory response induced by the LPS inoculation, as the effect of animal groups (NC and C) was not significant on the SCC ratio traits. In terms of systemic response, the NC/C group effect was significant for all the systemic traits considered in this study except r_IL-6 and the chemokine r_IL-8. However, considering the results observed for the time points across the experiment, we consider that only some of the studied markers, such as IL-10, IL-36RA and the modulator VEGF-A could be of interest to further assess the potential effect of a protein restriction on the systemic response to the inflammatory challenge of the mammary gland later in life. The preliminary results reported here settle the basis to better understand how the nutritional plane at early life can influence the local and systemic inflammatory responses in dairy sheep.

Acknowledgements

This study has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 772787 (*SMARTER*) and the Spanish Ministry of Science and Innovation through the project RTI2018-093535-B-I00 (*EPIMILKSHEEP*).

References

Bochniarz, M., Zdzisińska, B., Wawron, W., Szczubiał, M., Dąbrowski, R. (2017). J. Dairy Sci. 100, 9674–9680. https://doi.org/10.3168/JDS.2017-13552

Ford, J.A., Park, C.S. (2001). J. Dairy Sci. 84, 1669–1678. https://doi.org/10.3168/jds.S0022- 0302(01)74602-4

Kaplanski, G., Marin, V., Montero-Julian, F., Mantovani, A., Farnarier, C. (2003). Trends Immunol. 24, 25–29. https://doi.org/10.1016/S1471-4906(02)00013-3

Naylor, D., Sharma, A., Li, Z., Monteith, G., Sullivan, T., *et al.* (2020). J. Dairy Sci. 103, 5501– 5508. https://doi.org/10.3168/JDS.2019-17718

O'Rourke, D. (2009). Irish Vet. J. 2009 624 62, 1–6. https://doi.org/10.1186/2046-0481-62-S4- S15

Pandey, A.K., Singhi, E.K., Arroyo, J.P., Ikizler, T.A., Gould, E.R., *et al.* (2018). Hypertension 71, E1–E8. https://doi.org/10.1161/HYPERTENSIONAHA.117.10271

Sejrsen, K., Huber, J.T., Tucker, H.A., Akers, R.M. (1982). J. Dairy Sci. 65, 793–800.

Sejrsen, K., Purup, S., Vestergaard, M., Foldager, J. (2000). Domest. Anim. Endocrinol. 19, 93– 104. https://doi.org/10.1016/S0739-7240(00)00070-9

Statistics, I.S. (2019). IBM SPSS Statistics for Windows, Version 26.0. Google Search. Towne, J.E., Garka, K.E., Renshaw, B.R., Virca, G.D., Sims, J.E. (2004). J. Biol. Chem. 279, 13677–13688. https://doi.org/10.1074/JBC.M400117200

Turk, R., Koledić, M., Maćešić, N., Benić, M., Dobranić, V., *et al.* (2017). Mljekarstvo 67, 91–101. https://doi.org/10.15567/mljekarstvo.2017.0201

Villeneuve, L., Cinq-Mars, D., Lacasse, P. (2010). Animal 4, 944–950. https://doi.org/10.1017/S1751731110000170

734. Effects of breeding for scrapie resistance in Bavarian sheep breeding populations

J. Buitkamp* and J. Semmer

Bavarian State Research Center for Agriculture, Institute of Animal Breeding, 85586 Grub, Germany; johannes.buitkamp@lfl.bayern.de

Abstract

Scrapie is a fatal disease of small ruminants that is not transmitted to humans. The disease became the focus of research and policy when it was shown that sheep infected with BSE develop symptoms indistinguishable from scrapie, and it could not be ruled out that humans could be infected via this route. Among other measures, breeding programs were established to improve resistance to scrapie infection. These were based on selection for a single genotype (PRNP). Here we investigate the effects of the breeding program on PRNP allele frequencies and the impact of the selection for resistance in Bavarian herdbook sheep. We show that selection was successful and no severe side effects occurred.

Introduction

Scrapie is a fatal, neurodegenerative disease of sheep and goats. It has been known for nearly 300 years and is characterized by the presence of vacuole formation in neurons and abnormally shaped prion protein molecules. The nature of the infectious agent and the mechanisms leading to pathology are still not fully understood. Prusiner studied and described this entirely new pathogen species (Prusiner 1982) and introduced the term 'prions,' i.e. nucleotide-free, small proteinaceous pathogens that are highly resistant to environmental influences. Meanwhile, the family of diseases caused by prions is called transmissible spongiform encephalopathies (TSEs). TSEs include, for example, chronic wasting disease (CWD) in deer and elk and Creutzfeldt-Jakob disease in humans.

In 1987, a disease with similar pathology had been described in cattle in Great Britain (Wells *et al.* 1987). Wells *et al.* already suspected a pathogen related to that proposed for scrapie. The disease attracted considerable attention when it was found to cause a fatal TSE disease when transmitted to humans (vCJD, Hill *et al.* 1997), triggering a wave of research activity and regulatory control measures. Eventually, the disease was successfully controlled by interrupting the oral route of infection in cattle and removing the at-risk material from the carcass. However, because it could not be ruled out that some scrapie cases were caused by the BSE prion and thus posed a potential threat to human health, TSE in sheep and goats also became the focus of regulatory action. These were TSE surveillance programs and breeding for TSE resistance.

It has been recognized that there are genetic differences in susceptibility to scrapie in sheep, both between and within breeds. Later on it was shown that susceptibility was largely related to variants of the PRNP gene in sheep (Goldmann *et al.* 1994). These results allowed very efficient breeding for scrapie resistance. Therefore, in 2002, selective breeding programs in sheep were initiated in the EU in addition to other TSE control measures. Different breeding programs have been implemented depending, for example, on the endangerment status of the breed and the frequency of resistance/susceptibility PRNP alleles. In effect, these lead to sharp selection according to the genotype of one particular gene with possible side effects such as limiting the number of animals available for breeding and associated effects on performance or health traits. Here we investigated the effect of PRNP selection in Bavarian sheep. We monitored the change of allele frequencies from the year 2002 to 2008 by genotyping more than 2000 herdbook animals for the PRNP gene and investigated potential side effects.

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_734, © J. Buitkamp and J. Semmer 2022

Materials & methods

DNA Samples. EDTA-Blood samples from 2286 herdbook registered Bavarian sheep were collected in the years 2002 to 2008 in the course of routine parentage control. DNA was isolated by using the E.Z.N.A. Blood DNA Mini Kit (#12-3482-03, PEQLAB Biotechnologie GmbH, Erlangen, Germany) and stored at -20 °C. The largest number of samples (1825) was from Merinolandschaf (MLS), a smaller number of samples was from Suffolk (SUK), Schwarzköpfiges Fleischschaf (SKF), and Alpines Steinschaf (AST) with 146, 258, and 57 samples.

Genotyping. All DNA samples were genotyped using a multiplex ARMS system that allows rapid genotyping of PRNP amino acid positions 136, 154, 171 and the determination of haplotypes in the case of double heterozygote individuals (Buitkamp and Semmer 2004). ARMS primers were designed for multiplex allele specific amplification. 5' Primers for alleles 136_A and 136_V were fluorescence labeled (TET and Fam) enabling the evaluation of fragments on an ABI PRISM® 310 genetic analyser (AppliedBiosystems, Foster City, CA) and automated genotype designation using the the Genotyper® software (V. 2.5, Applied Biosystems).

Results & discussion

PRNP allele frequencies for all 2286 sheep are summarized in Table 1 for 2002-03 and from 2004 onward. The baseline frequency at the beginning of the scrapie resistance breeding program for the 'resistance allele' ARR was relatively low in MLS (0.2), while it was already very high in SUK and SKF (0.82, 0.83). This high prevalence corresponds to the results for most meat sheep breeds (Drögemüller *et al.* 2001). In the three economically relevant breeds, ARQ and ARR were the dominant alleles. The 'susceptibility allele' VRQ either did not occur or occurred with very low frequency. In the endangered sheep breed Alpine Stone Sheep (AST, listed on the Red List of the Federal Agency for Agriculture and Food (BLE) under 'conservation status', which means that an effective population size between 50 and 200 animals exists) all 5 alleles including ARH occur. Accordingly, the selection based on the PRNP genotype led to an increased frequency of the ARR allele in MLS, whereas in SUK and SKF the high frequencies were maintained. In AST there was only a moderate frequency increase for ARR, probably because the breed specific selection program is adapted to the situation for small populations.

Figure 1 shows the PRNP allele frequencies from year 2002 to 2017 for all MLS and the MLS breeding rams. The allele frequency of ARR rises from below 0.1 to above 0.6 within 2 years when the resistance breeding program starts. A slightly more moderate increase is observed in the genotypes from the whole MS population. This could be expected since the PRNP based selection program operates mainly on the rams. From the year 2004 on the allele frequencies stabilizes at 0.65 for ARR. This reflects the governmental

Table 1. *PRNP* allele frequencies summarized for the years 2002-2003 and as of 2004 for four Bavarian sheep breeds.

Breed	MLS		SUK		SKF		AST	
Year	2002-2003	2004-	2002-2003	2004-	2002-2003	2004-	2002-2003	2004-
Allele								
ARR	0.20	0.45	0.82	0.75	0.83	0.83	0.36	0.56
ARQ	0.71	0.51	0.13	0.23	0.16	0.013	0.40	0.32
ARH	0.00	0.00	0.00	0.00	0.00	0.00	0.34	0.07
AHQ	0.09	0.04	0.00	0.00	0.00	0.03	0.04	0.05
VRQ	0.00	0.00	0.06	0.01	0.01	0.01	0.08	0.00

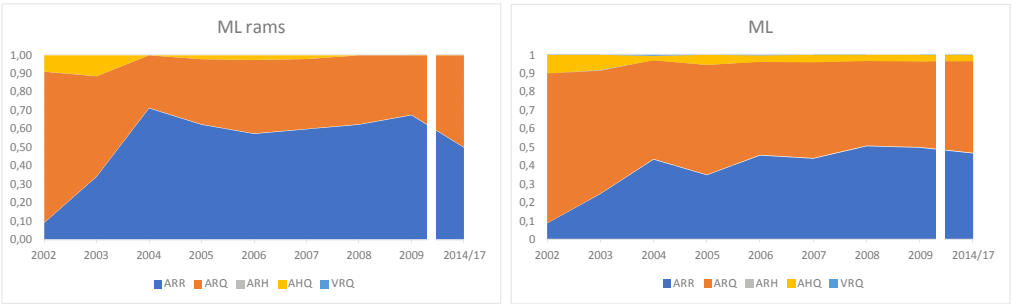


Figure 1. *PRNP* allele frequencies from all MLS sheep (bottom) and MLS breeding rams (above) in the period 2002-2017.

regulations that allow the use of rams carrying ARR/ARR or ARR/ARQ, or ARR/not VRQ respectively. These data show that the selection using a single gene genotype works highly efficient as expected.

In Germany, the numbers of BSE and scrapie cases sharply dropped with the governmental measures (ban of risk material from animal nutrition, TSE surveillance, etc.) in place and the start of breeding for scrapie resistance (Figure 2).

The side effects of the selection program are more difficult to investigate, and data are limited. By direct investigation of the allele effect no main adverse associations were found (e.g. Prokopová *et al.* 2002; Buitkamp *et al.* 2005; Vitezica *et al.* 2006; Pritchard *et al.* 2008).

In 2002 and 2003, the exclusion of MLS rams from breeding due to undesirable PRNP genotypes reduced the number of herd stock rams for selection on other traits. Therefore, it is expected that genetic progress in these years may be limited. Table 2 shows the mean values for daily gain and dressing out percentage for the test station progeny groups from the genotyped MLS rams. The values from 2002, the year with the strongest selection effect, are lower than those from all other years. This may indicate that there has been a negative effect on breeding progress. In this case, the adverse effect would be mainly due to the restriction of the selection of breeding animals, but not to undesirable effects of the ARR allele. There are few examples of sharp selection based on alleles of one gene in breeding animals. Prominent examples are selection for

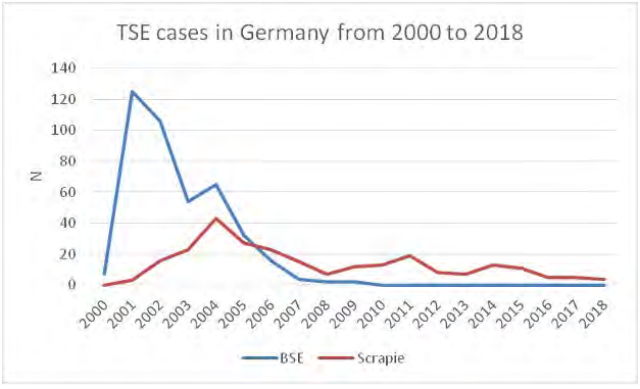


Figure 2. Number of BSE and Scrapie cases as registered by the Bundesministerium für Ernährung und Landwirtschaft ('Tierseuchenbericht').

Table 2. Means from the average values per year of the progeny groups from the progeny test station in Grub https://www.lfl.bayern.de/itz/schaf/031163/index.php

Year	Daily gain [g/d]	Dressing out %
2001	410.93	46.92
2002	397.98	46.69
2003	411.51	47.26
2004	435.72	48.19
2005	445.17	48.29
2006	431.00	48.32
2007	415.99	49.27
2008	451.34	47.80

MHS status in pigs or selection against Weaver in Brown Swiss, both of which led to a decrease in breeding progress in important traits.

Conclusions

The scrapie breeding program was successful and there could be a small side effect on economic traits due to the reduced number of breeding rams.

References

Buitkamp J., C. Mendel and K.-U. Götz (2005) DGfZ-Schriftenreihe 39:55-60.

Buitkamp J. and J. Semmer (2004) BMC Infect Dis 4(1):30.

Drögemüller C., T. Leeb and O. Distl (2001) Vet Rec 149(12):349-352.

Goldmann W., N. Hunter, G. Smith, J. Foster and J. Hope (1994) J Gen Virol 75(Pt 5):989-995.

Hill A.F., M. Desbruslais, S. Joiner, K.C. Sidle, I. Gowland, *et al.* (1997) Nature 389(6650):448-450, 526.

Pritchard T.C., C.M. Cahalan and I. Ap Dewi (2008) Animal 2(10):1421-1426.

Prokopová L., R.M. Lewis, W.S. Dingwall and G. Simm (2002). Scrapie genotype: a correlation with lean growth rate? 7th World Congress on Genetics Applied to Livestock Production, August 19-23, 2002, Montpellier, France.

Prusiner S.B. (1982) Science 216(4542):136-144.

Vitezica Z.G., C.R. Moreno, L. Bodin, D. Francois, F. Barillet, *et al.* (2006) J. Anim Sci. 84(6):1317-.

Wells G.A., A.C. Scott, C.T. Johnson, R.F. Gunning, R.D. Hancock, *et al.* (1987) Vet Rec 121(18):419-420.

735. Options for the selective breeding for resistance to gastro-intestinal nematodes of Sarda breed sheep

S. Salaris[1*], S. Casu[1], M.G. Usai[1], A. Scala[2] and A. Carta[1]

[1]Research Unit: Genetics and biotechnology, AGRIS Sardegna, 07040 Olmedo, Italy; [2]Department of Veterinary Medicine, University of Sassari, 07100 Sassari, Italy; slsalaris@agrisricerca.it

Abstract

In this paper we summarize results of previous studies on the resistance to gastro-intestinal nematodes obtained in a nucleus of Sarda ewes in terms of heritability estimate, QTL detection, functional annotation, association analysis based on imputed sequence data and prediction of GEBV for male selection candidates of the Herd Book. Heritability of log-transformed faecal egg counts was 0.21±0.015. Ten QTL regions were defined on nine chromosomes with an LDLA approach. In the most significant QTL regions, we identified functional candidate genes that contain high or moderate impact polymorphisms. The expected accuracies of GEBV are related to the amount of relationships with the nucleus. The Sarda breeders are assessing the feasibility of a selection program for nematode resistance based on fecal egg counting and on the genotypes in the nucleus flock and of selection candidate males bred in Herd Book farms genetically connected with the experimental flock.

Introduction

Gastrointestinal nematodes (GIN) are a major health problem in grazing animals that causes important yield reductions and increased production costs due to medical treatments and higher culling rates (Mavrot *et al.*, 2015). Fecal egg count (FEC), i.e. the number of parasite eggs per gram of faeces has been largely used as proxy trait to measure resistance to GIN. In Sardinia, sheep farming systems are based on grazing natural pastures and forage crops where GIN infection is unavoidable. Thus, most farmers administrate anti-helminthics often with an empirical approach in terms of individual diagnosis, doses and frequency of treatments (Sechi *et al.*, 2010). Genetic variation between individuals and breeds has been documented (Aguerre *et al.*, 2018). However, the inclusion of GIN resistance in current breeding scheme is difficult due to the laboriousness of the FEC recording at Herd Book (HB) scale. For the Sarda breed, it is available a female nucleus in an experimental farm in which the ewes generated by HB rams, have been pedigree recorded, genotyped and repeatedly measured for FEC for twenty years. The genetic links of these animals with those of the HB make it possible to use their data to study the genetic determinism of resistance to GIN at breed level. In this paper, we summarize the results obtained in previous studies on this experimental population in terms of heritability estimate, QTL detection, functional annotation, association analysis based on imputed sequence data (Casu *et al.*, 2022) and prediction of GEBV for male selection candidates of the HB (Salaris *et al.*, 2021). Results will be discussed focusing on the feasibility of genetic improvement strategies based on genomic tools.

Materials & methods

Experimental population. The nucleus flock (FRP) was set in 1999 with the first generation of ewes consisting in 928 back-cross Lacaune × Sarda. The following generations were obtained by mating adult ewes with rams coming from the artificial insemination center of the HB. In total, 5,386 ewes from 218 rams (10 F1 and 208 Sarda) were generated until 2020. The farming system was similar to that commonly applied in Sardinia with most of the adult ewes lambing in autumn and yearlings lambing between January and March. The feeding regime was based on grazing of natural pastures and forage crops, supplemented by hay and concentrates in winter and late spring. From 2000 to 2020 individual FEC were recorded from 1 to 3 times per year on adult ewes, more frequently in September and July. After editing, 16,530 records from

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_735, © S. Salaris *et al.* 2022

4,097 adult ewes recorded on 24 separate dates were retained. FEC presented a skewed distribution and were log-transformed using lnFEC=ln(FEC+14) for genetic analyses. All ewes of the nucleus, their sires and available male ancestors (272 rams) were genotyped with the Illumina Inc. OvineSNP50 Beadchip. SNP editing was performed using call rate and MAF thresholds of 90 and 1%, respectively. After quality control, 43,390 SNP across 26 autosomes were retained for the analysis. For QTL mapping the ovine genome assembly v4.0 was used to construct the genetic map by assuming 1 Mb =1 cM.

Genetic parameter estimation and QTL mapping. lnFEC data were analysed by a repeatability animal model including the permanent environment and additive genetic random effects of ewes and using the ASReml-R 4.1 software (Butler *et al.*, 2018). Environmental fixed effects were the date of sampling, the age of the ewe (from 1 to 4 years) and the interaction between lambs born and physiological status (periparturient, lactating, non-pregnant nor lactating ewes) at the date of sampling. Genetic relationships between 4,547 animals, including the recorded ewes and their sires (194) and genotyped ancestors (78), were taken into account by calculating the genomic relationship matrix (GRM) based on 50k genotypes, following Yang *et al.* (2010). On average, the number of daughters per sire was 22±25 from 1 to 118. Pseudo-phenotypes for QTL detection were calculated as the average performance deviation (APD) of each ewe, by averaging single animal residuals and summing-up the genetic and permanent environment random predictions. QTL mapping was performed with an LDLA method as described by Usai *et al.* (2019). Only QTL segregating within gametes of Sarda origin were tested by F-test at each SNP position. The 5% genome-wide significance threshold was determined by a Bonferroni correction for the 43,390 tests, which resulted in $-\log_{10}(P\text{-value})=6$.

Analysis of sequence data. Identified QTL regions (QTLR) were imputed from whole-genome sequences of 56 target animals of the population. An association analysis and a functional annotation of imputed polymorphisms in the identified QTLR were performed to pinpoint functional variants with potential impact on candidate genes identified from ontological classification or differentially expressed in previous studies. The NCBI *Ovis aries* 4.0 reference genome and annotation (release 102) and the SnpEff software v4.3.t (Cingolani *et al.*, 2012) were used for the analyses.

Prediction of GEBV for selection candidate rams. Single step genomic BLUP (ssGBLUP) was performed using the package AIREMLf90 included in the BLUPf90 suite programs (Misztal *et al.*, 2002) using FEC and genotypes of the female nucleus including their sires and male ancestors. Genotypes of further 734 HB Sarda rams were also available. The final pedigree file included 8,565 animals. The genomic relationship matrix was created as in VanRaden (2008). GEBV were calculated for the 1,006 genotyped rams classified according to their age and level of relationship with the females of the nucleus: 272 sires or ancestors; 489 old HB rams with daughters in HB flocks; 245 young HB rams (including 27 sons of FRP) considered as selection candidates. The expected accuracy was calculated as: $r_{GEBVi}=\{1-[SE_i^2/(h_{ii}\times\sigma_g^2)]\}^{1/2}$, where SE_i is the standard error of prediction and h_{ii} is the diagonal element of the H relationship matrix for the animal i.

Results

The overall mean and standard deviation of FEC and lnFEC were 332±727 and 4.82±1.43 respectively. The heritability and repeatability estimates of lnFec were 0.21±0.015 and 0.27±0.012 respectively. In the LDLA analysis 202 SNPs encompassed the 5% genome-wide significant threshold. With the exception of OAR 1, on which only one significant location was found, many significant SNPs mapped to the same chromosome. After clustering the significant locations on the same chromosome, ten QTL regions were defined on nine chromosomes. The ratio between the ANOVA estimators of the QTL variance and the total phenotypic variance ranged from 0.0087 to 0.0176 (Table 1).

Table 1. QTL regions from LDLA analysis (Casu *et al.*, 2022).

OAR	Range (kb)[1]	Significant SNP (n)	Highest peak[1]		
			Pos (bp)	-$\log_{10}$(*P*-value)	s_{qtl}^2/s_{APD}^2
1	-	1	99988914	6.256	0.0087
2	-	1	135598790	6.078	0.0093
2	213,089-213,166	3	213135270	6.554	0.0093
4	5,252-9,837	6	5386849	7.404	0.0111
6	12,638-12,695	3	12695314	6.864	0.0099
7	88,040-88,085	2	88040710	6.762	0.0098
12	35,973-41,153	19	39430517	9.191	0.0125
15	34,024-34,081	2	34024406	7.362	0.0110
19	18,933-31,423	11	18983777	7.302	0.0104
20	16,797-36,098	154	21170858	12.861	0.0176

[1] Pos = position based on the ovine genome assembly v4.0; s_{qtl}^2/s_{APD}^2 is ratio between ANOVA estimator of QTL variance and total phenotypic variance.

The association analyses on imputed polymorphisms confirmed that the QTLs on OAR 4, 12, 19 and 20 were the most significant. None of the most significant SNPs showed a functional effect on gene transcription. However, in the most significant QTLRs, we identified genes that: contain high or moderate impact polymorphisms; have been found differentially expressed in previous studies; contribute to enrich the most represented GO process (regulation of the immune system process, defense response). Among them, the most likely candidate genes were: TNFRSF1B and SELE on OAR12, IL5RA on OAR19, IL17A, IL17F, TRIM26, TRIM38, TNFRSF21, LOC101118999, VEGFA and TNF on OAR20.

Concerning the GEBV prediction, r_{GEBV} of HB rams ranged between 0.31 and 0.95 with a mean of 0.76±0.15 for sires or ancestors of FRP; 0.49±0.06 for old HB rams and 0.50±0.08 for young HB rams. Among the selection candidates, the highest values of r_{GEBV} were obtained for 27 sons of FRP ranging from 0.62 to 0.73.

Discussion

The heritability estimate of lnFec was moderate and consistent with previous studies in adult ewes (Aguerre *et al.*, 2018; Sechi *et al.*, 2009). Five QTL regions (OAR4, 7, 12, 19, 20), identified in the LDLA analysis, overlapped with regions that were shown to be associated to traits related to GIN resistance by others (Riggio *et al.*, 2014). The combination of association analysis and functional annotation of the sequence data did not lead to identify putative causative mutations. Overall, these results show that selective breeding may be an option to limit the issues related to infestation of gastro-intestinal nematodes in sheep. On the one hand, the heritability estimate and QTL detection results confirm that both traditional progeny testing and marker-assisted selection are realistic options. However, the laboriousness of fecal egg counting on a large scale makes marker-assisted selection potentially more profitable in terms of cost-benefits. Indeed, the portion of the total phenotypic variance explained by the ten significant markers identified in our study and already available on the commercial Illumina, was about 11%. As far as the potential application of a genomic selection approach is concerned, our results show that the expected accuracies are strongly related to the amount of relationships with the nucleus. In this sense, Carta *et al.* (this congress) proposed a procedure based on pedigree information to predict the expected accuracies of the HB matings which can be used to design the flow of animals from and toward the nucleus to achieve good accuracies for most selection candidates. As a perspective, accuracies of genomic predictions may be also improved by including in the evaluation model the ten highly significant identified SNPs. In the short term, the Breeders Association is assessing the feasibility of a selection program for nematode resistance based on fecal egg counting and on the genotypes described in this study for the nucleus flock and combined with

the genotyping of selection candidate males that are bred in Herd Book farms and are genetically connected with the experimental flock.

References

Aguerre S., Jacquiet P., Brodier H., Bournazel J. P., Grisez C., Prévot F., *et al.* (2018). Vet Parasitol, 256, 16–23. https://doi.org/10.1016/j.vetpar.2018.04.004

Butler D.G., Cullis B.R., Gilmour A.R., Gogel B.J., Thompson R. (2018). Hemel Hempstead: VSN International Ltd; 2018. p. 188.

Casu S., Usai M.G., Sechi T., Salaris S., Miari S., *et al.* (2022) Genet Sel Evol 54(1) https://doi.org/10.1186/s12711-021-00690-7

Cingolani P., Platts A., Wang L.L., Coon M., Nguyen T. *et al.* (2012). *Fly*, 6(2) 80-92. https://doi.org/10.4161/fly.19695

Mavrot F., Hertzberg H., and Torgerson P. (2015). Parasit. Vectors. 8(1):557. https://doi.org/10.1186/s13071-015-1164-z

Misztal I., Tsuruta S., Strabel T., Auvray B., Druet T., *et al.* (2002). Proc. of the 7th WCGALP, Montpellier, France.

Riggio V., Pong-Wong R., Sallé G., Usai M.G., Casu S., *et al.* (2014). J Anim Breed Genet. 131:426–36 https://doi.org/10.1111/jbg.12071

Salaris S., Casu S., Usai M.G., Scala A. and Carta A. (2021). ICAR Techn Series no. 25,147-150

Sechi S., Giobbe M., Sanna G., Casu S., Carta A., *et al.* (2010). Small Rumin Res. 88(2–3), 145–150 https://doi.org/10.1016/j.smallrumres.2009.12.022

Sechi S., Salaris S., Scala A., Rupp R., Moreno C., *et al.* (2009). Ital J Anim Sci 8(2), 156-158 https://doi.org/10.4081/ijas.2009.s2.156

Usai M.G., Casu S., Sechi T., Salaris S.L., Miari S., et al (2019). *Genet Sel Evol* 51(65) https://doi.org/10.1186/s12711-019-0508-0

VanRaden P.M. (2008). J dairy sci 91(11):4414-4423. https://doi.org/10.3168/jds.2007-0980

Yang J., Benyamin B., McEvoy B.P., Gordon S., Henders A.K., Nyholt D.R., *et al.* (2010). Nat Genet; 42:565–9 https://doi.org/10.1038/ng.608

736. Investigation of the Scrapie haplotypes in the South African Merino population

J.J. Reding* and B.E. Mostert

SA Stud Book and Animal Improvement Association, 118 Henry Street, Westdene, 9300, Bloemfontein, South Africa; jason@studbook.co.za

Abstract

Scrapie, a transmissible spongiform encephalopathy disease is described as a fatal, degenerative disease that affects the nervous system of sheep and goats and is considered an endemic disease in the European Union. Breeding programs for genetic improvement of Scrapie resistance have resulted in the identification of multiple haplotypes in the Prion protein gene that can be selected for. In this study, the occurrence and frequency of Scrapie haplotypes that are known to show varying levels of susceptibility were determined using genomic information obtained from the Illumina Ovine v2 array in 1,061 Merino sheep across five flocks in South Africa. Five haplotypes were observed in the South African Merino genotyped population and include ARQ (46.1%), ARR (31.48%), AHQ (14.47%), VRQ (6.83%) and ARH (0.61%). Of the genotyped animals, 87% have slight to high genetic resistance to Scrapie, while 13% are genetically susceptible to the disease, as they carry the VRQ haplotype.

Introduction

Scrapie is a transmissible spongiform encephalopathy (TSEs) disease, like bovine spongiform encephalopathy (BSE), i.e. a fatal, degenerative disease that affects the nervous system of sheep and goats (Acín *et al.*, 2004). Scrapie develops due to an accumulation of the Prion protein (PrP) in neuronal cells and is highly transmissible from animal to animal. This disease was first described in the UK around the 18th century. Affected animals show progressive changes in behaviour such as body tremors, stumbling and may have irritated skin which they try to scratch, resulting in excessive wool loss and/or damage to the skin.

Scrapie is considered an endemic disease in the European Union, with first incidences reported in 1878 in Iceland and has subsequently been reported all over the world in Canada, the USA, Australia, Norway, India, South Africa, Kenya, Germany, Brazil, Yemen and Sweden (Yaman and Ün, 2017) throughout the 20th century.

Breeding programs for genetic improvement of Scrapie resistance have resulted in the identification of multiple haplotypes in the PrP gene (Acín *et al.*, 2004). Studies in the Merino sheep breed have been done to identify the frequency of these haplotypes segregating in certain populations (Wiśniewska and Mroczkowski, 2010; Stepanek and Horin, 2017; Palomino *et al.*, 2019).

Haplotype assessment of animals that tested positive for classical Scrapie has indicated low occurrence rates in animals carrying the ARR haplotype (0-2.21%), while the majority of the positive animals carried the ARQ and VRQ haplotypes (13.31-43.78%; Yaman and Ün, 2017).

According to Cosseddu *et al.* (2007), the last Scrapie incidence in South Africa was reported in 1972. This is confirmed by the Animal Disease Status in South Africa Report of the Department of Agriculture, Forestry and Fisheries in 2015. In this study, the occurrence and frequency of Scrapie haplotypes that are known to show varying levels of susceptibility were determined from genomic information obtained from the South African (SA) Merino population.

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_736, © J.J. Reding and B.E. Mostert 2022

Materials & methods

Blood samples from 1,106 sheep from five flocks (2 strong-wool, 1 fine-wool, 1 polled and 1 horned) of the SA Merino population were genotyped using the Ovine Illumina v2 array. During post-genotyping quality control, all genotypes with an individual call rate of less than 0.9 were excluded, leaving a total of 1,097 genotypes.

Single nucleotide polymorphisms (SNP) located in codons 136, 154 and 171 of the PrP gene known to be responsible for causing classical Scrapie were annotated using ENSEMBL. This resulted in a total of six SNPs being analysed to create the three amino acids which encode the haplotypes that have been derived for Scrapie resistance. As it required all six SNPs to be called for a successful haplotype consolidation, this excluded thirty-six animals and therefore a final population of 1,061 SA Merino sheep were available for the study. The number of genotyped animals per flock was 122 animals for Flock 1 (strong-wool), 248 animals for Flock 2 (strong-wool), 238 animals for Flock 3 (horned), 236 animals for Flock 4 (fine-wool) and 217 animals for Flock 5 (polled). Haplotype pairings were grouped under five distinct levels of resistance in accordance with the study of Palomino *et al.* (2019). This grading system ranks haplotypes with resistance level 1 (R1) being highly resistant to developing Scrapie to resistance level 5 (R5) being the haplotypes that are highly susceptible to developing Scrapie as described in Table 1.

Results & discussion

The frequencies of the five haplotypes observed in the SA Merino genotyped population are indicated in Figure 1. These include ARR (31.48%), AHQ (14.47%), VRQ (6.83%), ARH (0.61%) and ARQ (46.1%). This is in accordance with the ARQ haplotype being considered as the 'wild type'. The ARH haplotype is quite scarce (0.6%), while the highly susceptible VRQ haplotype is present in 6.8% of the genotyped animals.

Figure 2 illustrates the number of animals for the different Scrapie haplotype pairings. One hundred and seventeen of the animals (11%) carry the highly resistant haplotypes (R1), 384 (36%) animals carry the resistant haplotypes (R2), 420 animals (39.5%) carry the haplotypes that render slight resistance (R3), 50

Table 1. Grading system for classification of Scrapie resistance and susceptibility genotypes adapted from Palomino *et al.* (2019).

Level	Genotype	Outcome
1	ARR / ARR	Most resistance to Scrapie.
2	AHQ / ARR	Resistance to Scrapie. Careful selection needed in breeding.
2	ARR / ARH	
2	ARR / ARQ	
3	AHQ / AHQ	Slight resistance to Scrapie. Careful selection needed in breeding.
3	AHQ / ARQ	
3	AHQ / ARH	
3	ARQ / ARQ	
3	ARQ / ARH	
3	ARH / ARH	
4	ARR / VRQ	Susceptible to Scrapie. Should not be used for breeding.
5	AHQ / VRQ	Highly susceptible to Scrapie. Should not be used for breeding.
5	ARH / VRQ	
5	ARQ / VRQ	
5	VRQ / VRQ	

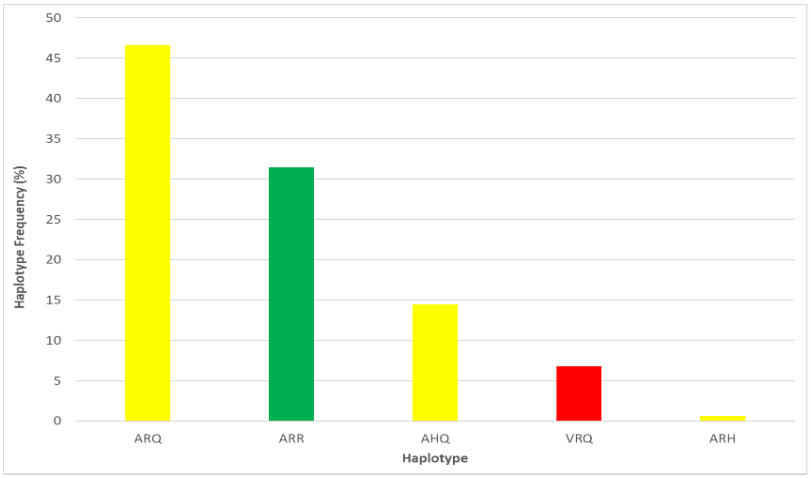


Figure 1. Frequency of the different haplotypes involved in Scrapie resistance and susceptibility of the South African Merino breed.

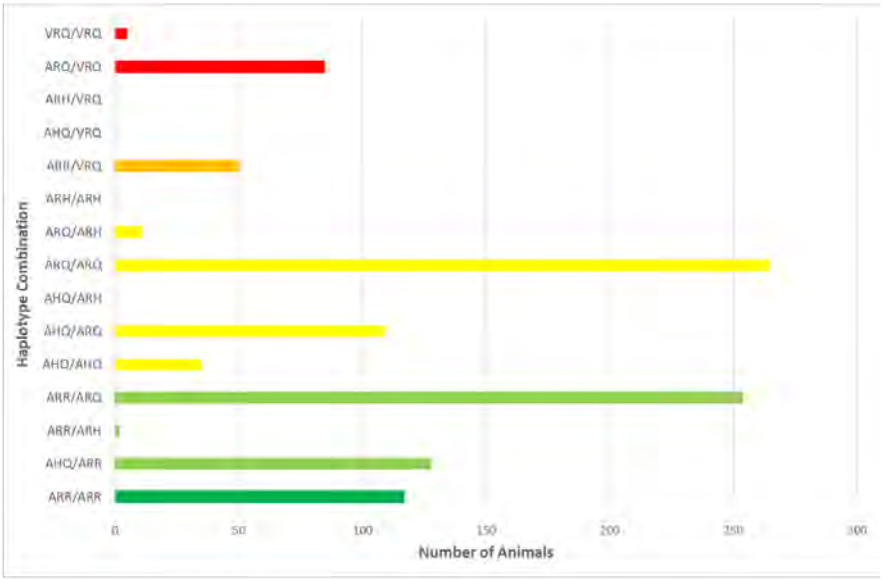


Figure 2. Number of genotyped animals with regards to the different haplotypes for Scrapie resistance and susceptibility.

animals (5%) carry the susceptible haplotypes (R4) and 90 (8.5%) carry the highly susceptible haplotypes (R5). No haplotype pairings for ARH/VRQ, AHQ/VRQ, AHQ/ARH and ARH/ARH were identified. This might be due to the small sample size of animals being genotyped at the time of the study, with representation from only five flocks in South Africa.

The distribution of the genotyped animals with regards to resistance levels over flocks are indicated in Table 2. From Table 2, all flocks have animals that are carriers of the highly susceptible haplotypes. This is an early

indication that these unwanted haplotypes are already commonly distributed across the flocks of the SA Merino population.

Looking at Figure 3 and in accordance with the grading system adapted from Palomino *et al.* (2019), 87% of the genotyped animals have slight to high genetic resistance to Scrapie, while 13% are genetically susceptible to the disease. Variation in risk of Scrapie infection is evident in the SA Merino breed, with 13% of the genotyped animals carrying haplotypes that are associated with susceptibility to high susceptibility for the disease. The exact susceptibility of the SA Merino breed still needs to be established and will only be possible once Scrapie incidences are being reported in South Africa. It is also important to consider that genetic factors other than the PrP gene might be involved in modulating animals' susceptibility to Scrapie in different populations. This might explain why Scrapie incidences in the South African population are not observed, although the highly susceptible haplotypes are present in the population. However, when Scrapie arises in a flock, the animals that are most susceptible will be those carrying the VRQ/VRQ haplotypes.

It is therefore recommended to the South African Merino breeders to consider a strategy where all stud rams are genotyped and all carriers of the VRQ haplotype are excluded from breeding. Furthermore, encouraging breeders to only use ARR/ARR rams in their flocks will increase the frequency of the highly resistant haplotype in the population. This may, however, have an adverse effect on the genetic diversity of the breed, which is also of importance.

Table 2. Distribution of genotyped animals over flocks with regards to resistance levels for Scrapie.

Level	Flock 1		Flock 2		Flock 3		Flock 4		Flock 5	
	Nr	%	Nr	%	Nr	%	Nr	%	Nr	%
Highly Resistant (R1)	28	23.0	8	3.2	53	22.3	7	3.0	21	9.7
Resistant (R2)	49	40.2	65	26.2	102	42.9	83	35.2	85	39.2
Slightly Resistant (R3)	28	23.0	95	38.3	51	21.4	141	59.7	95	43.8
Susceptible (R4)	10	8.2	25	10.1	19	8.0	1	0.4	5	2.3
Highly Susceptible (R5)	7	5.7	55	22.2	13	5.5	4	1.7	11	5.1

Figure 3. Percentage of genotyped animals with regards to the various levels of Scrapie resistance and susceptibility.

References

Acín, C., Martín-Burriel, I., Goldmann, W., Lyahyai, J., Monzón, M., *et al.* (2004) J. Gen. Virol. 85:2103–2110. https://doi.org/10.1099/vir.0.80047-0

Cosseddu, G.M., Agrimi, U., Pinto. J., & Schudel, A.A. (2007) Rev. Sci. Tech. Off. Int. Epiz. 26(3):657-668.

Palomino, P., Calle, L., Proença, J., Maestre, E., & Fernández-Garcia, J. L. (2019) Small Rumin. Res. 173:36–41. https://doi.org/10.1016/j.smallrumres.2019.02.010

Stepanek, O., & Horin, P. (2017) J. Appl. Genet. 58:111–121. https://doi.org/10.1007/s13353-016-0354-5

Wiśniewska, E., & Mroczkowski, S. (2010) Rocz. Nauk. Pol. Tow. Zootech. 6:115–123.

Yaman, Y., & Ün, C. (2017) Prion - An Overv. https://doi.org/10.5772/66916

737. Toward a genetic evaluation for functional longevity in French dairy sheep breeds

D. Buisson[1*], J.M. Astruc[1], L. Doutre[2] and I. Palhière[2]

[1]Institut de l'Elevage, BP 42118, 31321 Castanet-Tolosan, France; [2]GenPhySE, Université de Toulouse, INRAE, ENVT, 31326 Castanet Tolosan, France; diane.buisson@idele.fr

Abstract

The increase in genetic gain due to genomic selection in French dairy sheep breeds represents an opportunity to introduce new traits in selection objectives. As breeders have selected pure local breeds in their respective production area, robustness appeared to be an important criterion to improve adaptation to breeding systems. Considering functional longevity reflects the probability of a ewe to be culled for other criteria than production level, we have chosen to study Length of Productive Life, which is the number of days between the first lambing of the ewe and its culling. The heritabilities calculated using the Survival Kit are similar to those obtained in cattle and ranged from 0.07 to 0.16. Estimated breeding values were calculated but 30 to 40 daughters uncensored are necessary to reach an accuracy of 0.7. These first results are encouraging; however, this study should be continued to integrate functional longevity in selection objectives.

Introduction

In France, genomic selection implemented since 2015 has permitted an extra genetic gain compared to classical programs (Astruc *et al*., 2022). This extra gain represents an opportunity to introduce new traits in selection objectives. To face new challenges of sustainability and improve resilience and adaptation of ewes to breeding systems mainly based on pasturing, transhumance and rangelands, breeding organisations were interested in functional longevity. Indeed, functional longevity is a synthetic trait reflecting the probability of a ewe to be culled for other criteria than production level (Ducrocq 1992). As culling in dairy sheep breeds is mainly due to production level (20-25%), mammary problems (25-30%) and reproduction problems (10-20%), integrating functional longevity into selection objectives would allow breeders to have a better control of culling (economic interest) by mitigating involuntary culling, and therefore become more efficient to manage and improve their flocks (technical interest). Among several selection criteria of functional longevity, we have chosen to study the Length of Productive Life (LPL), which is defined by the number of days between the first lambing of the ewe and its culling. It is a simple indicator which doesn't require any additional measurement. The aim of this study is to estimate genetic parameters of functional longevity in French dairy sheep breeds in order to implement a genetic evaluation in routine.

Materials & methods

Data. This study focused on four dairy sheep breeds: Lacaune (LAC), Manech Tête Noire (MTN), Manech Tête Rousse (MTR), and Basco-Bearnaise (BB). For each breed, pedigree, reproductive and flock information (insemination, lambing, animal movements, flock's size) were extracted from the French national database SIEOL. Only ewes born in selection flocks between 2010 and 2020 and with a known sire were kept for the analysis.

Censoring. Survival analysis allows to consider a proportion of ewes whose end of lifetime is unknown: these are called 'censored data'. Ewes still alive on the 31st of August of 2020 were considered as censored at that date. Records of ewes with more than 6 lactations, who quit milk recording, who are sold to other flock, or who are in flock with a size increasing or decreasing by more than 50% in one year were considered as censored. As culling or censoring dates are not always recorded, it was decided to assign a new date based

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_737, © D. Buisson *et al.* 2022

on the last known event of each animal (lambing or not, milk production or not). LPL is then defined as the number of days between first calving and culling or censoring date.

Model. Taking these censored data into account requires the use of survival analyses. These analyses were carried out using the 'Survival Kit' (SK) developed by Ducrocq and Sölkner (1994) for animal populations under selection.

A Weibull model was used (Ducrocq *et al*., 2005; Terawaki *et al*., 2006):

$$h(t) = h_{0,n}(\tau) * \exp\{ \sum_m f_m(t) + hy_k(t) + s_i + 0.5\, mgs_j \} \quad (1)$$

$h_{0,n}(\tau)$ is the baseline hazard function assuming a different hazard function for nth lactation. τ denotes the number of days since the most recent lambing.

$\Sigma_m f_m(t)$ represents the sum of all fixed effects, time-dependant (t) or not, affecting the instantaneous culling risk. Considering these effects is key to focus on functional longevity rather than true longevity and correct for voluntary culling (milk production, flock strategy, etc.). The effects in the model are the age at first lambing (1, 2 or 3 years-old), the within-flock milk production (t – 11 classes), the flock size by variation in flock size interaction effect (t – 3 or 4 classes of flock size depending on the breed considered, and 3 classes of flock variation size ([-50%;-5%[, [-5%;5%],]5%;50%])), and the litter size (t – 0, 1, 2 or 3 lambs).

$hy_k(t)$ is the random log-gamma distributed flock-year effect. A single fixed parameter γ was assumed.

$s_i + 0.5\ mgs_j$ represents the additive genetic contribution of sire i and maternal grand-sire j of the ewe.

Heritabilities have been estimated using the formula of Mészáros *et al.* (2010):

$$h^2\ (sire - mgs) = \frac{4 * \sigma_s^2}{\frac{5}{4}\sigma_s^2 + \frac{1}{p} + \psi(\gamma)} \quad (2)$$

Where σ_s^2 is the sire's genetic variance, p is the percentage of uncensored data, and $\psi(\gamma)$ the variance of random effect herd×year.

Evaluation. Estimated breeding values (EBV) and coefficient of determination (CD) of all sires in our pedigree files were evaluated with the SK. EBV correlations between EBVs were also calculated to assess the relationship between LPL and the other traits under selection for all these sires. Finally, survival function of the best (top 10% – LPL+) and worse (bottom 10% – LPL-) rams, whose reliabilities were above 0.5, have been represented in order to validate our results.

Results & discussion

Data structure. Descriptive statistics for each breed are presented in Table 1. The number of ewes strongly varies according to the breed. The proportion of censored records ranges from 38.8 % in LAC to 51.1% in MTN.

Factors influencing productive life. Figure 1 shows relative risk ratios for three effects estimated in our model. Risk ratios for the different breeds are very similar for all effects. Risk ratios are higher when age at first lambing increases. Very high-risk ratios are found for the lowest milk production classes. Higher risks ratios were also observed for larger litter size. However, the risk ratio is much higher for ewes without lambs in LAC compared to the other breeds. Indeed, due to differences in sexual precocity, AI fertility, population

Table 1. Descriptive statistics of dataset.

	BB	MTN	MTR	LAC
Number of ewes	26,683	8,431	114,691	400,958
Sires in pedigree	892	399	2,851	4,595
Censored records (%)	50.7	51.1	43.3	38.8
Average LPL (days)	1,032	1,086	1,065	915

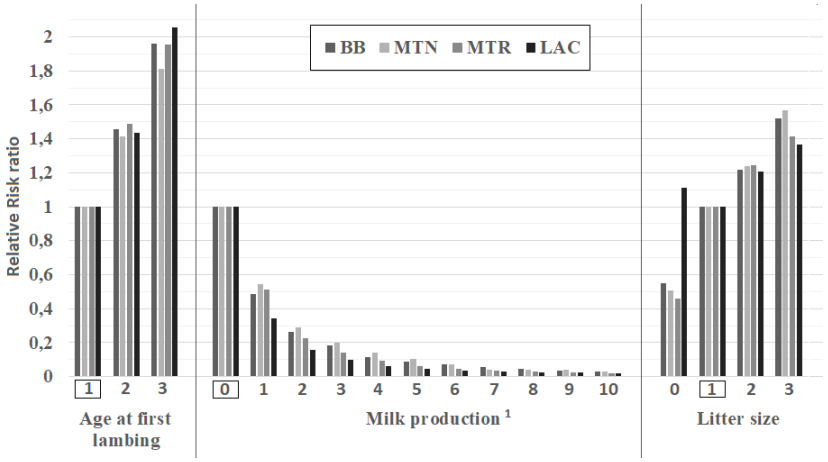


Figure 1. Relative risk ratios for three effects in the model. For each trait, these relative risk ratios are calculated in comparison to a reference level (surrounded value). Milk production: 11 classes from 0 (lower milk yield) to 10 (higher milk yield).

size and system management, culling non-pregnant ewes is more systematic in Lacaune than in Pyrenean breeds.

Genetic parameters. Table 2 shows the genetic parameters for LPL. Heritabilities range between 0.07 to 0.16 depending on the breed. These values are close to those obtained by Jenko *et al.* in 2013 in Slovenian Brown cattle (0.09-0.11) or Phocas *et al.* in 2006 (0.14-0.15) in French Charolais.

EBV. The survival curves of ewes born from the 10% sires with extreme EBV (LPL- or LPL+) are represented in Figure 2. Ewes born from LPL+ rams present a significantly higher survival rate than ewes born from LPL- rams: the survival rate at 1,400 days (almost 4 years old) is 11 to 19% superior in LPL+ compared to LPL- according to the breed. These results tend to validate EBV obtained with the SK and are encouraging

Table 2. Genetic parameters of length of productive life.

	Sire variance	Flock × year variance	Heritability	Genetic std
BB	0.0717±0.0103	0.3785	0.115	0.2867
MTN	0.0480±0.0153	0.4310	0.076	0.1918
MTR	0.0694±0.0045	0.2737	0.131	0.2776
LAC	0.0813±0.0001	0.3837	0.154	0.3252

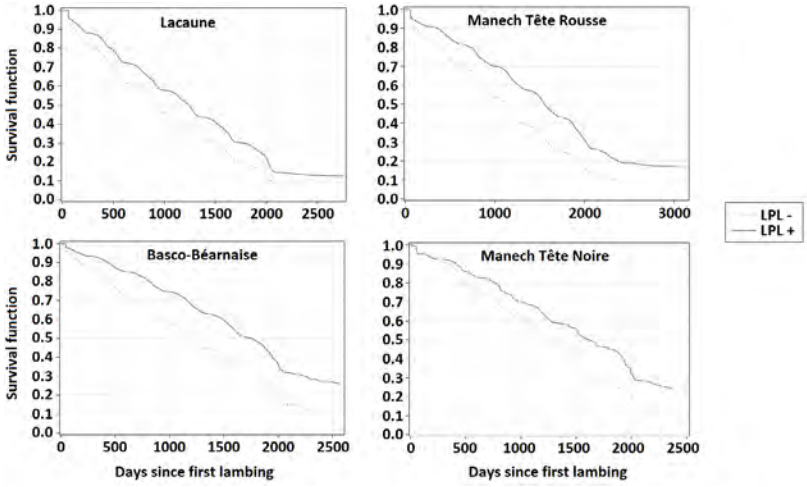


Figure 2. Survival functions of ewes born of rams with high longevity EBV (LPL+) or low longevity EBV (LPL-) in French dairy sheep breeds.

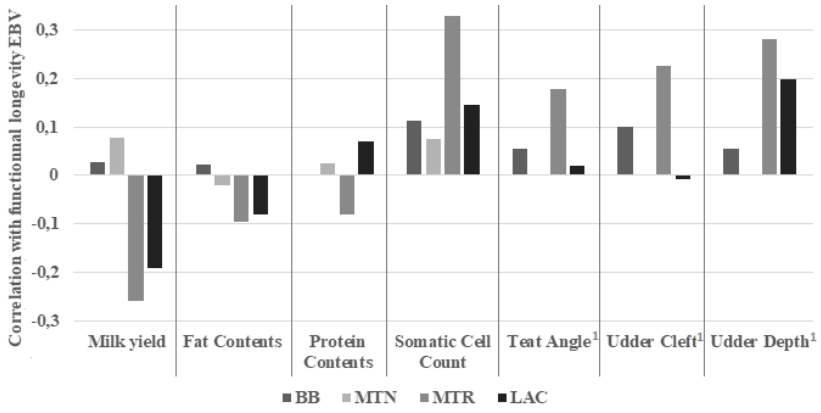


Figure 3. Correlations between longevity EBV and other traits under selection in French dairy sheep breeds. [1] No data available in MTN.

for a future genetic evaluation. However, the reliability of sires' EBV is strongly related to the number of uncensored daughters: the more uncensored daughters, the more reliable is the sire's EBV. In MTR, around 10 uncensored daughters are necessary to obtain a reliability superior or equal to 0.50, whereas only 4-5 are sufficient in LAC. In average, these accuracies can be obtained when the sire is 4 years-old in MTR and BB or 2-3 years old in LAC.

EBV correlations between LPL and other traits. Correlations between LPL and Milk Yield (MY) are negative in LAC and MTR (-25% and -19% respectively) and almost zero in BB and MTN (Figure 3). The negative correlations observed in LAC and MTR underline the antagonism between MY and LPL. Very low correlations are observed between LPL and fat and protein contents. The higher correlations are obtained with Somatic Cell Count (SCC) and udder morphology traits: improving SCC and udder morphology

would therefore improve functional longevity. However, this study should be completed by an estimation of the genetic correlations between LPL and the other traits (methodology of Tarrés *et al.,* 2006).

Conclusions

This approach about functional longevity indicates that LPL has a heritability similar to those obtained in cattle (0.07-0.16) and that it is possible to perform a genetic evaluation, even though accuracy remains low (CD<0.5) during the first years of use of rams. However, given the positive correlation between LPL and functional traits, introducing early predictors of LPL in the genetic evaluation could increase the accuracy. Further study is necessary to investigate which predictors should be used (SCC, udder morphology) to implement a multi-trait evaluation of functional longevity, as done currently in French dairy cattle (Ducrocq *et. al,* 2001).

References

Astruc J.M., Lagriffoul G., Legarra A. and Buisson D. (2022) Proc. of the 12th WCGALP, Rotterdam, the Netherlands (submitted).

Cox D.R. (1972) Journal of the Royal Statistical Society, Series B 34:187-220. https://doi.org/10.1111/j.2517-6161.1972.tb00899.x

Ducrocq V. (1992) INRAE Productions Animales, 5(HS):205-207. https://doi.org/10.20870/productions-animales.1992.5.HS.4290

Ducrocq V. and Sölkner, J. (1994) Proc. of the 5th WCGALP, Ontario, Canada.

Ducrocq V. (2005) Animal Science 80(03):249-256. http://dx.doi.org/10.1079/ASC41720249

Ducrocq V., Boichard D., Barbat A. and Larroque H (2001). Implementation of an approximate multitrait BLUP evaluation to combine production traits and functional traits into a total merit index. Proc. 52nd Annual Meeting EAAP, Budapest, Hungary.

Jenko J., Gorjanc G., Kovač M. and Ducrocq V. (2013) J. Dairy Sci., 96: 8002-8013. http://dx.doi.org/10.3168/jds.2013-6830

Mészáros G., Pálos J., Ducrocq V. and Sölkner J. (2010) Genet. Sel. Evol., 42:13. https://doi.org/10.1186/1297-9686-42-13

Phocas F. and Ducrocq V. (2006). Proc. of the 8th WCGALP, Belo Horizonte, Brazil.

Tarrés J., Piedrafita J. and Ducrocq V. (2006) Genet. Sel. Evol, 38(1): 65-83. https://doi.org/10.1186/1297-9686-38-1-65

Terawaki Y., Katsumi T. and Ducrocq V. (2006) J. Dairy Sci., 89(10): 4058-4065. https://doi.org/10.3168/jds.S0022-0302(06)72449-3

738. Maturity, a new indicator to improve longevity of Saanen dairy goats in France

M. Arnal[1,2*], I. Palhière[2] and V. Clement[1]

[1]Instit t de l'élevage, 24 chemin de Borde Rouge 31320 Castanet Tolosan, France; [2]GenPhySE, Université de Toulouse, INRAE, INPT, ENVT, 31326 Castanet Tolosan, France; mathieu.arnal@idele.fr

Abstract

The objective of this study was to obtain a maturity EBV to improve longevity in dairy goat. Several multi-parities models were developed in this study: a multi-trait model (MULTI) that used lactational production and a test day model (TDM) that used directly test day data. They were compared to the current repeatability model routinely used in France. There were no major differences for heritabilities and EBVs. It was found that the ability to produce milk in first parity was not the same as the ability to produce milk in third lactation. The second eigenvector of the (co)variance matrix permitted to discriminate the first and the third parity and to obtain a maturity EBV. This EBV was linked positively with longevity EBV and could be used to predict longevity. The multi trait model could be used to modify the way to calculate milk yield EBV used routinely in France by putting more emphasis on third parity EBVs, which is less negatively linked to longevity than first parity EBVs.

Introduction

Longevity is declining in dairy goats in France: goats born in 1991 produced 346 days more than those born in 2011 (Palhiere *et al.*, 2018). Current selection is based on production and udder morphology traits (Clément *et al.*, 2015). Longevity evaluation is under development in France. However, the longevity EBV accuracy becomes high several years (±7 years) after the start of AI bucks' semen diffusion. The milk yield EBV is routinely obtained using a repeatability model in dairy goats in France (Clément *et al.*, 2015). However, it is possible to use a multi trait model to obtain a milk yield EBV by parity. In doing an eigenvalues decomposition to the (co)variance matrix, the first component combines the EBVs by parity to obtain a 3-parity-milk-yield EBV. The second eigenvector permits to discriminate the first and the third parity milk yield EBV. This synthetic EBV is called maturity. A high maturity EBV means that the animal had a low first parity milk yield EBV and a high third parity milk yield EBV. The objective of this paper was to compare different models to calculate this maturity EBV. In a second part, the correlation between longevity and maturity was studied.

Materials & methods

Data. Data consists in 419,550 test-days from 65,953 lactations produced by 34,322 Saanen dairy goats from 118 herds measured between 2002 and 2021. Only the first three parities were kept. The data were extracted from the goat milk yield recording scheme used for the national French genetic evaluations. 67,359 animals were in the pedigree.

Models. Several models were compared.

Lactation model (LACT). LACT was a model very close to the one used routinely for milk yield evaluation in France (Clément *et al.*, 2015). It was a repeatability model. It was developed to be a reference for milk yield EBV in this study. The phenotype used was the lactation milk yield calculated with Fleischmann method. The genetic evaluation model was:

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_738, © M. Arnal *et al.* 2022

$y_{rnjlkmpi} = H_{rji} + A_{jkp} + M_{rjlp} + D_{rjmp} + a_n + p_n + e_{rnjlkmpi}$, where $y_{rnjlkmpi}$ was the lactation observation of goat n, in production year j, in parity r, belonging to kidding age class k, kidding month class l, class m of dry period length, in region p and herd i. H_{rji} is the fixed effect of herd; A_{jkp} is the fixed effect of age at kidding; M_{rjlp} was the fixed effect for the kidding period; D_{rjmp} the fixed effect of length of the dry period; a_n was the additive genetic breeding value that followed a normal distribution (mean: μ_{LACT}=0, variance= σ_{LACT}^2A) with A the additive genetic relation matrix based on pedigree information. p_n was the permanent environment value that followed a normal distribution. The genetic parameters were estimated using AIREMLF90 (Misztal *et al.*, 2002). TOT_LACT was the EBV obtained with LACT.

Test-day model (TDM). A TDM model was previously developed in Arnal *et al.* (2020). TDM was developed to obtain milk yield EBV, persistency EBV and maturity EBV using Legendre polynomial of degree 2. Here, a model using Legendre polynomial of degree 0 was used to be simpler than LEG2; persistency being not necessary in this study. The three parities were analyzed simultaneously in a multiple-parity model. The genetic and permanent environmental effects were modeled by a Legendre polynomial of order 0 for each parity: The phenotype used was directly the milk yield test day.

$$y_{rijklmdn} = HTD_{ri} + \mathrm{A}_{jk} + \mathrm{M}_{rjl} + \mathrm{D}_{rjm} + \sum_{o=1}^{6} \theta_{jrp} N_{(o,d)} + a_{rn0}\varphi_{(0,d)} + p_{rn0}\varphi_{(0,d)} + e_{rijklmdn}$$

where $y_{rijklmdn}$ was the observation of day in milk (DIM) d (7, …, 270) of goat n in production year j, in parity r, belonging to kidding age class k, kidding month class l, class m of dry period length, in region p, herd × test-date class i. HTD_{ri} is the fixed effect for the herd test-date; A_{jk} is the fixed effect for the age at kidding; M_{rjl} is the fixed effect for the kidding period; D_{rjm} is the fixed effect for length of the dry period; θ_{ko}, τ_{rlo}, π_{rmo} are fixed regression coefficients for age at kidding, the kidding month, dry period length respectively; $N_{(o,d)}$ is the o^{th} covariate at time d of a cubic natural spline function with 6 knots; a_{rn0} and p_{rn0} are respectively the random additive genetic and permanent environmental regression coefficients for the term of a Legendre polynomial of order *0*, $\varphi_{(0,d)}$ is the value of the term of the Legendre polynomial at time d and $e_{rijklmdn}$ is the residual term.

TOT_TDM which represents the milk production for the three first parities was obtained multiplying the EBVs matrix by the first eigenvector of the (co)variance matrix. MAT_TDM which represents the maturity EBV was obtained multiplying the EBVs matrix by the second eigenvector of the (co)variance matrix. The genetic parameters were estimated using WOMBAT (Meyer, 2007).

Multi-trait model (MULTI). A multi-trait model was developed to obtain milk yield EBV and maturity EBV. The phenotype used was the lactation milk yield calculated with Fleischmann method.

$y_{rnjlkmpi} = H_{rji} + A_{jkp} + M_{rjlp} + D_{rjmp} + a_{rn} + e_{rnjlkmpi}$, where $y_{rnjlkmpi}$ was the lactation observation of goat n, in production year j, in parity r, belonging to kidding age class k, kidding month class l, class m of dry period length, in region p and herd i. H_{rji} is the fixed effect of herd; A_{jkp} is the fixed effect of age at kidding; M_{rjlp} was the fixed effect for the kidding period; D_{rjmp} the fixed effect of length of the dry period; a_{rn} was the additive genetic breeding value. TOT_MULTI which represents the milk production for the three first parities was obtained multiplying the EBVs matrix by the first eigenvector of the (co)variance matrix. MAT_MULTI which represents the maturity EBV was obtained multiplying the EBVs matrix by the second eigenvector of the (co)variance matrix. The genetic parameters were estimated using AIREMLF90 (Misztal *et al.*, 2002).

Results

Heritabilities from the different models were estimated. The heritability for LACT was equal to 0.30±0.002. For MULTI, the heritabilities were equal to 0.34±0.006, 0.30±0.008 and 0.29±0.006 for first, second and

third parities respectively. For TDM, they were equal to 0.30±0.006, 0.31±0.007, and 0.32±0.008 for first, second and third parities respectively.

The genetic correlations between parities were slightly superior with MULTI than TDM for all the correlations. The highest correlation was between second and third parity (0.89±0.017 for TDM and 0.95±0.058 for MULTI). The smallest correlation was between first and third parity equal to 0.63±0.029 for TDM and 0.75±0.048 for MULTI. The correlation between first and second parity was intermediate: equal to 0.78±0.018 for TDM and 0.87±0.044 for MULTI.

The first eigenvalue representing the level of milk production for the three lactations represented the biggest part of the variance (91% for MULTI and 87.2% for TDM). The second eigenvalue represented only 8.3% for MULTI and 10.2% for TDM). Table 1 presents the first two eigen vectors for MULTI and TDM. The first eigenvector represented the production level for the first three lactations. For TDM, the value of the eigenvector for third parity (0.7) was higher than second one (0.6), higher than first parity (0.38). For MULTI, the value for second and third parity were equal to 0.6 and higher than first parity (0.52). The second eigenvector opposed the first and the third parity (-0.82 for first parity and +0.55 for third parity for TDM, and a value close to 0 for the second parity (-0.12)). For MULTI, the second eigenvector was like TDM (-0.8 for first parity, 0.1 for second parity and 0.59 for third parity).

Table 2 presents the correlations between the different EBVs and longevity EBV calculated in Palhière's study (2018) for AI bucks with at least 21 progeny (n=371). As expected, the correlation between TOT_LACT and TOT_MULTI was close to 1 (0.98). The correlation between TOT_LACT and TOT_TDM was smaller: equal to 0.89. The correlation between the two multi traits models, TOT_MULTI and TOT_TDM, was equal to 0.92. The correlation between TOT_LACT and first parity EBV were high: equal to 0.97 for TOT_MULTI and 0.95 for TDM. The correlation between TOT_LACT and third parity EBVs were smaller: equal to 0.91 for MULTI and 0.77 for TDM. The correlation between the two maturity EBVs, MAT_MULTI and MAT_TDM, was high: equal to 0.86. The correlations between TOT and MAT were close to 0 for TDM and MULTI as expected (-0.12 for MULTI and -0.02 for TDM) because they were derived from eigenvectors.

Table 1. First two eigenvectors (EV) for TDM and MULTI models.

	TDM		MULTI	
Parity	**1st EV (TOT)**	**2nd EV (MAT)**	**1st EV (TOT)**	**2nd EV (MAT)**
1st	0.38	-0.82	0.52	-0.80
2nd	0.60	-0.12	0.60	0.10
3rd	0.70	0.55	0.60	0.59

Table 2. EBVs correlations for AI bucks with at least 21 progeny (n=371).

	TOT_LACT	TOT_MULTI	MAT_MULTI	TOT_TDM	MAT_TDM
TOT_MULTI	0.98				
MAT_MULTI	-0.28	-0.12			
TOT_TDM	0.89	0.92	0.03		
MAT_TDM	-0.40	-0.27	0.86	-0.02	
LONGEVITY	-0.24	-0.18	0.37	-0.12	0.33

However, the correlation between MAT and TOT_LACT were negative (-0.28 for MULTI and -0.4 for TDM). The correlation between the longevity and milk EBVs were negative (TOT_LACT, TOT_MULTI and TOT_TDM: -0.24, -0.18, -0.12 respectively). The correlation between the first parity and the longevity was negative (-0.30 for TDM and MULTI), whereas the correlation was smaller between the third parity EBV and the longevity (-0.09 for Multi and -0.05 for TDM). The correlation between longevity EBV and MAT_MULTI was positive (0.37) close to the correlation between longevity EBV and MAT_TDM (0.33).

Discussion

The heritabilities estimated for milk yield from the different models were coherent and all around 0.3 which is the value used in the current official evaluation (Clément *et al.*, 2015). The correlation between TOT_LACT and MULTI that used same phenotypes were close to 1 and a bit smaller with TDM (significant). This difference could be due to the difference of phenotype used. The genetic correlations and consequently the eigenvectors and values were similar between TDM and MULTI model. Because MULTI and TDM gave similar EBVs and genetic parameters, MULTI, which is easier to used (less phenotypes with MULTI: several TD with TDM vs 1 lactational milk yield with MULTI), will be used to obtain maturity EBVs.

The genetic correlation between first and third parity are around 0.7 which seems to indicate that it is not totally the same trait to produce in first and third parity. The first eigenvalue is not representing 100% of the variance and the maturity that highlight the difference between first and third parity represents around 10% of the variance. This difference in genetic capacity to produce in first and third parity seems to be linked to longevity, with a smaller negative correlation with third parity than with first parity and consequently a positive correlation with maturity. This correlation strengths the utility of this new trait, the maturity, because it is linked to a trait of major interest for future breeding objectives. Indeed, longevity is a difficult trait to catch because an animal needs to have its career achieve to estimate well its EBV. Maturity could be a good predictor as SCS or udder morphology traits as showed in Palhière *et al.* (2018).

The current milk yield EBV is closer to first parity than third parity. It could be interesting to put more emphasis to third parity than first parity in the milk yield EBV. In this way, selecting for milk yield will be less damageable for longevity.

One limit of multi-traits models (MULTI and TDM) is that only the first three parities are considered whereas the LACT model used the first 10 parities which can be an advantage when maturity and difference of capacity to produce between the beginning of the career and the end of the career are studied.

Funding

This study was funded by APIS-GENE under MALO program.

References

Arnal, M., H. Larroque, H. Leclerc, V. Ducrocq, and C. Robert-Granié. 2020. J of Dairy Science 103:4517–4531.

Clément, V., I. Palhiere, and H. Larroque. 2015. collection résultats.

Meyer, K. 2007. Journal of Zhejiang University-Science B 8:815–821.

Misztal, I., S. Tsuruta, T. Strabel, B. Auvray, T. Druet. 2002. Pages 743–744 in Proc 7th World Congr Genet App Livest Prod.

Palhiere, I., C. Oget, and R. Rupp. 2018. Proc 10th World Congr Genet App Livest Prod 165.

739. The relevance of spermatozoal RNA and quality in livestock production

M.J. Hodge[1,2], S.J. Rindfleish[2], C.P. Stephen[1], S. de las Heras-Saldana[3*] and S.D. Pant[1*]

[1]School of Agricultural, Environmental and Veterinary Sciences, Charles Sturt University, Wagga Wagga, NSW 2678, Australia; [2]Apiam Animal Health, Apiam Genetic Services, Dubbo, NSW 2830, Australia; [3]School of Environmental and Rural Science, University of New England, Armidale, NSW 2351, Australia; sdelash2@une.edu.au; spant@csu.edu.au

Abstract

Spermatozoal cells are known to contain a wide variety of both coding and non-coding transcripts, but the physiological role of these transcripts is not fully understood. Therefore, the overall aim of this study was to characterize both coding and non-coding transcripts contained in ovine spermatozoa, and identify differences in expression patterns between three different sheep breeds (Merino, Dohne and Poll Dorset), and ejaculates that vary in quality. A variety of coding (e.g. *MVK, PAMR1, ATP2A3, KLK12,* etc.) and non-coding transcripts (e.g. XLOC_025216, XLOC_048757, XLOC_001809, XLOC_025093, etc.) were found to be differentially expressed in different sheep breeds and ejaculates of contrasting quality. Some of these transcripts have been reported to be associated with spermatozoal function or physiology in the literature. Overall, these results provide novel insights into spermatozoal transcriptomes in sheep, and also potentially indicate their functional relevance.

Introduction

Good reproductive performance is a crucial factor that influences the economics of livestock production, and successful conception is a key determinant of the reproductive performance of livestock. Consequently, artificial insemination is widely used in a variety of livestock species to maximise conception rates, while simultaneously exerting greater control over breeding, which in turn, facilitates increased genetic gain.

While the quality of semen used in artificial insemination is widely believed to influence conception outcomes during artificial breeding, poor conception outcomes can occur even when good quality semen is used (Hall 1981). This indicates that semen quality is not the sole determinant of successful conception, and additional molecular mechanisms functioning on the ram side may exist. The role of spermatozoal transcripts is interesting in this regard, as spermatozoa are known to contain thousands of transcripts, which are transferred to ova during fertilisation. It is generally understood that activation of embryonic genome is preceded by degradation of maternal transcripts post fertilisation. Therefore, the transcripts contained in spermatozoa that are transferred to the ova upon fertilization, could have an important role to play in these events.

Therefore, characterising spermatozoal transcripts, both mRNA and long non-coding RNA (lncRNAs) can be useful in increasing the likelihood of successful conception, which in turn has the potential to increase the efficiency and profitability of livestock farming.

Materials & methods

The methodology involved in semen collection, assessment, RNA isolation, library preparation, sequencing, and differential gene expression analysis to identify coding transcripts has already been published (Hodge *et al.* 2021). RNA samples extracted from spermatozoa were subjected to next-generation sequencing using the Illumina Hiseq2000 platform following manufacturer's instructions. Following sequencing, the quality of reads was assessed with FastQC v0.11.5. Poor quality bases (Phred score Q<30), adaptors, and overrepresented sequences were filtered out with trimmomatic v0.36 (Bolger *et al.* 2014).

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_739, © M.J. Hodge *et al.* 2022

Identification of Long non-coding RNAs (lncRNAs). Analyses to identify lncRNAs involved aligning cleaned reads to the reference genome (Ovis aries, Oar v.4) via STAR v2.7.7 (Dobin *et al.* 2013). The assemblies were merged with cuffmerge from Cufflink v2.2.1 and the GTF file produced was used in FeatureCounts function from Subread v2.0.3 software to obtain the transcript counts in each sample. A differential expression analysis was performed with DESeq2 v1.30.1 using edgeR v3.32.1 (Robinson *et al.* 2010) to filter the lowly expressed genes based on the logarithmic counts per million (lcpm) and TMM normalised gene counts (trimmed mean of M values). Ejaculate quality was used as a fixed effect in the statistical model to identify statistically significant changes in transcript abundances assessed between breeds (FDR<0.05 and logFC>1). Subsequently, a total of 9 Merino, 9 Dohne and 9 Poll Dorset rams, which passed quality control in edgeR, were included in the differential gene expression analysis. Four contrasts were performed, comparing transcript expression between breeds (Poll Dorset vs Merino, Dohne vs Merino, and Dohne vs Poll Dorset), and ejaculates of relatively high- and low-quality.

Transcripts longer or equal to 200 nt were identified as lncRNAs, and Cuffcompare v2.2.1 was used to classify transcripts. Only transcripts classified as 'u', 'o', and 'x' were retained for further analysis. Putative lncRNA candidates were then identified if their coding potential score was calculated as <0 via CNCI v2 and their coding probability was calculated as <0.5 via CPC2 v1.0.1. A miRNA database (http://rumimir.sigenae.org) on sheep was then used to remove lncRNAs that had a potential miRNA precursor (E-value 0.001) in BLAST v2.12.0. Transcripts with ORF larger than 300nt were identified using getorf function of EMBOSS (Rice *et al.* 2000) and removed. Transcript sequences were cross-matched against: (1) a protein sequence database (Pfam library) using hmmscan software; (2) the UniProt/SwissProt database; and (3) the annotated proteins for *O. aries* v4 using blastx. In all steps, any transcripts with coding potential were eliminated and remaining transcripts with an average FRKM>0.1 per breed, containing more than one exon were considered as lncRNAs, and retained for further analysis. The identified lncRNAs were used to find target genes (cis-target) within a 10 kb window of the lncRNA. Trans-target genes were identified based on a co-expression analysis using Pearson's correlation (r>±0.95 with *P*-value<0.05) between the mRNA genes and lncRNA.

Results

Differential gene expression analysis identified a number of Differentially Expressed Genes (DEGs) in each contrast that was performed. A summary of the total number of DEGs, as well as the gene symbols of the top 10 DEGs identified in each contrast, is presented in Table 1.

Analyses involving the identification of long non-coding RNAs identified a total of 1,205 lncRNAs in sheep spermatozoa, of which 150 transcripts passed the quality filtering for low expression and only 6 lncRNA were found to be differentially expressed. A summary of key differentially expressed lncRNAs along with their target genes is presented in Table 2.

Table 1. Number of differentially expressed genes in breed specific contrasts, and between ejaculates of varying quality (Hodge *et al.* 2021).

Contrast	No. of DEGs	Top DEGs
Dohne vs Merino	72	*MVK, TNS3, ALAS1, CCN1, MSRA, RRP15, PPARD, MED6, CARMIL1, FLRT2*
Dohne vs Poll Dorset	73	*PAMR1, LY6E, ARGLU1, CDO1, VARS2, TP53I11, POLK, FBXL14, MAN1A1, CHI3L1*
Merino vs Poll Dorset	570	*ATP2A3, SLC35A5, BORCS5, MSRA, KRT4, FAM210B, NKPD1, ITM2C, ANAX2, SLC2A3*
High vs Low quality	39	*KLK12, SPEM2, LYPLA1, TNNC1, LYRM4, OXCT2, SMKR1, FAM57A, SRGN*

Table 2. Differentially expressed target genes and their co-expressed lncRNAs across the comparisons of breeds sampled and ejaculate quality.

Contrast	DE lncRNA (Target Genes)
Dohne vs Poll Dorset	XLOC_048235 *(C1H1orf141, SLC38A4, XPNPEP3, NUP43, ELP4, CHRM1, PTH2R, FBXO7, ABCC9)*
	XLOC_072802 (LOC105607053, LOC105609923)
Merino vs Poll Dorset	XLOC_001101 *(CSF3R, CLEC5A, TRNAC-GCA, GRAMD1A, TREML2)*
	XLOC_001727 *(CLIC6, EX26)*
	XLOC_066444 *(FBXO7, CABS1, PPM1E, MED31)*
High vs Low quality	XLOC_023268 (LOC105611383, *SOBP)*

Discussion

There are a variety of factors on both the ram and ewe sides that influence conception outcomes in sheep, particularly when artificial breeding technologies like artificial in-semination are used. While a significant emphasis is placed on semen quality in artificial breeding programs, it is clear that other determinants influencing conception outcomes remain uncharacterized. The role of spermatozoal transcripts is of significant interest in this regard, because while spermatozoa have been shown to contain a large repertoire of transcripts which include both protein-coding mRNA and non-coding RNAs, very little is known about their influence on ejaculate quality and fertilisation.

This current study provides evidence indicating that a variety of mRNA and lncRNAs are expressed in sheep spermatozoa. A very high number of spermatozoal cells were needed for RNA extraction, because each spermatozoa conceivably carries a small amount of RNA relative to normal somatic cells. This supports the notion that spermatozoal cells are likely transcriptionally inactive for the most part.

However, in our study, we have identified several transcripts in ovine spermatozoa (e.g. *CABS1*, *CD82*, *ATP1B3*, *EFHB,* etc.), that have been previously associated with spermatogenesis, sperm morphology and ejaculate quality parameters. Therefore, it is likely that such spermatozoal transcripts play a physiological role in the development of spermatozoa, and could therefore be used as markers. Our study also identified several transcripts that were differentially expressed between spermatozoa sourced from ejaculates that varied in quality. Therefore, it is also possible that some of these transcripts can be used as markers to identify good quality semen that is likely to yield desirable conception outcomes.

Furthermore, the spermatozoal transcriptomes in different species have been found to contain thousands of transcripts, which is consistent with our findings (Selvaraju *et al.* 2017; Prakash *et al.* 2021). These transcripts are transferred from spermatozoon to the ova during fertilisation, where they could potentially play a crucial role in directing early embryonic development, thereby influencing pre-natal growth that may be correlated with economically important traits like birth weight, etc.

Overall, our findings accord with a variety of studies focused on spermatozoal transcripts in other species that also support a functional role for these transcripts. Therefore, our study offers several novel insights indicating spermatozoal transcripts could have crucial physiological roles and could potentially influence conception outcomes, and even early embryonic development. Therefore, future in depth investigations aiming to characterise the physiological role of spermatozoal transcriptomes are warranted.

References

Bolger A.M., Lohse M. and Usadel B. (2014) Bioinformatics 30(15): 2114-2120. https://doi.org/10.1093/bioinformatics/btu170

Dobin A., Davis C.A., Schlesinger F., Drenkow J., Zaleski C. *et al.* (2013) Bioinformatics 29(1): 15-21. https://doi.org/10.1093/bioinformatics/bts635

Hall J.L. (1981) Fertil Steril 35(4): 457-463. https://doi.org/10.1016/s0015-0282(16)45444-0

Hodge M.J., De Las Heras-Saldana S., Rindfleish S.J., Stephen C.P. and Pant S.D. (2021) Genes (Basel) 12(2). https://doi.org/10.3390/genes12020203

Prakash M.A., Kumaresan A., Ebenezer Samuel King J.P., Nag P., Sharma A. *et al.* (2021) Front Cell Dev Biol 9(647717. https://doi.org/10.3389/fcell.2021.647717

Rice P., Longden I. and Bleasby A. (2000) Trends Genet 16(6): 276-277. https://doi.org/10.1016/s0168-9525(00)02024-2

Robinson M.D., Mccarthy D.J. and Smyth G.K. (2010) Bioinformatics 26(1): 139-140. https://doi.org/10.1093/bioinformatics/btp616

Selvaraju S., Parthipan S., Somashekar L., Kolte A.P., Krishnan Binsila B. *et al.* (2017) Sci Rep 7(42392. https://doi.org/10.1038/srep42392

740. Exploring the vomeronasal organ transcriptome in Rasa Aragonesa rams with different sexual behaviour

K. Lakhssassi[1,2], I. Ureña[3], B. Marín[4], M.P. Sarto[1], B. Lahoz[1], J.L. Alabart[1], J. Folch[1], M. Serrano[3] and J.H. Calvo[1,5*]

[1]Unidad de Producción y Sanidad Animal, Centro de Investigación y Tecnología Agroalimentaria de Aragón (CITA), Instituto Agroalimentario de Aragón (IA2), Avenida Montañana, 50059 Zaragoza, Spain; [2]Institut National de la Recherche Agronomique, INRA BP 6356 Rabat, Morocco; [3]Instituto Nacional de Investigación y Tecnología Agraria y Alimentaria, Ctra. de La Coruña, 28040 Madrid, Spain; [4]Centro de Encefalopatías y Enfermedades Transmisibles Emergentes, Universidad de Zaragoza-IA2, 50013 Zaragoza, Spain; [5]ARAID, 50018 Zaragoza, Spain; jhcalvo@cita-aragon.es

Abstract

The vomeronasal organ (VNO) has long been of interest regarding the sensory control of social and sexual behaviour. In this study, high-throughput RNA sequencing was utilized to investigate transcriptional changes in the VNO between rams of Rasa Aragonesa with different sexual behaviour (active and not active). The RNA-Seq analysis generated 812,065,820 clean reads and identified six differentially expressed genes, of which *GFRA3*, *CEACAM*, and *LRIT3* were upregulated, whereas *SERPINB12, TP53INP2* and *CES4A* genes were downregulated in the active compared to the not active rams group. The Gene Set Enrichment Analysis (GSEA) of all the expressed genes in the VNO detected 128 overrepresented pathways at FDR 5%, predominantly related to biological process such as positive regulation of protein phosphorylation and cellular response to growth factor stimulus pathways. Studying these genes and the processes they control will improve our understanding of the genomic regulation of sexual behaviour in rams.

Introduction

In small ruminants, the vomeronasal organ (VNO) is morphologically constructed for the detection of pheromones, which are related to social and reproductive behavioural responses (Villamayor *et al.*, 2021). The sensory information interferes with the nervous control of gonadotropin secretion (Signoret, 1991). These chemosensory signals stimulate male sexual arousal and behaviour (Keverne, 2004). In fact, Ungerfeld *et al.* (2006) pointed out that the perception of pheromones by the VNO stimulates the sexual performance of rams, which determine a faster response to the stimulus of oestrous ewes. In addition, mountings and ejaculations increased without the need of more courtship behaviour. However, the blockage or removal of the VNO leads to a reduction of number of mounts and services in bulls (McGrath, 1981) and a depression of sexual activity in male rats (Saito and Moltz, 1986). In sheep, the ram effect is commonly used to improve the out of season reproduction. It has been observed that males with greater sexual behaviour produce a greater stimulus during seasonal anoestrous, which lead into a higher percentage of mated ewes and higher fertility. However, there is a considerable variation in sexual behaviour between rams, hypothesizing that VNO features may differ not only among species but also within the same species at mating (Villamayor *et al.*, 2021). In this study, we carried out an RNA sequencing approach to

investigate transcriptional changes in the VNO by comparing two groups of Rasa Aragonesa rams with different sexual behaviour (active and not active).

Materials & methods

Ethics statement. All experimental procedures including care of animals and euthanasia were performed in accordance with the guidelines of the European Union regulations for the use and care of animals in

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_740, © K. Lakhssassi *et al.* 2022

research (Directive 2010/63/EU) and approved by the Animal Ethics Committee of the Research Centre (protocol number 2017/02).

Animals, tissues samples and RNA extraction. The experiment was conducted with 59 rams, submitted to individual sexual behavioural pen tests, twice for each ram. Each ram was exposed to three synchronized adult ewes for 20 minutes to observe their behaviour, and count the frequency of mounts and services. Three serial blood samples were taken one hour apart the day before, in lithium heparin tubes, to measure testosterone. In addition, individual live weight (LW) and body condition score (BCS) were assessed as well as testicular size. Two ram homogeneous groups in LW and BCS were identified after carrying out a decision tree using the package 'party' in R project, considering all the variables for the mounts and services phenotypes: active (A) (7.93±3.56, average mounts ± sd) and not active (NA; without any mount) rams. Then, six rams of each group were sacrificed. VNO tissue samples were extracted and stored at -80 °C until RNA isolation by RNeasy Lipid Tissue mini kit (QIAGEN, Madrid, Spain).

RNA-Seq data processing. Sequencing was carried out generating Illumina paired-end reads of 151 bp. Quality control of sequences was checked by FastQC. The Trimmomatic program was used to remove adaptors, low quality reads and overrepresented sequences. Then, the trimmed reads were mapped to the reference genome Oar_rambouillet 1.0 (GCA_002742125.1) using STAR v2.7.8a and gene level quantification was estimated using HTSeq. Differential expression analysis were performed with EdgeR using the threshold of CPM=0.5 in at least five samples. Multiple-testing corrections by Benjamini and Hochberg false-discovery rate (FDR 5%) controlling procedure were performed. Genes with an adjusted *P*-value≤0.05 were identified as DEGs. Enrichment Analysis were examined by GSEA. All the pipeline analysis was done using the OmicsBox platform v2.0.36 from BioBam (https://www.biobam.com/omicsbox).

Results

Active and non-active rams had similar LW and BCS, and no significant differences were found in testosterone levels or testicular size. RNA-seq data were obtained for VNO tissue samples of Rasa Aragonesa rams (n=12), ranging the number of total raw reads per sample from 60,537,894 to 83,449,630. After filtering by Trimmomatic, the total number of clean reads ranged from 41,240,636 to 78,958,466. The reads uniquely mapped by STAR to the reference genome varied from 55.4 to 84.5% among samples. These findings indicated good quality data that were suitable for subsequent analysis. After counting the reads mapped in each gene by HTseq, we next investigated the differences in gene expression data between active and not active rams by EdgeR software. As a result, 15,227 genes were expressed in the VNO. EdgeR identified six DEGs at FDR 5%. The *GFRA3*, *CEACAM* and *LRIT3* were up-regulated in the active rams whereas the *SERPINB12*, *TP53INP2* and *CES4A* were found down-regulated (Table 1). Gene set enrichment analysis included the 15,227 expressed genes. We detected 128 overrepresented pathways at FDR q-value 5%, predominantly related to biological process (BP) and molecular function (MF). The top 30 GO are shown in Figure 1.

Discussion

The principal goal of this study was to determine the possible molecular mechanisms underlying the sexual behaviour of Rasa Aragonesa rams by investigating transcriptional changes in the VNO between active and not active rams. The analysis revealed fewer DEGs, suggesting little difference in the VNO. The three up-regulated DEGs identified belong to signal term and seems to affect the sexual behaviour via network interaction with other genes. In fact, *GFRA3* correspond to the glial-cell line-derived neurotrophic factor Family Receptor Alpha 3 (GDNF), among its related pathways are RET signaling. GDNF is recognized to regulate the development of small-diameter nociceptors and may be involved in modulating

Table 1. Differentially expressed genes between the active and not active rams in the vomeronasal organ.

Gene ID	Gene name	logFC[1]	logCPM[2]	*P*-value	FDR[3]
ENSOARG00020014530	*GFRA3*	1.3	3.40	4.46E-06	0.03
ENSOARG00020022556	*CEACAM*	2.8	0.92	7.65E-06	0.03
ENSOARG00020023342	*LRIT3*	2.4	-0.03	7.64E-06	0.03
ENSOARG00020025883	*SERPINB12*	-2.2	0.04	9.69E-06	0.03
ENSOARG00020013838	*TP53INP2*	-1.6	1.78	2.09E-05	0.05
ENSOARG00020018190	*CES4A*	-1.7	0.91	2.02E-05	0.05

[1] Log2 fold changes.

[2] The average log2 counts per millions.

[3] False discovery rate.

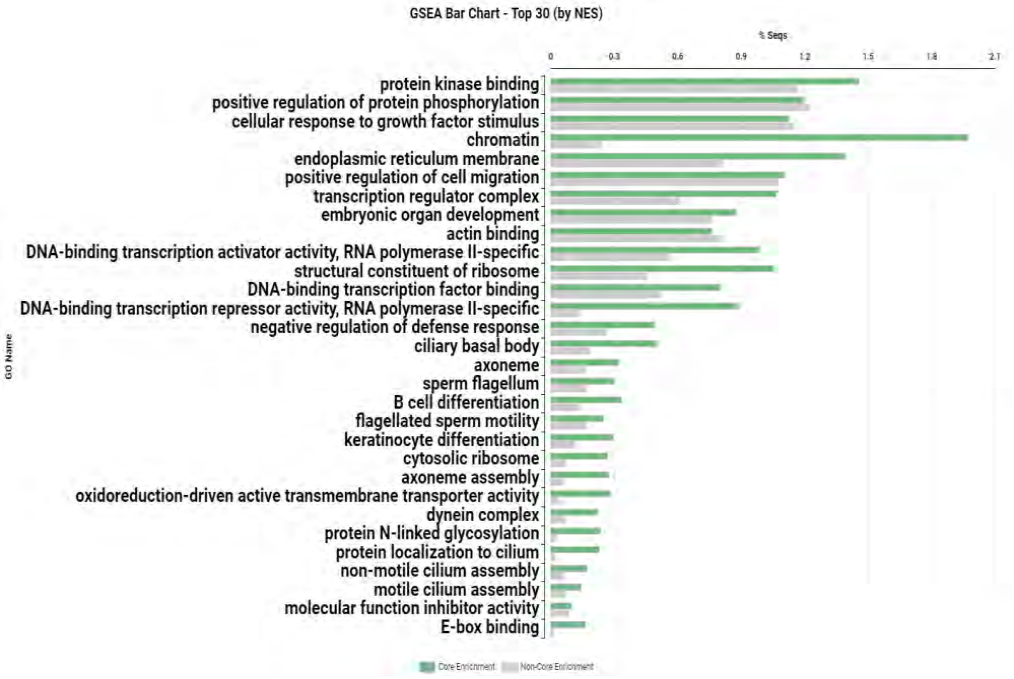


Figure 1. Top 30 GO between the A vs NA comparison in the VNO.

mechanosensation (Albers *et al.*, 2006). Low expression of this gene deteriorates reproductive potential via accelerated neurodegeneration (Vasiliev *et al.*, 2021). However, GFRA3 aberrant expression can increase human reproductive potential because of improvements in neural regeneration (Jankowski *et al.*, 2010). For the *CEACAM* gene, Gu *et al.* (2020) reported that CEACAM proteins disrupt transforming growth factor beta (TGFB) signaling pathway. Kim *et al.* (2013) indicated that *CEACAM* was one of the de-repressed gene families in the brain of *Peg3* heterozygous embryos and adult mice. This gene was reported to play important roles in controlling foetal growth rates and sexual behaviour in males and females. Swaney *et al.* (2008) indicated that the *Peg3* regulates sexual experience dependent preferences for oestrous odours. This is to say that mutation in *Peg3* could affect the expression of *CEACAM* gene family, which might result in a defect in reproduction or ability for mating. Similarly, *LRIT3* was reported as a modulator of Fibroblast Growth

Factor Receptor 1 (*FGFR1*) that is essential for the development of the gonadotropin-releasing hormone (GnRH) system (Tata *et al.*, 2012). Furthermore, the GSEA identified 128 overrepresented pathways, predominantly related to BP such as positive regulation of protein phosphorylation pathway, involved in the developing male and female rat brain as well as normal endocrine and behavioural processes in adulthood (Auger, 2003). Another pathway of interest was related to MF such as the protein kinase-binding pathway whose signaling pathways have been implicated in the estradiol regulation of sexual receptivity (Dewing *et al.*, 2008). These findings encourage deeper investigation towards the importance of these genes and their associated mechanisms to understand better the genomic regulation of sexual behaviour in Rasa Aragonesa rams.

References

Albers K.M., Woodbury C.J., Ritter A.M., Davis B.M., and Koerber H.R. (2006). J. Neurosci. 26:2981. https://doi.org/10.1523/JNEUROSCI.4863-05.2006

Auger A.P. (2003). J. Neuroendocrinol. 15:622–627. https://doi.org/10.1046/j.1365-2826.2003.01041.x

Dewing P., Christensen A., Bondar G., and Micevych P. (2008). Endocrinology. 149:5934–5942. https://doi.org/10.1210/en.2008-0847.

Gu S., Zaidi S., Hassan M.I., Mohammad T., Malta T.M., *et al.* (2020). Gastroenterology. 158:238. https://doi.org/10.1053/J.GASTRO.2019.09.023

Jankowski M.P., Rau K.K., Soneji D.J., Anderson C.E., and Koerber H.R. (2010). J. Neurosci. 30:16272–16283. https://doi.org/10.1523/JNEUROSCI.2195-10.2010

Keverne E.B. (2004). Physiol. Behav. 83:177–187. https://doi.org/10.1016/j.physbeh.2004.08.013

Kim J., Frey W.D., He H., Kim H., Ekram M.B., *et al.* (2013). PLoS One. 8. https://doi.org/10.1371/journal.pone.0083359

McGrath T.J. (1981). Texas A&M University. https://hdl.handle.net/1969.1/ETD-TAMU-1981-THESIS-M147

Saito T.R., and Moltz H. (1986). Physiol. Behav. 37:507–510. https://doi.org/10.1016/0031-9384(86)90215-5.

Signoret J.P. (1991). J. Steroid Biochem. Mol. Biol. 39:639–645. https://doi.org/10.1016/0960-0760(91)90263-5

Swaney W.T., Curley J.P., Champagne F.A., Keverne and E.B. (2008). Behav. Neurosci. 122:963–973. https://doi.org/10.1037/a0012706

Tata B. K., Chung W.C.J., Brooks L.R., Kavanaugh S.I., and Tsai P.S. (2012). Fibroblast. Biol. Reprod. 86:119–120. https://doi.org/10.1095/biolreprod.111.095992

Ungerfeld R., Ramos M.A., and Möller R. (2006). Appl. Anim. Behav. Sci. 99:248–252. https://doi.org/10.1016/j.applanim.2005.10.016

Vasiliev G., Chadaeva I., Rasskazov D., Ponomarenko P., Sharypova E., *et al.* I. Int. J. Mol. Sci. 22:1–17. https://doi.org/10.3390/ijms22052346

Villamayor P.R., Robledo D., Fernández C., Gullón J., Quintela L., *et al.* (2021). Genomics. 113:2240–2252. https://doi.org/10.1016/j.ygeno.2021.05.007

741. Elucidating transcriptomic differences in perirenal fat between male and female suckling lambs

M. Alonso-García*, J.J. Arranz, H. Marina, R. Pelayo, C. Esteban Blanco, B. Gutiérrez-Gil and A. Suárez-Vega

Department of Animal Production, University of León, Campus de Vegazana, s/n, 24007 León, Spain; maalg@unileon.es

Abstract

The aim of this study was to evaluate the sex-specific effects of a diet restriction in prepuberal ewes on the perirenal fat transcriptome of their offspring. For this research, we used RNA-Seq data from 11 lambs, 2 females and 3 males from control ewes, and 3 females and 3 males from ewes subjected to a nutritional challenge (NC) at the prepuberal stage. After a differential expression analysis, 116 genes showed an opposite expression pattern due to the NC between male and female lambs. These genes were functionally involved in mitochondrial activity, characteristic of brown adipose tissue. From our results, it could be inferred that the nutritional restriction at the prepuberal stage of the mothers may have an opposite effect in the transition from brown to white adipose tissue in each sex.

Introduction

Differences in fat deposition between males and females have been studied over the years in numerous species. Several factors, such as sex hormones, aging, hereditary factors and epigenetic effects, influence the fat distribution (Frank *et al.*, 2019). In sheep models, the perirenal adipose tissue (PAT) depot has been shown to be rich in brown adipocytes during the early stages of life (Henry *et al.*, 2017; Symonds *et al.*, 2015). In rodents and humans, it has been confirmed that females have more brown adipose tissue than males (Karastergiou *et al.*, 2012). In sheep, a recent study using the RNA-Seq methodology to determine differentially expressed genes (DEG) affecting fat metabolism in adipose tissue according to tail size and sex pointed out that genes regulating fat metabolism may be different for the two sexes (Fan *et al.* 2019). Besides that, also in sheep, it has been observed that maternal nutrition has a sex-dependent effect on offspring growth and carcass quality and yield (Ithurralde *et al.*, 2019). Regarding adipose tissue, both maternal under and overnutrition at the prenatal stage predisposed lambs to have visceral obesity by interfering with adipogenesis and adipose tissue functional traits (Khanal *et al.*, 2020). RNA-Seq analysis has been performed to elucidate how pre- and postnatal malnutrition can affect adipose tissue functional development in lambs (Ahmad *et al.*, 2021). However, little is known about how early life nutrition in ewes can affect the adipose tissue metabolism of lambs in a sex-dependent way. Therefore, the present study aimed to evaluate sex-specific effects of a maternal diet restriction at the prepuberal growth period on the PAT transcriptome of the lambs.

Materials & methods

Animal selection and RNA extraction. Twelve Assaf suckling lambs were selected for this study. Specifically, those animals belonged to four different groups: (1) Control (C) female (n=3), (2) Nutritional challenge (NC) female (n=3), (3) C male (n=3) and (4) NC male (n=3). Animals belonging to the NC group were descendants of dams subjected to dietary restriction during their pre-pubertal stage (from 3.5 to 5.5 months of age). After lambing, new-born lambs were kept with their mothers for the first 4 to 8 hours to suck colostrum. Then, they were fed artificially to avoid any maternal effect. Lambs were slaughtered when they reached the commercial weight established in Spain for suckling lamb meat (average live weight was 10.68±0.26 kg for females and 11.31±0.73 kg for males). All the animals were slaughtered at a local slaughterhouse, where a sample of PAT was collected from each suckling lamb. The tissue was preserved

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_741, © M. Alonso-García *et al.* 2022

in RNA-stabilization solution (Ambition RNAlater, Life Technologies) and stored 24h at 4 °C. After that, the solution was removed, and the samples were frozen at -80 °C until the RNA extraction. This extraction was performed using the miRNeasy Mini KIT (Qiagen) modified to its use in adipose tissue. Once the ribonucleic acid was obtained, the cDNA library construction and sequencing were conducted at Novogene in Cambridge (UK). For RNA-Seq, a Novaseq 6000 (Illumina) sequencer was used, generating stranded paired-end reads of 150 bp.

Bioinformatic pipeline. Quality control of raw sequences was performed using *Trimmomatic* (v.0.39) (Bolger *et al.*, 2014). We used *STAR* (v.2.7.6a) (Dobin *et al.*, 2013) software to map the lectures against the ovine reference genome (Oar_rambouillet_v1.0_r101). The quantification of the reads was performed by *HTSeq-counts* (v.0.6.1p1) (Anders *et al.*, 2015). *DESeq2* R package (v.1.32.0) (Love *et al.*, 2014) was used to perform the differential expression analysis (DEA) between the conditions analysed in the study. Using the normalized rlog transformed gene counts from *DESeq2,* we performed a principal component (PC) analysis to evaluate the distribution of the samples for the whole transcriptome and detect possible outliers. The model for *DESeq2* was fitted assuming an interaction between the NC and the sex:

$$y = NC + sex + NC:sex \quad (1)$$

We focused the analysis on the interaction between both conditions to establish the different responses to the treatment between the sexes. Differentially expressed genes (DEG) were considered at a false discovery rate (FDR)<0.05. Network analysis of DEG was performed using the *STRING* database (v.11.5) (Szklarczyk *et al.*, 2019) to identify protein-protein interactions (PPI). We used the 'co-expression' interaction as the source for information criteria, and interactions were considered relevant using a high confidence (0.07) score as defined by *STRING*. Gene Ontology (GO) terms from the functional enrichment analysis of the network were considered significantly enriched when the FDR<0.05.

Results

For the perirenal fat transcriptome of the 12 suckling lambs, the RNA-Seq analyses yielded an average of 32 million (32,147,225.17±1,125,254.99) paired-end reads per sample. Of those, ~87% paired-end reads per sample aligned to unique loci in the Oar_Ram_v1.0_r101 reference genome. The PC analysis helped us to identify an outlier in the C female group. After discarding this sample, the PC analysis was repeated with the remaining 11 samples. The three first PCs explained only 47% of the variance, being the 70% of the variance explained by the first six PCs. Moreover, no clustering of the samples into defined groups was observed. We detected 163 DEG when we analysed the interaction between sex and NC. Of them, 116 DEG showed an opposite expression pattern due to the NC in females and males. Thus, these genes had an augmented expression in C females compared to the NC females, while for males, they were upregulated in the NC group in comparison to the control group. In the gene network analysis, a total of 298 co-expression interactions (PPI enrichment *P*-value<1.0e-16) were identified among 94 of the 116 DEG showing an opposite expression pattern between females and males lambs from NC ewes. Functional enrichment analysis indicated that these co-expressed DEG were significantly enriched in 63 terms in the GO:Biological Process category, 16 in the GO:Molecular Function, and 38 in the GO:Cellular Component, being the top enriched GO terms for each category oxidative phosphorylation (37 genes), oxidoreductase activity (43 genes) and mitochondrion (74 genes), respectively.

Discussion

Results from our study highlight a set of genes within the PAT transcriptome whose expression patterns are opposite for male and female suckling lambs when their mothers have been challenged with a dietary restriction at the prepuberal growth stage. PPI network analysis showed that most of these genes were co-

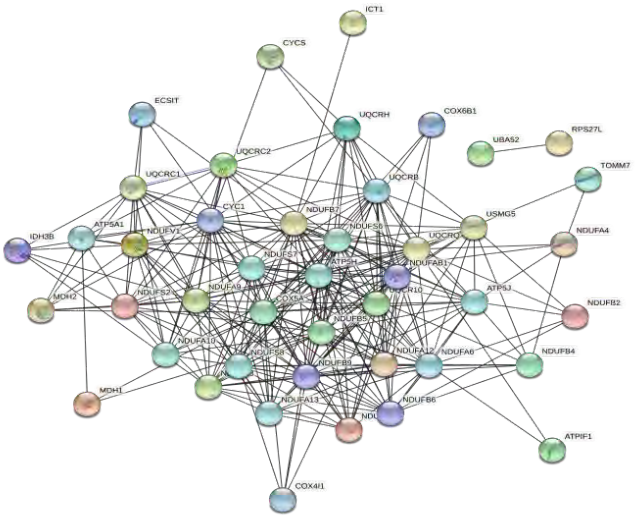


Figure 1. Protein-protein interaction network constructed with 94 co-expressed genes from the 116 DEG showing an opposite expression pattern between the male and female lambs here analyzed.

expressed and involved in functions related to mitochondrion activity (Figure 1). Brown adipose tissue (BAT) is characterized by a higher number of mitochondria compared to white adipose tissue. Mitochondrial energy expenditure in BAT is used primarily by non-shivering thermogenesis (Lee *et al.*, 2019). There is controversy about BAT deposits in suckling lambs; while some studies have observed that at 20 days of age, there is no BAT (Basse *et al.*, 2015), other studies have shown that there are BAT gene markers in lambs of the same age and older (Suárez-Vega *et al.*, 2018). One of the principal molecular markers for BAT is *UCP1* expression, whose expression in our samples was 369.5±394.63 FPKM in females and 98.06±111.03 FPKM in males. In our research, those females whose mothers were subjected to the nutritional challenge and the males whose mothers were from the control group showed a higher expression of genes that participate in oxidative phosphorylation and ATP synthesis compared with the alternative groups. In women, it has been reported that the activity and mass of BAT are greater than in men (Perkins *et al.*, 2013). Justo *et al.* (2005) detected a higher oxidative capacity in rat females. Cytochrome c oxidase (COX) is the enzyme responsible for oxygen consumption by mitochondrial oxidative phosphorylation. Our study found several COX subunits (*COX4I1*, *COX5A*, *COX8A*, *COX6B1*) upregulated in NC females and C males.

From our results, it could be inferred that a nutritional restriction in dairy ewes at the prepuberal stage has a sex-dependent effect on the transition from BAT to white adipose tissue of their offspring. While this process may be accelerated in NC male lambs, it seems to be delayed in female lambs. Other studies in rodent models and humans have showed that the influence of the maternal diet on the offspring is sex-dependent (Almeida *et al.*, 2017, Eriksson *et al.*, 2010). Overall, this is the first study in sheep showing the potential sex-dependent effects of maternal nutrition on fat deposition metabolism.

Acknowledgements

This study has received funding from the Spanish Ministry of Science and Innovation through the project RTI2018-093535-B-I00.

References

Ahmad, S., Drag, M.H., Salleh, S.M., Cai, Z., Nielsen, M.O. (2021) BMC Genomics 221(22): 1–21. https://doi.org/10.1186/S12864-021-07672-5

Almeida, M.M., Dias-Rocha, C.P., Souza, A.S., Muros, M.F., Mendonca, L.S. *et al.* (2017) Br. J. Nutr. 118: 788–803. https://doi.org/10.1017/S0007114517002884

Anders, S., Pyl, P.T., Huber, W. (2015) Bioinformatics 31: 166–169. https://doi.org/10.1093/bioinformatics/btu638

Basse, A.L., Dixen, K., Yadav, R., Tygesen, M.P., Qvortrup, K. *et al.* (2015) BMC Genomics 1–19. https://doi.org/10.1186/s12864-015-1405-8

Bolger, A.M., Lohse, M., Usadel, B. (2014) Bioinformatics 30: 2114–2120. https://doi.org/10.1093/bioinformatics/btu170

Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C. *et al.* (2013) Bioinformatics 29: 15–21. https://doi.org/10.1093/bioinformatics/bts635

Eriksson, J.G., Kajantie, E., Osmond, C., Thornburg, K., Barker, D.J.P. (2010) Am. J. Hum. Biol. 22: 330–335. https://doi.org/10.1002/AJHB.20995

Fan, H., Hou, Y., Sahana, G., Gao, H., Zhu, C. *et al.* (2019) Anim. 9: 655. https://doi.org/10.3390/ANI9090655

Frank, A.P., De Souza Santos, R., Palmer, B.F., Clegg, D.J. (2019) J. Lipid Res. 60: 1710–1719. https://doi.org/10.1194/JLR.R086975

Henry, B.A., Pope, M., Birtwistle, M., Loughnan, R., Alagal, R. *et al.* (2017) Endocrinology 158: 2212–2225. https://doi.org/10.1210/EN.2017-00081

Ithurralde, J., Pérez-Clariget, R., Corrales, F., Fila, D., López-Pérez, Á. *et al.* (2019) Livest. Sci. 221: 105–114. https://doi.org/10.1016/J.LIVSCI.2019.01.024

Justo, R., Frontera, M., Pujol, E., Rodríguez-Cuenca, S., Lladó, I. *et al.* (2005) Life Sci. 76(10): 1147-1158. https://doi.org/10.1016/j.lfs.2004.08.019

Karastergiou, K., Smith, S.R., Greenberg, A.S., Fried, S.K. (2012) Biol. Sex Differ. 3: 1–12. https://doi.org/10.1186/2042-6410-3-13

Khanal, P., Pandey, D., Binti Ahmad, S., Safayi, S., Kadarmideen, H.N. *et al.* (2020) Physiol. Rep. 8: e14359. https://doi.org/10.14814/PHY2.14359

Lee, J.H., Park, A., Oh, K.J., Lee, S.C., Kim, W.K. *et al.* (2019) Int. J. Mol. Sci. 20. https://doi.org/10.3390/IJMS20194924

Love, M.I., Huber, W., Anders, S. (2014) Genome Biol. 15. https://doi.org/10.1186/s13059-014-0550-8

Perkins, A.C., Mshelia, D.S., Symonds, M.E., Sathekge, M. (2013) Nucl. Med. Commun. 34: 168–174. https://doi.org/10.1097/MNM.0B013E32835BBBF0

Suárez-Vega, A., Arranz, J.J., de la Fuente, L.F., Mateo, J., Gutiérrez-Gil, B. (2018) Anim. Genet. 49: 605–617. https://doi.org/10.1111/age.12725

Symonds, M.E., Pope, M., Budge, H. (2015) Annu. Rev. Nutr. 35: 295–320. https://doi.org/10.1146/annurev-nutr-071813-105330

Szklarczyk, D., Gable, A.L., Lyon, D., Junge, A., Wyder, S. *et al.* (2019) Nucleic Acids Res. 47: D607–D613. https://doi.org/10.1093/nar/gky1131

742. Measures of diversity based on autosomal and sex chromosomes of Valachian sheep

M. Mészárosová[1*], N. Moravčíková[1], G. Mészáros[2], I. Pavlík[3], M. Margetín[1] and R. Kasarda[1]

[1]Slovak University of Agriculture in Nitra, Tr. A. Hlinku 2, 949 76 Nitra, Slovakia; [2]University of Natural Resources and Life Sciences, Vienna, Gregor-Mendel-Straße 33, 1180 Vienna, Austria; [3]Research Institute for Animal Production Nitra, Hlohovecká 2, 951 41 Lužianky, Slovakia; xmeszarosovm@uniag.sk

Abstract

The genomic inbreeding and heterozygosity levels were analysed in the Original and Improved Valachian sheep, and eight other breeds from open access data sets. The autosomal diversity levels were comparable between breeds. The markedly higher heterozygosity in the Improved Valachian was likely due to breed formation history, which ended ca. 40 years ago, and involved crossbreeding strategies. The heterozygosity levels in the X chromosome were clearly lower from those on the autosomes in all examined breeds. In particular, the dairy and meat Lacaune, and the Border Leicester breeds exhibited almost no heterozygosity here. In the Slovak autochthonous breeds, the heterozygosity levels were comparable to autosomes. The results of this study will be used to further improve the diversity of Valachian sheep.

Introduction

The use of single nucleotide polymorphism (SNP) markers has become common practice in sheep breeding. One of the most popular ways to utilize genomic data is to assess genomic diversity. This is especially important in small breeds, as the proper management of genetic diversity is the most important tool that keeps these breeds from the downward spiral in an extinction vortex.

The aim of our paper was to analyse the genomic diversity in two autochthonous Slovak national breeds, the Original Valachian (OVA) and Improved Valachian (IVA) sheep.

Materials & methods

Heterozygosity and inbreeding levels were computed separately on various parts of the genome to assess the diversity of OVA and IVA sheep, which are of triple production focus (milk, meat, wool). Therefore, Lacaune – milk (LMI) and East Friesian Brown (EFB) as dairy breeds, Lacaune – meat (LME) and German Texel (GTX) as meat breeds, Australian Merino (AUM) and Border Leicester (BRL) as wool breeds (Kijas *et al.* 2012) were used for comparison. Sheep outgroups, European Mouflon (*Ovis orientalis orientalis*, EMU) and Argali (*Ovis ammon*, OAM) were also used for the comparison (Ciani *et al.* 2015).

SNP data from Illumina OvineSNP50v1 chips were used for the analysis. The data set before quality control consisted of 103 LMI and 39 EFB; 78 LME and 46 GTX; 50 AUM and 48 BRL; and 24 EMU and 6 OAM individuals. The Valachian sheep were genotyped using the GeneSeek GGP Ovine 50K chip, with 96 OVA and 69 IVA animals genotyped. Due to only a partial overlap in the SNP sets, the data sets were not merged together, but the computations were done separately for each breed.

In all cases, quality control consisted of removing SNPs and animals for which more than 10% of genotypes were missing. After the quality control 46,812 SNPs remained for the open access data, and 42,595 for the Valachian data set. Only one IVA animal was removed due to missing genotype data. The genomic diversity was assessed separately on autosomal chromosomes and the X sex chromosome.

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_742, © M. Mészárosová *et al.* 2022

The PLINK (Chang *et al.*, 2015) software was used for general data management, quality control, and to compute heterozygosity and allele frequency based inbreeding levels on the autosomal genome using the --het option. This option was limited to autosomal genome only, so allele frequencies were computed with in-house scripts. The ROH based inbreeding was computed using the cgaTOH (Zhang *et al.*, 2013) software, considering ROH with minimal length of 1 Mb, and not allowing for any missing or heterozygous SNPs.

Results & discussion

The overall inbreeding coefficients based on the autosomal genome were widely distributed (Figure 1). The IVA population appeared to be much less inbred compared to most of the other analysed sheep, likely due to the targeted crossbreeding activities that were used to form this breed. The IVA breed was a result of the upgrade of the autochthonous OVA breed, when its overall high-altitude adaptation and disease resistance traits were augmented with increased production, using other breeds such as Texel, Hampshire Down, Lincoln Longwool, Leicester Longwool, Lacaune, and East Friesian. Such activities, carried out between 1950s and 1980s clearly resulted in a higher number of haplotypes and lower chance of even small ROH to occur. The OVA breed on the other hand, consisted of animals with numerous ROH regions, hence a very wide range of inbreeding levels.

The heterozygosity levels across the autosomes had an expected tendency, compared to inbreeding levels. The correlation coefficient between the ROH based inbreeding and heterozygosity was -0.87, between allele frequency-based inbreeding and heterozygosity -0.99. The difference between the two sets of inbreeding results was likely based on the difference in computation methodologies. Not all homozygotes were considered in ROH estimations, as part of the runs was broken by missing or heterozygous SNPs.

The expected and observed heterozygosity was 0.391 and 0.388 in OVA, and 0.425 and 0.429 in IVA breeds. The overall observed heterozygosity levels in autosomes of livestock breeds ranged on average from 0.299 in BRL to 0.429 in IVA. The heterozygosity levels were much lower in the outgroups overall, with just 0.200 in EMU and 0.106 in OAM. It is not clear, however, if the increased number of homozygotes were due to real inbreeding, or due to ascertainment bias.

The diversity of X chromosome seems to be an under researched topic based on the overall scarcity of publications, yet it was suggested as a unique tool for population-genetic studies (Schaffer, 2003). The heterozygosity levels on the X chromosome were very different from those on the autosomes, however, similarly to findings of Chen *et al.* (2018). Here, a large proportion of the sheep from open access data sets had heterozygosity level of zero, i.e. they were entirely homozygous. Also, smaller portion of OVA animals showed a very low heterozygosity level. The rest of the data showed a similar, or even higher heterozygosity on X chromosome, compared to autosomes.

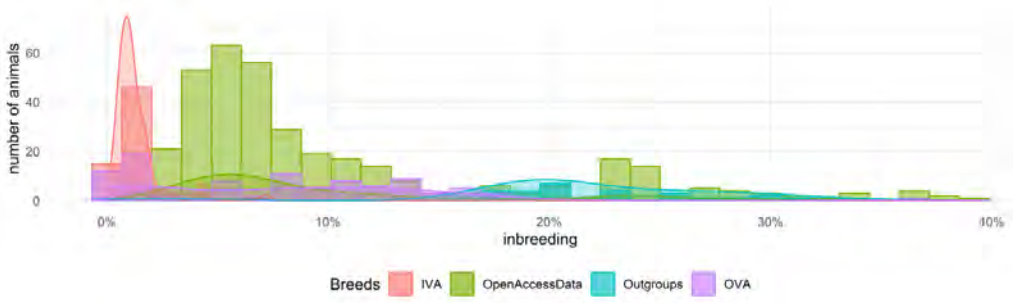


Figure 1. Genomic inbreeding coefficients based on ROH.

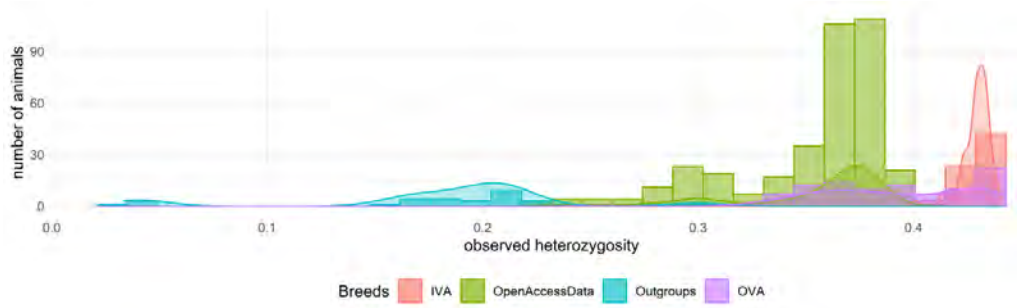


Figure 2. Observed heterozygosity levels on the autosomal chromosomes.

The reason for such difference in homozygosity levels was unclear. One possible reason could be the existence of long conserved regions on the X chromosome, as shown e.g. in pigs (Tong *et al.*, 2020), with potential candidate genes for body size. Of course, ascertainment bias in outgroups could also play a role, which would explain the low heterozygosity levels on the X chromosome of EMU and OAM (Table 1). On the other hand, even lower heterozygosity levels were found in established breeds, such as LME, LMI, and BRL. Also, the heterozygosity in other open access data was on average lower compared to autosomes in the same breed. In this respect, there was a notable variation in all breeds showed by high standard deviations, except BRL. Interestingly, both OVA and IVA had comparable average heterozygosity levels on the X chromosome and autosomes. Also here, the higher variation indicated that few animals had much lower heterozygosity levels, as also shown on Figure 3.

In conclusion, the inbreeding and heterozygosity levels in OVA population was comparable to other major sheep breeds. The higher heterozygosity levels in IVA were likely a consequence of the crossbreeding activities upon breed formation about 40 years ago, which still seem to have lasting effect in the population diversity metrics. These results will be used to further improve the diversity of the Slovak autochthonous breeds.

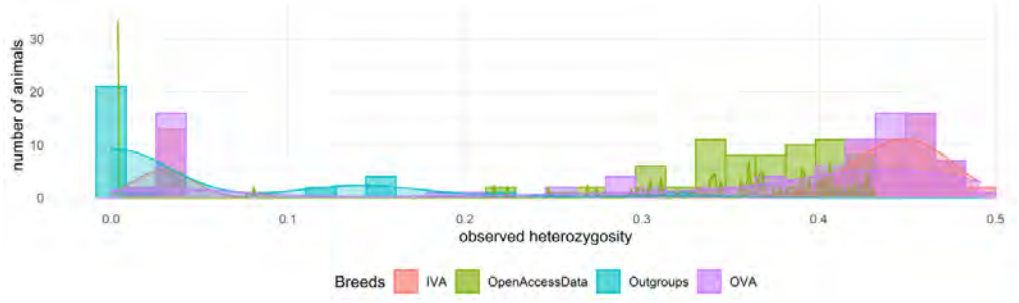


Figure 3. Observed heterozygosity levels on the X chromosomes. The y axis was truncated to improve clarity – the number of individuals with zero heterozygosity from the open access was ~300.

Table 1. Measures of autosomal and X chromosome diversity in sheep populations – mean (standard deviation).[1]

	Slovak breeds		Meat breeds		Dairy breeds		Wool breeds		Outgroups	
	OVA	IVA	LME	GTX	LMI	EFB	AUM	BRL	EMU	OAM
F (PLINK)[2]	0.066	-.028	0.057	0.098	0.065	0.237	0.064	0.247	0.496	0.733
	(0.087)	(0.016)	(0.032)	(0.069)	(0.016)	(0.118)	(0.042)	(0.029)	(0.074)	(0.275)
F (ROH)[2]	0.077	0.013	0.055	0.111	0.060	0.211	0.070	0.238	0.208	0.237
	(0.063)	(0.011)	(0.028)	(0.063)	(0.013)	(0.108)	(0.038)	(0.028)	(0.047)	(0.127)
Heterozygosity[2]	0.390	0.429	0.374	0.358	0.371	0.303	0.327	0.299	0.200	0.106
	(0.036)	(0.006)	(0.013)	(0.027)	(0.006)	(0.047)	(0.016)	(0.012)	(0.029)	(0.109)
F (ROH)[3]	0.317	0.295	0.851	0.536	0.837	0.497	0.630	0.886	0.456	0.352
	(0.364)	(0.386)	(0.169)	(0.386)	(0.198)	(0.375)	(0.392)	(0.005)	(0.137)	(0.191)
Heterozygosity[3]	0.339	0.352	0.016	0.189	0.023	0.180	0.126	0.001	0.044	0.059
	(0.161)	(0.175)	(0.076)	(0.186)	(0.088)	(0.168)	(0.185)	(0.001)	(0.083)	(0.091)

[1] Leading zeros were omitted due to space limitations.

[2] Considering all 26 autosomes.

[3] Considering only the X chromosome.

Acknowledgements

The authors wish to acknowledge the support of Slovak Research and Development Agency via the funding of the projects APVV-17-0060 and APVV-20-0161, which made the publication of this paper possible.

References

Chang C.C., Chow C.C., Tellier L.C.A.M., Vattikuti S., Purcell S.M. *et al.* (2015) Gigascience 4(1):s13742-015-0047-8. https://doi.org/10.1186/s13742-015-0047-8

Chen Z.H., Zhang M., Lv F.H., Ren X., Li W.R. *et al.* (2018) Genome Biol. Evol. 10(5):1282–1297. https://doi.org/10.1093/gbe/evy085

Ciani, E., Lasagna, E., D'Andrea, M., Alloggio I., Marroni F. *et al.* (2015) Genet. Sel. Evol. 47:64. (2015). https://doi.org/10.1186/s12711-015-0139-z

Kijas J.W., Lenstra J.A., Hayes B., Boitard S., Porto Neto L.R. *et al.* (2012) PLoS Biol. 10(2): e1001258. https://doi.org/10.1371/journal.pbio.1001258

Schaffner, S. (2003) Nat. Rev. Genet. 5:43–51. https://doi.org/10.1038/nrg1247

Tong, X., Hou, L., He, W., Mei C., Huang B. *et al.* (2020) Sci. Rep. 10:9433. https://doi.org/10.1038/s41598-020-66061-2

Zhang L., Orloff M.S., Reber S., Li S., Zhao Y. *et al.* (2013) PLoS ONE 8(3):e57772. https://doi.org/10.1371/journal.pone.0057772

743. The data quality score: objective assessment of data quality for Australian sheep breeders

D.J. Brown[1*], E.J. McCrabb[2], P.E. Bradley[2], I.J. Rose[2], R.G. Banks[1] and S.Z.Y. Guy[1]

[1]AGBU, a joint venture of NSW Department of Primary Industries and University of New England, 2351, Armidale, Australia; [2]Meat and Livestock Australia, The Short Run, University of New England, 2351, Armidale; dbrown2@une.edu.au

Abstract

Data quality influences the accuracy of estimated breeding values, and hence the accuracy of selection and genetic progress. This paper describes how a data quality score (DQS) was created to characterise the overall data quality for Australian sheep flocks. The data quality metrics investigated captured data quantity (e.g. degree of performance/pedigree recording), accuracy, structure (e.g. sire representation and linkage) and timeliness of data submission. These metrics were combined to calculate the overall DQS, with weightings dependant on breed type, variation and relationships between metrics. The DQS was well received by industry, with overwhelming support to publish the DQS after an initial producer education campaign. The DQS will help identify and highlight breeders who collect high quality data, help breeders improve their data quality, and increase information available to ram buyers. This research supports Australia's sheep industry to value phenotypes in an objective manner.

Introduction

Data is the cornerstone of genetic evaluation systems. Maximising data quality enhances accuracy of estimated breeding values and of selection decisions, which drives higher rates of genetic progress. The level and breadth of recording, and the availability and accuracy of supporting information (pedigree, fixed effects and genotypes), also impact selection accuracy and genetic progress. The genetic evaluation for the Australian sheep industry, Sheep Genetics, primarily relies on data submitted by ram breeders. While there are standards and guidelines for data submission, there is wide variation in data recording and submission across the industry.

Sheep Genetics make available 'RAMping Up Genetic Gains' (RUGG) reports, which provide breeders with feedback on their data (Stephen *et al.*, 2018). While these RUGG reports are available to breeders and have been used as an engagement tool for service providers and consultants, uptake by industry has been varied. As an incentive to seedstock breeders to improved recording, a data quality score (DQS) was developed to describe (and potentially rank) the quality of data submitted for genetic evaluation. This paper describes metrics that characterise data quality, the development of the DQS, and the response to the DQS by Australian sheep breeders.

Materials & methods

Data quality metrics. The following data quality metrics were calculated for each flock. Full Pedigree (%): the overall proportion of animals from the flock in the analyses with full pedigree. Contemporary group variation in ages (%): proportion of animals recorded that are in contemporary groups with variation in age. Variation in age within contemporary groups is expected with accurate birth date recording. Proportion of effective progeny (%): calculated as the proportion of progeny each sire has in the group, averaged across sires and all contemporary groups. Proportion of animals recorded for the weight, reproduction, wool and carcase trait groups (%). Average proportion of animals recorded that are directly linked to external flocks, by trait group (%): weight, carcase scan, reproduction and wool. Timeliness of data submission was calculated as interval between the date of recording and actual submission date (days). Calculation of the

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_743, © D.J. Brown *et al.* 2022

data quality metrics that capture the above factors are provided in more detail in Guy and Brown (2021a). These metrics were calculated for 441 Terminal, 304 Merino, and 96 Maternal flocks.

The data quality score. The metrics above were combined into an overall DQS for each flock. All metrics were weighted considering breed type (Terminal, Maternal or Merino), variation and relationships within the data quality metrics and relationships between data quality metrics and metrics describing genetic gains (Guy and Brown, 2021b). Due to the different scales and variances of each data characteristic metric, they were scaled to unit variance and a mean of zero. The resulting DQS for each method was also then scaled to between 0 and 100 for ease of interpretation.

Consultation with industry. The reporting of the data quality metrics and DQS was evaluated at 6 industry events involving 96 flocks. DQS reports were generated and incorporated into the existing RUGG reports. Reports were provided for each participating flock, followed by group discussion and collection of anonymous feedback.

Results

Data quality metrics. The data quality metrics with notable variation are shown in Figure 1.

While distributions for different breed types overlapped, Merino flocks had markedly lower full pedigree recording and more variation across flocks (mean ± SD, 43.6±35.3%) compared to Maternal (89.3±18.6%) and Terminal (87.2±21.6%). There was also less variation in ages within Merino contemporary groups. Given that the majority of lambs originate from natural mating this implies less accurate recording of birth dates.

The data quality metrics relating to data structure were fairly consistent across breeds. Average effective progeny numbers were 67.5±18.2% for Merino, 64.0±20.4% for Maternal, and 57.5±21.1% for Terminal flocks. The level of linkage for weight traits was 76.0±19.7% in Merino, 68.1±13.4% in Terminal and 64.5±16.9% in Maternal flocks.

All breed types had a high proportion of animals recorded for weight traits, with Merino flocks having the lowest level of recording, and the most variability across flocks within the breed type (75.2±25.1% compared to 79.8±15.9% for Maternal and 81.4±17.2% for Terminal flocks). Maternal flocks recorded more

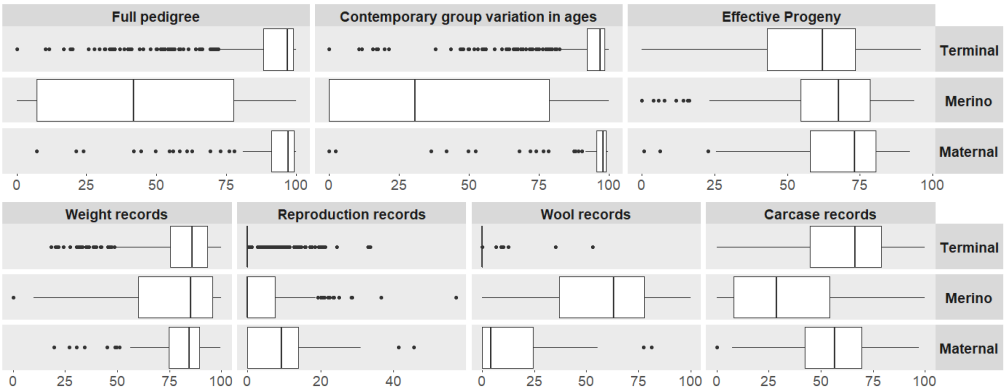


Figure 1. Distributions of data quality metrics for Terminal, Merino and Maternal flocks.

reproduction traits and had the most variation compared to other breed types (8.7±8.9%, compared to 4.2±7.6% for Merino and 1.8±4.9% for Terminal flocks). Merino flocks had a greater proportion of animals with wool traits recorded, although this varied greatly (59.3±25.9%). Terminal flocks had the highest level of carcase phenotype recording (61.4±20.1%, compared to 54.2±20.1% for Maternal and 33.2±28.4% in Merino flocks).

There was also variation in how timely data were submitted for analysis. Merino flocks took longer to submit pedigree records (average age of animals to appear in the pedigree was 376 days, compared to 240 days in Terminal and 243 days in Maternal flocks). The time taken to submit phenotypic data was similar across breeds for weight traits (submission time ranging from 92 days to 109 days post-measurement) and carcase data (range 38 days to 55 days).

The data quality score. Metrics relevant to each breed type were included in the calculation of the DQS (for example, level of recording and linkage for wool traits were not included for the DQS calculations for Terminal flocks). Specific metrics were also given higher weights to target improvements for each breed type (for example, degree of full pedigree recording was lowest for Merino flocks, so more emphasis was placed on this metric for breeders in that group).

The DQS was categorised as 'star ratings' by equally splitting the DQS range (score of 0-20: 1 star, 20-40: 2 stars, 40-60: 3 stars, 60-80: 4 stars, 80-100: 5 stars; Figure 2).

Consultation with industry. The DQS was well-received by industry. Constructive feedback was provided to further enhance the usefulness of the DQS and the reports developed to support it. The majority of anonymous survey respondents thought the DQS and star rating were useful and relevant to industry, and also a source of motivation to improve their own recording strategies (Figure 3).

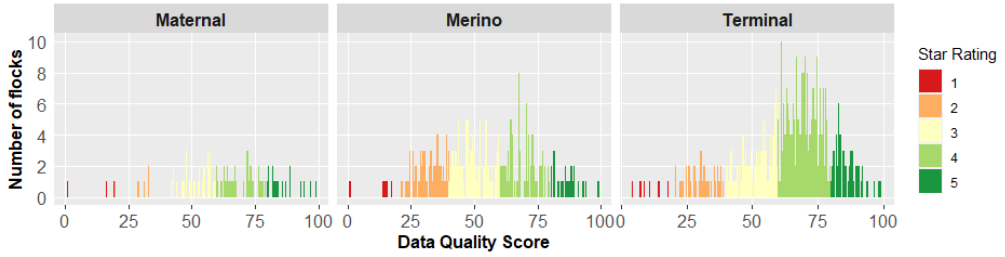


Figure 2. Distributions of data quality score for Terminal, Merino and Maternal flocks.

Figure 3. Level of agreement (green) and disgreement (red) for the usefulness of the Data Quality Score for improved recording, and relevance for industry (63 responses).

There was also overwhelming support to report the DQS to breeders, and to publish the data quality star ratings for each flock following an extension effort to help ram breeders and buyers understand the new information.

Discussion

Data quality metrics. The level of performance recording for each trait aligned with the general breeding objectives of each breed type. The high proportion of weight trait recordings across all breeds is likely to reflect the relative ease of recording compared to other traits. There was significant variation across breeds and flocks for all other trait groups, which highlights the opportunity for improvement in all breed types.

Delivering genetic progress through data quality. Data quality metrics were related to metrics describing genetic gains (average index accuracy, average index value and index trend) in Merino flocks (Guy and Brown, 2021b), as well as Maternal and Terminal flocks (unpublished). This suggests that improvement in data quality should lead to improvement in the rates of genetic progress.

Some aspects of data quality are captured in estimated breeding value (EBV) accuracies, which are reported alongside EBVs. While the accuracy of genetic merit estimation is influenced by the level of recording and management group structure (Brown *et al.* 2001; Swan and Brown, 2007), EBV accuracy is calculated using the amount and structure of information utilised (i.e. quantity), and not explicitly the precision and timeliness of information supplied to the analysis. The reporting of additional data quality metrics beyond EBV accuracy provides breeders with targeted feedback to help them improve their data recording programs.

In addition to the general reporting of these new metrics to industry, more extension and decision support tools are required to help breeders implement sensible performance recording improvements. Further work planned is to conduct cost-benefit analysis of different recording options and embed this into tools which can be used by service providers as the basis for objective breeding program advice.

Benefits to industry. The DQS will help identify and highlight breeders who collect high quality data, help breeders improve their data quality, and increase information available to ram buyers. This research supports Australia's sheep industry to value phenotypes in an objective manner, and can be further expanded to evaluating a flock's data contribution to the reference population.

Acknowledgements

The development of the data quality metrics and DQS has been a long-term project and has received significant funding from MLA and input from Sheep Genetics staff and industry ram breeders. In particular the authors would like to acknowledge the significant contributions of Caris Cornwall (nee Jones), Luke Stephen, Gabrielle Sherring and Andrew Swan.

References

Brown D.J., Tier B and Banks R.B (2001) Proc. Of the 14th AAABG, Queenstown, New Zealand.

Guy S.Z.Y., and Brown D.J. (2021a) Proc. Of the 24th AAABG, Adelaide, Australia.

Guy S.Z.Y., and Brown D.J. (2021b) Proc. Of the 24th AAABG, Adelaide, Australia.

Stephen L.M., Brown D.J., Jones C.E. and Collison, C.E. (2018) Proc. of the 11th WCGALP, Auckland, New Zealand.

Swan A.A. and Brown D.J. (2007) Proc. Of the 17th AAABG, Armidale, Australia.

744. Impact of genotypic information on genetic evaluation accuracy for mastitis and footrot in the UK Texel Sheep

K. Kaseja[1], S. Mucha[1], E. Smith[2], J. Yates[2], G. Banos[1] and J. Conington[1]

[1]SRUC Easter Bush, Roslin Institute Building, Edinburgh EH25 9RG, United Kingdom; [2]The British Texel Sheep Society, Stoneleigh Park, Warwickshire, CV8 2LG, United Kingdom; karolina.kaseja@sruc.ac.uk

Abstract

Estimation of breeding values (EBVs) is combined with estimation of accuracies for particular traits. These accuracies are important as they can tell the breeder how trustworthy the EBVs are. For hard to measure, low heritability traits or those measured at the end of life, EBV accuracy values tend to be low. This study aimed to examine the impact of including genotypic information in genetic evaluations to improved EBV accuracy. We analysed California mastitis test (CMT) and footrot (FR) records and over 10,000 genotypes collected on more than 30 UK Texel sheep farms. Results showed that the inclusion of genomic information based on Single Nucleotide Polymorphism (SNP) may increase the accuracy of prediction, especially for non-phenotyped animals.

Introduction

Mastitis and footrot are endemic diseases causing significant losses to the UK sheep industry (Nieuwhof and Bishop, 2007), impacting on productivity and often causing poor welfare of the animals. Phenotypes for both these traits can only be collected either late in life on lactating ewes (mastitis) and/or animals under disease challenge (footrot).

Heritabilities estimated for these or similar health traits tend to be small (McLaren *et al.*, 2018; Mucha *et al.*, 2015) meaning the obtained accuracies for EBVs remain low, if insufficient phenotypes recorded on close relatives are available. Higher accuracy means less room for bias, making the EBV more reliable, which is extremely important for the breeders. The systematic collection of these records in a cohort of animals (reference population or a network of genetically linked phenotype farms), delivers EBVs with appropriate high accuracies to help identify animals that are more resistant to disease and have higher production value. The collection of these disease or gender-specific records can be difficult and expensive. Therefore, having the phenotyped population of animals that are genetically linked to the wider population under the selection is key to the dissemination of genetically resistant livestock to disease. The development of genomic breeding values (GEBVs) by using phenotypes with corresponding genotypes to perform Best Linear Unbiased Prediction (BLUP) allows more animals to obtain more accurate breeding values for specific traits. This is especially important for young animals that not yet had a chance to be phenotyped due to their young age or to animals for which the phenotype can be collected on the opposite sex only. It is also important because some flocks may not have the capacity to accept levels of disease prevalence required for phenotyping, yet would benefit from being able to identify more resistant animals for further breeding. The aim of this study was to determine the difference in EBV accuracy values comparing genomic prediction (Single Step Genomic BLUP – ssGBLUP) with a conventional approach (BLUP).

Materials & methods

Genotypes. Texel sheep from participating flocks within the UK were genotyped between January 2015 and March 2019 using four Single Nucleotide Polymorphism (SNP) chip arrays, namely Illumina OvineHD BeadChip with 606,006 SNPs (HD), Illumina OvineSNP50 with 54,241 SNPs (50K), Illumina OvineLD BeadChip with 15,000 SNPs (LDv1) and Illumina OvineLD BeadChip with 16,560 SNPs (LDv2). Dataset contained 10,193 genotypes, which were subject to standard quality control. Firstly, genotypes that failed

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_744, © K. Kaseja *et al.* 2022

on call rate threshold were removed if the number of SNPs receiving the proper value (AA, AB, BA or BB) divided by the total number of SNPs was below 89.4%, which is the standard threshold used for the UK livestock genomic evaluations. Secondly, genotypes that were successfully assigned to registered animals in the UK National Texel (iTexel) database and had their pedigree verified were kept. After removing duplicate entries, the final dataset contained 9,391 genotypes which were subject to standard checks at the SNP level, investigating SNP call rate, minor allele frequency and Hardy-Weinberg equilibrium. This led to subsets of 45,686, 36,654, 12,427 and 10,725 SNPs for HD, 50K, LDv2 and LDv1 SNP arrays, respectively. These SNP subsets were then imputed using Findhap V3 software (VanRaden *et al.*, 2011) to 50K. For genotyped animals, five generations pedigree (n=48,244) were built using information from the iTexel database.

The imputed genotypes were used to assess the population structure of the breed using Principal Component Analysis (PCA) (Macciotta *et al.*, 2010; S Mucha *et al.*, 2015) in R software (R Core Team, 2021).

Phenotypes. Data collection was undertaken between 2015 and 2019 on 32 farms across Great Britain. Trained technicians visited farms at least once per year to score breeding females on presence of footrot on a five-point scale, where 0= no infection (healthy) and 4= severe infection (Conington *et al.*, 2008). Animals were scored between one and five times in total. Mastitis phenotypes were collected as California Mastitis Test (CMT) scores, which are highly correlated with somatic cell count causing mastitis (González-Rodríguez and Cármenes, 1996; McLaren *et al.*, 2018). In order to normalise the data, the collected scores were summed either for two halves of the udder for CMT or four feet for footrot (FRT). The sum was increased by one, to avoid sum of zero, and log transformed (McLaren *et al.*, 2018).

Variance components, heritabilities and breeding values. Genetic parameters used in this study were estimated using an animal model in ASReml (Gilmour *et al.*, 2009), with the following model: $y = Xb + Za + Wc + e$ where: y is the vector of phenotypic observations (CMT or FRT); b the vector of fixed effects, containing lambing number, litter size born, scorer, contemporary groups (month-year and farm-year) for CMT and dam age, scorer, foot vaccination status and flock for FRT; a is the vector of random animal effect; c is a vector for permanent environment (for CMT only); e is the vector of random residual effects; X, Z and W are the designed matrices for fixed (X) and random effects (Z and W).

Estimated Breeding Values were derived with BLUP (Henderson, 1949), while ssGBLUP (Legarra *et al.*, 2009; Misztal *et al.*, 2009) was used to obtain genomic EBVs, where the Genomic Relationship Matrix was constructed using VanRaden's first method (VanRaden, 2008). Calculations were undertaken in MiX99 software (Lidauer *et al.*, 2015). Breeding values, both conventional and genomic, were estimated using the same dataset including 821,692 animals from the iTexel database, containing genotyped and/or phenotyped animals with their pedigree data. In both cases, EBV accuracies were estimated with Apax99 software using a two-step method which in the first instance calculates information amount due to observations coming from defined model and secondly uses Misztal and Wiggans (1988).

Results

The breed composition of the studied population was examined by plotting first and second principal components of the genomic relationship matrix. Results, showing no major outlier groups, indicates that this population is mainly homogeneous, hence genomic analyses can be performed for all the animals included in this study. Over 35% of the variation can be explained by the first ten principal components, with 14.8 and 4.7% explained by first and second component, respectively.

Heritability estimates were 0.07 (±0.03) and 0.12 (±0.02) for CMT and FRT, respectively.

Figure 1 illustrates potential gain in accuracy for CMT and FRT for genotyped animals from ssGBLUP compared to BLUP. Animals with genotype available were divided into two groups: phenotyped and non-phenotyped; the average increase in accuracy was 0.06 and 0.08 for phenotyped and 0.21 and 0.14 for non-phenotyped animals for mastitis and footrot, respectively. Maximum accuracy gain was observed for non-phenotyped animals and was 0.52 for CMT and 0.47 for FRT. These results indicate that with a properly designed reference (genotyped and phenotyped) population, the use of genomic information is beneficial for non-phenotyped animals.

Discussion

Estimated heritabilities are in line with expectations for the health traits and other similar research (McLaren *et al.*, 2018; Mucha *et al.*, 2015; Kaseja *et al.*, 2018). The results from this study show that the estimated accuracy values increase both for CMT and FRT when genome-wide SNP genotypes were incorporated in the genetic evaluation for these health traits. This study proves that the incorporation of all available information for the prediction of genomic breeding values is highly valuable leading to more robust predictions of merit for breeding animals. It will also increase the trust in breeding values by farmers and breeders and reduce the risks associated with selection. The estimated heritabilities for disease traits were low, indicating that non-additive genetic and environmental effects are playing an important role for mastitis and footrot in sheep, although selection to reduce these diseases is valuable not least to improve animal health and welfare but also to reduce transmission and improve flock profitability and efficiency.

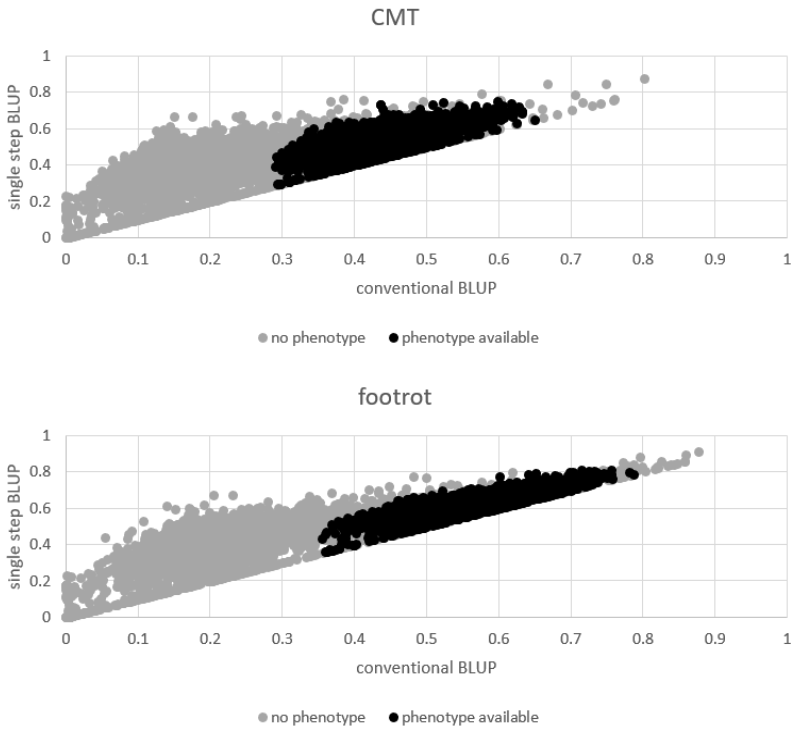


Figure 1. Accuracy values of EBVs obtained for genotyped animals for mastitis (CMT) and footrot while using conventional BLUP (horizontal axis) and single-step genomic BLUP (vertical axis).

Acknowledgements

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 772787 (SMARTER), InnovateUK project ID 102646 (BBSRC project no. BB M02833X/1) and InnovateUK project ID 131791.

References

Conington, J., Hosie, B., Nieuwhof, G.J., Bishop, S.C., Bünger, L. (2008) Veterinary Research Communications 32(8), 583–589. https://doi.org/10.1007/s11259-008-9062-x

Gilmour, A.R., Gogel, B.J., Cullis, B.R., Thompson, R. (2009). ASReml user guide release 3.0. VSN International Ltd. https://doi.org/10.1017/CBO9781107415324.004

González-Rodríguez, M.C., Cármenes, P. (1996) Small Ruminant Research 21(3), 245–250. https://doi.org/10.1016/0921-4488(95)00826-8

Kaseja K., McLaren, A., Yates, J., Mucha, S., Banos, G., Conington, J. (2018) Proc. of the 11th WCGALP, Auckland, New Zealand

Legarra, A., Aguilar, I., Misztal, I. (2009) Journal of Dairy Science 92, 4656–4663. https://doi.org/10.3168/jds.2009-2061

Lidauer M., Matilainen K., Mäntysaari E., Pitkänen T., Taskinen M. (2015) Technical Reference Guide for MiX99 Solver

Macciotta, N.P.P., Gaspa, G., Steri, R., Nicolazzi, E.L., Dimauro, C., *et al.* (2010) Journal of Dairy Science 93(6), 2765–2774. https://doi.org/10.3168/jds.2009-3029

McLaren, A., Kaseja, K., Yates, J., Mucha, S., Lambe, N., Conington, J. (2018) Animal 12(12), 2470–2479. https://doi.org/10.1017/S1751731118000393

Misztal, I., Wiggans G.R., (1988) Journal of Dairy Science 71: 27–32. http://dx.doi.org/10.1016/S0022-0302(88)79976-2

Misztal, I., Legarra, A., Aguilar, I. (2009) Journal of Dairy Science 92, 4648–4655. https://doi.org/10.3168/jds.2009-2064

Mucha, S, Mrode, R., MacLaren-Lee, I., Coffey, M., Conington, J. (2015) Journal of Dairy Science 98(11), 8201–8208. https://doi.org/10.3168/jds.2015-9682

Mucha, S., Bunger, L., Conington, J. (2015) Genetics Selection Evolution 47(1), 35. https://doi.org/10.1186/s12711-015-0119-3

Nieuwhof G., Bishop S. (2005) Animal Science 81(1), 23-29. https://doi.org/10.1079/ASC41010023

VanRaden, P.M., O'Connell, J.R., Wiggans, G.R., Weigel, K.A. (2011) Genetics Selection Evolution 43(1), 1–11. https://doi.org/10.1186/1297-9686-43-10

Van Raden, P. M. (2008) Journal of Dairy Science 91:4414–4423. https://doi.org/10.3168/jds.2007-0980

745. Characterization of homozygous haplotype deficiency compromising fertility traits in Manech Tête Rousse dairy sheep

M. Ben Braiek[1*], C. Moreno-Romieux[1], C. Hozé[2], J-M. Astruc[3] and S. Fabre[1]

[1]GenPhySE, Université de Toulouse, INRAE, ENVT, 31326 Castanet-Tolosan, France; [2]Allice, 149 rue de Bercy, 75595 Paris, France; [3]Institut de l'Elevage, 24 chemin de Borde-Rouge, 31321 Castanet-Tolosan, France; maxime.ben-braiek@inrae.fr

Abstract

With the objective to identify deleterious recessive mutations, we have scanned the genome of 5,271 French Manech Tête Rousse (MTR) dairy sheep using 50k SNP phased genotypes and pedigree data. Five deficient homozygous haplotypes, named MTRDHH1 to MTRDHH5, were identified. These haplotypes showed a highly significant deficit of homozygous animals ranging from 84 to 100%, and a carrier frequency ranging from 7.8 to 16.6%. By comparing at-risk matings (between carriers) and safe matings, some of these haplotypes were associated with a reduced success of artificial insemination (MTRDHH2), and/or an increased stillbirth rate (MTRDHH1 and 2) suggesting the segregation of recessive variants leading to embryo and/or perinatal lethality affecting fertility. Other haplotypes are likely to be associated with lamb morphological disorders or counter-selection based on breed standard. For each haplotype, the most probable candidate genes were highlighted based on their roles in lethal phenotypes, genetic disorders, hornless or coat color.

Introduction

In livestock under selection, the small effective population size and a non-negligible inbreeding rate could increase homozygosity and thereafter lead to the emergence of genetic disorders under the control of recessive deleterious variants. When homozygous, these variants very often lead to early embryo death, or to developmental disorders that affect fetuses and subsequently young individuals altering health and welfare (Georges *et al.* 2019). Detection of homozygous haplotype deficiency using reverse genetic screen is a useful method for mapping specific haplotypes associated with deleterious recessive mutations (VanRaden *et al.* 2011). The analysis is based on high throughput genotyping data usually available in the framework of genomic selection. For the first time in sheep, we recently validated this approach in Lacaune dairy sheep by the identification of eight independent deficient homozygous haplotypes (Ben Braiek *et al.* 2021). Complementary whole genome sequence studies have led to identify the causal variant in the *CCDC65* gene within one of these haplotypes (LDHH6, OMIA 002342-9940). Homozygous lambs for the *CCDC65* mutation suffered from a lethal respiratory syndrome (Ben Braiek *et al.* 2022). The objective of the present study is to search for homozygous haplotype deficiency in the Manech Tête Rousse (MTR) dairy sheep also having a large set of genotyping data from genomic selection. This opens the opportunity to discover new deleterious recessive mutations in sheep.

Materials & methods

Animal and genotyping data. The total dataset consists of 6,845 genotyped Manech Tête Rousse animals (82% male and 18% female) born between 1993 and 2021. Genotypes were obtained in the framework of the French dairy sheep genomic selection from Ovine SNP50 BeadChip (MD; n=3,889) and SheepLD (LD; n=2,956) (Illumina Inc., San Diego, USA). Quality control for each SNP was based on: (1) a call frequency >97%; (2) a minor allele frequency >1%; and (3) the respect of the Hardy-Weinberg equilibrium ($P>10^{-5}$). Phasing and imputation (from LD to MD SNP chip) of genotypes were done using *FImpute v2.2.* After quality control, the remaining 38,523 autosomal SNPs were mapped onto the *Ovis aries* genome assembly Oar v3.1.

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_745, © M. Ben Braiek *et al.* 2022

Detection of homozygous haplotype deficiency. The genome of 5,271 genotyped animals belonging to trios (77 offspring with both parents genotyped and 4,799 offspring with sire and maternal grandsire genotyped) was scanned by a 20 SNP sliding window as described (Ben Braiek *et al.* 2021). Consecutive 20 SNP windows with the same parameters (haplotype frequency, significant homozygous animal deficit) were clustered to define larger region called 'Manech Tête Rousse Deficient Homozygous Haplotype' (MTRDHH). For each MTRDHH, the genotypical status (homozygous non-carriers, heterozygous and homozygous carriers) was determined for all genotypes available (n=6,845).

Analysis of fertility traits. Trait records of MTR matings between 2006 and 2019 were obtained from the national database and based on ewes with a genotyped sire and genotyped rams (maternal sire and ram both having a known status at each MTRDHH; n=330,844 matings). We analyzed two fertility-associated traits, artificial insemination success (AIS) and stillbirth rate (SBR). We have considered 'at-risk' a mating between a carrier ram and a ewe from a carrier sire and the other combinations were 'safe' mating. AIS and SBR were considered to differ significantly when the fixed effect 'mating type' had a *P*-value lower than 1.0×10^{-3} after correcting for multiple testing with a level of significance at 1% following the SAS GLIMMIX procedure previously described (Ben Braiek *et al.* 2021).

Identification of candidate genes in MTRDHH regions. We extracted all annotated genes located in the MTRDHH region extended by 1 Mb from each side from the ovine genome Oar_v3.1 using CLC export annotation function (QIAGEN CLC Main Workbench 7.9). All genes related to mortality/aging phenotypes when knocked-out in mice and/or associated to mammalian genetic disorders were extracted from Mouse Genome Informatics (MGI, http://www.informatics.jax.org), Online Mendelian Inheritance in Man (OMIM, https://omim.org) and Online Mendelian Inheritance in Animal (OMIA, https://omia.org) databases.

Results

Identification of deficient homozygous haplotypes in MTR dairy sheep. Using a reverse genetic screen in MTR genotyped population, we identified 5 independent regions showing a significant deficit of homozygous animals and named MTRDHH (Table 1). Among those regions, three haplotypes (MTRDHH1, 2 and 3) showed a complete deficit of observed homozygous animal while 9 to 13 were expected. The two other haplotypes (MTRDHH4 and 5) presented a partial deficit between 84 and 91% (some homozygous animals were genotyped). The length of these haplotypes ranged from 1.8 to 4.2 Mb and carrier frequencies

Table 1. List of Manech Tête Rousse deficient homozygous haplotypes.

Haplotype	OAR	Number of markers	Position (Mb)[1]	Carrier frequency (%)[2]	Number of homozygotes			
					Exp[3]	Obs[4]	Deficit	Poisson *P*-value
MTRDHH1	20	32	21.4-23.2	9.7	13	0	100%	2.9×10^{-6}
MTRDHH2	1	66	229.2-233.4	8.7	10	0	100%	3.8×10^{-5}
MTRDHH3	1	39	97.2-99.8	7.8	9	0	100%	9.6×10^{-5}
MTRDHH4	10	26	28.2-29.9	8.7	11	1	91%	4.5×10^{-4}
MTRDHH5	13	29	61.3-64.0	16.6	49	8	84%	5.3×10^{-13}

[1] Position on ovine genome assembly Oar_v3.1.
[2] Whole genotyped population (n=6,845).
[3] Expected.
[4] Observed.

ranged from 7.8 to 16.6% in the whole genotyped population. MTRDHH2 and 3 are located on the same chromosome (OAR1) but are not in linkage disequilibrium. The other haplotypes are located on different chromosomes, MTRHH1 on OAR20, MTRDHH4 on OAR10 and MTRDHH5 on OAR13. Consequently, the five MTRDHH detected are likely to host five recessive deleterious variants.

Impact of MTRDHH on the success rate of AI and on stillbirth rate. In order to identify the putative lethal effect of the MTRDHH, we studied two fertility binary traits (AIS and SBR) by comparing at-risk matings to safe matings. AIS (population mean= 60.9%) is a good proxy for embryonic loss, while SBR (population mean=7.5%) is a good proxy for perinatal lethality. As shown in Figure 1, MTRDHH1 has a significant negative impact on SBR (+7.5% in at-risk matings) and MTRDHH2 have negative effects on both AIS (-3.3%) and SBR (+4.3%). The other haplotypes showed no significant effect on the two tested traits.

Candidate genes for MTRHH regions. Among the 459 genes annotated in the five MTRDHH regions, 61 are implicated in lethal phenotype when knocked-out in mice. Among them, several are associated also with mammalian genetic disorders (Table 2). The remaining candidate genes are involved in important functions such as metabolism (*MMUT* and *SSR3*), DNA/RNA processing or transcription (*TFAP2B*, *DHX36* and *BRCA2*), or cell signaling (*PKHD1*, *SV2A*, *ECM1*, *KL* and *RXFP2*).

Discussion

In this study, we have confirmed the usefulness of reverse genetic screen to detect homozygous haplotype deficiency in sheep, even with a more reduced dataset of genotyped trios (MTR~5,000) compared to our previous work (Lacaune ~15,000, (Ben Braiek *et al.* 2021)). However, using the same parameters we identified less significant haplotypes (5 in MTR vs 11 in Lacaune) with the lower haplotype carrier frequency at 7.8% compared to 3.7% in Lacaune.

Among the five MTRDHH, three of them are totally deficient in homozygous animals leading to the hypothesis that these haplotypes host recessive lethal mutations. Accordingly, MTRDHH1 and 2 are associated with defect in fertility traits, AIS and/or SBR. This lets us think that when homozygous these haplotypes induce embryonic or fetal/neonatal lethality. To support this hypothesis, we found obvious candidate genes in these regions (*MMUT, TFAP2B, PKHB1, DHX36* and *SSR3*) all associated with

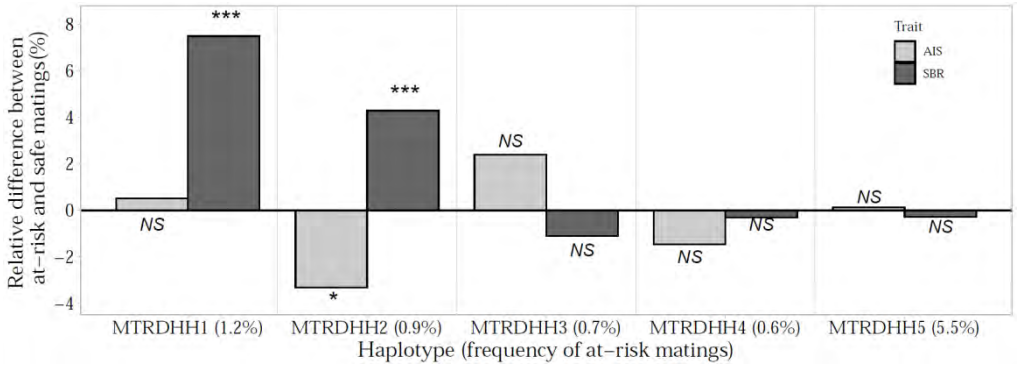


Figure 1. Effects of MTRDHH on the success rate of artificial insemination and on the stillbirth rate in at-risk matings compared to safe matings. *AIS*: artificial insemination success, *SBR*: stillbirth rate. Significant effects are indicated by the corrected *P*-value for multiple tests with a threshold set at α=1%: *$P>1\times10^{-3}$; **$P>1\times10^{-4}$; ***$P>1\times10^{-5}$. NS: not significant.

Table 2. Candidate genes located in MTRDHH regions.

Haplotype (phenotype)[1]	Gene name/Phenotype disorder in mouse and human
MTRDHH1 (Neonatal lethality)	*MMUT:* neonatal, postnatal lethality [MGI:97239]/ methylmalonic aciduria [OMIM# 609058]; *TFAP2B:* neonatal, postnatal, preweaning lethality [MGI:104672]/ patent ductus arteriosus, Char syndrome [OMIM# 601601]; *PKHD1:* perinatal, postnatal, preweaning lethality [MGI:2155808]/ polycystic kidney/ liver disease [OMIM# 606702]
MTRDHH2 (Embryo/fetus lethality)	*DHX36:* embryonic lethality [MGI:1919412]; *SSR3:* neonatal lethality [MGI:1914687]
MTRDHH3 (Juvenile lethality)	*SV2A:* premature death [MGI:1927139]; *ECM1:* preweaning lethality, premature death [MGI:103060]; skin/mucosa/viscera lipoid proteinosis [OMIM# 602201]
MTRDHH4 (Juvenile lethality, morphological disorder)	*KL:* premature death [MGI:1101771]/ tumoral calcinosis [OMIM# 604824]; *BRCA2:* postnatal, prenatal lethality, premature death [MGI:109337]/ Wilms tumor, Fanconi anemia D1 [OMIM# 600185]; *RXFP2:* Polled/Horns phenotype in sheep [OMIA 000483-9940]
MTRDHH5 (Morphological disorder)	*ASIP:* Coat color in sheep [OMIA 000201-9940]; skin/hair/eye pigmentation [OMIM#600201]

[1] Supposed deleterious phenotype in sheep.

embryonic, perinatal or neonatal lethality when altered in mouse or in human (Table 2). In contrast, MTRDHH3 showing also a complete deficit in homozygous animals was not associated with alteration of AIS or SBR. Since lambs are genotyped around three months of age, this haplotype could affect lambs after birth in the preweaning period as supposed by premature death caused by the alteration of the two candidate genes in this region (*SV2A* and *ECM1*, Table 2).

MTRDHH4 and 5 showed a partial deficit in homozygous animals and no significant effect on fertility traits. We hypothesize that these deficits are likely due to either a defect after birth in the juvenile period (alteration of *KL* or *BRCA2* genes for MTRDHH4) or counter-selection based on breed standard. Indeed, horned female and black animals are unwanted in MTR selection scheme. Interestingly, MTRDHH4 hosts the *RXFP2* gene known to be implicated in horn/polled phenotype (Wiedemar and Drögemüller 2015), and *ASIP,* known to be involved in coat color determinism, locates in MTRDHH5 (Norris and Whan 2008).

Acknowledgements

This research has received funding from the European Union's Horizon 2020 research and innovation program under the Grant Agreement No. 772787 (SMARTER), PRESAGE project (CASDAR No. 20ART1532777) and APIS-GENE and Région Occitanie.

References

Ben Braiek M, Fabre S, Hozé C, Astruc J-M and Moreno-Romieux C. (2021) *Genet Sel Evol* 53(1):41. https://doi.org/10.1186/s12711-021-00634-1

Ben Braiek M, Moreno-Romieux C, Allain C, Bardou P, Bordes A, *et al.* (2022) *Genes* 13(1):45. https://doi.org/10.3390/genes13010045

Georges M, Charlier C and Hayes B. (2019) *Nat Rev Genet* 20(3):135-56. https://doi.org/10.1038/s41576-018-0082-2

Norris BJ and Whan VA. (2008) *Genome Research* 18(8):1282-93. https://doi.org/10.1101/gr.072090.107

VanRaden PM, Olson KM, Null DJ and Hutchison JL. (2011) *Journal of Dairy Science* 94(12):6153-61. https://doi.org/10.3168/jds.2011-4624

Wiedemar N and Drögemüller C. (2015) *Animal Genetics* 46(4):457-61. https://doi.org/10.1111/age.12309

746. Genetic analysis of colostrum and passive immune transfer in divergent goat lines selected for functional longevity

M. Wicki[1,2*], C. Huau[1], T. Fassier[3] and R. Rupp[1]

[1]GenPhySE, Université de Toulouse, INRAE, Castanet Tolosan, France; [2]Institut de l'Elevage, Castanet-Tolosan, France; [3]Domaine de Bourges, INRAE, Osmoy, France; marine.wicki@inrae.fr

Abstract

The aim of this study is to find early indicators of longevity in two divergent goat lines selected for high or low functional longevity. Passive Immune Transfer (PIT) that is transfer of immunoglobulin (IgG) from mother to kid through colostrum could be one of these indicators. The first step of this project was to study the impact of PIT (n=422) on goat kid's health. Success of PIT had a positive influence on kid's health, which resulted in a decrease in the occurrence of several health problems. Then we estimated genetic parameters for colostrum quality (n=201): heritability was 0.19±0.16 for IgG concentration and 0.37±0.16 for IgG yield. PIT heritability (IgG in blood of kids) was lower: 0.04±0.07. Altogether, the study led to promising results regarding the possibility of integrating the colostrum quality traits into selection.

Introduction

In goat breeding, longevity can be defined as the number of days between first kidding and reform (Palhiere *et al.*, 2018). Functional longevity corresponds to longevity corrected for production level. Among the factors influencing longevity, genetics and breeding conditions in the young animals are critical. Goat kids health status is largely based on the success of passive immune transfer (PIT) that is transfer of immunoglobulin (Ig) from mother to kid. The main consequence of a failure of PIT is the increased risk of mortality; demonstrated in cattle (Imbert, 2005) and sheep (Bekele *et al.*, 1992) mainly due to digestive disorders. In small ruminants, PIT is strongly linked to the ingestion of a quality colostrum in sufficient quantity. Thus, one of the goals of this study was to determine (and quantify) the existence of a genetic component in the observable inter-individual variability in PIT and colostrum quality. Some literature data exist in sheep (Gilbert *et al.* (1998) and cattle (Cordero-Solorzano *et al.* (2021); Conneely *et al.* (2013)). However, such genetic variability has not yet been demonstrated and estimated in goats.

Materials & methods

Animals and phenotypes. This study is based on data from two divergent lines of Alpine goat. Following the method developed by Palhière *et al.*(2018), we made a hyper-selection of dairy goats based on the genetic evaluation for functional longevity of 8,787 Alpine AI bucks. From 2017 to 2021, divergent line goats were bred, at the same conditions,in an INRAE experimental Farm (Bourges, France): 221 low longevity line (low_LGV) and 228 high longevity line (high_LGV). At birth, vigor and weight of the goat kid, litter size and time between birth (actual or estimated hour) and colostrum intake were measured. A blood sample is taken between 1 and 8 days of age in order to proceed with the assay of IgG by radial immuno diffusion (n=422) and asses PIT, which is considered successful for an IgG concentration higher than 10 g/l in the serum. From birth to 4 weeks of age, sanitary measurements are carried out every week on the kids of both sexes (n=441). Finally, at birth the colostrum of the goats was collected during the first two milkings. At this time, the weight of colostrum and the time between calving and milking recorded. Individual samples of colostrum are collected for the assay of IgG by radial immune diffusion (97 low_LGV, 92 high_LGV and 12 other lines). Milk yield, fat and protein content, and SCC were measured every three weeks in first lactation, as part of the official milk records (n=201). Lower numbers are related to fertility or colostrum collection problems.

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_746, © M. Wicki *et al.* 2022

Animals were cared and handled in accordance with the French legislation on animal experimentation and European Convention for the Protection of Vertebrates Used for Experimental and Other Scientific Purposes (European Directive 86/609) under specific authorization number APAFIS#8613-2017012012585646.

Effect of immune traits on health status. The influence of PIT (fit as a covariate) and genetic line on goat kids health status was studied using a logistic regression model. The events tested were the occurrence (binary 0/1) of several health problems during the first four weeks of goat kids life: diarrhea, nasal discharge, ocular discharge, abnormal body temperature and obstructed breathing. A hindquarters cleanliness score was also studied using the dag score (http://www.flyboss.com.au/sheep-goats/breeding-and-selection/dag-scoring-and-selection.php, X3 classes), which reflects adhesion of fecal material to the breech area. No respiratory problems occurred so we did not analyze this health indicator. In addition to fixed effect of PIT and genetic line, a fixed effect of year of birth was included in the model for dag score.

Genetic parameters estimation. The genetic parameters were estimated for two traits related to colostrum quality (n=201): IgG concentration (IgGc) and yield (IgGy); and three traits related to milk production (n=201): milk yield (MY), the ratio between fat and protein contents (FC/PC) and the lactation somatic cell score (LSCS). We also analyzed the serum IgG concentration (PIT) measured in kids (n=422).

Since the data were not repeated (only one record per animal), all variables were analyzed using the following mixed polygenic model. The model included fixed effect of the year (4 levels from 2018 to 2021) for IgGc, FC/PC, the fixed effects of lactation number (3 levels) and litter size (3 levels) for IgGy, birth year (5 levels from 2017 to 2021) for PIT, and, in addition, age at first calving (3 levels) for MY, FC/PC and LSCS. The random effects included in the model were the additive genetic effect of the animal and the residual. For all these variables, the relationship matrix A was built tracing back 5 generations. The matrix A contains between 950 and 1,474 individuals depending on the trait analyzed. Heritabilities, the different variance components, genetic and phenotypic correlations between dairy and colostrum traits were estimated by bivariate analyses while PIT was studied by univariate analysis All the analyses were done using Wombat (Meyer, 2007).

Results & discussion

Health status. The results of logistic regressions on health indicators are summarized in Table 1. For the majority of health indicators, we note a significant effect of PIT which results in a decrease in the probability of health problems with increasing IgG concentration. For the occurrence of diarrhea, the odd ratio is 0.62, i.e. an increase of 1 g/l of serum IgG decreases by 38% the probability of occurrence of diarrhea during the first 4 weeks of life of the kids. This result agrees with the study of Basoglu and Camkerten (1999) who compared a group of calves with diarrhea and a group of healthy calves. Among the latter, none of them had a PIT failure compared to 35% for the group of diarrheic calves. For ocular and nasal discharges, the odds ratios are 0.41 and 0.51, respectively.

Table 1. Results of analyses of variance (logistic regression) for health indicators.[1]

	Dag score	Diarrhea	Nasal discharge	Ocular discharge	Abnormal temperature
PIT	$<10^{e}$-5	<0.005	$<10^{e}$-6	$<10^{e}$-5	NS
Year of birth	<0.005	NS	*	*	*
Genetic line	NS	NS	NS	<0.1	NS

[1] NS: The effect is non-significant. * effect not be tested because of occurrence disparity between years of birth.

Genetic parameters. Genetic parameters for colostrum and milk production traits are summarized in Table 2. For dairy traits, the heritabilities obtained are in agreement with those found in previous studies in the alpine breed: 0.3 for milk yield (Arnal *et al.*,2019) and 0.20 for the LSCS (Rupp *et al.*, 2011). The heritability of 0.19±0.16 obtained for IgG concentration is close to that found in sheep (0.19±0.12 in Gilbert *et al.*, 1988) and not significantly different from those of cattle (0.20±0.09 in Cordero-Solorzano *et al.* (2021). To our knowledge, there is no literature on genetic parameters for IgG yield. However, Conneely *et al.* (2013) found a heritability of 0.21±0.08 for colostrum weight in cattle, which is not significantly different from our result. Furthermore, the modest positive genetic correlation between MY and IgG yield of 0.36 suggests that genetic selection for milk production applied over many years has not led to a degradation of this second trait, and has even improved it. However, MY seems to be genetically independent of the IgG colostrum concentration. The positive genetic correlation between LSCS and colostrum traits is not negligible (0.58±0.93 for IgG concentration and 0.53±0.43 for IgG yield). The large estimation errors associated with these correlations question the reliability of these results, but if they are confirmed, it will be necessary to better understand why improved colostrum quality is genetically linked to increased mastitis risk. Finally, based solely on the relatively high heritabilities of IgG concentration and IgG yield we can assume that genetic selection on these two traits would be feasible. From an operational point of view, it seems possible to develop a large scale collection system for the concentration of IgG in colostrum using the refractometer (%Brix) for example, which is easier to measure than IgG yield. However, we can partly explain our high estimation errors by the small numbers studied. Therefore, it would be important to refine this estimate of heritability with more data, using additional animals and/or lactations. Indeed, this heritability may be weaker, as found (h^2=0.10) in another study in suckling cattle (Conneely *et al.*, 2013).

Passive immune transfer. The heritability of serum IgG concentration in kids was estimated to be 0.04±0.07. This low value may be due to many factors influencing this trait that we did not capture and analyze in our study, thus increasing the residual variability. Essentially, we can underline that our study does not consider the quality of the colostrum distributed which is likely to be very variable. Despite these untested or uncontrolled factors, the heritability obtained in our study agrees with other studies. Donovan *et al.* (1986) found a heritability of 0.12±0.08 for the IgG concentration and 0.09±0.07 for the total protein content of serum in calves. Haagen *et al.* (2021) estimated the heritability of total serum protein content to be 0.06±0.01 also in cattle. Gilbert *et al.* (1988) nevertheless found a higher heritability in sheep of 0.18±0.06, perhaps due to the large number of lambs (n=964) and the control of the distribution mode (suckling) and the age at blood sampling (36h after birth). Maltecca *et al.* (2007) also found a heritability of 0.18 for this trait in sheep.

Table 2. Genetic parameters for milk production and colostrum traits in Alpine goats. Heritabilities are on the diagonal, genetic correlations in the upper right part and the phenotypic correlations in the lower left part.[1]

	IgG concentration	IgG yield	MY	FC/PC	LSCS
IgG concentration	0.19±0.16	0.73±0.39	-0.15±0.49	-0.03±0.66	0.58±0.93
IgG yield	0.29±0.07	0.37±0.16	0.36±0.31	-0.05±0.51	0.53±0.43
MY	0.01±0.08	0.27±0.08	0.46±0.21	0.42±0.45	0.11±0.6
FC/PC	-0.11±0.08	0.10±0.08	0.06±0.08	0.19±0.16	*
LSCS	-0.05±0.08	-0.07±0.08	-0.05±0.08	*	0.13±0.15

[1] An asterisk indicates software did not reach the desired convergence; results are not interpretable.

Conclusions

Our study confirmed the positive impact of PIT on the health of the kids, resulting in a decrease in the probability of occurrence of various health problems, including digestive disorders. Moreover, IgG concentration and yield in colostrum showed moderate heritabilities. However, the heritability was lower for serum IgG concentration in kids. IgG yield and milk production seem to be genetically positively correlated with each other while IgG concentration has little correlation with milk production. Therefore, it could be useful to improve this trait by selection.

Acknowledgements

This study has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 772787 (SMARTER) and the national APIS-GENE (Paris, France) funded projects ActiveGoat and Resilait. The authors thank the staff of the INRAE experimental unit at Bourges for technical support in producing, raising, and monitoring the animals.

References

Arnal M., Larroque H., Leclerc H., Ducrocq V. Robert-Granié C. (2019) Genetics, Selection, Evolution 21(2):205-215. https://doi.org/10.1186/s12711-019-0485-3

Basoglu A., Camkerten I.M.S. (1999) Israel Journal of Veterinary Medicine 54: 9-10. https://www.researchgate.net/publication/233985645

Bekele T., Otesile E.B., Kasali O.B. (1992) Small Ruminants Research 9 (3): 209-215. https://www.sciencedirect.com/science/article/abs/pii/092144889290151S

Conneely M., Berry D.P., Sayers R., Murphy J.P., Lorenz I. *et al.* (2013) Animal 7(11):1824-1832. https://doi.org/10.1017/S1751731113001444.

Cordero-Solorzano J., De Koning D.J., Traven M., De Haan T., Jouffroy M. *et al.* (2021) bioRxiv. https://doi.org/10.1101/2021.08.08.455591

Donovan G.A., Badinga L., Collier R.J., Wilcox C.J., Braun R.K. (1986) Journal of Dairy Science 69:754-759. https://doi.org/10.3168/jds.S0022-0302(86)80464-7.

Gilbert R.P., Gaskins C.T., Hillers J.K., Parker C.F., McGuire T.C. (1988) Animal Science 66(4):855-863. https://doi.org/10.2527/jas1988.664855x

Haagen I.W., Hardie L.C., Heins B.J., Dechow C.D. (2021) Journal of Dairy Science 104(2):2018-2026. https://doi.org/10.3168/jds.2020-19080.

Imbert A. (2005) Ecole Nationale Vétérinaire d'Alfort, thèse pour le doctorat de vétérinaire.

Maltecca C., Rossoni A., Nicoletti C., Santus E., Weigel K.A. *et al.* (2007) Journal of Dairy Science 920(10):4814-4820. https://doi.org/10.3168/jds.2007-0183

Meyer, K. (2007). Journal of Zhejiang University Science B, 8(11): 815-821.

Palhière I., Oget C., Rupp R. (2018) Proc. of the 11th WCGALP, Auckland, New Zealand.

747. Selection signatures in South African indigenous sheep using the HAPFLK approach

I.N. Malan[1,2], S.W.P Cloete[1] and A.H. Molotsi[1*]

[1]Department of Animal Sciences, Stellenbosch University, Matieland, Private bag X1, 7600 Stellenbosch, South Africa; [2]Ministry of Agriculture, Water and Land Reform, Directorate of Agricultural Research and Development, Private Bag 13184, Namibia; annelind@sun.ac.za

Abstract

Selection sweeps have been identified in indigenous sheep breeds using the hapFLK approach. Five indigenous fat-tailed sheep breeds, the Karakul (70), Damara (35), Pedi (30), Zulu (30) and Namaqua Afrikaner (48), one composite fat-tailed breed, the Meatmaster (20), a commercial meat breed, the Dorper (46) and one exotic sheep breed, Merino (98) were genotyped using the 50K Ovine SNP beadchip. Thirteen regions spanning 11 chromosomes were under selection across the breeds. Putative candidate genes associated with the regions under selection involved immune response (BPIFC, CLDN12, CLDN14, TRL4), ultraviolet exposure and thermoregulation (KCNJ5, FZD1, ACADSB, SCNA11), energy metabolism (MYH9) and meat production (PDEZA, UTRN). These results confirm natural selection for adaptive traits such as immunity, heat tolerance and meat production in indigenous sheep breeds.

Introduction

Indigenous sheep breeds are known for their robustness and adaptiveness, potentially superseding exotic breeds especially in semi-arid environments in Southern African regions (Molotsi *et al.* 2017). To unpack the underlying genetic mechanisms associated with the adaptiveness of indigenous sheep; diversity studies, genome-wide association studies, selection sweeps and whole genome sequencing has been implemented in various breeds across the globe (Kijas *et al.* 2009; Kijas *et al.* 2012). Selection sweeps is a method to unveil regions in the genome that has been subjected to strong selection. Candidate genes linked to heat stress (DNAJC28), spermatogenesis and pigmentation (MREG) have been observed in South African indigenous sheep breeds (Molotsi *et al.* 2018). Selection sweeps associated with adaptive traits such as coat colour (ASIP, MC1R) have also been observed by Rochus *et al.* (2018). Selection sweeps in Iranian sheep have been analysed by Manzari *et al.* (2019), identifying candidate genes putatively linked to skeletal development (HOXB13), energy metabolism (CAV1), reproduction (RARG) and immune responses (MRAP) have been identified.

There are different methods used to identify selection sweeps, some approaches use runs of homozygosity, the Bayesian approach uses Fst values, as well as HapFLK. HapFLK, quantifies the differences in frequencies of the haplotypes between the populations (Fariello *et al.* 2013). The hapFLK method allows for analysis between multiple populations and it considers the hierarchical structure of the populations involved (Fariello *et al.* 2013). Therefore, the aim of this study was to use HapFLK in an attempt to identify selection signatures in South African indigenous sheep breeds.

Materials & methods

SNP genotyping data from the 50K Ovine SNP beadchip was sourced for five indigenous fat-tailed sheep breeds, the Karakul (70), Damara (35), Pedi (30), Zulu (30) and Namaqua Afrikaner (48), one composite fat-tailed breed, the Meatmaster (20), a commercial meat breed, the Dorper (46) and one exotic sheep breed, Merino (98). Binary files (bed, bim and fam) generated in PLINK version 1.9 software (Purcell *et al.* 2007) were used to run an analysis based on the hapFLK method and using the Soay sheep breed as the outgroup. A python script was used to generate output files including Reynolds distance matrix,

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_747, © I.N. Malan *et al.* 2022

population tree, kinship matrix, allele frequencies and the FLK results. To visualize plots in the R program, negative log *P*-values were used to generate a Manhattan plot (Figure 1) with the aid of the ape, phangorn, ggplot2 and qqman packages. The suggestive significant line was set at $-\log_{10}(10^{-2})$. The hapFLK values were standardized and estimates were used to generate *P*-values from a normal distribution. Candidate genes (Table 1) within 1 Mb of the SNP of interest were identified by making use of *Ovis aries* gene assemblies on the National Center of Biotechnology Information (NCBI) platform.

Results

The Manhattan plot (Figure 1) depicts thirteen regions that have been under selection, spanning a total of 11 chromosomes. HapFLK *P*-values for individual SNPs were converted to -log values for better representation of the plot.

Discussion

The putative role of those genes listed in Table 1 in ovine metabolism needs consideration in cases where annotated. MSMO1 (chromosome 17) is linked to lipid metabolic processes, particularly cholesterol synthesis (Guan *et al.* 2020). The RAB32 gene located on chromosome 8 is involved in milk fat globule synthesis (Guan *et al.* 2020). Despite these candidate genes being identified in another livestock species (goat), the genetic marker might be in a similar position on analogous chromosomes in the ovine species. Milk production in sheep is important for rearing lambs and is positively associated with increased weaning weights. Most indigenous sheep breeds are kept for meat production. Two genes playing a role in meat production have been under selection, namely PDEZA on chromosome 15 and UTRN on chromosome 8 (Xu *et al.* 2021).

The gene TRL4 on chromosome 2 is related to genes involved in adaptive immune responses unique to sheep (Ha *et al.* 2015). Several genes that are involved in thermoregulation, ultraviolet exposure, skin development and ocular protection have also been identified in this study. The gene ASCL4 on chromosome 3 is responsible for skin barrier function and development, while KCNJ5 on chromosome 1 plays a role in water balance and salt retention (Chebii *et al.* 2020). This finding concurs with recent findings that indigenous sheep breeds (Pedi and Damara) were more water efficient than exotic breeds (Mupfiga, 2022). Several genes (FZD1, ACADSB, and SCNA11 on chromosomes 4, 22, and 19) were related to thermoregulation, temperature control and thermal stress (Ababaikeri *et al.* 2020; Librado *et al.* 2015). Maintaining traits linked to thermal stress, pigmentation is vital to the survival of sheep breeds in hot and arid environments. Heat stress from exposure could compromise production. Farming with animals that can cope with natural ambient temperatures could avert this. In conclusion, this study confirms natural selection for adaptive traits such as immunity, heat tolerance and meat production in indigenous sheep breeds.

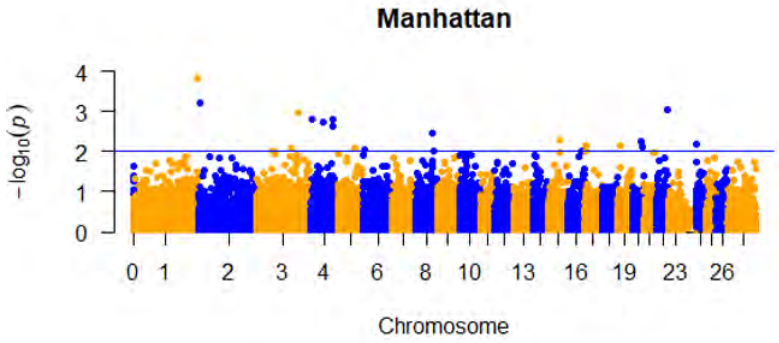


Figure 1. The hapFLK Manhattan Plot indicating regions under selection.

Table 1. HapFLK output highlighting SNP's that have putatively been under selection and their position within the genome.

Chr	Start	stop	Size (Mb)	Peak SNP	*P*-value	No of genes	Peak gene
1	291571232	294645906	3.07	OAR1_293056939.1	0.000145677	14	DYRK1A
2	4157609	7853921	3.70	s74535.1	0.00061206	6	TRL4
22	46001708	50501359	4.50	s11730.1	0.000911338	26	ACADSB
3	188042199	192951584	4.91	OAR3_190755374.1	0.001094956	23	MYH9, BPIFC
4	9140058	12572663	3.43	OAR4_10789078.1	0.001532869	16	FZD1
4	101955323	106631398	4.68	OAR4_104805602.1	0.001590593	20	PLXNA4
4	53982098	58240432	4.26	OAR4_56957796.1	0.001869332	20	CAV1
8	74979574	79687356	4.71	OAR8_77261220.1	0.003355182	18	RAB32, UTRN
15	53170403	55211450	2.04	OAR15_53158448.1	0.005305967	20	PDEZA
20	37217857	40782315	3.65	OAR20_38027997.1	0.005788378	7	E2F3
24	2630546	7031645	4.40	s54800.1	0.006270075	41	CREBBP
19	10181419	13907035	3.73	s40605.1	0.00705739	40	SCNA11
17	761527	3875182	3.11	s27585.1	0.007199631	13	MSM01

References

Ababaikeri, B., Abduriyim, S., Tohetahong, Y., Mamat, T., Ahmat, A. *et al.* (2020) Front. Zool. 17 (31): 1–15. https://doi.org/10.1186/s12983-020-00379-5.

Chebii, V. J., Oyola, S. O., Kotze, A., Entfellner, J. B. D., Musembi Mutuku, J. *et al.* (2020) Animals. 10 (11): 1–14. https://doi.org/10.3390/ani10112181.

Fariello M.I, Boitard S., Naya H., SanCristobal M., and Servin B. (2013) Genetics. 193 (3): 929–941. https://doi.org/10.1534/genetics.112.147231.

Guan, D., Landi, V., Luigi-Sierra, M. G., Delgado, J. V., Such, X. *et al.* (2020) J. Anim. Sci. Biotechnol. 11 (35):1–19. https://doi.org/10.1186/s40104-020-00435-4.

Ha, M., Sabherwal, M., Duncan, E., Stevens, S., Stockwell, P. *et al.* (2015) PLoS ONE 10 (10): 1-19. https://doi.org/10.1371/journal.pone.0139774.

Kijas, J.W., Townley, D., Dalrymple, B.P., Heaton, M.P., Maddox, J.F. *et al.* (2009) PLoS ONE 4: e4668. https://doi.org/10.1371/journal.pone.0004668

Kijas, J.W., Lenstra, J.A., Hayes, B., Boitard, S., Porto Neto, L.R. *et al.* (2012) PLoS Biol 10: e1001258.

Librado, P., Sarkissian, C. D., Ermini, L., Schubert, M., Jónsson, H. *et al.* (2015) PNAS, 112 (50): 6889–6897. https://doi.org/10.1073/pnas.1513696112.

Manzari, Z., Mehrabani-Yeganeh, H., Nejati-Javaremi, A., Moradi, M. H., and Gholizadeh, M. (2019) Anim. Genet. 50 (3): 298–302. https://doi.org/10.1111/age.12772.

Molotsi, A.H., Dube, B., Oosting, S., Marandure, T., Mapiye, C. *et al.* (2017) Sustain. 9: 1225. https://doi.org/10.3390/su9081225.

Molotsi, A. H., Cloete, S. W. P., Taylor, J. F., Whitacre, L. K., and Dzama, K. (2018) Proc of the 11th WCGALP, Auckland, New Zealand.

Mupfiga, S. (2022) MSc thesis. Stellenbosch University.

Purcell S., Neale B., Todd-Brown K., Thomas L., Ferreira M.A.R. *et al.* (2007) Am. J. Hum. Genet. 81(3): 559–575.

Rochus, C. M., Tortereau, F., Plisson-Petit, F., Restoux, G., Moreno-Romieux, C. *et al.* (2018) BMC Genomics 19 (71): 1–17. https://doi.org/10.1186/s12864-018-4447-x.

Xu, S.-S., Gao, L., Shen, M., and Lyu, F. (2021) Front. Genet. 12: 1–7. https://doi.org/10.3389/fgene.2021.738879.

748. An across-flock genetic analysis of components of reproduction in four South African Merino resource flocks

C.L. Nel[1,2*], A.A. Swan[3], K. Dzama[2], W.J. Olivier[4], M.A. Snyman[4] and S.W.P. Cloete[1,2]

[1]Directorate: Animal Sciences, Western Cape Department of Agriculture, Muldersvlei Road, Elsenburg, 7607, South Africa; [2]Department of Animal Sciences, Stellenbosch University Merriman Avenue, Matieland, 7602, South Africa; [3]AGBU, a joint venture of NSW Department of Primary Industries and University of New England, Armidale, NSW, 2351, Australia; [4]Grootfontein Agricultural Development Institute, Middelburg, 5900, South Africa; neliusn@elsenburg.com

Abstract

It is well accepted that reproduction is a key profit driver in small stock production systems. The traditional approach for targeting reproductive output in many countries has been through composite trait selection, such as number of lambs weaned per ewe mated (NLW). However, there could be benefit in considering the components of NLW individually. This study reports across-flock analyses of NLW and its components, conception rate, litter size and ewe rearing ability in South African Merino resource flocks. Direct genetic effects were small, but all traits were heritable, estimates ranging between 0.02 for ewe rearing ability to 0.06 for litter size. The only noteworthy unfavourable genetic correlation was between litter size and ewe rearing ability. The traits reported here should be suitable for incorporation into commercial breeding programs of the wider industry.

Introduction

Selection for reproduction both in Australia and South Africa has in the past been based on composite traits such as the number of lambs born or weaned (NLW; Brien *et al.*, 2014). Bunter and Brown (2015) proposed using the components of NLW, such as conception rate, litter size, and rearing ability individually, since sole emphasis on composite traits could risk unpredictable outcomes as correlated responses (Bunter *et al.*, 2020). Targeting selected components could deliver more flexible selection objectives (Swan and Bunter, 2021). For example, the high requirements of ewes bearing multiples could give reason to emphasize conception rate, but not litter size in challenging production environments. In Australia, a pipeline for the evaluation of such component traits has since been developed (Bunter *et al.*, 2020) and incorporated into the national genetic evaluation of Sheep Genetics (Bunter *et al.*, 2019, 2020).

The prospects of similar developments should be evaluated for South African Merinos, but the present data structure in recorded flocks poorly supports the decomposition of NLW. A lack of mating lists restricts NLW to a 'per ewe lambed' basis and thus lacks conception rate. The sporadic recording of dead lambs in pedigrees also leads to little variation between lambs born and weaned, and thus does not appropriately reflect ewe rearing ability. However, South Africa hosts multiple Merino resource flocks maintained under well-developed recording programs by science and technology institutions (Schoeman *et al.*, 2010). Data from these flocks could serve as an example to evaluate these component traits, but studies on these flocks have thus far been restricted to within-flock analyses of only selected traits (Cloete *et al.*, 2017). The aim of this study was thus to conduct a genetic evaluation of the components of NLW from an across-flock analysis extended to four SA resource flocks.

Materials & methods

The records used in this study originated from the Elsenburg (n=6,807), Grootfontein (n=7,825), Cradock (n=8,940) and Langgewens (n=1,230) Merino resource flocks maintained across the Western and Eastern Cape provinces of South Africa (Schoeman *et al.*, 2010). Older records and pedigrees were truncated to

1986 to align with the start of the Elsenburg selection experiment. Reproduction traits were recorded as (n=24,802) repeated records on 7,173 naturally joined ewes (3.45 records per dam). This included the composite trait NLW which was decomposed into the underlying component traits, namely: conception rate (CON: 0 = did not conceive, 1 = conceived), litter size (1 to N lambs born) and ewe rearing ability (NLW/number of lambs born), where the latter two traits were censored in the case where CON = 0.

The age of the ewe (2 to 7+ years) was tested for significance ($P<0.05$) as a fixed effect, while contemporary groups in which the ewe was weaned as a lamb and also in which specific ewes lambed were fit as sparse fixed. The contemporary groups were generally defined by concatenating flock × year × season, but special cases like the selection line groups of the Elsenburg flock (Cloete *et al.*, 2004; 2017) and the Langgewens ewes mated to non-Merino sires was also accounted for by extended contemporary group definitions. Random effects considered were the direct genetic (σ^2_a) and ewe permanent environmental (σ^2_{ewe}) effects. Criteria for the retention of random effects were based on significant ($P<0.05$) differences according to the log-likelihood ratio test. All analyses commenced in ASREML V4.2 (Gilmour *et al.*, 2015). A single-trait repeatability model was used to derive the variance components of NLW, and pairwise two-trait models were used to estimate the relevant correlations of NLW with CON, LS and ERA, respectively. A three-trait analysis was used to derive the parameters and correlations among CON, LS and ERA.

Results

Ewe age was significant ($P<0.01$) in all traits considered in this study, with the best performance in ewes of intermediate (3-4 years) age. The coefficients of variation showed that component traits were less variable than NLW, but still exhibited useful variation to facilitate selection (Table 1). The addition of the σ^2_a and σ^2_{ewe} variance components to the operational model that only included systematic effects was associated with significant ($P<0.05$) increases in log-likelihood values for all traits. However, the h^2 of reproduction traits was generally low to very low (Table 1), ranging from 0.02 for ERA to 0.06 for LS. With the exception of LS, the ewe permanent environmental ratio (pe^2) was generally of larger magnitude that the corresponding h^2. Genetic correlations of NLW with the component traits were moderate to high at 0.59 to 0.70 (Table 2). This was expected given the part-whole nature underlying these relationships. Correlations among the component traits were either low to moderate or estimated with a low accuracy. Of some importance was the negative and unfavourable relationship of LS with ERA, which was negative on both the genetic and phenotypic levels.

Table 1. Summary statistics as well as variance components and ratios for number of lambs weaned (NLW), conception (CON), litter size (LS) and ewe rearing ability (ERA).

	Trait			
	NLW	**CON**	**LS**	**ERA**
Mean	0.99	0.82	1.57	0.79
S.D.	0.76	0.39	0.57	0.35
CV	77.2%	47.2%	36.3%	44.8%
Variance components				
σ^2_a	0.016±0.004	0.004±0.001	0.018±0.003	0.002±0.001
σ^2_{pe}	0.036±0.004	0.011±0.001	0.011±0.003	0.009±0.001
σ^2_e	0.463±0.005	0.11±0.001	0.258±0.003	0.11±0.001
σ^2_p	0.515±0.005	0.126±0.001	0.287±0.003	0.12±0.001
Variance ratios				
h^2	0.03±0.007	0.03±0.008	0.06±0.009	0.02±0.007
pe^2	0.07±0.008	0.09±0.009	0.04±0.009	0.07±0.009

Table 2. Genetic (above diagonal) and phenotypic (below) correlations among number of lambs weaned (NLW), conception (CON), litter size (LS) and ewe rearing ability (ERA).

	NLW	CON	LS	ERA
NLW		0.70±0.08	0.66±0.08	0.59±0.12
CON	0.59±0.004		0.16±0.13	0.25±0.22
LS	0.55±0.01	-0.05±0.01		-0.37±0.17
ERA	0.76±0.003	0.07±0.01	-0.15±0.01	

Discussion

All traits in this study were heritable and could be targeted by genetic selection, but pe^2 effects were generally of a considerably larger magnitude, with the exception of that observed for LS. Generally, similar pe^2 estimates were reported for the literature for the component traits CON (0.01-0.14) and LS (0.02-0.06) (Bunter and Brown, 2015; Bunter *et al.*, 2016, 2020; Cloete *et al.*, 2017; Walkom and Brown, 2017; Bunter and Swan, 2021) as well as ERA (0.05) (Bunter *et al.*, 2016). The low h^2 for the composite and component reproduction traits also agreed with what is generally expected from the literature for NLW (0.03-0.10) (Cloete *et al.*, 2004, 2017; Bunter and Brown, 2015; Bunter *et al.* 2016; Walkom and Brown, 2017) as well as CON (0.02-0.09) (Bunter and Brown, 2015; Walkom and Brown, 2017; Bunter *et al.*, 2020; Bunter and Swan, 2021). The h^2 of LS was slightly higher, but on the lower end of the range reported by previous studies (0.04-0.13) (Bunter and Brown, 2015; Bunter *et al.*, 2016, 2020; Cloete *et al.*, 2017; Walkom and Brown, 2017; Bunter and Swan, 2021), whereas a very low h^2 for ERA was well supported by the literature (0.01-0.08) (Bunter *et al.*, 2016, 2020; Walkom and Brown, 2017; Bunter and Swan, 2021). With a similar h^2 for NLW compared to the component traits, there should be little advantage in the use of the composite trait. In turn, targeting component traits could be beneficial, as they will more closely reflect the aetiology underlying specific reproduction outcomes. Overall, the correlations reported in Table 2 did not suggest strong or unfavourable responses from emphasis on component traits, and the low genetic correlations between CON and LS supported the premise that these traits could be targeted individually. The negative relationship between LS and ERA was expected, given the well-known higher susceptibility of multiples to adverse conditions. However, if also considering the higher h^2 of LS, it remains important to note that LS should not be targeted without additionally considering ERA (Bunter *et al.*, 2020), or lamb survival (Nel *et al.*, 2021), which is not expressed by NLW explicitly.

In conclusion, a higher resolution focus on specific characteristics will be limited by what information can be practically gathered on a large scale, but the traits reported here were all derived under relatively standard recording procedures. Thus, it should be feasible for commercial programs to accommodate the protocols followed in the resource flocks.

References

Brien, F. D., Cloete, S. W. P., Fogarty, N. M., Greeff, J. C., Hebart, M. L., *et al.* (2014) Anim. Prod. Sci. 54(6):667–693. https://doi.org/10.1071/AN13140.

Bunter, K. L., and Brown, D. J. (2015) Proc. of the 21st Association for the Advancement in Animal Breeding and Genetics meeting, Victoria, Australia.

Bunter, K. L., and Swan, A. A. (2021) Proc. of the 24th Association for the Advancement in Animal Breeding and Genetics meeting, Adelaide, Australia.

Bunter, K. L., Swan, A. A., Gurman, P. M., Boerner, V., McMillan, A. J., *et al.* (2019) Proc. of the 23rd Association for the Advancement of Animal Breeding Genetics meeting, Armidale, Australia.

Bunter, K. L., Swan, A. A., Gurman, P. M., and Brown, D. J. (2020) Anim. Prod. Sci. 61(3): 333-336. https://doi.org/10.1071/AN20444.

Bunter, K. L., Swan, A. A., Purvis, I. W., and Brown, D. (2016) Anim. Prod. Sci. 56(4):679–689. https://doi.org/10.1071/AN14630.

Cloete, S. W. P., Gilmour, A. R., Olivier, J. J., and Van Wyk, J. B. (2004) Anim. Prod. Sci. 44(8):745–754. https://doi.org/10.1071/EA02225.

Cloete, S. W. P., Van Wyk, J. B., and Olivier, J. J. (2017) Proc. of the 22nd Association for the Advancement in Animal Breeding and Genetics meeting, Townsville, Australia

Gilmour, A. R., Gogel, B. J., and Welham, S. J. 2015. ASReml User Guide Release 4.1 Structural Specification.

Nel, C. L., Swan, A. A., Dzama, K., Scholtz, A. J., and Cloete, S. W. P. (2021) Anim. Prod. Sci. 61(18):1965–1981. https://doi.org/10.1071/AN21198.

Schoeman, S. J., Cloete, S. W. P., and Olivier, J. J. (2010) Livest. Sci. 130(1-3):70–82. https://doi.org/10.1016/j.livsci.2010.02.012.

Swan, A. A., and Bunter, K. L. (2021) Proc. of the 24th Association for the Advancement of Animal Breeding and Genetics meeting, Adelaide, Australia.

Walkom, S. F., and Brown, D. J. (2017) Anim. Prod. Sci. 57(9):1952–1961. https://doi.org/10.1071/AN15574.

Part 41

Companion animals

749. Diversity status and future breeding strategy for Tatra hound in Slovakia

R. Kasarda* and N. Moravčíková

Slovak University of Agriculture in Nitra, Tr. A. Hlinku 2, 949 76 Nitra, Slovak Republic;
radovan.kasarda@uniag.sk

Abstract

The aim of this study was to analyse the level of genetic diversity and inbreeding based on DNA using Canine SNP Bead Chip. The total number of genotyped individuals was 52. Relative low average minor allele frequency (MAF=0.25) indicates an increase of homozygotes proportion. The level of genomic inbreeding derived from homozygous excess ranged from -0.12 to 0.23 (F_{HOM}=-0.0113). Observed results showed that the inbreeding in the current generation is at the level 1.89%. Due to closed herdbook and higher probability of co-selection of relatives the future increase of the inbreeding in the Tatra hound population could be expected. Tatra Hound Breeders Club in Slovakia intends to apply family and line-breeding to manage increase in inbreeding per generationon sustainable level (bellow 1%) to prevent negative effect of inbreeding depression.

Introduction

Significant changes can be observed in Slovakia in the ownership and the structure of the farmland. Due to the current development, natural areas become more fragmented (building and construction, highway network) with low acreage, which copes with hunting areas. Due to the changes in the structure of hunting areas, there is pressure on breeds structure and dog performance.

In 2009, a group of dog breeders defined the breeding goal of Tatra Hound (TH). TH is a composite breed based on autochthonous non-pedigreed females, local Slovakian Hound (FCI standard No. 244), Bavarian Mountain Hound (FCI 217) and Standard Dachshund (FCI 148). The breeding program was established to create a new breed that will fulfil the present needs of hunting practice in Slovakia. At present, more than 1,400 individuals are registered in the herd book with fully traced pedigrees minimum up to the 4th generation of ancestors. As the number of founders is limited and the herd book closed, the mating program currently plays an important role.

In general, dogs are recognised as one of the most inbreed domestic animals, which is coming from the fact that a limited number of ancestors (often derived from common founders) and mating of close relatives to fix specific traits are relatively common (Sams and Boyko, 2019). Such mating results in the formation of long homozygous segments across DNA known as runs of homozygosity (ROH) (Gibson *et al.* 2006). Not surprisingly, ROH segments were identified even in grey wolves (Randi 2011, Gómez-Sanchez *et al.* 2018).Identification of ROH segments is essential for diversity analyses not only in the farm but also in companion animals as prevention of negative effect of inbreeding depression in small local populations under strong selection pressure (Kasarda *et al.* 2016, Kukučková *et al.* 2017, Moravčíková *et al.* 2019). Therefore, the aim of this study was to analyse the level of genetic diversity and genomic inbreeding in a small local population of the Tatra Hound breed.

Materials & methods

The total number of genotyped individuals was 52. Canine – Ilumina 230k HD Chip containing 220,853 SNPs was used for genotyping (211,256 autosomal SNPs with known position). Standard manipulation and

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_749, © R. Kasarda and N. Moravčíková 2022

pruning of SNP data have been made in PLINK 1.9 (Chang *et al.* 2015). After quality control (--geno 0.1, --mind 0.1 and --maf 0.01), 162,776 SNP and 52 individuals were left for the analysis.

Diversity parameters. PLINK 1.9 (Chang *et al.* 2015) was used to calculate minor allele frequency (MAF), observed (H_o) and expected heterozygosity (H_e) and genomic inbreeding derived from the genome-wide homozygous excess (F_{HOM}). Package *detectRUNS* (Biscarini *et al.* 2015) was used to calculate genomic inbreeding derived from ROH segments distribution in the genome (F_{ROH}) under software R (R Core Team, 2021).

Screening of ROH of different lengths (0-2, 2-4, 4-8, 8-16, and >16 Mbp) was used to test the effect of the mating program in Tatra hound on the distinction between genetic information inherited from different generations of ancestors. The ROH were defined as genomic segments with a minimum 500 kb (Sams and Boyko, 2019) containing at least 46 consecutive homozygous SNP genotypes no more than 1000 kb apart. Missing or heterozygous genotypes in ROH were not allowed. In addition, each ROH had to have a minimum of one SNP per 5000 kb. Lencz *et al.* (2007) formula was used to calculate the minimum number of SNPs in the ROH. F_{ROH} was calculated by dividing the sum of the ROH length (kb) in the length class by the total length of the autosomal genome (kb) covered by the SNPs.

Genetic structure of the population. The population stratification degree was quantified by calculating Nei's genetic distances on an individual level using the R package *StAMPP* (Pembleton *et al.*, 2013) and visualised as a heatmap in program R (R Core Team, 2021).

Results

The average minor allele frequency (MAF=0.248) was low, indicating an increase of proportion of homozygotes. Similarly, the average observed H_o=0.336 and expected heterozygosity H_e=0.332 indicated a higher proportion of homozygous genotypes in the gene pool of the Tatra Hound population. The average level of genomic inbreeding F_{HOM} was -0.011, ranging from -0.118 to 0.229. The results of F_{ROH} calculated across five different length classes are summarised in Table 1. The higher the length of ROH segments, the lower F_{ROH} observed, i.e. higher historical than recent inbreeding in population. Observed results showed that the inbreeding in the current generation is at the level 1.89%. The population stratification degree is visualised as a heatmap in Figure 1. As indicated in Figure 1. the population is clustered into two main clusters (red circles) according to the breeding strategy and the fact that there are two basic coat colours (black and red) in the Tatra hound inherited from source breeds: Slovakian Hound and Standard Dachshund (black) and Bavarian Mountain Hound (red). Further clustering indicates basic family and line structure in the population.

Table 1. Summary statistics for ROH per class.

ROH class (Mbp)	No. of ROH	Distribution of ROH (%)	Mean length ROH	F_{ROH} (%)
0-2	3,503	0.642	1.064	12.09
2-4	1,027	0.188	2.808	8.84
4-8	611	0.112	5.558	6.45
8-16	255	0.047	10.519	3.97
>16	58	0.011	20.067	1.89

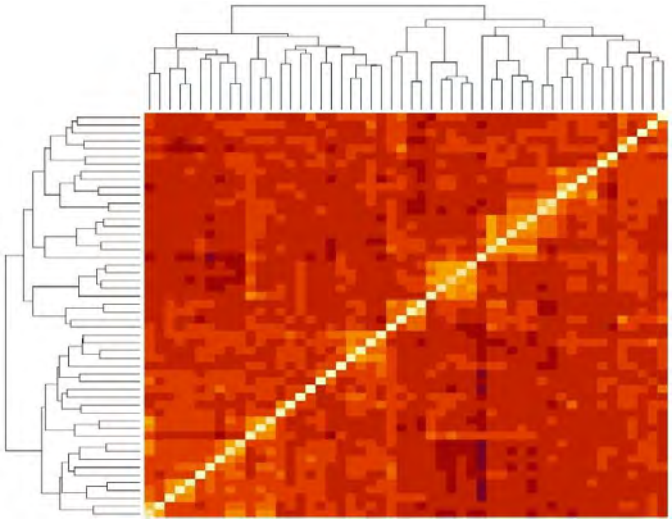


Figure 1. The Tatra hound population stratification degree (dark red = un-related; light yellow=related).

Discussion

Most dog breeds are kept for pleasure, show, sport, and hobby. However, a special place belongs to hunting dog breeding. Keeping dogs for hunting purposes is traced back to the early beginning of domestication. Today, humans better understand genetics and animal breeding and are developing new strategies using biotechnology and progressing in genetic engineering. As well documented in farm animals, progress in biotechnology and genetic engineering could speed up the selection progress. However, especially in the case of local populations, intensive selection can be accompanied by a risk of diversity damage (Curik *et al.* 2017). Therefore, this study intended to view Tatra hound population status and stratification to set up future breeding strategy better.

According to the calculated H_o and H_e the population is in equilibrium (P=0.95, χ^2-test). The negative values of F_{HOM} reflect the fact that F_{HOM} cannot distinguish IBD and IBS alleles, therefore overestimating inbreeding in population (Shi *et al.* 2020). Negative F_{HOM} values indicate less homozygous individuals in population than expected, i.e. observed heterozygosity is larger than expected heterozygosity.

The longer the ROH segments, the closer generation to the common ancestor, i.e. short ROH segments indicate the level of historical inbreeding (founders). Long segments reflect the state of inbreeding in current generations. In Tatra hound, historical and present inbreeding is generally lower than in well-established breeds with a long history like German shepherd, Belgian Shepherd or Czechoslovakian Wolfdog (Moravčíková *et al.* 2021), however higher compared to the wild ancestor of all dogs, grey wolf. Due to closed herdbook and higher probability of co-selection of relatives the future increase of the inbreeding in the Tatra hound population could be expected.

Tatra Hound Breeders Club in Slovakia intends to apply family and line-breeding to manage increase in inbreeding per generationon sustainable level (bellow 1%) to prevent negative effect of inbreeding depression. Furthermore, the Tatra hound breeders club is following regulations and recommendations of FCI to breed high performance, healthy dogs according to the highest standards and current scientific knowledge.

Acknowledgements

This publication was supported by the Slovak Research and Development Agency (Project No. APVV-20-0161 and No. APVV-17-0060).

References

Biscarini F., Nicolazzi E.L., Stella A., Boettcher P.J. and Gandini G. (2015) Front Genet 6:33. https://doi.org/10.3389/fgene.2015.00033

Curik I., Ferenčaković M. and Sölkner J. (2017) R. Bras. Zootec. 46(9).https://doi.org/10.1590/S1806-92902017000900010

Gibson J., Morton, N. and Collins, A. (2006) Hum. Mol. Genet. 15:789-795. https://doi.org/10.1093/hmg/ddi493

Gómez-Sánchez D., Olalde I., Sastre N., Enseñat C., Carrasco R. *et al.* (2018) Mol. Ecol. 27:3599-3612. https://doi.org/10.1111/mec.14824

Chang C.C., Chow C.C., Tellier L.C.A.M, Vattikuti S., Purcell S.M. *et al.* (2015) GigaScience 4:7. https://doi.org/10.1186/s13742-015-0047-8

Kasarda R., Vostrý L., Moravčíková N., Vostrá-Vydrová H., Dovč P., *et al.* (2016) Acta Agric. Scand. A Anim. Sci. 66:67-74. https://doi.org/10.1080/09064702.2016.1249400

Kukučková V., Moravčíková N., Ferenčakovič M., Simčič M., Mészáros G. *et al.* (2017) Conserv. Genet. 18:893-910. https://doi.org/10.1007/s10592-017-0935-9

Lencz T., Lambert C., DeRosse P., Burdick K.E., Morgan T.V. *et al.* (2007) Proc. Natl. Acad. Sci. USA 104:19942-19947. https://doi.org/10.1073/pnas.0710021104

Moravčíková N., Kasarda R., Vostrý L., Krupová Z., Krupa, E. *et al.* (2019) Czech J. Anim. Sci. 64:491-503. https://doi.org/10.17221/226/2019-CJAS

Moravčíková N., Kasarda R., Židek R., Vostrý L., Vostrá-Vydrová H. *et al.* (2021) Genes 12:832. https://doi.org/10.3390/genes12060832

Pembleton L.W., Cogan N.O. and Forster J.W. (2013) Mol. Ecol. Res. 13:946-952. https://doi.org/10.1111/1755-0998.12129

R Core Team (2021) R: A Language and Environment for Statistical Computing; R Foundation for Statistical Computing: Vienna, Austria. Available online: http://www.R-project.org/

Randi E. (2014) Mammal. Rev. 41:99-111. https://doi.org/10.1111/j.1365-2907.2010.00176.x

Sams A.J. and Boyko A.R. (2019) G3 Genes Genomes Genet. 9:117-123. https://doi.org/10.1534/g3.118.200836

Shi L.Y., Wang L.G., Liu J.X., Deng T.Y., Yan H. *et al.* (2020) J Anim Sci Biotechnol. 11:46. https://doi.org/10.1186/s40104-020-00447-0

750. Genome-wide association study for ectopic ureter in three dog breeds

A. Stella[1*], V. Cassani[2], G. Majolino[3], P. Cozzi[1], S. Biffani[1], F. Biscarini[1] and G. Minozzi[4]

[1]Institute of Agricultural Biology and Biotechnology, Italian National Research Council – CNR, Via Edoardo Bassini 15, 20133 Milan, Italy; [2]Independent DVM, via Manara 7 Castano Primo, Italy; [3]Independent DVM, Vicolo del Forno 5, 43044 Collecchio PR, Italy; [4]Department of Veterinary Medicine DIMEVET, University of Milan, Via dell'università, 26900 Lodi, Italy; alessandra.stella@cnr.it

Abstract

The objective of the study was the identification of genomic regions associated with the occurrence of canine ectopic ureter, a congenital abnormality of the urinary system that requires surgical intervention. The study was applied to three breeds: Labrador Retriever, Golden Retriever and Newfoundland. A case: control design was used, comprising 263 animals. Numbers of cases/controls for the three breeds were Labrador Retriever (48/50), Golden Retriever (60/60) and Newfoundland (15/20). After quality control, 137,244 markers were used in the analysis. Data were analysed with a mixed model that accounted for breed and sex. After applying a Bonferroni correction, 4, 9 and 5 significant markers were observed in three regions on canine chromosomes CFA 5, 27 and 30 respectively. A preliminary GO analysis showed enrichment in processes related to regulation of signalling and system development, among others.

Introduction

Ectopic ureter (EU) is a congenital defect of the urinary system in which the ureter inserts anywhere other than the vesical trigone (Osborne *et al.*, 1995). This anatomical change can be unilateral or bilateral. The occurrence of ectopic ureter in the dog has been estimated to be 0.016% (Dean *et al.*, 1988), however, veterinarians indicate an increasing and prevalent incidence in some popular breeds. Dogs with ectopic ureters usually present urinary incontinence, male and females may present symptoms at different ages, due to difference in anatomy, and different breeds seems to exhibit a different female to male cases ratio (Hayes, 1984). It is thought to occur because of disruption in normal embryogenesis and is commonly associated with other abnormalities of the urogenital tract. Gallana *et al.* (2018) estimated the heritability of the phenotype EU (classified in three classes based on anatomical diagnosis and severity) to be very high (0.71), which may indicate involvement of one or a few major genes. The same authors applied genome-wide association analysis (GWAS) to a subsample and identified 6 loci on canine chromosomes 3, 17, 27 and 30 with 16 SNPs significantly associated with the defect. The study was performed using data from the Entlebucher breed, a small population with high inbreeding rate. This may have affected the results. The aim of our study is to investigate genome-wide association for EU in three breeds that have recorded a growing number of cases in the last decade: Labrador retriever, Golden Retriever and Newfoundland.

Materials & methods

Samples of EU cases were collected from 2017 through 2021 by veterinarians, with the informed consent of the owners (breeders, in some cases). Affected individuals were: 48 Labrador Retrievers (LAB), 60 Golden Retriever (GOL) and 15 Newfoundlands (NEW). Pedigree information was not available. Cases were matched with data from healthy unrelated individuals for each breed: 50 LAB, 60 GOL and 20 NEW yielding a total dataset of 263 individuals. Blood samples were collected in EDTA vacutainers (BD Vacutainer Systems, Plymouth, UK), refrigerated at 4 °C and then stored at -20 °C. DNA was extracted from whole blood samples using the NucleoSpinVRBlood kit (Macherey-Nagel, Duren, Germany). The Illumina CanineHD bead chip was used for genotyping the samples.

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_750, © A. Stella *et al.* 2022

Data underwent quality control using Plink v1.9 (Chang *et al*., 2015). The final dataset was pruned for low minor allele frequency (0.05), genotyping efficiency (SNP with more than 5% missing) and failure to meet Hardy-Weinberg equilibrium (1×10^{-6}). The CanFam3.1 assembly served as the reference genome. SNP with more than 5% of data missing, or with MAF≤1% were removed. Genotyping efficiency for samples was also verified, and samples with more than 5% missing data were removed. After quality control, 137,244 markers were included in the GWAS analysis. Only autosomes were included in the GWAS. The software GEMMA v0.98 (Zhou *et al.*, 2012) was used to perform association analysis using a mixed model approach with breed and sex included in the model. Bonferroni correction was applied to estimate 0.05 significance level. ClueGO (Bindea *et al.*, 2009) was used for biological interpretation of genes underlying regions of interest.

Results

A total of 4, 9 and 5 significant markers were observed in three regions on canine chromosomes CFA 5, 27 and 30 respectively (Figure 1). Interestingly, the most significant SNP identified in our study located on chromosome 27 (position 23,004,532) overlaps closely with the region identified by Gallana *et al.* (2018) on chromosome 27 defined by 4 highly significant SNPs in positions 22,783,765 to position 23,114,194. Among the genes underlying the significant regions on chromosome 5, PDE4B and FGGY are associated with congenital organ defects, although mainly cardiac.

A preliminary GO analysis of a subset of genes (n=30) underlying the regions of interest on chromosome 5, 27 and 30 showed enrichment in biological processes related to regulation of signalling, negative regulation of biological processes, system development and regulation of cell communication.

Discussion

In the only previous case-control genome wide association study conducted on canine ectopic ureter Gallana *et al.* (2018) identified 16 significantly associated SNPs. Of these, 13 SNPs fall in the same chromosomes identified in our study, namely chromosomes 27 and 30. In detail 11 SNPs are located on chromosome 27 and 2 on chromosome 30.

Despite indicating the HOX gene cluster and the SOX5 gene as potential positional candidates on chromosome 27 and *SMAD6* and *SMAD3* on chromosome 30 Gallana *et al.* (2018) indicate that none of the genes are strong functional candidates.

However, recent gene expression and protein levels studies conducted in women affected by urinary incontinence and pelvic prolapse identified that SMAD3 gene expression levels were significantly lower in affected women compared to the control patients (P=0.008) (Akin *et al.* 2021). The reduction in the

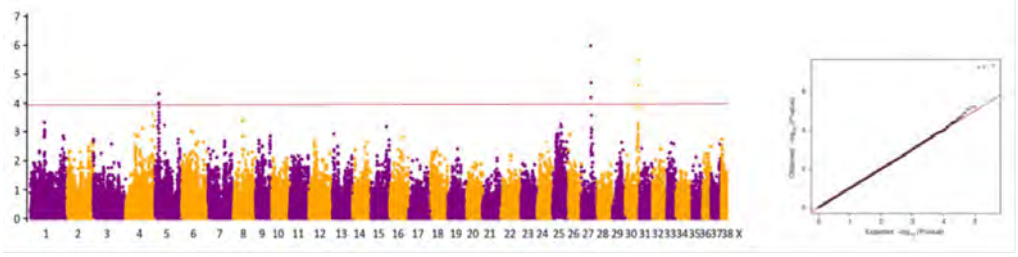


Figure 1. Manhattan plot (a) showing -log10 of the raw *P*-values for each genotyped SNP by chromosome (*x*-axis). The horizontal line denotes genome-wide significance based on Bonferroni-corrected *P*-values. (b) Q–Q plot of samples used in GWAS showing the -log10 of the expected versus the observed *P*-values.

expression levels of SMAD3 may indicate a functional importance and possible variability in its close regulatory regions should be further explored.

References

M.N Akin, A.A. Sivaslioglu, T. Edgunlu, Kasap B., Celik S.K. et al (2021). Mol Biol Rep. 2021;48(2):1401-1407. https://doi.org/10.1007/s11033-021-06220-4.

G. Bindea, Mlecnik B., Hackl H., Charoentong P., Tosolini M., *et al.* (2009) Bioinformatics. 15;25(8):1091-3. https://doi.org/10.1093/bioinformatics/btp101.

Chang C.C., Chow C.C., Tellier L.C., Vattikuti S., Purcell S.M., *et al.* (2015). Gigascience. 2015 25;4:7. https://doi.org/10.1186/s13742-015-0047-8.

Dean P. W., Bojrab M. J., Constantinescu G. M. (1988) Contin. Educ. Vet., 10, 146–162.

Gallana M., Utsunomiya Y.T,, Dolf G., Pintor Torrecilha R.B., Falbo A.K., *et al.,* (2018). Anim Genet. 49(6):645-650. https://doi.org/10.1111/age.12728.

Hayes H. M. Jr. (1984) J. Small Anim. Pract., 25, 501–504.

Osborne C. A., Johnston G. R., Kruger J. M. (1995) Ectopic ureters and ureteroceles. In: Osborne C. A., Finco D. R. (eds.) Canine and feline nephrology and urology. 1st ed. Williams & Wilkins, Philadelphia, PA. pp 608–622.

Zhou X., Stephens M. (2012) Nat Genet. 17;44(7):821-4. https://doi.org/10.1038/ng.2310.

751. Linkage disequilibrium variability in the Czechoslovakian wolfdog genome as a result of long-term artificial selection

N. Moravčíková[1*], R. Kasarda[1], L. Vostrý[2], H. Vostrá-Vydrová[2], J. Vašek[2] and D. Čílová[2]

[1]Slovak University of Agriculture in Nitra, Tr. A. Hlinku 2, 94976 Nitra, Slovakia; [2]Czech University of Life Sciences Prague, Kamýcká 129, 16500 Praha – Suchdol, Czech Republic; nina.moravcikova@uniag.sk

Abstract

The genetic makeup of any dog breed is significantly affected by mating close relatives carrying preferred phenotypic traits by breeders. Such mating simultaneously leads to changes in the linkage disequilibrium, mainly in genomic regions controlling preferred phenotypes. Following this assumption, the study aimed to analyse the variability in linkage disequilibrium (LD) in the genome of the Czechoslovakian wolfdog compared to its wild ancestor grey wolf and identify genomic regions significantly subject to artificial selection. The database included genomic information of 60 animals genotyped using the CanineHD chip. The variability in linkage disequilibrium was tested by comparing the LD level in target regions in the Czechoslovakian wolfdog and grey wolf genomes and quantified by calculating varLD score. The strongest signal of artificial selection was found on autosome 17. Inside genomic regions under selection pressure were located 59 protein-coding genes with diverse functions in the organism.

Introduction

When considering the genetic makeup of any dog breed, repeated close mating of relatives was used to fix breed-specific phenotypic traits preferred by breeders. Such a mating usually resulted in various alterations in the genome, including changes in allele frequencies (Cagan and Blass, 2016), the formation of long runs of homozygosity (Caniglia *et al.*, 2018) or changes in linkage disequilibrium (Gray *et al.*, 2009).

The Czechoslovakian wolfdog is a large dog reminiscent of a wolf. Because the Czechoslovakian wolfdog was created by crossing Carpathian grey wolf and German Shepherd dog, different patterns are manifested mainly in its behaviour than in other breeds of dogs. This breed retains the anatomical advantages of the wolf, thanks to which it can run steadily, adapt to climatic conditions and orient itself well in the field. Animals belonging to this breed are temperamental and active dogs with extremely developed senses and orientation in the field. In the current breeding program, great emphasis is placed on animals trainability. The genomic diversity of Czechoslovakian Wolfdog was analysed in several previous studies (e.g. Smetanová *et al.*, 2015; Canaglia *et al.*, 2018; Moravčíková *et al.*, 2021).

The aim of this study is to compare the level of linkage disequilibrium in genomes of Czechoslovakian wolfdog and grey wolf to find genomic regions that were subject to artificial selection and identify protein-coding genes located inside genomic regions under strong selection pressure.

Materials & methods

The genomic information of Czechoslovakian wolfdog (30 animals) and grey wolf (30 animals) was used to analyse the effect of artificial selection on the genome-wide level of linkage disequilibrium (LD) to identify those genomic regions that have been most frequently subjected to selection in the Czechoslovakian wolfdog genome. In the case of the Czechoslovakian wolfdog, nuclear DNA was genotyped by the CanineHD 230k chip in a commercial lab. The genomic information of grey wolves genotyped by the CanineHD 130k chip was retrieved from a web repository (Shannon *et al.*, 2016). The standard quality control of genotyping data was performed by Plink v1.9 (Chang *et al.*, 2015) to filter out non-autosomal SNPs, loci with unknown chromosomal positions, all animals and SNPs with a call rate lower than 90%. In addition, the minimum

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_751, © N. Moravčíková *et al.* 2022

minor allele frequency was set up to 1%. After SNP pruning, 136360 SNP markers remained and all 60 animals.

Differences in the LD level in the autosomal genome of the Czechoslovakian wolfdog and grey wolf were quantified using the VarLD program (Ong and Teo, 2010). This program compares LD in target regions in the genome between two groups and is based on the assumption that selection of specific genetic variants in the target region of one population changes the LD level in that region at the same time, while in another population not affected by the same selection pressure the level of LD remains unchanged. LD values (expressed by r_{LD}^2) calculated for windows containing 50 SNPs within each autosome were transformed into standardised varLD scores (S_i^*), based on which selection signals were identified. Genomic regions significantly affected by artificial selection were recognised as regions containing 99.9% of signals. The distribution of varLD standardised scores across the autosomal genome was visualised by R package qqman (Turner, 2018). Protein-coding genes located inside selection signals were detected by R package ensembldb (Rainer *et al.*, 2019).

Results

The analysis of LD level showed a total of 19 regions under strong selection pressure in the Czechoslovakian wolfdog genome spread across seven autosomes (Table 1). The largest region covering 6.08 Mbp (from 30,064,957 to 36,146,893 bp) was found on autosome 17, which at the same time showed the strongest signal of selection (Figure 1). On the other hand, the shortest region was identified on autosome 1 (from 24,486,423 to 24,717,376 bp). Overall, 125 protein-coding genes with diverse biological functions were identified within the detected selection signals.

Table 1. Genomic regions showing significant differences in LD level.

CFA	Start (Mbp)	End (Mbp)	Protein-coding genes
1	24.49	24.72	MC5R, FAM210A, LDLRAD4
9	39.91	40.62	ASIC2, SPACA3, TMEM98, MYO1D, CDK5R1, PSMD11, C9H17orf75
17	2.50	5.35	RSAD2, RNF144A, ROCK2, E2F6, GREB1, NTSR2, LPIN1, TRIB2, RMDN2, CYP1B1, ATL2, HNRNPLL, GALM, SRSF7, GEMIN6, DHX57, MORN2, ARHGEF33, SOS1, CDKL4, MAP4K3, TMEM178A, THUMPD2, SLC8A1, PKDCC, EML4, COX7A2L, KCNG3, MTA3, ARID5A, NEURL3, NCAPH, SNRNP200, CIAO1, TMEM127, STARD7, DUSP2, ASTL, ADRA2B, KCNIP3, GPAT2, PROM2, ZNF2, ZNF514, MRPS5, MAL, MALL, NPHP1, TPC3, BUB1, ACOXL, BCL2L11, CTNNA2, LRRTM1, REG3A, LRRTM4, SYCP1, TSHB, TSPAN2
17	8.15	9.71	
17	13.21	15.07	
17	30.06	36.15	
17	42.25	46.39	
17	52.49	52.70	
19	7.33	7.91	PCDH10, C19H4orf33, SCLT1, JADE1, PGRMC2, LARP1B, ABHD18, MFSD8, PLK4, HSPA4L, INTU, FAT4, ANKRD50, HS6ST1, UGGT1, SAP130, AMMECR1L, POLR2D, WDR33, LIMS2, GPR17, MYO7B, IWS1, PROC, MAP3K2, ERCC3, BIN1, GYPC, CNTNAP5
19	8.50	13.59	
19	13.94	14.28	
19	14.49	16.03	
19	16.28	17.61	
19	21.76	24.53	
19	25.57	26.40	
24	23.14	24.39	RALY, EIF2S2, ASIP, AHCY, ITCH, DYNLRB1, PIGU, TP53INP2, NCOA6, GSS, ACSS2, MYH7B, TRPC4AP, EDEM2, PROCR, MMP24, EIF6, FAM83C, UQCC1
25	36.76	37.39	-
32	6.54	6.96	TMEM150C, SCD5, SEC31A, THAP9, LIN54, COPS4, UNC5C, STPG2
32	18.21	20.10	

[1] Canine autosome.

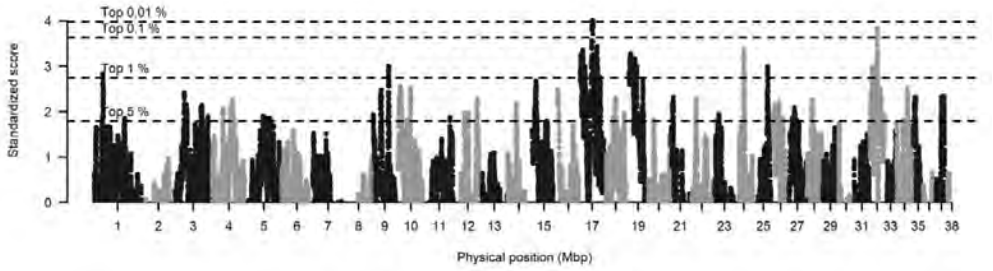


Figure 1. Distribution of varLD standardised score across autosomal genome.

Discussion

The analysis of linkage disequilibrium variability in the genome of the Czechoslovakian wolfdog confirmed that the artificial selection acted in specific genomic regions controlling different phenotypic traits, mainly coat colour, reproduction, growth, muscle and trainability.

Two genes (MC5R and ASIP) responsible for coat colour in dogs were detected within selection signals on autosomes 1 and 24. The MC5R gene was found to be related to the shedding and coat length of dogs (Hayward *et al.*, 2016) and the ASIP gene cooperates with the MC1R gene for switching between red and black pigments in dog hair (Dreger *et al.*, 2020).

Some of the genes located in regions under strong selection pressure were previously associated with reproduction in dogs, e.g. PGRMC2 (Tahir *et al.*, 2012). On the other hand, the ASIC2 located near signal on autosome 9 was found to be related to dog herding, predation, temperament and trainability traits (Shan *et al.*, 2021). On autosomes 9, 19 and were identified genes linked to muscle formation (MYO1D, MYO7B and SCD5).

On autosome 17 within selection signals, two genes causing eye diseases in dogs were found. The CYP1B1 gene was in previous studies associated with primary open-angle glaucoma in beagles (Kato *et al.*, 2009), and the NPHP1 gene was linked to retinal disease (Ahonen *et al.*, 2014). This indicates that artificial selection can affect genomic regions controlling preferred phenotypes as well as genomic regions that interact with them on a DNA level.

Acknowledgements

This study was funded by Slovak Research and Development Agency (grant numbers APVV-20-0161 and APVV-17-0060) and by the Ministry of Agriculture of the Czech Republic (grant number QK1910156).

References

Ahonen S.J., Arumilli M., Seppälä E., Hakosalo O., Kaukonen M.K. *et al.* (2014) PLoS One 9(12):e114552. https://doi.org/10.1371/journal.pone.0114552

Cagan A. and Blass T. (2016) BMC Evol Biol 16(10). https://doi.org/10.1186/s12862-015-0579-7

Caniglia R., Fabbri E., Hulva P., Bolfíková B.Č., Jindřichová M. *et al.* (2018) BMC Genomics 19:533. https://doi.org/10.1186/s12864-018-4916-2

Chang C.C., Chow C.C., Tellier L.C.A.M, Vattikuti S., Purcell S.M. *et al.* (2015) GigaScience 4:7. https://doi.org/10.1186/s13742-015-0047-8

Dreger D.L., Anderson H., Donner J., Clark J.A., Dykstra A. *et al.* (2020) Genes (Basel) 11(7):739. https://doi.org/10.3390/genes11070739

Gray M.M., Granka J.M., Bustamante C.D., Sutter N.B., Boyko A.R. *et al.* (2009) Genetics 181(4):1493-505. https://doi.org/10.1534/genetics.108.098830

Hayward J.J., Castelhano M.G., Oliveira K.C., Corey E., Balkman C. *et al.* (2016). Nat Commun 7:10460. https://doi.org/10.1038/ncomms10460

Kato K., Kamida A., Sasaki N. and Shastry B.S. (2009) Mol Vis 15:2470-2474. PMID: 19956412; PMCID: PMC2786889.

Moravčíková N., Kasarda R., Židek R., Vostrý L., Vostrá-Vydrová H. *et al.* (2021) Genes (Basel) 12(6):832. https://doi.org/10.3390/genes12060832

Ong R.T. and Teo Y.Y. (2010) Bioinformatics 26(9):1269-1270. https://doi.org/10.1093/bioinformatics/btq125

Rainer J., Gatto L. and Weichenberger C.X. (2019). Bioinformatics 35(17):3151-3153. https://doi.org/10.1093/bioinformatics/btz031

Shan S., Xu F. and Brenig B. (2021) Front Vet Sci 8:693290. https://doi.org/10.3389/fvets.2021.693290

Shannon L.M., Boyko R.H., Castelhano M., Corey E., Hayward J.J. *et al.* (2015) Proc Natl Acad Sci U S A 112(44):13639-13644. https://doi.org/10.1073/pnas.1516215112

Smetanová M., Černá Bolfíková B., Randi E., Caniglia R., Fabbri E. *et al.* (2015) PLoS One 10: e0143807. https://doi.org/10.1371/journal.pone.0143807

Tahir M.Z., Reynaud K., Grimard B., Thoumire S., Chastant-Maillard S. *et al.* (2013) Reprod Fertil Dev 25(7):1065-1076. https://doi.org/10.1071/rdv24n1ab70

Turner S. (2018) The Journal of Open Source Software. https://doi.org/10.1101/005165

752. Applications of single-molecule real-time isoform sequencing (Iso-Seq) for unravelling complexity of dog transcriptomes

W. Zhang*, L. Eory, M. Salavati, E. Clark, J. Smith, A.L. Archibald and J.J. Schoenebeck*

The Roslin Institute and Royal (Dick) School of Veterinary Studies, University of Edinburgh, Easter Bush, Midlothian, EH25 9RG, United Kingdom; wengang.zhang@ed.ac.uk; jeff.schoenebeck@roslin.ed.ac.uk

Abstract

Here we present a summary of the generation of dog Single-molecule Real-time Isoform Sequencing (Iso-Seq) data, and its application for transcriptome annotation. Iso-Seq libraries from six tissues were sequenced, and read clean-up was implemented with IsoSeq V3 tools. A high retention rate in demultiplexing and refining steps, as well as a high mapping rate (90%) of Full-length non-concatemer (FLNC) reads, guaranteed enough useful information for subsequent analyses. Finally, 17,766-45,170 clustered transcripts from the six libraries were aligned to the ROS_Cfam_1.0, a male Labrador retriever dog reference genome assembly. In terms of annotation, we identified 14,059 to 26,263 transcripts and 4,985 to 10,457 genes with the sequence data in each of the libraries, with 32 to 68% unique transcripts being expressed. Coding transcripts in each tissue are expressed in a relatively stable manner, from 60-70%. This study provides an improved dog genome annotation resource and offers new insights into tissue-specific transcriptome pattern.

Introduction

Characterization of the transcriptome is crucial to understanding gene function, biological processes, and cellular states. The most popular and easily accessible sources of transcriptome information are those provided by RefSeq and Ensembl using complex annotation pipelines and multiple data sources. However, these annotations have a portion of mis-annotated genes and are far from complete (Frankish *et al.* 2015, Uszczynska-Ratajczak *et al.* 2018).

In recent years, the development of Single-molecule Real-time Isoform Sequencing (Iso-Seq) has enabled generation of transcript reads of more than 10 Kb, which allows for full length sequencing of native RNA isoforms including messenger and noncoding transcripts.

Many studies have discussed novel transcript detection with the aid of Iso-Seq (Leung 2021, Ali 2021). In short, the production of Iso-Seq data using Pacific Biosciences (PacBio) SMRT (Single-molecule real-time) technology begins with a circularised cDNA library. Library molecules contain cDNA, primer sites for polymerase, and barcodes for sequencing pooled samples. Sequencing occurs when fluorescent signals emitted during the replication of the library's cDNA molecules are imaged and coded as bases. The raw data produced by sequencing are called circular consensus sequencing (CCS) reads. Further processing from pooled experiments includes demultiplexing CCS reads according to their barcode to generate full length (FL) reads. Of note, library construction can produce concatemeric cDNA molecules; to remove these artefacts, FL reads are processed to split concatemeric reads into individual transcripts termed full length non-concatemeric reads (FLNC). Finally, FLNC reads are collapsed via clustering algorithms to provide the sequences of unique transcripts. With post-sequencing processing, the counts of CSS, FL, FLNC, and clusters should become successively smaller as artefacts and redundancy are removed. Typically, the processed isoform data is presented in fasta file format, which can be aligned to a genome assembly to assign positional context. To discriminate coding from non-coding, transcripts are typically aligned with protein databases such as UniProt (The UniProt Consortium, 2021) or their alignment processed by tools such as FEELnc, a program developed to annotate genes that encode long noncoding RNA (Wucher *et al.*, 2017).

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_752, © W. Zhang *et al.* 2022

In this study, we have employed the Iso-Seq technology to generate a full-length transcriptome from multiple tissues of dogs. We explore the characteristics of the Iso-Seq data, and show its capacity for dog transcriptome annotation in the future.

Materials & methods

Iso-Seq data generation. RNA from three dogs was pooled by tissue to maximise detection of tissue-specific transcripts. We generated SMRT cell libraries based on TeloPrime Full-Length cDNA Amplification Kit V2, which selects the methylated G capped transcripts in amplification. A total of 8 SMRT cells were sequenced corresponding to 6 types of tissues: adrenal gland, cerebellum, liver, testis, tracheal cartilage, and scapular lymph node. Both libraries construction and sequencing on the Pacific Biosciences Sequel platform were implemented at the Vienna BioCenter Core Facilities. Sequencing data was deposited under the accession PRJEB44465.

Initial Iso-seq QC and analysis. Iso-Seq raw data were processed using the IsoSeq v3 pipeline to generate transcript sequences. The reads were aligned to the ROS_Cfam_1.0 genome with minimap2 (Li 2018), and transcriptome annotation was produced by running cDNA_Cupcake collapse.

Results & discussion

Six cDNA libraries were sequenced using SMRT technology on the PacBio Sequel platform (Figure 1A) and yielded high quality clustered sequences after processing with the IsoSeq3 pipeline. The average length of circular consensus sequencing (CCS) reads in the six libraries ranged between 1,526 bp to 2,013 bp and follow an approximately normal distribution with 90% of the reads ranging from 614 bp to 3,016 bp (Figure 1B). Table 1 shows the number of reads after each IsoSeq V3 step. In the demultiplexing step, 93.8% reads were retained on average, and in the refining step, on average 96.3% reads were maintained. The mapping ability of the Full-length Non-Concatemer (FLNC) reads in the six libraries was between 87 and 94%, with lymph node library having a slightly higher proportion of unmapped reads (13.1%) and multiple mapping reads (2.3%) than the other libraries (Figure 1C. Thus, the high retention rate in IsoSeq V3 processing and the high mapping rate of the FLNC reads indicated the good quality of the Iso-Seq data generated in this study.

Among the six tissues, brain cerebellum and liver were both sequenced on two SMRT cells and therefore have the most CCS, FL, and FLNC reads. Because the final clustering step is related to gene expression abundance in tissues, the number of clustered sequences in liver dropped dramatically. Finally, the clustering step for the six libraries yielded 17.8K to 45.2K polished sequences for subsequent annotation.

Table 1. Number of reads for each tissue after each step in the IsoSeq V3 analyses.

Tissue	CCS[1]	FL[2]	FLNC[3]	Clustered sequences
Adrenal gland	451,150	426,713	413,301	23,534
Brain cerebellum	735,425	694,546	677,261	45,170
Lymph node	454,776	425,398	407,469	18,338
Liver	642,032	608,423	593,046	25,663
Tracheal cartilage	396,669	369,628	352,364	17,766
Testis	415,720	383,018	362,361	26,782

[1] Circular consensus sequencing.
[2] Full-length reads.
[3] Full-length non-concatemer reads.

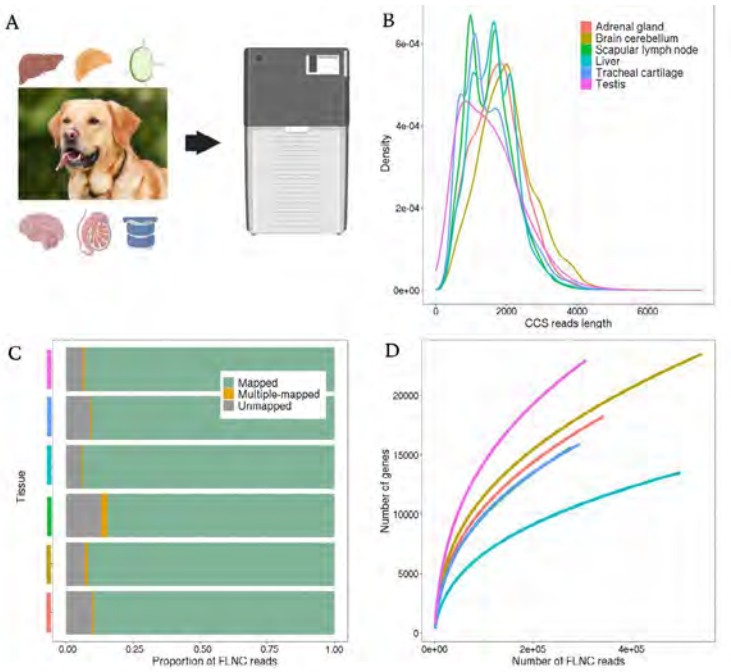


Figure 1. Overview of generation and processing of Iso-Seq data. (A) RNA from six tissues (adrenal gland, cerebellum, liver, testis, tracheal cartilage, and scapular lymph node) were sequenced on the PacBio Sequel platform. Graphic was created with BioRender.com. (B) Circular consensus sequencing (CCS) read length distribution by tissue. (C) Rarefaction curve showing gene-level does not reach saturation in the six libraries. (D) Mapping statistic of the Full-length Non-Concatemer (FLNC) reads aligned against the ROS_Cfam_1.0 genome.

Table 2 presents the number of transcripts and genes annotated in each tissue. Testis was the tissue in which the detection of expressed genes and transcripts was highest; 10,457 and 22,576, respectively. In contrast, for tracheal cartilage the expression of only 4,985 genes and 14,059 transcripts was detected. Rarefaction curve analysis at the gene level showed that sampling was still not saturated when reaching the maximum capacity of one or two SMRT cells, indicating the complexity of transcriptome (Figure 1D).

In terms of unique transcripts detected by each tissue, the proportion ranges from 32.3% (tracheal cartilage) to 68.1% (testis), with an average of 46.7%. We also predicted the coding ability of transcripts by aligning sequences against the Uniprot database and by FEELnc estimation. The transcripts, which showed coding potential in both methods, were regarded as coding transcripts. The brain cerebellum is seen to express the most coding transcripts (16,689) and tracheal cartilage has the least number (9,279). Although the brain cerebellum has nearly twice the number of coding transcripts than the least expressed tissue, the proportion of coding transcripts compared with total transcripts is relatively stable, ranging from 60 to 70%.

In summary, we have generated a high quality Iso-Seq dataset that can be used for annotation of the dog genome. In addition, knowledge of tissue-specific genes or isoforms can also provide us with a fundamental understanding of gene functions in future studies.

Table 2. Summary of Iso-Seq annotation by tissues.

Tissue	Transcripts	Genes	Unique transcripts	Coding transcripts[1]
Adrenal gland	18,541	7,632	6,938 (37.4%)	12,185 (65.7%)
Brain cerebellum	26,263	9,152	14,547 (55.3%)	16,689 (63.5%)
Lymph node	14,649	6,827	5,329 (36.4%)	9,704 (66.2%)
Liver	14,943	5,402	7,524 (50.4%)	10,500 (70.3%)
Tracheal cartilage	14,059	4,985	4,543 (32.3%)	9,279 (66.0%)
Testis	22,576	10,457	15,374 (68.1%)	13,683 (60.6%)

[1] Coding transcripts are defined as transcripts aligned with Uniprot database.

References

Ali, A., Thorgaard, G. H., and Salem, M. (2021) Front. Genet 12. https://doi.org/10.3389/fgene.2021.683408

Frankish, A., Uszczynska, B., Ritchie, G. R., Gonzalez, J. M., Pervouchine, D., *et al.* (2015) BMC Genom. 16(8):1-11. https://doi.org/10.1186/1471-2164-16-S8-S2

Leung, S. K., Jeffries, A. R., Castanho, I., Jordan, B. T., Moore, K., *et al.* (2021) Cell Reports 37(7):110022. https://doi.org/10.1016/j.celrep.2021.110022

Li, H. (2018). Bioinformatics, 34(18):3094-3100.

The UniProt Consortium. Nucleic Acids Res. (2021), 49(D1): https://doi.org/10.1093/nar/gkaa1100

Uszczynska-Ratajczak, B., Lagarde, J., Frankish, A., Guigó, R., and Johnson, R. (2018) Nat. Rev. Genet. 19(9):535-548. https://doi.org/10.1038/s41576-018-0017-y

Wucher V, Legeai F, Hedan B, Rizk G, Lagoutte L, *et al.* (2017) Nucleic Acids Res., 5;45(8):e57. https://doi.org/10.1093/nar/gkw1306

753. Genome-wide signals of positive selection identified in Livestock Guardian Dogs

M. Shihabi[1*], V. Glavaš[1], V. Cubric-Curik[1], M. Zorc[2], P. Dovč[2] and I. Curik[1]

[1]University of Zagreb, Faculty of Agriculture, Svetošimunska 25, 10000 Zagreb, Croatia; [2]University of Ljubljana, Biotechnical Faculty, Jamnikarjeva 101, 1000 Ljubljana, Slovenia; mshihabi@agr.hr

Abstract

Livestock Guardian Dogs (LGDs, *Canis familiaris*) are large dogs adapted to harsh mountain environments and selected to protect grazing livestock from predators in free-ranging grazing systems. Signals of positive selection in a meta-population represented by nine LGD breeds (116 individuals) were analysed based on information from 104,618 SNP genotypes using the extreme Runs of Homozygosity islands (eROHi) and the integrated Haplotype Score (iHS) approach. We found that 28 genomic regions (11 identified using the eROHi approach and 17 using the iHS approach) had highly significant signals for positive selection. Some of the most pronounced selection signals are associated with genes such as *LRIG3*, *TRIP11*, *CNTNAP2* and *EFNA5*. The presented results will contribute to the understanding of the synergistic behavioural interaction of LGDs with humans and grazing animals, especially sheep and goats, and the mammals' adaptation to harsh mountain conditions, as well as to their breeding management.

Introduction

Livestock Guardian Dogs (LGDs, *Canis familiaris*) are large dogs specialised in protecting mostly sheep and goat flocks from large predators (wolves, bears, etc.) in free-ranging grazing systems. The phylogenetic position of LGD is described in Talenti *et al.* (2018), while a more detailed genomic characterization can be found in Janeš *et al.* (2021). They are adapted to the harsh environment in the mountains where they help people protect their flocks, usually without supervision. At the same time, the long-term artificial selection process has shaped the behaviour of LGDs toward intelligence, trustworthiness, attentiveness, and protectiveness (Yilmaz *et al.*, 2015). The oldest clear written reference to LGDs is found in Aristotle's History of Animals from around 2,360 BP and it is quite possible that the origin of LGDs dates back even to the early beginnings of domestication of sheep and goats (Linnell and Lescureux, 2015).

The main objective of this study was to improve our understanding of the mechanisms of environmental adaptation and response to artificial selection in LGDs by identifying positive selection signals. Therefore, we used the extreme Runs of Homozygosity islands (eROHi) and the integrated Haplotype Score (iHS) approaches to analyse the available high-throughput genomic data.

Materials & methods

Data and genotyping. Our analyses referred to the meta-population represented by nine LGD breeds considered as a single group, a total of 120 dogs and 23 grey wolves genotyped for 172,115 SNPs using the Illumina CanineHD BeadChip (Illumina Inc., San Diego, CA, USA). Grey wolves, Sharplanina and Tornjak dog genotypes were generated in previous study (Janes *et al.*, 2021), the remaining LGDs (Caucasian Shepherd, Fonni's Dog, Great Pyrenees, Kuvasz, Maremma Sheepdog, Mannara Dog, and Pastore della Sila) and the grey wolves (19) were taken from publicly available genotypes published in Parker *et al.* (2017), Talenti *et al.* (2018), Shannon *et al.* (2015), and Dreger *et al.* (2016). Only autosomal SNPs with known chromosomal positions were used. During quality control performed with SAS 9.4. (SAS Institute, Cary, NC) and the software PLINK v1.9 (Chang *et al.*, 2015), we removed all SNPs with a GenTrain score <0.4, with a GenCall score ≤0.8, with a call rate <90% and SNPs that deviated from Hardy-Weinberg equilibrium

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_753, © M. Shihabi *et al.* 2022

with $P<10^{-7}$ as well as all individuals with a call rate <95%. All further analyses were performed with a final data set that included 104,618 SNPs and 116 dogs and 23 wolves.

Extreme runs of homozygosity islands (eROHi) and integrated haplotype score (iHS). ROHs were estimated for each individual using the SNP & Variation Suite (SVS) v8.7.0 software package (Golden Helix, Inc., Bozeman, MT, USA) according to the parameters described in Ferenčaković *et al.* (2013). For each SNP, ROH frequency was normalised to mean chromosomal frequency, while the transformed value was represented as -log10(P). Signals with -log(P)>10 were considered significant. The iHS analysis was calculated according to Voight *et al.* (2006) after chromosome-to-chromosome haplotype phasing was performed using Shapeit2 software (Delaneau *et al.*, 2014), with the ancestral allele option set up according to the SNP status observed in 23 grey wolves. iHS values were calculated for each autosomal SNP in the genome using the R package 'rehh' (Gautier and Vitalis, 2012). SNPs with -log(P) >3 were considered outliers, whereas non-overlapping 1 Mb windows (sliding for 250 kb) with >10% outliers were considered genomic regions with signals of positive selection. In the eROHi approach, signals are ordered by the highest -log(P) value within a region, whereas in the iHS approach, signals are ordered by the proportion of outliers within a window.

Results & discussion

Genome-wide visualisation of regions (eROHi) or windows (his) with significant positive selection signals is shown in Figure 1 and 2, while a more detailed description of the genes located within the 10 most significant regions/windows can be found in Table 1. Two different complementary approaches were used: his, which is more powerful in detecting soft sweeps, and eROHi, which is able to detect near and fully fixed selective sweeps.

It is more useful to look for windows of consecutive SNPs that contain numerous extreme values because selective sweeps tend to generate clusters of extreme values in the sweep region, whereas in a neutral model the extreme values are more evenly distributed (Voight *et al.*, 2006). Therefore, all individual iHS outliers have been coloured grey in Figure 2.

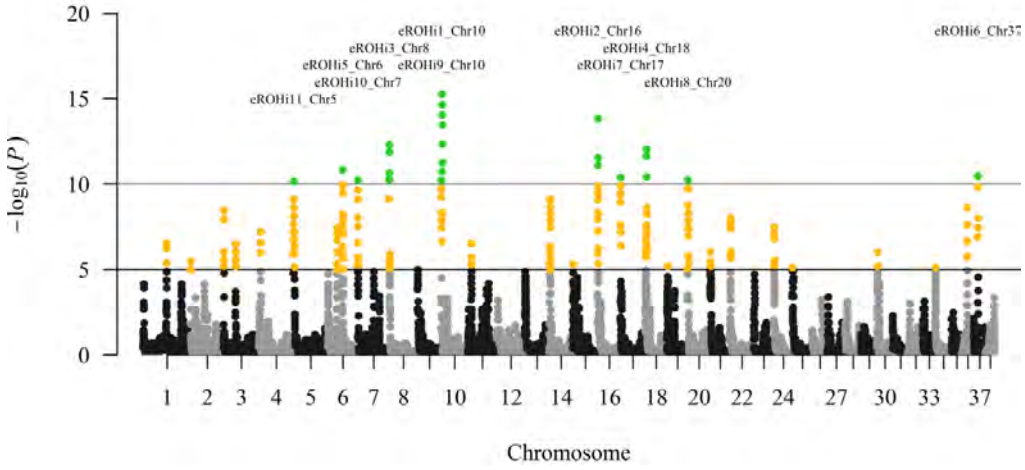


Figure 1. Genome-wide signals of positive selection analysed by eROHi approach.

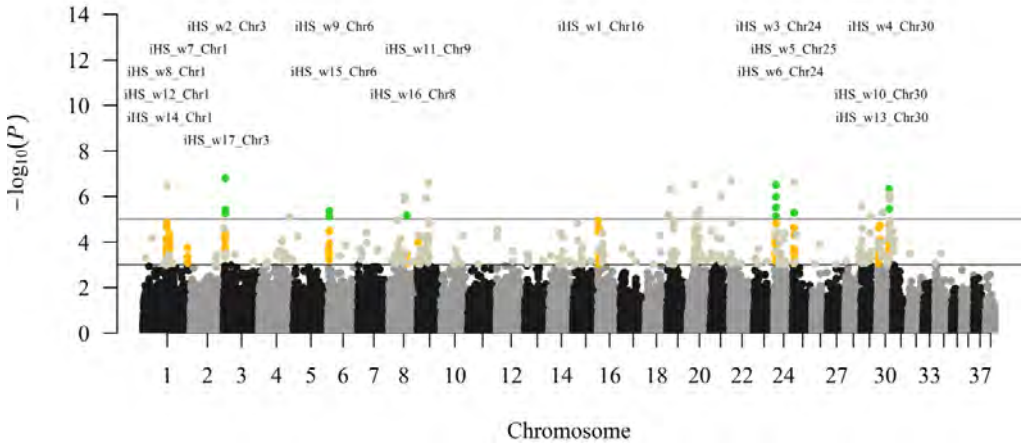


Figure 2. Genome-wide signals of positive selection analysed by iHS approach.

Table 1. Most important genes detected as positive selection signals.

Signal name	Signal region (Mb)	#SNPs[1]	Most important genes
eROHi1_Chr10	2.61-3.76	15	*LRIG3, SLC16A7*
eROHi2_Chr16	1.04-1.92	8	*IGFBP1, ADCY1, MIR8853, PDIA4, …*
eROHi3_Chr8	0.31-1.36	29	*ATXN3, TRIP11, TC2N, FBLN5, …*
eROHi4_Chr18	3.48-3.92	9	/
eROHi5_Chr6	39.4-39.42	2	*UNKL, GNPTG*
eROHi6_Chr37	11.6-11.98	8	*ICA1L, WDR12, CARF, NBEAL1*
eROHi7_Chr17	0.74-1.1	5	*SNTG2, TPO, PXDN, MYT1L*
eROHi8_Chr20	1-1.02	2	*CHCHD6*
eROHi9_Chr10	1.23-1.44	6	*STAT6, LRP1, STAC3, R3HDM2, …*
eROHi10_Chr7	1.06	1	*LMOD1*
iHS_w1_Chr16	2-3	12	*CUL1, CNTNAP2*
iHS_w2_Chr3	3.5-4.5	8	*FBXL17, EFNA5*
iHS_w3_Chr24	4.5-5.5	8	*RIN2, SLC24A3, SCP2D1, DTD1*
iHS_w4_Chr30	30-31	9	*DENND4A, MEGF11, TIPIN, RAB11A, …*
iHS_w5_Chr25	3.75-4.75	10	*CCNA1, SPART, SOHLH2, DCLK1, …*
iHS_w6_Chr24	3-4	7	*CRNKL1, NAA20, RIN2, SLC24A3, …*
iHS_w7_Chr1	114.75-115.75	5	*ZFP30, ZNF527, ZNF420, ZNF790, …*
iHS_w8_Chr1	59.75-60.75	8	*TBC1D32*
iHS_w9_Chr6	4-5	6	/
iHS_w10_Chr30	3-4	7	*DPH6, MIR8792*

[1] Significant SNPs in the region: (1) -log(P)>10 for eROHi; or (2) -log(P)>3 for iHS approach.

We have commented only on the most interesting genes identified as potential signals for positive selection. For example, according to the results of the eROHi approach, the region containing *LRIG3* (Leucine Rich Repeats and Immunoglobulin Like Domains 3), associated with aggression against foreign dogs in Zapata *et al.* (2020), showed the strongest signal. *LRIG3* may play a role in skull and inner ear morphogenesis during embryonic development (UniProtKB). The selection signal related to *TRIP11* (thyroid hormone receptor

interactor 11), gene associated with Golgi apparatus, encodes a protein that is interacting with thyroid hormone receptor beta, was also very strong. The strongest signals for positive selection identified by iHS point to *CNTNAP2* (Contactin Associated Protein 2) and *EFNA5* (Ephrin A5). Both are protein-coding genes associated with brain development and behaviour. In humans, *CNTNAP2* is associated with several neurodevelopmental disorders, while *EFNA5* prevents axon bundling in cocultures of cortical neurons with astrocytes, a model for late development and differentiation of the nervous system. *EFNA5* has been linked to the domestication process (Shin *et al.*, 2014). It is possible that some of the signals identified are related to adaptation to the harsh mountain environment and could be shared with wolves. However, this is an interesting topic for our future research.

The presented results will contribute to the understanding of the synergistic behavioural interaction of LGDs with humans and grazing animals, especially sheep and goats, and the mammals' adaptation to harsh mountain conditions, as well as to their breeding management. Nevertheless, further analyses are needed to provide a more complete picture of positive selection signals in LGDs.

Funding

This study was supported by project ANAGRAMS-IP -2018-01-8708 funded by the Croatian Science Foundation.

References

Chang C.C., Chow C.C., Tellier L.C., Vattikuti S., *et al.* (2015) Gigascience 4(1): s13742-015. https://doi.org/10.1186/s13742-015-0047-8

Delaneau O., and Marchini, J. (2014) Nature communications 5(1): 1-9. https://doi.org/10.1038/ncomms4934

Dreger D.L., Davis B.W., Cocco R., Sechi S., *et al.* (2016) Genetics 204(2): 737-755. https://doi.org/10.1534/genetics.116.192427

Ferenčaković M., Sölkner J., Čurik I. (2013) Genetics Selection Evolution 45: 42. https://doi.org/10.1186/1297-9686-45-42

Gautier M., and Vitalis R. (2012) Bioinformatics 28(8): 1176-1177. https://doi.org/10.1093/bioinformatics/bts115

Janes M., Zorc M., Ferenčaković M., Curik I., *et al.* (2021) Sustainability 13(4): 2289. https://doi.org/10.3390/su13042289

Linnell J.D.C., and Lescureux N. (2015) Livestock guarding dogs–cultural heritage icons with a new relevance for mitigating conservation conflicts. Norwegian Institute for Nature Research, Trondheim, Norway, 76.

Parker H.G., Dreger D.L., Rimbault M., Davis B.W., *et al.* (2017). Cell reports 19(4): 697-708. https://doi.org/10.1016/j.celrep.2017.03.079

Shannon L.M., Boyko R.H., Castelhano M., Corey E., *et al.* (2015) PNAS 112(44): 13639-13644. https://doi.org/10.1073/pnas.1516215112

Shin D.H., Lee H.J., Cho S., Kim H.J., *et al.* (2014) BMC genomics 15(1): 1-16. https://doi.org/10.1186/1471-2164-15-240

Talenti A., Dreger D.L., Frattini S., Polli M., *et al.* (2018) Ecology and evolution 8(5): 2911-2925. https://doi.org/10.1002/ece3.3842

Voight B.F., Kudaravalli S., Wen X., and Pritchard J.K. (2006) PLoS biology 4(3): e72. https://doi.org/10.1371/journal.pbio.0040072

Yilmaz O., Coskun F., and Ertugrul M. (2015) Journal of Livestock Science 6: 23-35.

Zapata I., Lilly M.L., Herron M.E., and Alvarez C.E. (2020) bioRxiv. https://doi.org/10.1101/2020.08.13.249805

754. Implementation of single-step GBLUP for complex behaviour traits in Labrador Retrievers used as service dogs

M.M. Riser[1*], E. Leighton[2], J. Russenberger[2], C. Moser[3] and B. Fragomeni[1,4]

[1]Department of Animal Science, University of Connecticut, 17 Manter Rd, Unit 4040, Storrs, CT 06268, USA [2]International Working Dog Registry, 4103 Hillglen Way, San Antonio, TX 78253, USA; [3]Guide Dogs New South Wales, 7-9 Albany Street, St Leonards, New South Wales, Australia [4]Institute for System Genomics, 67 North Eagleville Road, Storrs, CT 06269-3003, USA; molly.riser@uconn.edu

Abstract

The objective of this study was to evaluate the potential of genomic selection to improve behavioural traits in a population of Labrador Retrievers using the single-step GBLUP (ssGBLUP) methodology. Data consistent on phenotypes of 17 behaviour traits on approximately 5,000 individuals, their complete pedigree, and whole-genome sequence (WGS) for 463 animals. 150k SNPs were selected from WGS data. Heritabilities were low to moderate on most traits, and genomic estimated breeding values were in general more accurate than those computed with pedigrees. Gains in accuracy show that genomic selection can help the improvement of working dogs by identifying superior animals at younger ages. Gains are expected to increase as more animals are genotyped and more phenotypes are collected. The current results urge the need to continue recording behaviour traits in guide dogs and to expand the number of genotyped animals. Future steps include evaluation of multi-trait animal models and weighted ssGBLUP.

Introduction

Working dogs are highly trained in specific tasks related to their roles, including guiding people with visual impairment, servicing people with disabilities, among other services. However, not all puppies bred for pursuing their specific task will graduate from their programs due to health or behavioural issues. Genetic evaluation, like that used in ranking livestock, has been used successfully in only a small number of working dog breeding programs (Zhang *et al.*, 2009, Wilson *et al.*, 2013, Leighton *et al.*, 2019). Especially, behaviour traits pose a challenge because of the low heritability that is expected for such phenotypes in addition to difficulties regarding phenotypic coding that should be uniform across different subpopulations. The International Working Dog Registry data (IWDR; www.iwdr.org) provides an internet cloud-based database accessible from anywhere in the world for collecting and uniformly storing health and behaviour records on working dogs. The objective of the present study is to evaluate a genomic evaluation program in service dogs with single-step GBLUP using the IWDR data.

Materials & methods

Phenotypic data were collected from 4,841 Labrador Retrievers for 17 traits from IWDR behaviour checklist. A summarized description of the traits is available in Table 1. All traits were recorded with a score that varied from 1 (severe) to 5 (absent). Pedigree information included 23,593 animals starting from 1991 to 2016, with most of the data being from after 2008. Records were collected from year to year from number of different organizations, including the Guiding Eyes for the Blind (New York, USA), Guide Dogs New South Wales (New South Wales, AUS).

A total of 65,849 SNPs with minor allele frequency (MAF) less than 0.05 were removed from the data, including 22,774 monomorphic SNPs. No SNPs or animals presented call rate less than 0.9. Additionally, 43 SNPs and 8 animals were removed due to Mendelian conflicts. After quality control, a total of 137,088 SNPs were available for 457 animals.

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_754, © M.M. Riser *et al.* 2022

Table 1. Description of traits and abbreviation used.

Behaviour trait	Description	Abbreviation
Anxiety in new situations	Anxious during first few visits to unfamiliar locations	AnxNewSit
Noise sensitivity	Startle, tense body language, escapes or displacement behaviours when exposed to loud noises excluding thunder	Noise
Self-modulation	It takes the dog a long time to return to a productive emotional state following exposure to an arousing or stressful stimuli	Selfmod
Handle sensitivity	Drops rear quarters when harness handle lays on back	Hsens
Fear of under footings	Fearful or apprehensive of various walking surfaces. Scored 1(severe) to 5 (absent)	UF
Body Sensitivity	Inhibited by physical contact with objects other than harness handle on back	Bsens
Activated by stress	Copes with stressful situations poorly evidenced by becoming more active with faster movements, taking treats harder, more distracted	Actstress
Inhibited by stress	Copes with stressful situations poorly evidenced passively by shutting down, avoidance and withdrawal as a response to stress	Inhibstress
Socially inappropriate with people	Exhibits poor social manners with people	SocInap
Distracted by scent	Distracted by olfactory stimuli	Sniff
Distracted by dogs	Persistent interest in and high excitability level with other dogs	Dogdist
Excited by movement	Easily distracted by non-animal movement and has trouble redirecting attention	Movexit
Excitable	Increases energy and arousal level without observed stress signals in response to stimuli	Excit
Novel objects	Fearful avoidant or suspicious of new or novel objects in the environment	NovelObj
Retreats from known person	Moves head or face away when reached for by known person	Retreats
Fear of strangers	Fearful, nervous, apprehensive of strangers	Strange
Body handling	Avoidant, anxious, or aggressive when being handled for noninvasive activities like a vet exam/grooming/nail trim	BHand

Variance components were estimated under the average information restricted maximum likelihood (AIREML) method. All the traits followed the model:

$$y_{ijkl} = sex_i + yob_j + cg_k + a_l + e_{ijkl} \quad (1)$$

Where y_{ijkl} is the behaviour phenotype for animal (a_l), of $sex_{i,}$ born in year of birth (yob_j), and with data collected in contemporary group (cg_k), and e_{ijkl} is the residual for this observation. The effects of sex, cg, and yob were considered fixed, while animal and residual were random. The contemporary group was a concatenation of the organization that carried out the animal behaviour assessment with the month and year of collection.

Once variance components were estimated, the best linear unbiased predictors (BLUP) were calculated by solving the Henderson system of equations (Henderson, 1975). Such values, calculated using pedigree to model the additive genetic effect will be hereby termed estimated breeding values (EBV). Genomic information was incorporated to calculated genomic estimated breeding values (GEBV) for all traits under a single step GBLUP (ssGBLUP) approach (Aguilar *et al.*, 2010, Christensen and Lund, 2010).

Accuracy of breeding values were assessed using a forward in time cross-validation approach. Such a method mimicked the selection of young animals. In the first step, GEBVs were calculated using the complete data. These were the benchmark and termed $GEBV_{COM}$. In the second round of analysis, 100 validation individuals were selected randomly from a list of 147 dogs that had their own phenotypes and at least ten progeny. Those animals and their progeny had their records masked, and their (G)EBV calculated with the

reduced data and termed $GEBV_{RED}$. The accuracy for (G)EBV was the average correlation of $GEBV_{COM}$ and $GEBV_{RED}$ after five replicates.

Results

Heritablities were on average 0.13 and varied from 0.02 (Retreats) to 0.37 (Hsens) (Figure 1). The accuracy was on average 0.64 for the pedigree BLUP and ranged from 0.53 (Movexit) and 0.76 (Sniff). The average accuracy for ssGBLUP was 0.70 and ranged from 0.58 (Bsens) to 0.81 (Sniff).

Discussion

Heritabilities were low to moderate for all traits, except for handle sensitivity. Such values suggest that genetic progress may be achieved for the traits in the study at a possible slow pace. Values in the present study are lower than health traits measured in a similar population (Leighton *et al.*, 2019). Schultz *et al.* (2013) estimated heritability for behaviour traits assessed through a questionnaire, and most of the heritabilities were reported to be under 0.10, which agrees with the current findings. Genomic selection should be used to overcome the low heritability limitation, as it is expected to increase the accuracy of prediction for breeding values of young animals, especially for traits that are expensive and hard to measure and traits with low heritability (VanRaden *et al.*, 2009, Harris and Johnson, 2010, Su *et al.*, 2012).

Gains in accuracy achieved by the inclusion of genomic data ranged from 0.22 (anxiety in new situations) to -0.03 (fear of strangers). Those results illustrate the potential of ssGBLUP to increase the genetic gain in a breeding program for most traits. It is obvious that the number of genotyped animals is still lower than what it should be, and gains are expected to increase as more individuals are genotyped. Additionally, the validation approach is not optimal due to only 147 animals out of 467 had at least ten progeny. This limitation likely overestimated the accuracies, as the breeding values with complete and reduced data did not change as much if more genotypes were available in the study. However, the trend of improvement encourages more effort in implementing a genomic selection program in this population.

While heritabilities were low and gains in accuracy were modest for most traits, the results suggest that a genomic selection program for behaviour in guide dogs is feasible. The quality of phenotypes and their consistency across populations allows the program to continue to expand. More animals should be genotyped so that accuracies can reach their potential, which will maximize genetic gain. The next steps for this project are to test multiple trait models and weighted ssGBLUP and evaluate their impact in accuracies

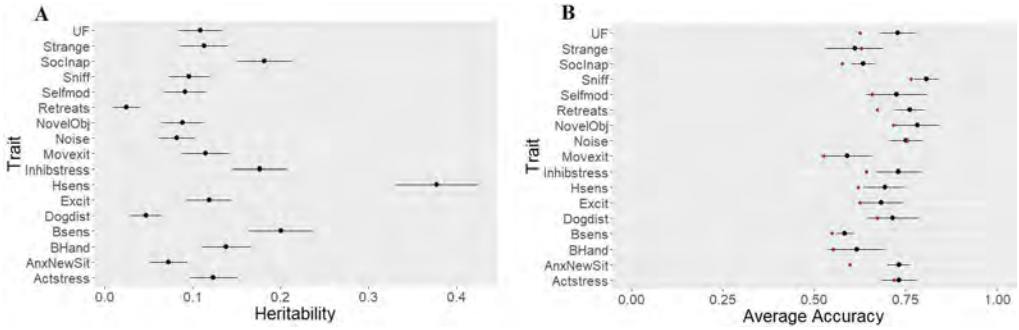


Figure 1. (A) Individual trait heritabilities. The horizontal line represents the standard error of the traits. (B) Black dots represent the average accuracy of GEBV with ssGBLUP across five replicates, and the horizontal bar is the standard deviation. The red dots represent the pedigree BLUP average accuracy.

and breeding strategies. We aim to reduce the dimensionality of the analysis by clustering traits with high correlations, which are hypothesized to be high. Additionally, we will investigate which traits are the best predictors of success of graduating from training. Finally, we intent to use this information to create an index to select animals for higher graduation likelihood.

References

Aguilar, I., Misztal, I., Johnson, D.L., Legarra, A., Tsuruta, S. *et al.* (2010) J. Dairy Sci., 93(2), pp.743-752. https://doi.org/10.3168/jds.2014-9125

Christensen, O.F. and Lund, M.S., (2010.. Genet. Sel. Evol. 42(1), pp.1-8. https://doi.org/10.1186/1297-9686-42-2

Danecek, P., Auton, A., Abecasis, G., Albers, C.A., Banks, E., *et al.* (2011) Bioinformatics, 27(15), pp.2156-2158. https://doi.org/10.1093/bioinformatics/btr330

Henderson, C.R., (1975). Biometrics, pp.423-447. https://doi.org/10.2307/2529430

Leighton, E.A., Holle, D., Biery, D.N., Gregor, T.P., McDonald-Lynch, M.B., *et al.* (2019). PloS one, 14(2), p.e0212544. https://doi.org/10.1371/journal.pone.0212544

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A., *et al.* (2007). Am. J. Hum. Genet., 81(3), pp.559-575. https://doi.org/10.1086/519795

Schulz, K.M., Moser, D.W., Bormann, J.M., Leighton, E.A. and Schulz, K.M., (2013) Proc. Of the 62nd Annual National Breeders Roundtable, St Louis, MO. p.26.

Team, R.C., 2013. R: A language and environment for statistical computing.

Zhang, Z., Zhu, L., Sandler, J., Friedenberg, S.S., Egelhoff, J., *et al.* (2009). Am. J. Vet. Res 70(4), pp.483-492. https://doi.org/10.2460/ajvr.70.4.483Wilson, B.J., Nicholas, F.W., James, J.W., Wade, C.M. and Thomson, P.C., (2013).. PLoS One, 8(10), p.e77470. https://doi.org/10.1371/journal.pone.0077470

Zheng, X., Levine, D., Shen, J., Gogarten, S.M., Laurie, C. *et al.* (2012). Bioinformatics, 28(24), pp.3326-3328. https://doi.org/10.1093/bioinformatics/bts606

755. Longevity and major causes of death are heritable in Irish Wolfhound

C.J.A. van Gemert[1,2], J.J. Windig[2] and P. Berg[1*]

[1]Norwegian University of Life Sciences, P.O. Box 5003, 1432 Ås, Norway; [2]Wageningen University & Research, P.O. Box 338, 6700 AH Wageningen, the Netherlands; peer.berg@nmbu.no

Abstract

Irish Wolfhound (IW) is a large dog breed developed several centuries ago with an average longevity of 7 years The genetics behind the short lifespan is poorly understood. Based on data from the Irish Wolfhound database (IWDB) on 118,095 pedigreed IW and data from the IW Longevity Study including 10,122 dogs with known longevity and 6,057 of these having a known cause of death (COD), longevity and the two most common COD were analysed. Heritability, maternal effects, and correlations were estimated accounting for sex, birthyear, country of origin and inbreeding coefficient. Both longevity (0.27) and COD (0.17-0.20) were moderately heritable. Longevity and COD showed low correlations, both genetic and residual correlations. We argue that this is likely due to short longevity being caused by early aging and COD being specific expressions of early aging. In addition, inbreeding was found to significantly reduce longevity. Implications for increasing longevity in IW are discussed.

Introduction

Dogs are popular pet animals, displaying a large variation in size, color, temperament and behaviors. Dogs are the species with the most recorded genetic defects (OMIA 2022), which is also reflected in differences in health and longevity of different breeds. Larger breeds tend to have a shorter lifespan than smaller breeds because larger breeds tend to age more quickly (Kraus *et al*., 2013). Larger breeds spend relatively more energy to reach adult size than small breeds. Thereby they would have relatively less energy left for health maintenance (Fan *et al*., 2016). Giant and large breed dogs are also at an increased risk of osteosarcoma (bone cancer), relative to smaller breeds (Breur *et al*., 2001). So size has an important influence on longevity and disease risk in dog breeds.

The Irish wolfhound is a rough-coated dog known as the largest and tallest of all coursing breeds, though not the heaviest breed. Their minimum size is defined by the American Kennel Club as 81 cm for mature males and 76 cm for mature females. They are also characterized by a median longevity lower than expected. Leading causes of death for Irish Wolfhounds are dilated cardiomyopathy and bone cancer (Urfer *et al*., 2007), which are also the most reported causes of death for dog breeds in general, besides age (Lewis *et al*., 2018).

The objective of this paper is to evaluate the potential for selecting for a longer and healthier life for the Irish Wolfhound breed, by estimating genetic variation in longevity, main causes of death and their correlation.

Materials & methods

Material. Data were obtained from the Irish wolfhound database (IWDB 2021), including an estimated 98% of all Irish wolfhounds ever registered. We used the complete and public IWDB dataset and the private Irish Wolfhound Longevity Study dataset. The pedigree used consists of 167,132 dogs, from 1859 to 2020, with 11,558 dogs having longevity records and 6,057 of these having a recorded cause of death.

There are two possible causes of bias in the longevity records. First, the older animals have a potential recording bias since recording started in 2012, and recordings back in time are likely to favor surviving and

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_755, © C.J.A. van Gemert *et al.* 2022

important older dogs. Secondly, animals born in recent years do not all have longevity records yet. This is illustrated in Figure 1, with a reduced longevity in recent years, solely due to incomplete recordings. We evaluated these causes of bias by analysing different subsets of data, including years 1859-2011, 1979-2013 and 1979-2020 respectively. Results of these analyses are not included here. Further analyses presented here are based on the 1979-2013 cohort, containing 118,095 dogs in total, of which 10,122 have longevity records and 6,067 a known cause of death. There are 133 different classifications for cause of death (COD) recorded. Here we include the two most frequent categories, summarizing heart disease and cancer, recorded as 0/1. Data is summarized in Table 1.

Statistical analysis. For genetic analyses the following model was used:

$$Y_{ijklm} = Sex_i + Country_j + BY_k + \beta \times F_i + a_l + c_m + e_{ijklm}$$

Where Y is longevity or COD, Sex is male or female, Country is one of 18 countries of birth, BY is birth year, F is individual inbreeding coefficient calculated from pedigree, and a is random additive genetic effect, c is random common environmental dam effect and e is the residual.

Data were analysed using R 4.0.3 (R Core Team, 2020) and ASReml 4.1 (Gilmour *et al.*, 2015) with uni- or bi-variate models. In bivariate models the common environmental dam effect c was excluded.

Results

Irish Wolfhound is characterised by a short average lifespan of 7.25 years, see Table 1. More than half of the dogs dies from one of the two most common categories heart disease and cancer. Average lifespan is slightly shorter in dogs categorized as having died from heart disease or cancer, with an average of 6.74 and 6.86 years respectively. Longevity is however highly variable with observations ranging from 0 to 5,479 days (15 years). More specifically 18% were 5 years or younger, 57.7% were 7 years or older and 27.6% were 9 years or older. Figure 1 shows average longevity by birth year and a slight negative trend in longevity can be observed.

Heritability of longevity was estimated at 0.27 (Table 2), slightly higher than for the two main causes of death with heritabilities of 0.17 and 0.20, respectively. Thus, there is a large genetic variance in longevity, a genetic standard deviation of 482 days. Common environmental dam effects were small. Common environmental dam effects were left out of the model in bivariate analyses due to convergence problems.

The genetic correlation between cancer and longevity was positive but unfavorable, with increasing cancer resulting in increased longevity however the genetic correlation was not significantly different from zero. In contrast the genetic correlation between heart disease and longevity was favorable (-0.24) and significantly different from zero. Residual correlations between longevity and causes of death were low but unfavorable.

Table 1. Number of dogs with observations on longevity, cause of death (COD), prevalence of COD and mean longevity (days + SD) of all dogs and those diagnosed with a specific COD.

Trait	No. observations	Prevalence	Mean longevity
Longevity	10,114		2,648±925
COD Heart disease	6,057	18%	2,459±804
COD Cancer	6,057	34%	2,503±709

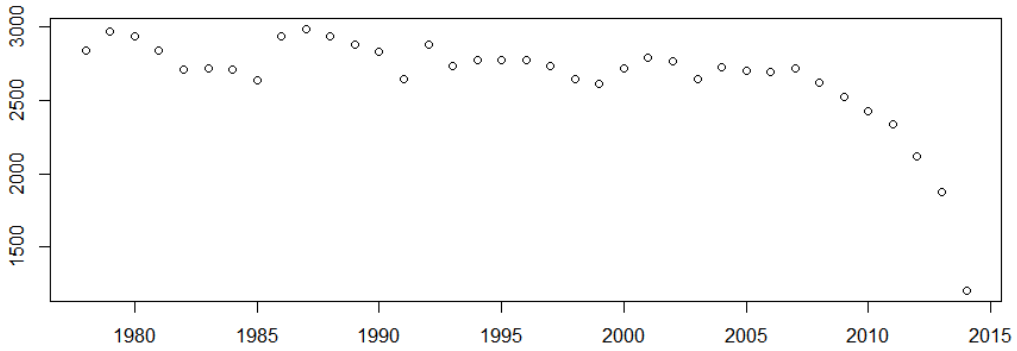


Figure 1. Mean longevity versus year of birth for Irish Wolfhound.

Table 2. Heritability, common litter effects, genetic (above diagonal) and residual (below diagonal) correlations (+ SE).

Trait	h^2	c^2	Correlations		
			Longevity	COD heart	COD cancer
Longevity	0.27±0.03	0.04±0.01		-0.24±0.09	0.10±0.09
COD heart	0.17±0.03	0.02±0.02	0.06±0.02		-[1]
COD cancer	0.20±0.03	0.02±0.02	0.01±0.02	-[1]	

[1] Not estimated.

Inbreeding significantly affected longevity. An increase of the inbreeding coefficient of 1% unit (0.01 increase) decreased longevity by 9.6 (±2.6) days, equivalent to a 0.4% reduction in longevity. Inbreeding did not significantly affect the two COD categories.

Discussion

Longevity observed in this material is in agreement with longevity observed in previous studies on Irish Wolfhound (Simpson, 2016; Urfer *et al.* 2007).

We found a significant heritability and a large genetic standard deviation of longevity of 482 days. This indicates a large potential for increasing longevity by selection. However, there are at least two challenges. First, the current relatively low longevity is likely caused by previous selection for breed characteristics and/or the severe bottlenecks that the breed has experienced during its history. Size is an important breed characteristic, but generally size is known to be negatively related to longevity across breeds and the main effect seems to be an earlier onset of aging (Kraus *et al.* 2013). Thus, a better understanding of the relationship between breed standard and health/longevity is needed. Secondly, both longevity and cause of death are observed late in life, most often after reproductive life. Thus, selection of young individuals can only be done with low accuracy, unless genetic evaluation utilizing information from relatives are implemented. Even more promising would be the implementation of genomic selection which would allow for increased accuracy on young dogs (Meuwissen *et al.* 2001). Jointly, these two challenges point to the need for early indicators of health and longevity. Promising indicators could likely be related to early indicators of aging.

Genetic and residual correlations between longevity and the two main causes of death were low and generally close to zero. This can be explained by the specific causes of death being a consequence of faster aging, but that the specific cause of death is partly caused by other loci in addition to environmental effects.

A specific and common heart disease, dilated cardiomyopathy, has been found to be affected by three loci on chromosomes 1, 21, and 37 of Irish Wolfhound (Simpson *et al*., 2016; Philipp *et al*., 2012).

We found a significant and unfavorable effect of inbreeding on the longevity of an individual. This implies that there are deleterious recessive alleles for longevity segregating in the population. Targeting these specific loci would require a large genotyping effort and would take a long time since samples for genotyping of ancestral individuals does not exist. Alternatively, the effect of inbreeding can be counteracted by the use of dogs from other breeds. However, this requires a continued gene flow from one or several other breeds to have a lasting effect on genetic diversity (Windig & Doekes, 2018).

Conclusions

Our results document a large potential for genetically increasing health and longevity in Irish Wolfhound. In the short term the most promising route seems to be identification of early indicators of aging, and in the longer term the development of a genotyped and phenotyped reference population to allow accurate genetic evaluation of younger dogs.

References

Breur G. J., Lust G., & Todhunter R. J. (2001). Genetics of canine hip dysplasia and other orthopaedic traits. In: *The Genetics of the Dog. Ruvinsky A and Sampson J, ed CABI Publ, Wallingford, UK*, pp 267-98.

Fan R., Olbricht G., Baker X., & Hou C. (2016). *Aging (Albany NY)*, *8*(12), 3209. https://doi.org/10.18632/aging.101081

Gilmour, A. R., Gogel, B. J., Cullis, B. R., Welham, S. J. and Thompson, R. (2015). ASReml User Guide Release 4.1 Functional Specification. https://www.vsni.co.uk

IWDB (2021). https://iwdb.org/

Kraus C., Pavard S., & Promislow D. E. (2013). *The American Naturalist*, *181*(4), 492-505. https://doi.org/10.1086/669665

Lewis T. W., Wiles B. M., Llewellyn-Zaidi A. M., Evans K. M., & O'Neill D. G. (2018). *Canine genetics and epidemiology*, *5*(1), 1-17. https://doi.org/10.1186/s40575-018-0066-8

Meuwissen T.H.E., Hayes B.J., and Goddard M.E. (2001) Genetics 157(4):1819-1829. https://doi.org/10.1093/genetics/157.4.1819

OMIA, Online Mendelian Inheritance in Animals. Available at: https://www.omia.org/.

Philipp U., Vollmar A., Häggström J., Thomas A., & Distl O. (2012). *PLoS One*, *7*(6), e36691. https://doi.org/10.1371/journal.pone.0036691

R Core Team (2020). https://www.R-project.org/.

Simpson S., Dunning M. D., Brownlie S., Patel J., Godden M., *et al.* (2016). *BioMed research international*, *2016*. https://doi.org/10.1155/2016/6374082

Urfer S. R., Gaillard C., & Steiger A. (2007). *Veterinary quarterly*, *29*(3), 102-111. https://doi.org/10.1080/01652176.2007.9695233

Windig J. J., & Doekes H. P. (2018). *Journal of Animal Breeding and Genetics*, *135*(3), 238-248. https://doi.org/10.1111/jbg.12330

756. The Vaika project: a multidisciplinary study of canine aging and the genetic mechanisms influencing the aging process

H.J. Huson[1*], E.L. Andrianova[2], D.I. Fleyshman[3], L. Brodsky[4], J.P. Loftus[5], J.J. Wakshlag[5] and A.V. Gudkov[3]

[1]Department of Animal Science, Cornell University, Ithaca, NY 14853, USA; [2]Vaika, Inc., East Aurora, NY 14052, USA; [3]Roswell Park Comprehensive Cancer Center, Buffalo, NY 14263, USA; [4]Tauber Bioinformatics Research Center, University of Haifa, Haifa, Israel; [5]Department of Clinical Sciences, Cornell University, Ithaca, NY 14853, USA; hjh3@cornell.edu

Abstract

The Vaika Project (https://www.vaika.org/) formed a colony of 103 retired Alaskan sled dogs (ASD) ages 8-11 years old to study the genetics of mammalian aging. The aim was to study genome variation and destabilization and link this with clinical signs of aging. Retired ASD provide a unique canine model: genetically distinct but more diverse than modern breeds. These dogs are exposed to the natural environment throughout their life and adaptable to our research colony setting. Extensive evaluation of physical performance, immune response, cognitive function, and health are used to assess the progression of aging. Illumina Canine high-density single-nucleotide polymorphism (SNP) arrays and whole-genome sequencing (WGS) data are used to identify genome variation among the dogs in the Vaika population and retrotransposons expansion in somatic cells over time. Preliminary genetic associations with arthritis were found, and retrobiome expansion with age and cancer was characterized.

Introduction

Aging is the functional decline of various physiological systems, leading to a decrease in resilience with age. Current theories of aging can be separated into two complementary groups that explain aging as either: (1) the accumulation of biological errors; or (2) a pre-determined genetic program. Regardless of the theory, genome destabilization plays a vital role in aging, and specific variants are often implicated in the clinical variation of symptoms associated with aging. Canines offer the opportunity to study aging in an organism that ages similarly to humans. They are exposed to the same environmental factors as humans, develop similar age-related diseases, have comparable medical treatments (vaccines, medications, etc.), and have an established body of knowledge on their physiology and genetics. However, privately-owned dogs introduce undesired 'noise' related to varying environmental factors and owner compliance in research protocols. Conversely, the use of purpose-bred laboratory dogs eliminates their natural exposure to environmental factors.

Aging can be viewed as a genetically programmed process that is associated with the accumulation of cells with damaged genome and dysregulated epigenome in somatic tissues that serve as a source of progressive chronic inflammation ('inflammaging') (Franceschi *et al.*, 2007). The activity of endogenous retrotransposons (i.e. SINEs, LINEs, and endogenous retroviruses collectively forming 'retrobiome') are major genome destabilizing factors that are especially active in dogs (Fleyshman *et al.*, 2021). Therefore, differences in the aging of individual dogs are influenced by inherited genetic content and are expected to be manifested by the accumulation of cells with expanded retrobiome. Both can be analysed using genomic methodologies. The Vaika Project aims to use WGS and SNP data: (1) to identify genetic variants associated with the decline of immune function, physical condition, cognitive function, and other aging parameters; and (2) evaluate genome destabilization via quantification of SINE and LINE expansion. While both aims seek to identify genome variation associated with aging parameters, they are distinct yet complementary approaches. Aim 1 identifies genomic regions highlighted by genome-wide association studies (GWAS) likely influencing phenotypic variation of aging parameters. It compares the individual dogs' genomes

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_756, © H.J. Huson *et al.* 2022

and looks for common variants within the population to aging symptoms. Aim 2 focuses more on genome variation occurring within an individual over time, specifically quantifying the accumulation of retrotransposons within an individual's cells and the association of this retrobiome expansion with aging parameters. Aim 2 will also look for common population patterns of retrobiome expansion over time regarding where the expansion occurs in the genome and to what degree it expands.

To this end, the Vaika Project created a research colony of 103 retired ASD, limiting the drawbacks and capitalizing on the benefits of privately owned and laboratory-based dogs (Fleyshman *et al.*, 2021). The ASDs have been exposed to similar environmental factors given their lifestyle and are now housed in a controlled research setting allowing for consistent diet and care with rigorous multi-disciplinary and complementary assessments of aging and resiliency in a longitudinal study. While this study of canine aging is advantageous for improving our knowledge of human aging, it may also help extend our canine companions' lifespan and quality of life.

Materials & methods

Canine colony. A cohort of 103 retired ASD ages 8-11 years old from 23 kennels spanning North America was adopted into a research colony housed at Cornell University for the remainder of their lifespan in a longitudinal study (6+ years) known as the Vaika Project. Basic health, pedigree, and performance data were collected upon admission. All dogs are maintained on the same diet (excluding medical diets) at optimal body condition, have routine medical care, and participate in outdoor exercise and enrichment programs daily. The Cornell IACUC approved the protocol (2018-0022) with Cornell's Center for Animal Resources and Education overseeing animal husbandry and veterinary care.

Genotypic and phenotypic characterization. All dogs were genotyped on the Illumina Canine HD array with 220,661 SNPs. Quality assessment and association studies were performed using Golden Helix SNP and Variation Suite v8.x (Golden Helix, Inc., Bozeman, MT, www.goldenhelix.com). DNA samples from healthy lung tissue and spontaneously developed tumours of several dogs were subjected to WGS with an average of 50x coverage. WGS libraries were prepared using the TruSeq DNA PCR-free preparation kit (Illumina, Inc.) and performed sequencing on a NovaSeq6000. A novel bioinformatics methodology, named CLARET (for classification of retroelements), has been developed to identify and quantify somatic integrations of SINE elements (in preparation) and applied to determine the degree of retrobiome expansion in individual dogs and their tumours. Phenotypic data spanning immune function, physical condition, cognitive function, and health parameters are collected annually, every nine months, or biannually. These parameters are correlated with the outcome of genetic analyses to identify potential connections between inherited genetic variants and retrobiome expansion with aging.

Here we present a preliminary GWAS for canine arthritis, meeting the objective of aim 1 by identifying common inherited genetic variants within the population associated with this aging symptom. Individual dog (n=102) computed tomography (CT) scans taken 12 months after trial initiation were used to create a categorical score for arthritis based on presence and severity. Scores were determined by a single radiography veterinarian and ranged from 0, unaffected, to 17, denoting dogs severely affected with arthritis in multiple areas of the body. In addition, we compared the number of reads for repeated elements quantified from WGS data from DNA samples. The DNA was obtained from peripheral blood of the same Labrador Retriever at 5 and 9 years of age (obtained from Cornell Biobank) and from an ASD tumour tissue (osteosarcoma) and healthy lung tissue obtained from the same dog as a control.

Statistical analyses. Genome-wide association studies (GWAS) were performed on the SNP array data using a single-locus additive mixed model incorporating a genomic relationship matrix (EMMAX) to

correct for population structure. A Bonferroni corrected *P*-value<0.05 established significance. CLARET technology enables the statistical estimation of retrobiome expansion scales in individual DNA samples. With the main effect and pair-interaction factorial design for parameters (factors) of CLARET determining the detection of new SINE repeat integrations, we compared differences in counts of collected somatic integrations between younger and older ages of dogs and between normal and tumour tissues. The Illumina read-coverage normalized whole-genome distribution for concentrations of hybrid-reads (each concentration indicates one germline integration) was approximated by combining two Beta distributions. The expectation-maximization algorithm performed an approximation in each sample.

Results

After quality control filtering using a SNP call rate <0.9, >2 alleles, and a minor allele frequency <0.03, 164,580 SNPs remained for analysis. All samples had a call rate >99%. For the GWAS on arthritis score, 75 dogs were determined arthritis-free, while 28 dogs had scores ranging from 1-17 with an average severity score of 4.5 within affected dogs (population average 1.3). The two most significant associations were on chromosomes 26 and 15, with multiple SNPs associated on chromosomes 18 and 27 (Figure 1). This included two SNPs spanning a 50 kb region on chromosome 18 within the Succinyl-CoA:Glutarate-CoA Transferase (*SUGCT*) gene.

CLARET methodology was developed and adapted to analyse germline and somatic contents of dog SINE elements. Preliminary results revealed a progressive increase in somatic cell mosaicism in acquiring new integrations of SINEs belonging to the most evolutionary recent and transcriptionally active subfamilies of C elements. A similar expansion of specific subfamilies of these SINE elements was seen in aging Labrador Retriever dogs (Figure 2A) and in the comparison of tumour cells to healthy somatic cells from ASDs (Figure 2B).

Discussion

The Vaika Project offers a multidisciplinary approach to studying aging using ASD as a canine model. The extensive phenotypic assessment of aging of this cohort of dogs provides the opportunity to identify genetic variants associated with aging symptoms. Notably, the *SUGCT* gene has been implicated in human musculoskeletal abnormalities (Nakabayashi *et al.*, 2002), making it a promising candidate gene for canine arthritis. With time, we will have multiple CT scans of the same dogs, plus physical exam and treadmill gait analysis data that may complement and add to our arthritis phenotype. Retrobiome analysis showed similar SINE subfamilies multiplying with aging and tumorigenesis. The estimated speed of this process is measured in hundreds of new integration events in somatic cells and even faster invasion of retrobiome in

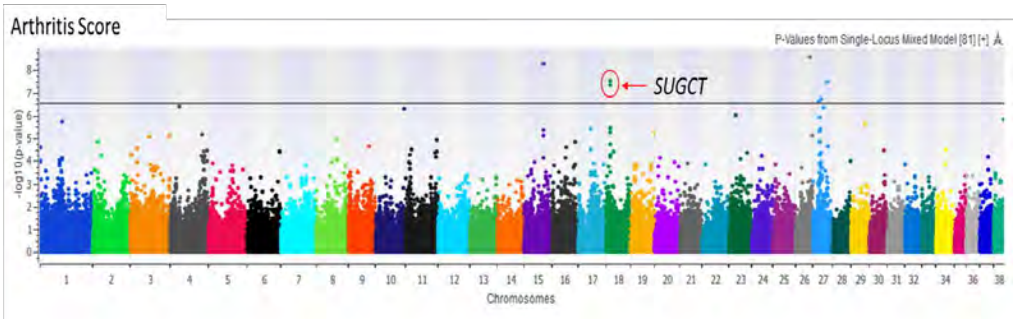


Figure 1. Manhattan plot of arthritis score genome-wide association study. Arthritis was categorically scored from 0 (unaffected) to 17 (severely affected) based upon CT scan readings of 102 dogs. Horizontal black lines represents a Bonferroni *P*-value=0.05.

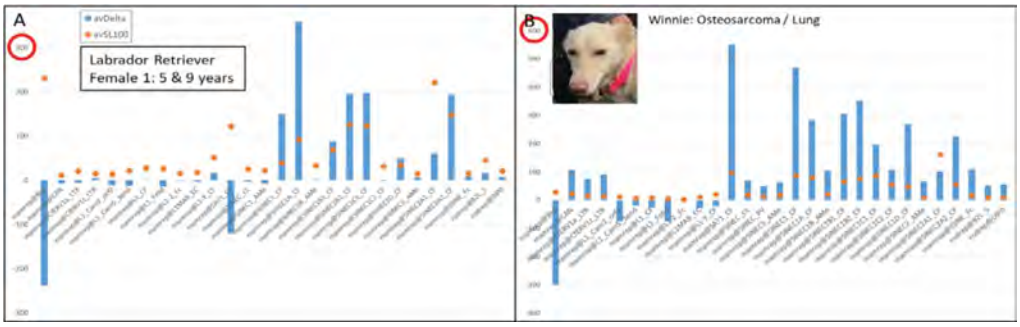


Figure 2. Subfamilies of SINE elements that underwent similar degrees of expansion in somatic cells in aging dogs. (A) Difference in copy number of SINE elements detected in peripheral blood DNA collected from the same Labrador Retriever dog at ages 5 and 9 years. (B) Difference in copy number of SINE elements detected in DNA of healthy lung tissue compared to osteosarcoma tumour of the same Alaskan sled dog. X-axis highlights SINE subfamilies; Y-axis corresponds to the difference in copy number variation.

carcinogenesis, thus, reflecting a high degree of genomic instability accompanying progression of canine tumours. As years and dogs pass, we will continue to analyse retrobiome activity in peripheral blood, and normal and tumour tissues. The Vaika Project allows us the opportunity to explore the integration of retroelements in an individual's genome over time and to identify population variation that is inherited or acquired, which affects aging.

References

Fleyshman, D. I., Wakshlag, J. J., Huson, H. J., Loftus, J. P., Olby, N. J., *et al.* (2021). Aging 13(18), 21814-21837. https://doi.org/10.18632/aging.203600

Franceschi, C., Capri, M., Monti, D., Giunta, S., Olivieri, F., *et al.* (2007). Mech Ageing Dev, 128(1), 92-105. https://doi.org/10.1016/j.mad.2006.11.016

Nakabayashi, K., Fernandez, B.A., Teshima, I., Shuman, C., Proud, V.K., *et al.,* (2002). Genomics 79(2);186096. https://doi.org/10.1006/geno.2002.6695

757. Genomic diversity in selected Austrian dog populations

G. Mészáros[1*], M. Mészárosová[2] and A. Geretschläger[3]

[1]University of Natural Resources and Life Sciences, Vienna, Gregor-Mendel-Straße 33, 1180 Vienna, Austria; [2]Slovak University of Agriculture, Tr. Andreja Hlinku 2, 949 76 Nitra, Slovakia; [3]FERAGEN GmbH, Strubergasse 26, 5020 Salzburg, Austria; gabor.meszaros@boku.ac.at

Abstract

Genomic diversity was assessed in four Austrian dog populations using SNP marker data. There was considerable variation in inbreeding and heterozygosity levels within and between breeds. The correlations for these two metrics for individual dogs was r=-0.68, with considerable variation between breeds. The differences were likely caused by different considerations of IBS and IBD markers. In practice a few breeders tend to base their choice of mate pairs only on heterozygosity levels. In our paper we show this approach as counter-productive, leading to on average higher inbreeding levels in the offspring.

Introduction

Overall genetic diversity is one of the most important concerns of dog breeding organizations, and many professional breeders. With the availability of SNP marker data, consistent and standardized screening of various diversity parameters is possible (see e.g. Janeš *et al.*, 2021). The analysis of diversity based on dense genomic markers has several advantages over traditional systems, most importantly it does not rely on pedigree records, which might be missing or incomplete in some cases. Even though precise pedigree records might exist in the established traditional breeds, they might be inaccessible due to the overall secrecy surrounding especially the elite canine breeding scene.

In this work we focus on selected aspects of genetic diversity, such as heterozygosity, inbreeding levels and genomic relationships among potential mating pairs in four Austrian dog populations (Boxer, Leonberger, Rhodesian Ridgeback and Standard Poodle). The aim of our work is to explore these characteristics in the four dog populations from Austria, and highlight the relationships, or possible misconceptions related to them.

Materials & methods

Parameters of genomic diversity were assessed for 75 Boxer, 152 Leonberger, 46 Rhodesian Ridgeback and 58 Standard Poodle dogs, genotyped with Illumina CanineHD Whole-Genome Genotyping BeadChip. Within the data quality control, SNPs and individuals with more than 10% missing data and SNPs with less than 5% minor allele frequency were removed. After quality control 162,879 SNPs and 329 dogs remained for the analysis. A separate quality control was performed for the computation of genomic inbreeding coefficients, without limitation on minor allele frequency, resulting in 207,029 SNPs kept for the analysis.

The overall data management and quality control was done with PLINK 1.9 (Chang *et al.*, 2015). The genomic inbreeding was computed based on runs of homozygosity (ROH), using the cgaTOH software (Zhang *et al.*, 2013). For the purposes of this analysis the ROH segments were determined based on minimal length of 8 Mb. In order to account for non-complete genotypes in such long segments, up to two missing, but no heterozygous SNPs were allowed.

$$F_{ROH} = \frac{\sum_k ROH_k}{L} \quad (1)$$

where $\sum_k ROH_k$ is the sum of the ROH length for each segment k, and L is the length of the autosomal genome covered by SNPs.

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_757, © G. Mészáros *et al.* 2022

PLINK was used to compute heterozygosity levels for each dog, based on non-missing genotypes. Comparisons between heterozygosity levels and genomic relationships within data subsets, and all visualizations were done with R.

Results

As the first step a principal component analysis was performed (figure not shown). Here the breeds were clearly distinguished, with the first two principal components explaining about 67% of the variability between breeds. The breed clusters were well separated, with the Rhodesian Ridgeback and the Standard Poodle breeds being the closest to each other.

The inbreeding levels in the four analysed populations ranged from 0 to 31.1%. While there was certain variation in inbreeding levels in all breeds (Figure 1), the Rhodesian Ridgeback and the Standard Poodle appeared to be less inbred, while the Boxer population was shown to be the most inbred (Table 1).

The overall heterozygosity levels for each breed, compared to the entire analysed data set (Figure 2A-D) show a similar picture as the inbreeding levels. The correlation coefficient between the heterozygosity and ROH based inbreeding was r=-0.68 for the entire data set. Surprisingly, there were considerable differences between breeds, with correlation coefficients of -0.25, -0.52, -0.52 and -0.83 for the Boxer, Leonberger, Rhodesian Ridgeback and Standard Poodle breeds, respectively.

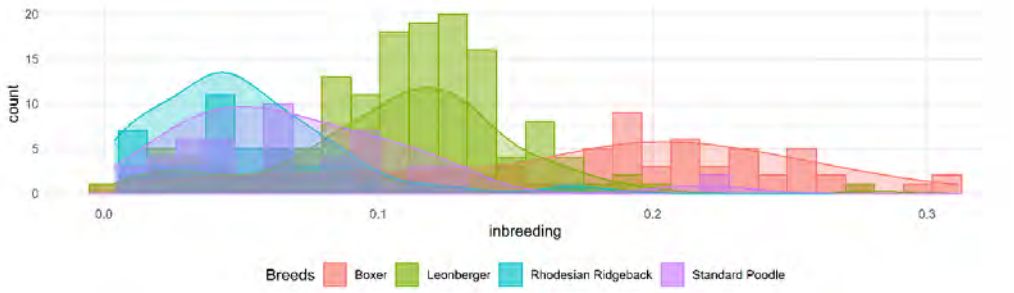


Figure 1. Distribution of inbreeding coefficients (up to 6 generations, ROH>8 Mb) in the selected populations. Bar plots show the number of dogs; the curves are the densities within the breed.

Table 1. Means and standard deviations (in brackets) for inbreeding, heterozygosity and genomic relationships[1].

	Boxer	Leonberger	Rhodesian Ridgeback	Standard Poodle
Inbreeding in all dogs	0.173 (0.0761)	0.109 (0.424)	0.049 (0.0329)	0.070 (0.0439)
Heterozygosity in all dogs	0.236 (0.0156)	0.271 (0.0124)	0.304 (0.0172)	0.316 (0.0190)
Genomic relationship in all dogs	0.229^{a} (0.0749)	0.161^{a} (0.0685)	0.092^{a} (0.0705)	0.046^{a} (0.0933)
Genomic relationship in dogs with heterozygosity > mean	0.355^{b} (0.0539)	0.269^{b} (0.0585)	0.173^{b} (0.0607)	0.092^{b} (0.1127)
Genomic relationship in dogs with heterozygosity > mean + 1sd	0.346^{b} (0.0506)	0.269^{b} (0.0635)	0.205^{b} (0.1542)	0.129^{c} (0.0771)

[1] For each column differences in alphabetic superscripts indicate significant differences.

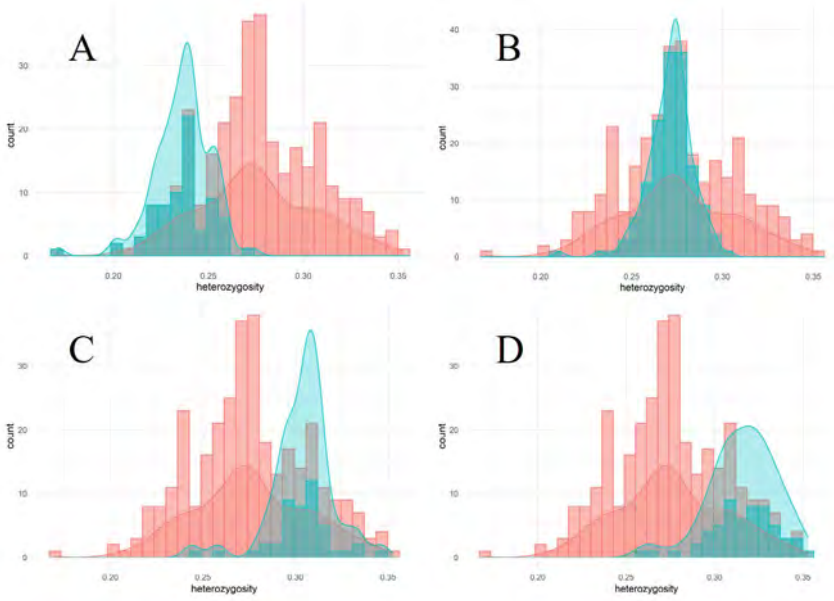


Figure 2. Heterozygosity levels in the Boxer (A), Leonberger (B), Rhodesian Ridgeback (C), and Standard Poodle (D) populations in green, compared to joint heterozygosity levels from all populations (red). Bar plots show the number of dogs; the curves are the densities within the breed.

It also came to our attention, however, in some cases, that the dog breeders use the simple heterozygosity levels in an unexpected way. Such unexpected use was, when the only criterion for selection of mating pairs was that the desired mating partner was from the more heterozygous part of the spectrum for the given population. Surprisingly, no further check on the (genomic) relatedness was performed. The hope of the breeders was that the offspring will be also highly heterozygous, as the parents. Of course, there is no guarantee for this to happen, indeed there is a real possibility for exactly the opposite outcome of more inbred offspring, with even lower heterozygosity.

In order to simulate such a scenario, we computed the average and standard deviation for heterozygosity and genomic relationship in the entire data set for all breeds (Table 1). In a follow-up we limited the data set to dogs with above average heterozygosity. We note that the genomic relationships were not recomputed between the scenarios. Also, the sex information was not considered for potential mate pairs, as the goal was a proof of principle, i.e. if the dogs with above average heterozygosity were more or less related as the average of the population.

Discussion

The inbreeding levels were computed based on the ROH methodology that considers long, continuously homozygous segments on the genome. It was assumed that the homozygosity status of SNPs along these segments with millions of nucleotides in length was not coincidental, but due to the inheritance of the same ancestral haplotypes from both the maternal and paternal side. Thus, the proportion of these ROH segments on the genome was the direct and exact measurement of inbreeding for each dog. Similarly, to other methods used to assess inbreeding levels, the number of generations of ancestors one wants to consider can be chosen using the ROH metric. The reason for considering such long segments of 8 Mb was that it corresponds to ancestry up to 6 generations in the past, similarly to the pedigree depth usually

considered by the breeders. There was a considerable variation in inbreeding levels in each breed, which was likely due to individual decisions of breeders. The differences in average inbreeding coefficients between breeds, however, were a likely consequence of breed management by breeding organizations.

The reason for the differences in heterozygosity levels (Figure 2) could be that while heterozygosity was considered on SNP-by-SNP basis, the inbreeding was measured with whole genome segments that must fulfil certain criteria (missingness and heterozygosity). Also, the overall heterozygosity, i.e. 1- homozygosity, also takes into account SNPs that are identical by state (IBS), and not just those which are IBD, and thus contribute to true inbreeding levels. As the IBS markers could be homozygous by pure coincidence, they might not properly characterize the differences between individuals. Thus, the overall heterozygosity might not be the most accurate measure of genomic diversity. Despite this slight limitation, and perhaps due to the strong correlation to overall inbreeding levels, the heterozygosity is being often used to assess diversity in practical dog breeding.

The simulation of the breeding practice of mating highly heterozygous dogs without considering their genomic relationship uncovered a flaw in this practice. Interestingly, in all breeds the genomic relationship was significantly higher in the restricted set than the population average (Table 1.), indicating that such selection of mate pairs would lead to on average higher inbreeding levels in offspring. The genomic relationship was similarly high when the populations were further restricted to the most heterozygous dogs, i.e. those with one standard deviation above the mean.

In conclusion the genetic diversity measured via genomic inbreeding and overall heterozygosity, differs considerably between breeds. It was also shown that a choice of mate pairs based purely on heterozygosity, without considering the genomic relationships was counter-productive.

References

Chang C.C., Chow C.C., Tellier L.C.A.M., Vattikuti S., Purcell S.M., Lee J.J. (2015) GigaScience, 4(1):s13742-015-0047-8. https://doi.org/10.1186/s13742-015-0047-8

Janeš M., Zorc M., Ferenčaković M., Curik I., Dovč P., Cubric-Curik V. (2021) Sustainability 13(4):2289. https://doi.org/10.3390/su13042289

Zhang L., Orloff M.S., Reber S., Li S.., Zhao Y, Eng C. (2013) PLoS ONE 8(3):e57772. https://doi.org/10.1371/journal.pone.0057772

758. Breed differences in the use of tools for dog breeding

M. Janes[1*], T.W. Lewis[2] J.J. Ilska[3], G. Gorjanc[1] and J.A. Woolliams[1]

[1]The Roslin Institute, Royal (Dick) School of Veterinary Studies, University of Edinburgh, Easter Bush, Midlothian, EH25 9, RG, Scotland, United Kingdom; [2]The Guide Dogs for the Blind Association, Hillfields, Burghfield Common, Reading, Berkshire, RG7 3YG, United Kingdom; [3]The Kennel Club, Clarges Street, London W1J 8AB, United Kingdom; mateja.janes@ed.ac.uk

Abstract

Inbreeding occurs when relatives are mated together, possibly decreasing fitness. Tools to assess the potential for inbreeding in a population from selection and mating decisions could provide more robust populations. The Kennel Club (KC) tool Mate Select provides support for making informed mating decisions by offering inbreeding information spanning 222 breeds. Information on usage was assessed from 2011 to 2018 in relation to dog registrations. Breeds with up to 70 dams had an increase from 2 to 10 searches per dam registered, while breeds with 20,000 registered dams declined towards 2 searches per dam. There was significant variation between breeds of the same census size, and breeds participating in recording schemes leading to the publication of EBVs had more searches per dam (P<0.001), however this was not directly related to the availability of the EBVs.

Introduction

Inbreeding has long been used in companion animals and livestock, and has been part of the process that has led to the establishment of breeds. Inbreeding inevitably increases in restricted populations, such as dog breeds, diminishing genetic variation within the breed (Leroy *et al.*, 2011). To track genetic variation of such populations, we commonly calculate coancestry coefficients between pairs of individuals in the population, average coancestry coefficients, as well as individual inbreeding coefficients. These coefficients have been utilised in many studies to characterise inbreeding and population structure within and across dog breeds (e.g. Karjalainen *et al.*, 1997). While these coefficients are informative, it is more important to manage the rate of increase in the coefficients (ΔF) to manage genetic diversity.

The concept and definition of inbreeding were developed by Wright (1931) and it is widely recognized that lower breed census sizes and selection intensity increase ΔF (Robertson 1961). To manage ΔF, breeders use methods such as optimum contribution selection (Brisbane *et al.* 1995), which is well-established in leading animal breeding programmes (Howard *et al.*, 2017). However in dog populations this technology is more challenging due to a dispersed population and many individual breeders. This structure leads to demands for closer collaborative interactions for breeders and their breed and kennel groups that are often unmet (Fernández *et al.*, 2017). As a step towards managing inbreeding in UK dog breeds by promoting better breeding decisions KC created the Mate Select tool in collaboration with the Animal Health Trust. This provides inbreeding coefficients on individuals in the pedigree and on putative offspring of proposed matings, together with essential KC registration information. Further data comprised DNA testing for deleterious mutations, phenotypic results of health screening (such as eye tests and hip scores), and, when available, Estimated Breeding Values (EBVs) for complex traits like hip and elbow dysplasia (Lewis *et al.*, 2013). As a result, breeders may reduce propagation of lethal recessive alleles, improve complex traits, and avoid matings between close relatives. Since the impact of any tool will depend strongly on its usage, the aim of this contribution is to analyse the usage of Mate Select in UK dog breeding.

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_758, © M. Janes *et al.* 2022

Materials & methods

Data. The information for this paper was obtained from two databases. The first database was an anonymised log of KC's Mate Select searches from May 2011 to September 2018. In this context, searches are requests for information about potential matings of two dogs. During this period, 946,632 searches were made for 213 breeds. There was also information about the user's regional location for each search. The second database was a pedigree for 10,182,582 dogs from 224 breeds. This database also included phenotypes such as coat colour, elbow and hip scores (where accessible), and sex. A total of 1,857,384 dogs were registered during the period covered by Mate Select. A unique KC dog identification number was used to link the two databases. During the study period, EBVs for dysplasia were reported, supplementing the documented phenotypes. These breeds are referred to as 'EBV-track' breeds.

Data analysis. To characterise the extent and mode of use over time, several aspects were analysed and compared. In this paper, we focus on the evaluation of disparities in the utilisation of Mate Select by breed, as well as the relationship between breed popularity and the availability of EBVs.

Searches in relation to breed. The number of searches per breed in Mate Select was compared to the breed popularity as determined by the database of pedigree registrations. The number of Mate Select searches per breed was calculated over the study period. The number of active dams was defined as those appearing in the registration of offspring across the period covered by Mate Select data. The number of dams per breed was then compared to the number of searches per breed over the same period. Breeds with less than 10 active dams in the period were excluded. The EBV-track breed subset was also considered as a breed classifier. We fitted the following mixed linear model was fitted using ASReml-R (VSN International, UK):

$$y = 1m + xb + f(x) + zc + e \quad (1)$$

where y is the base 10 logarithm of total searches for a breed; 1 is a vector of 1's, m is the population mean, x is the base 10 logarithm of the number of dams registered, b the associated regression coefficient, f(x) the smoothing spline fitted within the mixed model for x, z a binary indicator variable whether the breed had EBVs published by KC at any point in the study period, c its effect, and e is a residual vector. Since the model is on a logarithmic scale, Equation 1 can be transformed into a model for the ratio of searches to registered dams, by subtracting x from both sides. The convex hull was used to identify outlier breeds in our dataset, and is the subset of breeds that define the smallest convex polygon enclosing all the breeds in the data.

Results

The results analyse how breeders use Mate Select's 'kinship' data, which may be retrieved for any potential mating between two registered dogs. This includes information on the inbreeding coefficient of the bitch and the dog, as well as their potential offspring. In the following, a search is defined as a request for information, and a bitch is defined as a registered female with recorded offspring.

Figure 1 illustrates the association between Mate Select searches expressed as a multiple of the number of registered dams per breed and the log of the total number of registered dams for the 197 breeds with at least 10 dams registered with the KC throughout the study period. The predicted number of searches increases from 2.2 per dam for breeds with 10 dam census to 10.5 per dam for breeds with 100 dam census, then decreases to 3.5 and 2.4 respectively for 1000 and 10,000 dam censuses. The trend line in Figure 1 peaks at 11.5 searches per dam for breeds with 70 dams. For the extreme census size of the Labrador Retriever the search rate was 4.32 searches per dam. The red line shows the trend obtained from the mixed-model

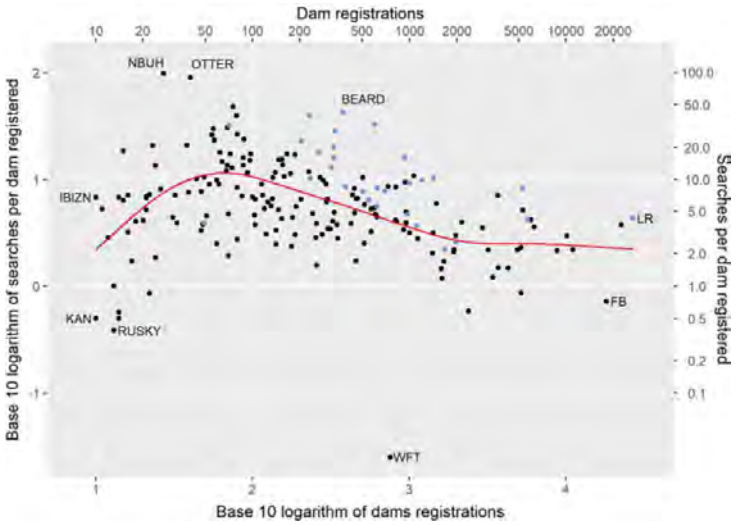


Figure 1. The relationship of frequency of searches per breed with dams registered.

Table 1. The breeds forming the convex hull of the distribution of search frequency and breed size, with the tag used to identify them in Figure 1 and the text.

Breed	Tag	Registered dams	Searches in mate select	Searches/dam	EBV-track
Otterhound	OTTER	40	3,572	89.3	No
Norwegian Buhund	NBUH	27	2,650	98.15	No
Ibizian Hound	IBIZN	10	5	0.5	No
Turkish Kangal Dog	KAN	10	5	0.5	No
Russian Toy	RUSKY	13	5	0.38	No
White Fox Terrier	WFT	757	19	0.03	No
French Bulldog	FB	18.119	13.105	0.72	No
Labrador Retriever	LR	26.723	115.451	4.32	Yes
Bearded Collie	BEARD	378	15.907	42.08	Yes

smoothing spline fitted with Equation 1. The points shown in light blue are the EBV-track breeds and these breeds have more searches per dam ($P<0.001$). The labelled breeds in Figure 1 are the 9 breeds that form the convex hull and Table 1 identifies these breeds with their labels giving further information. These breeds are on the 'edge' of the logarithmic distribution of breed size and searches per breed. For example, LR and IBIZN had a higher number of searches per dam whereas KAN, WFT and FB are extreme in having a relatively low number of searches given their census sizes, with WFT being particularly extreme having 40 times more registered dams than searches. In contrast, the NBUH had 98 times more searches than registered dams. If WFT is excluded then it is replaced in the hull by the Yorkshire Terrier.

Discussion

Inbreeding was originally used in formation of breeds, but increasingly threatens populations' genetic health. Dog breeders can use tools like KC's Mate Select to manage breeding decisions. The tool leverages data from 222 breeds with over 230,000 yearly registrations overall. Although there was large variation between breeds, Mate Select was well used by all KC recognized breeds for the given period. In principle,

more searches per dam for smaller breeds would be better for genetic management, however this was not evident for the lowest census breeds of the study. There are several reasons to doubt this expectation. First, a breed with few dams and few dogs available, restricts mating options. Second, dams may be owned by a small number of breeders, thus limiting information about accessible dams. Thirdly, it is likely the UK populations for a number of these breeds is a subset of larger (but still small) global breed populations that separately keep their own disjoint pedigree databases. Beyond the numerically smallest breeds, smaller breeds showed more searches with a peak of 70 registered dams (Figure 1) and a fall as breed populations grew larger.

However, among breeds of similar size, there is considerable variation due on breed culture or structure. The reasons for the Wire Fox Terrier's extremely low search frequency are unknown, while the high search frequency for the Norwegian Buhund coincides with the publication of a DNA test for the cerebellar ataxia that affects this breed (Jenkins *et al.*, 2020). While a multitude of DNA testing for diverse breeds have emerged in recent decades, no additional extreme search frequency coincidences have been observed. The finding concerning EBV-track breeds raised the hypothesis that providing EBVs may have prompted more searches per dam. This hypothesis was testable as EBVs for a subset were launched during the study period. Disappointingly, subsequent analysis (not shown) showed no differences in search frequency before and after the launch of EBVs, and so leaves open the question of what aspect of breed culture promotes the greater use of tools such as Mate Select.

References

Brisbane J.R., Gibson J.P. (1995) Theor Appl Genet. 91:421–31.

Fernández J., Bennewitz J. (2017) Chapter 2. Defining genetic diversity based on genomic tools; 2017. p. 49–76. https://doi.org/10.3920/978-90-8686-850-6_2

Howard J.T., Tiezzi F., Huang Y., Gray K.A., Maltecca C. (2017) J Anim Sci. 95:4318–32.

Jenkins C.A., Kalmar L., Matiasek K., Mari L., Kyöstilä K., *et al.* (2020) PLoS Genetics 16(1):1-20. https://doi.org/10.1371/journal.pgen.1008527

Karjalainen L., Ojala M. (1997) J Anim Breed Genet.; 114(6):33–41. https://doi.org/10.1111/j.1439-0388.1997.tb00489.x

Leroy G., Baumung R. (2011) Anim Genet. 42(1):66–74.https://doi.org/10.1111/j.1365-2052.2010.02079.x

Lewis T., Blott S.C., Woolliams J.A. (2013) BMC Genet. 14(16):1–11. https://doi.org/10.1186/1471-2156-14-16

Robertson A. (1961) Genet Res. 2(2):189–94. https://doi.org/10.1017/S0016672300000690

Wright S. (1931) Genetics. 16(2):97–159.

759. Inbreeding and litter size in Dutch pedigreed dogs

J.J. Windig[1,2*], M.L. Margarita[1] and H.P. Doekes[1]

[1]ABG, Wageningen UR, P.O. Box 338, 6700 AH Wageningen, the Netherlands; [2]Centre for Genetic Resources the Netherlands, P.O. Box 338, 6700 AH Wageningen, the Netherlands; jack.windig@wur.nl

Abstract

Inbreeding may negatively affect fertility, especially when mating closely related individuals. We evaluated the frequency of these matings in Dutch pedigreed dogs between 1997 to 2018 and their effect on litter size in 15 breeds. The percentage of matings between half-cousins and closer has decreased since 1999. Litter size of individual matings tended to be significantly lower only for full sib and parent offspring matings. Inbreeding depression was only evident when analysed at the population level. In two breeds, a 1% increase in inbreeding of the dam resulted in an estimated reduction of litter size of 0.05 and 0.028 pups. These findings emphasize the importance for breeders and owners to avoid mating closely related dogs.

Introduction

Inbreeding is the result of mating related animals. Inbreeding can lead to the expression of genetic defects and inbreeding depression resulting in, e.g. a lower fertility (Falconer and Mackay 1996; Doekes *et al.* 2021). In many pedigreed dog breeds, inbreeding rates have been >0.5% per generation (Lewis *et al.*, 2015). To counteract the increase in inbreeding, mating between close relatives such as parent – offspring mating has been banned in breeds and kennel clubs. This does, however, not prevent inbreeding. Since pedigreed dogs form (small) closed populations high inbreeding rates may still occur and lead to inbreeding depression.

The effects of inbreeding may be (partly) compensated by natural and artificial selection. The heritability of fertility traits in dog populations tends to be low to moderate, i.e. in the range of 0.05 to 0.30 (Hare and Leighton, 2005; Leroy *et al.*, 2015). A decrease in fertility due to an increase in inbreeding may not be observed due to simultaneous selection for better fertility. Only when inbreeding rates are high, genetic drift is the main driver of allele frequency changes and selection cannot counteract the decrease in fertility over time due to inbreeding.

Pedigreed dogs provide an opportunity to investigate the effect of inbreeding on fertility. Litter sizes and relationships can be readily determined from pedigrees. Here, we used national pedigree data from the Dutch Kennel club to: (1) evaluate the frequency of close relative matings in the Netherlands from 1999 to 2018; (2) study the effect of these matings on litter size in 15 breeds; and (3) determine the amount of inbreeding depression in two of the breeds while accounting for a detailed pedigree.

Materials & methods

Data and classification of close relative matings. A data base was provided by the Dutch Kennel club. This data-base contained breed, birth date and father and mother of all registered pups of pedigreed dogs born in the Netherlands between 1997 and 2018. The following close relative matings were determined, with in brackets the corresponding relationship coefficients: parent – offspring (0.5), full sibs (0.5), half sibs (0.25), grandparent – grandchild (0.25), full uncle – niece or full aunt – nephew (0.25), half uncle – niece or half aunt – nephew (0.125), full cousins (0.125), and half cousins (0.0625). The first three could be determined only when all 4 grandparents were known, which was the case for 917,465 pups born from 1999 onwards. The other relationships could be determined if all 8 great-grandparents were known, which was the case for 103,785 pups born from 2004 onwards.

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_759, © J.J. Windig *et al.* 2022

The effect of close relative matings on litter size. Close relative matings were grouped according to the relationship coefficient, resulting in 5 categories: no close relationship, 0.0625, 0.125, 0.25 and 0.5. Litters and litter sizes were determined by grouping pups with the same parents born at the same date. Only pups from the 13 most numerous breeds containing 47% of all registered dogs were analysed. This ensured that there were at least 20 animals in each relationship category, year and parity. The effects of relationship category, parity, year and interactions on litter size were determined in a 3-way ANOVA. The 13 breeds were supplemented with the Saarloos Wolfhond and Schapendoes breeds for which the breeding organisations provided detailed records over a longer period.

Inbreeding depression for litter size at population level in two breeds. The amount of maternal inbreeding depression for litter size was estimated for the databases of the Saarloos Wolfdog and Schapendoes provided by the breed clubs. These databases contained ancestry up to the start of the breeds in the 1940s. Only litters since the 1990s were used for the analysis, because before the 1990s sex ratios were skewed, suggesting that before the 1990s not all pups were registered but rather those that were later used for breeding. The following model was run with ASReml (Gilmour *et al.*, 2015):

$$LS_{ij} = u + Par_j + bF_i + A_i + Pe_i + e_{ij} \quad (1)$$

with LS_{ij} being the number of registered pups (i.e. alive at 2 months of age) for the j^{th} litter of dam i, Par_j the parity of the litter, b the coefficient of regression of LS on the inbreeding coefficient F_i of the dam i, Ai the random genetic effect of dam i, Pe_i the random permanent environmental effect of the dam i across all her litters and e_{ij} the random residual.

Results

The frequency of close relative matings decreased over the years (Table 1). Father-daughter, Mother-son, Full sib and Grandparent – grandchild matings have not occurred since 2011.

Litter size was smaller in small-sized breeds, e.g. the Chihuahua with a mean of 2.7 pups, than in large-sized breeds, e.g. the Golden Retriever with a mean of 6.6 pups (Table 2). Differences between years were significant for 11 of the 15 breeds (Table 2), but only for the Saarloos Wolfdog there appeared to be a trend, with litter size decreasing from above 6 before the 1990s to 3.5 in 2008 after which it recovered to close to 5 pups.

Differences in litter sizes over all parities were significant in all breeds except the Saarloos Wolfdog. When compared to the mean litter size within breeds, litter size was on average 6% lower for first parity litters and 8% lower for parities above 4 (Figure 1). Differences in litter size between relationship categories were

Table 1. Percentage pups from close relative matings for all pedigreed pups registered in the Netherlands with known grandparents (1999 onwards) and with known great-grandparents (2004 onwards).

Years	Half Cousins	Full Cousins	Half uncle Niece	Full uncle Niece	Half Sibs	Grand parent g.child	Full Sibs	Parent Child
1999-2003	NA[1]	NA[1]	NA[1]	NA[1]	1.54%	NA[1]	0.17%	0.33%
2004-2008	11.03%	7.09%	9.07%	1.46%	1.38%	0.22%	0.09%	0.27%
2009-2013	8.05%	3.93%	7.19%	1.02%	0.83%	0.15%	0.03%	0.05%
2014-2018	6.95%	2.95%	6.17%	0.71%	0.60%	0.00%	0.00%	0.00%

[1] Not enough litters with great-grandparents known to determine percentage of matings.

Table 2. Analysed breeds, number of litters analysed, mean litter size and *P*-values for the effect of year, parity and relationship category on litter size in 3-way ANOVA.

Bree[d1]	#Litter	Litter size (S.D)	Year	Parity	Relationship	Interactions[2]
LAB	13,808	6.3 (2.8)	0.000	0.000	0.053	
GSD	12,767	5.0 (2.8)	0.000	0.000	0.045	Y×P
DAH	12,742	3.9 (2.0)	0.000	0.000	0.000	P×R
GOR	7,486	6.6 (2.9)	0.004	0.000	0.000	
BMD	6,235	5.6 (2.9)	0.001	0.000	0.213	
BOX	5,327	4.9 (2.4)	0.018	0.000	0.502	
CKCS	5,638	3.7 (1.7)	0.001	0.000	0.820	
BCL	3,985	4.6 (2.6)	0.104	0.000	0.013	
BDF	3,076	5.9 (2.9)	0.067	0.000	0.102	
FBD	4,497	4.0 (2.2)	0.058	0.000	0.163	
CHI	6,629	2.7 (1.4)	0.000	0.000	0.048	
EBD	3,869	4.5 (2.5)	0.000	0.000	0.451	
ECS	3,703	4.6 (2.3)	0.161	0.000	0.030	Y×P, P×R
SWD	699	4.3 (2.6)	0.000	0.068	0.444	
SDS	4,643	5.9 (2.3)	0.000	0.000	0.000	

[1] LAB, Labrador Retriever; GSD, German Shepherd Dog; DAH, Dachshund; GOR, Golden Retriever; BMD, Bernese Mountain Dog; BOX, Boxer; CKCS, Cavalier King Charles Spaniel; BCL, Border Collie; BDF, Bouvier des Flandres; FBD, French Bulldog; CHI, Chihuahua; EBD, English Bulldog; ECS, English Cocker Spaniel; SWD, Saarloos Wolfdog; SDS, Schapendoes.

[2] Only significant interactions ($P<0.05$) are indicated.

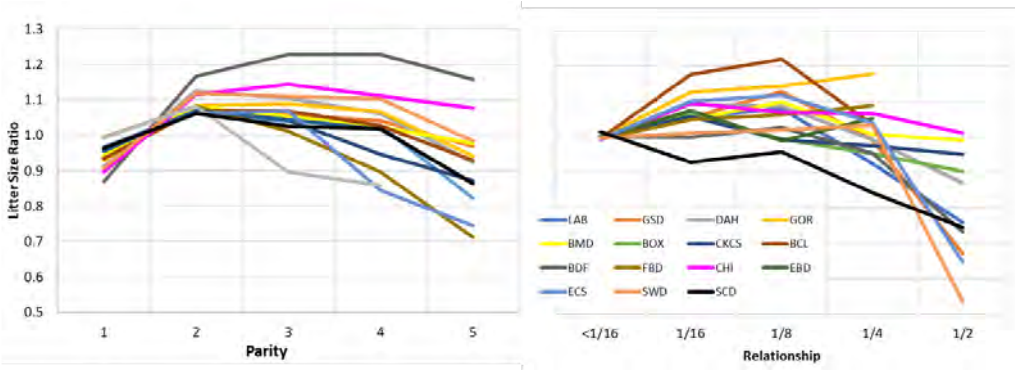


Figure 1. Litter size ratio (average within group litter size over breed average) over parity and relationship category. Breed abbreviations see Table 2.

significant for 7 of the 15 breeds. Litter size was on average 16% lower in the highest relationship category (Figure 1). Litter sizes were on average 8% higher for relationship category 0.0625 and 0.125.

Estimated heritabilities for litter size in the Saarloos Wolfdog and Schapendoes were moderate with 33 and 24% (Table 3). A 1% increase in the inbreeding coefficient of the mother resulted in an estimated reduction of litter size of 0.05 and 0.028 pups. This was equivalent to 1.16 and 0.47% of the trait mean, and 1.90 and 1.18% of the standard deviation.

Table 3. Genetic parameters (SE), inbreeding depression and *P*-value of effect maternal inbreeding for litter size in the Saarloos Wolfdog and Schapendoes.

Breed	V_A [1]	V_{PE} [2]	V_R [3]	h^2	+1% F [4]	p
Saarloos Wolfdog	2.16 (0.56)	0.47 (0.46)	3.96 (0.32)	0.327 (0.074)	-0.050	0.035
Schapendoes	1.29 (0.18)	0.56 (0.15)	3.61 (0.10)	0.236 (0.031)	-0.028	0.011

[1] Additive genetic variance.

[2] Permanent environmental variance.

[3] Residual variance, h^2 = heritability.

[4] Change in number of pups per 1% increase of inbreeding coefficient.

Discussion

The number of matings between closely related dogs clearly decreased over the years. This is probably due to increased awareness of the detrimental effects of inbreeding and in line with observations that the rate of inbreeding in most dog breeds have gone down (e.g. Lewis *et al.* 2015). The negative effect of inbreeding on litter size in individual matings was only evident for the highest relationship category (matings between full sibs or parent offspring), and only significant for about half of the breeds. For lower relatedness categories (0.0625 to 0.125) litter size was even above the breed's average for most breeds. Various studies have previously reported (insignificant) favourable effects of inbreeding in animal populations (e.g. Doekes *et al.*, 2021). Apparently, the effect of inbreeding depression, when present, is so small that it could be masked by other factors. When the effect of maternal inbreeding was analysed at the population level it was highly significant for both breeds. Leroy *et al.* (2015) found clear inbreeding depression as well, with similar analyses as in this paper in seven breeds, and of similar size (0.007 to 0.042 per 1% F) as we found (0.028 and 0.050). One advantage of genetic analysis at the population level is that genetic effects other than inbreeding can be taken into account as well. Selective breeding, e.g. mating dogs from large litters to each other, may for example lead to relatively large litters for more inbred litters. For breeders and owners that have to select mates for their dogs, it is thus important to avoid mating closely related dogs. The effect of mating less related dogs is less obvious. Genetic management of breeds is, however, important to prevent a permanent reduction in litter size due to fixation of genetic defects with small effect due to genetic drift, i.e. an imbalance between inbreeding rate and selection/purging. The reduction in litter size as observed in the Saarloos Wolfhond may well have been the result of its high inbreeding rate over a long time.

Acknowledgements

We would like to thank the Dutch Kennel Club (RvB, Laura Roest) for providing the national database, the ALVS (Saarloos Wolfhond) and VNS (Schapendoes) for providing pedigrees, and Mila Stigter for analysing number of matings between closely related individuals.

References

Falconer D.S., and Mackay T.F.C. (1996) Introduction to quantitative genetics. Longman Group, Harlow, UK

Doekes H.P., Bijma P., and Windig J.J. (2021) Genes 12, 21. https://doi.org/10.3390/genes12060926

Gilmour A. R., Gogel B. J., Cullis B. R., Welham S., and Thompson R. (2015). ASReml user guide release 4.1 structural specification. Hemel hempstead: VSN international ltd.

Leroy G., Phocas F., Hedan B., Verrier E., and Rognon X. (2015) Veterinary Journal 203, 74-78. https://doi.org/10.1016/j.tvjl.2014.11.008

Lewis T.W., Abhayaratne1 B.M., and Blott S.C. (2015) Canine Genetics and Epidemiology 2:13. https://doi.org/10.1186/s40575-015-0027-4

Part 42

Equine

760. Genetic analysis of three horse breeds of Greece

A. Kominakis[1], N. Kritikos[2], A. Simpson[2], T. Sdourlias[3], E. Iliopoulou[4], V. Karagianni[5], D. Pappas[6] and A.L. Hager-Theodorides[1*]

[1]Department of Animal Science, Agricultural University of Athens, Iera Odos 75, 11855 Athens, Greece; [2]Skyrian Horse Society, 34007 Skyros, Greece; [3]Thessaly Horse Breeders' Association, Lampi 2, Agria, 37300 Volos, Greece, [4]Directorate of Animal Genetic Resources and Veterinary Infrastructure, Ministry of Rural Development and Food, Veranzerou 46, 10438 Athens, Greece; [5]Centre of Animal Genetic Improvement of Athens, Ministry of Rural Development and Food, Veranzerou 46, 10438 Athens, Greece; [6]Centre of Animal Genetic Improvement of Ioannina, Ministry of Rural Development and Food, 45332 Ioannina, Greece; a.hager@aua.gr

Abstract

The present study aimed to provide insights into the genetic diversity, structure and relationships of three rare horse breeds (Pindos, Skyros and Thessaly) of Greece. To this end, we estimated measures of genetic diversity and conducted admixture and principal components analysis using equine microsatellite genotypic data on 475 animals. All populations displayed high levels of genetic variability, including the Skyrian horse that has the lowest census size and is bred in isolation. Present results confirmed the genetic differentiation of the Skyrian horse from the other two populations and the genetic proximity of the Pindos and Thessaly horses.

Introduction

In Greece, seven native horse breeds have been recognized so far: Crete, Pinias, Andravidas, Thessaly, Pindos, Zakynthos and Skyros (Zafrakas, 1991, Alifakiotis, 2000). Of these, the latter is native to the homonymous Aegean island and presents a unique phenotype with lowest withers height (in the range from 95 to 112 cm). Because of its distinct phenotype and geographic isolation, the breed has been hypothesized to be genetically distinct from mainland horse breeds (Zafrakas, 1991). This hypothesis was confirmed by Bömcke *et al.* (2011) showing the distant genetic relationship of the Skyrian horse to any other Greek and international horse or pony breed, including small sized (e.g. Exmoor, Shetland and Welsh) breeds and the Przewalski horse (feral horse). Previous reports (Bömcke *et al.*, 2011, Kominakis *et al.* 2013) have shown that the Skyrian horse displays exceptionally high levels of genetic diversity, despite its low census size (n~250). On the Thessalian plains of central Greece, the homonymous horse breed is raised mounting up to 450 animals. The famous horse of Alexander the Great, named 'Bucephalus' (meaning 'ox-head') was a Thessalian horse. It is a light riding and draft breed with withers height ranging from 135 to 145 cm. Finally, the Pindos horse is native to the Pindos mountain range spanning from West Macedonia to Thessaly and Epirus with a census population size of about 500 animals and withers height ranging from 110 to 135 cm. Like most other Greek horse breeds, this breed is also sure-footed, adaptable and hardy, being able to thrive in harsh environments. The present study aimed to provide insights into the genetic diversity, structure and relationships between and island horse breed, Skyrian, and two of the most populous breeds of the Greek mainland, Thessaly and Pindos horses.

Materials & methods

Animal sampling and data collection. The initial data set included records from a total number of n=687 animals registered with the respective Skyrian, Thessaly or Pindos horse breeding associations. Of these, a subset of n=475 records with complete information on origin, sex, place and year of birth and genotypes of microsatellite markers were used for further analysis. The total numbers of complete records per breed were n=242 (130 females, 112 males) for Skyrian, n=211 (164 females, 47 males) for Thessaly and n=22 (17 females, 5 males) for Pindos horses. Animals were genotyped using DNA extracted from hair samples

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_760, © A. Kominakis *et al.* 2022

and the StockMarks™ for Horses 17-Plex Genotyping Kit (Applied Biosystems, Thermo Fischer Scientific) that included 17 short tandem repeat (STR) markers: AHT4, AHT5, ASB2, ASB17, ASB23, CA425, HMS1, HMS2, HMS3, HMS6, HMS7, HTG4, HTG6, HTG7, HTG10, LEX3 and VHL20. This STR panel included the 12 markers of the official ISAG panel recommended for parentage analysis (AHT4, AHT5, ASB2, ASB17, ASB23, HMS2, HMS3, HMS6, HMS7, HTG4, HTG10, and VHL20). The PCR amplicons were analyzed on an ABI 310 Genetic Analyzer and to extract STR allele sizes, raw data fragment analysis was carried out using the GeneMarker® software (Softgenetics®, Pennsylvania, USA).

Genetic analysis. Estimates of the number of alleles per locus (n_A) along with expected (He) and observed (Ho) heterozygosity rates were obtained with the *Arlequin* software (ver. 3.5.2.2). We tested for Hardy-Weinberg equilibrium (HWE) or disequilibrium (HWD) across all STR markers and heterozygote deficit using exact U score tests and the online version of *GenePop* ver.4.7 software. The same software was used to test for pairwise allelic and genotypic differentiation between populations using exact G tests. Fis estimates were obtained with *GENETIX* (ver. 4.05.2). Admixture analysis was conducted using the *STRUCTURE* software (ver. 2.3.4), assuming K (ancestral populations) in the range from K=1 to K=3. Each time, we run 500K for burn in and 750K for 10 rounds of MCMC at each of K=1 to K=3, and the optimum K value was assessed through the analysis of the LnP(D) distribution plots and delta K values from *Structure Harvester.* Bar plots were constructed with *CLUMPP*, to align the replicates, and *distruct*. Finally, we performed Principal Components Analysis (PCA) of genotypic data using the *GENETIX* program and plotted the first two PCs using *SAS* (ver. 4.2).

Results

Measures of genetic diversity. Heterozygosity was highest in the Pindos and lowest in the Skyrian horse populations. The highest mean number of alleles per locus (n_A) was 8.82 and the lowest 6.65 and were found in the Thessaly and Skyrian populations, respectively (Table 1).

While Pindos population was in HWE (P=0.325, U global test), the remaining two populations were in HWD (P<0.001, U global test) due to heterozygote deficit (Fis=0.019 and 0.109, respectively). Furthermore, the three populations differentiated from each other, both at the allelic and genotypic level (exact G tests, results not shown). The optimum value for ancestral populations (K) was 2 (Skyros vs Pindos and Thessaly) (Figure 1). For K=3, the Pindos (orange color) and the Skyros (blue color) populations were still presented as distinct genetic entities while individuals of the Thessaly breed either differentiated (dark blue color) or were close or shared ancestry proportions with the Pindos horse (Figure 1, K=3).

PCA showed a distinct cluster including the Skyrian horses, while Thessaly and Pindos horses overlapped (Figure 2). These results were consistent with the admixture analysis, suggesting a higher degree of differentiation of the Skyrian horse breed and a shared genetic background for Pindos and Thessaly populations.

Table 1. Means and standard deviations (in parentheses) of measures of genetic diversity in three rare horse breeds of Greece.

Breed	n_A	He	Ho	Fis (95% CI)
Pindos	7.647 (2.344)	0.784 (0.091)	0.787 (0.090)	-0.0035 (-0.0711-0.0112)
Skyros	6.648 (1.579)	0.640 (0.172)	0.630 (0.177)	0.0198 (-0.0014-0.0392)
Thessaly	8.824 (2.789)	0.767 (0.086)	0.683 (0.121)	0.1099 (0.0865-0.1281)

[1] n_A = number of alleles per locus; He = expected heterozygosity; Ho = observed heterozygosity; Fis = Fisher's fixation index.

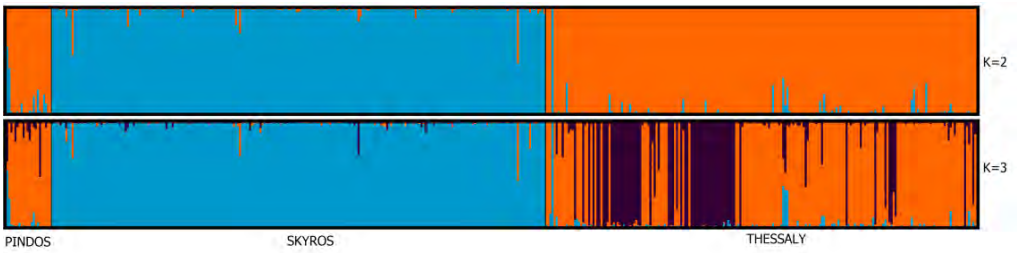


Figure 1. Ancestry proportions of three horse breeds of Greece. The x-axis represents individuals grouped according to breed. Each colour represents a population and each individual is represented by a vertical line partitioned into coloured segments, with segment lengths representing the admixture proportions from K=2 to K=3 ancestral populations.

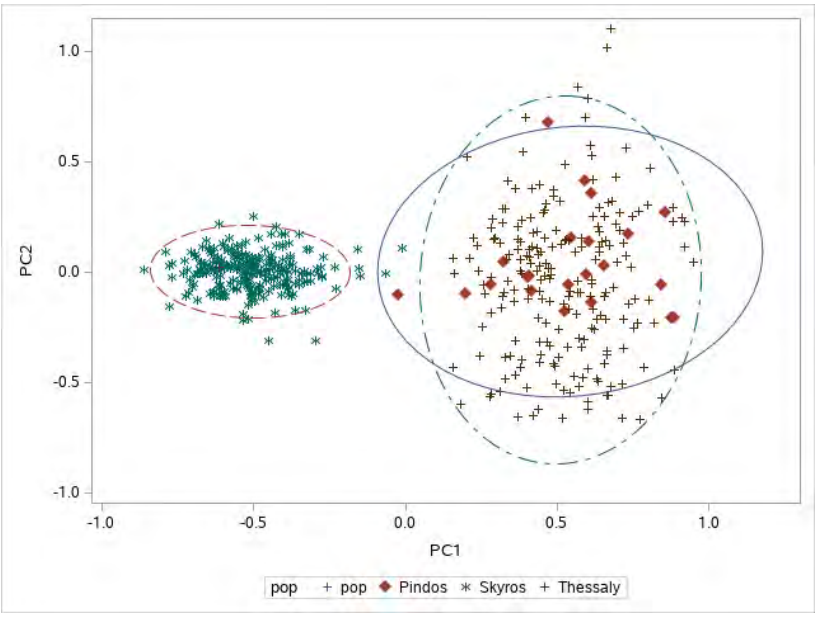


Figure 2. PCA based on genotypic data of 475 horses belonging to Skyros, Thessaly or Pindos rare breeds of Greece. PC1 and PC2 explained 5.9 and 2.62% of the variance of genotypic data, respectively.

Pairwise estimates of the genetic differentiation measure, i.e. F_{ST}, further confirmed the previous findings showing highest genetic differentiation between Skyros – Pindos (F_{ST}=0.113) and Skyros – Thessaly (F_{ST}=0.104) and genetic closeness between the Pindos and Thessaly horses (F_{ST}=0.0103).

Genetic diversity by sex and time. Estimates of genetic diversity by sex and year of birth for the Skyrian and Thessaly horses are shown in Table 2 and 3, respectively. These analyses were not performed for the Pindos horse due to limited number of available records. Interestingly, in both breeds considered contrasting estimates of the most distant time periods within breeds resulted in no statistically significant differences in either population (results not shown). In addition, there was no difference for the genetic diversity measures between males and females (results not shown).

Table 2. Means and standard deviations (underneath) of measures of genetic diversity by sex and year of birth in the Skyrian horse.

Time period	n_A	Ho	He	Sex	n_A	Ho	He
1983-2000 (n=54)	5.941	0.65	0.642	Males (n=112)	6.353	0.64	0.636
	1.749	0.19	0.178		1.498	0.193	0.176
2001-2010 (n=105)	6.118	0.622	0.641	Females (n=130)	6.235	0.618	0.636
	1.691	0.172	0.169		1.64	0.169	0.164
2011-2019 (n=83)	5.941	0.627	0.637				
	1.435	0.19	0.177				

[1] n_A = number of alleles per locus; He = expected heterozygosity; Ho = observed heterozygosity.

Table 3. Means and standard deviations (underneath) of measures of genetic diversity by sex and year of birth in the Thessalian horse.

Time period	n_A	Ho	He	Sex	n_A	Ho	He
1988-2000 (n=40)	7.765	0.69	0.776	Males (n=47)	7.824	0.664	0.768
	2.538	0.13	0.082		2.767	0.207	0.085
2001-2010 (n=128)	8.588	0.684	0.766	Females (n=164)	8.588	0.689	0.765
	2.575	0.117	0.088		2.476	0.114	0.088
2011-2014 (n=43)	7.118	0.674	0.754				
	2.369	0.098	0.098				

[1] n_A = number of alleles per locus; He = expected heterozygosity; Ho = observed heterozygosity.

Discussion

The present study provides estimates of measures of genetic diversity for the Pindos and the Thessaly horses for the first time, while validating previous findings for the Skyrian horse. Unexpectedly, all populations displayed levels of genetic variation comparable or higher than those reported for more populous breeds (e.g. Bjørnstad *et al.*, 2000; Van de Goor *et al.*, 2011). This was especially remarkable for the Skyrian horse that has the smallest census size and is bred in isolation for decades. Estimates on observed heterozygosity and numbers of alleles per locus in the three breeds included in this study are comparable to earlier reports on the Skyrian horse (Avdi and Banos, 2008; Bömcke *et al.* 2011; Kominakis *et al.* 2013) and to figures on other international breeds (e.g. Colling and Kelly, 1996; Bjørnstad *et al.*, 2000; Van de Goor *et al.*, 2011). The results presented here further support the genetic uniqueness of the Skyrian horse when compared to the other two breeds and revealed a close genetic relationship between the Pindos and the Thessaly horses. Findings of the present study provide scientific evidence to support the continuation of the *in situ* conservation programs for these indigenous horses of Greece.

References

Alifakiotis T.A. (2000) The Indigenous horse of Greece. Ministry of Agriculture, Athens, Greece.

Avdi M., Banos G. (2008) Livest. Sci. 114: 362–365. https://doi.org/10.1016/j.livsci.2007.11.007

Bjørnstad G., Gunby E., and Røed K.H. (2000) J. Anim. Breed. Genet. 117: 307–317. https://doi.org/10.1046/j.1439-0388.2000.00264.x

Bömcke E., Gengler N., and Cothran E.G. (2011). Genet. Mol. Biol. 34: 68–76. https://doi.org/10.1590/S1415-47572010005000113

Colling D., and Kelly S. (1996) Anim. Genet. 27 (Suppl. 2): 32. https://doi.org/10.1111/j.1365-2052.1996.tb01585.x

Kominakis A., Kritikos N., Dalaklidis Th., Karagianni V., Voulgari M., *et al.* (2013) Proceedings of the 4th Pan-Hellenic Congress in Technology of Animal Production, Thessaloniki, Greece.

Van de Goor L.H.P., van Haeringen W.A., and Lenstra J.A. (2011) Anim. Genet. 42: 627-633. https://doi.org/10.1111/j.1365-2052.2011.02194.x

Zafrakas A. (1991) The Equine and its Feeding. Kyriakidis, Thessaloniki, Greece.

761. Genetics of new conformation traits in Italian Heavy Draught Horse foals

R. Mantovani[1*#], E. Mancin[1#], B. Tuliozi[1], G. Gomez Proto[1], G. Pigozzi[2] and C. Sartori[1]

[1]University of Padova, Department of Agronomy, Food, Natural Resources, Animals & Environment, Viale dell'Universita' 16, 35020 Legnaro (PD), Italy; [2]Italian Heavy Draught Horse Breeders' Association, Via Verona 90, 37068 Vigasio, Italy; roberto.mantovani@unipd.it; [#]these authors contributed equally

Abstract

This study aimed to estimate genetic parameters for five novel conformation traits indirectly linked to dynamic aptitude (i.e. trot) and to body condition in Italian Heavy Draught Horse. Established under the European Agricultural Fund for Rural Development, these traits are now routinely recorded on young foals. Novel traits recorded in 2,283 foals (7,895 horses in pedigree) were analysed and correlated with two traits included in the selection index. Traits showed low to medium heritability levels (from 0.084 for croup direction to 0.213 for croup length), and positive genetic correlations (from 0.073 for body condition score vs croup direction to 0.901 for neck length vs neck direction). Medium genetic correlations were found with already established traits temperament (average r_g=0.500) and fleshiness (average r_g=0.588). Results suggest the possibility to successfully include the novel traits into the selection index to improve functionality for movement (through assessed neck and croup traits) and individual well-being (considering body condition score).

Introduction

In recent years, the Italian horse breeders' associations have focused on new breeding goals under the Equinbio breeding program, financed under the European Agricultural Fund for Rural Development by the Italian Ministry of Agriculture, Food and Forestry. Among the new breeding goals, new phenotypes have been identified and recorded with the aim to exploit a possible future use in selection. Particularly, the breeders' association of the Italian Heavy Draught Horse (IHDH) has focussed on five new phenotypes mainly addressing functional aspects relating to the dynamic aptitude (mainly trot) and the body condition. Data were recorded on young foals aged between 2 and 9 months. The IHDH, originally created for both military purposes and agricultural work, was converted toward the meat production during the 1980's, but today the breed is considered suitable for both meat production and heavy draught activities, like agricultural work or use in holiday farms. Additionally, the IHDH can exert even an important role in landscape management because of its easy adaptation to low-grade pastureland (Mancin *et al.*, 2020). The aim of this study was to assess the genetic structure of the five new phenotypes recorded on IHDH foals over a period of 5 years and particularly, the genetic correlations among these new phenotypes and with two further phenotypes already under selection, i.e. the temperament and fleshiness.

Materials & methods

The five new phenotypes recorded on young foals of IHDH breed were four body measurements, the neck length (NL), the neck direction (ND), the croup length (CL), and the croup direction (CD), and the body condition score (BCS). All these traits were obtained by a linear score using a 1-5 scale and half points (i.e. 9-point scale) by skilled classifiers. Data were recorded from 2017 to 2021 on a total number of 3,664 foals. In addition, temperament (TE) and fleshiness (FL) were considered among the traits routinely collected in the IHDH breed and included in the total selection index (described in Folla *et al.*, 2019). After editing, the final dataset used for genetic parameter estimation included 2,283 single-record data for young foals, accounting for 7,895 animals in the pedigree file up to the 8th generation. A previously validated model (Folla *et al.*, 2019) included the random stud-year-classifier (SYC) as single or groups of SYC in case of single observation assembled by geographical area, type of housing, production target, vaccination of foals

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_761., © R. Mantovani *et al.* 2022

and stud's mares mean body condition to identify nutritional management). Further factors accounted were the fixed effects of sex, age at evaluation (months, 9 classes), and, as surrogate of parity number, the age of the mare at foal's scoring (5 classes: <4 yrs.,…,≥14 yrs.). Univariate models were run to estimate variance components of each trait, and bivariate models allowed to estimate covariances and derive genetic correlations among traits. All analyses used AIREMLf90 (Misztal, 2008). Approximate standard errors were calculated using the Monte-Carlo technique of Houle and Meyer (2015).

Results

Descriptive statistics of traits are reported in Table 1. Averages of traits ranged from 2.81 (BCS) to 3.29 points (ND). Apart from BCS, all traits were right skewed. A low to medium heritability (h^2) was found for the five traits, ranging from h^2=0.084 (SE=0.038) for CD to h^2=0.213 (SE=0.047) for CL (Table 2). Medium to high genetic correlations (Table 3) were found across the four conformation traits for neck and croup (average r_g=0.600) with minimum for ND vs CD (r_g=0.07) and maximum for NL vs ND (r_g=0.901). Phenotypic correlations ranged from r_p=0.07 (NL vs CD) to r_p=0.717 (NL vs ND). BCS showed genetic correlations ranging from r_g=0.171 (with NL) to r_g=0.322 (with CL) with the other four novel traits, and phenotypic correlations from r_p=0.073 (with CD) to r_p=0.154 (with CL). The established traits temperament and fleshiness, already included in the selection index of the breed, showed medium correlations with the novel traits (average r_g=0.500 and 0.588, respectively; Table 4). The highest r_g estimates were found for ND vs both TE (r_g=0.586) and FL (r_g=0.601).

Discussion

Breeding for conformation involves aspects of morphology and functionality, and it is important for breeding decisions in livestock breeds. Traits may be measured or scored (in points), through a subjective evaluation based on valuating scoring or even by linear type evaluation (Koenen *et al.*, 1995). Current genetic evaluation of IHDH relies on a linear evaluation for 11 type traits scored in foals, and five of these traits are included in the total selection index (Folla *et al.*, 2019). In recent years, breeding programs financed under the European Agricultural Fund for Rural Development have led a wave of novelties in

Table 1. Descriptive statistics of the five novel linear traits considered in the study.

Trait	Descriptor range		Mean (SD)	Skewness	Kurtosis
	Min. (score 1)	Max. (score 5)			
Neck length (NL)	Short	Long	3.26 (0.48)	0.43	0.02
Neck direction (ND)	Horizontal	Vertical	3.29 (0.44)	0.66	0.19
Croup length (CL)	Short	Long	3.19 (0.38)	0.86	1.11
Croup direction (CD)	Horizontal	Sloping	3.02 (0.31)	0.55	5.04
Body condition score (BCS)	Scrawny	Fleshy	2.81 (0.56)	-1.07	2.28

Table 2. Additive genetic (σ_a^2), stud (σ_s^2), residual variance (σ_r^2), and heritability (h^2; SE in brackets) for the five novel linear traits considered in the study.

Trait	σ_a^2	σ_s^2	σ_r^2	h^2 (SE)
Neck length (NL)	0.0234	0.1139	0.0942	0.101 (0.033)
Neck direction (ND)	0.0233	0.0879	0.0818	0.121 (0.036)
Croup length (CL)	0.0315	0.0495	0.0669	0.213 (0.047)
Croup direction (CD)	0.0082	0.0240	0.0653	0.084 (0.038)
Body condition score (BCS)	0.0608	0.1117	0.1404	0.194 (0.043)

Table 3. Genetic (above diagonal) and phenotypic (below diagonal) correlations for the five novel linear traits considered in the study.

Trait	NL	ND	CL	CD	BCS
NL	-	0.901 (0.094)	0.640 (0.151)	0.322 (0.271)	0.171 (0.191)
ND	0.717 (0.013)	-	0.755 (0.129)	0.245 (0.296)	0.186 (0.180)
CL	0.482 (0.021)	0.509 (0.020)	-	0.737 (0.194)	0.322 (0.151)
CD	0.268 (0.024)	0.312 (0.023)	0.361 (0.022)	-	0.285 (0.235)
BCS	0.112 (0.028)	0.141 (0.027)	0.154 (0.026)	0.073 (0.025)	-

[1] NL=Neck length; ND=Neck direction; CL=Croup length; CD = Croup direction; BCS=Body condition score.

Table 4. Genetic and phenotypic correlations between the five novel linear traits and two established linear traits.

	Temperament		Fleshiness	
Trait	r_a	r_P	r_a	r_P
Neck length	0.426 (0.163)	0.230 (0.026)	0.551 (0.165)	0.300 (0.023)
Neck direction	0.608 (0.136)	0.237 (0.025)	0.785 (0.126)	0.340 (0.022)
Croup length	0.475 (0.141)	0.146 (0.025)	0.683 (0.113)	0.287 (0.023)
Croup direction	0.586 (0.214)	0.046 (0.025)	0.601 (0.196)	0.104 (0.024)
Body condition score	0.407 (0.148)	0.070 (0.026)	0.322 (0.139)	0.197 (0.024)

the phenotypic recordings for breeding in many European livestock breeds, including horses. The IHDH has been traditionally evaluated for the dual purpose of heavy draught and meat. Recently the breed has experienced a renewed attention for the original heavy draught purpose now devoted to the activity of didactic and organic farms, and to leisure pursuits such as team races (Mancin *et al.*, 2020). The novel traits for neck and croup are aimed to improve the movement ability of the horses. The neck is used for balancing while trotting, and too short or too long necks, as well as necks not in-line with the body, are less suitable to well balance the movement. Moreover, a too short and horizontally placed croup may cause problems to the movement of the limbs. However, a very sloped croup is more an aesthetic defect than in conflict with the thrust in different gaits. Linear type evaluations for croup and neck are found in many horses bred for sport activities, e.g. in Pantaneiro horse (neck length and croup length, r_g=0.46; McManus *et al.*, 2002) and in Dutch Warmblood (Koenen *et al.*, 1995). In this latter breed, patterns of genetic parameters for length and position of neck (h^2=0.21 and h^2=0.12, respectively; r_g=0.42) as well as croup length and shape (h^2=0.15 and h^2=0.28; r_g=0.40) were similar to this study. In Belgian Warmblood horses, length (h^2=0.27) and orientation (h^2=0.24) of neck, and length (h^2=0.30) and slope (h^2=0.30) of croup were also considered (Rustin *et al.*, 2009). BCS is linked to individual welfare of an animal: The trait is reflecting subcutaneous individual fat deposition and is routinely used in many livestock breeds including horses to assess body fat reserves and energy status. Development of energy balance over time can be crucial for metabolism, reproduction, and health (Coffey *et al.*, 2001). Data on evaluations of BCS in horses can be found, e.g. for Pura Raza Español horse (Sánchez-Guerrero *et al.*, 2019; h^2=0.15), Dutch Shetland Pony mares (Schurink *et al.*, 2010; h^2=0.38) and Italian Haflinger (Samoré *et al.*, 1997; h^2=0.09 for BCS). Horse literature is scarce about studies considering neck, croup, and BCS traits in parallel (e.g. Samoré *et al.*, 1997), and the genetic correlations have been studied even less. The present study can offer interesting findings about linear conformation traits related to dynamic aptitude and body status (energy supply). Positive genetic relationships have been found, but the traits have intermediate optima, so possible selection should maintain intermediate values whilst improving traditional productive traits. Temperament could also help in improving the movement aptitude, since it is scored to obtain lively individuals with favourable movement abilities (Folla *et al.*, 2019).

The high genetic correlation of temperament with new phenotypes indirectly related to dynamic attitudes could be explained considering the great attention put by classifier in the evaluation of temperament, that is in young foals strictly linked also to the observed animal movements. Fleshiness evaluates muscle masses underlying the meat production traits, and the genetic correlations indicating no antagonism with movement related conformation traits suggest the possibility to proficiently improve both meat and draught aptitude.

Acknowledgements

The authors thank the National Breeder Association for Italian Heavy Draught Horse breeds (ANACAITPR) which funded the study within Equinbio (PSRN 2014-2020 FEASR, sottomisura 10.2) and Equinbio.2 (PSRN 2020-2023 FEASR, sottomisura 10.2) projects.

References

Coffey M.P., Emmans G.C., and Brotherstone, S., (2001). Anim Sci 73, 29–40. https://doi.org/10.1017/S1357729800058021.

Koenen E.P.C., van Veldhuizen A.E., and Brascamp E.W. (1995). Livest Prod Sci 43(1), 85–94, https://doi.org/10.1016/0301-6226(95)00010-I

Folla F., Sartori C., Guzzo N., Pigozzi G., and Mantovani R. (2019). Livest Sci 219, 91-96. https://doi.org/10.1016/j.livsci.2018.11.019

Houle D., and Meyer K. (2015). J Evol Biol 28(8), 1542-1549. https://doi.org/10.1111/jeb.12674

Mancin E., Ablondi M., Mantovani R., Pigozzi G., Sabbioni A., and Sartori, C. (2020). Animals, 10(8), 1310. https://doi.org/10.3390/ani10081310

McManus C., Miserani M., Santos S.A., Silva J.A., … and Sereno J.R. (2002). Proc. of the 7th WCGALP, Montpellier, France (Vol. 5, pp. 5-18)

Misztal I. (2008). J Anim Breed Genet 125(6), 363-370. https://doi.org/10.1111/j.1439-0388.2008.00774.x

Rustin M, Janssens S, Buys N, and Gengler N. (2009). J Anim Breed Genet, 126(5), 378-386. https://doi.org/10.1111/j.1439-0388.2008.00798.x

Sánchez-Guerrero M.J., Ramos J., Valdés M., and Valera M. (2019). Livest Sci, 230, 103851. https://doi.org/10.1016/j.livsci.2019.103851

Samoré A.B., Pagnacco G., and Miglior F. (1997). Livest Prod Sci, 52(2), 105-111. https://doi.org/10.1016/S0301-6226(97)00143-7

Schurink A., Vogelzang R.H., and Ducro B.J. (2010). Proc. of the 9th WCGALP, Leipzig, Germany (Vol. 243).

762. Effects of selection on breed composition of the Spanish Sport Horse population

E. Bartolomé[1], S.T. Rodríguez-Ramilo[2*], J. Fernández[3] and M. Valera[1]

[1]Departamento de Agronomía, ETSIA, Universidad de Sevilla, 41013 Seville, Spain; [2]GenPhySE, INRAE, INPT, ENVT, 31326 Castanet Tolosan, France; [3]Departamento de Mejora Genética Animal, INIA, Madrid, Spain; silvia.rodriguez-ramilo@inrae.fr

Abstract

The Spanish Sport Horse (CDE) is a recently developed composite breed selected mainly for show jumping performance since 2002, when the CDE studbook was officially founded. The main aim of this study was to highlight the effects of selection on this breed, regarding the breeds composition, and its repercussion in the breeding program. The CDE pedigree, with 47,884 animals (21,163 registered as CDE), was used for the analyses, including 10 breed groups corresponding to the major horse breeds present in the CDE Studbook. Two periods were considered to analyse the genetic contribution of founders for each breed group. Estimated breeding values (EBV) were available for show jumping discipline. Results showed that Other Horse Breeds group had the highest genetic contribution to CDE, followed by the Spanish Purebred. A small genetic progress was observed, as founders from most breeds showed lower EBVs than their descendants. This indicates that the genetic selection made has been effective, even though the management was not perfect.

Introduction

Recently, the equine breeding industry has become a very profitable activity, selecting not only pure horse breeds, but creating composite breeds with a focus on the horses' ability for a certain sport performance (Bartolomé *et al.,* 2011). Examples are the Irish Sport Horse, the Brazilian Sport Horse and recently, the Spanish Sport Horse (CDE). Particularly, this last breed is bred for dressage, eventing and endurance with a special focus on show jumping. The breed is managed by the Spanish Sport Horse Breeders Association (ANCADES, www.ancades.com) and its breeding program was officially approved in 2005. Previous studies on the CDE breed have shown improvements on the sport performance (Bartolomé *et al.,* 2018; Solé *et al.,* 2017), which could be a signal for certain selection in the breed. However, no studies have been conducted on CDE breeders' preferences when selecting reproducers for the next generation (EBV values for show jumping performance or breed composition) and neither on the effectiveness of the CDE Breeding Program. Therefore, the aim of this study was to highlight the effects of selection on this breed and whether it has been effective and adequate regarding the CDE main breeding objectives. This objective has been evaluated from two points of view: (1) partitioning CDE breed composition; and (2) evaluating EBV for show jumping discipline.

Materials & methods

The CDE Official Stud Book was created in 2002 and the foundational registry was closed in 2004 (Bartolomé *et al.,* 2011). The pedigree information used in this study included 47,884 animals (18,799 males and 29,085 females) with 21,163 registered as CDE horses (44.7% males and 55.3% females). According to the composite nature of this breed, different horse breeds were included in the CDE studbook via relatives of the CDE animals (55.8% of the whole pedigree). Thus, 10 breed groups were assessed corresponding to the major horse breeds present in the CDE Stud Book: Anglo-Arab Horse breed (AAH; n=1,561), Arab Horse breed (AH; n=1,835), Hanoverian breed (HANN; n=1,742), Holsteiner breed (HOLST; n=1,269), Dutch Warmblood (KWPN; n=2,120), Spanish Purebred (PRE; n=5,434), Thoroughbred (TH; n=2,924), Selle Français (SF; n=2,311), Other Horse Breeds group (OHB; n=7,525) and the Spanish Sport Horse breed group (CDE; n=21,163). The OHB group included a total of 69 breeds with less than 1000 animals each. In order

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_762, © E. Bartolomé *et al.* 2022

to study the evolution of these breeds over time (in terms of the proportion of individuals belonging to each breed) and their influence on CDE pedigree, two different periods were considered according to the year of birth of the animals. A first period (*P0*), including non-active animals born before 2000, was settled before the Foundational Registry of the CDE Stud Book was closed, in 2004. Consequently, no CDE animals were found in this period. Then, a second period was considered (*P1*) including active (foreign or not) animals, born from 2000 to 2021. Table 1 shows the CDE's breed composition according to the period considered.

Probability of gene origin of the CDE horse breed. In order to check the genetic representation of the founder individuals in the CDE population and in all the breed groups considered, the *genetic contribution of founders* to the descendant gene pool of the population (Lacy, 1989) was assessed for each breed group in every period considered. Following Moureaux *et al.* (1996), ancestors with both unknown parents in the available database were considered founders. These pedigree analyses were computed using the program ENDOG v.4.8 (Gutierrez and Goyache, 2005).

Genetic selection for Show Jumping discipline. Performance ability for Show Jumping was evaluated using the information from the routine genetic evaluations of CDE breeding program. The estimated breeding values (EBV) of the Show Jumping sport discipline of 12,197 horses in the CDE pedigree were available for this study (Bartolomé *et al.,* 2018). EBVs were standardized for an interval of 80-120 with a population average of 100. Genetic parameters for the Show Jumping sport discipline were estimated using a multivariate BLUP analysis carried out with the VCE software – version 6.0.2 (Groeneveld *et al.,* 2010), using the following genetic model:

$$SJ_{ijklmnopq} = b_q + Br_j + Gen_k + Ev_l + Hf_m + Cti_n + Rhi_o + pe_p + a_q + e_{ijklmnopq}$$

where $SJ_{ijklmnopq}$ was the Show Jumping score of animal q with Age i, from Breed j, of Gender k, recorded at the Event l, where Height of fences was m, Category and type of Test interaction n, Rider-horse interaction o and permanent environment p. b_q = linear covariate for the age effect of the q animal; Br_j = systematic effect of j^{th} breed (j = 1, …, 19); Gen_k = systematic effect of k^{th} gender (k = male, female, gelding); Ev_l = systematic effect of l^{th} event (l = 1, …, 1,131); Hf_m = systematic effect of m^{th} height of fences (m = 1, …, 10); Cti_n = systematic effect of n^{th} category and type of test interaction (n = 1, …, 214); Rhi_o = random effect of o^{th} rider-horse interaction (o = 1, …, 22,320); pe_p = random effect of p^{th} horse permanent environment (p = 1, …, 10,508); a_q = random effect of q^{th} animal (q = 1, …, 10,508); and $e_{ijklmnopq}$ = random residual effect.

Results

Breed groups' distribution appearing in Table 1 showed PRE and OHB as majority horse breeds in *P0*, with 4,387 and 6,357 animals born in this period, respectively. Afterwards, CDE was the majority horse breed in

Table 1. Distribution of horse breeds from the Spanish Sport Horse Official Studbook according to the period considered.[1]

	Breed groups										TOTAL
	AAH	AH	CDE	HANN	HOLST	KWPN	OHB	PRE	TH	SF	
P0	1,415	1,548	0	1,445	1,111	1,676	6,357	4,387	2,790	2,145	22,874
P1	146	287	21,163	297	158	444	1,168	1,047	134	166	25,010

[1] *AAH* is Anglo-Arab Horse, *AH* is Arab Horse, *CDE* is Spanish Sport Horse, *HANN* is Hannoverian Horse, *HOLST* is Holsteiner Horse, *KWPN* is KWPN Horse, *OHB* is Other Horse Breeds, *PRE* is Spanish Purebred, *TH* is Thoroughbred, *SF* is Selle Français, *P0* is period with animals born before 2000 and *P1* is period with animals born between 2000 and 2021.

P1. On the other hand, OHB, PRE and KWPN, breeds showed the higher representation after CDE in *P1*, with 1,168, 1,047 and 444 animals, respectively. The rest of the breeds contributed less than 300 animals.

Effect of genetic selection on the probability of gene origin of the CDE horse breed. In order to ascertain the genetic history of the CDE breed, the average genetic contribution of founders according to breed and period, was represented in Figure 1 including mean EBV for each breeds' founders and for their descendants calculated for Show Jumping discipline. Breeds were ordered according to their accumulated genetic contribution. Results showed that OHB horses had the highest genetic contribution, with almost 23% followed by PRE and TH breeds, with 14.8 and 12.6%, respectively. Whereas, AH, AAH and KWPN breeds showed the lowest, with an accumulated contribution less than 6%. On the other hand, mean EBV of founders was below the mean (EBV<100) for most breeds except for HOLST, with KWPN, SF and AAH obtaining EBVs on the mean. However, when analysing mean EBV of their descendants, most of them were over the mean, with SF, HOLST and KWPN showing EBV>101. Whereas only CDE, PRE and AH' descendants showed EBV means below 100. Furthermore, only PRE and AH founder's breeds showed higher EBVs than their descendants, indicating that breeders have not preference for PRE and AH breeds.

Discussion

The selection and mating are processes of conscious decisions for short-term achievements in intensively managed domesticated species (Medugorac *et al.*, 2011). Results from pedigree structure in the CDE population over the two periods considered, indicated that there are still some horse breeds such as the PRE and OHB that, despite their decreased influence from the founder population to the active CDE population,

Figure 1. Average genetic contribution of founders classified by breed (each column), per period (different colours within columns), with their mean estimated breeding value (EBV Founders; yellow squares) and the mean estimated breeding value of their descendants (EBV Descendants; blue triangles) for show jumping equestrian discipline, ordered according to contribution level of the founders. Where *AAH* is Anglo-Arab Horse, *AH* is Arab Horse, *CDE* is Spanish Sport Horse, *HANN* is Hannoverian Horse, *HOLST* is Holsteiner Horse, *KWPN* is KWPN Horse, *OHB* is Other Horse Breeds, *PRE* is Spanish Purebred, *TH* is Thoroughbred, *SF* is Selle Français, *P0* is period with animals born before 2000 and *P1* is period with animals born between 2000 and 2021. The red dots line indicates the mean *EBV* = 100.

they are still the largest contributing breed group. The reason for PRE retaining its position could be that it is the main national horse breed (Solé *et al.,* 2018), thus making it a very handy option for the breeders. Accordingly, the breed group OHB includes several foreign sport horse breeds, all of them with high sport standards, which have been selected for sport performance before CDE breed was created (Koenen *et al.,* 2004). And, thus, a very interesting option for CDE breeders as potential reproducers for their animals. The evolution of founder genetic contributions and the EBVs for Show Jumping, per breed group and period, showed, first, a general tendency of the breeders to follow the EBVs from the genetic evaluations of the breed published by ANCADES each year, to select the best breeding stock for their animals, thus increasing the genetic progress of the breed. And secondly, that CDE breeders have been also doing a selection against PRE and AH animals on their CDE composition. This was probably due to these horse breeds were not specialized in show jumping discipline, but in dressage (Sánchez-Guerrero *et al.,* 2014) and endurance (García-Ballesteros *et al.,* 2018), respectively, thus being not desirable breeds for Show Jumping's CDE breeders. Finally, results showed a light genetic progress, indicating that the genetic selection made is being effective, even though the management is not perfect.

Acknowledgements

The authors wish to thank the National Association of Spanish Sport Horse Breeders (ANCADES) for providing the data used in this study.

References

Bartolomé E., Cervantes I., Valera M., and Gutiérrez J.P. (2011) Liv Sci, 142: 70-79. https://doi.org/10.1016/j.livsci.2011.06.021

Bartolomé E., Menéndez-Buxadera A., Molina A., and Valera M. (2018) J Anim Breed Genet, 135 (2): 138-148. https://doi.org/10.1111/jbg.12315

García-Ballesteros S., Varona L., Valera M., Gutiérrez J.P., and Cervantes I. (2018) Animal, 12(1): 20–27. https://doi.org/10.1017/S1751731117001331

Groeneveld E., Kovac M., and Mielenz N. (2010) VCE Users's Guide and Reference Manual Version 6.0 University of Ljubljana.

Gutiérrez J.P., and Goyache F. (2005) J Anim Breed Genet, 122: 172-176. https://doi.org/10.1111/j.1439-0388.2005.00512.x

Koenen E.P.C., Aldridge L.I., and Philipsson J. (2004) Liv Prod Sci, 88(1-2): 77-84. https://doi.org/10.1016/j.livprodsci.2003.10.011

Lacy R.C. (1989) Zool Biol, 8: 111–123. https://doi.org/10.1002/zoo.1430080203

Medugorac I., Veit-Kensch C.E., Ramljak J., Brka M., Marković B., *et al.* (2011) Ecol Ev, 1(3): 408–420. https://doi.org/10.1002/ece3.39.

Moureaux S., Verrier E., Ricard A., and Meriaux J.C. (1996) Genet Sel Evol, 28: 83–102. https://doi.org/10.1186/1297-9686-28-1-83

Sánchez-Guerrero M.J., Cervantes I., Valera M., and Gutiérrez, J.P. (2014) J Anim Breed Genet, 131(5): 395–402. https://doi.org/10.1111/jbg.12088

Solé M., Bartolomé E., José Sánchez M., Molina A., and Valera M. (2017) Livest Sci, 200: 23-28. https://doi.org/10.1016/j.livsci.2017.03.019.

Solé M., Valera M., and Fernández, J. (2018) J Anim Breed Genet, 135(6):460-471. https://doi.org/10.1111/jbg.12352

763. Early selection criteria for improved longevity of jumping horses

A. Ricard[1,2*], S. Deretz[1] and B. Dumont Saint Priest[2]

[1]GABI, INRAE, AgroParisTech, Université Paris Saclay, Centre de Recherche Ile-de-France-Jouy-en-Josas-Antony, 78352 Jouy-en-Josas, France, [2]Institut Français du Cheval et de l'Equitation, Pôle développement, innovation et recherche, 61310 Exmes; France, anne.ricard@inrae.fr

Abstract

In order to find early selection criteria to improve the longevity of show jumping horses in competition, a specific protocol was constructed. Before entering competition, young horses were measured for many traits. These horses were offspring of two groups of sires selected as having the highest and lowest estimated breeding values for functional longevity in jumping competition calculated from progeny. Functional longevity was defined as the time spent in competition corrected for the level of performance. Initial results on subsamples of 346 horses for blood markers and 529 horses for surface temperature from infrared thermography are reported. Key markers of oxidative stress had significant effect on group of sires, especially superoxide dismutase. Surface temperature showed high heritability (up to 0.70) for body and contrast of body with specific locations (fetlocks, hocks, foot, eyes), but only asymmetry in right and left temperature of feet was linked to group of longevity of the sire.

Introduction

Improved longevity is beneficial for the welfare and health of the horse, as well for the satisfaction of a committed rider. To select for this objective is however difficult: heritability is low (0.08 in Ricard and Blouin, 2011) and estimated breeding values (EBV) for longevity are only obtained late in life of breeding animals, as a large number of uncensored phenotypes of relatives are required. This is further compromised as average length of jumping career is about 6 years with maximum over 20 years while horses begin competition at 4 years old. Therefore, possibilities of indirect early selection have been investigated in literature. Selection criteria in these studies were often based on traditional health status, morphological and gait measurements (Jonsson *et al.*, 2014; Seiero *et al.*, 2016; Wallin *et al.*, 2001). That is why a specific protocol was constructed in this study. Two groups of stallions were created (high and low) according to their EBVs for functional longevity in jumping competition calculated from progeny performances. Prior to start of competitive life, young horses (£4 years) descendants from these stallions were measured for different criteria: welfare, temperament, linear profiling of morphology, physiological parameters, blood biomarkers, body composition through impedance measurements (McKeen and Lindinger, 2004), gait accelerometric measurements, heart rate (during gaits and temperament tests), surface temperatures (from infrared images). Initial results on subsamples of 346 horses for blood markers and 529 horses for surface temperature are reported. The aims of the study are: (1) prediction of sire group (high longevity and lower longevity); and (2) heritability estimates for surface temperature measurements.

Materials & methods

Blood markers. A total of 91 variables were obtained from blood tests on 346 horses. Horses were mainly Selle Français (94%) aged 2 years (9%), 3 years (44%) and 4 years (47%). There were 45% females, 14% geldings and 41% stallions. A proportion of 66% descended from the group of favourable sires for longevity and 34% from the unfavourable. They were measured during 15 events (combination of location and date). Difference between the two groups of sires for longevity was equal to 0.46 genetic standard deviation, the risk ratio between the two groups was 1.2. Blood variables were divided into eight categories: liver function, kidney function, muscle function, oxidative stress, osteoarticular function, inflammation markers, blood counts and a miscellaneous section. Principal component analysis (PCA) was used to describe relationships

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_763, © A. Ricard *et al.* 2022

between variables. General linear models were used to identify environmental factors. Logistic regression was used to predict the group of sires from blood variables with stepwise selection. Quality of logistic regression was assessed by measuring area under curve (AUC).

Surface temperature. Infrared images from the Trotec IC085LV thermal imaging camera were used to measure surface temperature. For each horse, six images were taken: back, left profile, right profile, front legs, back legs, head. The software IC-IR Report allows to draw geometrical figures on the image and to extract the minimum, average and maximum temperature of a region of interest (ROI). For each image a rectangular reference area was defined to measure the temperature of the whole visible horse. Then specific morphological ROIs were defined. For each horse, the average, maximum and minimum temperatures of six reference areas and 28 ROIs were measured. Data were available for 529 horses, 3,138 images and 17,328 areas. Horses were mainly Selle Français (95%) aged 2 years (18%), 3 years (54%) and 4 years (28%). There were 48% females, 15% geldings and 37% stallions. The horses were sired by 88 sires, 343 by sires classified as favourable for longevity (65%) and 186 (35%) by sires classified as unfavourable. Difference in longevity for the two groups was the same as for the blood variables. The average number of offspring per sire was 6.0, ranging from 1 to 50. Horses were measured at 23 different combinations of location and date between 2018 and 2021 (12 to 34 horses per event). Variables were transformed to better describe the specificity of the temperatures of each horse. We defined six types of variables:

- body, overall images and per image;
- differences between body and specific areas: fetlocks, hocks, foot;
- differences between areas: fetlock to foot, hock to foot;
- differences antero-posterior for foot;
- differences right and left for fetlock, hock, back, foot and eyes;
- differences between median and lateral values for back.

A univariate animal mixed model was used with the fixed effects of age (3 levels), gender (3 levels), event (23 levels), and a random effect of the genetic value of the horse. Relationship matrix included 5,014 horses (pedigree back over 5 generations). The effect of the group of sires for functional longevity on temperature was tested. We performed a principal component analysis (PCA) of the variables corrected for the effects of age, gender and event estimated by the previous model.

Results

Blood tests. PCA confirmed the categories of blood variables according to functions with higher absolute phenotypic correlations between variables within functions. Heritabilities could not be estimated because of too low sample size, but the random effect of sire was significant for 44 variables, implying that most blood variables were probably heritable. Event effect was significant for 45% of variables, age for 25% and gender for 85%.

Blood test: logistic regression to predict sire group. The AUC of the logistic regression was 0.76, not fully satisfactory. Table 1 shows the significant variables (*P*-value£0.1). Horses descending from sires with favourable longevity had a higher level of erythrocyte superoxide dismutase (SOD in U/g Hb) as well as lower haemoglobin and bile acid levels than horses descending from sires with unfavourable longevity.

Surface temperature: environmental factors and heritability. Event effect was significant for all variables except for some left versus right asymmetry variables. Gender was rarely significant, especially when measuring asymmetry. Age was significant for almost half of the variables. Two years olds and three-years olds were cooler overall than four years olds. The temperature of the fetlocks, feet and eyes was higher relative to the body in 2- and 3-year-olds than in 4-year-olds. Youngest horses were more asymmetrical

Table 1. Blood test variables selected in logistic regression used to differentiate between the longevity group of sires: mean, standard deviation, *P*-value, estimate.

	Mean	SD	*P*-value	Estimate
Bile acids (µmol/l)	5.426	2.428	0.0130	0.24
Bilirubin (mg/l)	13.62	5.72	0.0068	-0.113
Total protein (g/l)	64.49	7.00	0.0410	-0.086
Globulin beta 1 (g/l)	9.882	3.480	0.0850	0.15
SOD (U/g Hg)	1,959	525	0.0006	-0.00148
Copper (mg/l)	0.9741	0.1807	0.0682	-2.3
Leucocyte (/mm^3)	8.754	1.946	0.0028	-0.69
Small lymphocytes (/mm^3)	3,626	948	0.0111	0.00081
Monocytes (/mm^3)	375.7	130.3	0.0582	0.0039
Haemoglobin (g/dl)	12.55	1.29	0.0030	1.104

(fetlock, back, eyes). Females had warmer feet and fetlocks relative to the body than geldings and stallions. Eyes were coolest relative to the body in stallions, followed by females and then geldings. Table 2 shows heritabilities for variables with estimates significantly different from zero for average and/or maximum temperature.

Surface temperature: PCA. The first principal component explained only 17% of the variance. Twenty-three variables were needed to explain 80% of the variance, according to the deliberate choice to avoid redundant variables by using temperature differences. The first axis discriminated between horses that had a rather cold whole body and warm eyes, hocks, fetlocks and feet relative to the body and the opposite (warm body, cold extremities and joints relative to the body). The former had accordingly greater temperature contrasts. These horses also had warm feet relative to the hocks and fetlocks. The second axis gathered the right/left asymmetry variables for fetlocks and feet. The variables calculated from the average or maximum temperatures were on the same axes and indicated the same. The third axis discriminated horses that had warmer hocks than feet, warmer fetlocks than feet and asymmetry in hock temperatures. The fourth axis highlighted the antero-posterior asymmetries of foot temperatures.

Table 2. Heritabilities (SE) of variables derived from measured surface temperatures on 6 images: back, left profile, right profile, front legs, back legs, head.

Measurement	Region of interest	Temperature	
		Average	Maximum
Reference	Overall mean	0.68 (0.19)	-
	Head – overall mean	0.31 (0.14)	-
Specific relative to Reference	Fetlocks	0.58 (0.17)	0.58 (0.16)
	Hocks	0.26 (0.15)	0.27 (0.16)
	Feet	0.64 (0.18)	0.64 (0.17)
	Eyes	0.09 (0.12)	0.70 (0.18)
Specific	Hocks relative to Fetlocks	0.25 (0.13)	0.29 (0.14)
Asymmetry antero-posterior	Feet	0.15 (0.11)	0.13 (0.10)
Asymmetry right-left	Hocks	0.22 (0.13)	0.11 (0.10)
Asymmetry median-lateral	Back polygons	0.70 (0.18)	0.00

Surface temperature: effect of longevity group of sires. Only one variable showed significant difference for EBV-based sire group for longevity (*P*-value 0.03): right and left asymmetry in posterior feet. Asymmetry was lower for the group of favourable sires.

Discussion

Results highlighted the first steps to find early selection criteria for longevity. Oxidative stress and asymmetry of feet temperature were the first interesting criteria. It is possible to relate the amount of SOD to longevity in two ways: either this antioxidant protects against the deterioration of cellular components and thus helps to maintain the cellular integrity necessary for intense exercise; or a decrease in SOD in horses with poor longevity is linked to certain diseases (muscular / respiratory) that may cause the horse to stop competing. Asymmetry in temperature may indicate inflammatory, vascular, or neurological disorder (Soroko and Howell, 2018). Even if surface temperature is highly influenced by environmental factors (Jansson *et al.*, 2021), high heritability of infrared temperatures has already been found in horses. Heritability was 0.26 (0.16) for differences in eyes temperature during effort phase and 0.52 (0.07) during recovery phase in a jumping test (Bartolome *et al.*, 2021). Heritability of eye temperature was 0.38 (0.17) before and 0.50 (0.22) just after dressage competition (Sanchez *et al.*, 2016). Eyes temperature may be related to the level of stress (Cook *et al.*, 2001). When the study protocol will be completed, and 1000 horses measured, more reliable results regarding potential early indicators for longevity can be produced.

Funding

This work was funded by INRAE, Institut Français du Cheval et de l'Equitation and Fonds Eperon.

References

Bartolome E., Perdomo-Gonzalez D.I. Sanchez-Guerrero M.J., and Valera M. (2021) Animals 11. https://doi.org/10.3390/ani11030832

Cook N., Schaefer A., Warren L., Burwash L., Anderson M *et al.* (2001) Can. J. Anim. Sci. 81:621.

Jansson A., Lindgren G., Velie B.D., and Sole M. (2021) Physiol. Behav. 228. https://doi.org/10.1016/j.physbeh.2020.113218

Jonsson L., Egenvall A., Roepstorff L., Nasholm A., Dalin G. *et al.* (2014) Journal of the American Veterinary Medical Association 244:1449-1461.

McKeen G., and Lindinger M.I. (2004). Equine and Comparative Exercise Physiology 1:199-209. https://doi.org/10.1079/ecep200422

Ricard A., and Blouin C. (2011) J. Anim. Sci. 89:2988-2994. https://doi.org/10.2527/jas.2011-3931

Sanchez M.J., Bartolome E., and Valera M. (2016) Appl. Anim. Behav. Sci. 174:58-65. https://doi.org/10.1016/j.applanim.2015.11.006

Seiero T., Mark T., and Jonsson L. (2016) Livest. Sci. 184:126-133. https://doi.org/10.1016/j.livsci.2015.12.010

Soroko M., and Howell K. (2018) J. Equine Vet. Sci. 60:90-96. https://doi.org/10.1016/j.jevs.2016.11.002

Wallin L., Strandberg E., and Philipsson J. (2001) Livest. Prod. Sci. 68:97-105. https://doi.org/10.1016/s0301-6226(00)00244-x

764. Recent advances in genomics of equine health, welfare and performance

S.C.M. Mikko

Swedish University of Agricultural Sciences, Dept of Animal Breeding and Genetics, P.O. Box 7023, 750 07 Uppsala, Sweden; sofia.mikko@slu.se

Abstract

Equine research and breeding have encountered major changes due to the rapid development of molecular genetic technology. Today, there are 254 equine traits listed in Online Mendelian Inheritance in Animals (OMIA), while the *Horse*QTLdb lists 2,605 QTLs representing 64 traits. Strong selection increases the risk of inbreeding and multiplication of deleterious variants. Analysis of genomic regions under selection, and trio sequencing could complement Genome-wide Association Studies (GWAS) in the search for genetic background of traits and diseases. The increased possibilities to test for a variety of disease-causing mutations, or performance-associated QTLs, support breeders in informed breeding decisions. SNP genotyping will most likely be implemented for horse parentage analysis in the near future. This large amount of genomic data will open new possibilities to address research questions, and will make it possible to carry out marker-assisted and/or genomic selection also in horses.

Introduction

Equine research and breeding currently encounter major changes due to the rapid development of molecular genetic technology. The horse was among the first mammals to be sequenced, and the trace files, and first gene build of the EquCab2 assembly, was released in already in 2007 (Wade *et al.* 2009). This resulted in an increase of likely causal variants detected and reported to the database Online Mendelian Inheritance in Animals (OMIA) (Figure 1).

The availability of the horse genome assemblies likewise increased the number of QTLs reported to the Horse Quantitative Trait Locus Database (*Horse*QTLdb) (Figure 2; Hu *et al.* 2007, 2022). In 2018 the EquCab3.0 assembly was released providing the community with updated genomic data (Kalbfleisch *et al.* 2018), which resulting in even more variant- and QTL-associations reported.

Equine genomic research has created new possibilities to effectively identify and study genetic risk factors for hereditary diseases in order to develop new and efficient treatments. The gain in genomic understanding about complex traits has further facilitated the selection process for performance in sport horses. Breeding

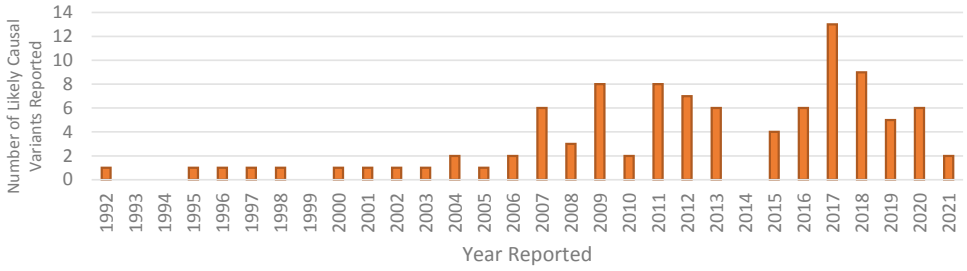


Figure 1. Number of likely causal variants reported in OMIA per year.

R.F. Veerkamp and Y. de Haas (eds)
Proceedings of 12th World Congress on Genetics Applied to Livestock Production (WCGALP)
DOI: 10.3920/978-90-8686-940-4_764, © S.C.M. Mikko 2022

organizations are obliged to monitor mutations unfavourable to animal welfare, and should be proactive in providing breeding advise promoting sound and healthy individuals. Offering a scientifically supported platform will ensure that the Animal Welfare Act and breeding advice are adapted to current and future technological developments in genomics. However, for horse breeding in practice, the new technologies present both opportunities and risks with the entire genome of a horse sequenced in an effort to detect disease-causing mutations, and at the same time reveal information about other genetic variants in this particular individual. Hence, we are facing a new era where breeders will need and demand assistance and guidance by geneticists, veterinarians and other professionals to interpret the plethora of data available so they can make informed choices in their breeding scheme. The structure of the equine breeding business mainly differs from the breeding of production animal with very few large internationally operating commercial breeding companies. Instead, the equine breeding mostly relies on smaller private companies, or single breeders gathering a quite disparate group of people with a diverse base of knowledge. To produce competitive and healthy horses, breeders therefore need reliable information based on scientific evidence.

Strong selection pressure, as in selective breeding of elite performing horses, will result in chromosomal regions with lower genetic variation, that could be referred to as 'regional inbreeding'. Identification of genetic markers or causative mutations in such regions provide new selection tools in horse breeding. By calculating breed divergence, we revealed long homozygous regions in equine chromosomes, comprising genes possibly important for performance in sport horses (Ablondi *et al.* 2019a). Further, we revealed significantly divergent selection pressure on chromosome regions comprising genes controlling cell signalling in the central nervous system in horses not bred for show jumping, compared to show jumping horses. Our analysis discovered positive selection on chromosome regions with genes involved in mobility, relaxed locomotion and sensitivity in non-show jumping horses. In show jumping horses, positive selection, to a larger extent involve regions with genes associated with mentality, postsynaptic signalling, reward system, neuromuscular control of coordination, muscle growth and -function (Ablondi *et al.* 2019b).

Important monogenic and multifactorial diseases, and traits in the horse

Strong selection increases the risk of inbreeding and multiplication of deleterious disease-causing variants, that do not only negatively affect health and welfare of individual horses, but also have negative economic consequences for breeders. Knowledge of association between traits and genotypes with major influence will therefore aid breeders to produce healthy, sustainable, and better performing horses.

Currently, OMIA lists in total 254 Equine traits, and disorders (https://omia.org/home/). The absolute majority of them are health related (82%), followed by exterior (8%), and reproduction (6%), while there are fewer growth, performance, or behaviour traits reported. A likely causal variant is known in 99 of

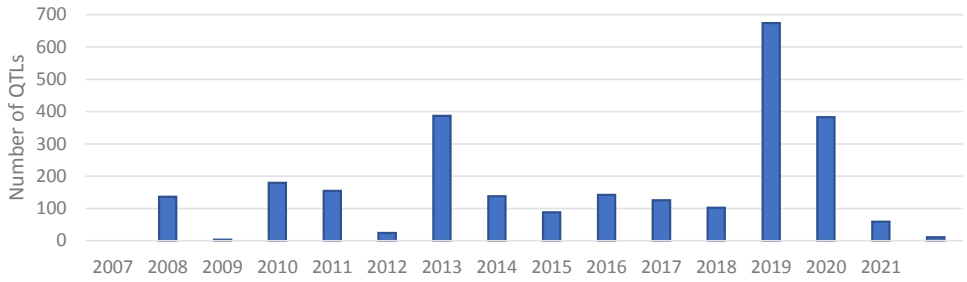


Figure 2. Number of QTL/associations published and reported in the *Horse*QTLdb per year.

these traits or disorders, where 54% of them represent coat colours, 41% are disease-causing variants, and 4% of the variants are associated with performance traits. Among the most recent studies of important monogenic diseases are Fragile Foal Syndrome (Ablondi *et al.* 2022), Dwarfism and joint laxity in Frieasian horses (Leegwater *et al.* 2016), and Ocular Squamous Cell Carcinoma (Chen *et al.* 2021).

Since 2007, 2,605 QTLs/associations representing 64 traits, have been reported to the *Horse*QTLdb (https://www.animalgenome.org/cgi-bin/QTLdb/EC/index). Also in this database, health traits are the most common (48%), followed by growth traits (24%), performance traits (17%), and exterior (6%). Reproduction and behaviour traits are the least reported in the *Horse*QTLdb. Among the most recent studies of important Equine QTLs are; Osteochondrosis (Drabbe *et al.* 2022), Height at Withers (Vosgerau *et al.* 2022), and gait and racing performance (Ricard and Duluard 2022). The difficulties in recording behaviour phenotypes probably reflect the small number of genetic associations reported in both OMIA and *Horse*QTLdb.

Inherited congenital disorders are not always reported by the breeders to the breeding association, and the frequencies of traits and disorders differ largely between breeds, why it is challenging to calculate prevalences. Breeds with an economic value may be studied more often than native and feral horse breeds kept mainly for conservation purposes. In race- and sport horses, a functional and healthy musculoskeletal system is critical for their use, while endangered breeds might instead face fertility problems due to inbreeding. Such factors will influence what specific traits and disorders will be studied, and will not always reflect the prevalence or importance of a trait. In general, health, size and performance traits are the most studied reflecting the functional use of horses.

From disease to genetic testing

To find an assumed genetic association, it is of utterly importance to obtain a detailed clinical description reflecting the correct disease classes associated with the genotypes. A GWAS could be performed using any kind of large-scale genotyping methodology; SNP arrays, 'Genotyping by Sequencing' (GBS, Elshire *et al.* 2011), or 'Skim Sequencing' (Golicz *et al.* 2015). These technologies are also useful for detection of signatures of selection, or specific haplotypes carrying disease-causing mutations. In rare diseases with few cases available, and diseases with autosomal recessive inheritance, causative variants can be detected by Whole Genome Sequencing (WGS) of family trios where candidate variants with the expected genotypes are filtered bioinformatically. In a current project we use WGS of family trios to unravel mutations causing microphthalmia where one or both eyes are underdeveloped or missing. Once a candidate variant is detected it is crucial to validate the finding in a larger sample set, and by functional studies.

Today, the increased possibilities to test for a variety of disease-causing mutations, or performance-associated QTLs, support breeders in informed breeding decisions. Such genetic tests will indeed prevent production of disease-affected foals while unaffected carriers can still be kept in the breeding pool and maintain the genetic variation of the population. Large scale genotyping, or WGS is likely to replace today's genetic tests on individual or panels of mutations. Inevitably, this technology-driven efficient mutation identification will affect the future of horse breeding.

Genomics and the breeders' perspective

Despite the continuous increase in the amount of equine genomic data, genomic selection has not yet been fully implemented in horse breeding. The main reasons are high costs, difficulties to obtain large enough reference populations, and a high effective population size or short linkage disequilibrium in many horse breeds, compared with for example cattle where genomic selection has been applied for many years. For more than 20 years, a few microsatellites have been the standard markers for parentage analysis of horses, however SNP genotyping will most likely be implemented for parentage in the near future. Such a large

amount of genomic data from a huge pool of horses comprising many breeds around the globe, will open new possibilities to address research questions, and will make it possible to implement marker-assisted and/or genomic selection also in horses.

References

Ablondi M, *et al.* (2019a). Signatures of selection in the genome of Swedish Warmblood horses selected for sport performance. *BMC Genomics* 20:717:1-12. https://doi.org/10.1186/s12864-019-6079-1.

Ablondi, M., *et al.* (2019b). Genomic divergence in Swedish Warmblood horses selected for equestrian disciplines. *Genes* 10(12):E976. https://doi.org/10.3390/genes10120976.

Ablondi M., *et al.* (2022) Performance of Swedish Warmblood fragile foal syndrome carriers and breeding prospects. *Genet Sel Evol* 54(4):1-16. https://doi.org/10.1186/s12711-021-00693-4

Chen, L., *et al.* (2021). A novel DDB2 mutation causes defective recognition of UV-induced DNA damages and prevalent equine squamous cell carcinoma. *DNA Repair (Amst)* 97:103022:1-9. https://doi.org/10.1016/j.dnarep.2020.103022.

Drabbe, A., *et al.* (2022). Genome-Wide Association Analyses of Osteochondrosis in Belgian Warmbloods Reveal Candidate Genes Associated With Chondrocyte Development. *J Eq Vet Sci* 111:103870:1-6. https://doi.org/10.1016/j.jevs.2022.103870

Elshire, R.J., *et al.* (2011). A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS One* 6(5):e19379. https://doi.org/10.1371/journal.pone.0019379.

Golicz, A.A., *et al.* (2015) Skim-Based Genotyping by Sequencing. In: Jacqueline Batley (ed.), Plant Genotyping: Methods and Protocols, Methods in Molecular Biology, vol. 1245, Springer Science & Business Media, New York, USA. https://doi.org/10.1007/978-1-4939-1966-6_19.

Hu Z-L., and Reecy J.M. (2007). Animal QTLdb: Beyond a Repository - A Public Platform for QTL Comparisons and Integration with Diverse Types of Structural Genomic Information. *Mamm Gen,* 18(1):1-4. https://doi.org/10.1007/s00335-006-0105-8

Hu Z-L., *et al.* (2022). Bringing the Animal QTLdb and CorrDB into the future: meeting new challenges and providing updated services. *Nucl Acid Res,* 50(D1):D956–D961. https://doi.org/10.1093/nar/gkab1116.

Kalbfleisch, T.S., *et al.* (2018). Improved reference genome for the domestic horse increases assembly contiguity and composition. *Commun Biol* 1(197):1-8. https://doi.org/10.1038/s42003-018-0199-z

Leegwater, P.A., *et al.* (2016). Dwarfism with joint laxity in Friesian horses is associated with a splice site mutation in B4GALT7. *BMC Genomics* 17:839. https://doi.org/10.1186/s12864-016-3186-0.

Online Mendelian Inheritance in Animals, OMIA. Sydney School of Veterinary Science, 11 March 2022. Available at: https://omia.org/

Ricard, A., and Duluard, A. (2021). Genomic analysis of gaits and racing performance of the French trotter. *J Anim Breed Genet* 138(2):204-222). https://doi.org/10.1111/jbg.12526

Vosgerau, S., *et al.* (2022). Genetic and genomic characterization followed by single-step genomic evaluation of withers height in German Warmblood horses. *J Appl Genet.* Epub ahead of print. https://doi.org/10.1007/s13353-021-00681-w.

Wade, C.M., *et al.* (2009). Genome Sequence, Comparative Analysis, and Population Genetics of the Domestic Horse. *Science* 326:865-867. https://doi.org/10.1126/science.1178158

765. Microsatellite allele imputation from SNP genotypes for parentage verification in sport horses

H. Crichan[1*], C. Engler[1], E. Goulas[1], S. Dhorne-Pollet[2], M. Addes[1] and A. Ricard[1,2]

[1]Institut Français du Cheval et de l'Equitation, Pôle développement, innovation et recherche, La jumenterie, 61310 Exmes, France; [2]GABI, INRAE, AgroParisTech, Université Paris Saclay, 78350 Jouy en Josas, France; harmony.crichan@ifce.fr

Abstract

In the major livestock species, parentage verification is done using Single Nucleotide Polymorphism (SNP) markers. ISAG is working to produce a SNP panel for parental verification in horses. There will be a transition period between microsatellites (MS) and SNPs. Without any tool for the transition, breeders will have to double genotype foals or re-genotype mares and stallions with SNPs. We wanted to produce a tool to predict MS genotypes using SNP markers that could be included on a chip with the new ISAG panel. To achieve this we built a correspondence table between MS alleles and SNP haplotypes. We validated it using kfold cross-validation. With the highest SNP density, we obtained 3.4% of incompatibilities over 2 or more MS or 96.6% of parental verification achieved over the major breeds in our study, Selle-Français (SF, 54%), French Trotters (TF, 22%), Arabs (AR, 17%), Anglo-Arabs (AA, 3%) and other saddlebreds (SE, 4%).

Introduction

Nowadays, parental verification in horses is done using MS genotypes. ISAG is currently working on a new SNP panel for parentage testing that will soon replace the microsatellite panel. Lab technologies to produce microsatellite and SNP genotypes are different so genotypes of these markers cannot be obtained at once. The objective of our study is to predict MS genotypes from SNP haplotypes using commercial SNP chips to perform parental verification.

Materials & methods

Data. IFCE (Institut Français du Cheval et de l'Equitation), INRAe and Fonds Eperon funded research programs such as Genequin, JumpSNP, SelGenEqui, GenEndurance and SoGen that provided horses with SNP genotypes used in this study. MS genotypes of the SNP genotyped horses and their parents from routine parental verification came from SIRE (equine information database). Horses with incompatibilities with their dam or sire or occurring as a parent of an incompatible foal were removed from the reference population. There were 5,892 horses with a genotype over at least one known MS and with SNP genotypes. The breeds were mainly Selle-Français (SF, 54%), French Trotters (TF, 22%), and Arabs (AR, 17%), with few Anglo-Arabs (AA, 3%) and other saddlebreds (SE, 4%).

As the data came from different research programs, horses were genotyped with three different SNP chips: 54K (54,602 SNPs, 29%), 65K (65,157 SNP, 15%) and 670K (670,806 SNP, 56%). SF were mostly genotyped with the 670K SNP chip, half of the TF were genotyped with the 54K and the other half with the 670K, and AR mostly with the 65K.

By investigating the MS genotypes, we observed that three MS over the 14 were missing for most of the horses, we had genotypes for only 11.4% of the horses for ASB17, 11.4% for ASB23 and 10% for HSM2. Result were therefore obtained for the 11 MS remaining

Method – correspondence table building. The method was based on McClure *et al.* (2012) and McClure *et al.* (2013) who worked on cattle. SNPs located within 500,000 base pairs on each side of the MS were

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_765, © H. Crichan *et al.* 2022

selected. We then phased MS alleles together with SNP genotypes using Fimpute3 (Sargolzaei *et al.*, 2014) by converting MS alleles into pseudo-SNP (Karimi *et al.* 2018, Teissier *et al.* 2020). This transformation was useful because FImpute3 only works bi-allelic markers. From the imputation, we obtained SNP haplotypes and MS alleles alongside. A correspondence table was then created between MS alleles and SNP haplotypes. The table was built so that one MS allele could have multiple correspondences with SNP haplotypes but a SNP haplotype had to correspond to only one MS allele. A second rule was that only the closest SNP constituted the haplotypes. Eight to 40 SNP located on either side of the MS constituted the haplotypes.

Method – validation. Horses were randomly split in five groups of equal sizes; each group was in turn the validation set and the training set. Using the training set, we built the correspondence table. The MS genotypes of the validation set were predicted using the correspondence table only. The partitioning of the validation sample was repeated 10 times. We then compared the predicted MS alleles with the actual MS alleles information. We measured ineffectiveness using the percentage of MS genotypes attributed with the correspondence table that were different from the original MS genotypes. The second measure of ineffectiveness was the number of MS found incompatible using MS parental information.

Results

Genotypes description. Over the 5,892 horses, parental verification was achievable for 96.1% of them (5,661). Of the 5.9% of the 5,661 horses with at least one MS incompatible, 5.3% had one and 0.6% had two MS incompatible. MS had a variable number of alleles in our dataset, ranging from six to 12, and recruited SNPs around them varied from 111 to 314. 1,908 SNPs were used in this study, the vast majority were in the 670K SNP chip (1,902), and only 10% of them were in the 65K SNP chip.

Training and validation sets. Rates of false genotypes were 0.6% for MS HTG6 and 8.1% for MS HMS3 with a 32 SNPs long haplotype. With the k-fold validation, false genotypes varied with the MS considered and with the length of the haplotype with a threshold observed at 32 SNPs (Figure 1). The proportion of horses with more than one MS incompatible increased from 0.6% initially to 12.4% with a 32 SNP

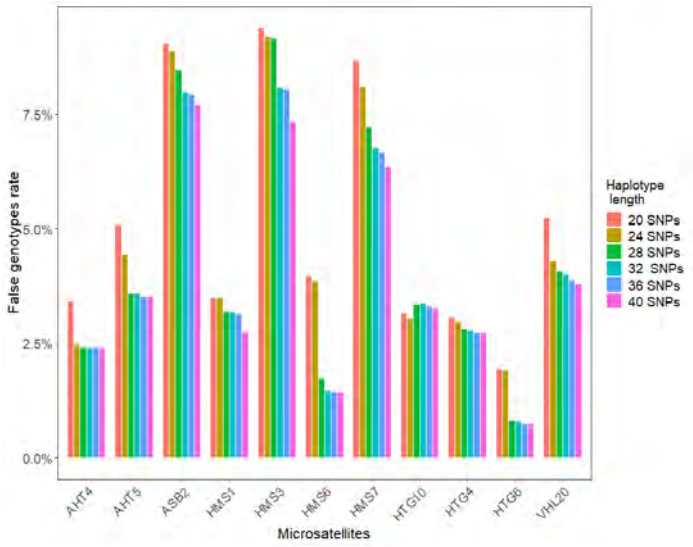


Figure 1. Average rates of false genotypes by MS and by haplotype length in the validation set (10-fold cross validation).

haplotype, the standard deviation of theses proportions between the random draws was 0.3%. For a 32 SNPs haplotype, we obtained 25% of one MS incompatible and 12% of 2 MS incompatible.

The error rate depended on the breed. Horses with more than one MS incompatible varied from 27.3% for AR to 7.3% for SF. A major difference between these breeds was the chip mainly used for the SNP genotyping.

As Figure 2 shows, the incompatibility differences were related to the density of the SNP chip used to genotype the horses. Only 3.4% of horses genotyped with the 670K SNP chip had more than one MS incompatible, whereas this percentage greatly increased to 32.1 and 21% with horses genotyped with the 65K SNP chip and the 54K SNP chip respectively. For horses genotyped with the 670K SNP chip, results were promising on the three breeds with only 4% for FR, 3.4% for AR and 3.2% for SF of horses with more than one MS incompatible.

Discussion

The method of SNP based MS imputation is recommended as a transition tool to save genotyping extra costs. There are some limitations regarding the results obtained from the correspondence table. First, there were three breeds in our study and we saw that the distribution of MS alleles differed across breeds; if we want to be able to predict MS alleles over all breeds, we will need more genotypes from different breeds to improve the corresponding table. This procedure may be difficult to manage and to automate, as some horses will need re-genotyping with MS. It will need to be accepted by breeders who are the primary clients for parental verification. If not, parental verification will need to be done by re-genotyping and/or double genotyping during the transition period. One of the questions is also the value of these results internationally. Horses that produce offspring or perform internationally will most likely be genotyped with both types of marker sets. Moreover, ISAG recommends only the re-genotyping in the first place as it was eventually done with cattle during their transition from MS to SNP. Finally, breeders' organizations will have to encourage their members to SNP genotype their horses so that the transition period is reduced to a minimum.

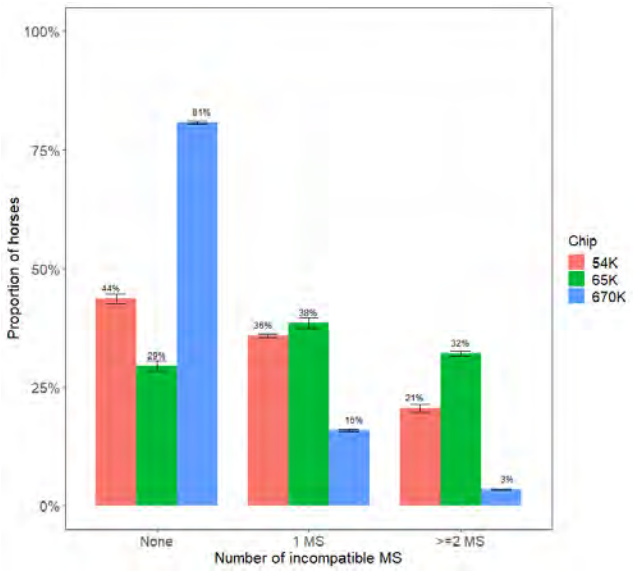


Figure 2. Average rates of incompatible MS by SNP chips in the validation set (10-fold cross-validation).

References

Karimi Z, Sargolzaei M, Robinson J.A.B, and Schenkel F.S. (2018). Can. J. Anim. Sci. 4:98. 750-759 https://doi.org/10.1139/cjas-2018-0009

McClure M, Sonstegard T, Wiggans G, Van Tassell CP. (2012) Front Genet. 3:140. https://doi.org/10.3389/fgene.2012.00140

McClure MC, Sonstegard TS, Wiggans GR, *et al.* (2013) Front Genet. 4:176. https://doi.org/10.3389/fgene.2013.00176

Sargolzaei M, Chesnais J.P, Schenkel F.S. (2014) BMC Genom. 15:478. https://doi.org/10.1186/1471-2164-15-478

Teissier M, Larroque H., Brito L.F., Rupp R., Schenkel F.S. et al (2020) JDS. 103:12:11559-11573 https://doi.org/10.3168/jds.2020-18662

766. Equine parentage control: bridging the gap from a microsatellite to a SNP based approach

W. Nolte[1,2], H. Alkhoder[3], M. Wobbe[3,4], K.F.S. Stock[3,4], E. Kalm[5], S. Vosgerau[5], N. Krattenmacher[5], G. Thaller[5], J. Tetens[6] and C. Kühn[1,7*]

[1]Institute of Genome Biology, Institute for Farm Animal Biology, 18196 Dummerstorf, Germany; [2]Saxon State Office for Environment, Agriculture and Geology, 01468 Moritzburg, Germany; [3]IT Solutions for Animal Production, 27283 Verden, Germany; [4]Institute for Animal Breeding and Genetics, University of Veterinary Medicine Hannover, 30559 Hanover, Germany; [5]Institute of Animal Breeding and Husbandry, Christian-Albrechts-University, 24118 Kiel, Germany; [6]Department of Animal Sciences, Georg-August-University, 37077, Göttingen, Germany; [7]Faculty of Agricultural and Environmental Sciences, University Rostock, 18059 Rostock, Germany; kuehn@fbn-dummerstorf.de

Abstract

In equine parentage control, the current common practice is the use of an internationally agreed upon microsatellite (MS) marker panel. In German horse breeding the goal is to transition to a SNP-based parentage control system. To facilitate the change from one method to another, an intermediate solution is the imputation of MS markers based on SNP genotypes. The feasibility and accuracy of such an imputation has been tested with a learning dataset of horses (n=2,607) from five different breeds. Horses have been genotyped for 11 to 17 MS markers, depending on studbook policies, and have been SNP-genotyped with an Illumina BeadChip (GGP Equine 70k or GGP Equine plus). The imputation has been conducted with Beagle 5. Results indicated that an average imputation accuracy 95.69% could be achieved for the 12 MS markers that belong to the mandatory panel of the International Society for Animal Genetics (ISAG).

Introduction

For more than two decades, the routine equine parentage control in Germany has been conducted via the microsatellite (MS) panel endorsed by the International Society for Animal Genetics (ISAG, https://www.isag.us/) (Bowling *et al.*, 1997). In the context of the planned introduction of genomic selection in German Warmblood horses by the International Association for Future Horse Breeding (IAFH), founded in 2017, a switch to a SNP based panel has become necessary. The advantage of using an SNP array lies in the diversity of its applications. Conditional on an appropriate number and selection of SNP markers, the information can be used not only for parentage control, but also for the determination of genomic breeding values for various traits. Compared to MS, SNPs have a better standardisation in genotyping and a lower mutation rate. At the same time, they are characterised by a lower cost per genotype, a higher throughput of data and an easier process to automate in the laboratory (McClure *et al.*, 2013).

In parentage control, it is essential that marker genotype information of the same category is available for both, parents and foals, i.e. MS or SNP information is required for both generations. To avoid double typing and the associated costs, time and effort in the future, foals should only be genotyped with a SNP array. As a consequence, it is necessary to obtain the MS information by imputation in order to be able to compare it with the parental generation. Imputation is the statistical determination, based on the SNP data, of the MS genotypes that an animal is most likely to carry. Before this process may be used routinely in practice, it is necessary to create a large learning dataset genotyped for both SNPs and MS. In cattle, the switch from the old MS system to a SNP panel has already been demonstrated and successfully applied (McClure *et al.*, 2012). The aim of this study was to impute MS genotypes in horses based on SNP genotypes and to assess the accuracy level of this approach in equines.

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_766, © W. Nolte *et al.* 2022

Materials & methods

The learning dataset used in this study consisted of 2,706 female horses that belong to the warmblood horse breeds Holsteiner (n=988), Trakehner (n=333), Westphalian (n=196), Oldenburg (n=926) and Oldenburg International (n=164). Hair samples have been collected by the studbooks during or as part of routine horse registration, identification or performance tests. Contract laboratories of the participating studbooks performed the MS genotyping according to their agreed upon panels, ranging from 11 to 17 MS markers. DNA extraction for the SNP genotyping was performed according to standard protocols at the Werlhof Institute (Hanover, Germany). Horses have been SNP genotyped with the Illumina GGP Equine 70k (n=715) or the GGP Equine plus (n=1,892) beadchip. All animals included here had a SNP call rate of ≥95%. SNPs were filtered for their minor allele frequency (≥0.01), a GC-Score (≥0.6), a call frequency (≥0.90), and Hardy-Weinberg-Equilibrium (≤0.001). MS genotypes and SNP genotypes were combined into one file in VCF-format. The imputation process was conducted with Beagle version 5.0 (Browning *et al.*, 2018). Default parameters were kept except for window size (200) and effective population size (30,000). All SNPs that are located on the same chromosome as the target MS were used for the imputation. In order to assess error rates of this method, a tenfold cross validation was performed in which 10% of the horses were masked for their MS genotype and imputed by means of the other 90% of the learning dataset. Ten percent of the horses from each of the five breeds were pooled to form a representative validation group. The imputed MS genotypes were then compared to the original (masked) data and scored for accuracy on a scale from zero to 100 as follows: 0 (both MS alleles incorrectly imputed compared to the original), 50 (one MS allele correctly imputed) and 100 (both MS alleles correctly imputed). Average accuracy scores were calculated per replicate, MS marker and breed. The error rate was calculated as 100 minus the respective imputation accuracy.

Results

The overall average imputation accuracy for all 17 MS markers was 96.17% (SD: 5.55), which corresponds to an error rate of 3.83% (Table 1). The error rates were higher in the set of mandatory markers (4.31%) than in the optional markers (2.68%). When evaluating the results separately for each breed, the Trakehner population had the lowest error rates with 2.40% for all 17 markers. The observation that the mandatory markers had a higher average error rate than the optional markers was made for each breed except for Westphalians. Overall, the group of Westphalian horses stood out because of comparatively high error rates for both mandatory markers (mean: 4.98%, SD: 6.37) as well as optional markers (mean: 22.45%, SD: 22.62). An assessment of the error rates on a MS marker level (Figure 1) revealed three markers that had high error rates in all breeds (ASB17, ASB23, AHT5) and three markers that only underperformed in Westphalian horses (HSM1, CA425, LEX3). An in-depth analysis of error rates and genotyping rates (GTR) for the MS markers showed that five out of the six problematic MS markers had low GTR (zero to 65%).

Table 1. Average imputation error rates for microsatellite markers by horse breed.

	Average imputation error rates in percent (%) for MS[1]-markers		
Horse breed	**17 markers**	**12 mandatory markers**	**5 optional markers**
all breeds	3.83	4.31	2.68
Holsteiner	3.32	4.47	0.58
Oldenburger	3.04	3.68	1.52
Oldenburg International	6.65	8.85	1.38
Trakehner	2.40	3.01	0.94
Westphalian	10.12	4.98	22.45

[1] MS = microsatellite.

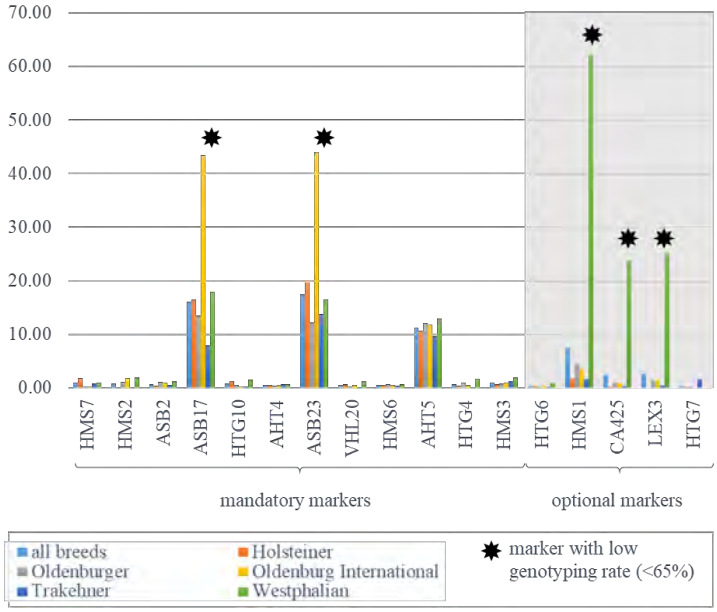


Figure 1. Imputation error rates in percent (%) for 17 equine microsatellite markers.

In Westphalians the GTR for HMS1, CA245, and LEX3 were 98.98, 22.96, and 22.45%, respectively. These markers had GTR below 1.5% in all other breeds. The markers ASB17 and ASB23 had GTR>98% in all breeds except for the Holsteiner horse, where this marker was only genotyped in single cases (GTR=0.20%.). As for AHT5, the GTR ranged between 99.40% (Trakehner) and 100% (all other breeds). An assessment of the chromosomal position of AHT5 showed that is marker is located at the telomeric end of the equine chromosome ECA8. Consequently, only eight SNPs were available for imputation upstream of the MS marker.

Discussion

Results from a ten-fold cross-validation scheme for a MS-marker imputation based on SNP-genotyped revealed that average imputation accuracies of 98% and beyond are feasible for certain markers. However, error rates increase when GTR are at a low level, especially when GTR are very heterogeneous between subgroups of a dataset. For Westphalian horses it could be observed, that the imputation accuracy for horses of this breed is greatly diminished when only few data points from horses of other breeds are also introduced, potentially representing haplotypes that are not present in the Westphalian group. Using an across breed imputation approach for markers of very unequal and low GTR cannot be recommended. We conclude that a high quantity of information leads to better imputation accuracies. However, optional markers of high GTR and high imputation accuracy, e.g. HTG6 and HTG7, are a valuable substitute for problematic markers such as AHT5. Compared to previous results in cattle, similar imputation accuracies were achieved within our study, particularly when considering MS markers with GTR above 90%. This is remarkable, especially because the reference panel in each replicate contained less than 2,500 animals, while McClure *et al.* (2014) used a reference panel of over 7,000 animals and reached accuracy levels of >98% in Irish cattle, which is similar to results in our study. With over 97% imputation accuracy, Marina *et al.* (2021) have also reported comparable magnitudes from Spanish Assaf sheep, following an experimental design, which was highly similar to the design in our study. Overall, our results can be considered highly encouraging and for a future routine application in practice.

References

Bowling A.T., Egglestone M.L., Byrns G., Clark R.S., De Leanis S. *et al.* (1997). Validation of microsatellite markers for routine horse parentage testing. Anim. Genet. 28: 247-252. https://doi.org/10.1111/j.1365-2052.1997.00123.x

Browning B. L., Zhou Y., Browning S. R. (2018). A one-penny imputed genome from next generation reference panels. Am. J. Hum. Genet. 103(3): 338-348. https://doi.org/10.1016/j.ajhg.2018.07.015

Marina H., Suarez-Vega A., Pelayo R., Gutiérrez-Gil B., Reverter A. *et al.* (2021). Accuracy of Imputation of Microsatellite Markers from a 50K SNP Chip in Spanish Assaf Sheep. Animals 11(1): 86. https://doi.org/10.3390/ani11010086

McClure M., Sonstegard T., Wiggans G., Van Tassell C. P. (2012). Imputation of microsatellite alleles from dense SNP genotypes for parental verification. Front. Genet. 3: 140. https://doi.org/10.3389/fgene.2012.00140

McClure M. C., Sonstegard T. S., Wiggans G. R., Van Eenennaam A. L., Weber K. L. *et al.* (2013). Imputation of microsatellite alleles from dense SNP genotypes for parentage verification across multiple Bos taurus and Bos indicus breeds. Front. Genet. 4: 176. https://doi.org/10.3389/fgene.2013.00176

McClure M., Treacy M., O'Connor K., McCarthy J., Kearney J.F. (2014). Proc. of 10th WCGALP, Vancouver, Canada.

Yang J., Lee S. H., Goddard M. E., Visscher P. M. (2011) GCTA: a tool for Genome-wide Complex Trait Analysis. Am. J. Hum. Genet. 88(1): 76-82. https://doi.org10.1016/j.ajhg.2010.11.011

767. Haplotype blocks and heterozygosity rich regions on ECA2 in Swedish Warmblood horses

M. Ablondi[1*], S. Eriksson[2] and S. Mikko[2]

[1]Department of Veterinary Science, Università degli Studi di Parma, 43126 Parma, Italy; [2]Dept. of Animal Breeding and Genetics, Swedish University of Agricultural Sciences, P.O. Box 7023, 750 07 Uppsala, Sweden; michela.ablondi@unipr.it

Abstract

The Fragile Foal Syndrome (FFS) is a disease caused by a recessive lethal missense mutation in the *PLOD1* gene (ECA2). Despite its harmful effect, a relatively high frequency of the FFS allele carriers has been found in the Swedish Warmblood (SWB), suggesting a heterozygous advantage. Balancing selection can be further studied in haplotype blocks and increased heterozygosity around the target of selection. In this study we evaluated the presence of haplotype blocks and Runs of Heterozygosity on ECA2 in 380 SWB horses divided by sport discipline. In horses mainly bred for dressage a haplotype block comprising the FFS variant was found. On average 11.7 Runs of Heterozygosity were identified per horse on ECA2, with no significant difference in numbers between the sport disciplines. This study provides a preliminary characterization of haplotype blocks and heterozygosity rich regions on ECA2 which may further corroborate the potential presence of balancing selection for the FFS allele.

Introduction

Fragile Foal Syndrome (FFS) is a monogenetic disease caused by a recessive lethal missense point mutation in the *PLOD1* gene located on ECA2. Most FFS recessive homozygous foals are assumed to be lost by abortion during late gestation, whereas liveborn affected foals show severe skin fragility, joint hyper-elasticity, and need to be euthanized (Aurich *et al.*, 2019). Despite its harmful effect, a relatively high frequency of FFS carriers (7.4%) was found in the Swedish Warmblood breed (SWB), suggesting the presence of a heterozygous advantage. The presence of balancing selection in SWB was corroborated in a recent study where significant positive effects of the FFS allele were found mainly on movement related traits (Ablondi *et al.*, 2022). Balancing selection causes various types of signatures in the genome such as: increased diversity around the target of selection, differentiation between populations departing from the same genome wide background, and increased linkage disequilibrium around the target of selection (Fijarczyk and Babik, 2015). The advances in high throughput genotyping procedures permit further investigation of the presence of such signatures within animal species. In this study we aimed to explore the presence of haplotype blocks and heterozygosity rich regions on ECA2 in SWB horses to investigate if there is balancing selection acting on the FFS variant or on another closely located variant.

Materials & methods

A total of 380 SWB horses born 2010-2011 were genotyped using the 670 K Affymetrix® Axiom® Equine Genotyping Array. These horses were also genotyped for the XM_001491331: c.2032G > A variant in the *PLOD1* gene by a TaqMan genotyping assay on a StepOnePlus™ instrument (Applied Biosystems, Foster City, CA, USA). For the scope of this work, only SNPs on ECA2 (52,075 SNPs) were retrieved. The quality control (QC) was performed in PLINK v1.90 (Purcell *et al.*, 2007), based on the criteria: minor allele frequency (MAF) (<0.05), missing genotype per single SNP (GENO) (>0.10) and missing genotype per individual (>0.10). Haplotype blocks were defined as in Gabriel *et al.* (2002), considering the region including the *PLOD1* gene ± 1 Mbp, and detected using the Haploview software.

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_767, © M. Ablondi *et al.* 2022

Runs of Heterozygosity (ROHet) were detected using the settings suggested by (Santos *et al.*, 2021) in the R package 'detectRUNS' (Biscarini *et al.*, 2019). ROHet within the ECA2 were considered as indicator of heterozygosity rich regions if they were shared by over 25% of the analysed horses. Haplotype blocks and ROHet detection was performed separately in two subgroups, defined as in Ablondi, *et al.* (2019), where horses with higher or lower EBV for show jumping than the reference population (mean EBV of 100) were classified as show jumping (SJ) and non-show jumping (NS) horses, respectively. Most NS horses had a dressage index above average. The Ensembl gene annotations EquCab3 was used to identify genomic elements extending 250 kb up- and downstream of the potential heterozygosity rich regions.

Results

A total of 31,122 SNPs and 380 horses passed the QC (189 NS and 191 SJ). Overall, 47 and 50 haplotypes were found in the NS and SJ horses, respectively, considering the region including the *PLOD1* gene ± 1 Mbp. These haplotype blocks were between 2 Kb to 115 Kb long. The haplotype with the greatest number of SNPs in the NS horses comprised 18 SNPs (40 Kb long) and was located 836 Kb upstream the FFS variant. For SJ horses, the highest number of SNPs within a haplotype was 19 (115 Kb long) and located 349 Kb downstream the FFS variant. In the NS, but not in the SJ, the FFS variant was in strong LD (D' equals to 1) with the closest downstream SNP (7 Kb apart), located within the *MFN2* gene (Figure 1).

On average 11.7 ROHet per horse were identified on ECA2, which was equal in both groups (SJ and NS horses). The ROHets had a length ranging between 12.6 Kb and 324 Kb and contained from 15 to 30 SNPs. On average ROHets were 69.5 Kb long and included 15.8 SNPs. A total of three ROHets were shared among

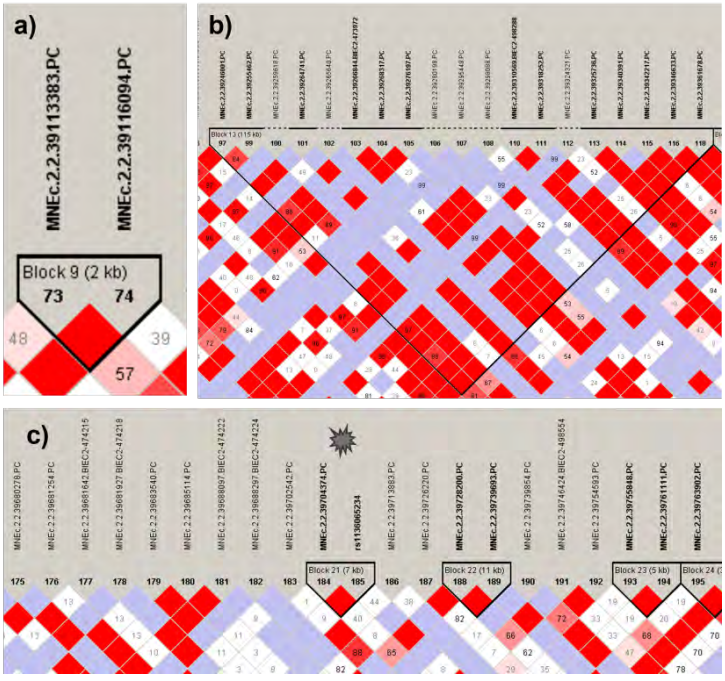


Figure 1. Examples of haplotype blocks in SWB within the *PLOD1* gene ± 1 Mbp: (a) haplotype block with minimum block size of 2 Kb, (b) haplotype block with maximum block size of 115 Kb in SJ horses, (c) haplotype block between FFS variant and the closest downstream SNP in NS horses which is highlighted with a grey star.

25% of the SWB horses: one in the SJ subgroups and two in the NS subgroup (Figure 2 and Table 1). In the case of NS horses, six lncRNA, five protein coding genes and one snRNA were found within the two heterozygosity rich regions shared in over 25% of the horses. In contrast in SJ, seven protein coding genes, two lncRNA and one miRNA were found.

Discussion

In NS horses, we found a haplotype block containing the FFS variant which was not detected in the SJ horses. A potential reason behind the failure to detect the haplotype in the SJ may be the lower frequency (5.8% in SJ vs 14.6% in NS) of the FFS variant in SJ horses as previously found by Ablondi *et al.* (2022) or that the selection pressure for heterozygosity in *PLOD1* is lower in SJ horses. In NS horses, the FFS variant was in strong LD with a SNP located within the Mitofusin 2 (*MFN2*) coding gene. The *MFN2* is involved in the regulation of vascular smooth muscle cell proliferation and has been found associated with disorders of the peripheral nervous system (Engelfried *et al.*, 2006), but is not known to be associated with joint

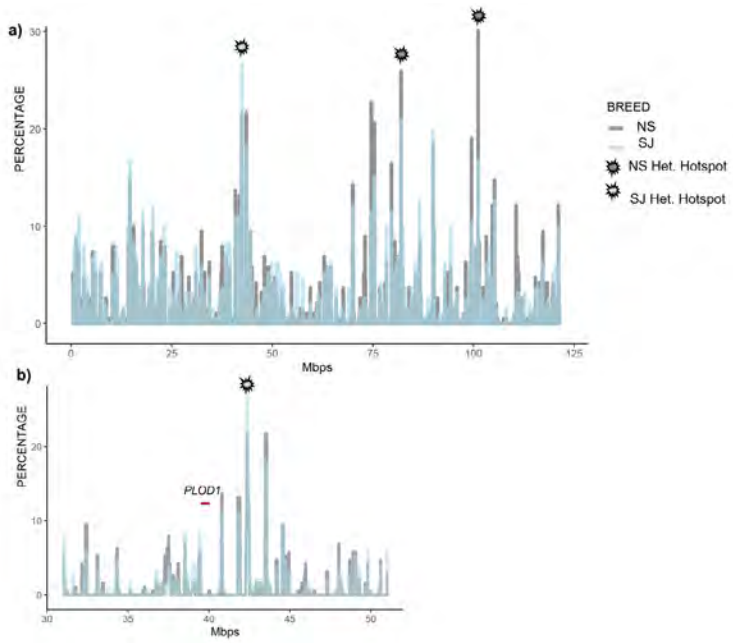


Figure 2. Proportion of times each SNP falls inside a ROHet in the two subgroups (SJ and NS) plotted against their position along the ECA2, with (a) showing the whole ECA2 and (b) showing a zoom-in representation on the *PLOD1* region ±10 Mb is shown.

Table 1. Heterozygosity rich regions on ECA2 shared in over 25% of the SWB horses divided into subgroups (SJ and NS) based on sport discipline.

Subgroup	Start (bp)	End (bp)	N. SNPs	Kb length	Percentage shared
NS[1]	81,932,632	81,972,128	16	39.5	25%
NS	101,054,766	101,088,692	12	33.9	32%
S[J]2	42,333,297	42,360,263	3	26.9	27%

[1] Non-show jumping SWB horses.
[2] Show jumping SWB horses.

elasticity or fragile skin. Interestingly, although for the *MNF2* SNP, the reference allele is the most common allele in the SWB, all FFS carriers horses among the NS horses carried the variant allele.

Compared to the length of runs of homozygosity (ROH) segments found in the SWB breed by Ablondi, *et al.* (2019) the ROHet investigated in this study were much fewer and detected more scarcely. This result was expected since polymorphic sites only account for a small portion of the whole genome. Nevertheless, the average number of ROHet found in the SWB ECA2 was more than three times higher compared to the one found in Mangalarga Marchador horse breed (Santos *et al.*, 2021). We did not find any clear signs of heterozygous rich regions next to the FFS variant. This can be expected if strong selection pressure is present around this locus, and heterozygosity advantage is only present for the FFS variant itself. We will further compare our findings on SWB ECA2 with the rest of the SWB genome to evaluate the potential role of balancing selection for the FFS variant.

References

Ablondi M., Eriksson S., Tetu S., Sabbioni A., Viklund Å., *et al.* (2019) Genes, 10(12), 976. https://doi.org/10.3390/genes10120976

Ablondi M., Johnsson M., Eriksson S., Sabbioni A., Gelinder Viklund Å., *et al.* (2022) *in press* Genet. Sel. Evol. https://doi.org/10.1186/s12711-021-00693-4

Ablondi M., Viklund Å., Lindgren G., Eriksson S., and Mikko S. (2019) BMC Genomics, 20(1), 717. https://doi.org/10.1186/s12864-019-6079-1

Aurich C., Müller-Herbst S., Reineking W., Müller E., Wohlsein P., *et al.* (2019) Anim. Repr. Sci., 211, 106202.

Biscarini F., Cozzi P., Gaspa G., and Marras G. (2019) detectRUNS: an R package to detect runs of homozygosity and heterozygosity in diploid genomes. Available at: https://cran.r-project.org/web/packages/detectRUNS/vignettes/detectRUNS.vignette.html#references

Engelfried K., Vorgerd M., Hagedorn M., Haas G., Gilles J., *et al.* (2006) BMC Med. Genet., 7(1), 53. https://doi.org/10.1186/1471-2350-7-53

Fijarczyk A., and Babik W. (2015). Mol.Ecol., 24(14), 3529–3545. https://doi.org/10.1111/mec.13226

Gabriel S.B., Schaffner S.F., Nguyen H., Moore J.M., Roy J., *et al.* (2002) 296(5576), 2225–2229. https://doi.org/10.1126/science.1069424

Purcell S., Neale B., Todd-Brown K., Thomas L., Ferreira M.A.R., *et al.* (2007). Am. J. Hum. Genet., 81(3), 559–575. https://doi.org/10.1086/519795

Santos W., Schettini G., Fonseca M.G., Pereira G.L., Chardulo L.A., *et al.* (2021) J. Anim. Breed. Genet., 138(2), 161–173. https://doi.org/10.1111/jbg.12508

768. Single step genomic evaluation for horses based on a multi-breed reference population

M. Wobbe[1,2*], H. Alkhoder[1], Z. Liu[1], S. Vosgerau[3], N. Krattenmacher[3], M. von Depka-Prondzinski[4], E. Kalm[3], R. Reents[1], W. Nolte[5], C. Kühn[5], J. Tetens[6], G. Thaller[3] and K.F. Stock[1,2]

[1]IT Solutions for Animal Production (vit), Heinrich-Schroeder-Weg 1, 27283 Verden, Germany; [2]University of Veterinary Medicine Hannover (Foundation), Institute for Animal Breeding and Genetics, Buenteweg 17p, 30559 Hanover, Germany; [3]Kiel University, Institute of Animal Breeding and Husbandry, Olshausenstr. 40, 24098 Kiel, Germany; [4]Werlhof-Institut MVZ, Schillerstr. 23, 30159 Hanover, Germany; [5]Leibniz Institute for Farm Animal Biology (FBN), Institute of Genome Biology, Wilhelm-Stahl-Allee 2, 18196 Dummerstorf, Germany; [6]University of Goettingen, DNTW Functional Breeding, Burkhardtweg 2, 37077 Goettingen, Germany; mirell.wobbe@vit.de

Abstract

The structure of the equine sector implies the need for collaborative approaches to implement genomic applications in sport horse breeding. Based on a multi-breed reference population of 5,000 horses from five breeds, a single step genomic evaluation system for linear conformation and performance traits was developed. Using two different validation approaches, forward and cross validation, medium (around 0.6) to high (over 0.9) correlations of genomic estimated breeding values (GEBVs) indicated stability of the developed system. Different relatedness between the breeds was reflected in the results, but the benefit of available genotypic data with regard to the variance of the GEBVs was found for each breed. According to the findings of this study, single-step genomic evaluation using a multi-breed reference population is a feasible approach for routine implementation of genomic selection in sport horses.

Introduction

The introduction of genomic selection in animal breeding can be referred to as paradigm shift (Meuwissen *et al.*, 2016). For horse breeding, this method has great potential because it allows shortening the long generation interval and integrating traits with low heritability or which are difficult to record into the breeding programs. Although this has been recognized already years ago (Haberland *et al.*, 2012, Mark *et al.*, 2014), there is until today only one studbook for sport horses with officially published genomic breeding values for a single trait (Knaap, 2017). The major challenges of successfully implementing genomic selection in horse breeding relate to definition of suitable target traits and compiling of a sufficiently large reference population. The structure of the equine sector implies particular interest in genomic systems which allow use of data across studbooks and breeds of riding horses: individual breeding populations are often small, but breeding goals are overlapping and exchange of genetic material is existing worldwide. The aim of this study was to validate the across breed approach for single step genomic evaluation and perform detailed distribution analyses of GEBVs with focus on possible breed differences.

Materials & methods

The data basis for developing a genomic evaluation system for equine conformation and performance traits was from five breeding organizations of German riding horses: Oldenburg (OL), Oldenburg International (OS), Trakehner (TRAK), Holstein (HOL), and Westphalian (WESTF). Phenotypes were available from routine linear description of the respective studbooks for which they used the same comprehensive linear scheme. The about 41,600 linear profiles of foals and adult horses, which could be used for this work, had been recorded between 2012 and 2019 in connection with regular evaluation events. To implement a single step genomic evaluation system, pedigree data of all phenotyped horses plus four ancestral generations

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_768, © M. Wobbe *et al.* 2022

were considered resulting in a relationship matrix of around 116,500 horses. The references population included 5,000 linearly described horses.

Genotyping was performed with commercially available single nucleotide polymorphism (SNP) arrays of medium density from Illumina (Illumina, 2015), GGPEquine (65,157 SNPs; 788 horses), and GGPEquinePlus (71,947 SNPs; 4,212 horses). After quality control with minor allele frequency of 0.01, call frequency of 0.1, call rate of 0.05 and a threshold of $P<0.001$ for Hardy Weinberg equilibrium test, n=4,964 horses and n=61,211 SNPs remained for analyses. The distribution of the different breeds in the dataset was: n=1,565 horses from OL, n=1,367 horses from HOL, n=829 horses from WESTF, n=778 horses from TRAK, and n=425 horses from OS. The genomic relationship between the five breeds included in the reference population was analyzed by principal component analysis (PCA plot, python version 3.8.1).

The 23 conformation and 18 performance-related linear traits were defined within age group and analyzed in pairs of analogous traits using the following linear animal models for foals (1) and adult horses (2).

$$Y_{ijklop} = \mu + SB_i + EVENT\text{-}TEAM_j + AGE_M_k + SEX_l + animal_o + e_{ijklop} \quad (1)$$

$$y_{ijmnop} = \mu + SB_i + EVENT\text{-}TEAM_j + AGE_Y_m + PTYP_n + animal_o + pe_o + e_{ijmnop} \quad (2)$$

where y is the phenotype; μ is the overall mean; SB is the fixed effect of the i^{th} studbook (i=1-5); EVENT-TEAM is the fixed effect of the combination of the event and team performing the linear description; AGE_M and AGE_Y are the fixed effects of the age of the foals in months (k=1-5; <1 month, 1-2 months, 2-3 moths., 3-4 months, >4 months) and of the adult horses in years (m=1-4; <3 years, 3-4 years, 4-5 years, >5 years); SEX is the fixed effect of sex for the foals (l=1-2; male female; distinct events for mares and stallions in the group of adult horses); PTYP is fixed effect of the presentation type for the adult horses (n=1-3; in hand, free, under saddle); animal is the random additive genetic effect; pe is the random permanent environmental effect; and e is the random residual effect. We used the software package MiX99 release XI/2019 version 19.1129 (Vuori *et al.*, 2006) for solving the equation model. GEBVs were standardized to a mean of 100 and a genetic standard deviation of 20.

The system was validated in two ways: Ten-fold cross-validation was performed by repeatedly choosing 15% of the genotyped and non-genotyped animals randomly per studbook, sex and trait as validation animals (VA); forward validation considered all horses linearly described in 2018 or later as VA. In each case, phenotypic data of the VA were excluded for the validation runs, and Pearson correlations between the GEBVs from the full run considering all available data and the validation run were determined (python version 3.8.1).

Distribution of the GEBVs were analyzed in more detail in the phenotyped horses born in 2007-2017 (approx. n=28,000). Descriptive statistics (mean, standard deviation, minimum, maximum) were generated for all horses and stratifying by studbook and availability of genotypic data.

Results

The principal component analysis of the genomic relationships among the five breeds in the reference population revealed three clusters. The first cluster comprised most of the horses from WESTF and OL, the second most of the horses from HOL and OS, and the third cluster included the horses from TRAK (Figure 1).

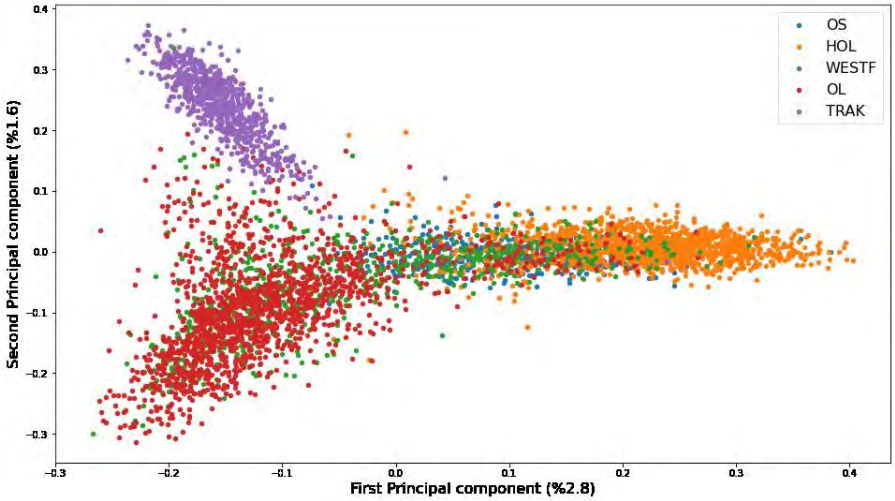


Figure 1. Principal component analysis of the n=4,964 horses in the reference population after quality check, showing the first two principal components.

The mean correlation coefficients between the GEBVs from the full run and the ten cross-validation runs were medium to high, with ranges of 0.60-0.95 for all VA and 0.86-0.99 for genotyped VA. Analyses of correlation coefficients for the genotyped VA by studbook revealed mostly minor differences, but with consistently higher figures for OL (0.88-0.98) and WESTF (0.86-0.99), followed by HOL (0.80-0.97) and OS (0.72-0.96), than for TRAK (0.63-0.94). The forward validation resulted in slightly lower GEBV correlations ranging between 0.56 and 0.89 for all VA and between 0.83 and 0.97 for the genotyped VA.

Analyses of the GEBV distributions showed clearly higher standard deviations in the genotyped horses than in the non-genotyped horses (on average +2.18 for the linear conformation traits and +2.30 for the linear performance traits). Similar results were obtained for each studbook with average differences between standard deviations of GEBVs in genotyped and non-genotyped horses ranging between +1.19 (TRAK) and +2.17 (OS) for conformation and between +1.30 (TRAK) and + 3.02 (OS) for performance traits. The standard deviations in all horses within studbook ranged between 8.68 (HOL) and 13.38 (TRAK) for conformation and between 7.84 (OS) and 11.47 (WESTF) for performance traits (Table 1).

Discussion

To meet the challenge of getting a sufficiently large reference population, German horse breeding organizations collaborated to assemble a 5,000 horse multi-breed training set for linear conformation and performance traits. The principal component analysis of the genomic relationships among the five included breeds reflected their relatedness. Different breeding focusses of studbooks on either of the two main disciplines of riding sport can explain the formation of two of the clusters, with one cluster including horses from studbooks with dressage as important breeding goal (WESTF, OL) and another cluster including horses from studbooks focusing on show-jumping (HOL, OS). The remaining cluster included TRAK horses, which is consistent with the breeding policy of a closed studbook where direct genetic links mainly exist through commonly used Thoroughbred or Arabian bloodlines and TRAK horses used in other studbooks. The expected closer relation of TRAK to the dressage cluster than to the cluster of studbooks with special focus on show-jumping was in accordance with other studies (Nolte *et al.*, 2019) also in our sample of genotyped horses.

Table 1. Means and standard deviations (std.) of GEBV across all performance traits (n=18) and all conformation traits (n=23) for the subset of phenotyped horses born between 2007-2017 by studbook and availability of genotypes.

	all	OL	HOL	WESTF	TRAK	OS
Conformation						
n	27,928	11,209	7,516	2,671	1,487	3,296
mean	100.19	100.73	99.95	100.06	100.15	99.26
std.	9.95	9.82	8.68	8.78	13.38	8.74
Conformation + genotype						
n	4,796	1,556	1,231	824	763	422
mean	100.28	101.02	100.10	100.10	99.80	99.31
std.	12.13	11.71	10.66	10.42	14.57	10.91
Performance						
n	28,560	11,343	7,599	2,874	1,497	3,356
mean	100.46	100.21	102.19	100.53	94.09	100.69
std.	11.27	9.93	8.07	11.47	11.01	7.84
Performance + genotype						
n	4,846	1,556	1,281	824	763	422
mean	100.12	100.67	102.22	101.22	93.67	101.18
std.	13.57	12.04	9.97	13.25	12.31	10.86

The reasonable variation of relatedness between the breeds included in the joint reference population implied the need of additional checks of the single step genomic evaluation system. Detailed analyses of the validation results by studbook showed lower GEBV correlations for TRAK than for the other contributing breeds. However, correlations were still in a range, indicating also for this studbook a satisfactory performance of the system for most of the linear target traits. In other species a multi-breed reference population has already proven to be a possible way for smaller populations to start with genomic selection if the respective breeds are related to each other (Lund *et al.*, 2014). This precondition is fulfilled for German warmblood horses. The analyses of the GEBV distributions showed in all breeds higher standard deviations for genotyped than for non-genotyped individuals indicating that genotyping benefits individual characterization of and differentiation between horses in a multi-breed single step genomic evaluation system for linear conformation and performance traits. With the transition to routine SNP genotyping with a medium density array as basis of parentage control, increasing proportions of genotyped horses will support future breeding progress in sport horse breeding.

References

Haberland A.M., König v. Borstel U., Simianer H., König S. (2012) Animal 6: 1369-1376

Illumina GGPEquine (2015) available at: https://www.illumina.com/

Knaap, J. (2017) 8. Pferde-Workshop, Bad Bevensen, Germany

Lund M. S., Su G., Janss L., Guldbrandtsen B., Brøndum R.F. (2014) Livest. Sci. 166:101-110

Mark T., Jönsson L., Holm M., Christianssen K. (2014) Proc of the 10th WCGALP, Vancouver, Canada.

Meuwissen T., Hayes B. and Goddard M. (2016) Anim. Front. 6(1):6-14.

Nolte W., Thaller G., Kühn C. (2019) PLoS ONE 14(4):e0215913.

Vuori K., Strandén I., Lidauer M. and Mäntysaari E. A. (2006) Proc. of the 8th WCGALP, Belo Horizonte, Brazil.

769. Methodology to integrate pedigrees of two Belgian Warmblood studbooks and its importance for genetic evaluation

L. Chapard*, N. Buys and S. Janssens

KU Leuven, Department of Biosystems, Center for Animal Breeding and Genetics, Kasteelpark Arenberg 30, 3001 Leuven, Belgium; lea.chapard@kuleuven.be

Abstract

Most of the Warmblood studbooks allow for exchanges of genetic material with other studbooks. However, unique horse identification numbers are not consistently used across them, and an approved horse may have multiple identification numbers. This might cause problems because genetic evaluations are based on the use of pedigree information. In particular, the development of an across-breed evaluation becomes inaccurate when horse identification numbers are not unique. In this study, two pedigrees are concatenated and a semi-automated method based on string comparisons is developed to obtain an integrated pedigree. In test runs of genetic evaluations for a competition trait, this integrated pedigree is compared with an uncorrected and concatenated version of the pedigree. This uncorrected version of the pedigree resulted in a slightly higher heritability and also in differences in estimated breeding values compared to the integrated pedigree.

Introduction

European Warmblood studbooks which aim to breed successful show jumping horses are genetically connected (Thorén Hellsten *et al.*, 2008; Ruhlmann *et al.*, 2009). Hence, exchanges of genetic material are very common across these populations, and many well-known breeding sires (and also mares) are represented in several studbook pedigrees, i.e. they have offspring in different countries (Rovere, 2016). As most of the European studbooks are open studbooks and have similar breeding objectives concerning jumping performance (Koenen *et al.*, 2004), the development of joint genetic evaluations across two or more studbooks was often discussed. However, to develop such an evaluation for several breeding organizations, unique identification numbers used across organizations are essential. Indeed, unique numbers are needed to connect pedigrees and link phenotypes to individuals. For this, the horse identification number (Unique Equine Life Number, UELN) was created (as of 1999), but in practice they are often not unique across and even within studbooks. For example, once a foreign horse is approved for breeding in a studbook, often a new identification number is created which is different from its original one. So, when it comes to working with pedigree data coming from different studbooks with genetically connected horses, having an integrated pedigree free of duplicates and of horses with multiple UELN becomes a tedious task. The aims of this study were to present a method to obtain an integrated pedigree by detecting horses with multiple UELN and deleting duplicates, and to assess the impact of an inaccurate pedigree on genetic evaluation by comparing estimated variance components and breeding values obtained with a good quality pedigree on the one hand and with a pedigree that contains horses with multiple UELN and duplicates on the other hand. These methods were applied here to two Belgian Warmblood studbooks: the Belgian Warmblood horse (BWP) and Zangersheide (Z).

Materials & methods

Pedigree data. To perform a genetic evaluation across the studbooks BWP and Z, a combined pedigree which contains both BWP and Z horses is required. Although many horses were common to both pedigrees, several had different UELN. So for the sake of the study, two pedigrees were created: a joined pedigree (p_joined) which is BWP (314,582 records) and Z (115,760 records) pedigrees concatenated without making any edits (i.e. horses with multiple UELN and duplicates were kept) and an integrated pedigree (p_integr)

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_769, © L. Chapard *et al.* 2022

which is a combination of BWP and Z pedigrees, corrected for horses with multiple UELN. To create these two pedigrees, BWP and Z pedigrees were concatenated (p_joined). Then, to obtain p_integr, corrections were made on p_joined by creating a cross-reference (translation table) containing all the variants of UELNs present in the pedigree. In this table, the UELN variants, the name of the horse and the UELN considered as the reference key were stored. Duplicates were detected by pairwise comparisons based on string comparisons of horses' names and names of their parents. Those comparisons were performed with the R package RecordLinkage (Sariyar and Borg, 2020), and the records with a probability of match above 0.90 were kept. A check of the duplicates found via the pairwise comparisons was performed semi-manually by looking at the horse's name, its birthdate, its gender and its parents' names using a custom R script (R Core Team, 2020). Briefly, records with no matches were removed from this dataset, the confirmed matches were kept and a choice of the reference UELN of each horse was manually made and included in the translation table. Variants of UELNs in the pedigree were replaced by their reference UELN using the information from the cross-reference table. Next, duplicates based on UELNs were removed from the pedigree. This resulted in a cleaned pedigree free of duplicates: p_integr.

Competition data. K.B.R.S.F (Royal Belgian Sport Equestrian) provided 2,436,461 show jumping competition records from 50,913 Belgian events which were held between 2004 and 2019. These records contained the sport number of the horse (a unique number given by K.B.R.S.F), its UELN, its sex, its age, its ranking, the competition height, the fault(s), the time, the rider number, the year and place of the competition. UELNs in the competition data were 'cleaned' by using the cross-reference table. Then, the following criteria were applied to edit the raw data: competition records that could not be linked to a horse via a UELN in the pedigree file were discarded and records of horses that were ≤3 years old or older than 18 years were removed. In total, 674,527 records were kept for the analysis. Rankings in competition were transformed in a normally distributed variable using the Blom transformation as described by Foran *et al.* (1995). In this study, success in competitions was defined as the variable 'corrected fence height' (CFH) which is a 'combination' of fence height and ranking in competition, using the Blom scores previously defined with a factor 1.32 (Chapard *et al.*, unpublished data):

$$\text{CFH} = \text{Fence height} + 1.32 \times \text{Blomscore} \quad (1)$$

Statistical model. The following model was used for estimating genetic parameters of CFH using a restricted maximum likelihood approach implemented in the program REMLF90 (Mitszal I. *et al.*, 2018):

$$y_{ijkl} = \mu + sex_i + age_j + animal_k + c_k + rider_l + e_{ijkl} \quad (2)$$

Where y_{ijkl} is the value of the trait CFH of the k^{th} animal; μ is the population mean; sex_i is the fixed effect of the horse's sex (i= 1, 2, 3); age_j is the fixed effect of the age (j=4,…, 18); $animal_k$ is the random additive genetic effect of the animal (k=1,…, 80,897 for p_integr; k=1,…, 82,719 for p_joined); c_k is the random permanent environmental effect, $rider_l$ is the random effect of the rider (l= 1,…, 8,410) and e_{ijkl} is the random residual effect.

Results

The integrated pedigree contained less horses (393,719) than the joined one (430,342) as duplicates and redundant registries of horses with multiple UELN were removed from p_joined. 54,092 horses had multiple UELN in p_joined and were added to the cross-reference. The number of UELNs detected per horse in the concatenated pedigree ranged from 1 to 12. Due to the integration step, the number of 'informative' horses decreased from 82,719 (p_joined) to 80,897 (p_integr). Table 1 shows the estimated variance components of CFH obtained with each pedigree (p_joined and p_integr). Estimated heritability of CFH was higher with

Table 1. Estimated variance components for jumping performance expressed as corrected competition height (CFH), using two pedigrees of different quality (p_joined and p_integr).

	Variance				Share in the variance[1]		
	Residual	Rider	Permanent environment	Genetic	Rr^2	c^2	h^2
p_joined	47.03	103.10	25.74	28.40	0.51	0.13	0.14
p_integr	47.03	102.10	28.46	24.35	0.51	0.14	0.12

[1] Standard errors on the ratio's varied from 0.0048 to 0.0064.

p_joined (h^2=0.14) than with p_integr (h^2=0.12). Also a difference regarding the permanent environmental effect for CFH was seen: it was 0.13 with p_joined, whereas c^2 was slightly higher with p_integr (0.14). However, the proportion of the rider effect (Rr^2) for CFH was similar with the two pedigrees p_joined and p_integr. Therefore, an inaccurate pedigree led, in this case, to a slight over-estimation of the heritability of the trait of interest and a slight under-estimation of the permanent environmental effect. The estimation of the rider effect was not impacted.

Estimated breeding values (EBVs) of horses were also impacted by an inaccurate pedigree. EBVs of different subsets of animals were studied: all horses, horses with own performance(s) and stallions that had at least 5 offspring. Correlations between EBVs obtained with p_integr and EBVs obtained with p_joined were 0.91 for all horses, 0.97 for horses with own performance(s) and 0.94 for stallions with at least 5 offspring (Figure 1). The highest correlation was found for horses with own performance(s).

Differences were also seen in the ranking of the 10 best stallions with at least 5 offspring (Table 2). Eight stallions out of ten are common between the EBV-based rankings obtained with p_joined and p_integr. However, not all of them got the same rank with p_joined and p_integr, implying that a published index obtained with a non-edited pedigree would differ from an index obtained with an accurate one.

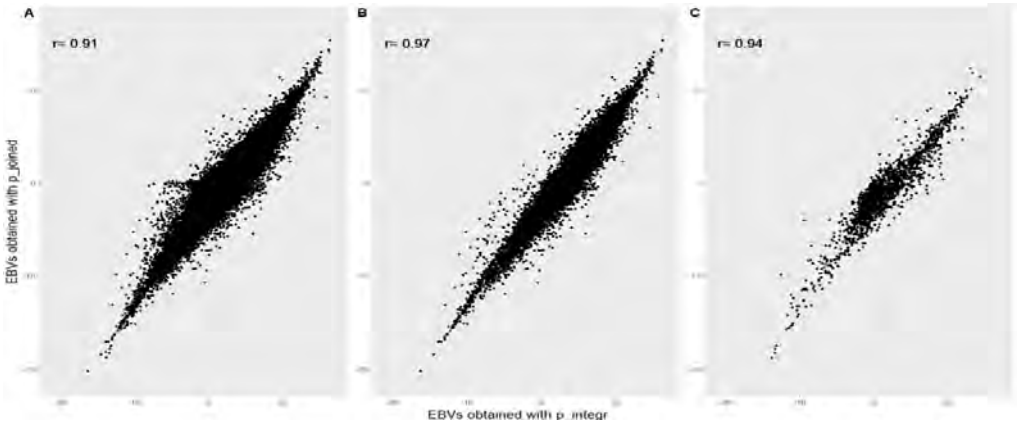


Figure 1. Relationships between EBVs obtained with two pedigrees (p_joined and p_integr). (A) all horses; (B) horses with own performance(s); (C) stallions with ≥5 offspring.

Table 2. Ranking of stallions with at least 5 offspring based on their EBVs.

Ranking stallions	CFH (p_integr)	CFH (p_joined)
1	Stallion 1	Stallion 2
2	Stallion 2	Stallion 1
3	Stallion 3	Stallion 6
4	Stallion 4	Stallion 4
5	Stallion 5	Stallion 3
6	Stallion 6	Stallion 11
7	Stallion 7	Stallion 5
8	Stallion 8	Stallion 21
9	Stallion 9	Stallion 9
10	Stallion 10	Stallion 10

Discussion

String comparisons of the names of horses and their parents proved to be a valuable tool to integrate pedigrees of different studbooks, although it was not always possible to detect matches. This was especially the case when the names of the horse were too different across studbooks and semi-manual editing was needed to obtain an integrated pedigree file. Hence, this method could be applicable to other horse populations and be an asset to develop across-breed evaluations. This study also showed that an inaccurate pedigree containing duplicates and animals with multiple identification numbers not only affected EBVs of individuals but also estimates of variance components which is in accordance with Senneke *et al.* (2004). EBVs can differ substantially between evaluations (accurate pedigree vs inaccurate pedigree) depending on the number of records of own performance linked to each UELN or/and numbers of offspring of each animal in the pedigree. Consequently, problems would certainly arise when publishing indices annually as there will be rank modifications caused by varying pedigree quality, and stallions may appear with multiple EBVs in the lists.

References

Foran M.K., Reilly M.P., Kelleher D.L., Langan K.W. and Brophy P.O. (1995). Proc. of the 46th Annual meeting of the EAAP, Prague, Czech Republic.

Koenen E.P.C., Aldridge L.I. and Philipsson J. (2004). Livestock Production Science. 88(1-2): 77-84. https://doi.org/10.1016/j.livprodsci.2003.10.011

Misztal I., Tsuruta S., Lourenco D. A. L., Masuda Y., Aguilar I., *et al.* (2018). Manual for BLUPF90 family programs. University of Georgia, Athens, USA.

R Core Team (2020). R: A language and environment for statistical computing. Available at: https://www.R-project.org/

Rovere G. (2016). Joint PhD thesis, Aarhus University, Aarhus, Denmark and Wageningen University, Wageningen, the Netherlands.

Ruhlmann C., Bruns E., Fraehr E., Philipsson J., Janssens S., *et al.* (2009). Livestock Science. 120(1-2):75-86. https://doi.org/10.1016/j.livsci.2008.04.016

Sariyar M., and Borg A., (2020). RecordLinkage: Record Linkage Functions for Linking and Deduplicating Data Sets. R package version 0.4-12.1. Available at: https://CRAN.R-project.org/package=RecordLinkage

Senneke S.L., MacNeil M.D. and Van Vleck L.D. (2004). Journal of Animal Science. 82(8):2307-2312. https://doi.org/10.2527/2004.8282307x

Thorén Hellsten E., Jorjani H., and Philipsson J. (2008). Livestock Science. 118(1-2):147-156. https://doi.org/10.1016/j.livsci.2007.12.028

Part 43

Porcine

770. Estimation of genetic parameters and trends of reproduction traits of Korean Landrace and Yorkshire pigs

S.S. Lee*, M. Alam, H.K. Chang and T.J. Choi

Animal Breeding and Genetics Division, National Institute of Animal Science (NIAS), P.O. Box 31000, Cheonan-si, South Korea; genemap@korea.kr

Abstract

The study aimed to investigate the genetic parameters of the reproduction traits of Korean Landrace and Yorkshire breeds. Three reproduction traits, namely age at first farrowing (AFF), total number of piglets born (TNB) and number of piglets born alive (NBA), were analyzed. Genetic parameters were estimated from individual datasets using a multiple-trait animal model in AIREMLF90 software. The heritability values for AFF, TNB and NBA were 0.07-0.14, 0.09-0.11 and 0.09-0.10, respectively. Strong genetic correlations existed between TNB and NBA (0.90 to 0.96). The genetic responses were identified no distinct trends for TNB-EBVs in Landrace and Yorkshire pigs prior to 2014, after which they displayed noticeable upward trends.

Introduction

The Korean pig breeding industry has been developed by small and medium sized breeding farms, with variable population sizes and inbreeding frequencies (Seo *et al.* 2011). Korean purebred pigs are generally selected according to their reproduction traits in order to optimize commercial pig performance. It is crucial to assess the genetic parameters of these traits to understand their genetic potential. From 2012, the National Institute of Animal Science (NIAS) has organized a 'National swine genetic improvement network' with 17 breeding farms and implemented a nationwide pig breeding program for sire lines of Duroc and dam lines of Landrace and Yorkshire. The aims of this program are selection of superior breeding stock suitable for Korean breeding objectives and increasing the genetic connectedness across the breeding farms by allowing exchange of genetic resources.

Materials & methods

Animals, phenotypes, and pedigree. Data on three reproduction traits (age at first farrowing, AFF; total number of piglets born, TNB; number of piglets born alive, NBA) were provided by the Korea Animal Improvement Association, and Korea Pork Producers Association. The dataset of the three reproduction traits included first-parity phenotypes from gilts only. Reproduction records of 11,758 Landrace and 40,171 Yorkshire gilts, born between 1992 and 2020, were analyzed. The AFF of all of these gilts ranged between 324 and 438 days. We also defined TNB as the total number of piglets born during farrowing, which included all of the normal, stillborn, mummified, deformed and underweight piglets. However, NBA only accounted for live piglets during farrowing and excluded both stillborn and mummified piglets. The pedigrees for reproduction traits comprised 37,538 and 15,483 animals that were Landrace and Yorkshire pigs, respectively. In all three breeds, the number of generations traced back in the pedigree for reproduction datasets was 26, respectively.

Statistical analyses. The (co)variance components of traits were estimated through multiple-trait models using the AIREMLF90 software package (Misztal *et al.* 2021). The animal models for reproduction traits included a similar fixed effect of the herd, birth year and birth season (HYS) and a fixed effect of farrowing year-season, irrespective of breeds. Each model also included the random additive genetic effects of dams. We further obtained the approximated standard error (SE) of genetic parameter estimates from (co)variance components using the AIREMLF90 option (se_covar_function), which uses a Monte Carlo method for the

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_770, © S.S. Lee *et al.* 2022

computation of SE, as proposed by Houle and Meyer (2015). The ratio of a trait's genetic standard deviation and its mean were expressed as the additive genetic coefficient of variation (CV_g), according to Houle (1992). We also obtained the genetic trend in traits using the averaged estimated breeding value (EBV) of animals based on their birth year.

Results

Table 1 presents the descriptive statistics of AFF, TNB and NBA for Korean Landrace and Yorkshire pigs. Landrace pigs showed the lowest average AFF value of 362.73±23.08 days. Despite some differences in average AFF, its CV values were smaller and similar across breeds. For TNB, the average of 12.07±3.40 piglets in Yorkshire was the highest. For the same trait, Landrace showed an intermediate mean to the other two breeds. Variations in TNB ranged between 27.60 and 28.13%. As for TNB, the mean NBA of 11.04±3.18 piglets was the highest in Yorkshire pigs.

Table 2 shows the variance components (genetic and residual), heritability (h^2) and genetic correlation (r_g) estimates of AFF, TNB and NBA traits. The h^2 value for the AFF trait 0.07±0.02 in Landrace and 0.12±0.01 in Yorkshire pigs. Genetic variation in AFF was relatively low, i.e. CV_g: 1%–2%. For TNB, our h^2 values in all pigs were low, ranging between 0.10 and 0.11. The h^2 of NBA was low in both breeds. The AFF-CV_g value was 8% in both breeds. For AFF and TNB, Yorkshire showed low, negative r_g values, such as -0.01, which were also not significantly different from zero. Landrace pigs exhibited a low and negative but slightly stronger association: -0.23±0.14. Landrace pigs also displayed a negative r_g between AFF and NBA (-0.21±0.14), in which another breed showed no significant association. On the other hand, the association between TNB and NBA was positive and strong in all pigs, i.e. 0.94 to 0.96.

Table 1. Descriptive statistics for reproduction traits in Korean Landrace and Yorkshire pigs.[1]

Breed	Trait	n	Mean	SD	Minimum value	Maximum value	CV
Landrace	AFF	11,758	362.73	23.08	324	438	6.36
	TNB	11,758	11.53	3.18	1	20	27.60
	NBA	11,757	10.63	2.98	0	20	28.02
Yorkshire	AFF	40,171	368.82	22.95	324	438	6.22
	TNB	40,171	12.07	3.40	1	20	28.13
	NBA	40,167	11.04	3.18	0	20	28.85

[1] AFF = age at first farrowing (day); TNB = total number of piglets born; NBA = number of piglets born alive; n = total sample size; SD = standard deviation; CV = coefficient of variation.

Table 2. Estimates of heritability (diagonals) and genetic correlations (upper diagonal), phenotypic correlation (lower diagonal) of reproduction traits with standard error in Landrace and Yorkshire pigs.[1]

Breed	Trait	Genetic parameter estimates			Genetic variance	Residual variance	CV_g
		AFF	TNB	NBA			
Landrace	AFF	0.07±0.02	-0.21±0.14	-0.23±0.14	21.08	261.40	0.01
	TNB	-0.01±0.01	0.10±0.02	0.94±0.02	0.87	7.83	0.08
	NBA	-0.03±0.01	0.90±0.00	0.10±0.02	0.77	7.08	0.08
Yorkshire	AFF	0.12±0.01	-0.01±0.06	-0.04±0.06	37.88	277.00	0.02
	TNB	0.00±0.01	0.11±0.01	0.96±0.01	0.99	8.35	0.08
	NBA	0.00±0.01	0.90±0.00	0.09±0.01	0.77	7.75	0.08

[1] AFF = age at first farrowing; TNB = total number of piglets born; NBA = number of piglets born alive; CV_g = genetic coefficient of variation.

The genetic response to AFF-EBVs was mainly inconsistent in all breeds, with no apparent visual trends. Similarly, we identified no distinct trends for TNB-EBVs in Landrace and Yorkshire pigs prior to 2014, after which they displayed noticeable upward trends. Moreover, Yorkshire TNB-EBVs were slightly higher than Landrace TNB-EBVs after 2002. With the two breeds, their NBA-EBV trends were similar to their TNB-EBV trends, even though Yorkshire pigs mostly performed better than Landrace pigs. Figure 1 shows the trends in animal estimated breeding value (EBVs) for NBA in Korean pig.

Discussion

AFF was found to be lowly heritable (0.07 to 0.14) across Korean breeds in our analysis. The low h^2 of AFF could be due to the lower genetic variability in Korean pigs. We suggest that necessary precautions be taken while considering the present AFF estimate for breeding decisions. We also found low heritability of TNB in Korean breeds. Using a single-trait repeatability model, Ogawa *et al.* (2019) obtained an h^2 of 0.12 for TNB in Landrace and Large White pigs, implying consensus. The NBA had low heritability in this study. In this respect, Ogawa *et al.* (2019) corroborated our study by reporting an h^2 of 0.12 in Landrace and 0.10 in Large White pigs. Ye *et al.* (2018), using multi-parity records in Large White pigs, also showed a lower NBA h^2 (0.06) in their pigs. A faster genetic improvement in any of the reproduction traits is not expected in Korean pigs due to their lower heritability values. Therefore, Korean pig breeding strategies should focus on increasing animals' genetic variabilities to achieve desirable selection responses, especially for the AFF trait.

In Korean pig breeds, the genetic association estimates of AFF with both TNB and NBA were lowly negative. However, most of these estimates were not significantly different from zero, except for Landrace pigs. This suggests that selection of Landrace pigs against AFF could increase TNB or NBA to some extent. Moreover, the non-significant correlation of AFF with TNB or NBA in Yorkshire pigs suggest no selection improvements in these breeds. The correlation between TNB and NBA was 0.90 or higher in our work. For this reason, both TNB and NBA could be assumed to have almost similar genetic bases, and an improvement in TNB could have subsequent desired effects on NBA.

For AFF, we observed no noticeable trend in EBVs in Landrace and Yorkshire pigs. No direct selection for AFF could explain. Another reason could be the genetic parameters of AFF, such as the low heritability, meaning greater environmental influences, and the lack of genetic correlations with TNB or NBA. Since 2014, we have also observed consistent improvements in TNB and NBA as they are directly targeted for reproduction improvement. According to recent statistics (KAIA, 2020), the average annual import between

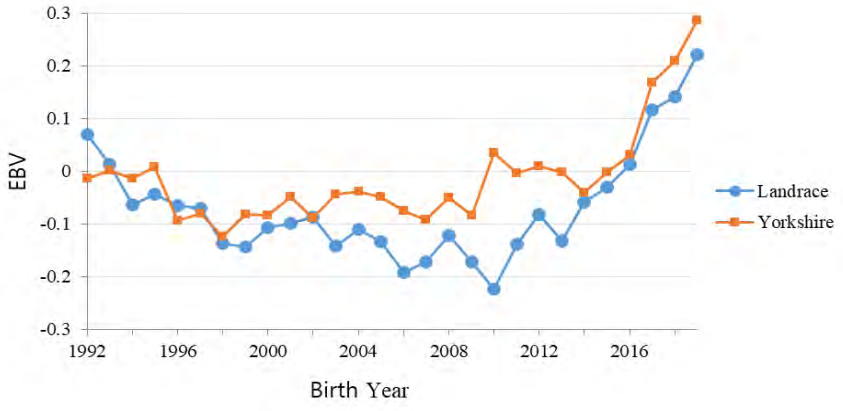


Figure 1. Trend of estimated breeding value (EBV) of NBA in Korean pigs.

2007 and 2010 was 1412 pigs, which increased by more than 4-fold between 2011 and 2012. Between 2011 and 2020, the average annual import was also reported as more than 2-fold greater than in previous years. In conclusion, most of our estimates suggested further improvement possibilities through selection. We argue that current genetic parameters could greatly assist future breeding and selection decisions for performance improvements in Korean pigs for specific purposes.

References

Houle, D. (1992) Genetics 130;195–204.

Houle, D., Meyer, K. (2015) Estimating sampling error of evolutionary statistics based on genetic covariance matrices using maximum likelihood. J. Evol. Biol. 28;1542–1549. https://doi.org/10.1111/jeb.12674

KAIA (Korea Animal Improvement Association) Annual Report (2020) Seoul, Korea, p. 156.

Misztal, I., Tsuruta, S., Lourenco, D., Aguilar, I., Legarra, A., Vitezica, Z.Z. Manual for BLUPF90 Family of Programs. Available at: https://nce.ads.uga.edu/wiki/lib/exe/fetch.%0Aphp?media=blupf90_all2.pdf.

Ogawa S., Konta A., Kimata M.; Ishii K., Uemoto Y. and Satoh M. (2019) Anim. Sci. J. 90;23–28. https://doi.org/10.1111/asj.13120

Seo J.H., Shin J.S., Noh J.K., Song C.E., Do C.H. (2011) Journal of Animal Science and Technology 53;397-408. https://doi.org/10.5187/ JAST.2011.53.5.397

Ye J., Tan C., Hu X., Wang A. and Wu, Z. (2018) Genetic parameters for reproductive traits at different parities in Large White Pigs. J. Anim. Sci. 96;1215–1220, https://doi.org/10.1093/jas/sky066

771. Genetic parameter estimates of heat tolerance in number of born alive in Landrace, Large White, and Duroc pigs

T. Okamura[1*], Y. Fukuzawa[1], M. Nishio[1], A. Arakawa[1], K. Ishii[1], H. Takahashi[2], S. Ogawa[3] and M. Satoh[3]

[1]Institute of Livestock and Grassland Science, National Agriculture and Food Research Organization, 2 Ikenodai, Tsukuba, Ibaraki 305-0901, Japan; [2]Global Pig Farms, Inc., 800, Kamihakoda, Hokkitsu-machi, Shibukawa-shi, Gunma, 377-0052, Japan; [3]Graduate School of Agricultural Science, Tohoku University, 468-1 Aramaki Aza Aoba, Aoba-ku, Sendai, Miyagi, 980-8572, Japan; okamut@affrc.go.jp

Abstract

This study aimed to investigate a suitable environmental indicator of heat stress on the number of born alive (NBA) and estimate the genetic variance components of a reproductive trait that reacts to heat stress. The data consisted of 26,770, 25,973, and 12,544 litters of Landrace, Large White, and Duroc sows, respectively, from four farms in Japan. The daily maximum ambient temperature was used as an indicator of heat stress. The correlation between the NBA and heat stress was estimated to determine a suitable period for heat load. The random regression model was fitted with heat load as random effects. Tolerance to heat stress was a heritable trait, but the variance was much smaller than trait performance in a standard environment. The heritability and additive genetic variance decreased as the heat load increased, implying that a lower heat load would be preferable for selection in NBA.

Introduction

Heat stress is one of the most important concerns in the pig industry, especially as global temperatures are rising. Heat stress decreases the number of piglets born alive (NBA), as well as the number of litters per sow per year. Therefore, it is necessary to improve genetically-determined tolerance to heat stress. The main objectives of this study were to investigate a suitable environmental indicator for expressing heat stress on NBA and to estimate the genetic parameters for heat stress.

Materials & methods

Data. The data consisted of 6,241 Landrace (26,770 litters), 5,861 Large White (25,973 litters), and 3,604 Duroc (12,544 litters) individuals from one breeding company with four farms in Japan. NBA was defined as the number of piglets alive at farrowing. Meteorological data were obtained from the Agro-Meteorological Grid Square Data, NARO, which can estimate the temperature and relative humidity for each 1-km square grid (Ohno *et al.*, 2016; https://amu.rd.naro.go.jp/). For each farrowing record, we obtained the daily maximum temperature (from days 140 to 0) and farrowing cycle of sows before the date of farrowing.

Statistical analysis. To identify the days in which heat stress during the reproduction cycle of the sow had the largest impact for each breed, NBA data were analysed for each breed using a three-step approach. First, the data were corrected for systematic and genetic effects using airemlf90 (Misztal *et al.*, 2018) with the following model:

$$y_{ijklmn} = parity_i + farm_j + year_k + cross_l + ss_m + a_n + pe_n + e_{ijklmn} \quad (1)$$

where y_{ijklmn} is the phenotypic value of NBA; $parity_i$ is the fixed effect of parity i (nine classes; one to eight and over nine); $farm_j$ is the fixed effect of farm j (four classes); $year_k$ is the fixed effect of the year of farrowing k (31 classes); $cross_l$ is the fixed effect of the type of crossing (two classes: pure or cross); ss_m is the random effect of service sire m; a_n is the random effect of the additive genetic effect of individual n; pe_n is

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_771, © T. Okamura *et al.* 2022

the random effect of the permanent environmental effect of individual n; and e_{ijklmn} is the random residual term. The following assumptions regarding random effects were made:

$$\begin{pmatrix} \boldsymbol{a} \\ \boldsymbol{pe} \\ \boldsymbol{ss} \\ \boldsymbol{e} \end{pmatrix} \sim N\left[\begin{pmatrix} \boldsymbol{0} \\ \boldsymbol{0} \\ \boldsymbol{0} \\ \boldsymbol{0} \end{pmatrix}, \begin{pmatrix} \sigma_a^2\boldsymbol{A} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \sigma_{pe}^2\boldsymbol{I} & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \sigma_{ss}^2\boldsymbol{I} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \sigma_e^2\boldsymbol{I} \end{pmatrix}\right]$$

where A is the additive relationship matrix, I is the identity matrix, and σ_a^2, σ_{pe}^2, σ_{ss}^2, and σ_e^2 are the additive genetic, permanent, environmental, service sire, and residual variances, respectively. The adjusted NBA records were obtained as estimated by e_{ijklmn}.

Second, to identify the day during the reproduction cycle of the sow heat stress that had a maximum influence on NBA performance, pairwise Pearson correlations between the adjusted observations and daily maximum temperature at each day of the cycle of the sow from 140 days before the day of farrowing were estimated.

Third, to identify the period (sequential days) during which the reproduction cycle of the sow heat stress had a maximum impact on NBA performance, pairwise Pearson correlations between adjusted NBA records and average daily maximum temperature in the interval centred on the estimated day of mating (114 days before farrowing) were estimated. The most impactful period that has the highest absolute value of the correlation coefficient was estimated for each breed.

To estimate the (co)variance components for the NBA and heat tolerance, the following random regression model was used:

$$y_{ijklmno} = \beta\varphi_{1o} + parity_i + farm_j + year_k + cross_l + ss_m + a_{0n} + a_{1n}\varphi_{1o} + pe_{0n} + pe_{1n}\varphi_{1o} + e_{ijklm} \quad (2)$$

where β is the fixed regression coefficient for the environmental covariate at the population level, φ_{1o} is the average daily maximum temperature during the best period for each breed as a covariate (standardised between 1 and 1) at value o, a_{0n} are a_{1n} random regression coefficients for the additive genetic effect of individual n for the intercept and slope terms, respectively, and pe_{0n} are pe_{1n} random regression coefficients for the permanent environmental effect of individual n for the intercept and slope terms, respectively. The following assumptions regarding random effects were made:

$$\begin{pmatrix} \boldsymbol{a_0} \\ \boldsymbol{a_1} \\ \boldsymbol{pe_0} \\ \boldsymbol{pe_1} \\ \boldsymbol{ss} \\ \boldsymbol{e} \end{pmatrix} \sim N\left[\begin{pmatrix} \boldsymbol{0} \\ \boldsymbol{0} \\ \boldsymbol{0} \\ \boldsymbol{0} \\ \boldsymbol{0} \\ \boldsymbol{0} \end{pmatrix}, \begin{pmatrix} \boldsymbol{G}\otimes\boldsymbol{A} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{P}\otimes\boldsymbol{I} & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \sigma_{ss}^2\boldsymbol{I} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \sigma_e^2\boldsymbol{I} \end{pmatrix}\right], \ \boldsymbol{G} = \begin{pmatrix} \sigma_{a0}^2 & \sigma_{a01} \\ \sigma_{a01} & \sigma_{a1}^2 \end{pmatrix}, \boldsymbol{P} = \begin{pmatrix} \sigma_{pe0}^2 & \sigma_{pe01} \\ \sigma_{pe01} & \sigma_{pe1}^2 \end{pmatrix}$$

where σ_{a0}^2 and σ_{a1}^2 are the additive genetic variances for the intercept and slope terms, respectively; σ_{a01} is the covariance between the two effects; σ_{pe0}^2 and σ_{pe1}^2 are the permanent environmental variances for the intercept and slope terms, respectively; and σ_{pe01} is the covariance between the two effects. Variance components were estimated using remlf90 (Misztal *et al.*, 2018). After obtaining the random-regression variance components, an additive genetic (co)variance matrix among the values of was calculated as follows:

$$\Gamma_a = \Phi'G\Phi, \Gamma_{pe} = \Phi'P\Phi$$

where G is the same as in model (2), and Φ is a matrix of the number of rows equal to the number of unique values of φ_1 in model (2) and two columns (a vector of 1 and φ_{1o}). The heritability of each value of o was calculated as follows:

$$h^2_{oo} = \frac{\Gamma_{aoo}}{\Gamma_{aoo} + \Gamma_{peoo} + \sigma^2_{ss} + \sigma^2_e}$$

where Γ_{aoo} and Γ_{peoo} are the o[th] values of the diagonal of Γ_a and Γ_{pe}, respectively.

Results

The relationship between the adjusted NBA and the daily maximum temperature on a given day of the reproduction cycle is shown in Figure 1. For Landrace and Large White pigs, the absolute values of the correlation coefficients were highest around the estimated mating date (114 days before farrowing). In contrast, Duroc had negative correlation coefficients throughout the gestation period. To estimate correlations between adjusted NBA and average daily maximum temperature in the interval centred on the estimated mating date, the period between 130 days to 110 days, 130 days to 80 days, and 130 days to 0 days before farrowing had the highest absolute values of correlation in Landrace, Large White, and Duroc, respectively. The average daily maximum temperature over these periods could be the best indicator of heat load for each breed. Table 1 shows the variance components estimated for the NBA and heat tolerance. The additive genetic variance for tolerance to a given heat stressor is much smaller than that for trait performance in standard environments.

Figure 2 shows the estimates of heritability, additive genetic variance, permanent environmental variance, and phenotypic variance across the range of the average daily maximum temperature. Although these were constant for Duroc, the heritability and additive genetic variance decreased as the heat load increased for Landrace and Large White populations.

Discussion

Heat stress around the mating date strongly affects the reproductive performance of pigs and this pattern is consistent with a previous study of the total number of births (Bloemhof *et al.*, 2013). Previous studies (Silva *et a.*, 2014;) have reported that the tolerance component was slightly less than that in the standard environment, which is not consistent with our results. It is possible that the populations selected in hot regions have already acquired tolerance to heat stress and, consequently, the variability is small. The decrease of heritability and additive genetic variance as the heat load increased consistent with the previous studies (Tizzie *et al.*, 2020), which implied that selection in low heat load (comfortable to sows) is suitable for NBA.

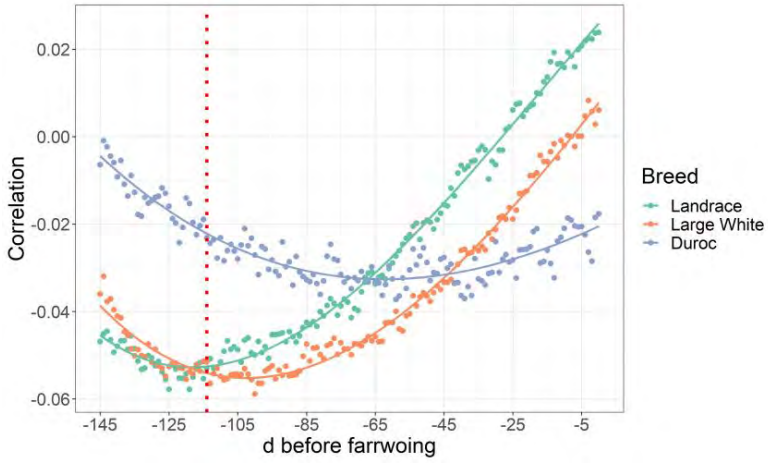


Figure 1. Correlations between adjusted number of born alive and daily maximum temperature.

Table 1. Variance components estimated for number of born alive and heat tolerance.

Components[1]	Landrace	Large White	Duroc
σ_{a0}^2	1.194	1.442	0.823
σ_{a01}	-0.167	-0.056	-0.029
σ_{a1}^2	0.160	0.104	0.022
σ_{pe0}^2	0.667	0.657	0.799
σ_{pe01}	0.154	0.061	-0.080
σ_{pe1}^2	0.094	0.145	0.038
σ_{ss}^2	0.262	0.289	0.116
σ_e^2	7.647	7.395	5.191

[1] See materials & methods section.

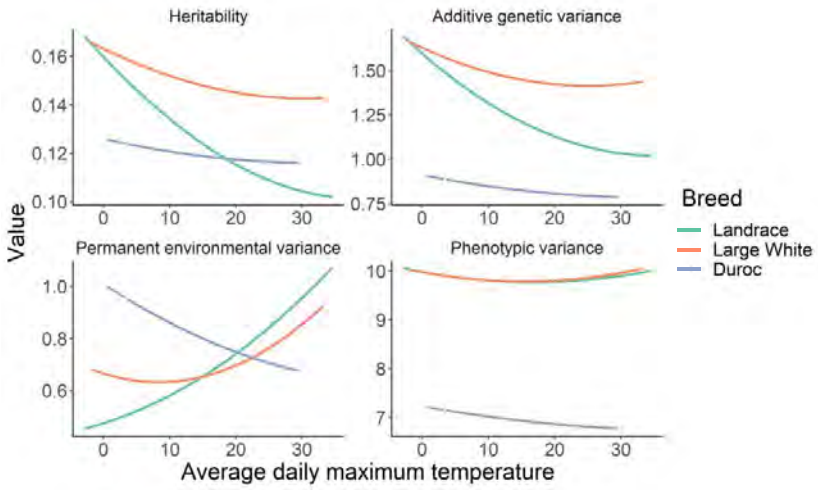


Figure 2. Estimates of heritability and variance components across the range of average daily maximum temperature.

References

Bloemhof, S., Mathur, P. K., Knol, E. F., and Van der Waaij, E. H. (2013) J. Anim. Sci. 91(6): 2667–2679. https://doi.org/10.2527/jas.2012-5902

Ohno, H., Sasaki, K., Ohara, G., Nakazono, K. (2016) Clim. in Bio. 16: 71-79. (in Japanese with English title) https://doi.org/10.2480/cib.J-16-028

Misztal, I., Tsuruta, S., Lourenco, D. A. L., Masuda, Y., Aguilar *et al.* (2018) Manual for BLUPF90 family programs. Available at: http://nce.ads.uga.edu/wiki/doku.php?id=documentation

Silva, F. F., Mulder, H. A., Knol, E. F., Lopes, M. S., Guimarães, S. E. F. *et al.* (2014) J. Anim. Sci. 92(9): 3825-3834. https://doi.org/10.2527/jas.2013-6486

Tiezzi, F., Brito, L. F., Howard, J., Huang, Y. J., Gray, K. *et al.* (2020) Front. genet. 11: 629. https://doi.org/10.3389/fgene.2020.00629

772. Bayesian analysis of paternal and maternal gametic effects in a reciprocal cross between two Iberian varieties

H. Srihi[1*], N. Ibáñez-Escriche[2], J. Casellas[3], J.L. Noguera[4], P. Hernández[2], M. Martín de Hijas[3], M. Vázquez-Gómez[3], S. Negro[5], J.P. Rosas[5] and L. Varona[1]

[1]Facultad de Veterinaria, Instituto Agrolimentario de Aragón (IA2). Universidad de Zaragoza. 50013 Zaragoza, Spain; [2]Universitat Politècnica de València, 46071 Valencia, Spain; [3]Universitat Autònoma de Barcelona, 08193 Barcelona, Spain; [4]Genètica i Millora Animal, IRTA, 25198 Lleida, Spain; [5]Programa de Mejora Genética 'Castúa', INGA FOOD S.A. (Nutreco) 06200 Almendralejo, Spain; houssemsrihi@unizar.es

Abstract

INGA FOOD, S. A. company has developed a crossbreeding scheme between two Iberian varieties (Retinto and Entrepelado) to generate a F_1 hybrid sow. A previous study has detected differences in the reproductive performance (litter size) across the two reciprocal crosses, suggesting the presence of imprinting effects. In this study, we developed a multivariate gametic model able to estimate the gametic correlations among the paternal and maternal effects generated from any of the two genetic origins. Gametic correlations significantly lower than one would explain the differences in performance across reciprocal crosses. The results indicated that, despite the small amount of available information, the estimates of the gametic correlation between paternal and maternal effects are smaller in the Entrepelado than in the Retinto populations. This would explain the differences in performance among the reciprocal crosses.

Introduction

The Iberian pig breed is recognized for their perfect adaptation to the 'Dehesa' environment and the high quality of its products (Serra *et al.*, 1998), Traditionally, the Iberian pig production was developed with purebred varieties and extensively. However, in recent decades, many traditional production systems have been replaced by intensive farms using crossbreeding with Duroc sires to improve growth and efficiency at final product (Serrano *et al.*, 2008). Regulatory norms of Iberian pig production allow this cross, as long as the sow is Iberian. The reproductive performance of the Iberian sows is lower than in white pig populations (Silio *et al.*, 2001). INGA FOOD, S. A. company developed a crossbreeding scheme between two Iberian varieties (Retinto and Entrepelado) to generate a F_1 hybrid sow associated with a relevant heterosis (Noguera *et al.*, 2019). However, that study identified strong differences in performance between both reciprocal crosses (Entrepelado × Retinto –ER- vs Retinto × Entrepelado –RE-). Therefore, the aim of this study is to develop a multivariate gametic model able to estimate the paternal and maternal gametic contributions of the Retinto and Entrepelado varieties in the ER and RE crosses and the correlations among them.

Materials & methods

Phenotypic and genomic data. The phenotypic data used in this study consisted of number of piglets born alive -NBA- from 206 Entrepelado × Retinto and 134 Retinto × Entrepelado sows. A summary of the data is presented in Table 1.

Genotypes for all sows were generated using the GeneSeek® GPP Porcine 70K HDchip (Illumina Inc., USA). The filtering consisted of excluding genotypes with a minor allele frequency lower than 0.05 and a SNP call rate less than 0.90 in the overall population. Subsequently, 33,875 SNP markers were used to calculate the paternal and maternal phases using the AlphaPhase software (Hickey *et al.*, 2011).

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_772, © H. Srihi *et al.* 2022

Table 1. Number of records (and sows between brackets), mean and standard deviation of the traits for Entrepelado × Retinto and Retinto × Entrepelado crosses.[1]

	Entrepelado × Retinto	Retinto × Entrepelado
N (NS)	1,258 (206)	700 (134)
NBA	8.77±2.24	8.84±2.37

[1] N = number of records; NS = number of sows; NBA = number born alive.

Statistical model. Data were analyzed with the following model:

$$\begin{bmatrix}\mathbf{y}_{ER}\\ \mathbf{y}_{RE}\end{bmatrix}=\begin{bmatrix}\mathbf{X}_{ER} & \mathbf{0}\\ \mathbf{0} & \mathbf{X}_{RE}\end{bmatrix}\begin{bmatrix}\mathbf{b}_{ER}\\ \mathbf{b}_{RE}\end{bmatrix}+\begin{bmatrix}\mathbf{W}_{ER} & \mathbf{0}\\ \mathbf{0} & \mathbf{W}_{RE}\end{bmatrix}\begin{bmatrix}\mathbf{s}_{ER}\\ \mathbf{s}_{RE}\end{bmatrix}+\begin{bmatrix}\mathbf{Z}1_{ER} & \mathbf{0}\\ \mathbf{0} & \mathbf{Z}1_{RE}\end{bmatrix}\begin{bmatrix}\mathbf{p}_{E}\\ \mathbf{p}_{R}\end{bmatrix}+\begin{bmatrix}\mathbf{Z}2_{ER} & \mathbf{0}\\ \mathbf{0} & \mathbf{Z}2_{RE}\end{bmatrix}\begin{bmatrix}\mathbf{m}_{\mathbf{R}}\\ \mathbf{m}_{R}\end{bmatrix}+\begin{bmatrix}\mathbf{e}_{ER}\\ \mathbf{e}_{RE}\end{bmatrix}$$

where y_{ER} and y_{RE} are the vectors of phenotypic records, b_{ER} and b_{RE} are systematic effects that include order of parity with 5 levels (first, second, third, fourth and five or more) and herd-year-season with 25 levels, s_{ER} and s_{RE} are the permanent sow environmental effects, p_E and p_R and m_E and m_R are the gametic paternal (p) and maternal (m) effects of Entrepelado and Retinto, respectively, and e are the vector of residuals. Further X, W, Z1 and Z2 are the corresponding incidence matrices. The prior distribution of the permanent sow environmental effects is:

$$\begin{bmatrix}\mathbf{s}_{ER}\\ \mathbf{s}_{RE}\end{bmatrix}=N(\mathbf{0},\mathbf{I}\otimes\mathbf{H}) \text{ where } \mathbf{S}=\begin{bmatrix}\sigma^2_{sER} & 0\\ 0 & \sigma^2_{sRE}\end{bmatrix}$$

and the prior distributions of the gametic effects are:

$$\begin{bmatrix}\mathbf{p}_{\mathbf{E}}\\ \mathbf{m}_{\mathbf{E}}\end{bmatrix}=N(\mathbf{0},\mathbf{G}_{\mathbf{E}}\otimes\mathbf{V}_{\mathbf{E}}) \text{ and } \begin{bmatrix}\mathbf{p}_{\mathbf{R}}\\ \mathbf{m}_{\mathbf{R}}\end{bmatrix}=N(\mathbf{0},\mathbf{G}_{\mathbf{R}}\otimes\mathbf{V}_{\mathbf{R}})$$

where G_E and G_R are the gametic relationship matrices of the Entrepelado and Retinto populations computed as defined by Nishio and Satoh (2015):

$$\mathbf{V}_{\mathbf{E}}=\begin{bmatrix}\sigma^2_{pE} & \sigma_{pmE}\\ \sigma_{pmE} & \sigma^2_{mE}\end{bmatrix} \text{ and } \mathbf{V}_{\mathbf{R}}=\begin{bmatrix}\sigma^2_{pR} & \sigma_{pmR}\\ \sigma_{pmR} & \sigma^2_{mR}\end{bmatrix}$$

where σ^2_{pX} and σ^2_{mX} are the paternal and maternal variances of the X={E,R} population, and σ_{pmX} is the covariance between them. The analysis was performed using the *gibbsf90* software (Misztal *et al.*, 2018) using a single chain of 5,000,000 iterations after discarding the first 1,000,000.

Results

The results of the posterior entail (and standard deviation) estimates of the variance components are presented in Table 2.

In addition, the posterior distribution of the genetic correlation between paternal and maternal effects in the Entrepelado and Retinto populations are presented in Figure 1.

The posterior mean (and mode) was 0.101 (0.245) for Entrepelado and 0.239 (0.461) for Retinto, and the posterior probability of a positive gametic correlation was 0.613 and 0.755 for Entrepelado and Retinto, respectively.

Discussion

The proposed bivariate model allows to estimate the gametic effects of each founder population acting as paternal or maternal as well as their correlation. In absence of paternal or maternal imprinting, the

Table 2. Posterior mean (and standard deviation) estimates of variance components.

Variance components[1]	Population	
	Entrepelado × Retinto	Retinto × Entrepelado
σ_s^2	0.119 (0.086)	0.344 (0.185)
σ_p^2	0.218 (0.099)	0.229 (0.101)
σ_m^2	0.288 (0.145)	0.273 (0.133)
σ_e^2	4.270 (0.186)	4.660 (0.277)

[1] σ_s^2 = sow permanent environmental variance; σ_p^2 = paternal gametic variance; σ_m^2 = maternal gametic variance; σ_e^2 = residual variance.

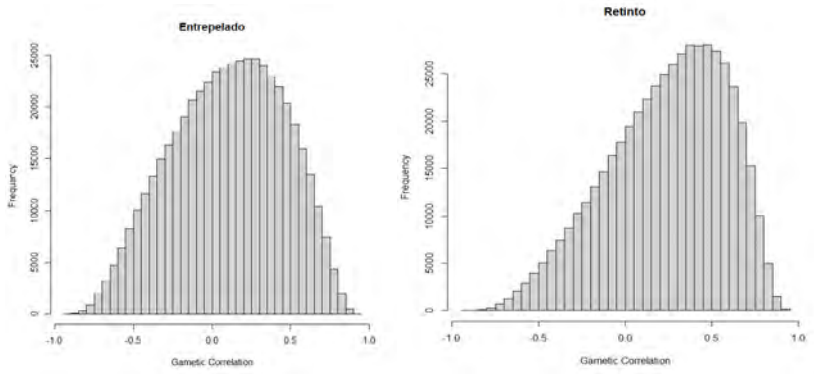


Figure 1. Posterior distributions of the correlations between paternal and maternal gametic effects in the Entrepelado and Retinto populations.

correlation among them should be perfect (one). A significant deviation from this perfect correlation would indicate differences in the performance of crossbred individuals depending on each of the population is acting as paternal (or maternal).

The results of the study are very preliminary, as the amount of phenotypic and genotypic information is very low. Differences in the posterior distribution of the correlation between gametic effects in Retinto and the Entrepelado varieties can be observed. The posterior distribution of the correlation in the Retinto variety is more positively skewed than the Entrepelado one. It suggests that the gametic contribution of Retinto individuals acting as boars or sows of the crossbred individuals are more similar, while the contribution of the Entrepelado individuals diverge greatly if they are used as boars or sows. This result may explain the differences in performance among the reciprocal crosses observed by Noguera *et al.* (2019) who propose using the Retinto variety as boar and the Entrepleado as sow providing a better performance that the opposite cross.

Acknowledgements

We thank the company Inga Food SA (Almendralejo, Spain) and its technicians (E. Magallón, L. Muñoz, P. Díaz, and D. Iniesta) for their cooperation and technical support. The research was partially funded by grants CGL2016-80155-R and PID2020-114705RB-I00 (Ministerio de Ciencia e Innovación), and IDI-

20170304 (CDTI). Srihi received funding from the European Union's H2020 research and innovation program under a Marie Sklodowska-Curie grant agreement, No. 801586.

References

Hickey J. M., Kinghorn B. P., Tier B., Wilson J. F., Dunstan N., and Van Der Werf J. H. J. (2011). Genetics Selection Evolution, 43(1), 1–13. https://doi.org/10.1186/1297-9686-43-12

Misztal I., Tsuruta S., Lourenco D., Aguilar I., Legarra A., and Vitezica, Z. (2018). Manual for BLUPF90 family of programs. Athens, USA.

Nishio M and Satoh M. (2015). Genetics Selection Evolution, 47(1), 1–10. https://doi.org/10.1186/s12711-015-0091-y

Noguera J. L., Ibáñez-Escriche N., Casellas J., Rosas J. P., and Varona, L. (2019). Animal. https://doi.org/10.1017/S1751731119001125

Serra X., Gil F., Pérez-Enciso M., Oliver M. A. Vázquez, J. M., *et al.* (1998). Livestock Production Science, 56(3). https://doi.org/10.1016/S0301-6226(98)00151-1

Serrano M. P., Valencia D. G., Nieto M., Lázaro R., and Mateos, G. G. (2008). Meat Science, 78(4). https://doi.org/10.1016/j.meatsci.2007.07.006

Silio L, Rodriguez M. C., Rodrigañez J., and Toro, M. A. (2001). La selección de cerdos ibéricos. In: Buxade, C., Daza A. (eds) Porcino Ibérico: Aspectos Claves, Mundi-Press, Madrid, Spain, pp. 125–149.

773. Heritabilities and genetic correlations of litter traits at farrowing and semen production traits in Duroc pigs

S. Ogawa[1*], M. Kimata[2], M. Tomiyama[2] and M. Satoh[1]

[1]Graduate School of Agricultural Science, Tohoku University, Sendai, Miyagi 980-8572, Japan; [2]CIMCO Corporation, Koto-ku, Tokyo 136-0071, Japan; shinichiro.ogawa.d5@tohoku.ac.jp

Abstract

We estimated heritabilities of semen production traits and litter traits and genetic correlations among them in purebred Duroc pigs. Semen production traits per ejaculate were semen volume, sperm concentration, proportion of morphologically normal sperms, total number of sperms, and total number of morphologically normal sperms. Litter traits at farrowing were total number born, number born alive, number still born, total litter weight, mean litter weight, and piglet survival rate. We used 45,913 records of 896 boars and 6,950 records of 1,400 sows. Estimated heritabilities ranged from 0.20 to 0.29 for semen production traits and from 0.06 to 0.25 for litter traits. Estimated genetic correlations between semen production traits and litter traits were weak, except for those of NSB and SVB with CON, NUM, and NUMF absolute values of which were higher than 0.3 in this study.

Introduction

Genetic parameters of semen production traits have been estimated (e.g. Wolf, 2010; Brinke *et al.*, 2020), but information on their genetic relationship with other economically important traits is limited. In particular, no study has reported the genetic parameters of semen production traits in Japanese pig population. Heere, we estimated the heritabilities and genetic correlations of semen production traits and litter traits at farrowing in a Japanese purebred Duroc pig population.

Materials & methods

Approval of Animal Care and Use Committee was not required for this study because the data were acquired from an existing database.

Phenotype and pedigree data. We analysed 45,913 semen production records of 896 Duroc boars and 6,950 farrowing records of 1,400 Duroc sows provided by CIMCO Corporation (Tokyo, Japan), which operates great-grandparent (GGP) and grandparent (GP) farms by Specific Pathogen Free system in northern to southern parts of Japan, were analysed. Pedigree data covered 67,993 animals. Sows were serviced by artificial insemination (typically three times using semen from the same boar).

Semen production traits per ejaculate were semen volume (VOL), sperm concentration (CON), proportion of morphologically normal sperms (PROP), total number of sperms (NUM) = VOL × CON, and total number of morphologically normal sperms (NUMN) calculated as NUM × PROP. Litter traits at farrowing were total number born (TNB), number born alive (NBA), number still born (NSB), total litter weight at birth (LWB), mean litter weight at birth (MWB) = LWB ÷ NBA, and piglet survival rate (SVB) = NBA ÷ TNB.

Numerical analyses. Genetic parameters were estimated, according to Wolf (2010) and Ogawa *et al.* (2019), using the following repeatability animal model:

$$y = Xb + Za + Wpe + e \quad (1)$$

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_773, © S. Ogawa *et al.* 2022

where y is the vector of phenotypic records; b is the vector of fixed effects; a is the vector of breeding values; pe is the vector of permanent environmental effects; e is the vector of random errors; and X, Z, and W are the design matrices. The vector b for semen production traits contains the effects of year at ejaculation (18 levels: 2000 to 2017), month at ejaculation (12 levels), farm (6 levels: two GGP and four GP farms), age at ejaculation (43 levels: 6 to 48 months), and interval between present and the previous semen collections (13 levels: up to 3, 4, …, 14 days, and ≥15 days). The vector b for litter traits contains the effects of farrowing year (18 levels: 2001 to 2018), farrowing season (4 levels: spring [March to May], summer [June to August], autumn [September to November], winter [December to February]), farm (6 levels: two GGP and four GP farms), and dam parity (10 levels: 1st to 9th and ≥10th).

Heritability was estimated by single-trait analysis. Genetic correlation was estimated by two-trait analysis. Non-zero covariances between two semen production traits and between two litter traits were assumed for all random effects. Non-zero covariance between semen production and litter traits was assumed for only for breeding values. Variance components were estimated using AIREMLF90 software (Misztal *et al.*, 2002). Standard errors were calculated according to Klei and Tsuruta (2008).

Results

Estimated heritabilities of semen production traits ranged from 0.20 for PROP to 0.29 for VOL, equal to or greater than those for litter traits, which ranged from 0.06 for SVB to 0.25 for MWB (Table 1). Estimated repeatabilities of semen production traits ranged from 0.41 for CON to 0.49 for VOL, greater than those of litter traits, which ranged from 0.08 for SVB to 0.32 for LWB.

Among semen production traits, estimated genetic correlations of VOL were positive with NUM and NUMN (0.48 and 0.51) and negative with CON (-0.66) (Table 2). That of PROP was positive with VOL (0.18) and negative with CON (-0.33). Those among litter traits agreed with those in Landrace and Large White populations of the same company (Ogawa *et al.*, 2019). Those between semen production and litter traits were weak in many cases, except for those of NSB and SVB with CON, NUM, and NUMF absolute values of which were higher than 0.3.

Table 1. Descriptive statistics of phenotypic measurements and estimated heritabilities (h^2) and repeatabilities (*rep*2) of traits studied.

Trait; abbreviation	Unit	Mean	SD	h^2	*rep*2
Semen volume; VOL	ml	126.7	56.8	0.29	0.49
Sperm concentration; CON	10^8 counts/ml	5.81	2.36	0.28	0.41
Proportion of morphologically normal sperm; PROP	-	0.89	0.05	0.20	0.46
Total number of sperm; NUM	10^8 counts	660.3	245.6	0.23	0.45
Total number morphologically normal sperm; NUMF	10^8 counts	588.3	221.5	0.22	0.45
Total number born; TNB	number	9.8	2.6	0.12	0.23
Number born alive; NBA	number	9.1	2.4	0.11	0.22
Number still born; NSB	number	0.7	1.0	0.06	0.09
Total litter weight at birth; LWB	kg	14.1	3.6	0.17	0.32
Mean litter weight at birth; MWB	kg	1.58	0.25	0.25	0.27
Survival rate at birth; SVB	-	0.94	0.09	0.06	0.08

Table 2. Estimated genetic correlations and their standard errors (below and above diagonal).[1]

Traits	Traits										
	VOL	CON	PROP	NUM	NUMN	TNB	NBA	NSB	LWB	MWB	SVB
VOL	-	0.07	0.12	0.10	0.10	0.15	0.15	0.17	0.13	0.12	0.17
CON	-0.66	-	0.11	0.11	0.11	0.13	0.13	0.15	0.12	0.11	0.16
PROP	0.18	-0.33	-	0.13	0.13	0.14	0.15	0.16	0.14	0.12	0.16
NUM	0.48	0.27	-0.12	-	0.00	0.14	0.15	0.16	0.14	0.12	0.17
NUMN	0.51	0.23	-0.02	0.99	-	0.15	0.15	0.17	0.14	0.12	0.17
TNB	0.09	-0.26	0.12	-0.19	-0.18	-	0.01	0.13	0.06	0.11	0.15
NBA	0.12	-0.20	0.18	-0.12	-0.09	0.96	-	0.16	0.07	0.11	0.16
NSB	-0.10	-0.34	-0.17	-0.39	-0.44	0.46	0.21	-	0.13	0.12	0.02
LWB	0.25	-0.31	0.18	-0.05	-0.03	0.71	0.65	0.48	-	0.08	0.14
MWB	0.16	-0.14	0.04	0.08	0.09	-0.12	-0.25	0.38	0.57	-	0.11
SVB	0.19	0.31	0.22	0.40	0.46	-0.20	0.06	-0.96	-0.27	-0.43	-

[1] See Table 1 for abbreviations of trait names.

Seasonal variations in phenotypic records were observed. Figure 1 shows examples of semen production traits; values of VOL and NUM were lower in hotter months (especially August to October). This might reflect in part the effect of heat stress on boars before semen collection.

Figure 2 shows the effects of farrowing season on NBA in different pig breeds; values were lower in autumn and winter, probably reflecting in part the effect of heat stress on sows just before service by artificial insemination.

Discussion

Previous studies using repeatability animal models estimated the heritability of semen production traits to range from 0.1 to 0.3, higher than those of TNB and NBA in different Landrace and Large White populations (Wolf, 2010; Brinke *et al.*, 2020). Genetic improvement of semen production traits, therefore, seems to be possible. As far as we know, this is the first study to estimate the genetic correlations between

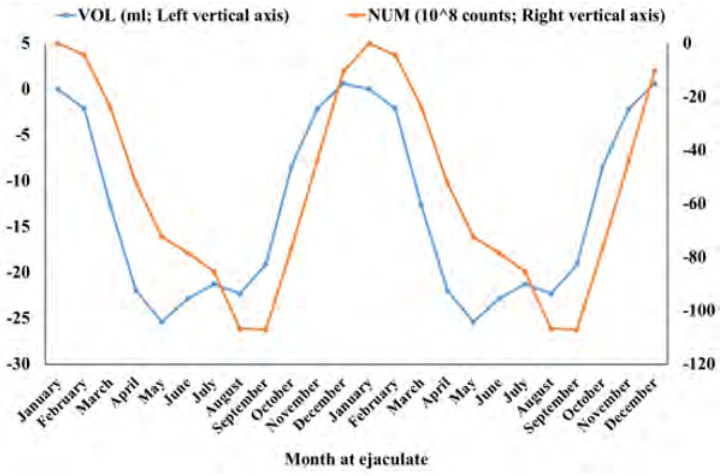


Figure 1. Estimated values of effects of month at ejaculate on semen volume (VOL) and total number of sperm (NUM).

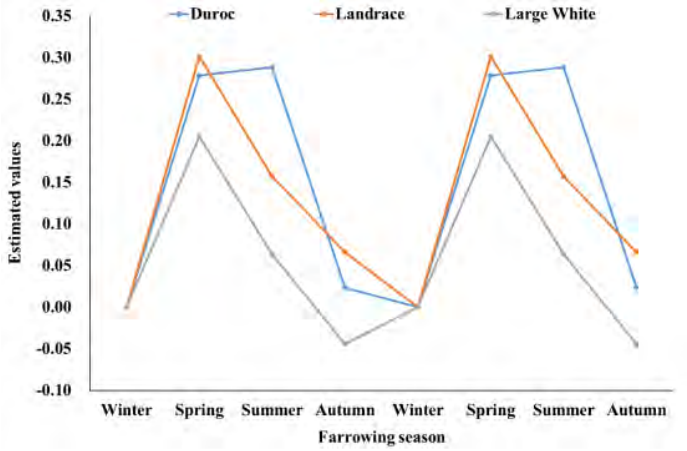


Figure 2. Estimated values of effects of farrowing season on number born alive for different pig breeds. Results for Landrace and Large White breeds were from Ogawa *et al.* (2019).

semen production and litter traits in purebred Duroc pigs. Wolf (2010) and Brinke *et al.* (2020) estimated genetic correlations between semen production and litter traits in different Landrace and Large White pig populations, the consistency of the results was poor, perhaps because, as Wolf (2010) stated, no animal has both records on semen production and litter traits, and therefore, genetic correlations could be estimated only via relatives.

Heat stress due to seasonal variation in climatic factors including temperature, as well as global warming, harms the efficient pork production system (Mayorga *et al.*, 2019). We to observed seasonal variation in semen production traits and NBA, which seems to be the heat stress. Therefore, we are now trying to establish a breeding scheme for improving heat tolerance in Japanese pig populations to confront the anticipated effects of heat stress.

References

Brinke I., Groβe-Brinkhaus C., Roth K., Pröll-Cornelissen M.J., Henne, H., *et al.* (2020) BMC Genet. 21:61. https://doi.org/10.1186/s12863-020-00865-z

Klei B., and Tsuruta S. (2008) Approximate variance for heritability estimates. Available at: http://nce.ads.uga.edu/~shogo/html/research/AI_SE.pdf

Mayorga E.J., Renaudeau D., Ramirez, B.C., Ross J., and Baumgard, L.H. (2019) Anim. Front. 9:54-61. https://doi.org/10.1093/af/vfy035

Misztal I., Tsuruta S., Strabel T., Auvray B., Druet T., *et al.* (2002) Proc. of the 7th WCGALP, Montpellier, France.

Ogawa S., Konta A., Kimata M., Ishii K., Uemoto Y., *et al.* (2019) Anim. Sci. J. 90:1510-1516. https://doi.org/10.1111/asj.13304

Wolf, J. (2010) J. Anim. Sci. 88:2893-2903. https://doi.org/10.2527/jas.2009-2555

774. Genomic analysis of quality and morphological sperm parameters in Piétrain breeding boars

C. Groβe-Brinkhaus[1,2*], Z.M. Kröhl[1], M. Friedrichs[3] and E. Tholen[1]

[1]University of Bonn, Institute of Animal Science, Endenicher Allee 15, 53115 Bonn, Germany; [2]Association for Bioeconomy Research (FBF), Adenauerallee 174, 53113 Bonn, Germany; [3]GFS-Cooperation, Zum Pöpping 29, 59387 Ascheberg, Germany; cgro@itw.uni-bonn.de

Abstract

Artificial insemination is essential for breeding progress and economic success in pig production. The aim of this study was the estimation of genetic parameter and to perform a genome-wide association study (GWAS) for traits recorded with the computer assisted semen analysis system in Piétrain breeding boars. Beside sperm volume, the fresh semen was evaluated for five sperm quality traits. In total 72,370 records of 958 Piétrain boars were available for variance component estimation. The h^2 ranged from 0.048 for motility to 0.284 for plasma droplets. The environmental effect of the repeated semen collection was highest for motility with 0.260. The genetic correlations met biological expectations and ranged from -0.912 to 0.933 among the sperm quality traits. First GWAS allowed to detect 21 chromosome-wide significant associations for motility, total number of sperm cells and sperm morphological abnormalities. Most of the identified regions have been already reported in other studies.

Introduction

Artificial insemination (AI) is essential for breeding progress and economic success in pig production. The selection of AI boars mainly based on breeding values for production and female reproduction traits (Wolf 2009), whereas the selection of young boars on semen quality is still not a common practice (Marques *et al.* 2017). In Germany, based on pedigree information an indirect selection is realized before an AI-boar is bought or performed based on phenotypes during semen production (M. Friedrichs, personal communication).

In the recent years the evaluation of the sperm quality in AI-stations has been changed from a subjective inspection to the computer assisted semen analysis (CASA). Compared to a subjective evaluation this system allows to record sperm motility and morphological abnormalities with a high measurement accuracy(e.g. Amann and Waberski 2014). Occurrence of morphological abnormalities and cytoplasmatic droplets can determine the quality of an ejaculate due to their negative impact on motility and fertility of sperm (Gruhot *et al.* 2019).

For semen quantity and sperm quality traits low to moderate heritabilities (h^2) have been reported (Marques *et al.* 2017; Wolf 2010). Several studies investigated candidate genes affecting subjective evaluated sperm quality traits (e.g. Kaewmala *et al.* 2011). But, because of the limited number of boars with phenotypic and genotypic data, only few studies have performed genome-wide association analyses (GWAS) for male reproduction traits (Marques *et al.* 2018; Gao *et al.* 2019; Diniz *et al.* 2014).

The aim of this study was the estimation of genetic parameters and to perform a genome-wide association study for CASA traits in Piétrain AI- boars, used as nucleus and terminal sires.

Materials & methods

All investigated animals belong to the breed Piétrain owned by the AI-station GFS. These boars are part of the breeding program of a German herdbook organisation (German Genetic) and were used as terminal

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_774, © C. Groβe-Brinkhaus *et al.* 2022

sires in production as well as breeding boars in the nucleus of the herdbook organisation. Data of 958 breeding boars housed on six AI-station of the GFS from March 2018 to March 2021 were available. In total 72,370 ejaculates were considered for investigation. Boars with less than ten ejaculates, were removed from the analyses.

The fresh semen was evaluated immediately after ejaculation using the CASA eFlow system Androvision® (Minitüb GmbH, Tiefenbach, Germany). Following five traits were recorded in a semen sample: sperm morphological abnormalities (ABN, %), cytoplasmic droplets (CD, %), total number of sperm cells (Ncell-T, Mrd), total number of intact sperm cells (Ncell-I, Mrd), sperm motility (MOT, %). In addition, the collected semen was weighted (semen volume, VOL, ml). Because of the skewness of ABN, the values were log-transformed into logABN.

Furthermore, boars were genotyped with the 60K Illumina bead chip. After quality control (call rate >0.95, minor allele frequency >0.05 and deviation from Hardy Weinberg Equilibrium $P<0.001$) 43,483 SNPs of 754 boars with 67,561 phenotypes were available for further investigations.

Variance component estimation. In a first step, environmental effects influencing the semen quality and sperm morphology were identified. Variance components were estimated using the VCE5 package (Kovač and Groeneveld 2003). The following multi-trait repeatability model in matrix form was used:

$$y = Xb + Za + Wpe + Ie \quad (1)$$

where y displays sperm quantity and -quality traits; X, Z, and W denoted the design matrices, I The identity matrix; b was the solution vector of the fixed effect including station-year-season of ejaculation, collection interval (days) and age in days of the boars when ejaculating (covariate, quadratic form); $a \sim N(0, A\sigma^2_a)$ was a vector of genetic values; $pe \sim N(0, I\sigma^2_{pe})$ was a vector of random permanent boar effect $e \sim N(0, I\sigma^2 e)$ denoted the residuals and I denoted the identity matrix. A series of tri-variate model analyses were applied, which comprised all possible three trait combinations. Heritability (h^2), permanent environmental effects (pe^2), phenotypic (r_p) and genetic (r_g) correlations are presented by calculating the mean of all particular estimations.

Genome-wide association study. For GWAS Equation 1 was extended by a single SNP as additional fixed regression coefficient, where the particular SNP was additively coded {0,1,2}. Possible effects through population stratification were visually proven by calculating a genomic relationship matrix following VanRaden (2008) and plotting the first two principal components. In order to control the relatedness between the animals, the additive genetic relationship was considered (Equation 1) and the genomic control method was applied. Significant associations were detected using a conditional Wald F statistic. Resulting *P*-values were transformed by calculating the false discovery rate to avoid error accumulation by multiple testing (Storey 2002). Markers with an adjusted *P*-value<0.05 were handled as chromosome wide significant. The GWAS was performed with ASreml (Gilmour *et al.* 2006).

Results

In total 72,370 ejaculation records of 958 Piétrain boars were available for analyses. The investigated traits revealed averaged mean values and standard deviations: VOL 245.69±72.62; MOT 88.83±5.31; ABN 2.78±2.8; PD 9.66±5.39; Ncell-T 73.74±25.61; Ncell-I 57.84±20.49.

In Table 1 and 2 an overview of the estimated h^2, pe^2 and correlations is given. The h^2 ranged from 0.048 for MOT to 0.284 for CD. The pe^2 reflects the permanent environmental repeatability of the measurements and

Table 1. Heritability and permanent environmental effects for sperm quantity and quality traits.

Traits (abbreviation, unit)	$h^2 \pm SE$	$pe^2 \pm SE$	ω^2
Semen volume (VOL, ml)	0.242±0.017	0.184±0.015	0.426
Motility (MOT, %)	0.048±0.015	0.260±0.015	0.308
Sperm morphological abnormalities (logABN, %)	0.149±0.026	0.201±0.023	0.350
Cytoplasmic droplets (CD, %)	0.284±0.032	0.166±0.025	0.450
Total number of sperm cells (Ncell-T, Mrd)	0.174±0.023	0.237±0.024	0.411
Number of intact sperm cells (Ncell-I, Mrd)	0.142±0.026	0.243±0.023	0.385

[1] SE: Standard error; h^2: heritability; pe^2 permanent environmental effect, ω^2: repeatability.

Table 2. Heritability, phenotypic and genetic correlations for sperm quantity and quality traits.[1]

	VOL	MOT	logABN	CD	Ncell-T	Ncell-I
VOL	0.242±0.017	0.144±0.116	-0.184±0.099	-0.226±0.092	0.220±0.074	0.325±0.091
MOT	0.002	0.048±0.015	-0.912±0.064	-0.821±0.085	-0.012±0.210	0.27±0.166
logABN	0.042	-0.336	0.149±0.026	0.581±0.034	0.061±0.066	-0.285±0.126
CD	-0.025	-0.418	0.765	0.284±0.032	0.175±0.042	-0.241±0.034
Ncell-T	0.253	0.214	0.099	0.0117	0.174±0.023	0.933±0.022
Ncells-I	0.258	0.493	-0.133	-0.207	0.907	0.142±0.026

[1] h^2 (± standard error) on the diagonal, r_p phenotypic correlation under the diagonal, r_g genetic correlation above the diagonal; VOL: volume; MOT: motility; logABN: log-transformed sperm morphological abnormalities; CD: cytoplasmic droplets; Ncell-T: total number of sperm cells; Ncell-I: number of intact sperm cells.

were moderate for all traits. The highest pe^2-values were observed for MOT (0.260) and Ncel-I (0.243). The effect of the general repeated effect ω^2 (h^2+pe^2) of the semen was highest for CD with 0.450.

The observed phenotypic and genetic correlations followed according to the sign biological expectations (Table 2). The semen amount revealed a positive r_g of 0.22 and 0.32 with Ncell-T and Ncell-I traits, but was negatively correlated the morphological traits ABN and PD. A strong negative r_g was found between MOT and ABN as well as MOT and CD of -0.91 and -0.82, respectively.

In a first GWAS, we identified 21 chromosome-wide significant SNPs comprising six QTL of three sperm quality traits. For MOT two QTL located on *Sus scrofa chromosome* (SSC) 6 and SSC7 were detected. On SSC5 13 SNPs located within three chromosomal regions revealed significant effects on logABN.

Discussion

The observes values for the mean and standard deviations of the investigated phenotypes are comparable to the reports in literature. The proportion of the CD were higher comparted to the ABN. This result can be expected, because CD are the most recent sperm abnormalities in boar semen (Briz and Fàbrega 2013). The trait ABN comprises several sperm abnormalities, like head or tail deformations, which can originate in the spermatogenesis or the maturation in the epididymidis (Briz *et al.* 1996). Plasma droplets, separated in cytoplasmic or terminal droplets and have been described as part of the sperm maturation, but have a negative impact on the fertilization (Briz and Fàbrega 2013).

The estimated h^2 and r_g of the investigated traits were in accordance with reports from literature (Wolf 2009; Diniz *et al.* 2014; Marques *et al.* 2017; Brinke *et al.* 2020). Only a limited number of studies investigated the genetics of CD so far (Zhao *et al.* 2019; Gruhot *et al.* 2019). In the recent work, the h^2 of CD (h^2: 0.28) was higher compared to ABN (h^2: 0.15).

The identified genomic regions have been reported in previous studies (Marques *et al.* 2018; Kaewmala *et al.* 2011). The limited number of six identified QTL met expectations, because the data set does not cover the complete phenotypic variance. The investigated boars were indirect selected for a high sperm quality and the number of low-quality ejaculates was below 5%.

References

Amann R. P., and Waberski D. (2014) Theriogenology 81 (1): 5-17.e1-3. https://doi.org/10.1016/j.theriogenology.2013.09.004

Brinke I., Groβe-Brinkhaus C., Roth K., Pröll-Cornelissen M. J., Henne H. *et al.* (2020) BMC Genet. 21 (1): 61. https://doi.org/10.1186/s12863-020-00865-z

Briz M., and Fàbrega A. (2013) The Boar Spermatozoon. In: Bonet S., Casas I., Holt W. V. and Yeste M. (eds) Boar Reproduction. Springer, Berlin, Heidelberg. pp. 3–47.

Briz M. D., Bonet S., Pinart B., Camps R. (1996) Anim. Reprod. Sci. 43 (4): 221–239. https://doi.org/10.1016/0378-4320(96)01470-4

Diniz D. B., Lopes M. S., Broekhuijse M.L.W.J., Lopes P. S., Harlizius B. *et al.* (2014) Anim. Reprod. Sci. 151 (3-4): 201–207. https://doi.org/10.1016/j.anireprosci.2014.10.014

Gao N., Chen Y., Liu X., Zhao Y., Zhu L. *et al.* (2019) BMC Genom 20 (1): 2893. https://doi.org/10.1186/s12864-019-6164-5

Gilmour R., Cullis B. R., Harding S. A., Thompson R. (2006) ASReml User Guide. Version 4: VSN International Ltd.

Gruhot T., Gray K., Brown V., Huang Y., Kachman S. D. *et al.* (2019) Anim. Reprod. Sci. 206: 85–92. https://doi.org/10.1016/j.anireprosci.2019.05.012

Kaewmala K., Uddin M. J., Cinar M. U., Grosse-Brinkhaus C., Jonas E. *et al.* (2011) Anim. Reprod. Sci. 125 (1-4): 170–179. https://doi.org/10.1016/j.anireprosci.2011.02.017

Kovač, M., and Groeneveld, E. (2003) VCE-5: user's guide and reference manual: version 5.1: Biotechnical Faculty, Department of Animal Science

Marques D. B. D., Bastiaansen J. W. M., Broekhuijse M. L. W. J., Lopes M. S., Knol E. F. *et al.* (2018) Genet. Sel. Evol. 50 (1): 164. https://doi.org/10.1186/s12711-018-0412-z

Marques D. B. D., Lopes M. S., Broekhuijse M. L. W. J., Guimarães S. E. F., Knol E. F. *et al.* (2017) J. Anim. Sci. 95 (10): 4251–4259. https://doi.org/10.2527/jas2017.1683

Storey J. D. (2002) J. R. Stat. Soc. Series. B Stat. Methodol. 64: 479–498. https://doi.org/10.1111/1467-9868.00346

VanRaden P. M. (2008) J. Dairy Sci. 91 (11): 4414–4423. https://doi.org/10.3168/jds.2007-0980

Wolf J. (2009) Animal 3 (8): 1094–1099. https://doi.org/10.1017/S1751731109004686

Wolf J. (2010) J. Anim. Sci. 88 (9): 2893–2903. https://doi.org/10.2527/jas.2009-2555

Zhao Y., Gao N., Cheng J., El-Ashram S., Zhu L. *et al.* (2019) Animals 9 (10). https://doi.org/10.3390/ani9100710

775. Differential expression in the porcine thymus upon infection by PRRSV, influenza B or their coinfection

K.M. Sarlo Davila[1,2*], Y. Sang[3], W. Ma[4] and L.C. Miller[2]

[1]Oak Ridge Institute for Science and Education, Oak Ridge Associated Universities, Oak Ridge, TN 37830, USA; [2]Virus and Prion Research Unit, National Animal Disease Center, USDA, Agricultural Research Service, Ames, IA 5061, USA; [3]Department of Agricultural and Environmental Sciences, College of Agriculture, Tennessee State University, Nashville, TN 37209, USA; [4]Department of Veterinary Pathobiology Department of Molecular Microbiology & Immunology, University of Missouri, Columbia, MO 65211, USA; kaitlyn.sarlodavila@usda.gov

Abstract

Porcine Reproductive and Respiratory Syndrome (PRRSV) causes large economic losses and suppresses the host immune system, impairing thymus function and leaving pigs susceptible to secondary infections such as zoonotic Influenza B (IBV). Seronegative pigs at 3-4 week-old were split into four treatment groups: control; infected with Type 2 PRRSV NPB strain; infected with B/Brisbane/60/2008 virus; or coinfected with both viruses. Pigs from each of the four treatment groups were necropsied at 3-, 5- and 7-days post-infection(dpi) and thymus samples were collected for transcriptomic analysis. Differentially expressed gene (DEG) analysis was carried out using DeSeq2 and g:Profiler was used to identify significantly enriched pathways. Several pathways related to immune function were detected in the PRRSV and coinfected but not IBV samples, indicating that PRRSV infection persist and subvert the host immune system by downregulating genes involved in cytokine and interferon signalling and monocyte and macrophage signalling with the thymus.

Introduction

Infectious diseases are one of the primary constraints of swine production, resulting not only in production and economic losses but also in some cases impacting human health (VanderWaal and Deen, 2018). Worldwide PRRSV is the most economically important infectious swine disease, estimated to result in a loss of $664 million annually in the US alone (Lunney *et al.*, 2016). One component of PRRSV economic devastation is its ability to suppress the host immune system, including impairing thymus function, leaving pigs susceptible to secondary infections (Wang *et al.*, 2020). Influenzas are common secondary infections to primary PRRSV infections: a serological survey of U.S. Midwest domestic swine herds from detected IBV in 38.5% of sampled farms and observed that herds infected with PRRSV had a higher prevalence of IBV compared to herds without PRRSV (Ran *et al.*, 2015). IBV is of concern not only to swine production but also human health as it is pathogenic to humans.

Materials & methods

Animals. Samples were provided by Dr Wenjun Ma, Kansas State University. The animal study was reviewed and approved by the Institutional Animal Care and Use Committee at Kansas State University (IACUC #3821). Briefly, 36 pigs (3-4 week-old) seronegative for influenza A, influenza B, and PRRSV were split into 4 groups: one blank control group and three infected groups (PRRSV, IBV, and PRRSV/IBV coinfection). Each pig from the PRRSV and coinfected group were intranasally infected with Type 2 PRRSV NB strain at a tissue culture infectious dose (TCID) of 2×10^5 $TCID_{50}$/ml. Control and IBV pigs were mock inoculated with virus free minimum essential medium (MEM) through the same route. At 14 days post infection with PRRSV each pig from coinfected and IBV group was inoculated intratracheally (1 ml) and intranasally (1 ml) with B/Brisbane.60/2008 virus at 10^6 $TCID_{50}$/ml. Pigs from the other two groups were inoculated with same volume of virus free MEM through the same routes. Three pigs from each of

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_775, © K.M. Sarlo Davila *et al.* 2022

the four groups were necropsied at 3-, 5-, and 7-days post IBV inoculation(dpi). The objective of this study was to investigate host transcriptomic response in the thymus following PRRSV and IBV infections as well as their coinfection

Sample preparation and sequencing. Thymus samples were collected from three pigs in each of the four treatment groups (Control, PRRSV, IBV and PRRSV/IBV) at each of the three time points (dpi, 5 dpi, and 7 dpi) with the exception IBV 7 dpi pigs at which time point samples were only collected from two pigs. Total RNA was extracted from the samples, preparation was performed the 3' Quantseq fwd kit, and 100 bp paired end reads were generated utilizing the Illumina Hiseq 6000.Tools present at galaxy.scinet.usdagov were utilized to analyze the sequenced reads. FastQC was used to perform quality control for reads and examine raw read data and counts. Trimmomatic (version 0.38.1) was used to remove adapters and reads with a phred score below 20 prior to alignment on the Sscrofa 11.1 genome with HiSat2 (version2.1.0). Raw counts were generated with the default parameters of FeatureCounts. Differential gene expression (DEG) was performed using DeSeq2 (version 2.11.40.6) utilizing a parametric fit type and poscounts to account for genes with zero counts. DEG analysis was based on the model treatment + dpi + treatment: dpi +E. Statistically significant (FDR<0.15) genes were reported for the interaction effect of treatment: dpi. Further analysis was performed on significant genes from each dpi and treatment group interaction for significantly enriched pathways using the g: Gost function of g: Profiler. A Benjamini Hochberg FDR<0.05 was used as the significance threshold for KEGG pathways. A network representing overlaps among enriched pathways was created utilizing the EnrichmentMap plugin of CytoScape. DEGs within KEGG pathways were visualized using PathView.

Results

PRRSV. Comparison of the PRRSV infected samples vs control samples resulted in a total of 3 statistically significant DEG (false discovery rate (FDR)<0.15) (2 downregulated,1 upregulated) for 3 dpi and did not result in any significantly enriched pathways. Expression levels peaked at 5 dpi with 253 significant DEG (196 downregulated, 57 upregulated) which were significantly enriched for 4 KEGG pathways: Influenza A, viral protein interaction with cytokine and cytokine receptor, cytosolic DNA-sensing pathway, and natural killer cell mediated cytotoxicity. Expression levels decreased at 7 dpi with 27 significant DEG (12 downregulated, 15 upregulated) which were not significantly enriched for any pathways.

Influenza B. Influenza B-infected samples showed the least expression overall but a similar pattern over time to PRRSV-infected samples, with the highest expression levels occurring at 5 dpi, which was also the only timepoint having pathways that were significantly enriched. Comparison of the Influenza B-infected samples vs control samples resulted in a total of 3 DEG (1 downregulated, 2 upregulated) for 3 dpi, 5 DEG (4 downregulated, 1 upregulated) for 5 dpi, and 4 DEG (1 downregulated, 3 upregulated) for 7 dpi. The 5 DEG at 5 dpi were over enriched for the KEGG pathway: autophagy-animal.

Coinfection. In contrast to either single infection when compared to the control samples the coinfected samples showed the greatest expression levels at 3 dpi. In total, 314 genes were differentially expressed (261 downregulated, 53 upregulated) at 3 dpi and were significantly enriched for 9 KEGG pathways: Influenza A, RIG-1-like receptor signalling pathway, viral protein interaction with cytokine and cytokine receptor, cytosolic DNA-sensing pathway, Hepatitis C, Herpes 1 simplex virus infection, Measles, Coronavirus disease-COVID 19 and Epstein-Barr virus infection. Only 4 DEG (1 downregulated, 3 upregulated) were significant at 5 dpi. Expression levels increased slightly at 7 dpi with 26 significant DEG (14 downregulated, 12 upregulated) but no pathways were significantly enriched at either time point.

Discussion

An enrichment map representing the overlap of enriched pathways across samples is shown in Figure 1. The PRRSV single infection and PRRSV/IBV coinfection share several enriched pathways yet neither infection overlaps with the IBV single infection, which may give insight into the ability of PRRSV infections persist in lymphoid tissues. This pattern from of expression differs significantly from that in the lung as Fleming *et al.* (2020) reported shared DEGS among the 3 infection groups in the lung. The DEGs from both PRRSV and coinfected samples were significantly enriched for the Influenza A pathway which has been very well annotated. The differential regulation of genes within this pathway for both the PRRSV and coinfected samples are shown in Figure 2. PRRSV has previously been reported to induce abnormal innate and adaptive immune responses, suppress interferon (IFN) induction, dysregulate cytokine expression and

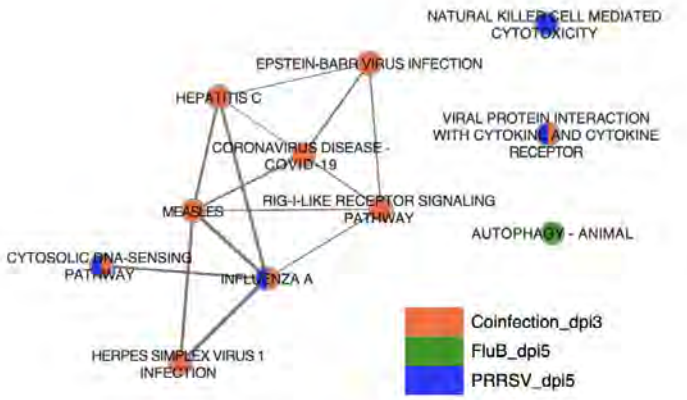


Figure 1. Cytoscape network of significantly enriched pathways. Pathways are represented as circles and connected with lines sized based on the number of genes shared by the connected pathways.

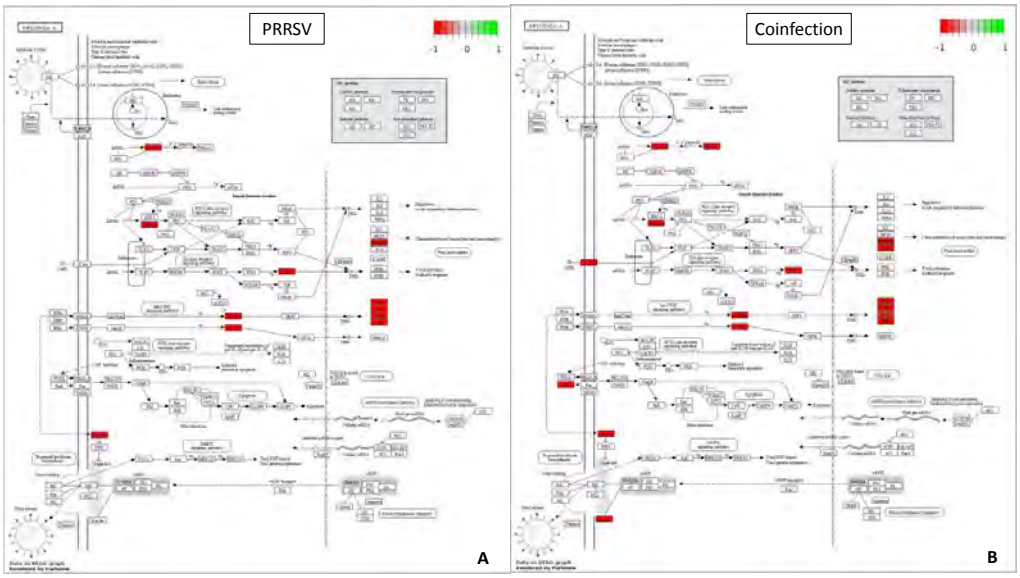


Figure 2. Enriched KEGG pathway Influenza A with downregulated genes highlighted in red and upregulated genes highlighted in green. (A) Differential regulation in the PRRSV samples. (B) Differential regulation in the coinfected samples.

atrophy the thymus but the mechanisms by which this occurs are not yet fully understood (Wang *et al.*, 2020). In both the PRRSV and coinfected samples several key genes in IFN induction were downregulated, including *MDA5* and *IRF7* as well as IFN induced genes *STAT1, MXA, OAS, ADAR, PML* and *Viperin,* all of which are key genes in antiviral activity (Stelzer *et al.*, 2016). *RANTES,* a chemokine signallign gene, was also downregulated. This downregulation may be unique to lymphoid tissues as Fleming *et al.* (2020) reported that *IRF* genes were upregulated in PRRSV infected lung samples across all three timpoints. Additionally *IP10, TLR4, FASL,*and *β-actin* were downregulated in the coinfection but not the PRRSV single infection. *IP10* is also an IFN induced gene which stimulates monocytes, *TLR4* is essential in pathogen recognition, *FASL* induces apoptosis and *β-actin* is crucial for intercellular signalling (Stelzer *et al.*, 2016). The downregulation of these genes provides insight into the mechanism by which PRRSV establishes multi-pathogen coinfection and persistent viral infection in lymphoid tissues.

References

Fleming DS, Miller LC, Tian Y, Li Y, Ma W, Sang Y.(2020). Pathogens. 9(11):934. https://doi.org/10.3390/pathogens9110934.

Lunney, J. K., Y. Fang, A. Ladinig, N. Chen, Y. Li, B. Rowland, *et al.* (2016) Annu. Rev. Anim. Biosci. 4:129–154. https://doi.org/10.1146/annurev-animal-022114-111025.

Ran, Z., H. Shen, Y. Lang, E. A. Kolb, N. Turan, *et al.* (2015) J. Virol. 89:4818–4826. https://doi.org/10.1128/jvi.00059-15.

Stelzer G, Rosen R, Plaschkes I, Zimmerman S, Twik M, Fishilevich S, *et al.* (2016) Curr. Protocols Bioinfo. https://doi.org/10.1002/cpbi.5.

VanderWaal, K., and J. Deen. (2018) Proc. Natl. Acad. Sci. U. S. A. 115:11495–11500. https://doi.org/10.1073/pnas.1806068115.

Wang, G., Y. Yu, X. Cai, E. M. Zhou, and J. J. Zimmerman. (2020) Trends Microbiol. 28:212–223. https://doi.org/10.1016/j.tim.2019.10.009.

776. Genetic associations between survival at different parities in commercial crossbred sows

B.G. Poulsen[1,2*], N.G. Leite[3] and D. Lourenco[3]

[1]Danish Pig Research Centre, Danish Agriculture & Food Council, Axelborg, 1609 Copenhagen W, Denmark; [2]Quantitative Genetics and Genomics, Aarhus University, Blichers Allé 20, 8830 Tjele, Denmark; [3]Department of Animal and Dairy Science, University of Georgia, Athens, GA 30602, USA; bgp@lf.dk

Abstract

We investigated whether sow survival is genetically associated across parities. Sow survival was investigated through five stayability traits and a five-trait animal model: The first stayability trait was survival from first service to first farrowing. The fifth stayability trait was survival from fourth farrowing to fifth farrowing. All sows were commercial and either F1 or rotationally crossbred between Landrace and Yorkshire. All stayability traits were favourably genetically correlated in both Landrace and Yorkshire. Stayability from first service to first farrowing was moderately to strongly genetically correlated with other stayability traits (0.26-0.73). The remaining four stayability traits were strongly genetically correlated with each other (0.63-0.88). Therefore, survival from first farrowing to fifth farrowing can be regarded as similar traits, while survival from first service to first farrowing is relatively different from later survival.

Introduction

Promoting meaningful genetic progress for sow survival is challenging. One reason is that sow survival is a biologically complex trait. Sows are culled for multiple reasons, and the risk for culling due to any of these reasons depends on the parity (Engblom *et al.*, 2007). Because frequencies of specific reasons for culling differ across parities, the rankings between genetic merits for sow survival likely also differ across parities. Genetic associations between sow survival at different time points have been estimated in purebred (Abell *et al.*, 2016; Plaengkaeo *et al.*, 2020; Ogawa *et al.*, 2021) and crossbred sows (Engblom *et al.*, 2009; Abell *et al.*, 2016; Poulsen *et al.*, 2020). However, results on sow survival from purebred populations may not be extrapolatable to crossbred populations (Abell *et al.*, 2016; Poulsen *et al.*, 2020). Furthermore, the studies on crossbred sows have neither provided comprehensive insights into genetic associations across parities (Poulsen *et al.*, 2020) nor used adequate sample sizes (Engblom *et al.*, 2009; Abell *et al.*, 2016). Improved knowledge on the genetic associations between commercial sow survival at different parities can contribute to better breeding procedures for sow longevity. The objective of this study was to investigate whether sow survival is genetically associated across parities in commercial and rotationally crossbred sows. This was examined by estimating (co)variance components for five survival traits with a multi-trait animal model. We hypothesized that the five survival traits are moderately to strongly genetically associated.

Materials & methods

Animal Care and Use Committee approval was not required for this study as data were obtained from an existing database.

Phenotype data. The sows were crossbred between the DanBred Landrace and DanBred Yorkshire populations. All sows produced on commercial breed-to-wean farms located in Denmark. The phenotype data consisted of farrowing records from sows born between January 1st, 2015, and December 31st, 2017. The farrowing records included service dates, farrowing dates, litter sizes, litter sire, and breed.

Traits. We investigated sow survival through five sequential stayability traits: stayability from the first service to first farrowing (Stay1), from first farrowing to second farrowing given survival in Stay1 (Stay2),

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_776, © B.G. Poulsen *et al.* 2022

from second farrowing to third farrowing given survival in Stay2 (Stay3), from third farrowing to fourth farrowing given survival in Stay3 (Stay4), from fourth farrowing to fifth farrowing given survival in Stay4 (Stay5). All traits have binary outcomes where '1' denotes survival and '0' denotes death. The survival rates ranged from 0.74-0.93 across traits (Table 1)

Partial additive relationship matrices. The additive genetic covariance between animals was modelled with the partial additive relationship matrices (Lo *et al.*, 1993; García-Cortés and Toro, 2006) because some sows and gilts with phenotypes were rotationally crossbred. With this method and two-way rotationally crossbred animals, additive genetic effects are split into three components: a breed-specific effect for Landrace, a breed-specific effect for Yorkshire, and a segregation effect between Landrace and Yorkshire. However, preliminary analyses showed that the segregation effects were not significant for this combination of population and trait. Therefore, segregation effects were not included in the statistical model.

Statistical model. Stay1, Stay2, Stay3, Stay4, and Stay5 were simultaneously analysed using a five-trait linear mixed model:

$$\mathbf{y} = \begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \\ \mathbf{y}_3 \\ \mathbf{y}_4 \\ \mathbf{y}_5 \end{bmatrix} = \mathbf{Xb} + \mathbf{Wh} + \mathbf{Ql} + \mathbf{Z}_L\mathbf{a}_L + \mathbf{Z}_Y\mathbf{a}_Y + \mathbf{e} \quad (1)$$

Where y_1, y_2, y_3, y_4 and y_5, are vectors of phenotypes for Stay1, Stay2, Stay3, Stay4, and Stay5, respectively; b is a vector of fixed effects for herd-year at birth, and month at birth, and regression coefficients of age at first service (linear and quadratic), fraction of Yorkshire, fraction of heterosis, and reproductive performance (permanent sow effect from a mixed linear model; zero for y_1); h is a vector of random effects for combinations of the herd, year, and month at birth; l is a vector of random effects for litters; a_L and a_Y are vectors of partial additive genetic effects for breed-specific terms from Landrace and Yorkshire, respectively; e is the residual; X, W, Q, Z_l and Z_y are design matrices of proper dimensions. In Equation 1, the effects within all vectors are nested within traits.

The statistical model contains two random environmental effects, two additive genetic effects, and one residual effect which were assumed to be distributed as:

$$\begin{bmatrix} \mathbf{h} \\ \mathbf{l} \\ \mathbf{a}_L \\ \mathbf{a}_Y \\ \mathbf{e} \end{bmatrix} \sim \text{MVN}\left(\begin{bmatrix} \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \end{bmatrix}, \begin{bmatrix} \boldsymbol{\Sigma}_H \otimes \mathbf{I}_H & & & & \\ \mathbf{0} & \boldsymbol{\Sigma}_L \otimes \mathbf{I}_L & & \text{Symm.} & \\ \mathbf{0} & \mathbf{0} & \boldsymbol{\Sigma}_{AL} \otimes \mathbf{A}_L & & \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \boldsymbol{\Sigma}_{AY} \otimes \mathbf{A}_Y & \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \boldsymbol{\Sigma}_E \otimes \mathbf{I}_E \end{bmatrix} \right) \quad (2)$$

where Σ_H, Σ_L, Σ_{AL}, and Σ_{AY} are 5×5 covariance matrices where each element is an estimable parameter; Σ_E is a 5×5 diagonal covariance matrix; I_H, I_L, and I_E are identity matrices; A_L and A_Y are partial additive

Table 1. Number of animals and phenotypic mean for each trait.

	#Animals	Mean
Stay1	112,373	0.93
Stay2	104,009	0.85
Stay3	88,501	0.85
Stay4	75,355	0.81
Stay5	60,691	0.74

relationship matrices for breed-specific terms from Landrace and Yorkshire, respectively; 0's are vectors or matrices with zeros; and ⊗ is the Kronecker product.

Covariance components were estimated using the average-information restricted maximum likelihood algorithm in the DMU software (Madsen *et al.*, 2006). The standard errors of covariance components were approximated using their asymptotic normal distribution. The standard errors of heritability statistics and correlations were approximated with the Delta Method (Dorfman, 1938; Oehlert, 1992). Heritabilities and correlations were assumed normally distributed when statistically tested.

Results

Estimated genetic correlations were favourable and moderate to strong (0.26±0.17-0.88±0.18; Table 2). Subsequent stayability traits tended to be more strongly associated than non-consecutive stayability traits (0.54±0.16-0.88±0.18 vs 0.26±0.17-0.84±0.11). Stay3, Stay4, and Stay5 were strongly associated (0.73±0.18-0.88±0.18).

The heritabilities were low for all traits (0.01±0.00-0.05±0.01). Heritabilities were larger for later-in-life stayability than those measured earlier in life, except for in Yorkshire where Stay3 was as heritable as Stay1.

Discussion

Most stayability traits were favourably and moderately to strongly genetically associated in the Landrace and Yorkshire populations. Thereby, the genetic background for sow survival is similar across most parities.

Other studies have estimated genetic associations between survival at different parities in commercial and crossbred animals (Engblom *et al.*, 2009; Abell *et al.*, 2016; Poulsen *et al.*, 2020). These studies also found that survivals at different parities are moderately to strongly genetically associated. This may be because of the overlap in culling reasons as reproductive failure persists as the predominant reason for culling until the fifth parity (Engblom *et al.*, 2007; Malanda *et al.*, 2019). Of the traits investigated, only Stay1 was not consistently strongly genetically correlated with the other stayability traits.

We found that Stay1 is often only moderately genetically associated with the later-in-life stayability traits. However, it is important to note that Stay1 is defined differently than the other stayability traits. In other

Table 2. Heritabilities (diagonal), phenotypic correlations (above diagonal), and genetic correlations (below diagonal) for Landrace.

		Stay1	Stay2	Stay3	Stay4	Stay5
Landrace	Stay1	0.01±0.00[1]	0.02±0.00	0.02±0.00	0.01±0.00[1]	0.02±0.00
	Stay2	0.54±0.16[1]	0.02±0.00	0.03±0.00	0.02±0.00[1]	0.03±0.00
	Stay3	0.73±0.15[1]	0.78±0.12	0.03±0.01	0.04±0.00[1]	0.04±0.00
	Stay4	0.40±0.15[1]	0.66±0.12	0.86±0.10	0.04±0.01[1]	0.06±0.01
	Stay5	0.35±0.15[1]	0.64±0.12	0.84±0.11	0.87±0.09[1]	0.05±0.01
Yorkshire	Stay1	0.01±0.00[1]	0.03±0.00	0.01±0.00	0.01±0.00[1]	0.02±0.00
	Stay2	0.70±0.15[1]	0.02±0.00	0.02±0.00	0.02±0.00[1]	0.02±0.00
	Stay3	0.63±0.21[1]	0.73±0.18	0.01±0.00	0.03±0.00[1]	0.02±0.00
	Stay4	0.26±0.17[1]	0.63±0.13	0.88±0.18	0.03±0.01[1]	0.04±0.00
	Stay5	0.40±0.17[1]	0.65±0.14	0.73±0.18	0.85±0.13[1]	0.03±0.01

[1] Not statistically significantly different from zero.

words, Stay1 is defined as the survival between a service and a farrowing, whereas the other stayability traits are defined as the survival between one farrowing and another farrowing. This difference in definition may contribute to a weaker genetic association between the stayability traits if the genetic background differs for voluntary and involuntary culling.

The heritabilities were low (<0.10) for investigated traits in both the Landrace and Yorkshire populations. Previous studies on genetic parameters for stayability do not agree on whether the heritability for stayability is low or high (0.01-0.98) (Engblom *et al.*, 2009; Abell *et al.*, 2016; Plaengkaeo *et al.*, 2020; Poulsen *et al.*, 2020; Ogawa *et al.*, 2021). However, there is a tendency that studies on stayability in purebred animals estimate higher heritabilities than studies in crossbred animals (purebred: 0.10-0.36; crossbred: 0.01-0.06) (Engblom *et al.*, 2009; Plaengkaeo *et al.*, 2020; Poulsen *et al.*, 2020; Ogawa *et al.*, 2021). Composite traits, such as survival, are expected to have low heritability (Price and Schluter, 1991). Furthermore, it is common procedure to use crossbred animals in commercial herds. Therefore, we argue that the lower heritabilities in crossbred animals are more representative of the heritability in commercial herds than heritabilities in purebred animals.

Most survivals at different parities were moderately to strongly genetically correlated. Therefore, the high parametrisation of multi-trait models may not contribute to an increase in prediction accuracy compared to parametrically simpler models, e.g. random regression- or repeatability models. However, it is worth noting that Stay1 was less genetically associated with later-in-life stayability.

References

Abell, C. E., R. L. Fernando, T. V. Serenius, M. F. Rothschild, K. A. Gray, *et al.* (2016) J Anim Sci Biotechnol 7(1):1-6. https://doi.org/10.1186/s40104-016-0112-x

Dorfman, R. A. (1938) Biometrics Bulletin 1(1):129-138.

Engblom, L., N. Lundeheim, A.-M. Dalin, and K. Andersson. (2007) Livest Sci 106(1):76-86. https://doi.org/10.1016/j.livsci.2006.07.002

Engblom, L., N. Lundeheim, M. D. Schneider, A. M. Dalin, and K. Andersson. (2009) Animal 3(6):783-790. https://doi.org/10.1017/S175173110900411X

García-Cortés, L., and M. Toro. (2006) Genet Sel Evol 38(6):601-615. https://doi.org/10.1186/1297-9686-38-6-601

Lo, L. L., R. L. Fernando, and M. Grossman. (1993) Theor Appl Genet 87(4):423-430. https://doi.org/10.1007/bf00215087

Madsen, P., P. Sørensen, G. Su, L. Damgaard, H. Thomsen, *et al.* (2006) Proc. of the 8th WCGALP, Belo Horizonte, Brazil.

Malanda, J., P. Balogh, and G. N. Dankó. (2019) Acta Agrar Debr 2(1):9-14. http://dx.doi.org/10.34101/actaagrar/2/3671

Oehlert, G. W. (1992) Am Stat 46(1):27-29. https://doi.org/10.2307/2684406

Ogawa, S., M. Kimata, K. Ishii, Y. Uemoto, and M. Satoh. (2021) Anim Sci J 92(1):1-10. https://doi.org/10.1111/asj.13599

Plaengkaeo, S., M. Duangjinda, W. Boonkum, K. J. Stalder, and J. W. Mabry. (2020) Livest Sci 233(1-8. https://doi.org/10.1016/j.livsci.2020.103970

Poulsen, B. G., B. Nielsen, T. Ostersen, and O. F. Christensen. (2020) J Anim Sci 98(6):1-8. https://doi.org/10.1093/jas/skaa183

Price, T., and D. Schluter. (1991) Evol 45(4):853-861. https://doi.org/10.1111/j.1558-5646.1991.tb04354.x

777. Piglet Survival is lowly heritable in DanBred Landrace pigs

X. Guo[1,2*], A.C. Sørensen[1], M. Henryon[1] and T. Ostersen[1]

[1]Danish Pig Research Centre, Danish Agriculture & Food Council, 1609 Copenhagen V, Denmark; [2]Center for Quantitative Genetics and Genomics, Aarhus University, 8830 Tjele, Denmark; xguo@lf.dk

Abstract

We tested the premise that piglet survival at day 5 is a lowly-heritable trait in DanBred Landrace pigs. We did this by fitting a bivariate animal model to piglet survival and total number of piglets born. Our results showed that piglet survival was lowly heritable in Landrace pigs. The direct and maternal heritabilities were 0.018 and 0.020; the total heritability was 0.030. Heritability for piglet survival is clearly low in DanBred Landrace pigs, but there should be enough genetic variation to increase piglet survival.

Introduction

The proportion of live piglets has increased since selection for litter size at day 5 after farrowing was implemented in DanBred's pig breeding program for Landrace and Yorkshire pigs (Nielsen *et al.*, 2013). Survival is an important trait in pig production systems for ethical and economic reasons (Amer *et al.*, 2014). Direct and maternal heritabilities for piglet survival have been reported as low. They tend to range from 0.01 to 0.24 when estimated using threshold-liability models (Su *et al.*, 2008; Ibáñez-Escriche *et al.*, 2009; Roehe *et al.*, 2010; Kapell *et al.*, 2011) and from 0.00 to 0.05 using linear models (Knol *et al.*, 2002). This suggests that piglet survival at day 5 is also a lowly-heritable trait in DanBred pigs. We tested this premise by fitting a bivariate animal model to DanBred Landrace pigs.

Materials & methods

Experimental design. We estimated the additive-genetic variation, heritability, and prediction accuracy for survival in DanBred Landrace pigs. We did this in two parts. First, we estimated additive-genetic variation and heritability by fitting a bivariate linear model to piglet survival at day 5 (SURV) as a trait of the piglet and total number of born piglets (TNB) as a trait of the sow. Second, we estimated prediction accuracy using breeding values derived from the bivariate linear model in a validation analysis. We did this using SURV and TNB records from 537,746 Landrace piglets from 35,127 first parity litters, and 568,629 animals were included in pedigree. All the piglets were born between January 2016 and April 2021. SURV was recorded on a binary scale, 0 for dead and 1 for alive. The descriptive statistics of the two traits are shown in Table 1.

Genetic variation and heritability. The bivariate model fitted to estimate additive-genetic variation and heritability was:

$$\begin{bmatrix} \boldsymbol{y}_S \\ \boldsymbol{y}_T \end{bmatrix} = \begin{bmatrix} \boldsymbol{X}_S & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{X}_T \end{bmatrix} \begin{bmatrix} \boldsymbol{b}_S \\ \boldsymbol{b}_T \end{bmatrix} + \begin{bmatrix} \boldsymbol{Z}_{a_S} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} \end{bmatrix} \begin{bmatrix} \boldsymbol{a}_S \\ \boldsymbol{a}_T \end{bmatrix} + \begin{bmatrix} \boldsymbol{Z}_{u_S} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} \end{bmatrix} \begin{bmatrix} \boldsymbol{u}_S \\ \boldsymbol{u}_T \end{bmatrix} + \begin{bmatrix} \boldsymbol{Z}_{m_S} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{Z}_{m_T} \end{bmatrix} \begin{bmatrix} \boldsymbol{m}_S \\ \boldsymbol{m}_T \end{bmatrix} + \begin{bmatrix} \boldsymbol{Z}_{l_S} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{Z}_{l_T} \end{bmatrix} \begin{bmatrix} \boldsymbol{l}_S \\ \boldsymbol{l}_T \end{bmatrix} + \begin{bmatrix} \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{Z}_{s_T} \end{bmatrix} \begin{bmatrix} \boldsymbol{s}_S \\ \boldsymbol{s}_T \end{bmatrix} + \begin{bmatrix} \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{Z}_{c_T} \end{bmatrix} \begin{bmatrix} \boldsymbol{c}_S \\ \boldsymbol{c}_T \end{bmatrix},$$

where y_S and y_T were observations of SURV and TNB, b was fixed effects including year-month, herd-year, and also regressions on age at first farrowing (age and age^2); a was direct genetic effects of the piglet, u was non-genetic effects of the piglet, m was maternal genetic effects for SURV and direct genetic effects of the sow for TNB, l was birth litter effects, s was service sire genetic effects, c was non-genetic effects of service sire; u_s and l_t were the random residuals. X and Z were incidence matrices relating the observations to the fixed and random effects, respectively.

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_777, © X. Guo *et al.* 2022

Table 1. Descriptive statistics.

Trait[1]	No. Obs	Mean	SD	Min	Max
SURV	537,746	0.794	0.404	0	1
TNB	35,127	15.309	4.201	1	32

[1] SURV: piglet survival at day 5; TNB: total number born.

The random effects were assumed to follow multi variate normal distributions:

$$\begin{bmatrix} \boldsymbol{a_S} \\ \boldsymbol{m_S} \\ \boldsymbol{m_T} \\ \boldsymbol{s_T} \end{bmatrix} \sim MVN \left(0, \begin{bmatrix} \sigma^2_{a_S} & \sigma_{a_S,m_S} & \sigma_{a_S,m_T} & \sigma_{a_S,s_T} \\ \sigma_{a_S,m_S} & \sigma^2_{m_S} & \sigma_{m_S,m_T} & \sigma_{m_S,s_T} \\ \sigma_{a_S,m_T} & \sigma_{m_S,m_T} & \sigma^2_{m_T} & \sigma_{m_T,s_T} \\ \sigma_{a_S,s_T} & \sigma_{m_S,s_T} & \sigma_{m_T,s_T} & \sigma^2_{s_T} \end{bmatrix} \otimes \boldsymbol{A} \right),$$

$$\begin{bmatrix} \boldsymbol{u} \\ \boldsymbol{l_S} \\ \boldsymbol{l_T} \\ \boldsymbol{c_T} \end{bmatrix} \sim MVN \left(0, \begin{bmatrix} \sigma^2_u & 0 & 0 & 0 \\ 0 & \sigma^2_{l_S} & \sigma_{l_S,l_T} & 0 \\ 0 & \sigma_{l_S,l_T} & \sigma^2_{l_T} & 0 \\ 0 & 0 & 0 & \sigma^2_{c_T} \end{bmatrix} \otimes \boldsymbol{I} \right),$$

where A was the matrix of additive genetic relationships constructed based on the pedigree, and I was an identity matrix. The model was fitted using the DMU software (Madsen and Jensen, 2013)

Variance components, genetic correlations and corrected phenotypic values (y_c) were estimated using the full data. Following a previous study (Eaglen and Bijma, 2009), phenotypic variance (σ^2_p) of SURV was calculated as $\sigma^2_{p_S} = \sigma^2_{a_S} + \sigma^2_{m_S} + \sigma_{a_S,m_S} + \sigma^2_u + \sigma^2_{l_S}$. The direct heritability of SURV ($h^2_{a_S}$) was calculated as $h^2_{a_S} = \frac{\sigma^2_{a_S}}{\sigma^2_{p_S}}$, the maternal heritability of SURV ($h^2_{m_S}$) was calculated as $h^2_{m_S} = \frac{\sigma^2_{m_S}}{\sigma^2_{p_S}}$, and the total heritability of SURV was calculated as $h^2_{t_S} = \frac{\sigma^2_{a_S} + \sigma^2_{m_S} + 2\sigma_{a_S,m_S}}{\sigma^2_{p_S}}$.

Prediction accuracy. Prediction accuracy was calculated by dividing the population into training and validation data sets. Breeding values for SURV were derived from the bivariate linear model fitted to SURV and TNB. The birthdate of 1st July, 2020 was chosen as the cut-off date. The predictive accuracy was evaluated as $\sqrt{cor^2(EBV, y_c)/h^2_{t_S}}$.

The y_c were raw phenotypic records corrected for fixed effects including year-month, herd-year and regressions on age at first farrowing (age and age^2).

Results

Survival. SURV was lowly heritable in Landrace pigs. The direct-genetic and maternal-genetic variances were similar (Table 2). Their heritabilities were 0.018 and 0.020. The total heritability was 0.030.

Genetic correlations. A negative genetic correlation between direct genetic and maternal genetic effects was observed at -0.215 for SURV. The unfavourable genetic correlation was significantly different from 0.

Prediction accuracy. As shown in Table 3, the accuracy was 0.17 for prediction of direct genetic values, 0.23 for prediction of maternal genetic values, and 0.31 for total genetic values.

Discussion

Our findings showed unequivocally that piglet survival is lowly heritable in the DanBred Landrace pigs. Even though it was low, the 3% total heritability that we estimated should allow selection to increase piglet survival. Our heritability estimates were generally supported by estimates for other pig populations

Table 2. Estimates of genetic (co)variance and heritabilities for piglet survival.

Parameter[1]	σ^2_{as}	σ_{a_s,m_s}	$\sigma^2_{m_s}$	h^2_{as}	h^2_{ms}	h^2_{ts}
Estimates	0.0030	-0.0007	0.0033	0.0183	0.0202	0.0300

[1] σ^2_{as}: variance of direct genetic effects; σ^2_{ms}: variance of maternal genetic effect; σ_{a_s,m_s}: covariance between direct and maternal genetic effects; h^2_{as}: direct heritability; h^2_{ms}: maternal heritability; h^2_{ts}: total heritability.

Table 3. Prediction accuracy of piglet survival.

Trait[1]	Direct	Maternal	Direct + Maternal
SURV	0.17	0.23	0.31

[1] SURV: piglet survival at day 5.

(Grandinson *et al.*, 2002; Knol *et al.*, 2002). We also found that approximately half of the genetic variation for piglet survival was direct-genetic variation, the other half was maternal-genetic. This is important because it means that breeders need to utilise both sources of variation to maximise genetic gains for piglet survival. The potential for increasing piglet survival by utilising both sources was highlighted by our prediction accuracies. Prediction accuracy of 0.31 is particularly high for a trait with a heritability of 3%. The high accuracy partly reflects the DanBred pig breeding system, where large litter sizes have resulted in each pig in the breeding population having many close relatives. So, even though the heritability for piglet survival is low, there should be enough genetic variation and sufficient prediction accuracy to increase piglet survival in DanBred Landrace pigs.

The high prediction accuracy probably also reflects the bivariate model we fitted to piglet survival. A preliminary analysis showed that it was worthwhile including TNB in this model because it resulted in larger estimates of direct and maternal genetic variances for piglet survival than a univariate model fitted to piglet survival. The most likely reason for the larger estimates was that the bivariate model fitted genetic and non-genetic correlations between SURV and TNB. For example, we detected a negative genetic correlation between direct and maternal-genetic effects for survival. Direct-genetic effects for survival also showed negative genetic correlations to both sow and service-sire effects for TNB (results not shown). These correlationsprovide additional information sources that increases the prediction accuracy for piglet survival. Therefore, fitting bivariate models that include TNB provide additional sources of information that increase prediction accuracy and help to utilise the genetic variation for piglet survival in pigs.

References

Amer P. R., Ludemann C. I., and Hermesch S. (2014) J. Anim. Sci. 92. https://doi.org/10.2527/jas.2014-7943.

Eaglen S. A. E., and Bijma P. (2009) J. Dairy Sci. 92:2229–2237. https://doi.org/10.3168/jds.2008-1654.

Grandinson K., Lund M. S., Rydhmer L., and Strandberg E. (2002) Acta Agric. Scand. A Anim. 52:167–173. https://doi.org/10.1080/090647002762381041.

Ibáñez-Escriche N., Varona L., Casellas J., Quintanilla R., and Noguera J. L. (2009) J. Anim. Sci. 87. https://doi.org/10.2527/jas.2007-0670.

Kapell D. N. R. G., Ashworth C. J., Knap P. W., and Roehe R. (2011) Livest. Sci. 135:215–224. https://doi.org/10.1016/j.livsci.2010.07.005.

Knol E. F., B. J. Ducro, J. A. M. van Arendonk, and T. van der Lende. (2002) Livest. Prod. Sci. 73. https://doi.org/10.1016/S0301-6226(01)00248-2.

Madsen P., and Jensen J. (2013) A user's guide to DMU, version 6, release 5.2.

Nielsen B., Su G., Lund M. S., and Madsen P. (2013). J. Anim. Sci. 91. https://doi.org/10.2527/jas.2012-5990.
Roehe R., Shrestha N. P., Mekkawy W., Baxter E. M., Knap P. W., Smurthwaite K. M., Jarvis S. Lawrence A. B., and Edwards S. A. (2010) J. Anim. Sci. 88. https://doi.org/10.2527/jas.2009-2287.
Su G., Sorensen D., and Lund M. S. (2008) Animal. 2:184–189. https://doi.org/10.1017/S1751731107001115.

778. Feasibility of reducing mortality of pigs from birth to slaughter by genetic selection

G. Su[1*], T. Liu[1,2], O.F. Christensen[1], M.S. Lund[1] and B. Nielsen[1,3]

[1]Center for Quantitative Genetics and Genomics, Aarhus University, Blichers Alle 20, 8830 Tjele, Denmark; [2]Institute of Animal Science, Guangdong Academy of Agricultural Sciences, Dafeng 1st Street 1, Wushan, Tianhe District, 510640 Guangzhou, China; [3]SEGES Pig Research Centre, Axelborg, Axeltorv 3, 1609 Copenhagen, Denmark; guosheng.su@qgg.au.dk

Abstract

Pig mortality is a big welfare and economic challenge. This study investigated the feasibility of reducing mortality of pigs from day 5 to slaughter and from birth to slaughter by breeding. Data of 163,782 female pigs were from DanBred Yorkshire breeding herds. Variances and breeding values were estimated using a linear, a logit and a probit model based on mortality record (0, 1) at individual level. Estimates of direct heritabilities ranged from 0.013 to 0.017 and maternal heritabilities from 0.005 to 0.017. Estimated breeding values (EBVs) from the three models were highly consistent. Although low heritabilities, validation accuracies of EBVs ranged from 0.415 to 0.420 for direct EBVs, from 0.192 to 0.392 for maternal EBVs, and from 0.357 to 396 for total EBVs. The results indicate that reducing pig mortality by breeding is feasible, and linear model is efficient for genetic evaluation of pig mortality.

Introduction

Pig mortality in modern pig production has become a major welfare and economic concern. It has been reported that pre-weaning mortality rates vary between 12 and 25% (Alonso-Spilsbury *et al.*, 2007). In addition to pre-weaning mortality, previous studies have shown that mortality rates ranged between 1.9 and 4.2% in nursery phase, and between 1.6 and 4.6% in finisher phase (USDA, 2015; Christiansen, 2018). In Denmark, selection for number of piglets alive at day 5 has been implemented since 2004, which is expected to improve both litter size and piglet survival before day 5 (Su *et al.*, 2007). However, the mortality from day 5 to slaughter has not directly been taken into account yet. So far, the knowledge in efficiency of breeding for reducing pig mortality is very rare, especially for mortality after weaning.

Pig mortality is usually recorded at litter level and considered as a trait of sow. However, mortality is a complex trait that is also affected by the pig's own genotype. It may therefore be more appropriate to perform genetic evaluation of mortality at pig individual level using a model including direct and maternal additive genetic effects. Furthermore, since pig mortality at individual level is a binary trait, a standard linear Gaussian model may be not appropriate for analysing such data, and more sophisticated models should be considered.

The objective of this study is to evaluate feasibility of reducing mortality of pigs before slaughter by estimating variance components and heritability of pig mortality and validating accuracy of genetic evaluation of mortality using a linear, a logit and a probit model, based on data from Danish Yorkshire breeding herds.

Materials & methods

Population and data. The data used in this study included 163,782 female pigs, collected from DanBred Yorkshire breeding herds from October 2018 to November 2021. Male pigs were not used in this study since relevant information was missing, e.g. castration. The traits in the analysis were mortality from day 5 to slaughter (Mort5S), and mortality from birth to slaughter (MortBS). Since all females were born in nucleus herds, slaughter age was set to 150 days for all pigs. A pig was scored 1 if the pig died before slaughter, otherwise 0. Based on the data, Mortality rate was 16.6% for Mort5S and 23.8% for MortBS in this population.

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_778, © G. Su *et al.* 2022

Statistical models. Variance components and breeding values of mortality were estimated using the following single-trait models.

Linear mixed model: $y = Xb + Z_c c + Z_l l + Z_d a_d + Z_m a_m + e$

Logit model: $\eta = Xb + Z_c c + Z_l l + Z_d a_d + Z_m a_m$

Probit model: $\rho = Xb + Z_c c + Z_l l + Z_d a_d + Z_m a_m$

In the models, y is the vector of binary observations $y_i = \{0,1\}$, η is the vector of logits with $\eta_i = \log(\pi_i / (1 - \pi_i))$ where π_i is probability of death of animal i, and ρ is the vector of expected liabilities of death with $\rho_i = \Phi^{-1}(\pi_i)$ where Φ is cumulative normal distribution function, b is the vector of fixed effects including parity, herd, and year-month, c is the vector of random herd-year-month effects, l is the vector of random litter effects, a_d is the vector of direct additive genetic effects, a_m is the vector of maternal additive genetic effects. The scales of b, l, c, a_d and a_m in the three models are different. It is assumed that $c \sim N(0, I\sigma_c^2)$, $l \sim N(0, I\sigma_l^2)$, $\begin{bmatrix}\mathbf{a_d}\\ \mathbf{a_m}\end{bmatrix} \sim N\left(\mathbf{0}, \begin{bmatrix}\sigma^2_{ad} & \sigma_{adam}\\ \sigma_{adam} & \sigma^2_{am}\end{bmatrix} \otimes \mathbf{A}\right)$. The residuals are assumed to $e \sim N(0, I\sigma_e^2)$ in the linear model, set to $e \sim N(0, I)$ in the probit model, and logistic distribution with variance ≈ 3.29 in the logit model. Phenotypic variance was defined as $\sigma_p^2 = \sigma_l^2 + \sigma_{ad}^2 + \sigma_{am}^2 + \sigma_{adam} + \sigma_e^2$ (Willham, 1972). The models were implemented using DMU package (Madsen *et al.*, 2010).

Validation of genetic prediction. For validation of genetic prediction using the above models, the full data were divided into reference data and validation data by dam's birth date with 2019-10-01 as a cut-off date. Accuracy of EBV was measured as the correlation between EBV and corrected phenotype (y_c) divided by accuracy of y_c. EBV from all three models were compared with the y_c from the linear model. The y_c was defined as $y_c = \hat{a}_d + \hat{a}_{m(dam)} + \hat{l} + \hat{e}$. By the definition, the accuracy of y_c was $r_{yc} = \sqrt{\frac{\sigma^2_{ad} + \sigma^2_{am} + \sigma_{adam}}{\sigma^2_l + \sigma^2_{ad} + \sigma^2_{am} + \sigma_{adam} + \sigma^2_e}}$. Dispersion bias for each model was measured by the regression of EBV from full data on EBV from reference data.

Results

For all traits and models, the estimates of direct and maternal heritabilities were very low, ranging from 0.005 to 0.017 (Table 1). For MortBS, direct heritabilities in underlying scale from the logit and the probit model were not higher than those in observed scale from the linear model, which was unexpected, indicating a challenge of analysing binary traits. Maternal heritabilities for MortBS was higher than those for Mort5S. The correlations between direct and maternal additive genetic effects was low and not significantly different from zero.

Table 1. Phenotypic variance, proportion of litter variance, direct (a_d) and maternal (a_m) heritability, correlation between a_d and a_m, as well as their standard error (±).

Trait	Model	σ_p^2	lit^2	h_{ad}^2	h_{am}^2	r_{adam}
Mort5S	Linear	0.134	0.051±0.0021	0.015±0.0026	0.005±0.0017	-0.119±0.1747
	Logit	3.648	0.071±0.0038	0.017±0.0033	0.010±0.0031	0.006±0.1884
	Probit	1.107	0.069±0.0039	0.016±0.0030	0.010±0.0031	0.032±0.1879
MortBS	Linear	0.177	0.061±0.0021	0.015±0.0026	0.010±0.0022	-0.182±0.1451
	Logit	3.665	0.076±0.0031	0.013±0.0024	0.014±0.0031	-0.069±0.1613
	Probit	1.128	0.084±0.0035	0.014±0.0025	0.017±0.0034	-0.050±0.1608

Although variance components and heritabilities were somewhat different among the three models, the EBVs from the three models are highly consistent (Table 2). For the validation animals, correlations between EBVs from logit and probit model were higher than 0.998. Correlations between EBVs from the linear model and the other two models were about 0.990 for direct EBVs and total EBVs, and ranged from 0.928 to 973 for maternal EBVs.

Regression coefficients of the EBVs obtained from full data on the EBVs from reference data for validation animals ranged from 0.932 to 1.194 (Table 3), indicating that dispersion bias was small for all models and traits. Regression coefficients for the logit and probit models were slightly less far from 1 than the counterparts from the linear model, indicating slightly less bias.

Despite low heritability, accuracy of EBV were not low (Table 4). Validation accuracies ranged from 0.415 to 0.420 for direct EBVs, from 0.192 to 0.392 for maternal EBVs, from 0.357 to 0.396 for total EBVs. The validation accuracies were very similar between the three models, except that accuracy of maternal EBV from the linear model was lower than those from the other two models. According to the results from the linear model, the total genetic variance for observed phenotypic values ($\sigma^2_{ad} + \sigma^2_{am} + \sigma_{adam}$) was 0.00248 for Mort5S and 0.00398 for MortBS. Selection on the sum of EBV_{ad} and EBV_{am} would reduce mortality by 1.95% for Mort5S, or 2.25% for MortBS, given selection intensity of 1.

Table 2. Correlations (upper diagonal for Mort5S and lower diagonals for MortBS) between EBVs from different models.

Model	EBV_{ad}			EBV_{am}			EBV_{ad+am}		
	Linear	Logit	Probit	Linear	Logit	Probit	Linear	Logit	probit
Linear	1	0.992	0.989	1	0.948	0.928	1	0.993	0.989
Logit	0.993	1	0.999	0.973	1	0.998	0.994	1	0.999
Probit	0.991	0.999	1	0.965	0.999	1	0.993	0.999	1

Table 3. Regression coefficients of EBVs from full data on EBVs from reference data.

Trait	Model	EBV_{ad}	EBV_{am}	EBV_{ad+am}
Mort5S	Linear	0.948	1.194	0.932
n=22,364	Logit	0.986	1.139	0.942
	Probit	1.002	1.103	0.946
MortBS	Linear	0.978	1.081	0.956
n=24,628	Logit	1.028	1.047	0.975
	Probit	1.038	1.038	0.978

Table 4. Validation accuracy of EBVs from different models.

Trait	Model	EBV_{ad}	$EBV_{am(dam)}$	$EBV_{ad+am(dam)}$
Mort5S	Linear	0.420	0.236	0.392
n=22,364	Logit	0.418	0.378	0.396
	Probit	0.419	0.392	0.395
MortBS	Linear	0.415	0.192	0.357
n=24,628	Logit	0.418	0.255	0.358
	Probit	0.419	0.259	0.357

Discussion

This study assessed feasibility of reducing pig mortality before slaughter by investigating genetic parameters and selection accuracy using three models. With the binary output of mortality at individual level, it was expected that the logit model and the probit model would perform better than the linear model in which the normality assumption is violated. However, the results showed that the EBVs from the linear model were highly consistent with the EBVs from the logit and probit models. The validation accuracies of the three models were also similar to each other. The negligible difference in prediction accuracy between linear model and probit or logit model have been found in simulation studies (e.g. Meijering and Gianola, 1985) and in the studies on mastitis in cattle (e.g. Koeck *et al.*, 2010). These results indicates that linear model is robust to violation of normality hypothesis, and thus it could be a good choice for genetic evaluation of pig mortality because it is simple to implement.

Heritabilities for the two mortality traits were lower than 0.02. However, accuracies of EBVs were moderate for direct EBV and total EBV, though relatively low for maternal EBV. This could be due to large information from relatives contributing to the estimation of breeding values. In the current study, an individual had, on average, about seven full-sibs and about 200 half-sibs in the data. According to the genetic parameters and accuracy of EBV obtained in this study, selection on total EBV would reduce mortality by $i\times1.95\%$ for Mort5S, and $i\times2.25\%$ for MortBS, where i is selection intensity. The realized reduction of mortality by including mortality in selection index will depend on the weight on mortality and the relationship of mortality with the other traits in the selection index.

Accuracies of EBV and expected genetic gains presented above are based on conventional selection. Genomic selection has been widely used in pig breeding. It has reported that selection accuracy and genetic gain increased up to 50% when changing conventional selection to genomic selection (e.g. Knol *et al.*, 2016). In most pig breeding companies, pure pigs in test are genotyped. When genomic information is used for prediction of pig mortality, the accuracies of EBV and the genetic progress in reducing mortality is expected to increase largely.

We conclude: even though heritability for mortality traits is low, accuracy of EBV is acceptable for selective breeding; linear mixed model is efficient for predicting breeding value of mortality traits; and selection to reduce pig mortality before slaughter is feasible.

References

Alonso-Spilsbury M., Ramirez-Necoechea R., Gonzlez-Lozano M., Mota-Rojas D. and Trujillo-Ortega M.E. (2007) J Anim Vet Adv 6: 76-86. https://medwelljournals.com/abstract/?doi=javaa.2007.76.86

Christiansen M.G. Seges Svineproduktion Rapport nr. 59. (2018) Available at: https:// svineproduktion.dk/-/media/PDF---Publikationer/Rapporter/Rapporter-2018/Rapport_59.ashx

Knol, E.F., Nielsen B., and Knap P.W. (2016) Animal Frontiers 6(1):15-22. https://doi.org/10.2527/af.2016-0003

Koeck A., Heringstad B., Egger-Danner C., Fuerst C., and Fuerst-Waltl B. (2010) J Dairy Sci 93:4351–4358. https://doi.org/10.3168/jds.2009-2955

Meijering A. and Gianola D. (1985) Genet Sel Evol 17(1): 115–132. https://doi.org/10.1186/1297-9686-17-1-115

Madsen P., Sørensen, P., Su, G., Damgaard L.H., Thomsen, H. *et al.* (2006) Proc. of the 8th WCGALP, Belo Horizonte, Brazil.

Su G., Lund M.S, and Sorensen D. (2007) J Anim Sci 85:1385-1392. https://doi.org/10.2527/jas.2006-631

USDA. Swine 2012 (2015) Available at: https:// aphis.usda.gov/animal_health/nahms/swine/downloads/swine2012/Swine2012_dr_PartI.pdf

Willham R.L. (1972) J Anim Sci 35(6):1288-93. https://doi.org/10.2527/jas1972.3561288x

779. Genotyping dead animals improves post-weaning survival of pigs in breeding programs

M. Sharif-Islam[1*], J.H.J van der Werf[2], M. Henryon[3], T.T. Chu[4], B.J. Wood[5] and S. Hermesch[1]

[1]AGBU, a joint venture of NSW Department of Primary Industries and University of New England, Armidale, NSW, 2351, Australia; [2]School of Environmental and Rural Science, University of New England, Armidale, 2351, Australia; [3]Danish Pig Research Center, SEGES, Axeltorv 3, 1609 Copenhagen V, Denmark; [4]Center for Quantitative Genetics and Genomics, Aarhus University, 8830 Tjele, Denmark. [5]School of Veterinary Science, University of Queensland, Brisbane, Queensland, 4072, Australia; mislam40@myune.edu.au

Abstract

A premise was tested that genotyping both surviving and dead pigs will realise more genetic gain in post-weaning survival (PWS) than genotyping only surviving animals. Stochastic simulation was used to estimate the rate of true genetic gain in different genotyping scenarios that differed in varying proportions of genotyping dead animals. Selection was for only PWS that had heritability of 0.02. Mortality was assumed 10%. The trait was controlled by 7,702 biallelic quantitative trait loci distributed across a 30 Morgan genome. We used 54,218 biallelic single nucleotide polymorphisms (SNPs) that were used in genomic prediction. Genotyping both surviving and dead animals realised 12 to 24% more genetic gain than genotyping only surviving animals. The power of detecting SNP effects increased when animals of extreme phenotypes are genotyped. Therefore, genotyping both surviving and dead pigs realised more genetic gain than genotyping only surviving animals.

Introduction

Post-weaning survival (PWS) is an economically important trait for growing pigs (Knol *et al.*, 2016; Hermesch *et al.*, 2014). The trait is easy to record but is lowly heritable (h^2=0.02-0.06) with low prediction accuracy (Leite *et al.*, 2021; Harper *et al.*, 2019). It can particularly benefit from genomic selection as shown by an increase in the prediction accuracy of genomically estimated breeding value (GEBV) for PWS from 20 to 50% when prediction included genomic information (Leite *et al.*, 2021; Knol *et al.*, 2016). However, one of the challenges when applying genomic selection to survival is that, for practical reasons, only surviving animals may be genotyped. This has the potential to introduce prediction bias and reduce accuracy (Leite *et al.*, 2021). In general, genotyping phenotypically contrasting animals can increase the genetic gain, accuracy of GEBV and decrease the bias of GEBV for continuous trait in comparison to genotyping only top animals because information on phenotypically extreme values results in more accurate estimates of SNP effects (Gowane *et al.*, 2019, Chu *et al.*, 2020). For the case of PWS, genotyping phenotypically contrasting animals refers to the scenario of genotyping both surviving and dead animals. This indicates that genotyping both surviving and dead animals is expected to realise more genetic gain than genotyping only surviving animals. Based on this reasoning, it was tested if genotyping all animals in the birth cohort realises more genetic gain (ΔG) for PWS than genotyping only surviving selection candidates.

Methods

Procedure. Stochastic simulation was used to estimate the ΔG_{true} realised for PWS. All surviving animals and 0, 20, 40, 60, 80, and 100% of randomly selected dead animals were genotyped. Genomic breeding values (GEBVs) were used in optimum contribution selection (OCS), and pedigree information was used to constrain inbreeding to 0.01 per generation. Post-weaning survival had an initial value of 90% and heritability of 0.02 on the observed scale (Harper *et al.*, 2019). It was controlled by 7,702 biallelic QTL distributed across a 30M genome. The genome contained 54,218 biallelic genetic markers that were used to calculate GEBV. Simulation of genomes in a founder population was detailed in the study by Henryon

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_779, © M. Sharif-Islam *et al.* 2022

et al. (2019). Breeding schemes were run for 10 discrete generations and replicated 30 times. Each replicate was initiated by sampling a unique base population from the founder population. Animals in the base populations were randomly selected in generation $t = 1$. In generations $t = 2 \ldots 10$, selection candidates were allocated matings by OCS.

Breeding scheme. Population consisted of 20 sires and 300 dams that were selected in every generation. Each sire was randomly mated with 15 dams. Each dam produced six progeny with a sex ratio of progeny of one to one.

Simulating phenotype. The true breeding value (TBV) was calculated as the sum of additive genetic effects at the 7,702 QTLs. Its residual value e_i was sampled from $e_i \sim N(0, \sigma_e^2 = 0.79$ on the underlying scale). The target observed-scaled heritability (h_o^2=0.02) was transformed to heritability on underlying scale (h_l^2 =0.06) according to the following formula (Dempster and Lerner, 1050):

$$h_l^2 = \frac{h_o^2 K(1-K)}{z^2} \quad (1)$$

Where K is the proportion of survivability incidence which was assumed 90% in this study, z is the height of normal distribution curve at threshold.

The liability of animal i was calculated as $l_i = a_i = e_i$, where a_i is the animal's TBV and e_i is the residual value. Phenotype of individual animal i, y_i was assigned as $y_i = 1$ when liability exceeded the threshold value for survivability of 90% and otherwise $y_i = 0$.

Genomic prediction. A liability threshold (probit) model was used to estimate the breeding value for the post-weaning survival trait.

$$l = Xb + Z_1 a + Z_2 c + e \quad (2)$$

where b is the fixed generation effect, a and c are vectors of random additive genetic effect and random common environmental effect respectively, e is the vector of random residual error. The co(variance) structure of the random effects was:

$$\begin{pmatrix} \boldsymbol{a} \\ \boldsymbol{c} \\ \boldsymbol{e} \end{pmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \boldsymbol{A}\sigma_a^2 & 0 & 0 \\ 0 & \boldsymbol{I}\sigma_c^2 & 0 \\ 0 & 0 & \boldsymbol{I}\sigma_e^2 \end{bmatrix} \right)$$

Where I is an identity matrix, σ_a^2=0.06 is the additive genetic variance on the underlying scale, σ_c^2=0.15 is the common litter variance and σ_e^2=0.79 residual variance on the underlying scale that was used to sample phenotype. Single-step genomic breeding value (ssGBLUP) was estimated using H matrix instead of A.

Comparisons. Rate of true genetic gain (ΔG_{true}) was calculated in each replicate as a linear regression of average true breeding values of animals born in generations $t = 5$ to 10 (G_t) on time t. Accuracy of GEBV was calculated in each generation as the correlation between true breeding values (TBVs) and GEBV of surviving animals. Generation averages were averaged across the 30 replicates. Bias of GEBV in each replicate was calculated as the regression of TBV on GEBV of surviving animals. All results are presented as the mean of 30 replicates.

The ADAM software (Pedersen *et al.*, 2009) was used for simulating a breeding program. The DMU package (Madsen *et al.*, 2006) was used for predicting GEBV and OCS was carried out by the EVA software (Berg *et al.*, 2009).

Results

Compared to traditional pedigree selection, the addition of genotype information on all surviving animals resulted in 70% more true genetic gain. Genotyping both surviving and dead pigs realised a higher ΔG_{true} compared with genotyping only surviving pigs (Table 1). Genotyping all surviving pigs and 20-100% of the dead pigs realized 12-23% more genetic gain than genotyping only surviving animals. Compared to only genotyping surviving animals, the greatest realised gains were made when 20% of dead animals were also genotyped (a 12% increase). Thereafter, increasing genotyping dead animals by 20% only resulted in a realised 4-6% increase.

Genotyping both surviving and dead animals increased accuracy of GEBVs. Genotyping all surviving pigs and 20-100% of the dead pigs generated GEBVs that were 12.5 to 18% more accurate respectively, than genotyping only surviving animals (Table 1). Meanwhile, bias was reduced by 7% when also genotyping all dead pigs.

Discussion

In this simulation study, the benefits of genotyping both surviving and dead pigs were assessed based on changes in genetic gain, prediction accuracy and bias of GEBV for PWS in surviving animals as compared to the genotyping only surviving pigs and pedigree selection. Since single-step method was used in this simulation, relationship information about ungenotyped dead pigs was available from the genotyped surviving pigs. As a result, genotyping only surviving pigs provided additional genetic gain and increased prediction accuracy compared to the pedigree selection. However, genotyping both surviving and dead pigs refers to the genotyping phenotypically contrasting animals and that realised more genetic gain and increased prediction accuracy in comparison to genotyping only surviving animals. The finding in this study is in agreement with other studies using continuous trait (Gowane *et al.*, 2019; Chu *et al.*, 2020). When animals of extreme phenotypes are genotyped, the power of detecting SNP effect increased because most of the genetic information is captured in animals of extreme phenotypes (Huang and Lin, 2007). This resulted in higher genetic gain in PWS when both surviving and dead animals were genotyped in this simulation study. One might argue that additional genetic gain in PWS due to genotyping dead pigs might be because of genotyping additional animals not because of dead animals specifically. To explore

Table 1. Rate of true genetic gain (ΔG_{true}, expressed on underlying scale), prediction accuracy and bias of estimated breeding value for post-weaning survival in different proportion of genotyping dead animals. Standard error is shown in parenthesis.[1]

Genotyping strategy	ΔG_{true}	Accuracy	Bias
Pedigree selection only	0.051 (0.002)	0.199 (0.007)	1.20 (0.05)
Genotype all surviving animals	0.087 (0.003)	0.264 (0.008)	1.22 (0.04)
Geno surviving + 20% dead	0.097 (0.004)	0.297 (0.007)	1.24 (0.03)
Geno surviving + 40% dead	0.101 (0.003)	0.317 (0.006)	1.29 (0.03)
Geno surviving + 60% dead	0.105 (0.002)	0.318 (0.006)	1.19 (0.03)
Geno surviving + 80% dead	0.107 (0.003)	0.322 (0.006)	1.23 (0.02)
Geno surviving + 100% dead	0.104 (0.003)	0.312 (0.006)	1.13 (0.02)

[1] Geno surviving = all surviving animals were genotyped. Results were the mean of 30 replicates.

this hypothesis, additional scenarios were simulated where equal number of live or dead animals were genotyped in addition to genotyping fixed numbers of live animals. Findings of additional simulations (results not shown) clearly indicated that additional genetic gain was achieved from genotyping both live and dead animals. Therefore, the hypothesis turned out to be true that genotyping both surviving and dead animals realised more genetic gain as compared to genotyping only surviving animals.

This study assumed PWS to be 90% so that the benefit of genotyping different proportions of dead animals could be shown. In reality, PWS is greater than 90% because of high biosecurity in nucleus populations. Variation of a binary trait depends on proportion of incidence. If the mortality rate ranges from 3 to 5% in the nucleus population, data will lack enough variation for breeding value estimation. The low incidence of PWS in nucleus populations could be circumvented by adding mortality information from commercial herds. Genotyping commercial dead pigs could be used to add crossbred pigs in the reference population, since relationships between animals can be efficiently estimated with genomic data instead of keeping pedigree. If the commercial dead pigs are genotyped and added in the reference population for selection in nucleus, prediction accuracy of GEBV for post-weaning survival may increase. However, the benefit of genotyping commercial population depends on the correlation between purebred and crossbred populations (Van Grevenhof and Van der Werf, 2015). Therefore, it is worth investigating the benefit of genotyping commercial dead pigs to increase the accuracy of GEBV for PWS in the nucleus.

References

Berg P. *et al.* (2006) Proc. Of the 8th WCGALP, Belo Horizonte, Brazil.

Chu T.T. *et al.* (2020) Front. Genet. 11(866). https://doi.org/10.3389/fgene.2020.00866

Dempster E.R. & Lerner I.M. (1950) Genetics. 35(2):212-236

Gowane G.R. *et al.* (2019) J Anim. Breed. Genet. 136:390-407. https://doi.org/10.1111/jbg.12420

Harper J., Bunter K.L. & Hermesch S. (2019) Proc. Of the 23rd AAABG, Armidale, Australia.

Henryon M., Liu H., Berg P. *et al.* (2019) Genet. Sel. Evol. 51(39). https://doi.org/10.1186/s12711-019-0475-5

Hermesch S., Ludemann C.I. & Amer P.R. (2014) J Anim. Sci. 92(12): 5358-5366. https://doi.org/10.2527/jas.2014-7944

Huang B.E. & Lin D.Y. (2007) Am. J Hum. Genet. 80(3):567-576. https://doi.org/10.1086/512727.

Knol E.F. *et al.* (2016) Anim. Front. 6(1):15-22 https://doi.org/10.2527/af.2016-0003

Leite N.G. *et al.* (2021) Anim. Sci. J. 99(8). https://doi.org/10.1093/jas/skab217

Madsen P. *et al.* (2006) Proc. Of the 8th WCGALP, Belo Horizonte, Brazil.

Pedersen L.D. *et al.* (2009) Livest. Sci. 121:343-344. https://doi.org/10.1016/j.livsci.2008.06.028

Van Grevenhof I.E.M. & Van der Werf J.H.J. (2015) Genet. Sel. Evol. 47(15). https://doi.org/10.1186/s12711-015-0104-x

780. GBP5 PRRSV resistance gene had no effect on pigs' infectivity or susceptibility in a trial simulating natural infections

M.E. Chase-Topping[1*], G. Plastow[2], J. Dekkers[3], Y. Fang[4], V. Gerdts[5], J. van Kessel[5], J. Harding[6], T. Opriessnig[1] and A. Doeschl-Wilson[1]

[1]The Roslin Institute, University of Edinburgh, Easter Bush, United Kingdom; [2]Department of Agricultural, Food and Nutritional Science, University of Alberta, AB, Canada; [3]Department of Animal Science, Iowa State University, IA, USA; [4]College of Veterinary Medicine, Department of Pathobiology, University of Illinois Urbana-Champaign, IL, USA; [5]Vaccine and Infectious Disease Organization-International Vaccine Centre, University of Saskatchewan, Saskatoon, Canada; [6]Department of Large Animal Clinical Sciences, University of Saskatchewan, Saskatoon, Canada; margo.chase@ed.ac.uk

Abstract

A bespoke small-scale porcine reproductive and respiratory syndrome virus 2 (PRRSV2) transmission experiment was designed to gather evidence on whether the previously identified *GBP5* PRRS resistance gene also confers lower susceptibility and/or infectivity of pigs under natural PRRSV transmission. Commercial full- and half-sib piglets (n=164) were assigned as either inoculators, shedders, or contact pigs. Pigs were grouped according to their relatedness structure and their *GBP5* genotype, and barcoding of the PRRSV2 strain (SD09-200) was applied to track pig genotype-specific transmission. Blood samples were collected on consecutive trial days and the area under the curve (AUC) of the $\log TCID_{50}$ viral load using rt-PCR was used as response variable general linear mixed model analyses. These provided no supportive evidence that the *GBP5* resistance gene confers differences in host susceptibility or infectivity. However, variation in AUC among sibling groups does provide evidence that PRRSV2 transmission is partly genetically controlled.

Introduction

Porcine reproductive and respiratory syndrome (PRRS) arguably remains one of the most important infectious diseases for the pig industry worldwide. Despite biosecurity measures and vaccination, PRRS prevalence continues to be high, with estimated production losses of around 664 million US dollars per year (Holtkamp *et al.*, 2013). The disease is commonly associated with reproductive failure, and respiratory disease, fever, reduced growth, and mortality in piglets and breeding animals. Quantitative genetics studies have identified substantial genetic variation in the resistance of pigs to PRRS, with a single gene on pig chromosome 4 (*GBP5*, encoding guanylate-binding protein 5) explaining 15% of the genetic variation in viral load and 11% of the genetic variation for growth rate in nursery pigs infected with PRRSV2 (Boddicker *et al.*, 2014; Koltes *et al.*, 2015). Although these results offer promising opportunities for mitigating PRRS through genomic selection, predicting the impact of such selection on PRRS prevalence is difficult, as the role of host genetics and in particular of the *GBP5* resistance gene in PRRSV transmission is currently not known.

The overall aim of this study was to determine the role of the *GBP5* resistance gene on PRRSV2 transmission. To this purpose, a bespoke PRRSV transmission experiment was designed. The experiment differed from conventional transmission experiments in two key aspects: (1) a specific genetic-relatedness structure and *GBP5* genotype composition of pigs in the different contact groups to maximise statistical power for estimating genotype effects on PRRSV transmission,;and (2) barcoding of the PRRS virus to trace pig genotype specific transmission routes.

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_780, © M.E. Chase-Topping *et al.* 2022

Materials & methods

Transmission experiment and statistical models. A total of 164 commercial cross-bred pigs (full-sibs or paternal half-sib progeny from ~70 sows, post weaning, up to 42 days old) were used in this study. Approximately half of the pigs from each sibling group carried (R+) or did not carry (R-), a copy of the dominant *GBP5* resistance allele. The study design consisted of two steps (Figure 1): step 1 involved infection of shedder pigs through natural exposure to PRRSV infected inoculation pigs, while step 2 was the main transmission trial to assess genotype effects on shedder pig infectivity and contact pig susceptibility. After considering parentage and *GBP5* genotype (Figure 1 and as outlined below), pigs were randomly assigned to one of the following roles: (1) Inoculation pig (n=20); (2) Shedder pig (n=48); (3) Contact pig (n=96). Full-sib inoculation pigs (all R-, from 10 unrelated parents) were split into two rooms, where they were inoculated with either a bar-coded (room A, blue, Figure 1) or a non-barcoded (room B, red, Figure 1) version of a virulent PRRSV2 strain circulating in North American pig farms (SD09-200; infection dose 10^5 $TCID_{50}$). Previous infection trials confirmed that barcoding had no effect on the infectivity or pathogenesis of the virus. One day later, 24 sibling pairs of shedder pigs were introduced into the two rooms (each consisting of equal proportion of pigs of each genotype and with sibs of the same R genotype split across

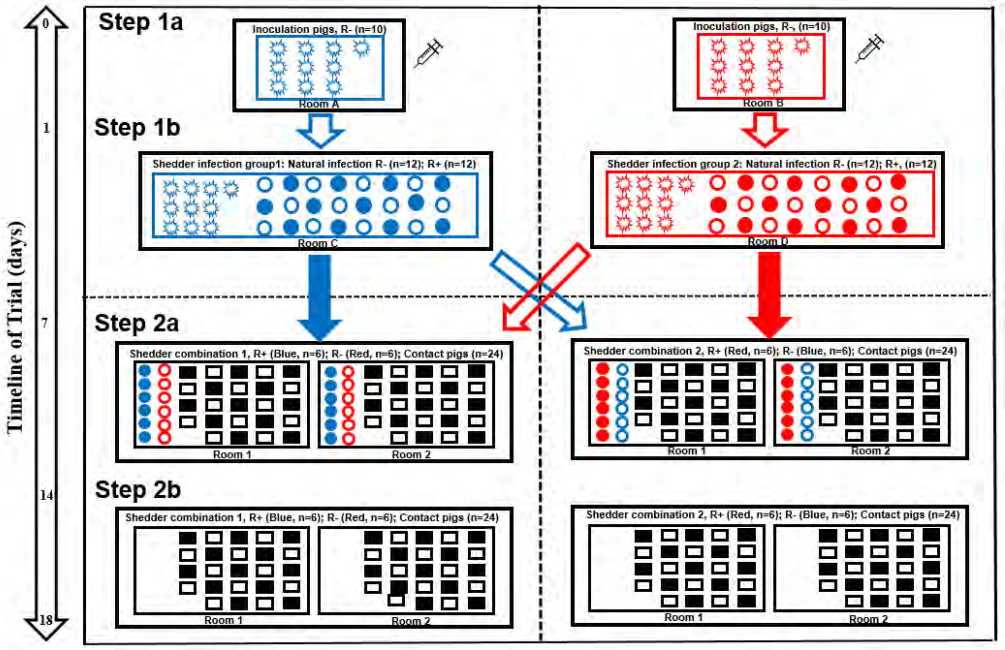


Figure 1. Experimental design. Schematic of the design of the trial comprising creation of the shedder pigs (Step 1) for use in the main transmission trial (Step 2). Step 1, a. direct infection of the inoculation pigs (day 0; challenge dose was 10^5 $TCID_{50}$ in a total volume of 4 ml, and the challenge route was intramuscularly (IM) and intra-nasally (IN) with 2 ml IM and 1 ml for each nostril) and b. infection of shedder pigs through natural transmission (day 1 to day 6); and Step 2, a. re-assortment of the naturally infected shedder pigs and placement in rooms with naive contact pigs (day 7) and b. removal of the shedder pigs (day 14). Jagged circles, directly inoculated pigs; Smooth circles, naturally infected shedder pigs; Squares, contact pigs; closed symbol R+ genotype, open symbol R-; Blue, Shedder infection group 1; Red, Shedder infection group 2. Arrows indicate the direction of the movement of pigs (R+ solid arrows; R- open arrows).

rooms), where they were in direct contact with the inoculated pigs. At day 7, infected shedder pigs were transferred into 4 new rooms, according to their genotypes and relatedness (sibs of the same R genotype split across rooms to avoid confounding between relatedness and R genotype effects), where they could transmit PRRSV2 to resident naïve contact pigs (24 per room, 12 pigs per R genotype; pigs of the same R genotypes from different sibling groups, and sibs of different R genotypes in the same room).

To minimise re-infection of shedder pigs, they were removed at day 14. The trial lasted only 18 days, enough time to ensure contact pigs were infected by shedders, and contact-contact pig transmission was limited. Sampling involved collecting blood samples on trial day (d) d6, d8, d9, d10, d14 and d17. For statistical analyses, the area under the curve (AUC) of the log$TCID_{50}$ viral load derived from rt-PCR of the serum samples of all shedder and contact pigs was generated using the spline rule (R version 4.1.0). To determine genetic and R genotype effects on contact pig susceptibility and shedder pig infectivity, multivariable models were fitted to the data using General Linear Mixed Models (GLMM, SAS Proc Mixed), with the shedder or contact pig's AUC as dependent variables and the shedder and contact pig R genotypes, the shedder infection group, shedder combination, and interactions as fixed effects. Sibling-group and room (step 2, contact pigs only) were added as random effects.

Results

Shedder pigs. All but 2 shedder pigs were viraemic at d6, d8, d10, and d14. All pigs were infected by pigs within the same shedder infection group (step 1b, Figure 1). Analysis of AUC showed that there was no significant R genotype effect on shedder infection level, regardless of the shedder infection group the pigs originated from (shedder infection group×shedder genotype, P=0.49; Figure 2A).

Contact pigs. All but 1 contact pig was viraemic at some point in the transmission trial. Only eleven (11%) were viraemic at d9 (2 days post contact (dpc) with the shedder pigs), however 96% (n=92) were viraemic at

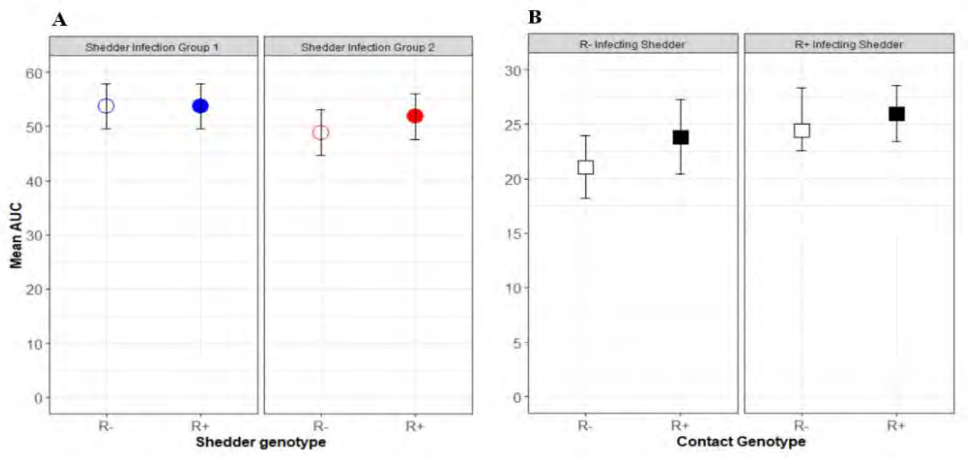


Figure 2. GLMM results for shedder pig (A) and contact pig (B) models of serum Area Under the Curve (AUC) representing the level of infection. For shedder pigs the results are displayed with respect to the shedder infection group they originated from and their *GBP5* genotype (R-, R+). For contact pigs the results are displayed with respect to the genotype of the shedder pig (R-, R+) they were infected by (indicating genotype effect on shedder infectivity) and their own genotype (R-, R+) (indicating genotype effect on contact pig susceptibility). Error bars are 95% confidence intervals.

d14 (7dpc). Multiple infections from shedder groups with different R genotypes were detected in only three pigs. Analysis of AUC showed that there was no significant R genotype effect on contact pig infection level, regardless of the R genotype of the contact pig or of the shedder pig that transmitted the infection (Infected by R+ shedder×contact genotype, P=0.46; Figure 2B).

Variation between sibling groups. The study design relied on the use of full- or half-sibs with contrasting R genotypes. Figure 3 shows the area under the curve for the shedder (Figure 3A) and contact (Figure 3B) sibling groups. There was considerable variation between sibling groups for both the shedder and contact pigs, and the addition of sibling-group improved the fit of the GLMM (lower BIC), significantly for contact pigs (log-likelihood ratio (P<0.025).

Discussion

This pilot study introduces a novel experimental design to assess genetic effects on disease transmission. However, it provides no supportive evidence that the previously identified *GBP5* resistance gene confers differences in host susceptibility or infectivity under natural PRRSV transmission. However, recent research (Jeon *et al.*, 2021) found that the *GBP5* resistance gene also confers differences in pigs' resilience when subjected to a natural polymicrobial disease challenge. Hence, given the relatively small scale of the present study, and the potential implications of the results to the pig industry, it is paramount for the results of this study to be validated. Finally, although the present study did not identify significant associations of *GBP5* with PRRSV transmission, it did provide strong evidence that PRRSV transmission is under host genetic control, as evidenced by the significant sibling-group effects in the models. Future studies should focus on investigating the genetic architecture underlying PRRSV transmission as an important component for reducing PRRS prevalence through genetic selection.

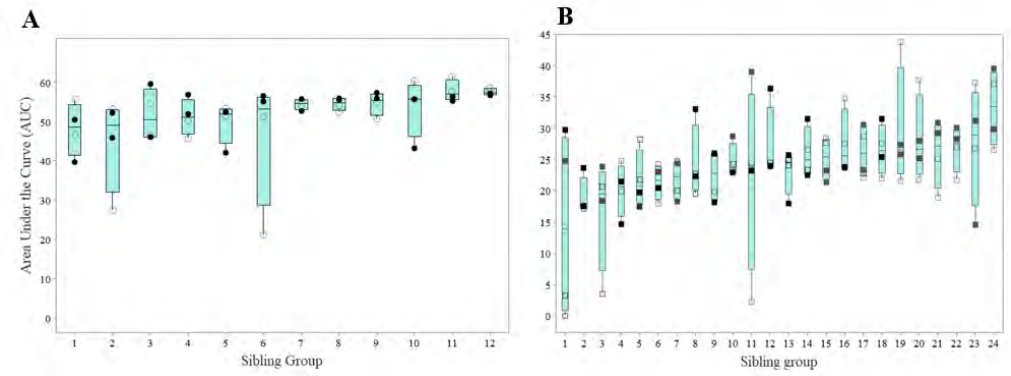


Figure 3. Differences in viral load as measured by Area Under the Curve (AUC) for the A. shedder (n=12) and B. contact pig (n=24) sibling groups. Data is organized based on increasing median AUC. Open circles/squares, R- genotype; closed circles/squares, R+ genotype

Acknowledgements

This study was funded by Genome Alberta, PigGen Canada and the Roslin Institute Strategic Programme grants (BBS/E/D/20002172 & /30002275). Pigs were supplied by Fast Genetics.

References

Boddicker N.J., Garrick D.J., Rowland R.R.R., Lunney J.K., Reecy J.M., *et al.* (2014) Anim Genet.45:48–58. https://doi.org/10.1111/ age.12079

Holtkamp D.J., Kliebenstein J.B., Neumann E.J., Zimmerman J.J., Rotto H.F. (2013) J. Swine Health Prod. 21: 72–84. https:// www.aasv.org/shap/issues/v21n2/v21n2p72.html

Jeon, R.L., Cheng, J., Putz, A.M., Dong, Q., Harding, J.C.S., *et al.* (2021) Livestock Sci. 244:104399. https://doi.org/10.1016/j.livsci.2021.104399

Koltes J.E., Fritz-Waters E., Eisley C.J., Choi I., Bao H., Kommadath A., *et al.* (2015) BMC Genomics. 16:412. https://doi.org/10.1186/s12864-015-1635-9

781. Unravelling the genetic basis governing the porcine metabolism

S. Bovo[1*], G. Schiavo[1], F. Fanelli[2], A. Ribani[2], F. Bertolini[1], V.J. Utzeri[1], M. Gallo[3], G. Galimberti[4], S. Dall'Olio[1], P.L. Martelli[5], R. Casadio[5], U. Pagotto[2] and L. Fontanesi[1]

[1]Department of Agricultural and Food Sciences, Division of Animal Sciences, University of Bologna, Viale Giuseppe Fanin 46, 40127 Bologna, Italy; [2]Department of Surgical and Medical Sciences, Endocrinology Unit, University of Bologna, Via Massarenti 9, 40138 Bologna, Italy; [3]Associazione Nazionale Allevatori Suini, Via Nizza 53, 00198 Roma, Italy; [4]Department of Statistical Sciences 'Paolo Fortunati', University of Bologna, Via Belle Arti 41, 40126 Bologna, Italy; [5]Biocomputing Group, Department of Pharmacy and Biotechnology, University of Bologna, Via San Giacomo 9/2, 40126 Bologna, Italy; samuele.bovo@unibo.it

Abstract

We present a first catalogue of genetic variants influencing the porcine metabolome and an *in silico* reconstruction of metabolic pathways in two pig breeds: Large White and Duroc. For about 1,300 pigs, targeted metabolomic profiles (186 metabolites each), were used for modelling metabolite relationships via Gaussian Graphical Models. Metabolic networks presented breed similarities as well as specific features. Coupling via genome scans the metabolomic and the genomic profiles (>60,000 variants each), we identified >50 DNA markers and >20 candidate genes controlling the plasma concentrations of about 40 metabolites. Overall, these results pointed out several similarities between the porcine and the human metabolism, that strength the usefulness of the pig as animal model. The obtained catalogue of genomic variants, in addition to the medium-high genomic heritability obtained for several metabolites (h^2>0.50-0.60), represents a relevant source of information for the use of metabolomic markers as intermediate phenotypes to explain more complex traits.

Introduction

The genome of an organism, its proteome and metabolome represent the simplest biological layers that interact each other and generates networks of biological functions (Fontanesi 2016). Processes and pathways established in this additional complex layer talk to each other and define cell types, organs and the organism itself. It is noticeable that also a small alteration of any biological layers can generate altered phenotypes. In livestock, these includes exterior, production, reproductive and pathological (or disease related) traits. Hence, there is a strong interest in disentangling such biological complexity to develop tools for nutrigenomics and genomic selection as well as for precision livestock feeding and breeding goals.

The introduction of high-throughput sequencing and phenotyping technologies has made it possible to rapidly obtain near complete readouts of the different molecules and features characterizing each layer. This in turn has led to the development of phenomics, that is the use of large scale approaches to study how genetic instructions translate into a final phenotype/trait (Houle *et al.* 2010; NIFA-NSF Phenomics Workshop Report 2011). In its simplest form, gene variants are responsible for protein (enzyme) forms that differ in the processing of metabolites lastly impacting metabolic pathways and defining a phenotype. Therefore, metabolites represent internal (or intermediate) phenotypes used to dissect more complex phenotypes like performance traits (Suhre and Gieger 2012). This changed the paradigm of linking genetic variants to performance traits (e.g. Bertolini *et al.* 2018) via genome wide association studies (GWAS) to linking genetic variants to metabolite levels (m) representing a final phenotype via mGWAS.

With a set of pilot studies, we previously showed that different porcine metabolic routes are established by the different sexes or pig breeds (Bovo *et al.* 2015, 2016). Here, by merging genomic and metabolomic profiles (>60,000 variants and >180 metabolites) of about 1,300 pigs, we present a first catalogue of genetic

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_781, © S. Bovo *et al.* 2022

variants influencing the porcine metabolism plus its *in silico* reconstruction. This catalogue represents a great source of information that can be used to improve and develop additional tools in the fields of precision livestock feeding and breeding.

Materials & methods

Animals. About 1,300 performance tested pigs, including 900 Italian Large White (ILW) and 400 Italian Duroc (IDU) animals, were used in this study. Animals were from triplets of siblings from the same litter (two females and one castrated male) which were individually performance tested for genetic evaluation of a boar from the same litter (sib-testing). Animals were fed with a standard commercial feed and at the end of the test they were slaughtered after overnight fasting. Blood was collected just after jugulation into a tube with EDTA. Samples were refrigerated, processed and stored at -80 °C.

Phenotypes, genotypes and data processing. The abundance level of 186 metabolites covering five compound classes (acylcarnitines, amino acids, biogenic amines, glycerophospholipids and sugars) was determined via mass spectrometry-based targeted metabolomics of plasma samples using the Biocrates AbsoluteIDQ p180 Kit. For each breed, data quality checks considered the percentage of missing points per sample, the presence of outliers and the evaluation of inter-plate coefficients of variation. Abundances were normalized via the Box-Cox transformation. Confounding effects were removed via linear models (LMs). Analyses run in R v.3.4.4 (https://www.r-project.org).

All pigs were genotyped with the Illumina PorcineSNP60 BeadChip v.2 (61,565 SNPs). PLINK v.1.7 (Purcell *et al.* 2007) was used for within breed quality checks, including the evaluation of the call rates, Minor Allele Frequency and Hardy Weinberg Equilibium.

Metabolic network generation. The dependence between metabolite concentrations was investigated with a Gaussian Graphical Model (GGM; Krumsiek *et al.* 2011). Briefly, residuals of the LMs were used to compute partial correlations coefficients (PCC). After having computed the statistical significance of each PCC, Bonferroni correction with α=0.05 was applied to retain reliable relationships. Networks were annotated with Cytoscape v.3.0.1 (Shannon *et al.* 2003).

Genome scans. GWAS were carried out via Linear Mixed Models (LMM) as implemented in GEMMA (Zhou and Stephens 2012). Briefly, residuals of the LM were used as input phenotype and additive genetic models were tested. A genomic relatedness matrix was included in the model. The proportion of variance in phenotype explained by a given variant (PVE%) and the genomic heritability were concurrently calculated. Single metabolites concentrations and metabolites ratios were evaluated. Bonferroni correction with α=0.05 was applied to retain reliable associations. For metabolite ratios we used the *p-gain* statistics (Petersen *et al.* 2012). Meta-GWAS were carried out with METAL (Willer *et al.* 2010).

DNA sequencing. Two DNA pools (one for ILW and one for IDU; 35 pigs each) and 50 single animals (25 ILW and 25 IDU) were subjected to 40× and 20× whole genome sequencing (WGS), respectively. Reads were aligned over the Sscrofa11.1 reference genome with BWA tool 0.7.17 (Li and Durbin, 2009) and variants were called with GATK4 (McKenna *et al.* 2010).

Results

The quality control led to retain 169 and 164 metabolites for ILW and IDU, respectively. The merged dataset counted 160 metabolites.

The porcine metabolic network. A total of 160 metabolites entered in the network reconstruction process. The resulting GGM (Figure 1) counted 283 edges. No singletons were obtained. PCC values ranged from 0.15 to 0.78 and were computed both within the single breed and in the whole population. Few PCC were statistically significant only within a specific breed: no. 5 for IDU and no. 20 for ILW. The network was characterized by modules that clearly represented the specific metabolism of the different metabolite classes, poorly interconnected each other. The highest PCC was obtained between metabolites separated by one reaction step.

Genome scans. Genome wide association analyses of single metabolite concentrations (Figure 2) identified at least one association for 29 and 23 metabolites in ILW and IDU, respectively. The most significant association (p=1.83×10^{-32}) was detected for Kynurenine levels in ILW pigs. Analysis of metabolites ratios (Figure 2) pointed out several other associations; the most significant one (p-$gain$=7.37×10^{47}) was a metabolite of the PC group. Genomic heritability was in the ranges 0.0-0.52 (mean=0.22) and 0.0-0.62 (mean=0.15) for ILW and IDU, respectively. Candidate genes were further globally evaluated via sequencing for the detection of possible causative mutations.

Discussion

The reconstruction of porcine plasma metabolite networks led to evaluate several features of the pig metabolism. One of them is its similarity (considering the top PCC) to the GGM presented in human by Krumsiek *et al.* (2011), that remarks how the porcine metabolism is genetically governed similarly to what is known in humans. The results open new research routes for translational sciences. It was interesting

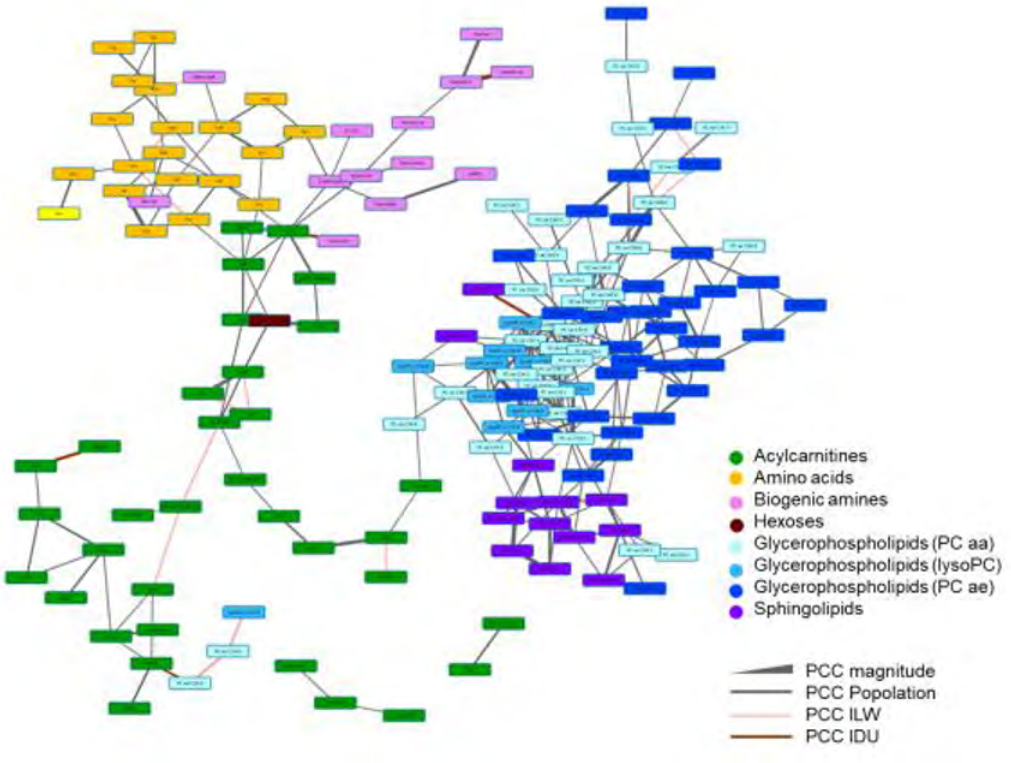


Figure 1. Reconstructed porcine metabolic network (Gaussian Graphical Model).

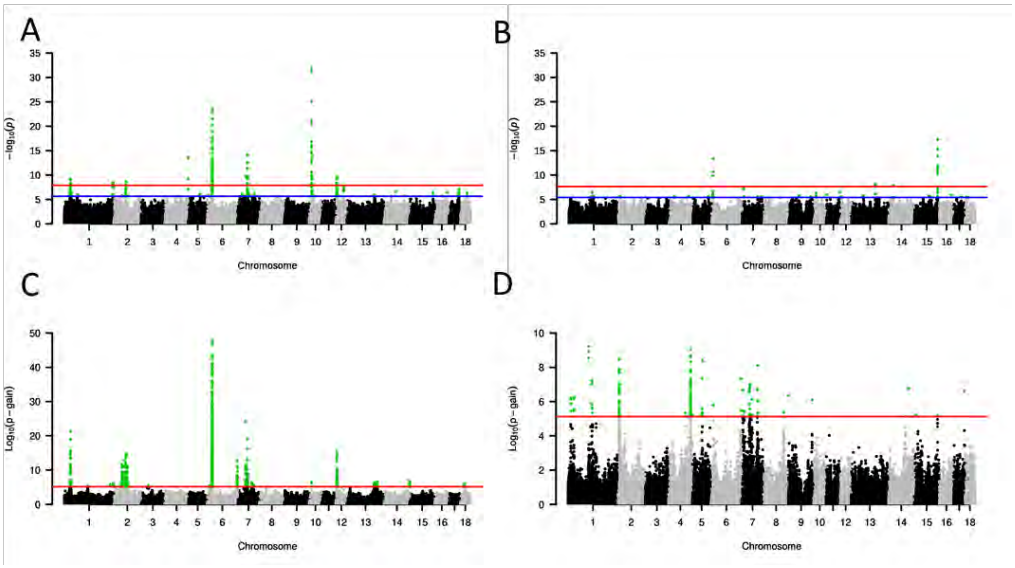


Figure 2. Overlaid Manhattan plots for the GWAS analysis of A-B) single metabolites in ILW and IDU pigs and C-D) metabolite ratios in ILW and IDU pigs. Green dots represent statistically significant associations.

to note the presence of breed specific metabolite relationships, suggesting that attention should be given when developing strategies and tools for precision livestock feeding and breeding. To further boost such strategies/tools, we have disclosed a several genomic markers, candidate variants and genes controlling up to 40 metabolite levels (21% of the metabolic profile). Together, the knowledge of both metabolite-metabolite relationships and the effect that variants can have on them gives the possibility to fine tune breeding programs using metabolites as internal phenotypes to improve animal productivity, welfare and obtain sustainable productions.

References

Bertolini F., *et al.* (2018) Animal 12(9):1777-1784. https://doi.org/10.1017/S1751731118000757

Bovo S., *et al.* (2015) J. Anim. Sci. 93(12):5681-5693. https://doi.org/10.2527/jas.2015-9528

Bovo S., *et al.* (2016) Animal 10(10):1741-1748. https://doi.org/10.1017/S1751731116000483

Fontanesi L. (2016) Animal Frontiers 6(1):73-79. https://doi.org/10.2527/af.2016-0011

Houle D., Govindaraju D.R., and Omholt S. (2010) Nat. Rev. Genet. 11:855-866 https://doi.org/10.1038/nrg2897

Krumsiek J., Suhre K., Illig T., Adamski J., Theis F. J. (2011) BMC Syst Biol. 5:21. https://doi.org/10.1186/1752-0509-5-21

Li H., Durbin R. (2009) Bioinformatics 25(14):1754-1760. https://doi.org/10.1093/bioinformatics/btp324

McKenna A., *et al.* (2010) Genome Res. 20(9):1297-1303. https://doi.org/10.1101/gr.107524.110

NIFA-NSF Phenomics Workshop Report. Available at: https://www.nsf.gov/bio/pubs/reports/phenomics_workshop_report.pdf

Petersen AK., *et al.* (2012) BMC Bioinformatics 13:120. https://doi.org/10.1186/1471-2105-13-120

Purcell S., *et al.* (2007) Am J Hum Genet. 81(3):559-575. https://doi.org/10.1086/519795

Shannon P., *et al.* (2003) Genome Res. 13(11):2498-2504. https://doi.org/10.1101/gr.1239303

Suhre K., Gieger C. (2012) Nat. Rev. Genet. 13:759–769. https://doi.org/10.1038/nrg3314

Willer C.J., Li Y., Abecasis G.R. (2010) Bioinformatics 26(17):2190-2191. https://doi.org/10.1093/bioinformatics/btq340

Zhou X., Stephens M. (2012) Nat. Genet. 44(7):821-824. https://doi.org/10.1038/ng.2310

782. The value of innate sow appetite as a model trait for maternal breeding objectives

L. Vargovic*, K.L. Bunter and S. Hermesch

AGBU a joint venture of NSW Department of Primary Industries and University of New England, 2351, Armidale, Australia lvargov2@une.edu.au

Abstract

The concept presented here evaluated benefits of including 'innate appetite' as an additional source of information to enhance response within a maternal breeding objective. The term 'innate appetite' refers to (heritable) feed intake (or comparable feeding behaviour) data recorded at any point in a sows' life-cycle, reflecting an animals desire to eat. When 'innate intake' was included as a selection criterion, the overall response increased from $41.98 to $47.31. The main proportion of the overall response was attributed to traits recorded on growing pigs. At the same time, farrowing and pre-weaning survival, wean-to-conception interval and longevity improved. Maternal breeding objectives in pigs can be extended by using innate appetite data, provided that uncertainty of genetic parameters, in particular between traits of growing pigs and sows is clarified.

Introduction

Data on feed intake is typically recorded on growing pigs from terminal lines only. As an alternative, data reflecting a sows' desire to eat (i.e. innate appetite), recorded during some period of the sows' life-cycle, can potentially provide information correlated to feed intake. Data recorded on sows can be obtained from electronic sow feeding systems, potentially providing an additional source of readily available information for all breeding female relatives of selection candidates. The usefulness of this information for breeding purposes depends on the genetic variability of these phenotypes and their relationship to feed intake. Feed intake in growing and mature animals is moderately correlated (Bunter *et al.* 2010), when circumstances enable appetite to be expressed. When feeding is *ad libitum,* feed intake traits are moderately heritable in both growing pigs and sows (Bunter *et al.* 2010; Bergsma *et al.* 2013). In contrast, when feed intake is restricted, the heritability for feed intake in growing pigs (Hermesch *et al.* 1999) and gestating sows is reduced but heritability of feeding behaviour traits is seen, which also contribute to feed intake (Bunter *et al.* 2018; Vargovic *et al.* 2021b).

Feed intake, or a correlated feeding behaviour trait, can be recorded at any stage of a sows' life-cycle, and therefore this study refers to this trait generically as reflecting 'innate' appetite. It was assumed that this appetite trait is present generally throughout an animals' entire life, and used it as a model trait. The aim of this study was to investigate the impact of recording an 'innate appetite' trait on sows for response in a maternal breeding goal. The hypothesis was that data recorded on dams would improve selection in a maternal breeding objective.

Materials & methods

Traits. A maternal breeding objective (BO), including production and reproductive traits, has been previously derived for the Australian pig industry (Amer *et al.* 2014). This BO was adapted to replace feed conversion ratio with average daily feed intake (Table 1). The BO traits were average daily gain (ADG: g/day), backfat thickness (BF: mm), average daily feed intake (ADFI: kg/day), piglets born alive (NBA: pigs/litter), farrowing survival (FS: 0/1 for survival), pre-weaning survival (PWS: 0/1 for survival), wean to conception interval between the first and second parity (WCI: days), longevity (LONG: parities) and sow mature weight (MWT: kg). A consensus of assumed genetic parameters was adapted from several

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_782, © L. Vargovic *et al.* 2022

Table 1. Economic weight (EW, $/gilt), phenotypic standard deviation (PSD), heritability (diagonal), phenotypic and genetic (above and below diagonal) correlations.

Trait[1]	ADG	BF	ADFI	NBA	FS	PWS	WCI	LONG	MWT	IAP
EW	2.65	-28.61	-597.48	91.93	107.17	1092.88	-3.61	86.90	-4.17	-
PSD	69.2	1.86	0.37	2.73	0.23	0.11	8.98	1.84	21.26	0.84
ADG	0.21	0.11	0.20	-0.04	-0.01	0.03	0.02	-0.05	0.32	0.14
BF	0.02	0.38	0.12	0.02	0.05	0.03	-0.02	0.10	-0.01	-0.04
ADFI	0.49	0.32	0.25	-0.01	0.00	0.00	-0.10	0.00	0.07	0.10
NBA	-0.19	-0.02	-0.15	0.09	0.07	0.07	-0.02	0.22	0.00	0.07
FS	-0.01	0.00	0.00	0.13	0.13	0.06	-0.15	0.12	0.18	0.10
PWS	0.27	0.07	0.00	-0.19	0.28	0.05	-0.21	0.00	0.16	0.25
WCI	-0.09	-0.24	-0.15	-0.20	0.09	-0.15	0.08	-0.05	0.00	-0.02
LONG	-0.28	0.35	-0.21	0.30	-0.25	0.18	-0.22	0.14	-0.03	0.15
MWT	0.30	-0.12	0.27	-0.21	0.09	-0.22	0.00	0.10	0.18	-0.16
IAP	0.29	0.20	0.39	0.01	-0.05	0.06	0.18	-0.20	0.20	0.24

[1] ADG = Average daily gain (g/day), BF = Backfat thickness (mm), ADFI = Average daily feed intake (kg/day), NBA = Born alive piglets (piglets/litter), FS = Farrowing survival (0/1 for survival), PWS = Pre-farrowing survival (0/1 for survival), WCI = Wean to conception interval (days), LONG = Longevity (parity), MWT = Sow mature weight (kg), IAP = Innate sow appetite (kg/day)

Australian studies (Tholen *et al.* 1996; Bunter *et al.* 2008; Bunter *et al.* 2010) and one Dutch study (Bergsma 2011). The initial genetic parameters for innate sow appetite (IAP, kg/day) were adapted from data for lactation feed intake (Bunter *et al.* 2010; Bergsma 2011).

Breeding schemes and analyses. Trait responses were obtained by using the MTIndex program (van der Werf 2021). Predicted responses were for a single generation of selection with a selection intensity of one. It was assumed that ADG and BF were available at selection on the selection candidate, dam, sire, six full-sibs and 40 half-sibs. For NBA, FS, PWS, WCI and IAP, dams had two records (one record for WCI) and three half-sibs. For LONG and MWT dams had one record. The first scenario (Base0) included recording all traits, except ADFI and IAP. The Base1 scenario included recording on all traits, except ADFI. The Base1 scenario was also used for simple sensitivity analyses, to investigate the impact of uncertain correlations between IAP and the other traits for responses in the individual traits and the overall index. Sensitivity analyses were conducted by changing genetic correlations (r_g) between IAP and other traits, keeping heritability constant, one at a time, by plus or minus 0.10, where possible, as shown in Table 3 and Figure 1. This change was only considered possible if correlations between all other traits combinations remained constant.

Results

Adding data for IAP increased the overall selection response by 13% (Base0 vs Base1: $41.98 vs $47.31/gilt, Table 2) with the assumed genetic parameters. Recording IAP (Base1) resulted in additional desired response for all traits, except ADG and NBA. The largest improvements were for ADFI (Base0 vs Base1: $22.69 vs $26.32) and LONG (-$5.43 vs -$2.61). Modifying r_g between IAP and BO traits was conducted within boundaries to keep the variance-covariance matrix positive definite, and for each trait was possible in only one direction (Table 3). These modifications of r_g between IAP and BO traits had an impact on the overall response relative to Base1 (Table 3). Changing r_g of IAP with ADG or ADFI had the largest impact on the overall response (97.13 and 104.55%) and for individual traits (Figure 1). Traits that were the most sensitive to this change were ADFI, BF, NBA and LONG (Figure 1). When r_g with other BO traits were changed, the overall responses varied less.

Table 2. Responses in breeding objective traits ($/gilt) for Base0 and Base1 scenarios.

Scenario	ADG[1]	BF	ADFI	NBA	FS	PWS	WCI	LONG	MTW	Total
Base 0	-3.51	15.78	22.69	11.33	0.07	0.44	-0.61	-5.43	1.23	41.98
Base 1	-4.82	14.54	26.32	9.88	0.28	1.08	-0.19	-2.61	2.84	47.31

[1] For trait abbreviations see Table 1.

Table 3. Responses in overall breeding objective (BO, $/gilt) with changes in the genetic correlations (r_g) between innate sow appetite (IAP) and other traits.

Scenario	ADG[1]	BF	ADFI	NBA	FS	PWS	WCI	LONG	MTW
Change in r_g[2]	+0.10	-0.10	+0.10	*-0.10*	+0.10	*-0.10*	-0.10	+0.10	-0.10
Response in BO[3]	-1.36	-0.13	2.15	1.29	-0.14	0.42	-0.16	-0.96	-0.56
Response in %[3]	97.13	99.72	104.55	102.73	99.70	100.89	99.67	97.97	98.82

[1] For trait abbreviations see Table 1.

[2] Italics represents less favourable correlations.

[3] Response relative to Base1.

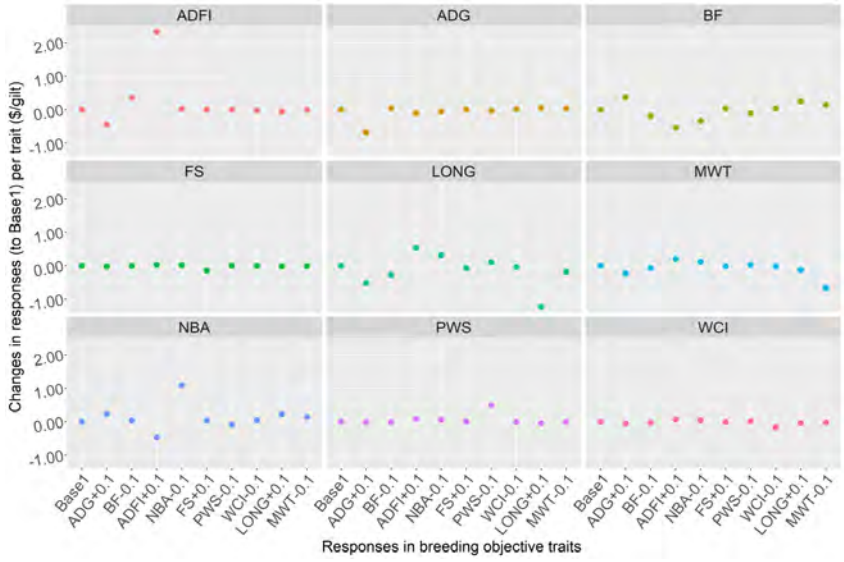


Figure 1. Changes in response for breeding objective traits ($/gilt), relative to Base1, when genetic correlations between innate sow appetite (IAP) and each trait change by ±0.10 (x-axis).

Discussion

A maternal breeding objective attempts to balance traits on growing animals with reproductive traits. With the assumed goal, data recording and Base1 parameters, results supported the hypothesis generally that traits representing IAP added as a selection criterion contribute to an increased overall response, due to genetic correlations between IAP and other traits. This supports the concept that sow feed intake can theoretically be recorded at any point in a sows' life-cycle, providing an opportunity to utilise already

recorded data. Therefore, the generic term 'innate' was used, reflecting any stage of a sows' life-cycle when the data can become available. In practice, the information on intake is not readily available in all stages of a sows' life, due to the limitations imposed by the high cost of recording. In this study lactation feed intake was used as an illustration, because of data availability (Bunter *et al.* 2010; Bergsma *et al.* 2013). The alternative traits reflecting 'innate appetite' are feed intake or feeding behaviour recorded during gestation (Vargovic *et al.* 2021b), or appetite before farrowing (Vargovic *et al.* 2021a). Changing the correlations between IAP and the other traits, whilst keeping the heritability constant, was done with the aim to mimic the alternative sow feed intake traits. None of the BO traits could have correlations changed by 0.10 in both directions as the variance-covariance matrix was no longer positive definite, demonstrating that the direction/magnitude of change was unlikely. Possible direction of change was not consistently favourable or unfavourable for each trait. Although the added information resulted in enhanced response, varying the correlations from the assumed baseline parameters affected the overall results. This demonstrated the effect of uncertain genetic parameters on the overall response, in particular between growing pigs and sows, and therefore requires clarification.

References

Amer, P., Ludemann, C., and Hermesch, S. (2014) J. Anim. Sci. 92(12): 5345-5357. https://doi.org/10.2527/jas.2014-7943

Bergsma, R. (2011) Genetic aspects of feed intake in lactating sows. Wageningen University, Wageningen, the Netherlands.

Bergsma, R., Mathur, P.K., Kanis, E., Verstegen, M.W.A., Knol, E.F. *et al.* (2013). J. Anim. Sci 91(8): 3601-3611. https://doi.org/10.2527/jas.2012-6200

Bunter, K.L., Lewis, C.R.G., Hermesch, S., Smits, R., and Luxford, B.G. (2010) 'Proc. of the 9th WCGALP', Leipzig, Germany.

Bunter, K.L., Smits, R., Luxford, B.G., and Hermesch, S. (2008) 'Pig Genetics Workshop', Armidale, Australia

Bunter, K.L., Vargovic, L., Athorn, R.Z., Henman, D., and Luxford, B.G. (2018) 'Proc. of the 11th WCGALP', Auckland, New Zealand.

Hermesch, S., Luxford, B.G. and Graser, H.U. (1999) 'Proc. of the 13th AAABG', Mandurah, Australia

Tholen, E., Bunter, K.L., Hermesch, S. and Graser, H.U. (1996) Aust. J. Agric. Res. 47(8), 1261-1274. https://doi.org/10.1071/AR9961261

Van der Werf, J.H.J. (2021) Multiple Trait Selection Index 20Trait version. Available at: https://jvanderw.une.edu.au/software.htm

Vargovic, L., Bunter, K.L. and Hermesch, S. (2021a) 'Proc. of the 24th AAABG', Adelaide, Australia

Vargovic, L., Hermesch, S., Athorn, R.Z. and Bunter, K.L. (2021b) J. Anim. Sci 99(1), 1-12. https://doi.org/10.1093/jas/skaa395

783. Genomic regions associated with backfat thickness show pleiotropic effect on osteochondrosis in pig

M. van Son[1*], M.F.L. Derks[2,3], M.S. Lopes[2,4], C.A. Sevillano[2], B. Harlizius[2] and E. Grindflek[1]

[1]Norsvin, Storhamargata 44, 2317 Hamar, Norway; [2]Topigs Norsvin Research Center, P.O. Box 43, 6640 AA Beuningen, the Netherlands; [3]Wageningen University & Research, P.O. Box 338, 6700 AH Wageningen, the Netherlands; [4]Topigs Norsvin, Visconde do Rio Branco 1310, 80.420-210 Curitiba, Brazil; maren.van.son@norsvin.no

Abstract

The aim of this study was to perform genome-wide association analyses for backfat thickness and osteochondrosis in Landrace pigs and to fine map pleiotropic genomic regions. In order to characterise genomic regions, phenotypic information of 5,000 animals with osteochondrosis scored from CT images and 40,000 animals with backfat thickness scored from CT or ultrasound images were analysed. All animals were genotyped with a medium density SNP chip and a subset of them were genotyped with a high-density SNP chip as well, allowing for imputation. Two genomic loci were found in common for osteochondrosis and backfat thickness, one on chromosome 5 and one on chromosome 14. For both regions, an antagonistic relationship was found. Fine mapping using an impact score approach identified the *CCND2* gene as the most likely causal gene on chromosome 5, whereas a mutation in *CRTAC1* had the highest impact score in the chromosome 14 region.

Introduction

Osteochondrosis (OC) is a developmental disease usually occurring in joints in an early stage of life in pigs as well as other domestic animals. It is characterised by a disturbed endochondral ossification in the articular cartilage and the epiphyseal growth plates and is most likely caused by failure of blood supply to the growing cartilage (e.g. Ytrehus *et al.*, 2007). OC is the most important cause of leg weakness in pigs and is therefore an important trait due to both animal welfare concerns and economic reasons. In Norwegian Landrace pigs, the heritability of OC was shown to vary from 0.06-0.21 depending on anatomical location, and the total OC score had a heritability of 0.31 (Aasmundstad *et al.*, 2013). Genomic studies have been conducted to elucidate the mechanisms causing OC. Quantitative trait loci (QTL) regions associated with OC have been identified on *Sus scrofa* chromosome (SSC) 2, 3, 5, 13 and 15 (Andersson-Eklund *et al.*, 2000; Laenoi *et al.*, 2011).

Backfat (BF) is an important carcass composition trait in pig production. Lean meat is the most valuable product so breeding programs have aimed for more growth of lean meat and reduced levels of BF (Lonergan *et al.*, 2001). Heritability of BF in Landrace pigs is 0.42 (unpublished data, H. C. Oliveira). Genomic studies have showed numerous candidate genes influencing backfat thickness and QTLs identified over the years can be found in the animal QTL database (https://www.animalgenome.org/QTLdb).

The aim of this study was to detect genomics regions contributing to BF in Landrace pigs and to fine map regions showing pleiotropy with OC.

Materials & methods

Animals. The animals included in this study were purebred Landrace boars and sows born between 2013 and 2020 in Norway. Animals used in the OC analysis were boars from the nucleus boar testing station in Hamar, Norway, that entered the testing station at approximately 30 kg live weight and lived in groups of 12 pigs per pen. At the end of the test period (at approximately 120 kg live weight), boars were subjected

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_783, © M. van Son *et al.* 2022

to a CT scan for measurement of OC as well as body composition traits. Prior to scanning, the boars were sedated using Azaperone (Stresnil Vet ®, Janssen-Cilag Ltd., Buckinghamshire, UK), which was injected intramuscularly. OC was scored from CT images as described in detail by Aasmundstad *et al.* (2013). For eight anatomical locations (distal and lateral humerus and femur, right and left side), a score for OC between 0 and 5 was assigned and the sum of phenotypes from all the locations, the total score, was used as the OC phenotype. A total of 5,000 animals were included for the OC association analysis.

Animals used in the BF analysis were boars from the boar testing station as well as boars and sows from Norwegian nucleus farms. BF measurements were done using CT data on boars from the boar testing station and ultrasound on the other animals. Backfat was measured in millimeter and 40,000 animals were included in the BF association analysis.

Genotyping. Ear samples for DNA extraction were collected and stored at -20 °C until used (Caisley ear sampling protocol). DNA for genotyping was extracted from the ear samples using BioSprint DNA Kit (Qiagen, Hilden, Germany) by BioBank (Hamar, Norway). DNA concentration and quality was measured using a NanoDrop ND-1000 spectrophotometer (NanoDrop Technologies, DE, USA). The genotyping was performed at CIGENE, University of Life Sciences, Norway and at NEOGEN, Scotland, UK. All the animals in this study were genotyped using Illumina porcine medium-density SNP chips (50K, 80K) (Illumina, San Diego, USA) or a custom made 25K SNP chip (Illumina and Neogen). Some of the boars were also genotyped using the Axiom porcine 660K array from Affymetrix (Affymetrix Inc., Santa Clara, CA, USA). All genotypes were filtered based on call rate>0.97 and minor allele frequency (MAF)>0.01. Imputation of all animals to 660K was performed using FImpute v3 (Sargolzaei *et al.*, 2014) with default settings and the complete pedigree as additional information. SNPs that were shared between the SNP chips were checked for matching genotypes and allele frequencies before imputation.

Genome wide association study and fine mapping. Genome-wide association study (GWAS) was performed using the GCTA software (Yang *et al.*, 2011). The total OC score and the backfat measurements were included as phenotypes using a linear animal model.

Fine mapping was conducted using the porcine combined annotation dependent depletion (pCADD) method (Derks *et al.*, 2021). The pCADD pipeline takes the top SNP as input and identifies SNPs from whole genome sequence (WGS) data that are in high linkage disequilibrium (LD; >0.7) with the top SNP. SNPs from WGS data are then ranked based on the pCADD scores, chromatin data information (liver) and gene expression data. The output from the pipeline is a list of likely causal variants. WGS and gene expression data from the Landrace population were available from previous studies (e.g. Van Son *et al.*, 2017a,b).

Results

GWAS were conducted using 660K imputed genotypes to identify genomic regions associated with OC and BF in Landrace boars. The prevalence of OC for the animals included in this study ranged between a total score of 0-19 with mean=2.81 (SD=2.54). BF ranged from 2 to 18 mm with mean=7.44 (SD=1.82). The GWAS detected significant loci for both OC (Figure 1) and BF (Figure 2) (corrected *P*-value<1.0^{-10}) on SSC5 at 66 Mb and on SSC14 at 109 Mb. For BF, additional significant loci were detected on SSC2 44 Mb (highly significant), SSC4 1 Mb, SSC5 15 Mb and SSC18 11 Mb.

The top SNP of the SSC5 QTL is AX-116685771 for both traits, showing opposing effects where the A allele increases BF and decreases OC. The SNP explains 0.5% of the genetic variance for OC and 0.1% for BF. Fine mapping of this region returned 47 variants in LD with the top SNP. The variant with the highest impact

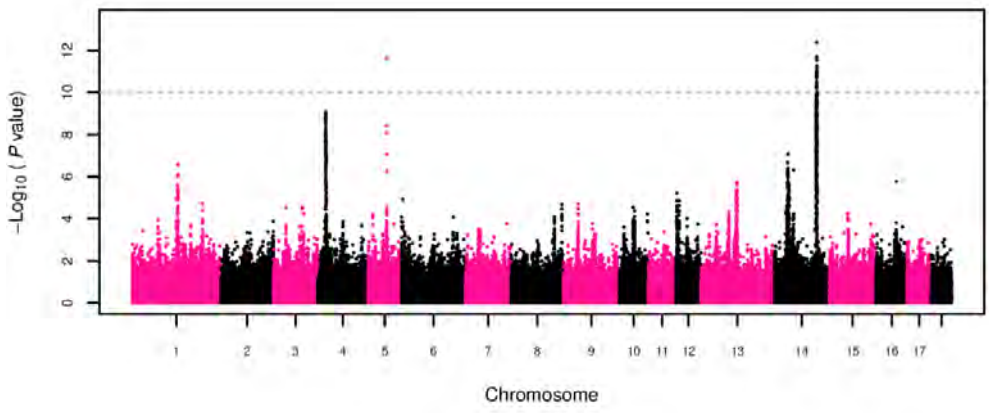


Figure 1. GWAS results of OC in Landrace pigs.

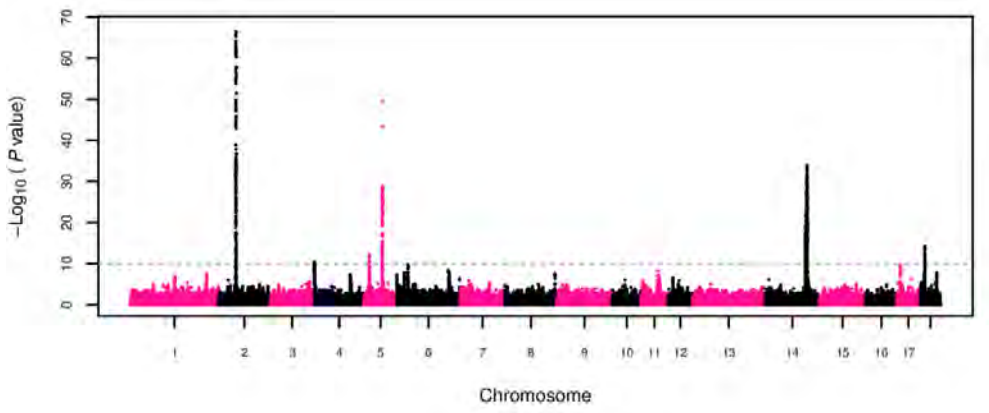


Figure 2. GWAS results of BF in Landrace pigs.

score was an intronic variant in the cyclin D2 gene (*CCND2*) with a pCADD score of 13.5. All 47 variants from this region were located in *CCND2* or downstream of this gene.

For the SSC14 QTL, AX-116521996 is the top SNP for OC whereas AX-116659727 is the top SNP for BF (LD between these two SNPs is 0.9). AX-116659727 is also highly significant for OC and the SNP shows opposing effects for the two traits in the same manner as on SSC5. The top SNPs explains 0.1% of the genetic variance for both OC and BF. Using AX-116659727 as top SNP, fine mapping of this region returned 319 variants in LD. The variant with the highest impact score was an intronic variant in cartilage acidic protein 1 (*CRTAC1*; pCADD=17.8).

Discussion

Breeding for increased growth rate and more lean meat has reduced the level of BF. However, pigs that grow faster are also more susceptible to OC and by including phenotypes on OC in the breeding goal, it is possible to balance the selection. Genomic information can be useful to identify genes showing pleiotropic effects on different traits and this study identified two QTL regions in common for OC and BF; on SSC5 and 14.

A previous study also identified the region on SSC5 for BF in four different breeds (Gozalo-Marcilla *et al.*, 2021). They suggested a causal role for the fibroblast growth factor 23 (*FGF23*) gene as its expression in osteocytes is linked to adipose tissue secretion of proteins. In this study, fine mapping identified *CCND2* as the most likely causal gene on SSC5. It is involved in the control of the cell cycle (Chiles, 2004) and knock-out studies in mice show numerous consequences in development of organs and functions. The two genes *FGF23* and *CCND2* are located just 50 kb apart and both genes are good biological candidates for both traits. It is also possible that the two genes are somehow interacting or that one gene is affecting BF and the other OC.

On SSC14, a mutation in *CRTAC1* showed the highest impact score. This gene is an extracellular matrix protein of articular cartilage (Steck *et al.*, 2007) and a promising candidate gene for OC, however, a function related to BF is not clear. It is also possible that two closely linked genes are causing the common QTL region on SSC14. The fine mapping of this region also listed several other potential candidate genes with high impact score, but further studies are needed to pinpoint the causal gene.

References

Aasmundstad T., Kongsro J., Wetten M., Dolvik N. I., and Vangen O. (2013) Animal 7(10):1576-1582. https://doi.org/10.1017/S1751731113001158

Andersson-Eklund L., Uhlhorn H., Lundeheim N., Dalin G., and Andersson, L. (2000) Genetical Research 75(2):223-230. https://doi.org/10.1017/s0016672399004371

Chiles T.C. (2004) The Journal of Immunology 173(5):2901-2907. https://doi.org/10.4049/jimmunol.173.5.2901

Derks M. F. L., Groβ C., Lopes M. S., Reinders M. J. T., Bosse M. *et al.* (2021) Genomics 113(4):2229-2239. https://doi.org/10.1016/j.ygeno.2021.05.017

Gozalo-Marcilla M., Buntjer J., Johnsson M., Batista L., Diez F. *et al.* (2021) Genetics Selection Evolution 53:76. https://doi.org/10.1186/s12711-021-00671-w

Laenoi W., Uddin M. J., Cinar M. U., Groβe-Brinkhaus C., Tesfaye D. *et al.* (2011) Genetics Selection Evolution 43:13 https://doi.org/10.1186/1297-9686-43-13

Lonergan S. M., Huff-Lonergan E., Rowe L. J., Kuhlers D. L., and Jungst S. B. (2001) Journal of Animal Science 79(8):2075-2085. https://doi.org/10.2527/2001.7982075x

Sargolzaei M., Chesnais J. P., and Schenkel F. S. (2014) BMC Genomics 15:478 https://doi.org/10.1186/1471-2164-15-478

Steck E., Bräun J., Pelttari K., Kadel S., Kalbacher H. *et al.* (2007) Matrix Biology 26(1):30-41. https://doi.org/10.1016/j.matbio.2006.09.006

Van Son M., Kent M. P., Grove H., Agarwal R., Hamland H. *et al.* (2017a) BMC Genetics 18:85(1) https://doi.org/10.1186/s12863-017-0549-8

Van Son M., Tremoen N. H., Gaustad A. H., Myromslien F. D., Våge D. I. *et al.* (2017b) BMC Veterinary Research 13(1):362 https://doi.org/10.1186/s12917-017-1279-x

Yang J., Lee S. H., Goddard M. E., and Visscher P. M. (2011) American Journal of Human Genetics 88(1):76-82. https://doi.org/10.1016/j.ajhg.2010.11.011

Ytrehus B., Carlson C. S., and Ekman S. (2007) Veterinary Pathology 44(4):429-448. https://doi.org/10.1354/vp.44-4-429

784. Non-genetic correlations between production traits and litter traits in pigs are important to account for

T. Ostersen[1*], M. Henryon[1], X. Guo[1,2], O.F. Christensen[2], G. Su[2] and A.C. Sørensen[1]

[1]*Danish Pig Research Centre, Danish Agriculture & Food Council, 1609 Copenhagen V, Denmark;*
[2]*Center for Quantitative Genetics and Genomics, Aarhus University, 8830 Tjele, Denmark; tao@lf.dk*

Abstract

We hypothesized that: (1) non-genetic correlations between litter size (LS) and daily gain are different from zero; and (2) modelling non-genetic correlations changes estimates of genetic (co)variances of the traits compared to when they are ignored. We tested these hypotheses using a new multi-trait model that separates genetic and non-genetic correlations by fitting correlations between the residual effects for one trait and non-residual effects of the other trait. We estimated variances by fitting models with and without these non-genetic correlations to LS and daily gain in DanBred Landrace pigs. We found that all non-genetic correlations between LS and daily gain were significantly different from zero. Estimates of genetic (co)variances changed significantly when non-genetic correlations were ignored. Clearly, multi-trait models with production traits and litter size traits should include correlations between residual effects for one trait and non-residual effects of the other trait to separate genetic correlations from non-genetic correlations.

Introduction

Multi-trait models are state-of-the-art for breeding value estimation (Mrode, 2014, p. 70). They provide more unbiased and accurate predictions than single-trait models when non-genetic correlations among traits differ from genetic correlations (Mrode, 2014, p. 70). Modelling non-genetic correlations is clearly important for prediction. However, some of these correlations are often neglected because the model framework has not been developed. An example is when a single phenotype is available for a trait(s) and multiple phenotypes are available for another trait(s). In pig breeding, daily gain and litter size are examples of such traits. Daily gain is often recorded as a single phenotype per animal. Litter size is recorded for each farrowing and each sow can have multiple phenotypes for litter size. A further complication is that two possible non-genetic correlations may exist. First, an animal's record of daily gain can be affected by the litter size in its birth litter. Second, a female's daily gain can affect its non-genetic permanent ability to produce large litters. Standard multi-trait models cannot cater for correlations between the residual effect for one trait and a non-genetic non-residual effect for another trait. The challenge is that neglecting them risks that they are captured by genetic correlations during REML estimation (Walter and Mao, 1985). Clearly, a new model framework that separates non-genetic correlations from genetic correlations is needed. We propose such a model and in doing so, hypothesize that: (1) the non-genetic correlations between litter size and daily gain are different from zero; and (2) modelling non-genetic correlations changes the estimates of genetic (co-)variances of the traits compared to when they are ignored.

Materials & methods

Experimental design. We estimated (co)variances using two trivariate models fitted to litter size at day 5 (LS), average daily gain until 30 kg (G1) and average daily gain from 30 to 100 kg (G2) in DanBred Landrace pigs. The two models were with and without non-genetic correlations between LS and the production traits.

Data. There were 93,408 records of LS from 45,973 purebred Landrace sows and 64% of the litters were crossbred with DanBred Yorkshire sire and the others were purebred litters. There were 203,411 records of G1 (g/day) and 212,191 records of G2 (g/day) all from purebred DanBred Landrace animals (Table 1).

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_784, © T. Ostersen *et al.* 2022

Table 1. Descriptive statistics.

Trait[1]	No. Obs	Mean	SD	Min	Max
LS	93,408	14.2	4.1	0	32
G1	203,411	382	37	295	579
G2	212,191	1,080	109	680	1,609

[1] LS: litter size at day 5; G1: average daily gain (g/day) from birth to 30 kg; G2: average daily gain (g/day) from 30 kg to 100 kg

The pedigree included 237,500 purebred Landrace animals. The data we used was from litters born between January 2016 and April 2021.

Model. The tri-variate model to estimate variance and covariance components was:

$$\begin{bmatrix} \boldsymbol{y_{LS}} \\ \boldsymbol{y_{G1}} \\ \boldsymbol{y_{G2}} \end{bmatrix} = \begin{bmatrix} \boldsymbol{X_{bLS}} & 0 & 0 \\ 0 & \boldsymbol{X_{bG1}} & 0 \\ 0 & 0 & \boldsymbol{X_{bG2}} \end{bmatrix} \begin{bmatrix} \boldsymbol{b_{LS}} \\ \boldsymbol{b_{G1}} \\ \boldsymbol{b_{G2}} \end{bmatrix} + \begin{bmatrix} \boldsymbol{Z_{aLS}} & 0 & 0 \\ 0 & \boldsymbol{Z_{aG1}} & 0 \\ 0 & 0 & \boldsymbol{Z_{aG2}} \end{bmatrix} \begin{bmatrix} \boldsymbol{a_{LS}} \\ \boldsymbol{a_{G1}} \\ \boldsymbol{a_{G2}} \end{bmatrix} +$$
$$\begin{bmatrix} \boldsymbol{Z_{cLSp}} & \boldsymbol{Z_{cLSc}} & 0 \\ 0 & 0 & 0 \\ 0 & 0 & \boldsymbol{Z_{cG2}} \end{bmatrix} \begin{bmatrix} \boldsymbol{c_{LSp}} \\ \boldsymbol{c_{LSc}} \\ \boldsymbol{c_{G2}} \end{bmatrix} + \begin{bmatrix} \boldsymbol{Z_{pLS}} & 0 & 0 \\ 0 & \boldsymbol{Z_{eG1}} & 0 \\ 0 & 0 & \boldsymbol{Z_{eG2}} \end{bmatrix} \begin{bmatrix} \boldsymbol{p_{LS}} \\ \boldsymbol{e_{G1}} \\ \boldsymbol{e_{G2}} \end{bmatrix} + \begin{bmatrix} \boldsymbol{Z_{eLS}} & 0 & 0 \\ 0 & \boldsymbol{Z_{lG1}} & 0 \\ 0 & 0 & \boldsymbol{Z_{lG2}} \end{bmatrix} \begin{bmatrix} \boldsymbol{e_{LS}} \\ \boldsymbol{l_{G1}} \\ \boldsymbol{l_{G2}} \end{bmatrix}$$

where y_{LS}, y_{G1} and y_{G2} are vectors of records for LS, G1 and G2; b_{LS} is a vector of fixed effects for LS including herd-year, year-month-litterbreed, year-parity, and first and second order regressions on the sow's age at first farrowing as well as time interval between farrowings; b_{G1} is a vector of fixed effects for G1 including herd-year-month, sex and birth-parity as well as a regression on the age of the sow at farrowing; b_{G2} is a vector of fixed effects for G2 including herd-year-stable, sex, and birth-parity. The vectors a_{LS}, a_{G1} and a_{G2} contain the genetic effects of the sow on LS, the genetic effect of the pig on G1 and the genetic effect of the pig on G2. The vectors c_{LSp} and c_{LSc} contain non-genetic random service sire effects on LS for purebred and crossbred litters. The vector c_{G2} contains random effects of pen on G2. The vector p_{LS} contains permanent non-genetic effects of the sow on LS, whereas the vectors l_{G1} and l_{G2} contained non-genetic random effects of birth-litter. Vectors e_{G1} and e_{G2} contain the residuals for G1 and G2 following the individual animal, whereas e_{LS} contains the residual for LS following the individual litter. All X and Z matrices are design matrices. The random effects are distributed as:

$$\begin{bmatrix} \boldsymbol{a_{LS}} \\ \boldsymbol{a_{G1}} \\ \boldsymbol{a_{G2}} \end{bmatrix} \sim MVN(0, \boldsymbol{G_0} \otimes \boldsymbol{A}); \ \boldsymbol{c_{LSp}} \sim N(0, \boldsymbol{I}\sigma^2_{cLSp}); \ \boldsymbol{c_{LSc}} \sim N(0, \boldsymbol{I}\sigma^2_{cLSc});$$

$$c_{G2} \sim N(0, \boldsymbol{I}\sigma^2_{cG2}); \begin{bmatrix} \boldsymbol{p_{LS}} \\ \boldsymbol{e_{G1}} \\ \boldsymbol{e_{G2}} \end{bmatrix} \sim MVN(0, \boldsymbol{RP_0} \otimes \boldsymbol{I}); \begin{bmatrix} \boldsymbol{e_{LS}} \\ \boldsymbol{l_{G1}} \\ \boldsymbol{l_{G2}} \end{bmatrix} \sim MVN(0, \boldsymbol{RL_0} \otimes \boldsymbol{I}),$$

where G_0 is a 3×3 additive-genetic covariance matrix; RP_0 is a 3×3 non-genetic covariance matrix related to animals; RL_0 is a 3×3 non-genetic covariance matrix related to litters; I is an identity matrix and A is the numerator relationship matrix.

The second model ignoring non-genetic correlations between production traits and litter size is as above, but with the off-diagonal elements of the first row and column of RP_0 and RL_0 equal to zero. (Co-)variance components in the above models were estimated with REML using DMU (Madsen and Jensen, 2013)

Results

Non-genetic correlations. All non-genetic correlations between LS and the two growth traits were significantly different from zero (Table 2). The correlations between residuals for the two growth traits and the permanent non-genetic effect of the sow on LS were favourable 0.21 and 0.12 for G1 and G2, whereas the correlations between the residual for LS and the non-genetic birth litter effects were unfavourable -0.29 and -0.11 for G1 and G2.

Genetic variances. The genetic variances changed significantly when the non-genetic correlations were ignored (Table 3). The genetic variance of LS increased from 1.03 to 1.30 and the genetic correlations between LS and the two growth traits weakened. The correlation between LS and G1 changed from -0.11 to -0.04, and the correlation between LS and G2 changed from -0.24 to -0.16.

Discussion

Our findings supported our hypothesis that non-genetic correlations between litter size and daily gain are significant. We found significant and favourable correlations between the residuals for the two growth traits and the non-genetic permanent effect of the sow on LS. We found significant and unfavourable correlations between the residual for LS and the non-genetic birth litter effects on the two growth traits. Our findings also supported our second hypothesis that modelling these correlations changes the estimates of genetic variances compared to when they are ignored. We found significant changes in both the genetic variance of LS and the genetic correlations between LS and the two growth traits. Clearly, multi-trait models with production traits and litter size traits should include correlations between residual effects for one trait and non-residual effects of the other trait to separate genetic correlations from non-genetic correlations.

Table 2. Non-genetic variances for residual, permanent non-genetic sow effect and birth-litter effect (correlations above diagonal).

	RP_0[1]			RL_0[2]		
	LS	**G1**	**G2**	**LS**	**G1**	**G2**
LS	1.15	0.21	0.12	11.5	-0.29	-0.11
G1	5.3	557	0.06	-12	148	0.28
G2	8.4	98	4,425	-7.6	71	421

[1] RP_0 contained variances for the animal-level non-genetic variances – i.e. permanent non-genetic sow effect for litter size and residual for the two growth traits. Standard errors of correlations were within the range 0.01 to 0.02, standard errors of diagonal elements were 0.07, 5 and 37.

[2] RL_0 contained variances for the litter-level non-genetic variances – i.e. residual for litter size and birth-litter variance for the two growth traits. Standard errors of correlations were within the range 0.01 to 0.02, standard errors of diagonal elements were 0.07, 3 and 14.

Table 3. Genetic variances for models with and without non-genetic correlations between litter size and production traits (correlations above diagonal).

	With non-genetic correlations[1]			**Without non-genetic correlations**[1]		
	LS	**G1**	**G2**	**LS**	**G1**	**G2**
LS	1.03	-0.11	-0.24	1.30	-0.04	-0.16
G1	-1.8	234	0.37	-0.74	230	0.37
G2	-13	299	2,787	-9.5	292	2,774

[1] Standard errors of correlations were within the range 0.02 to 0.04, standard errors of diagonal elements were 0.1, 10 and 80 for LS, G1 and G2.

The change in genetic variances and correlations was likely caused by both the non-genetic correlations on animal-level as well as on litter-level. Models, where only litter-level non-genetic correlations were ignored, caused inflated genetic variance for LS as well as very strong negative genetic correlations between LS and the two growth traits (results not shown). Oppositely, models, where only animal-level non-genetic correlations were ignored, did not cause inflated genetic variance for LS, but instead genetic correlations between LS and the two growth traits became closer to zero (results now shown). So, it seems that modelling both animal-level and litter-level non-genetic correlations are of importance for these traits.

The multi-trait model that we developed is novel because it handles non-genetic correlations between traits, where the levels of observation are different. We fitted our model to litter size, measured at the litter level, and daily gain, measured at the individual level. Not only is the model relevant for these traits, it may also be relevant for other traits on the same levels, for other mammalian species, where litter size is larger than one (Mínguez *et al*., 2015), and for traits on different levels. Examples of traits on different levels include non-genetic correlations between group recorded traits and individually recorded traits (Ma *et al*., 2021) and between longitudinally recorded traits and individually recorded traits (Shirali *et al*., 2017). So, our model is not only relevant to litter size and production traits in pigs, but also litter traits and individual traits in other species. It is also relevant for other scenarios, where the level of observation differs among traits analysed.

References

Madsen, P. and Jensen, J. (2013) A user's guide to DMU, version 6, release 5.2. Center for Quantitative Genetics and Genomics Dept. of Molecular Biology and Genetics, University of Aarhus.

Ma, X., Christensen, O.F., Gao, H., Huang, R., Nielsen, B. et al (2021). Heredity, 126(1):206-217. https://doi.org/10.1038/s41437-020-0339-3

Mínguez, C., Sanchez, J.P., El Nagar, A.G, Ragab, M. and Baselga, M. (2015) J. Anim. Breed. Genet. 133 (2016) 303–315. https://doi.org/10.1111/jbg.12197

Mrode, R.A. (2014) Linear Models for the Prediction of Animal Breeding Values. CABI, Oxfordshire, UK

Shirali, M., Strathe, A. B., Mark, T., Nielsen, B. and Jensen, J. (2017). Journal of animal science, 95(3), 1050-1062. https://doi.org/10.2527/jas.2016.0606

Walter, J.P. and Mao, I.L. (1985) J Dairy Sci 68:91-98. https://doi.org/10.3168/jds.S0022-0302(85)80802-X

785. Breeding for component traits of litter size at day 5 increases piglet survival while maintaining litter size at day 5

M. Henryon[1*], T. Ostersen[1], X. Guo[1], G. Su[2] and A.C. Sørensen[1]

[1]Danish Pig Research Centre, Danish Agriculture and Food Council, 1609 Copenhagen, Denmark; [2]Center for Quantitative Genetics and Genomics, Aarhus University, 8830 Tjele, Denmark; mahe@lf.dk

Abstract

We tested the hypothesis that selection for component traits of live piglets at day 5 (LP5) realises more genetic gain (ΔG) for piglet survival than selection for LP5 while maintaining ΔG for LP5. We tested this premise by stochastic simulation of the breeding scheme for Danbred's Landrace pigs. We estimated ΔG for LP5 and piglet survival realised by selection for: (1) an index for LP5; and (2) an index for TNB and piglet survival at 1% rate of inbreeding. Selection for index TNB and survival realised about 30% more ΔG for piglet survival than selection for the LP5 index while realising the same ΔG for LP5. Clearly, it is worthwhile separating LP5 into its component traits: it realises more ΔG for piglet survival while maintaining ΔG for LP5.

Introduction

Number of live piglets in a litter on day 5 after farrowing (LP5) is a key trait of DanBred's pig breeding program in Denmark (Su *et al.*, 2020). LP5 is considered a trait that is expressed by the sow of each litter. It is a composite of total number of piglets born (TNB) and survival of individual piglets within litters until day 5 after farrowing (piglet survival). Breeding for increased LP5 has been remarkably successful in DanBred's two maternal breeds: Landrace and Yorkshire (Nielsen *et al.*, 2013; Su *et al.*, 2020). Breeding has realised genetic gain (ΔG) for LP5 by allocating selection pressure to TNB and piglet survival. The drawback is that breeding for LP5 does not allow breeders to control the amount of selection pressure allocated to TNB and piglet survival. Both TNB and piglet survival are economically important because they contribute to LP5. However, there is good reason to allocate more selection pressure to piglet survival: piglet survival is also a proxy for animal welfare and consumer acceptance. One way that we could breed for LP5 while allocating more selection pressure to piglet survival is to separate LP5 into its component traits and select for an index of these components. We could shift selection pressure from TNB to piglet survival by increasing the index weight for piglet survival. This led us to hypothesise that selection for component traits of LP5 realises more ΔG for piglet survival than selection for LP5 while maintaining ΔG for LP5. We tested this hypothesis by stochastic simulation of the breeding scheme for Danbred's Landrace pigs.

Materials & methods

Procedure. We used stochastic simulation of the breeding scheme for Danbred's Landrace pigs to estimate ΔG for LP5 and piglet survival realised by selection for an index for LP5 – referred to as I_{LP5} – and an index for TNB and piglet survival (I_{TNB+S}) at 1% rate of inbreeding. Selection index I_{TNB+S} was tested with different relative index weights for TNB and piglet survival. We also selected for TNB (I_{TNB}) and piglet survival (I_S) as extreme indices of I_{TNB+S}. The simulations were carried out by: (1) sampling unobserved underlying liabilities for TNB and piglet survival; (2) expressing and phenotyping LP5, TNB, and piglet survival on their observed scales based on the liabilities; (3) predicting EBV for LP5, TNB, and piglet survival by fitting animal models to the phenotypes on the observed scales; (4) using the EBV to construct selection indices; and (5) selecting breeding animals by optimum-contribution selection (OCS) with the index values as estimates of genetic merit. All animals were phenotyped for piglet survival before EBV prediction and selection. Surviving animals were candidates for selection. Dams that were selected in each generation were phenotyped for LP5 and TNB after selection. Breeding schemes were run for 16 discrete generations ($t = 1$

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_785, © M. Henryon *et al.* 2022

… 16) and replicated 300 times. Selection for the indices was carried out in generations t = 12 … 16. The first 11 generations were used to establish a pedigreed population with phenotypes.

Breeding scheme. Each replicate was initiated by sampling a unique base population of 100 sires and 1000 dams. One-hundred sires and 1000 dams were randomly selected and mated in generations t = 1 … 11. In generations t = 12 … 16, 1000 matings were allocated to selection candidates by OCS. The number of matings allocated to each male varied from 0, 1, 2 … to 50 matings. One thousand females were allocated a single mating. The 1000 sire and dam matings were paired randomly. Litter size of each mating (dam) was the dam's phenotype for TNB on the observed scale. Offspring were assigned as males or females with probability 0.5.

Sampling underlying liabilities. Unobserved underlying liabilities for TNB and piglet survival (ℓ_T and ℓ_S) were sampled as:

$$\begin{bmatrix}\boldsymbol{\ell}_T\\ \boldsymbol{\ell}_S\end{bmatrix} = \begin{bmatrix}\mathbf{1} & \mathbf{0}\\ \mathbf{0} & \mathbf{1}\end{bmatrix}\begin{bmatrix}\mu_T\\ \mu_S\end{bmatrix} + \begin{bmatrix}\mathbf{Z}_{a_d} & \mathbf{Z}_{a_s} & \mathbf{0} & \mathbf{0}\\ \mathbf{0} & \mathbf{0} & \mathbf{Z}_{a_{id}} & \mathbf{Z}_{a_m}\end{bmatrix}\begin{bmatrix}\mathbf{a}_d\\ \mathbf{a}_s\\ \mathbf{a}_{id}\\ \mathbf{a}_m\end{bmatrix} + \begin{bmatrix}\mathbf{Z}_s & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{bmatrix}\begin{bmatrix}\mathbf{s}\\ \mathbf{0}\end{bmatrix} + \begin{bmatrix}\mathbf{Z}_l & \mathbf{0}\\ \mathbf{0} & \mathbf{Z}_{ls}\end{bmatrix}\begin{bmatrix}\mathbf{l}\\ \mathbf{ls}\end{bmatrix} + \begin{bmatrix}\mathbf{0} & \mathbf{0}\\ \mathbf{0} & \mathbf{Z}_e\end{bmatrix}\begin{bmatrix}\mathbf{0}\\ \mathbf{e}\end{bmatrix} \quad (1)$$

where m_T=15.3 and m_S=0 are mean liabilities for TNB and piglet survival, a_d and a_s are vectors of true genetic sow and service-sire effects for TNB, a_{id} and a_m are vectors of true direct-genetic effects and maternal-genetic effects for piglet survival, s is a vector of non-genetic service-sire effects for TNB, l is a vector of litter effects for TNB that can also be interpreted as a vector of residuals, ls is a vector of litter effects for survival, e is the vector of residuals for piglet survival, 1 are vectors of 1's, 0 are vectors and matrices of 0's, and Z_{a_d}, Z_{a_s}, $Z_{a_{id}}$, Z_{a_m}, Z_s, Z_l, Z_{ls}, and Z_e are design matrices. Effects for animals in the base populations were sampled as:

$$\mathbf{a} = \begin{bmatrix}a_d\\ a_s\\ a_{id}\\ a_m\end{bmatrix} \sim N\left(\begin{bmatrix}0\\0\\0\\0\end{bmatrix}, \mathbf{G} = \begin{bmatrix}1.367 & 0.231 & -0.036 & -0.094\\ 0.231 & 0.508 & -0.078 & 0.066\\ -0.036 & -0.078 & 0.003 & -0.008\\ -0.094 & 0.066 & -0.008 & 0.003\end{bmatrix}\right), \; s \sim N(0, 0.63), \begin{bmatrix}l\\ ls\end{bmatrix} \sim N\left(\begin{bmatrix}0\\0\end{bmatrix}, \begin{bmatrix}14.98 & -0.58\\ -0.58 & 0.14\end{bmatrix}\right),$$

and $e \sim N(0, 0.78)$. True genetic effects of animals in subsequent generations were sampled as $\mathbf{a}_i \sim N(½ \cdot (\mathbf{a}_{\text{sire}i} + \mathbf{a}_{\text{dam}i}), ½ \cdot (1 - \bar{F}_i) \cdot \mathbf{G})$, where $a_{\text{sire}i}$ and $a_{\text{dam}i}$ are additive-genetic values of the sire and dam of animal i and $\bar{F}_i$ is the average inbreeding coefficient of the sire and dam.

Observed phenotypes. TNB and piglet survival were expressed and phenotyped on their observed scales. TNB was transformed from the unobserved underlying liability scale to the observed scale by rounding elements of ℓ_T to the closest integer. Piglet survival was assumed to follow a threshold-liability model. Animal I survived (y_{Si}=1) if its liability, $\ell_{Si} > k$, and died (y_{Si}=0) if $\ell_{Si} \leq k$, where k=-0.82 is a fixed threshold derived from the mean probability of piglet survival 0.79 for DanBred Landrace. LP5 were expressed for each litter as $\text{LP5}_j = \sum_{i=1}^{\text{TNB}_j} y_{Sij}$, where y_{Sij} refers to piglet I in litter j.

Prediction. Animal models were fitted to the observed phenotypes to estimate (co)variances and predict EBV. The models were: (1) LP5: $\mathbf{y}_{\text{LP5}} = \mathbf{1}\nu_{\text{LP5}} + \mathbf{Z}_{a_d}\mathbf{a}_d + \mathbf{Z}_s\mathbf{s} + \mathbf{e}$; (2) TNB and survival: $\begin{bmatrix}\mathbf{y}_T\\ \mathbf{y}_S\end{bmatrix} = \begin{bmatrix}\mathbf{1} & \mathbf{0}\\ \mathbf{0} & \mathbf{1}\end{bmatrix}\begin{bmatrix}\nu_T\\ \nu_S\end{bmatrix} + \begin{bmatrix}\mathbf{Z}_{a_d} & \mathbf{0} & \mathbf{0}\\ \mathbf{0} & \mathbf{Z}_{a_{id}} & \mathbf{Z}_{a_m}\end{bmatrix}\begin{bmatrix}\mathbf{a}_d\\ \mathbf{a}_{id}\\ \mathbf{a}_m\end{bmatrix} + \begin{bmatrix}\mathbf{Z}_s & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{bmatrix}\begin{bmatrix}\mathbf{s}\\ \mathbf{0}\end{bmatrix} + \begin{bmatrix}\mathbf{Z}_l & \mathbf{0}\\ \mathbf{0} & \mathbf{Z}_{ls}\end{bmatrix}\begin{bmatrix}\mathbf{l}\\ \mathbf{ls}\end{bmatrix} + \begin{bmatrix}\mathbf{0} & \mathbf{0}\\ \mathbf{0} & \mathbf{Z}_e\end{bmatrix}\begin{bmatrix}\mathbf{0}\\ \mathbf{e}\end{bmatrix}$; (3) TNB: $\mathbf{y}_T = \mathbf{1}\nu_T + \mathbf{Z}_{a_d}\mathbf{a}_d + \mathbf{Z}_s\mathbf{s} + \mathbf{e}$; and (4) Survival: $\mathbf{y}_S = \mathbf{1}\nu_S + \mathbf{Z}_{a_{id}}\mathbf{a}_{id} + \mathbf{Z}_{a_m}\mathbf{a}_m + \mathbf{Z}_{ls}\mathbf{ls} + \mathbf{e}$, where y_{LP5}, y_T, and y_S are vectors of observed phenotypes for LP5, TNB, and piglet survival, $\hat{\nu}_{\text{LP5}}$, $\hat{\nu}_T$, and $\hat{\nu}_S$ are means for LP5, TNB, and piglet survival, e are residuals, and all other effects are as defined for Equation 1.

Selection indices. The selection indices were: (1) LP5: $\mathbf{I}_{\text{LP5}} = \hat{\mathbf{a}}_d$; (2) TNB and survival: $\mathbf{I}_{\text{TNB+S}} = \hat{\mathbf{a}}_d + \omega \cdot (½ \cdot \hat{\mathbf{a}}_{id} + \hat{\mathbf{a}}_m)$, ω=10, 15, 16, 18, 20, 25, 30, 40; (3) TNB: $\mathbf{I}_{\text{TNB}} = \hat{\mathbf{a}}_d$; and (4) Survival: $\mathbf{I}_S = ½ \cdot \hat{\mathbf{a}}_{id} + \hat{\mathbf{a}}_m$. EBV used in indices (1)-(4) were predicted by corresponding animal models (1)-(4). We use the circumflex '^' to denote EBV predicted by the models. The effects and weights in the selection

indices reflected Landrace's genetic contribution to Duroc × (Landrace × Yorkshire) pigs that are used in production.

Optimum-contribution selection. OCS solved $U_t(\mathbf{c}) = max\{\mathbf{c}'\mathbf{I}_j | \mathbf{c}'\mathbf{A}\mathbf{c} \leq C_{t+1}\}$ with respect to c, where c is a vector of genetic contributions to the next generation, I_j is a vector of index values ($I_j = I_{LP5}$, I_{TNB+S}, I_{TNB}, I_S), A is a numerator-relationship matrix, and C_{t+1} is the pre-defined average relationship in generation t+1 (Meuwissen, 1997). C_{t+1} was a function of average relationship in generation t and the pre-defined 1% rate of inbreeding.

Genetic gain and rate of inbreeding. We plotted ΔG for LP5 against ΔG for piglet survival for the four selection indices based on Landrace's genetic contribution to production pigs. ΔG for LP5 and piglet survival were calculated as the linear regression of $LP5_t$ and S_t on t, where $LP5_t$ and S_t are the average true breeding values for LP5 and piglet survival of animals born at generation t = 12 … 16. The true breeding values for LP5 and piglet survival of animal I born at generation t were estimated as $LP5_i = (\mu_T + a_{di}) \times S_i$ and $S_i = 1 - \Phi(k = -0.82, \text{mean} = ½ \times a_{idi} + a_{mi}, \text{sd} = 1)$, where a_{di}, a_{idi}, and a_{mi} are the i^{th} elements of a_d, a_{id}, and a_m, and Φ(.) is the cumulative standard-normal distribution. Rate of inbreeding was calculated as 1-exp(β), where β is the linear regression of $\ln(1-F_t)$ on t and F_t is the average coefficient of pedigree inbreeding for animals born at times t = 12 … 16.

Results

Selection for index TNB and survival, I_{TNB+S}, with index weight ω=20 applied to piglet survival realised about 30% more ΔG for piglet survival than selection for index LP5, I_{LP5}, while realising the same ΔG for LP5 (Figure 1). Increasing the weight in I_{TNB+S} to ω=25 realised about 73% more ΔG for piglet survival for only a 3.7% decrease in ΔG for LP5. Selection for I_{TNB+S} did not realise more ΔG for LP5 than I_{LP5} at the same ΔG for piglet survival. Selection for the first of our extreme indices, I_{TNB}, realised negative ΔG for piglet survival and 36% less ΔG for LP5 than selection for index I_{LP5}. Selection for I_s realised 174% more ΔG for piglet survival than I_{LP5}, but 65% less ΔG for LP5.

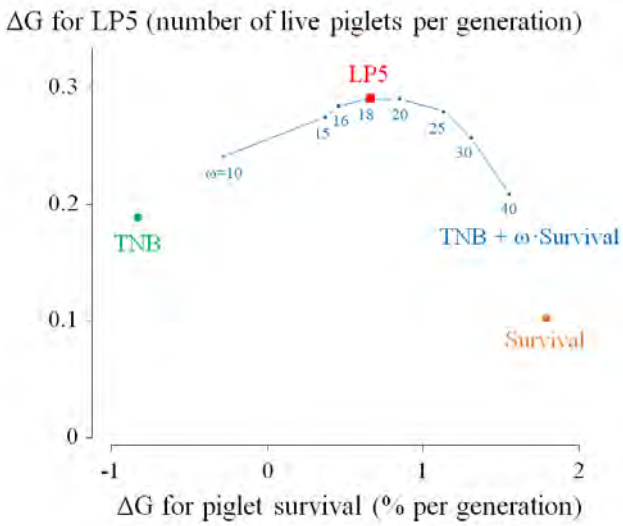


Figure 1. ΔG for LP5 against ΔG for piglet survival realised by selection for four indices at 1% rate of inbreeding. Rates are means of 300 simulation replicates. SD between replicates ranged from 0.059-0.069 (ΔG for LP5) and 0.27-0.46 (ΔG for piglet survival).

Discussion

Our findings supported our hypothesis that selection for component traits of LP5 realises more ΔG for piglet survival than selection for LP5 while maintaining ΔG for LP5. The additional 30% ΔG for piglet survival that we simulated in Landrace pigs while maintaining ΔG for LP5 makes it worthwhile separating LP5 into its component traits. We could realise even more ΔG for piglet survival if we are prepared to accept a small decrease in ΔG for LP5. Separating LP5 into its component traits and selecting for an index of these traits realised more ΔG for piglet survival because it shifted selection pressure from TNB to piglet survival. It must have shifted selection pressure, given that we realised more ΔG for piglet survival while maintaining ΔG for LP5. There will probably always be an incentive to increase ΔG for piglet survival given that piglet survival is economically important and a proxy for animal welfare and consumer acceptance. However, separating LP5 into its component traits does come with challenges. It can increase the number of traits that need to be recorded, increase the dimension of multi-trait prediction models, and require reliable estimates of heritabilities and genetic and residual correlations between the component traits and all other traits included in these prediction models. These challenges could persuade some pig breeders to continue selecting for LP5. So, provided it does not burden breeding schemes, it makes perfect sense to separate LP5 into its component traits because it realises more ΔG for piglet survival while maintaining ΔG for LP5.

We were surprised to find that separating LP5 into its component traits did not realise more ΔG for LP5 than selection for LP5 at the same ΔG for piglet survival. Clearly, the animal model that we fitted to TNB and piglet survival did not provide additional benefits beyond allowing us to shift selection pressure to piglet survival. The model was designed to cater for differences in heritabilities of the traits and to differentiate between genetic and residual correlations. TNB and piglet survival have similar heritabilities in Landrace pigs when assessed as traits of the sow of the litters; and the genetic correlation between the traits is similar to their residual correlation. However, the model could still provide additional benefits in other pig breeds and animal species where heritabilities of component traits are different and genetic correlations differ from residual correlations. This leads us to speculate that separating composite traits into component traits realises more ΔG for both component traits than selection for composite traits when the components differ for heritability and genetic and residual correlations.

References

Meuwissen T.H.E. (1997) J. Anim. Sci. 75:934-940. https://doi.org/10.2527/1997.754934x

Nielsen B., Su G., Lund M.S., and Madsen P. (2013) J. Anim. Sci. 91(6): 2575-2582 https://doi.org/10.2527/jas.2012-5990

Su G., Höglund J., Christensen O.F., and Lund, M.S. (2020) Possibilities for genetic improvement of pig survival until slaughter. DCA - Nationalt Center for Fødevarer og Jordbrug.

786. Phenotyping strategies for an efficient and holistic approach to reduced boar taint through genomic selection

A. Markey*, C. Burgeon and N. Gengler

Gembloux Agro-Bio Tech, University of Liège, Passage des Déportés 2, 5030 Gembloux, Belgium; alice.markey@uliege.be

Abstract

Surgical castration without anaesthesia or analgesia of boars is about to be banned in Europe. This ancestral practice avoids the development of boar taint (BT), an unpleasant smell found in cooked meat of some uncastrated male pigs, mainly due to the accumulation of skatole and androstenone in adipose tissue. Animal breeding and especially genomic selection in the context of whole-male rearing has been recognised as a potentially promising strategy to avoid BT. However, challenges exist to define and to record BT associated traits. Required phenotypes must be cheap, available in routine as close as possible to the slaughterline and repeatable. This review will also cover new opportunities through different and innovative aspects of BT phenotype definitions allowing the potential identification of underlying genetic architecture and its use in the breeding against BT.

Introduction

Boar taint (BT), found in uncastrated male pigs, is a complex hormonal odour caused in majority by androstenone and skatole (Rius *et al.*, 2005) that give respectively an urine and faecal odour to cooked meat (Font-i-furnols, 2012). Androstenone (i.e. *5α-androst-16-en-3-one*) is a testicular steroid that is synthesized in Leydig cells from pregnenolone (Squires *et al.*, 2020) and in high concentrations since puberty (Bonneau, 1998). Skatole is an alkaloid derived from tryptophan (Claus *et al.*, 1994; Jensen *et al.*, 1995a). They are produced in the colon by the breakdown of indigestible fermentable food and dead intestinal cells by microbiota (Jensen *et al.*, 1995b; Wesoly *et al.*, 2012). Androstenone and skatole can be found in various tissues and fluids such as in fat and blood (Wesoly *et al.*, 2012; Squires *et al.*, 2020). Other compounds could contribute to BT (e.g. 2-aminoacetophenone) (Rius Solé *et al.*, 2001; Fischer *et al.*, 2014) or could be correlated (e.g. presence of high levels of saturated fatty acids (Mörlein *et al.*, 2014)). Moreover, the odour intensity is affected by several factors such sexual maturity which particularly affects androstenone production, husbandry practices (e.g. nutrition particularly involves in skatole synthesis), the health and the raising environment, but also breed and genetics (Li *et al.*, 2019; Squires *et al.*, 2020), and lastly the interaction of all these. From one individual to another, the tainted characterisation is therefore highly variable. Despite extensive knowledge that was established about BT in the last years and related phenotypes based on biological and chemical facts, work is ongoing to advance on required phenotyping strategies for an efficient and holistic approach to do genomic selection for reduced BT while considering pig breeding as a whole.

Phenotypes

Direct classification of BT phenotypes. Phenotypes are characterised when boars are slaughtered as carcasses have to be checked for BT to determine what will be there use later on in the meat processing (e.g. consumption of pork chops for untainted boars or dilution of the meat in the case of tainted meat). Throughout Europe, the 'human nose' method is generally used by trained operators online or at line except in Denmark where a colorimetric method is used to detect and quantify indolic compounds (European Commission, 2019). From an animal breeding perspective this type of field data based on the current detection methods has the advantage of being fast, easily performed in routine and cheap, therefore on a large scale, but has the weakness of unreliability left by the subjective identification of an operator even if

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_786, © A. Markey *et al.* 2022

he is trained for 'human nose method' and the lack of specificity for the colorimetric method (Burgeon *et al.*, 2021a).

Detection of BT generating compounds in fat. An alternative to sensory analysis is accurate chemical analysis of odorous compounds. This implies normally that animal are dead but a biopsy tool has been developed by Baes *et al.* (2013) to test live boars and determine from their phenotype of BT compounds. Many chromatographic techniques have been developed that allow a high level of robustness, reliability, accuracy and sensitivity in the detection of BT compounds often found in trace amounts. Their disadvantages are in particular their rather high price, the difficulty of setting up such an instrument in terms of space and technical training in the slaughterhouse, and the lack of speed (including preparation). Portable instrumental methods such as biosensors, surface-enhanced Raman scattering, and Rapid Evaporative Ionisation Mass Spectrometry technologies are being studied but still required sample preparation. These research include the detection well-known BT compounds (e.g. androstenone and skatole) but can also be based on volatile organic compounds profile as Burgeon *et al.* (2021b) recently explored (Font-i-furnols *et al.*, 2020; Burgeon *et al.*, 2021a). Finally the Laser Diode Thermal Desorption–Tandem Mass Spectrometry is currently being tested in slaughterhouses in Denmark, probably to replace the colorimetric method, meeting the requirement of cost and reliability (Font-i-furnols *et al.*, 2020; Burgeon *et al.*, 2021a). Most of current detection methods are based on the detection of these compounds in fat matrix. However, other matrices could be studied such as blood or plasma and saliva since the presence of skatole and androstenone these matrices have been previously observed (Bone *et al.*, 2019; Dubois *et al.*, 2021).

Indirect traits used to detect risk of BT. Other traits related to sexual maturity were also studied although they were not always considered satisfactory for classification. Weight is used in the UK as an indicator of puberty implying slaughter below 110 kg (European Commission, 2019) but this method is at risk for pigs with early age maturity. Bernau *et al.* (2018) also correlated greater testicular volume and body fat with abundant androstenone content. If body fat is routinely assessed on slaughterlines measuring testis volume may present a greater challenge. Fatty acids have also been identified as potential indicators of BT. It would appear that tainted fats are significantly richer in saturated fatty acids and poorer in unsaturated fatty acids than an untainted subcutaneous adipose tissue (Mörlein *et al.*, 2014; Liu *et al.*, 2017). Measuring fatty acids using appropriate techniques may be reliable in routine using appropriate tools (e.g. near-infrared reflectance spectrometry) (Font-i-Furnols *et al.*, 2020).

Classical breeding against boar taint

Classical genetic tools. Breeding programmes need to use different precited phenotypes to identify boars with a low risk of transmitting odour development to their offspring (Drag *et al.*, 2017). In Europe, there are various BT selection programmes (e.g. Topigs Norsvin, Nucleus) (Squires *et al.*, 2020) often nowadays exploiting genomics (Drag *et al.*, 2017).

Holistic approach towards pig breeding. If the previously mentioned indirect traits were relatively, but not totally, i.e. body fat, in neutral relationships to breeding objectives, BT is suspected to have negative genetic correlations with meat quality, including intramuscular fat and fat thickness, and fertility and that selecting for reduced BT incidence would deteriorate the selection response in these traits. Genetic correlations of reproductive and aggressiveness traits, aspects concerning farmers for breeding, with BT compounds have been reported as related by Duarte *et al.* (2021). However, these correlations depend on the breed and the gender and their capacity to develop associated traits to BT compounds development.

Future opportunities and challenges

Innovative phenotypes. BT is complex. Its compounds, some of which are well known, are widely studied. Currently, genetic and genomic studies for selection against BT generally focus on the main compounds responsible for BT (skatole, androstenone and indole at a smaller extent) as well as on sexual maturity related traits. To characterise BT and optimize selection to make it fast and cheap for use in routine, research should open up to other ways of working or go down again paths that are sometimes abandoned due to lack of funds or technological resources. For example, the study of volatile organic compounds related to BT (e.g. Burgeon *et al.*, 2021b) is a path that is again being opened in particular through the development of more refined and potentially faster chemical techniques. Technological phenotyping has to explore the BT characterization in fat and other matrices which contain these compounds. BT compounds could be related in particular to the physiology of boars and sows and indicator traits (traditional and novel ones). Finally, this innovative phenotyping is a gateway to genetics, genomics and gene expression studies for known and currently undetermined BT molecules. Where fundamental science brings technologies to phenotyping, these could be linked to establish the most accurate genetic selection possible to respond to producers and consumers requirements.

Gene pathways. Skatole and androstenone have been extensively studied, including at genomic level to determine which genes are involved in the physiological pathways (e.g. Squires *et al.*, 2020). As genome wide association studies (GWAS) on skatole, androstenone, indole and other usual BT traits have already identified candidate genes, research could also find new gene pathways from exploiting these novel phenotypes. These new identified biomarkers could be then used.

Gene expressions. The study of these genes and their expression in different tissues could bring new opportunities to define expression quantitative trait loci (eQTL) of interest. Drag *et al.* (2018) have already characterised eQTLs function according to regulation systems of traditional BT compounds. However, since BT reduction is desired, it is necessary to monitor possible overlaps for other traits as highlighted by Drag *et al.* (2018).

Acknowledgements

We gratefully acknowledge the Service Publique de Wallonie (SPW) Agriculture of Région Wallonne (RW) for supporting and funding this work by the project NoWallOdor (D31-1396).

References

Baes C., Mattei S., Luther H., Ampuero S., Sidler X. *et al.* (2013) Animal 7(5):714–720, https://doi.org/10.1017/S1751731112002273

Bernau M., Schwanitz S., Kremer-Rücker P. V., Kreuzer L.S. and Scholz A.M. (2018) Livest. Sci. 213(2018):7–13. https://doi.org/10.1016/j.livsci.2018.04.008

Bone C., Anderson C., Lou Y. and Squires E.J. (2019) J. Steroid Biochem. Mol. Biol. 185(2019):218–224. https://doi.org/10.1016/j.jsbmb.2018.09.006

Bonneau M. (1998) Meat Sci. 49(S1):257–272. https://doi.org/10.1016/S0309-1740(98)90053-5

Burgeon C., Debliquy M., Lahem D., Rodriguez J., Ly A. *et al.* (2021a) Trends Food Sci. Technol. 112:283–297. https://doi.org/10.1016/j.tifs.2021.04.007

Burgeon C., Markey A., Debliquy M., Lahem D., Rodriguez J. *et al.* (2021b). Foods 10(6):1311 https://doi.org/10.3390/foods10061311

Claus R., Weiler U. and Herzog A. (1994) Meat Sci. 38(2):289–305. https://doi.org/10.1016/0309-1740(94)90118-X

Drag M., Skinkyté-Juskiené R., Do D.N., Kogelman L.J.A. and Kadarmideen H.N. (2017) Sci. Reports 7(12205):1–18. https://doi.org/10.1038/s41598-017-11928-0

Drag M., Hansen M.B. and Kadarmideen H.N. (2018) PLoS One 13(2):1–30 https://doi.org/10.1371/journal.pone.0192673

Duarte D.A.S., Schroyen M., Mota R.R., Vanderick S. and Gengler N. (2021) J. Appl. Genet. 62:137–150. https://doi.org/10.1007/s13353-020-00598-w

Dubois L., Delhaye M., Mayeres P., Gillard N., Delahaut P. *et al.* (2021) Proc. of 18th GCxGC Symposium.

European Commission Establishing Best Practices on the Production, the Processing and the Marketing of Meat from Uncastrated Pigs or Pigs Vaccinated against Boar Taint (Immunocastrated) (2019). Available online: https://ec.europa.eu/food/sites/food/files/animals/docs/aw_prac_farm_pigs_cast-alt_establishing- best-practices.pdf (accessed on 6 January 2022).

Fischer J., Gerlach C., Meier-Dinkel L., Elsinghorst P.W., Boeker P. *et al.* (2014) Food Res. Int. 62:35–42. http://dx.doi.org/10.1016/j.foodres.2014.02.045

Font-i-furnols M. (2012) Meat Sci. 92(4):319–329. https://doi.org/10.1016/j.meatsci.2012.05.009

Font-i-furnols M., Martín-Bernal R., Aluwé M., Bonneau M., Haugen J.-E. *et al.* (2020) Animals 10(1886):1–27. https://doi.org/10.3390/ani10101886

Jensen M.T., Cox R.P. and Jensen B.B. (1995a) Appl. Environ. Microbiol. 61(8):3180–3184. https://doi.org/10.1128/aem.61.8.3180-3184.1995

Jensen M.T., Cox R.P. and Jensen B.B. (1995b) Anim. Sci. 61:293–304. https://doi.org/10.1017/S1357729800013837

Li X., Jensen B.B. and Canibe N. (2019) Appl. Environ. Microbiol. 85(6). https://doi.org/10.1128/AEM.02327-18

Liu X., Trautmann J., Wigger R., Zhou G. and Mörlein D. (2017) Food Chem. 231:301–308. https://dx.doi.org/10.1016/j.foodchem.2017.03.112

Mörlein D. and Tholen E. (2014) Meat Sci. 99:1–7. https://dx.doi.org/10.1016/j.meatsci.2014.08.002

Rius M.A., Hortos M. and Garcıa-Regueiro J.A. (2005) Meat Sci. 71(4):595–602. https://doi.org/10.1016/j.meatsci.2005.03.014

Rius Solé M.A. and Garcia Regueiro J.A. (2001) J. Agric. Food Chem. 49(11):5303–5309. https://dx.doi.org/10.1021/jf010482d

Squires E.J., Bone C. and Cameron J. (2020) Animals 10(1665):1–22. https://dx.doi.org/10.3390/ani10091665

Wesoly R. and Weiler U. (2012) Animals 2:221–242. https://doi.org/10.3390/ani2020221

787. Genome-wide association study for oxidative stress indicators in replacement gilts

S.A. Breuer[1*], J. Chinchilla-Vargas[2], B.J. Kerr[3] and K.J. Stalder[1]

[1]Department of Animal Science, Iowa State University, 806 Stange Rd, Ames, IA 50011, USA; [2]PIC North America, 100 Bluegrass Commons Blvd #2200, Hendersonville, TN 37075 USA; [3]USDA-ARS-NLAE, 1015 N. University Blvd., Ames, IA 50011, USA; sabreuer@iastate.edu

Abstract

Oxidative stress (OS) may be an important contributor to sow longevity, but little is known about OS variation or its genetic implications for commercial swine production. The goal of this study was to assess OS levels in gilts prior to breeding and identify possible genomic relationships to OS using genome wide associations (GWAS). Blood was collected from 318 crossbred gilts to perform reactive oxygen species (ROS) and serum antioxidant capacity (SAC) analyses. Following a GWAS analysis, one window of interest explainied 0.67% of variation for SAC. Two biologically relevant genes, monoamine oxidase (MAO) A and B were identified. Behaviour traits may be associated with MAOA which may be impactful as the US swine industry moves to group sow housing. Estimates from this analysis should be viewed cautiously due to the small numbers of animals included. Further research is warranted to evaluate MAO's potential impact on sow behaviour and longevity.

Introduction

OS is described as an imbalance between oxidants and antioxidants at the cellular level (Lykkesfeldt and Svendsen, 2007). ROS, commonly referred to as free radicals, are produced at the cellular level through electron transport chain leakage as part of an animal's energy metabolism. However, there are instances that may exacerbate this phenomenon in swine production such as pregnancy (Covarrubias *et al.*, 2008), lactation (Zhao *et al.*, 2013), poor nutrition (Kim *et al.*, 2013), environment (Wang *et al.*, 2021), and management (Buchet *et al.*, 2017). Uncontrolled ROS production can cause damage to cellular components such as DNA, RNA, cholesterol, lipids, carbohydrates, and proteins (Celi and Gabai, 2015). In contrast, antioxidants are used to neutralize ROS before cellular damage occurs. Two key components associated with neutralizing ROS at a cellular level are antioxidant networks and SAC. To this end, the ratio of ROS to SAC has been used to calculate an oxidative stress index (OSi) which has been suggested to characterize the animals' overall stress level (Abuelo *et al.*, 2013).

Narrow-sense heritability estimates for the oxidative stress biomarker, glutathione peroxidase, has been reported to be high at 0.55 (±0.05) in studies involving slaughter pigs (Ntawubizi *et al.*, 2020). While narrow-sense heritability helps determine the proportion of phenotypic variance dictated by additive genetic variance, GWAS aim to identify specific relationships between genomic regions and phenotypes. Single nucleotide polymorphisms were used in this study to isolate specific genomic windows that might be useful in identifying a specific gene to improve traits that are economically and biologically important in commercial swine production. With growing concern for sow longevity, measuring stress that animals experience has become a focus for researchers to combat the associated negative effects of such stress, which is increasing because of the industry shift to group sow housing. Sows have shown a difference in various oxidative stress parameters due to different housing scenarios and social ranking. Zhao *et al.* (2013) reported that sows exhibiting a lower social ranking in a group housing system had increased levels of stress biomarkers indicating aggression and behaviour could contribute to oxidative stress status.

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_786, © S.A. Breuer *et al.* 2022

Because the relationship between oxidative stress status in breeding animals and genetics is lacking, the goal of this study was to determine baseline oxidative stress status for gilts, post puberty but prior to mating, and determine if any genomic regions are related to these measurements.

Materials & methods

The experiment was carried out according to the guidelines of the Institutional Animal Care and Use Committee of Carthage Innovative Swine Solutions for the humane use and care of swine in research.

Animals and biological sampling. A total of 318 crossbred replacement gilts (PIC 1050) on a commercial farm in Illinois (US) were involved in this study. Gilts were sourced from the same multiplier and born within the same week. All gilts were provided *ad libitum* access to the same diet, similar housing conditions, and managed (boar exposure, vaccinations, etc.) following the company's standard operating procedures. At 23 weeks of age, two blood samples were taken via jugular puncture from each gilt with one sample tube containing sodium heparin to analyse OS parameters and one sample tube containing EDTA for genetic analysis. Serum samples were obtained from the blood in sodium heparin tube after centrifugation at 2,500×*g* and 20 °C for 15 min and stored at -80 °C until analysis.

Oxidative stress status analyses. ROS were assayed using the spectrophotometric d-ROM test (Diacron International, Grosseto, Italy) as described by Trotti *et al.* (2001). The d-ROM assay determines hyper peroxides through a reaction with the chromogen. Results are expressed as 'Caratelli units' (CarrU) where 1 CarrU is equivalent to the oxidizing power of 0.08 mg H_2O_2/ml. SAC was assayed with the OXY-Adsorbent Test (Diacron International, Grosseto, Italy) (Trotti *et al.*, 2001). The SAC assay utilizes a hypochlorous acid (HClO) solution to oxidize antioxidant pools (albumin, bilirubin, uric acid, thiol groups, vitamins, glutathione peroxidase, superoxide dismutase, catalase, etc.) within the serum, yielding the cumulative antioxidant potential action relative to the sum of all antioxidants. Results are expressed as μmol HClO/ml. Directions for both assays were followed according to the kit booklet and adapted to a 96 well plate absorption spectrophotometer (BioTek Cytation15, Santa Clara, CA). While both methods have been used in bovine (Abuelo *et al.*, 2013) and human studies, they have not been used in swine.

Variance component estimation and genome-wide association study. ROS and SAC traits were analysed using a linear model that contained a single fixed effect of plate number consisting of 10 and 12 levels, respectively. The plate effect was included because of known differences in standardization of kit calibrators. In the case of OSi, two fixed effects accounting for the plate level for ROS and SAC respectively were included. A genomic relationship matrix was constructed as described by VanRaden (2008) and used to account for genetic correlations between genotypes and phenotypes. Given the small sample size, variance components were estimated using a Bayesian inference of 150,000 iteration, burning 25,000, and thinning every 50. These estimations were performed with Thrshgibbs1 and postgibbs1, members of the blupf90 suite of programs (Misztal *et al.*, 2018). A single step genomic best linear unbiased (GBLUP) method under a Bayesian approach was used to identify 1.0 Mb candidate genomic regions associated with the three traits based on the percentage of additive variance they explained, using PostGSf90 (Aguilar *et al.*, 2014). Because of the relatively small animal numbers involved in this study, it was decided that GO would not be beneficial so the focus would be on windows that explained a proportion of GV that was markedly greater than others. Additionally, the genes that were included in those windows along with 1 mb up and downstream from the specific gene location were evaluated in the present analyses.

Results

Descriptive statistics and genetic parameter estimates are provided in Table 1 for OS traits. As shown in Figure 1, window (19_38)[1] explained 0.67% of variation in SAC. There were two genes of interest annotated as MAO A and MAO B and will be referred to as the MAO system.

Discussion

Metabolism of neurotransmitters involved in aggressive behaviour include dopamine, adrenaline, noradrenaline, histamine, and serotonin (Kolla and Bortolato, 2020). These compounds are degraded by the enzymes, MAO A and B, encoded by the MAO genes. During MAO catalysis, H_2O_2, NH_4^+, and aldehydes are products that remain, where H_2O_2 is known for being an agonist of ROS production (Edmondson and Binda, 2018). MAO A and B have demonstrated to be related to stress in many trials involving mice (Cases *et al.*, 1995; Doyle *et al.*, 1996). Overall, OS is biologically associated with increased MAO expression in mammals and could be used to further investigate stress in swine production settings. However, low activity of MAO results in increased aggressive behaviour (Mentis *et al.*, 2021) indicating that a balance

Table 1. Descriptive statistics and heritability estimates for each OS trait.

Trait	Unit	N	Mean	Min	Max	SD	Heritability
ROS	CarrU	318	792.1	547.6	1,119.4	112.7	0.08
SAC	µmol HClO/ml	318	382.7	91.9	950.6	174.2	0.12
OSi	CarrU/(µmol HClO/ml)	318	2.5	0.9	8.7	1.3	0.16

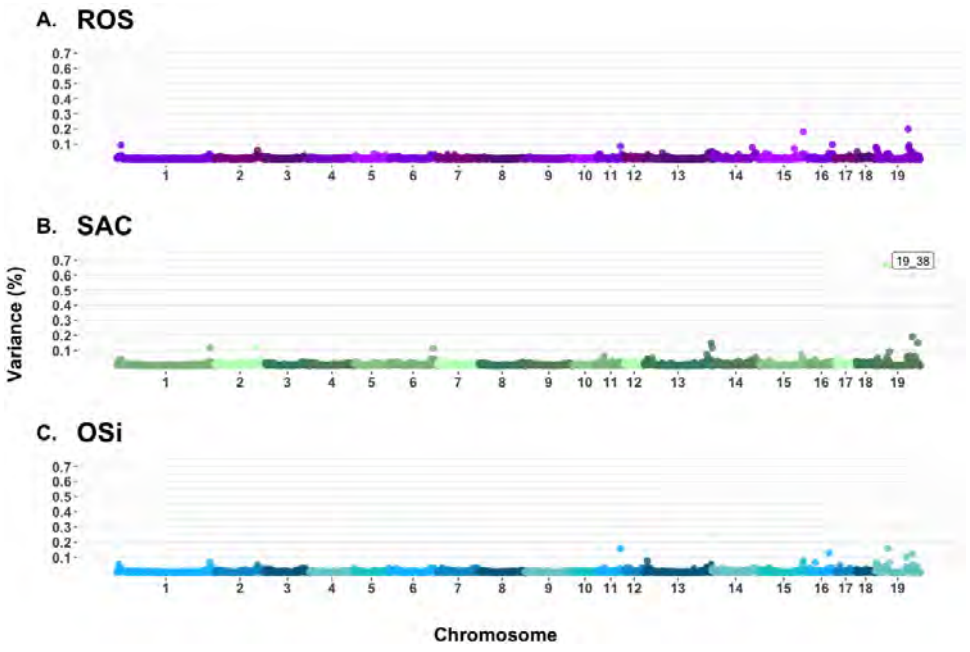


Figure 1. Manhattan plot indicating the proportion of window variance by genome location for ROS, SAC, and OSi traits.

[1] Chromosome X is written as 19 in the plot.

of MAO gene and enzyme expression is necessary to achieve optimal oxidative stress homeostasis and behavioural disposition. Estimates of this study, while encouraging, should be interpreted with caution due to the small sample size used in the genomic analysis. Therefore, research with an increased sample size is needed to determine how useful this information is in a breeding program to reduce oxidative stress and prolong cellular efficiency in breeding swine.

References

Abuelo, A., Hernandez, J., Benedito, J.L., and Castillo, C. 2013. Animal 7 (8):1374-1378. https://doi.org/10.1017/s1751731113000396

Aguilar, I., Misztal, I., Tsuruta, S., Legarra, A., and Wang, H.. 2014. Proc of the 10th WCGALP, Vancouver, Canada.

Buchet, A., Belloc, C., Sialelli, J., Lacoste, A., and Merlot, E. 2017. Journ. Rech. Por. 49:189-194.

Cases, O., Seif, I., Grimsby, J., Gaspar, P., *et al.* 1995. Science 268 (5218):1763-1766. https://doi.org/10.1126/science.7792602

Celi, P., and Gabai, G. 2015. Front. Vet. Sci. 2. https://doi.org/10.3389/fvets.2015.00048

Covarrubias, L., Hernandez-Garcia, D., Schnabel, D., Salas-Vidal, E., and Castro-Obregon, S. 2008. J. Dev. Biol. 320 (1):1-11. https://doi.org/10.1016/j.ydbio.2008.04.041

Doyle, A., Hucklebridge, F., Evans, P., and Clow, A. 1996. J. Life Sci 59 (16):1357-1362. https://doi.org/10.1016/0024-3205(96)00461-4

Edmondson, D.E., and Binda, C. 2018. Membrane Protein Complexes: Structure and Function, SPRINGER, New York, US. https://doi.org/10.1007/978-981-10-7757-9

Kim, S.W., Weaver, A.C., Shen, Y., and Zhao, Y. 2013. J. Anim. Sci Biotechnol 4 26:(2013). https://doi.org/10.1186/2049-1891-4-26

Kolla, N.J., and Bortolato, M. 2020. Prog. Neurobiol. 194:101875. https://doi.org/10.1016/j.pneurobio.2020.101875

Lykkesfeldt, J., and Svendsen, O. 2007. Vet. J. 173 (3):502-511. https://doi.org/10.1016/j.tvjl.2006.06.005

Mentis, A.F.A., Dardiotis, E., Katsouni, E., and Chrousos, G.P 2021. Transl. Psychiatry 11 (1). https://doi.org/10.1038/s41398-021-01257-2

Misztal, I., Tsuruta, S., Lourenco, D., Aguilar, I., *et al.* 2018. University of Georgia.

Ntawubizi, M., Raes, K., and De Smet, S 2020. J. Anim. Sci. 98 (1). https://doi.org/10.1093/jas/skz378

Trotti, R., Carratelli, M., Barbieri, M., Micieli, G., *et al.* 2001. Haematologica 86 (1):85-91.

VanRaden, P.M. 2008. J. Dairy Sci. 91 (11):4414-4423. https://doi.org/10.3168/jds.2007-0980

Wang, H., Zeng, X.Y., Zhang, X.X., Liu, H.G., and Xing, H.J. 2021. Ecotoxicol. Environ. Saf. 212. https://doi.org/10.1016/j.ecoenv.2021.111974

Zhao, Y., Flowers, W.L., Saraiva, A., Yeum, K.J., and Kim, S.W. 2013. J. Anim. Sci. 91 (12):5848-5858. https://doi.org/10.2527/jas.2013-6388

788. Genetic determinism of boar taint in the French Landrace pig breed

C. Larzul[1*], M.-J. Mercat[2], C. Hassenfratz[2], C. Carillier-Jacquin[1], R. Comte[3], B. Blanchet[4], I. Louveau[3], S. Boulot[2] and A. Prunier[3]

[1]Université de Toulouse, INRAE, ENVT, GenPhySE, 24 chemin de Borde-Rouge, 31326 Castanet-Tolosan, France; [2]IFIP-Institut du porc, Alliance R&D, La Motte au Vicomte, BP 35104, 35651 Le Rheu, France; [3]PEGASE, INRAE, Institut Agro, 16 Le Clos, 35590 Saint-Gilles, France; [4]INRAE UE3P, 16 Le Clos, 35590 Saint-Gilles, France; catherine.larzul@inrae.fr

Abstract

Since pig breeds differ in their risk of developing boar taint, it is important to consider line-by-line selection by estimating genetic parameters of odorant compounds in relation to sexual development and production traits. The present study focused on French Landrace, one of the main French maternal lines. More than 1000 non-castrated males were measured for growth rate, body composition and meat quality, fat-odorant compounds, androstenone and skatole, and four steroid hormones. As expected, heritability values were high for androstenone (h^2=0.69) and estradiol (h^2=0.79). Genetic correlations of boar taint compounds were moderate to high with testosterone and estradiol but close to 0 with progesterone and cortisol. Genetic correlations of these compounds were low with growth rate, ultimate pH or color, but moderate and favorable with carcass leanness. It is therefore possible to include fat-odorant compounds in selection objectives, either directly or via sex hormones, while controlling their impacts on production performance.

Introduction

In France, male piglet castration without anaesthesia has been prohibited since January 1st, 2022. One of the primary motivations for castration of male pigs is to improve the quality of the meat, mainly to prevent the appearance of an odor defect known as boar taint. The two main components responsible for this defect are skatole (produced by bacteria in the digestive tract) and androstenone (a steroid produced by testes) which accumulate in fatty tissue. Reducing the level of these two compounds largely resolves the occurrence of the defect.

To promote the breeding of uncastrated male pigs, several pig lines have been developed under the label 'low boar taint' (Larzul, 2021). Most of the effort has focused on the paternal lines. However, the influence of maternal lines should not be overlooked. In that respect, maternal lines are more at risk than paternal lines (Larzul, 2021). Specifying the genetic determinism of boar taint compounds in maternal lines is therefore an essential prerequisite for implementing selection strategies aimed at limiting boar taint without penalizing production traits. We estimated the genetic parameters of the boar taint compounds, in relation with growth, feed efficiency, body composition and meat. Steroid hormones have also been included in the study in order to clarify the genetic relationships between boar taint and testicular and adrenal metabolism and, possibly, to propose alternative selection criteria to reduce the risk of boar taint.

Materials & methods

Animals. Between two and three weeks of age, male French Landrace piglets were transferred from their birth farm to the Rheu phenotyping station in 14 piglet post-weaning pens (UE3P, https://doi.org/10.15454/1.5573932732039927E12). Around 26 kg, all the pigs of the pen were simultaneously transferred to the fattening pens to start the control at 35 kg live weight. The animals were slaughtered at the Cooperl Arc Atlantique slaughterhouse (Montfort, 35). At slaughter, the average body weight was 117±6 kg, for a median age of 168 days. During the control period, the pigs were fed *ad libitum* with a feed with a net

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_788, © C. Larzul *et al.* 2022

energy content (NE) of 9.6 MJ/kg, digestible lysine content of 0.94 g/MJ EN, and of digestible tryptophan of 1.7 g/kg. A total of 1,084 animals, from 249 sires, were bred in 33 batches between June 2018 and April 2021.

Measures. The average daily gain (ADG) was calculated from the weights at the start and at the end of the control period. The feed conversion ratio (FCR) and the average daily feed intake (DFI) were calculated from the consumption recorded by the feeding device. A day before departure to the slaughterhouse, a blood sample was collected in heparinized tubes from the jugular vein. The blood samples were then centrifuged and the plasma was stored at -20 °C until the concentrations of testosterone, estradiol, progesterone and cortisol were measured. Concentrations were measured using ELISA kits (ST AIA-Pack Testosterone, ST AIA-Pack hsE2, ST AIA-Pack CORT, ST AIA-PackPROGIII, Tosoh Corporation, Tokyo, Japan) developed for an automatic analyzer (AIATosoh Corporation, Tokyo, Japan). Dressing percentage (DY) is the ratio between the weight of the carcass with head 24 hours after slaughter and the weight before departure to the slaughterhouse after fasting. Carcass leanness (%M) was automatically estimated at the slaughterhouse (CSB Image-Meater®). The pH was measured (Sydel pHmeter, Xerolyt electrode) in the semimembranous muscle 24 hours post mortem (pH24). The day after slaughter, the colour parameters (L*a*b*) were measured (Minolta CR300 chromameter) on a fresh section of the *gluteus superficialis*. Finally, a sample of backfat was taken from the neck and frozen for subsequent determination of androstenone and skatole (Batorek *et al.*, 2012).

Statistical analyses. Androstenone and skatole contents lower than 0.2 and 0.03 µg/g, respectively, can't be precisely measured. All values below these thresholds have been brought back to the threshold values. To normalize the distributions, the measurements obtained from the assays were transformed before analysis with the log function, except for the measurement of testosterone (square root transformation). All variables were then reduced and centered. For the dosages and the meat quality characteristics, the model included the effect of the weight at the end of the control period and the date of slaughter, equivalent to the batch of measurement. For the growth and food consumption traits, the model included the weight at the start of the control, the batch effect and the fattening pen as a random effect. For body composition measurements, the model took into account the weight before departure and the batch effect. The animal effect was added as a random effect in order to estimate the additive genetic variance. Six generations were considered to build the kindship matrix. The analyses were carried out with the VCE6 software (Neumaier and Groeneveld, 1998). Estimates of the variance components were performed jointly for the six assays. Other genetic correlations were estimated by two-trait analyses including one of the traits obtained from the assays and one of the other traits.

Results

The estimated heritabilities for these traits are all moderate to high. The highest heritabilities are observed for plasma estradiol level (h^2=0.80) and androstenone level in back fat (h^2=0.83). The lowest heritabilities were observed for the level of testosterone (h^2=0.26) and progesterone (h^2=0.34). The estimated heritabilities for cortisol (h^2=0.54) and skatole (h^2=0.50) were medium. The genetic correlations between compounds and production traits are reported in Table 1.

Discussion

Heritabilities. The estimated heritability for androstenone is in the upper range of previously published values (Larzul, 2021). The heritability value of skatole content was expected to be lower, according to results published in the literature (Larzul, 2021). It is recognised that skatole is more subject to environmental effects than androstenone. The heritability of estradiol is higher than those estimated in a population of Landrace (h^2=0.65) in Norway (Grindflek *et al.*, 2011) and much higher than the values estimated in German Landrace (h^2=0.09) (Brinke *et al.*, 2021). The result can probably be explained by a stricter

Table 1. Genetic correlations between boar taint compounds, steroid hormones and production traits.

	Es	Pr	Co	An	Sk	ADG	FCR	DFI	%M	DY	L*	a*	b*	pH24
Te	0.70	0.12	0.00	0.60	0.64	-0.07	0.54	0.43	-0.51	-0.54	-0.07	0.44	0.34	-0.07
Es		0.23	0.04	0.87	0.51	0.16	0.45	0.43	-0.44	-0.38	-0.03	0.4	0.26	-0.1
Pr			0.44	0.34	0.15	0.23	0.36	0.47	-0.10	-0.13	-0.05	0.09	0.13	0.07
Co				0.08	-0.08	0.00	0.08	0.11	-0.09	0.06	0.17	-0.19	-0.02	-0.12
An					0.45	0.09	0.17	0.19	-0.34	-0.39	-0.16	0.31	0.09	-0.05
Sk						-0.02	0.28	0.23	-0.28	-0.48	0.03	-0.06	-0.02	0.01

[1] Te: testosterone; Pr:progesterone; Co: cortisol; An : androstenone; Sk : skatole; ADG : average daily gain; FCR : feed conversion ratio; DFI: daily feed intake; %M: carcass leanness; DY: dressing yield

standardisation of the sampling and/or by the higher age and weight of the animals in the present study. The lower heritability of testosterone compared to estradiol (Grindflek *et al.*, 2011; Dugué *et al.*, 2020), confirmed by this study, is explained by a production of testosterone that is pulsatile and more sensitive to changes in the environment and stress. Few published results exist for the heritability of plasma cortisol in pigs In Landrace, at the end of fattening, Kadowaki *et al.* (2012) estimated a plasma cortisol heritability value of 0.2 and Brinke *et al.* (2021) of 0.11. The differences with the present study could be explained by breed or age at sampling. Knowing that cortisol heritability seems higher in a stressful situation (Larzul *et al.*, 2015), it is possible that the measured level does not reflect the basal level, as at the time of blood sampling, the animals were already isolated, for their departure to the slaughterhouse.

Genetic correlations. As expected, the genetic correlations between testosterone, estradiol and androstenone are high and in agreement with the published values (Larzul, 2021). Grindflek *et al.*, 2011 estimated much higher correlation values between these traits (from 0.8 to 0.95) in Landrace populations than those of the present study. In contrast, Brinke *et al.* (2021) published correlations of the same magnitude (from 0.49 to 0.89). These high correlations reflect the common metabolic pathways for these steroids (Robic *et al.*, 2014). The genetic correlations between, on the one hand, the steroids produced by the testes (androstenone, estradiol and testosterone) and, on the other hand, the steroids produced mainly by the adrenal glands (progesterone and cortisol) are weak (from zero to 0.34). Considering Brinke *et al.* (2021) and present results, it can be assumed that the steroid activity of the adrenal glands and testes are poorly genetically related.

Growth rate is not genetically related to any of the measurements except with progesterone (r_g=0.23). FCR and DFI are positively correlated not only with testicular hormones but also with progesterone, suggesting an overall genetic relationship with steroid metabolism. In contrast, %M is negatively correlated with testicular hormones, and more particularly with testosterone. This result is contrary to expectations, given that testosterone is known to have an anabolic effect on muscle. However, it should be considered that the animals were in the process of puberty and low levels of testosterone and estradiol indicate an early stage of sexual maturity for individuals just starting to deposit fat. The negative genetic correlations between %M and androstenone or skatole are consistent with the fact that the leaner lines are less at risk of developing boar taint (Larzul, 2021). The negative and moderate genetic correlation between yield and skatole might be related to viscera development. The pH measured 24 hours post mortem is not genetically linked to steroid hormones and odorant compounds as already shown by Dugué *et al.* (2020) in Pietrain. As for the colour parameters, they are moderately correlated with estradiol and testosterone (around 0.3) as well as with androstenone for the red index (r_g=0.31). The links between steroids and meat quality could result from the pubertal stage of development of the animals. In general, less mature stages are associated with lighter

and less red meat, thus individuals more advanced in their pubertal development would have darker meat with a higher red index.

The estimation of the genetic parameters on boar taint related traits in one of the main French maternal lines shows that overall selection against boar taint will have indirect responses similar to those of the sire lines. However, in these maternal lines, it is also imperative to estimate the genetic relationships between boar taint risk (in males) and reproductive traits (in related females).

References

Batorek N., Škrlep M., Prunier A., Louveau I., Noblet J. *et al.* (2012) J. Anim. Sci. 90:4593-4603.

Brinke I., Grosse-Brinkhaus C., Roth K., Pröll-Cornelissen M.J., Klein S. *et al.* (2021) Animals 11:231.

Dugué C., Prunier A., Mercat M.J., Monziols M., Blanchet B. *et al.* (2020). Animal 1:1333–1341.

Grindflek E., Meuwissen T.H.E., Aasmundstad T., Hamland H., Hansen M. H. S. *et al.* (2011) J. Anim. Sci. 89:680-692.

Kadowaki H., Suzuki E., Kojima-Shibata C., Suzuki K., Okamura T., *et al.* (2012) Livest. Sci. 147:20–26.

Larzul C., Terenina E., Foury A., Billon Y., Louveau I., *et al.* (2015) Animal 9:1929-1934.

Larzul C. (2021) Animals, 11:699.

Neumaier A. Groeneveld E. (1998) Genet. Sel. Evol. 30:3-26.

Robic A., Faraut T., Prunier A. (2014) J. Steroid Biochem. Mol. Biol. 140:44-55.

789. A comprehensive overview of mitochondrial DNA insertions in the nuclear genome of the pig

M. Bolner*, M. Ballan, S. Bovo, G. Schiavo and L. Fontanesi

Department of Agricultural and Food Sciences, Division of Animal Sciences, University of Bologna, Viale Giuseppe Fanin 46, 40127 Bologna, Italy; matteo.bolner2@unibo.it

Abstract

The integration of nuclear mitochondrial DNA into the mammalian genomes is an ongoing but rare evolutionary process that produces nuclear sequences of mitochondrial origin (NUMTs). In this study we identified and analysed NUMTs of the porcine genome. We first constructed a comparative distribution map of NUMTs in the Sscrofa11.1 reference genome and in other 14 assembled and drafted pig genomes (of Asian and European origin). Then, a total of 581 whole genome sequencing datasets, from 35 different breeds or populations, were analysed for the discovery of recently integrated and polymorphic NUMT positions. Overall, we built a comprehensive catalogue of more than 500 NUMTs, some of which were not included in the assemble genomes. Breed-specific NUMTs as well as segregating NUMTs were also identified. This work lays the base for in-depth studies on the phylogeny of *Sus scrofa* populations.

Introduction

Nuclear sequences of mitochondrial origin (NUMTs) are insertions of mitochondrial DNA (mtDNA) into the nuclear genome of an eukaryotic organism through horizontal transfer mechanisms (Lopez *et al.* 1994). NUMTs are sequence fossils that provide phylogenetic information among related species, sub-species and lines and, for this reason, can be useful to reconstruct the history of animal domestication (Schiavo *et al.* 2017). The integration of NUMTs into the nuclear genome is still an ongoing, even if rare, evolutionary process.

The recent rise in genomic data available for livestock has allowed the discovery of NUMTs in different species (e.g. Pereira and Baker 2004; Schiavo *et al.* 2018, 2019). NUMTs in the genome of the pig (*Sus scrofa*) were recently investigated by Schiavo *et al.* (2017) using a LAST-based (Kiełbasa *et al.* 2011) pipeline and the Sscrofa10.2 reference genome. However, a new version of the pig genome (Sscrofa11.1), constructed from the same Duroc pig of the previous version, has been released in 2017. This is the current standard reference genome. Therefore, it would be needed to redefine the NUMT map in this new version and evaluate the differences between the two reference genomes (Sscrofa10.2 *vs* Sscrofa11.1) and obtain an updated catalogue of NUMTs present in the pig genome.

Over the last few years, many whole genome sequencing (WGS) datasets have been produced in all livestock species, including the pig. These datasets can be mined to detect additional NUMT regions, not included in the reference genome and other NUMT sequences that might have been recently inserted in the genome of some breeds, populations or individuals.

Here, we built a comprehensive catalogue of porcine NUMTs by analysing the novel Sscrofa11.1 reference genome, 14 additional drafted pig genomes (of Asian and European origin) and about 600 WGS datasets.

The results of this study could be useful for gaining a better understanding of the phylogenetic history of the domestic pig, the relationships between breeds and, eventually, the impact of polymorphic NUMTs in affecting phenotypic traits in this species.

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_789, © M. Bolner *et al.* 2022

Materials & methods

Reference genomes and WGS datasets. The reference pig genome (Sscrofa11.1) and the drafted genome reference sequence of 14 different pig breeds (Table 1) were downloaded from the National Center for Biotechnology Information (NCBI, https://www.ncbi.nlm.nih.gov/). Data included both the nuclear and mitochondrial genomes. Moreover, WGS data of 464 pigs from 35 pig breeds and populations were retrieved from the European Nucleotide Archive (ENA; https://www.ebi.ac.uk/ena/browser/). This dataset was enlarged by including additional 117 WGS data obtained from individual pigs of the three Italian heavy pig breeds (Italian Large White, Italian Landrace and Italian Duroc).

Detection of NUMTs in the reference and drafted genomes. The pipeline followed what previously proposed by Schiavo *et al.* (2017). Briefly, the mitochondrial genome was circularized by concatenating two copies of the complete mtDNA in order to capture matches extending over the boundaries of the linear sequence. LAST software (Kiełbasa *et al.* 2011) was used to query the Sscrofa11.1 reference genome and the additional 14 reference genomes. Then, alignments pointing out putative NUMTs were retained considering the E-value statistics. A threshold of E-value = 55 was obtained based on the method proposed by Tsuji *et al.* (2012). A comparative analysis and mapping of NUMTs included in Scrofa11.1, in Sscrofa10.2 and in all 14 drafted genome versions was obtained by aligning the flanking regions (1000 bp each) of each detected NUMT with all the other genomes.

NUMT detection from WGS dataset. The pipeline followed, with some adjustments, the approach described by Ju *et al.* (2015). Reads were initially mapped over the reference genome with BWA (Li and Durbin, 2009). Then, the genome was divided in windows of 500 bp that were evaluated for the presence of discordant reads (i.e. one mate aligned to mtDNA and the other one aligned to a nuclear chromosome). BLASTN (Altschul *et al.* 1990) was then used to realign reads contained in windows presenting discordant reads and to confirm those regions/windows representing NUMTs. Then, windows containing soft clipped reads were evaluated for the possible alignment of the clipped portion, via BLASTN, to the porcine mtDNA.

Table 1. NUMTs identified in the reference and drafted pig genomes.

Breed	GenBank ID	n. of NUMTs	Private NUMTs	NUMTs size (bp)	Coverage (%)[1]
Duroc[2]	GCA_000003025.6	488	1	297,503	0.012
Berkshire	GCA_001700575.1	493	0	279,525	0.011
Hampshire	GCA_001700165.1	451	1	269,114	0.011
Landrace	GCA_001700215.1	459	0	266,685	0.011
Large White	GCA_001700135.1	442	0	261,456	0.011
Nero Siciliano	GCA_006511355.2	445	0	281,837	0.011
Pietrain	GCA_001700255.1	455	1	270,699	0.011
Bama miniature	GCA_007644095.1	480	2	295,995	0.012
Bamei	GCA_001700235.1	465	3	266,519	0.011
Jinhua	GCA_001700295.1	463	0	266,877	0.011
Meishan	GCA_001700195.1	476	0	284,625	0.012
Ningxiang	GCA_020567905.1	479	0	293,148	0.012
Rongchang	GCA_001700155.1	460	1	266,020	0.011
Tibetan	GCA_000472085.2	468	2	255,068	0.010
Wuzhishan	GCA_000325925.2	529	2	301,494	0.012

[1] Fraction of the nuclear genome covered by NUMTs.
[2] Sscrofa11.1 reference genome.

This led to the identification of *de novo* insertions as well as the portion of inserted mtDNA and the location of the breakpoints at the nuclear level.

Results

Detection and comparative evaluation of NUMTs from reference genomes. The number of NUMTs detected in the reference genome (Duroc) as well as in the other drafted genomes is reported in Table 1. The number of matches resulted similar across all breeds, with an average ± standard deviation of 466±15. The size of these NUMTs ranged from 56 bp to 10,996 bp. Of the identified NUMTs, 293 are shared between all breeds. We also detected private NUMTs (i.e. breed specific) in the range of 0 to 3 per breed. NUMTs on the same chromosome (or scaffold) within 20 kb of each other were annotated as part of the same NUMT region, which is expected to be derived by the same insertional events followed by subsequent mutations that divided the integrated mtDNA. Across the 15 analysed reference or drafted genome sequences we detected from 199 to 261 NUMT regions. Figure 1 shows the distribution of the different NUMT length classes in the two reference genomes (Sscrofa10.2 and Sscrofa11.1). The two distributions resulted similar, with the majority of NUMTs covering the class of 100-500 bp.

Detection and evaluation of NUMTs from WGS datasets. The comprehensive evaluation of WGS datasets led to confirm the existence of most NUMTs (those retrieved from the evaluation of the reference genome) as well as the detection of *de novo* NUMT insertions (i.e. NUMTs not included in the reference or drafted genome versions). For example, the analysis based on Italian heavy pig breeds detected a total of 383 NUMTs, 80% of which were already annotated in the reference genomes (they overlapped with known NUMTs). Moreover, the evaluation of soft-clipped reads led to the discovery of about 50 breakpoints in the nuclear genomes pointing out novel NUMTs, not included in the reference genome. This pipeline also led to evaluate the zygosity state of the NUMTs as well as their segregation across breeds.

Discussion

We present for the first time a comprehensive catalogue of porcine NUMTs detected using the combination of different approaches. We started by annotating NUMTs in the Sscrofa11.1 reference genome and

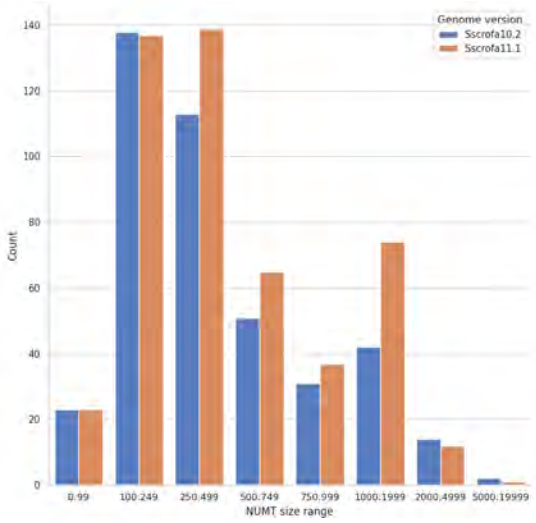


Figure 1. Comparison of NUMTs size distribution across the two pig reference genome versions.

then this information was integrated with NUMTs detected in drafted genomes of different breeds and unassembled WGS datasets. Most of NUMT positions were confirmed across all assembled genomes. In a few cases, partial overlapping of NUMT positions across the assembled genomes might be due to potential chimeric misassembled positions, incomplete sequences and polymorphic regions (i.e. presence/absence of the insertion), as determined when WGS datasets were used. All three zygosity states were identified using WGS datasets. That means that, as already demonstrated (Schiavo *et al.* 2017), some NUMT positions are polymorphic within and across breeds. Some of the polymorphic NUMT sites could be quite recent in terms of integration into the genome of some pigs whereas for others, the integration event probably occurred a few millions of years before present (MYBP). This might suggest that their polymorphic state is derived from relatively more recent (compared to the insertion event) interspecies admixture that contributed to shape the *S. scrofa* genome. These NUMTs can be very informative to reconstruct the genetic events that constituted the modern pig breeds and lines through different waves of gene flux and introgression, particularly, between European and Asian pig populations. Other NUMTs might be breed-specific. Direct PCR validation in a larger number of animals would be needed to confirm these results that can lead to the development of breed-specific markers.

References

Altschul S.F., Gish W., Miller W., Myers E.W. (1990) J Mol Biol 215(3):403-410. https://doi.org/10.1016/s0022-2836(05)80360-2

Chen J.M., Chuzhanova N., Stenson P.D., Ferec C. *et al.* (2005) Hum Mutat 25:207-221 https://doi.org/10.1002/humu.20133

Ju Y.S., Tubio J.M.C., Mifsud W., Fu B. *et al.* (2015). Genome Res. 25(6):814-824. https://doi.org/10.1101/gr.190470.115

Kiełbasa S.M., Wan R., Sato K., Horton P., Frith M.C. (2011) Genome Res. 21:487-93. https://doi.org/10.1101/gr.113985.110

Li H., Durbin R. (2009) Bioinformatics 25(14):1754-1760. https://doi.org/10.1093/bioinformatics/btp324

Lopez J.V., Yuhki N., Masuda R., Modi W., O'Brien, S.J. (1994). J. Mol. Evol. 39(2):174-190. https://doi.org/10.1007/BF00163806

Nergadze S.G., Lupotto M., Pellanda P., Santagostino M., Vitelli V. *et al.* (2010) Anim. Genet. 41(Suppl. 2):176-185. https://doi.org/10.1111/j.1365-2052.2010.02130.x

Pereira, S.L., Baker, A.J. (2004) BMC Evol Biol 4:17. https://doi.org/10.1186/1471-2148-4-17

Schiavo G., Hoffmann O.I., Ribani A., Utzeri V.J., Ghionda M.C. *et al.* (2017) DNA Res. 24(5):487-498. https://doi.org/10.1093/dnares/dsx019

Schiavo G, *et al.*, (2018) Anim Genet. 49(3):259-264 https://doi.org/10.1111/age.12648.

Schiavo G., Bovo S., Ribani A., Kazemi H., Fontanesi L. (2019) Anim. Genet. 51(1):149-151. https://doi.org/10.1111/age.12889

Tsuji J., Frith M.C., Tomii K., Horton P. (2012) Nucleic Acids Res., 40:9073-9088. https://doi.org/10.1093/nar/gks424

790. To Box-Cox or not to Box-Cox: phenotypic and genomic evaluation of litter size variability in Landrace pigs

G. Cieleń[1*], E.F. Knol[2], M.S. Lopes[2,3] E. Sell-Kubiak[1]

[1]Poznań University of Life Sciences, Department of Genetics and Animal Breeding, Wołyńska 33, 60-637 Poznań, Poland; [2]Topigs Norsvin Research Centre, Schoenaker 6, 6641 SZ Beuningen, the Netherlands; [3]Topigs Norsvin, R. Visc. do Rio Branco, 1310 Centro, Curitiba, PR, 80420-210, Brazil; cielengabriela@gmail.com

Abstract

This study aims to evaluate phenotypic and genomic background of variability of litter size obtained from data without and with Box-Cox transformation. We used 67,500 records on total number born (TNB) from Landrace pig population. Since the original data presented skewness the decision was made to perform Box-Cox transformation (bcTNB). Phenotypic variability was estimated as log-transformed variance of residuals (LnVar) for both TNB and bcTNB. EBVs obtained with LnVar were used to calculate the deregressed EBV, which were further used in genome-wide association study (GWAS) on ~5,000 sows and boars genotyped or imputed to 660K SNP-chip. The final GWAS results are still to be developed.

Introduction

Reproduction traits are highly important in livestock production and play an essential role in management and economic profits (Rempel *et al.*, 2010). Thus litter size in pigs has been intensively selected for decades in many breeding systems (Zhang *et al.*, 2019). Constant genetic improvement of this trait is possible thanks to existence of substantial genetic variation (Samorè *et al.*, 2016). However, due to confirmed genetic correlations (0.5) with increase of litter size also its variably is increasing (Sell-Kubiak *et al.*, 2015). The traits variability, can be measured by the heterogeneity of residual variance (Hill and Mulder, 2010).

Some authors (Rönnegård and Valdar, 2011; Yang *et al.*, 2011; Khaw *et al.*, 2016, Iung *et al.*, 2018) indicated that the data transformation can substantially affect the traits variability. Thus, it is suggested to use e.g. Box-Cox transformation instead of working with raw data. Moreover, the data transformation by affecting the estimates of trait variability can also lead to association with different genomic regions. So far only two studies on Large White pigs data (Sell-Kubiak *et al.* 2015;2022) indicated genomic regions involved in variability of litter size. Those previous studies, however, did not apply the data transformation.

Therefore, our study aims to evaluate the effect of Box-Cox transformation on trait variability and its phenotypic and genomic characteristics. In order to perform that we use Landrace pigs population and focus on total number born (TNB) and its variability estimated as log-transformed variance of residuals (LnVar).

Materials & methods

Phenotypes. Data for this study were obtained from the multiplication farm of the Topigs Norsvin (Vught, the Netherlands). In total 73,035 litter size (total number born, TNB) observations were available from 21,244 Landrace sows. After data editing, 67,500 litters from 16,273 Landrace sows remained for the analysis. The final TNB data presented skewness (Figure 1a), thus it was decided to perform the Box-Cox transformation. This was done using the MASS package in the R-program (R Development Core Team, 2011). For the estimation of λ, the fixed effects of parity of the sow, year and season of the farrowing and age at first insemination of a sow were used. Data after transformation are presented on Figure 1b and will be referred to as bcTNB.

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_790, © G. Cieleń *et al.* 2022

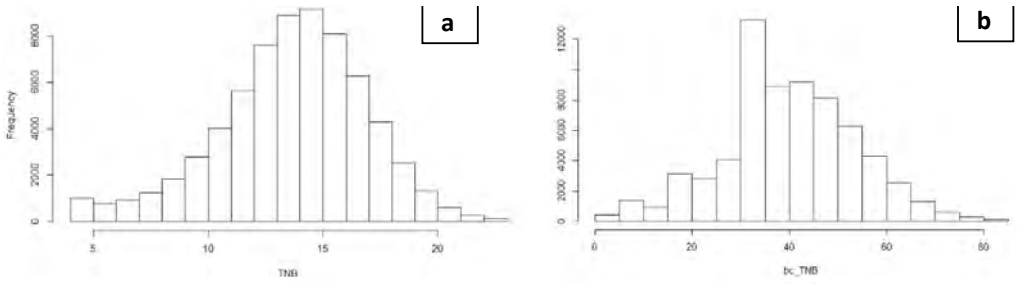


Figure 1. Distribution of raw litter size (TNB) and after Box-Cox transformation (bcTNB).

Statistical analysis. The animal model used for TNB and bcTNB in ASReml 4.1 (2015) was:

$$y = Xb + Za_d + Za_s + Upe + Vs + e \quad (1)$$

where y is a vector of observations on TNB or bcTNB; X, Z, U and V are known incidence matrices relating observations to fixed or random effects; b is a vector of fixed effects of parity of the sow, year and season of the farrowing and age at first insemination of a sow; a_d is a vector of dam additive genetic effect, with $a_d \sim N(0, A\,\sigma^2_{ad})$; a_s is a vector of random sire genetic effects with $a_s \sim N(0,A\,\sigma^2_{as})$; pe is a vector of random non-genetic effect of the sow (permanent sow effect), with $pe \sim N(0, I_{pe}\,\sigma^2_{pe})$; s is a vector of random non-genetic sire effect (service sire effect), with $s \sim N(0, I_c\,\sigma^2_c)$; e is a vector of residuals, with $e \sim N(0,I_e\,\sigma^2_e)$.

The residuals obtained for TNB and bcTNB were used to estimate log-transformed variance of residuals (LnVar). Then the variance components for LnVar were estimated with the model:

$$y_v = X_v b_v + Z_v a_{dv} + e_v \quad (2)$$

where y_v is a vector of observations on LnVar_TNB or LnVar_bcTNB; X_v and Z_v are known incidence matrices relating observations to fixed or random effects; b_v is a vector of fixed effects of year and season of the first farrowing; a_{dv} is a vector of dam additive genetic effect, with $a_{dv} \sim N(0, A\sigma^2_{adv})$ and e_v is a vector of residuals, with $e_v \sim N(0,I_{ev}\sigma^2_{ev})$. In this model variance was assumed heterogenous.

Genome-wide association study. The EBVs obtained for LnVar's were deregressed following Garrick *et al.* (2009) and used in GWAS. In total, ~5,000 sows and boars genotyped with Axiom porcine 660K array from Affymetrix (Affymetrix Inc., Santa Clara, CA, USA) or imputed to 660K using Fimpute v2.2 (Sargolzaei *et al.* 2014) were available. A single-SNP GWAS was performed with the following linear animal model in GCTA (Yang *et al.*, 2011):

$$y_k^* = \mu + X^*\hat{\beta} + u_k^* + e_k^* \quad (3)$$

where y_k^* is the deEBV of the k^{th} animal for LnVarTNB or varTNB; μ is the average of the deEBV; X^* is the genotype (0, 1, 2) of the k^{th} animal for the evaluated SNP; $\hat{\beta}$ is the unknown allele substitution effect of the evaluated SNP; u_k^* is the random additive genetic effect, assumed to be distributed as $\sim N(\mathbf{0}, \mathbf{G}\sigma^{2*}_a$, which accounts for the (co)variances between animals though the formation of a genomic numerator relationship matrix (G), constructed using the imputed 660K genotypes, σ^{2*}_a is the additive genetic variance; and e_k^* is the random residual effect, which was assumed to be distributed as $\sim N(\mathbf{0}, \mathbf{I}\sigma^{2*}_e)$.

Results

The significant difference in skewness, was observed after data transformation with the change from -0.469 to -0.02. Table 1 shows the variance components and heritability for litter size (TNB), whereas Table 2 for log-transformed variance (LnVar) of litter size before and after Box-Cox transformation (bcTNB). The data transformation did affect the variance components estimates for litter size and its variability. Heritability for TNB was 0.118 vs 0.125 for bcTNB, whereas for LnVar_TNB it was 0.0025 vs 0.0055 for LnVar_bcTNB. This indicates a larger proportion of additive genetic variance in entire variance based on transformed data. The change in magnitude of genetic variance was even more apparent when comparing GCV_{SDe} for LnVar's without and with transformation (2% vs 4%).

Figure 2 presents preliminary GWAS results for LnVar's that did not include in the litter size models variance components coming from the sire and before transformation of the data.

Discussion

As indicated by previous studies (Rönnegård and Valdar, 2011; Yang *et al.*, 2011; Khaw *et al.*, 2016; Iung *et al.*, 2018) the transformation of the data can have a substantial effect on the phenotypic variability of the trait of interest. This was confirmed here as the estimate of GCV_{SDe} for LnVar's changed from 0.02 to 0.044 after transformation. Moreover, the genetic correlation between mean TNB and LnVar's increased from 0.38 to 0.64 after data transformation (analysis not shown), which is surprising as usually Box-Cox transformation leads to decrease of correlation between mean and variability of a trait (Yang *et al.*, 2011; Iung *et al.*, 2018). Still the genetic correlations between LnVar's was 0.96 (analysis not shown). Data transformation also affect TNB variance components. Again the higher heritability was observed for bcTNB than TNB. This confirms that in case of skewed data of Landrace pigs, the transformation is necessary to reduce variation caused by distribution of the data.

Table 1. Variance components (with SE) for litter size (TNB) without or with Box-Cox transformation (bc) in Landrace sows.

Estimate	TNB	bcTNB
Additive genetic variance	1.13 (0.08)	21.33 (1.45)
Sire additive genetic variance	0.16 (0.08)	2.66 (1.37)
Permanent sow variance	1.11 (0.06)	19.92 (1.11)
Service sire variance	0.31 (0.08)	4.67 (1.39)
Residual variance	8.23 (0.05)	192.12 (1.47)
Heritability	0.118 (0.007)	0.125 (0.007)

Table 2. Variance components (with SE) for log-transformed variance (LnVar) of litter size (TNB) without or with Box-Cox transformation (bc) in Landrace sows.

Estimate	LnVar_TNB	LnVar_bcTNB
Additive genetic variance	0.005 (0.002)	0.008 (0.002)
Residual variance	0.32 (0.004)	0.30 (0.004)
GCV_{SDe} [1]	0.020	0.044
Heritability	0.0025 (0.0007)	0.0037 (0.0008)

[1] Genetic coefficient of variation at residual standard deviation level: $GCV_{SDe} \approx ½\sigma_{av}$.

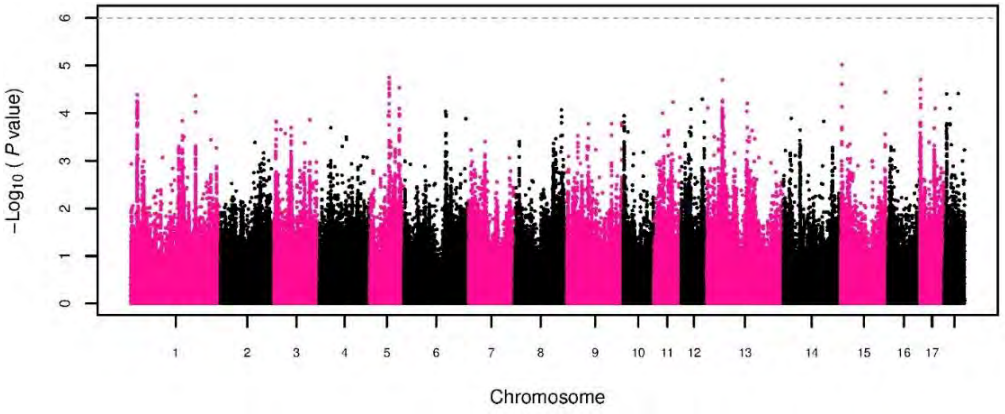


Figure 2. Preliminary GWAS results for the variability of litter size in Landrace pigs.

Previous studies focused on genomic background of litter size variability did not include data transformation since a Normal distribution of the raw data was indicated (Sell-Kubiak *et al.*, 2015; 2022). Moreover, variance components estimation for litter size in Landrace pigs included an additive genetic sire and non-genetic service sire components. This was not necessary e.g. in Large White pigs (Dobrzański *et al.*, 2019), where those variances were close to 0. Thus litter size data coming from Landrace pigs present substantially different properties than those of different breeds.

Our initial analysis did not include variance components coming from the sire (as genetic or non-genetic effect). Thus GWAS results present at the current stage need to be repeated using the deEBVs for LnVar's coming from models with the best fit to the litter size data and also separately for the data after Box-Cox transformation.

Conclusions

The results of this study indicate a substantial change in variability estimates when Box-Cox transformation is applied on data presenting skewness. Further analysis focused of genomic background of LnVar's will bring more insight into those differences.

Funding

This project was financed by the Polish National Science Centre (NCN SONATA grant no. 2016/23/D/NZ9/00029; ESK is the grant holder).

References

Dobrzański J, Mulder H.A., Knol E.F., Szwaczkowski T., and Sell-Kubiak E. (2020) J Anim Breed Genet 137:559–70. https://doi.org/10.1111/jbg.12465

Garrick D.J., Taylor J.F., and Fernando R.L. (2009) GSE 41:55. https://doi.org/10.1186/1297-9686-41-55

Gilmour A.R., Gogel B.J., Cullis B.R., Welham S.J., and Thompson R.. (2015) ASReml User Guide Release 4.1 Functional Specification, VSN International Ltd, Hemel Hempstead.

Guo X, Su G., Christensen O.F., Janss L., and Lund M.S. (2016) BMC Genomics 17(1):1-12. https://doi.org/10.1186/s12864-016-2806-z

Hill W.G. and Mulder H.A. (2010) Genet. Res. (Camb) 92:381–395. https://doi.org/10.1017/S0016672310000546

Iung L.H., Mulder H.A., Neves H.H, and Carvalheiro R. (2018) BMC Genomics 19:1–13. https://doi.org/10.1186/s12864-018-5003-4

Khaw H.L., Ponzoni R.W., Yee H.Y., Aziz M.A., Mulder H.A *et al.* (2016) Aquaculture 451:113–120. https://doi.org/10.1016/j.aquaculture.2015.09.003

R Development Core Team, 2011. R: A language and environment for statistical computing R Foundation for Statistical Computing, Vienna, Austria (http://www.R-project.org).

Rempel L.A., Nonneman D.J., Wise T.H., Erkens T. Peelman L.J. *et al.* (2010) J Anim Sci 88(1):1-15. https://doi.org/10.2527/jas.2009-1985

Rönnegård L., and Valdar W. (2011) Genetics 188:435–447. https://doi.org/10.1534/genetics.111.127068

Samorè, A.B., and Fontanesi, L. (2016) It J Anim Sci 15(2):211-232. https://doi.org/10.1080/1828051X.2016.1172034

Sargolzaei M., Chesnais J.P., and Schenkel F.S. (2014) BMC Genomics 15:478. https://doi.org/10.1186/1471-2164-15-478

Sell-Kubiak E., Duijvesteijn N., Lopes M.S, Janss L.L.G., Knol E.F. *et al.* (2015) BMC Genomics 16:1049. https://doi.org/10.1186/s12864-015-2273-y

Sell-Kubiak E., Knol E.F., and Lopes M.S. (2022) GSE 54:1. https://doi.org/10.1186/s12711-021-00692-5

Yang J., Lee S.H., Goddard M.E., and Visscher P.M. (2011) Am J Hum Genet 88:76–82. https://doi.org/10.1016/j.ajhg.2010.11.011

Yang Y., Christensen O.F., and Sorensen D. (2011) Genet Res (Camb) 93:33–46. https://doi.org/10.1017/S0016672310000418

Zhang Z., Chen Z., Ye S., He Y., Huang S. *et al.* (2019. Animals 9(10):732. https://doi.org/10.3390/ani9100732

791. Genetic parameters of feeding behaviour and feed efficiency traits in Pietrain

P. Núñez[1*], S. Gol[2], J. Reixach[2] and N. Ibáñez-Escriche[1]

[1]*Institute for Animal Science and Technology, Universitat Politècnica de València, 46022 València, Spain;* [2]*Selección Batallé S.A., 17421 Riudarenes, Spain; pnuez@posgrado.upv.es*

Abstract

Precision Livestock Farming technologies applied to pig production has enabled to measure individual consumptions and feeding behaviour. The objective of this study was to estimate the heritability of feeding behaviour and feed efficiency traits at different periods and their genetic correlations in a Pietrain sire line. The database contained records of 933 pigs. The feeding behavioural and feed efficiency traits analysed showed high heritability (from 0.35 to 0.56) except for daily feed intake (0.26) and average daily gain (0.10). The genetic correlation of feeding behaviour traits between periods was close to one. Also, relevant correlations between traits were found, such as feeding rate with average daily gain (-0.78) and with feed conversion (0.47). These partial results indicated that including feeding behavioural traits in the breeding programme could be beneficial. However, more accurate correlations are expected as more data from the test station will be added.

Introduction

Pork meat is one the most widely consumed and its production has been steadily increasing. Feed cost represents the major proportion of the total costs of pig production, so improving feed efficiency traits is key to obtaining acceptable profitability. Feed conversion rate (FCR) and residual feed intake (RFI) are the most widely used efficiency traits in selection indexes, but they are difficult to measure, while average daily gain (ADG) is easier to measure but does not consider the feed consumed. The development of precision livestock farming technologies has made possible to monitor individual weight and feeding behaviour, such as consumption, duration and frequency of visits. These behavioural traits can be measured early in life, and if they are correlated with feed efficiency traits can be used as an early selection criterion (Kavlak and Uimari, 2019). For instance, pigs with unfavourable feeding behaviour at the early stages of the test could be removed and replaced by other test pigs, and it could reduce costs considerably. The aim of this study was to estimate the heritability of feeding behaviour and the feed efficiency traits at different periods and their genetic correlations in a Pietrain sire line.

Materials & methods

Field data were obtained from the test station of the Pietrain selection program of Selección Batallé company (Girona, Spain) from June 2019 until April 2021. The pigs started the test with an initial body weight (IBW) of 53.6±7.6 kg at the age of 96.8±10.0 days and ended with 111.4±11.6 kg at the age of 166.0±7.6 days. Pigs were fed *ad libitum* on a high protein diet, and the feedings were recorded automatically with Pig Performing Test (PPT) stations (Nedap, Groenlo, the Netherlands) along with data of feed intake, duration, and body weight for each visit. Each pen was equipped with a single-space feeder, and the mean of pen occupation was 12.8±2.35 pigs per pen. The raw data included 523,195 observations of 933 animals. Data of feed intake equal to zero or animals with less than 95 days of age at the start of the test or less than 50 total days were removed. After filtering, 459,519 observations from 790 males and 32 females were grouped by animal to compute: total number of visits, the median number of visits per day (NVD), time spent in feeding per day (TPD), daily feed intake (DFI), median of time spent per visit (TPV), median feed intake per visit (FPV), median feeding rate per visit (FR), average daily gain (ADG), feed conversion rate (FCR) and residual feed intake (RFI). FR (g/min) was calculated as the median of the ratio between FPV and TPV.

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_791, © P. Núñez *et al.* 2022

The RFI was calculated as the residual term of the adjusted DFI (amount of feed consumed during the whole period, divided by the number of days):

$$DFI_{ijk} = sex_i + pb_j + b_1(IBW)_{ijk} + b_2(ADG)_{ijk} + b_3(BF)_{ijk} + b_4(LT)_{ijk} + e_{ijk}, \quad (1)$$

where sex_i is the sex effect (i= male, female), pb_j is the interaction between pen and slaughter batch (j= 1, …, 69), b_1, b_2, b_3, b_4 are partial regression coefficients of IBW, ADG, BF (Back-fat thickness) and LT (Loin muscle depth), respectively. BF and LT were measured at the end of the test with a PIGLOG 105® (Frontmatec, Kolding, Denmark), at the last rib level and 6.5 cm off the midline.

Bayesian univariate and bivariate models were performed for all studied traits and the same effects were considered in all models: sex (2 levels), pb (69 levels), common litter effects (346 levels) and the additive genetic effect (822 levels). The models assumed multinormal distributions for the common litter (l), additive genetic (a) and residual effects (e): $l \mid L_0 \sim N(0, L_0 \otimes I)$, $e \mid R_0 \sim N(0, R_0 \otimes I)$ and $a \mid G_0 \sim N(0, G_0 \otimes A)$, respectively, where I represents identity matrices, A is the numerator relationship matrix, $\otimes$ denotes Kronecker product, L_0, G_0 and R_0 are covariance matrices of each effect of one dimension for the univariate and two dimensions for the bivariate models. Bounded flat priors were considered for the systematic effects (sex and pb) and co(variances). The marginal posterior distributions of all unknown parameters were obtained using Gibbs Sampling with the programme gibbs1f90 (Misztal *et al.*, 2002). Chains of 1,000,000 samples were run, and the first 50,000 iterations were discarded. Then a sample was obtained every 100 iterations. The heritability and corresponding variance components for behavioural and efficiency traits were computed for the whole period and for two periods groups according to the phenotypic differences found between the 96-132 days of age (n=819) and 132-182 days of age (n=812).

Results

The heritability and HPD 95% for behavioural and production traits are listed in Figure 1. The posterior means of the heritability were higher for the first period than for the second period for NVD (0.48 to 0.35), DFI (0.32 to 0.26), TPV (0.52 to 0.37), FPV (0.59 to 0.38), FR (0.50 to 0.33); the only exception was TPD which increased from 0.41 to 0.52. The genetic correlations of each behavioural trait between the whole and first period were high with a range of 0.88 to 0.99. The posterior mean of the heritability for FCR, RFI, BF and LT were moderately high (0.35, 0.43, 0.43 and 0.41, respectively) and low for ADG (0.10). The heritability for the feed efficiency traits within period did not properly converge.

Genetic correlations between feed efficiency and behaviour traits are shown in Table 1. The most relevant genetic correlations (HPD95% not included zero) were found between NVD and FCR (-0.44), TPD and ADG (0.74), TPD and BF (0.37), DFI and FCR (0.81), DFI and RFI (0.78), DFI and BF (0.43), TPV and BF (0.39) and FR with ADG (-0.78). The genetic correlations within feeding behaviours traits were also calculated. NVD was inversely correlated with TPV, FPV and FR, their respective posterior mean values and HPD 95% (between brackets) were -0.64 (-0.85, -0.4), -0.92 (-1, -0.83) and -0.42 (-0.78, -0.08). On the contrary, FPV was positively correlated with TPV and FR, with values 0.68 (0.46, 0.87) and 0.46 (0.11, 0.75). For feed efficiency traits, FCR showed two important positive correlations with RFI, and BF with values of 0.70 (0.37, 0.96) and 0.53 (0.09, 0.99), respectively. The genetic correlation between ADG and FCR was not relevant with a value of -0.15 (-0.99, 0.67). All genetic correlations between traits exhibited a wide HPD95%.

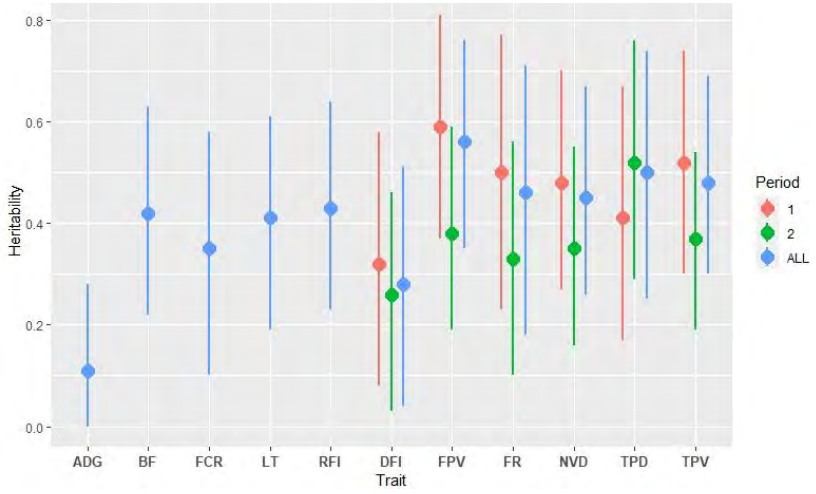


Figure 1. Posterior mean of the heritability and its HPD 95% for feed efficiency and behaviour traits. (Average daily gain (ADG), backfat thickness (BF), daily feed intake (DFI), feed conversion rate (FCR), feed intake per visit (FPV), feeding rate (FR), loin muscle depth (LT), number of visits per day (NVD), residual feed intake (RFI), time spent in feeding per day (TPD), time spent in feeding per visit (TPV).

Table 1. Posterior mean of genetic correlations [HPD95%] between feed efficiency and behaviour traits.

Trait[1]	FET[2]				
	ADG	**FCR**	**RFI**	**BF**	**LT**
NVD	0.33	-0.44	0.13	-0.19	0.07
	[-0.5, 0.9]	[-1.0, -0.0]	[-0.3, 0.5]	[-0.6, 0.2]	[-0.3, 0.4]
TPD	0.74	-0.03	0.16	0.37	-0.01
	[0.2, 1.0]	[-0.5, 0.5]	[-0.3, 0.6]	[0, 0.7]	[-0.5, 0.4]
DFI	0.31	0.81	0.78	0.43	-0.07
	[-1, 0.9]	[0.4, 1]	[0.5, 1]	[0, 0.8]	[-0.7, 0.5]
TPV	0.29	0.25	-0.07	0.39	0.08
	[-0.5, 0.9]	[-0.2, 0.7]	[-0.5, 0.4]	[0, 0.8]	[-0.3, 0.4]
FPV	-0.25	0.5	0.04	0.26	0.04
	[-1, 0.5]	[-0.1, 0.9]	[-0.4, 0.5]	[-0.1, 0.6]	[-0.3, 0.4]
FR	-0.78	0.47	0.23	-0.23	0.06
	[-1, -0.2]	[0, 0.9]	[-0.3, 0.7]	[-0.7, 0.2]	[-0.4, 0.5]

[1] Feeding behaviour traits (NVD=number of visits per day, TPD=time spent in feeding per day, DFI=daily feed intake, TPV= time spent in feeding per visit, FPV=median of feed intake per visit, FR=feeding rate).

[2] Feed efficiency traits (ADG=Average daily gain, FCR=Feed conversion rate, RFI= residual feed intake, BF=Backfat thickness, LT= loin muscle depth).

Discussion

The heritability of behaviour traits found in this study ranged from 0.28 (DFI) to 0.56 (FPV). These values were slightly higher than those obtained by Chen *et al.* (2010) in Duroc with estimates ranging from 0.18 (TPD) to 0.42 (FR), and similar to Kavlak and Uimari (2019) in Finnish Yorkshire with estimates ranging from 0.17 (DFI) to 0.47 (TPV). Nevertheless, there is a great variability of the heritability reported for

behaviour traits (Hall *et al*., 1999). The heritability of feed efficiency traits was similar to other studies except for ADG (0.10) which was lower than those obtained by Herrera-Caceres *et al.* (2019) and Kavlak *et al.* (2021) (0.19 and 0.29, respectively). However, there were also differences in the number of days of recording data. In our study were 65 days, whereas in Herrera-Caceres *et al.* (2019) and Kavlak *et al.* (2021) studies were 107 and 95 days. In contrast, the heritability for ADG was similar than in Chen *et al.* (2010) who obtained 0.21, 0.11 and 0.10 in three testing periods with 64 days of total recording. As expected, negative correlations were obtained for NVD with TPV, FPV and FR, (-0.64, -0.92 and -0.42, respectively) in agreement with Do *et al.* (2013) and Santiago *et al.* (2021). DFI was strongly positively correlated with FCR and RFI and moderately with BF (0.81, 0.78 and 0.43, respectively). These results are in line with those reported by Kavlak and Uimari (2019). A positive correlation was found for TPD with ADG (0.74) and BF (0.37), agreeing with Do *et al.* (2013) in the direction but with a different magnitude. In addition, a negative correlation was found between NVD and FCR (-0.44), in contrast to those positive correlations observed in Landrace, Yorkshire and Duroc by Do *et al.* (2013) and Duroc by Santiago *et al.* (2021) (0.35 and 0.34, respectively). These earlier observations suggest that Pietrain pigs with more NVD, have lower FR, TPV and FPV. At the same time, higher NVD may result in lower FCR. Considering the moderate to high heritability of NVD (0.45) and its strong correlation shown between the first period and the whole period (0.97), it could be useful to incorporate NVD in selection criteria to improve the accuracy of selection for FCR in Pietrain pigs. However, all the correlations between behavioural and feed efficiency traits in this study have a broad 95% HPD and will become more accurate as more data from the test station will be added.

References

Chen, C. Y., Misztal, I., Tsuruta, S., Zumbach, B., *et al.* (2010). J. Anim. Breed. Genet., 127, 230– 234. https://doi.org/10.1111/j.1439-0388.2009.00833.x.

Do, D.N., Strathe, A.B., Jensen, J., Mark, T., Kadarmideen, H.N. 2013. J. Anim. Sci. 91, 4069–4079. https://doi.org/10.2527/jas2012-6197.

Hall, A., Hill, W., Bampton, P., & Webb, A. (1999). Animal Science, 68, 43–48. https://doi.org/10.1017/S1357 72980 0050062.

Herrera-Cáceres, W., Ragab, M., Sánchez, J., 2019. Animal 14, 233–242. https://doi.org/10.1017/S1751731119002179.

Kavlak, A.T., Uimari,P., 2019. J. Anim. Breed. Genet. 136,484–494. https://doi.org/10.1111/jbg.12408.

Kavlak, A.T., Stranden, I.,Uimari, P., Lidauer, M.H., 2021. Animal 15 (2021), Issue 3, 100168. https://doi.org/10.1016/j.animal.2020.100168.

Misztal I, Tsuruta S, Strabel T, Auvray B, *et al.* (2002). Proc. of the 7th WCGALP, Montpellier, France, p. 28.

Santiago, K.G.; Kim, S.H.; Lopez, B.I.; Lee, D.H., *et al.* Agriculture 2021, 11, 850. https://doi.org/10.3390/agriculture11090850.

Schulze, V., Roehe, R., Bermejo, J. L., Looft, H., *et al.* (2003). Livest. Prod. Sci., 82, 105–116. https://doi.org/10.1016/S0301-6226(03)00034-4

792. Genetic analysis of hoof colour in Duroc pigs

J. Vila, O.O.M. Iheshiulor, M.C.A.M. Bink and A.E. Huisman*

Hendrix Genetics, P.O. Box 114, 5830 AC Boxmeer, the Netherlands; abe.huisman@hendrix-genetics.com

Abstract

Hoof colour is an important characteristic for the Spanish producers of the cured hams and shoulders. The colour of the hoof is one of the quality requirements for high-value product ranges. Within this niche industry the Duroc breed is used as terminal sire line. The blackness of the hooves is used as a selection criterion for Duroc boars entering into Iberian production. To accommodate for this in the Hypor breeding program, the colour of the hooves in the Duroc breeding lines is recorded. Based on the original phenotypes new colour traits were defined. A genetic analysis was performed to estimate variance components and their ratios to investigate the genetic background of the colour of hooves. Heritabilities were in the mid to high ranges (0.06-0.50), in addition a GWAS was performed, resulting in identified peaks on chromosomes 1, 5, and 7. The results presented here are preliminary results.

Introduction

Hoof colour is an important characteristic for the Spanish producers of the cured hams and shoulders. The colour of the hoof is one of the quality requirements for high-value product ranges. Within this niche industry the Duroc breed is used as terminal sire line to produce final products which are a cross between the Duroc breed and the Iberian breeds. According to Iberian producers, the most important characteristic is colour homogeneity, and colour intensity is less relevant. The blackness of the hooves is used as a selection criterion for Duroc boars entering Iberian production. Scientific reporting on inheritance of colour is scarce and dated, but in a series of seven papers Hetzer described colour inheritance for crosses between a variety of breeds (Hetzer, (1945-1948), reviewed by Legault, 1998). Remarkably, there was no mentioning of hoof colour in these papers.

Different classes of Iberian hams are produced, varying in degrees of Iberian breed present in the final product. The other breed making up the final product is the Duroc breed. Since 2014 the Spanish government is tagging products based on these specifications (http://www.aeceriber.es/). Although no direct relationship between hoof colour and ham quality has been reported, in the production of Iberian ham special emphasis is put on the colour of the hoofs attached to the hams. Black hoofs stand for quality, deviations from this black colour result in a perceived lower quality of the ham. Therefore, a requirement for the Duroc sires used to produce Iberian hams is the presence of black hooves. The purpose of this study is to investigate the genetic background of hoof colour in purebred Duroc animals.

Materials & methods

Phenotypes for hoof colour were collected at one of the Hendrix Genetics, Hypor nucleus facilities. The colour of the hoofs was assessed at birth for Duroc sire line. The phenotypes were defined as percentages black, completely black: 100%, the absence of black: 0%. For ease of registration, the percentages black does not cover the whole range of 0-100%, but five classes: 100, 75, 50, 25 and 0% (Figure 1). The scoring was done by a trained technician. To note, 0% does not mean completely white, it means that the black part is less than 25%, and for example 50% mean greater than 50% but less than 75%.

The original classification levels were arbitrarily chosen, and we apply a re-classification, by merging various levels, into fewer levels as this may further our understanding of the genetic background of the trait (Table 1).

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_792, © J. Vila *et al.* 2022

Figure 1. Images on hoof colour that illustrate the original five classification levels.

Table 1. Number of observations per class[1]

Trait[1]	0%	25%	50%	75%	100%	Levels
Original score	4,863	2,793	1,213	459	115	5
White versus rest	4,863	4,580				2
Black versus rest	9,328				115	2
Extremes narrow	4,863	4,465			115	3
Extremes wide	7,656		1,213	574		3

[1] Cells in grey represent merged classes.

The model used for the analysis of the defined traits was the same for all five traits. After testing several fixed and random effects, the final model:

$$y = Xb + Za + Wc + e$$

where y is vector of observations, b is vector of fixed effects, including breed, sex, year-quarter of birth, a is vector of random animal effects, c is vector of common environmental effects, e is vector of random residual effects, and X, Z, and W are incidence matrices relating records to fixed, animal and common environmental effects, respectively. It is assumed that the common environmental effects and the residual effects are independently distributed with means of zero and variance σ_c^2 and σ_e^2, respectively; therefore var(c) = $I\sigma_c^2$; var(e) = $I\sigma_e^2$; and var(a) = $A\sigma_a^2$.

The number of observations in total was 9,443; the number of observations per class is given in Table 1. Pedigree was traced back for 10 generations and included 13,656 animals. The number of observations per trait and class combination is presented in Table 1. Obviously, the number of observations reduces with the level of blackness of the hooves, i.e. black hooves are the least prevalent phenotype in the dataset.

Single trait analyses were performed, using the airemlf90 program out of the BLUPF90 suite of programs (Misztal *et al.* 2014), to estimate variance components for the different traits. While traits defined on a categorical scale, the models were analysed assuming a normal distribution. Subsequently a GWAS was performed using only data from 1,806 genotyped animals. The GWAS was performed using the GCTA software (Yang *et al.* 2014). A series of quality control steps (missing SNP, MAF and Hardy-Weinberg equilibrium) resulted in a genotype of 31,111 SNP available for 1,806 animals.

Results

The estimated variance components and their ratios are presented in Table 2. For trait Black versus the rest the results are unreliable, most likely due to the limited number of observations in the 100% class. Heritabilities for the different traits are in the mid to high ranges and vary from 0.322 for Extremes wide to 0.502 for Original score, except for trait 'Black versus the rest'.

Figure 2 shows the GWAS results for the different traits. For the original score 3 peaks were identified, located on chromosomes 1, 5 and 7 (Figure 2A). Not for all trait definitions peaks were found, i.e. for black versus rest (Figure 2C) no significant peaks were detected. Whereas for other trait definitions peaks appear to be in line with the peaks found for the original score.

Table 2. Phenotypic variance (σ^2) and estimated heritability (h^2) and common environment (c^2) for the five hoof colour traits[1]

Trait	σ^2	h^2	c^2
Original score	0.87	0.50	0.17
White versus rest	0.24	0.47	0.07
Black versus rest	0.01	0.06	0.10
Extremes narrow	0.27	0.46	0.08
Extremes wide	0.31	0.32	0.21

[1] Standard errors on h^2 ranged from 0.02 to 0.04; standard errors on c^2 from 0.01 to 0.02.

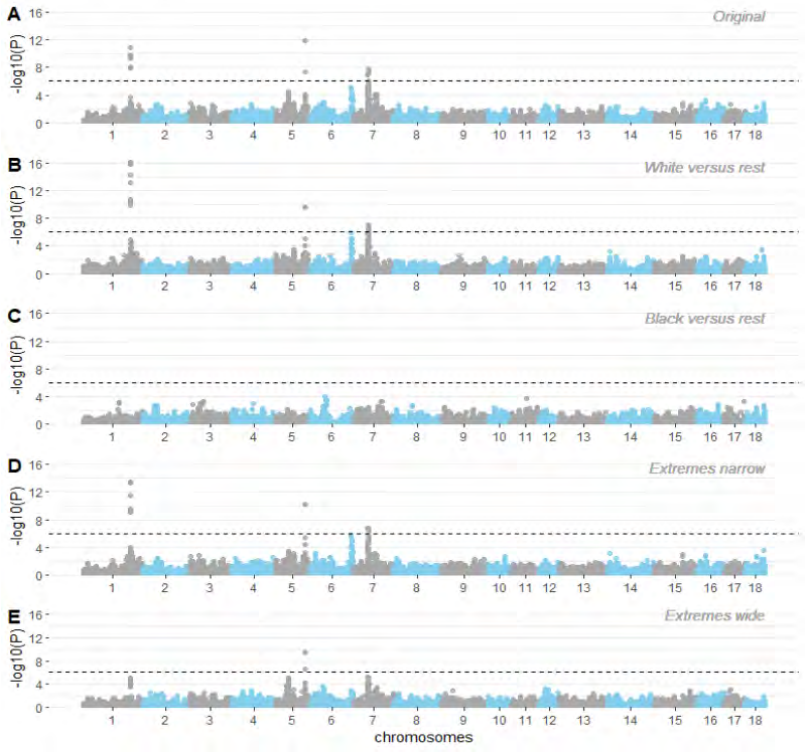


Figure 2. Manhattan plots on the GWAS results for the five hoof colour traits.

Discussion

The trait analysed has been recorded as a categorical trait, whereas there is a range in blackness of the hooves. The phenotypes have been scored visually, a possible better way of scoring could be the use of cameras or other visual tools. Given the nature of the trait definition a threshold model might have been the more appropriate choice of model. However, for ease of analysis and given the stage of this research we opted for the mixed linear model. Running a threshold model for the original score resulted in much higher heritability (0.68) whereas the litter component is similar (0.14). Other colour phenotypes such as coat colour and eye colour seem to be polygenic, even though there are a few genes with very large effects (KIT, MC1R, etc). Surprisingly, that's not what we found for hoof colour. The peak explaining the largest amount of variance was like ~5%. It seems to be rather polygenic. Incomplete penetrance could also play a role in the inheritance of hoof colour.

Next steps in the analysis will include investigating different modelling options and collecting more phenotypes, also at different ages to see how this trait evolves over time, i.e. is hoof colour seen at birth representative for hoof colour at older ages.

References

Hetzer, H.O. (1945a). Journal of Heredity 36(4): 121-128.

Hetzer, H.O. (1945b). Journal of Heredity 36(6): 187-192.

Hetzer, H.O. (1945c). Journal of Heredity 36(8): 255-256.

Hetzer, H.O. (1945d). Journal of Heredity 36(10): 309-312.

Hetzer, H.O. (1946). Journal of Heredity 37(7): 217-224.

Hetzer, H.O. (1947). Journal of Heredity 38(4): 121-124.

Hetzer, H.O. (1948). Journal of Heredity 39(4): 123-128.

Legault, C. (1998) Genetics of colour variation. In: Rothschild, M.F. and Ruvinsky, A. (eds) The Genetics of the Pig. CABI, Wallingford, UK, pp. 51–69.

Misztal, I., Tsuruta S., Lourenco D.A.L., Aguilar I., Legarra A., *et al.* (2014). Manual for BLUPF90 family of programs. Available at: http://nce.ads.uga.edu/wiki/lib/exe/fetch.php?media=blupf90_all2.pdf

Yang J., Zaitlen N.A., Goddard M.E., Visscher P.M., and Price A.L. (2014) Nat. Genet. 46(2):100-106

793. Filling information gaps in swine crossbreeding schemes by imputing non-genotyped F_1 animals to improve genetic gain

G.M. See[1], J.S. Fix[2], C.R. Schwab[2] and M.L. Spangler[1*]

[1]University of Nebraska-Lincoln, 1400 R St, Lincoln NE 68588, USA; [2]Acuity, 7475 State Road 127, Carlyle, IL 62231, USA; mspangler2@unl.edu

Abstract

Economically relevant traits are often not included in genetic evaluations because the animals that express them lack pedigree ties to selection candidates. Simulations produced three purebred (PB) breeds (A, B and C), an F_1 and a 3-way crossbred (CB). Breeds A, B and F_1 expressed lowly heritable traits with moderate genetic correlations between them. Selection was practiced for four generations with the goal of improving F_1 performance. Genotypes of F_1 dams were imputed using FImpute. Scenarios differed by number of CB genotyped per generation and knowledge of F_1 mates. When mates were known and >3 progeny were genotyped per generation, the phenotypic response in F_1 did not differ (P>0.05) from when the true genotypes were used. Un-genotyped F_1 animals can be imputed such that their phenotypes can be used in genetic evaluations but the benefit depends upon the number of CB progeny genotyped, parity and completeness of CB pedigree.

Introduction

With advancements in genotyping technologies, swine breeders are limited not by genotypes but by phenotypes which represent goal traits of the breeding program. Goal traits in swine breeding programs are often observed in animals outside of the nucleus and have a genetic correlation to nucleus traits which is often below unity. Previous reports have investigated CB genotyping and phenotyping strategies to improve genetic gain in PB selection candidates for CB performance with favorable results (Duenk *et al.*, 2021). Since the structure of most swine breeding programs consists of large full- and half-sib families (Badke *et al.*, 2012), a large degree linkage disequilibrium (LD) exists which can be exploited to impute genotypes of non-genotyped animals. Previous reports show that non-genotyped animals can have genotypes accurately imputed using greater than 10 genotyped progeny (Pimentel *et al.*, 2013; Berry *et al.*, 2018). Objectives of the present study were: (1) evaluate the accuracy of imputed genotypes of non-genotyped F_1 dams of CB animals; and (2) to quantify the response to selection for F_1 performance using imputed genotypes.

Materials & methods

Simulation and evaluation. Data were simulated using the AlphaSimR package (Gaynor *et al.*, 2019) in R. The genome consisted of three chromosomes each 1 CM in length with 11,100 evenly spaced bi-allelic markers. The historical population began with an effective population size of 12,000 which was gradually reduced to 250. Following historical founder development (n=1,500), founders were randomly assigned with equal numbers of males and females to three PB breeds (hereafter breeds A, B and C). Recent LD and PB breed differentiation was achieved by random mating during a genetic bottleneck (10 generations), with constant population size (30 generations), and followed by population expansion (10 generations). Males and females (25 and 425, respectively) were chosen to be breed founders. Dams produced one litter per generation and PB and CB matings produced 8 and 10 progeny, respectively. The 3-way crossbreeding program was achieved by crossing 25 sires from A with 400 dams from B to produce AB females. CB progeny were simulated by crossing AB females (n=1000) with sires from C (n=50).

A subset (n≈9,000) of markers was chosen to be included on the SNP chip used for imputation and genetic evaluations. Three phenotypes were simulated by randomly selecting 300 markers, which were not SNP,

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_793, © G.M. See *et al.* 2022

to be quantitative trait loci which acted additively. Simulated variance components for breeds A, B and AB were chosen to mimic reproductive traits in swine (h^2=0.2, 0.2 and 0.15, respectively) which were moderately correlated to each other (r_a=0.65). The only female animals which produced offspring expressed a phenotype and the phenotype was considered as expressed only at the first parity.

Selection and genetic evaluation. The goal of the selection scheme was to improve performance in AB by using estimated breeding values (EBV) produced from either single- or two-trait animal models. Prior to the simulation of CB animals, candidates from breeds A and B were chosen by EBV produced from single-trait models for within breed performance. After CB animals were produced, selection candidates were chosen by EBV for AB performance. EBV were produced by two-trait models including breeds A and AB, or breeds B and AB.

Single- and two-trait single-step genomic best linear unbiased prediction models were fitted using the BLUPf90 application programs. Single- (PB performance) and two-trait (PB and AB performance) models contained effects of the overall mean, generation (class effect), additive genetic (random) and residual (random). Variance components were those initially simulated.

Imputation strategies. Imputation of non-genotyped AB dams was done using FImpute V3 using both family- and population-based imputation (Sargozaei *et al.*, 2014). Genotypes (MAF>0.05) used to impute AB dams were from breeds A, B, C and CB progeny. Pedigree information used for imputation included relationships within PB breeds and AB dam relationships to CB progeny. Sire relationships with CB progeny were considered to be either known (knownSire) or unknown (unknownSire). The number of genotyped CB progeny per generation (2, 3, 4 or 6) was also investigated. Given the minimum number of progeny required to impute non-genotyped parents in FImpute is 4, genotyping 2 or 3 CB progeny per generation resulted in imputed genotypes after two parities resulting in fewer AB records in genetic evaluations. Comparisons across imputation strategies were made using mean genotype concordance (GC), defined as the proportion of correctly imputed genotypes. Response to selection using imputed genotypes was compared to selection when either no genotypes (noGeno) or the true genotypes (trueGeno) of AB animals were used. The cumulative phenotypic response in AB performance (CPR) was calculated per replicate as the difference between mean phenotype of AB dams from the final generation and mean phenotype in the first generation. Least squares (LS) means were then calculated for the CPR using the *emmeans* package in R by fitting a linear model including fixed effects of replicate, sire relationship status and number of genotyped CB. The analysis was then replicated 20 times.

Results

Summary of simulated animals. The degree of LD and distribution of MAF produced in breeds B and C was similar to A. In breed A, pairwise estimates of r^2, averaged across replicates, in the range of genomic distance in kb of 0-25, 50-120, 200-500, and 1000-2,000 was 0.49, 0.41, 0.28, and 0.14, respectively. The correlation of LD phase between PB breeds (Bladke *et al.*, 2012) averaged over pairwise PB breed comparison and replicate, within a range of genomic distance in kb of 0-100, 200-300, 500-600 and 700-800 was 0.72, 0.50, 0.32 and 0.23, respectively. In general, distributions of allele frequencies were near uniform. The number of alleles with extreme frequencies tended to be greater than those with intermediate frequency, but distributions varied by replicate.

Imputation accuracy. The mean GC of imputed AB genotypes in the final generation under knownSire using 2, 3, 4 or 6 CB progeny genotypes per generation was 0.96, 0.98, 0.96 and 0.97, respectively. However, relative to the cumulative number of genotyped CB, GC increased as the number of genotyped CB increased.

Once 9 CB progeny were genotyped little benefit in GC was achieved for an additional genotyped progeny (Figure 1A). In general, parity had little effect on GC under knownSire (Figure 1B).

The average GC of imputed AB genotypes in the final generation under unknownSire using 2, 3, 4 or 6 CB progeny genotypes per generation was 0.80, 0.82, 0.71 and 0.73, respectively. Breeding scheme factors such as the total number of genotyped CB progeny (Figure 1A) and parity (Figure 1B), impacted genotyped concordance using unknownSire. For every additional genotyped CB animal beyond 8, the corresponding gain in GC was negligible. Significant improvement in GC was achieved when genotyped CB progeny were from multiple parities. The average increase in GC when genotyped CB progeny were from 2 or 3 parities compared to one was 0.14 and 0.19, respectively. Parity 3 dams with less than 3 mates and parity 2 dams with one mate comprised 1.05 and 0.58% of all AB matings, respectively.

Response to selection. The CPR, expressed as LS means, is presented in Figure 1C. When noGeno and trueGeno were used the LS means for CPR were 0.32 and 1.26, respectively. As the number of genotyped CB progeny increased CPR also increased. Whether CB sires were included in the pedigree or not impacted the CPR differentially according to the number of genotyped CB progeny used for imputation. The CPR was not different when 2 or 3 CB progeny were genotyped per generation between knownSire and unknownSire, yet the increase in CPR was greater between generations 3 and 4 under knownSire compared to unknown sire. When the number of CB progeny genotyped per generation increased to 4 or 6, knownSire produced a greater CPR compared to unknownSire. CPR when sire was known and 4 or more CB genotypes per generation were used for to impute AB did not differ ($P>0.05$) compared to using trueGeno. However, using as few as 3 genotyped CB per generation, regardless of CB sire status, provided a greater response ($P<0.05$) in CPR than noGeno.

Discussion

The trend in LD phase between breeds in current study was similar to those previously reported in real swine breeds (Badke *et al.*, 2012). Previous reports suggest that imputation accuracy increases as MAF becomes more extreme (Pimentel *et al.*, 2013). Thus, imputation accuracy could be inflated upwards compared to previous reports with normally distributed allele frequencies.

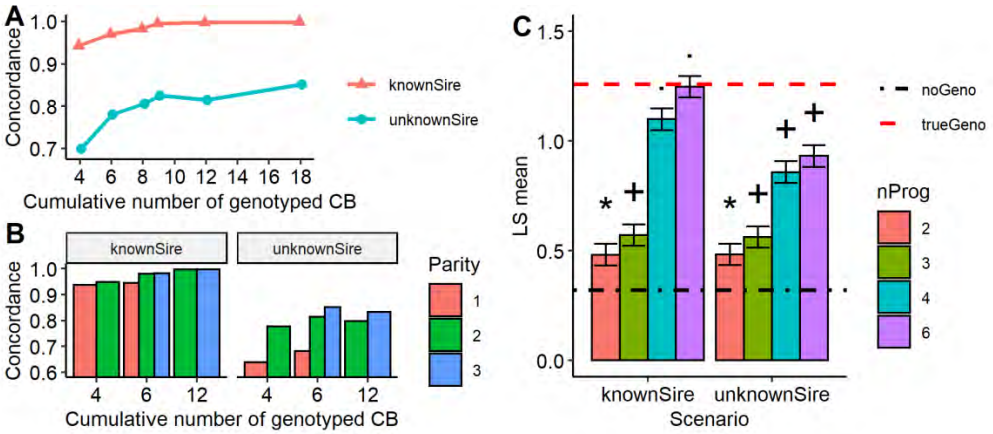


Figure 1. Factors affecting GC and CPR expressed as LS means. (A) Impact of the cumulative number of genotyped progeny on GC. (B) Impact of parity on GC. (C) CPR where * and + denotes significantly different ($P<0.05$) from noGeno, trueGeno and both noGeno and trueGeno, respectively.

In the current study, the simulation was designed to mimic a breeding scheme where CB genotyping decisions were made prior to the imputation of AB dams. Factors related to the design of the breeding scheme, including total number of genotyped progeny and parity of the dam (number of mates), contributed to GC. As expected, as the total number of CB progeny genotyped increased the accuracy of imputation also increased regardless of CB sire pedigree status, in agreement with previous reports (Berry *et al.*, 2018). The relationship between parity and GC has not been previously reported. The increase in GC due to parity is likely driven by the increase in the number of mates per AB dam. In the case where 2 or 3 CB progeny were genotyped per generation, AB dams with imputed genotypes had more parities and benefited from a larger reference population which led to an increase in GC compared to when 4 or 6 CB progeny were genotyped per generation. Increases in GC led to differences in CPR across imputation strategies. When genotyping 4 or 6 CB progeny per generation the CPR was greater than noGeno, regardless of CB sire pedigree inclusion. However, only under knownSire using 4 or 6 CB genotypes per generation was the CPR using imputed genotypes not different from trueGeno. Combined results suggest that imputed AB genotypes can be used to inform genetic evaluations and selection decisions. For entities that genotype CB, this study demonstrates an avenue to improve genetic gain in F_1 performance without the requirement of additional genotyping or pedigree tracking through the use of imputed genotypes.

References

Badke Y.M., Bates R.O., Ernst C.W., Schwab C., and Steibel J.P. (2012) BMC Genomics. 13(24):1-10. https://doi.org/10.1186/1471-2164-13-24

Berry D.P., McHugh N., Randles S., Wall E., McDermott K. *et al.* (2018) Animal. 12(2):191-198. https://doi.org/10.1017/S1751731117001653

Duenk P., Bijma P., Wientjes Y.C., and Calus M.P. (2021) J. Anim. Sci. 99:skab205. https://doi.org/10.1093/jas/skab205

Gaynor R.C.G., Gorjanc G., Wilson E., Money D., and Hickey J.M. (2019) AlphaSimR: Breeding Program Simulations. Available at: https://CRAN.R-project.org/package=AlphaSimR

Pimentel E.C., Wensch-Dorendorf M., König S., and Swalve H.H. (2013) Genet. Sel. Evol. 45:1-12. https://doi.org/10.1186/1297-9686-45-12

Sargolzaei, M., Chesnais J.P., and Schenkel F.S. (2014) BMC Genomics. 15:1-12. https://doi.org/10.1186/1471-2164-15-478

794. Distinctive gene-expression profiles characterise tail-biting precursors in pigs under dietary protein restriction

L. Couteller[1,2], L. Roch[2,3] and C. Kasper[2*]

[1]École Supérieure d'Agricultures, 55 Rue Rabelais, 49007 Angers, France; [2]Agroscope, Tioleyre 4, 1725 Posieux, Switzerland; [3]Animal Welfare Division, Veterinary Public Health Institute, University of Bern, Länggassstrasse 120, 3012 Bern, Switzerland; claudia.kasper@agroscope.admin.ch

Abstract

Tail biting is a common problem in pig production, but genetic studies are impeded by the difficulty of observing it. It is a broader phenomenon that begins before lesions manifest. Therefore, precursors of tail biting, such as oral and nasal manipulation of conspecifics, should be studied. However, the molecular physiology underlying those behaviours is not well understood. Here we show that gene expression profiles in the hypothalamus of 73 non-tail-docked pigs fed a protein-reduced diet differed between individuals showing abnormal and normal behaviours. We found that differences in the expression of genes involved in neurotransmission, notably in dopamine, serotonin and GABA, and G-protein coupled signalling, energy metabolism and appetite. Our results demonstrate that abnormal behaviours have distinct signatures in the brain, which could be partly caused by genomic variation. These insights can lead to a better understanding of the biological mechanisms involved and thus may ultimately inform genomic selection programmes.

Introduction

Damaging behaviours such as tail biting are a common problem in pig production, compromising animal welfare and causing economic losses. A previous study found low to zero heritability for tail biting (Breuer *et al.*, 2003). Genetic studies are impeded by the enormously complex phenotyping, because tail biting often cannot be observed directly. Tail biting is thought to be a broader phenomenon, and behavioural problems caused by various stressors begin before escalating into damaging behaviour where serious injuries occur (Valros, 2018). Therefore, it useful to study the precursors of tail biting, which are behaviours that are termed 'abnormal' such as belly nosing, ear biting and 'tail-in-mouth' (Brunberg *et al.*, 2011) to gain insights into the molecular physiology of tail biting.

Poor sanitary conditions in combination with dietary protein reduction increase the prevalence of tail biting (Van der Meer *et al.*, 2017). Since restricting the protein content in the diet is a promising way to reduce nitrogen emissions in pig manure, it is important to investigate whether this measure can lead to a reduction in animal welfare. Even if essential amino acids are substituted, the overall reduction in dietary protein could lead to an increase of abnormal behaviours due to neurotransmitter system dysfunction because of the reduced availability of amino acids required for neurotransmitter synthesis. Pigs differ in their ability to utilise dietary proteins, which has been shown to be heritable (Kasper *et al.*, 2020). Therefore, we expect that individuals are not equally affected by protein reduction. Here, we investigate whether, under protein-restricted feeding with essential amino-acid substitution, gene expression profiles reflect differences in behaviour. Furthermore, we explore the genes and their functions as well as the pathways that are affected by abnormal behaviours.

Materials & methods

Animals. Ninety-five non-tail-docked pigs were reared on a diet formulated to contain 80% of crude protein of the recommendation. Between day 98 and 115 after birth, all instances of nasal and oral manipulation behaviours towards conspecifics (biting, seizing and nosing), classified as 'abnormal' behaviours, were recorded during five-min observations at four occasions. The protein efficiency of each pig was calculated

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_794, © L. Couteller *et al.* 2022

as the amount of protein in the carcass after slaughter divided by the amount of protein ingested. The effect of protein efficiency and sex on abnormal behaviours was analysed with generalized mixed-effects models with a Poisson and binomial family. The individual ID as well as the sire ID were added as random effects to correct for multiple observations and the family structure.

Brain extractions, RNA preparation and sequencing. The brains of the remaining 94 pigs (one died before slaughter) were extracted around 20:33±01:50 min after the start of exsanguination. The hypothalamus was dissected, immediately frozen in liquid nitrogen and stored at -80 °C until RNA extraction. The brains of the first eight pigs and further six samples were discarded as the training set for the sampling procedure and because of bad sampling quality. RNA extraction and sequencing of 80 hypothalamus samples (from 38 females and 35 males) were carried out by an external provider (Alithea Genomics, Lausanne, Switzerland). RNA-seq was done using the BRB-seq method (Bulk RNA Barcoding and sequencing, Alpern *et al.*, 2019), in which samples are multiplexed after the first reverse-transcription step.

Differential gene expression and network analysis. We removed genes with zero variance, genes with zero counts in more than 90% of samples and genes with a low expression (i.e. less than 10 counts) and samples with sequencing depth of <1M. We applied a multiple regression design, which included manipulation behaviour towards conspecifics, protein efficiency and sex, to control for additional variance, in DESeq2 (Love *et al.*, 2014). *P*-Values and fold-changes of the differential gene expression analysis were computed using the negative binomial GLM fitting and Wald statistics. GO-term enrichment of differentially expressed genes (DEG) was performed in ToppGene Suite (https://toppgene.cchmc.org). Weighted gene co-expression network analysis in WGCNA (Langfelder and Horvath, 2008) was carried out on the filtered read count matrix after variance-stabilizing transformation. To divide the co-expressed genes into modules, we chose a maximum block size of 10,000, a soft power threshold of 10 and 'signed' as the topological overlap matrix type. The modules enriched for DEG related to abnormal behaviour were determined with least-square linear models, an empirical Bayes moderation of standard errors in the R package limma (Ritchie *et al.*, 2015). The same procedure for GO-term enrichment as described above was applied to the genes contained in the significant modules. All *P*-values were corrected with a Benjamini-Hochberg FDR (a=0.05).

Results & discussion

Behavioural observations. Only three pigs were involved in tail biting (3.2%, two biters, one individual was both biter and victim). Even though the majority of pigs performed nasal or oral manipulation behaviour towards conspecifics, 16 pigs (17%) did not show these behaviours in any of the four observations. The number of abnormal behaviours was not influenced by protein efficiency or sex (GLMM with Poisson family: P=0.194 for protein efficiency, P=0.836 for sex), nor was the probability to engage in those behaviours (GLMM with binomial family: P=0.190 for protein efficiency, P=0.672 for sex).

Differential gene expression. Sequencing produced a read count matrix of 31,908 genes by 80 individuals. Filtering resulted in the removal of 13,775 genes. Sequencing depths ranged from 0.56M to 3.57 M reads per sample (1,774,168±649,509 reads). Seven individuals were removed because their sequencing depth was <1M reads. We found 23 genes differentially expressed between behaviourally abnormal and normal pigs ($P \leq 0.05$), of which 17 were up-regulated in abnormally behaving pigs compared to normally behaving ones and six genes were down-regulated (Table 1). In the set of DEG, 23 GO terms were enriched, most notably pertaining to G protein-coupled activities, amino acid binding, neurotransmitter and synaptic vesicle transport. The heterotrimeric G-protein signalling pathway-Gq alpha and Go alpha mediated pathway and the 5HT2 type receptor mediated signalling pathway were also significantly enriched.

Table 1. Significantly differentially expressed genes across behavioural groups of pigs.

Gene	base mean[1]	log2-fold change	SE	*P*-value	adj. *P*-value[2]
TNNT1	24.94	1.80	0.52	6.7E-06	0.019
RGS16	228.38	1.86	0.60	1.5E-05	0.032
PRKCG	133.18	1.01	0.32	3.5E-05	0.041
PRKCH[3]	50.38	1.40	0.48	4.6E-05	0.041
RADIL	49.86	0.57	0.17	3.4E-05	0.041
PLXDC1	112.50	1.53	0.52	3.6E-05	0.041
SPOCK1	242.23	0.55	0.18	4.5E-05	0.041
KITLG	79.64	1.42	0.50	4.6E-05	0.041
NTNG1	177.89	1.12	0.39	7.0E-05	0.043
SIN3B	128.42	0.30	0.10	6.2E-05	0.043
GRM8	27.03	0.67	0.22	7.1E-05	0.043
AK8	21.96	0.78	0.25	6.3E-05	0.043
RIMS3	130.90	0.02	0.05	7.3E-05	0.043
GRM5	120.07	0.46	0.15	7.8E-05	0.044
GABRD	35.91	1.48	0.55	9.1E-05	0.047
PCP4	2,017.17	1.40	0.53	9.3E-05	0.047
NEXN	171.30	1.38	0.56	1.0E-04	0.048
CD44	19.34	-0.96	0.24	2.5E-06	0.019
PMP2	29.05	-1.32	0.34	5.5E-06	0.019
APLNR	47.80	-1.47	0.39	7.3E-06	0.019
DDC	51.37	-1.10	0.31	3.7E-05	0.041
CKK	69.87	-1.09	0.34	7.2E-05	0.043
SLC18A2	66.41	-1.31	0.41	1.0E-04	0.048

[1] Mean expression across samples.

[2] Hochberg-Benjamini adjusted FDR.

[3] Originally unknown gene in the Sscrofa11.1 assembly. A search for orthologs in the Expression Atlas (https://www.ebi.ac.uk/gxa/home) yielded PRKCHA.

Weighted gene co-expression network analysis. The genes were divided into 20 modules, of which only one was significantly enriched for DEG related to manipulation behaviour (logFC=0.11, t=3.08, adj. *P*=0.04). This module contained 30 genes, which were down-regulated in abnormally behaving relative to normally behaving pigs (Figure 1). Only three genes in the module overlapped with the list of significant DEG (*CCK*, *SLC18A2* and *DDC*). Among the genes contained in the module were six transcription factors (*NR2F2*, *FOXA1* and *FOXA2*, *EN1* and *EN2*, and *LMX1B*). Eight genes were involved in neurotransmitter metabolism (*DRD2*, *CHRNB3, CHRNA6, SLC6A3. SLC18A2, DDC*, *ALDH1A1* and *TH*). Three genes coded for proteins involved in the regulation of feed intake and energy metabolism (*RET*, *CCK* and *INS*). The other genes were related to cell migration (*PRPRU* and CD164), G-protein-coupled receptors (*GPR26* and *TRPC6*), dopaminergic neuron differentiation (*RSPO2*), regulating mitochondrial Ca2+ uptake in neurons (*PDZD8*), synaptic vesicle formation (*CPLX2*) and genes with other functions (*SLC10A4*, *ZAR1, LRRC3B, GSTO1, SFRP4* and *MXRA7*). We found 71 GO terms significantly enriched, most notably dopamine and catecholamine binding, neurotransmitter transport, neurotransmitter metabolic processes, synaptic vesicle transport, neuron generation and differentiation, behaviour and response to drugs.

Conclusions

As expected, tail biting was very rare. Classifying animals into groups based on their oral and nasal manipulation behaviour towards conspecifics (i.e. abnormal behaviour) is therefore useful and offers

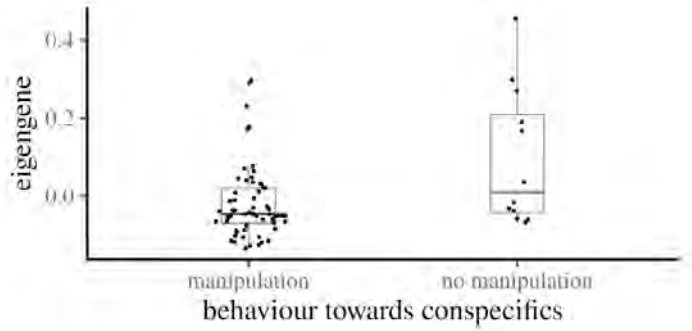


Figure 1. Eigengene values in the module that is significantly enriched for genes related to manipulation behaviour.

interesting insights. Here, we found evidence that abnormal behaviour in pigs is associated with gene expression profiles that are implicated in neuropsychiatric disorders in humans. Genes related to the synthesis, synaptic transmission and uptake of neurotransmitters, notably GABA, glutamate, dopamine, acetylcholine and serotonin, were differentially expressed in pigs showing abnormal behaviour. This was corroborated with the findings from pathway enrichment analysis. Interestingly, some of the DEG have been associated with energy metabolism, feed intake and appetite. Especially in the context of protein-efficiency studies, this warrants further investigation. As of now, it is unclear whether the differences in expression found in this study are caused by genetic variants, DNA methylation or other modifications, or short-term environmental influences. A better knowledge of these causes could inform genomic selection programs and allow the inclusion of welfare in the breeding goals.

References

Alpern, D, Gardeux, V, Russeil, J, Mangeat, B, Meireles-Filho, ACA, *et al.* (2019) - Genome Biol 20:71. https://doi.org/10.1186/s13059-019-1671-x

Breuer, K, Sutcliffe, MEM, Mercer, JT, Rance, KA, Beattie, *et al.* (2003) - Appl Anim Behav Sci 84:59–74. https://doi.org/10.1016/S0168-1591(03)00147-3

Brunberg, E, Jensen, P, Isaksson, A, Keeling, LJ (2013) - Genes Brain Behav 12:275–281. https://doi.org/10.1111/gbb.12002

Brunberg, E, Wallenbeck, A, Keeling, LJ (2011) - Appl Anim Behav Sci 133:18–25. https://doi.org/10.1016/j.applanim.2011.04.019

Kasper C, Ruiz-Ascacibar I, Stoll P, and Bee G (2020) J Anim Breed Genet 137(6):545-558. https://doi.org/10.1111/jbg.12472

Langfelder, P, Horvath, S (2008) - BMC Bioinformatics 9:559. https://doi.org/10.1186/1471-2105-9-559

Love, MI, Huber, W, Anders, S (2014) - Genome Biol 15:550–550. https://doi.org/10.1186/s13059-014-0550-8

Ritchie, ME, Phipson, B, Wu, D., Hu, Y., Law, C.W. *et al.* (2015) - Nucleic Acids Res 43:e47. https://doi.org/10.1093/nar/gkv007

Valros, A (2018). Chapter 5 - Tail biting, In: Špinka, M. (Ed.), Advances in Pig Welfare, Herd and Flock Welfare. Woodhead Publishing, pp. 137–166. https://doi.org/10.1016/B978-0-08-101012-9.00004-6

van der Meer, Y, Gerrits, WJJ, Jansman, AJM, Kemp, B, Bolhuis, JE (2017) – PLoS ONE 12:e0174688. https://doi.org/10.1371/journal.pone.0174688

795. Analyses of heat stress effects on the farrowing rate of Iberian sows

M. Muñoz[1*], G. Matos[2], G. Gómez[2] and J.M. García-Casco[1,2]

[1]INIA-CSIC, Animal Breeding Department, 28040 Madrid, Spain; [2]Sánchez Romero Carvajal, SRC, Jabugo, Huelva, Spain; [3]Centro de I+D en Cerdo Ibérico, INIA-CSIC, Ctra. EX101 km 4,7, 06300 Zafra, Spain; mariamm@inia.es

Abstract

Iberian pigs are well-adapted to high temperatures, however, detrimental effects on reproductive traits have been reported. We evaluate the effect of heat stress on farrowing rate (FR) of Iberian sows. Data included 13,290 first insemination records from 3,009 sows. Temperature-humidity index (THI) was calculated using records from the closest weather station. Heat stress was defined trough heat load, as the number of degrees surpassing a particular threshold (THI0) affecting farrowing rate. Cycle number, heat load (THI0=16) and breed sire effects on FR were assessed and heritability using a threshold model was estimated. Heritability of FR was low 0.039 (0.018/0.065) and heat load, cycle number and sire effects were detected. Heat stress affects the farrowing rate of Iberian sows since the liability for this trait was lower with higher heat loads. Therefore, THI has to be taken into account for improving the farrowing rate.

Introduction

It is well known that global warming affects livestock species. High environmental temperatures can produce heat stress on animals, modifying their physiology through a decrease of their prolificacy and an increase of the mortality at birth (Hoffmann, 2010). High temperatures can reduce oestrous behaviour, alter the ovarian follicular development, compromise the competence of oocyte and inhibit the development of embryo (Hansen *et al.*, 2001). Farrowing rate is an interesting reproductive trait in porcine livestock that can be defined as getting pregnant from first insemination and maintaining this pregnancy until farrowing. Bloemhof *et al.* (2012) estimated heritability values for farrowing ranged from 0.05 to 0.19 in different purebred and crossbred pig lines. They also observed that farrowing rate can be affected by high temperatures.

Iberian pigs are well adapted to high temperatures, since they have been living in the Southwest of the Iberian Peninsula where the temperatures on summer (June-August) can reach up during several days between 35 and 45 °C. However, negative effects of high temperatures on farrowing rate, total piglets born and piglets born alive have been reported (Piñán *et al.*, 2021).

Therefore, the objective of the current study was to analyse the effect of heat stress on farrowing rate of an Iberian commercial line and to estimate the genetic parameters for this trait.

Materials & methods

Data. In the present study, data of Iberian pigs from a closed commercial population (SRC line) fattened in an extensive traditional system was used. A total of 13,290 first insemination records from 3,009 sows inseminated between August 2009 and February 2020 with semen of Iberian or Duroc sires was available. Iberian semen was collected from sires living in the same farm that sows that were inseminated the same day or the day after the semen was collected. However, semen from Duroc was bought to artificial insemination centres, therefore the semen, after being collected, had to be cooled and shipped from the insemination centre to the farm. Farrowing rate was defined as a binary trait being 1 when the first insemination resulted in a pregnancy with a gestation length between 108 and 117 days and a total number of piglets born alive

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_795, © M. Muñoz *et al.* 2022

between 1 and 13. On the other hand, farrowing rate was defined as 0 when the first insemination did not result in a pregnancy. The records not fulfilling these conditions were removed and, therefore, 9,838 and 3,153 records registered as 1 and 0, respectively, were available.

The sows were allocated on farrowing houses on a farm situated in the north of Huelva province (Spain). Although the insemination was carried out into sheds, the sows were located in outdoor ground parks. Temperature (T°) and relative humidity (RH) at insemination was collected from the closest public weather station (*El Campillo;* https://n9.cl/g4xdl). The daily temperature-humidity index (THI) measured in degrees Celsius (°C) was estimated based on the following formula according to Ravagnolo and Misztal (2000):

$$THI(°C) = T(°C) - [0.55 - (0.0055 \times RH)] \times [T(°C) - 14.5]$$

Heat load quantification. Heat stress effects were estimated through heat load (H). Daily heat load is defined as the number of degrees surpassing a given threshold of THI (THI0) on the insemination day. Heat stress effects on farrowing rate were assessed using the following generalized mixed model with binomial distribution:

$$y_{ijk} = C_i + S_j + H_k + e \quad (1)$$

where y_{ijk} is the farrowing rate of the individual y at the oestrus cycle number i, sire j and heat load k, C_i is the insemination cycle number (1, 2, 3, 4, 5, 6, 7, ≥8), S_j is the sire type used for the insemination (Iberian or Duroc) and H_K is the heat load (up to 11 levels). Four models were fitted according to four different THI0 (16, 18, 20 and 22 °C) and the most adequate model was selected through an Akaike Information Criterion (AIC) analysis. The statistical analyses were carried out using *glmer* function in Rstudio (RStudio Team., 2020).

Genetic parameter estimation. The following Bayesian threshold animal model was fitted for liability (L)

$$L = Xb + Z_d a_d + W_p p_p + e \quad (2)$$

where L is the liability for farrowing rate, b are the systematic effects of the insemination cycle number, sire and heat load, a_d is the vector of the direct additive genetic effects, p is the vector of the maternal permanent random effects and e the vector with the residual effects. X, Z_d and W_p are the incidences matrix. The analyses were carried out using the software TM (Legarra *et al.*, 2011) and convergence were assessed by visual inspection and estimating the Z criterion of Geweke.

Results

Average farrowing rate was equal to 0.76 in the whole population. The farrowing rate was lower when Duroc sires were used for insemination (0.64) than when Iberian sires were used (0.79). A total of 19 sires did not have breed information available. It is worth to mention that Duroc sires were less used for insemination than Iberian ones (2,974 vs 10,002) and the use of these breeds was not uniform along the year (Figure 1A) or at a particular THI value (Figure 1B).

The AIC values for different THI0 ranged from 13,789.98 to 13,929.37. The optimal AIC value was the one for a THI0 equal to 16 °C. According to this, heat load fitting a threshold of 16 °C was used for estimating heritability and permanent environmental parameters for farrowing rate (liability). As a result, low values

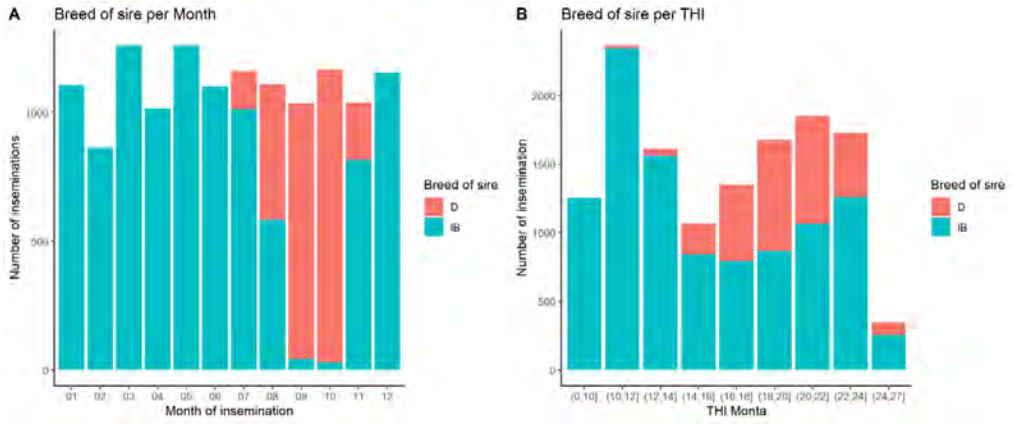


Figure 1. Number of inseminations per sire breed (Duroc or Iberian) (A) along months of insemination or (B) per THI.

were observed for both parameters, being h^2 equal to 0.039 (95% HPD: 0.018/0.065) and the permanent environmental effects of the sow (p^2) equal to 0.026 (95%HPD: 0.006/0.053).

Cycle number effects were inferred as difference from the first one. First cycle showed a higher liability than the second one, however, for cycle numbers greater than 6, their liability increased compared to the first cycle number.

The difference in liability for farrowing rate of Duroc respect Iberian sires was equal to -0.396 (95%HPD = -0.458/-0.334). There was also an effect of heat load on the liability for farrowing rate being lower when the heat load increases (Figure 2).

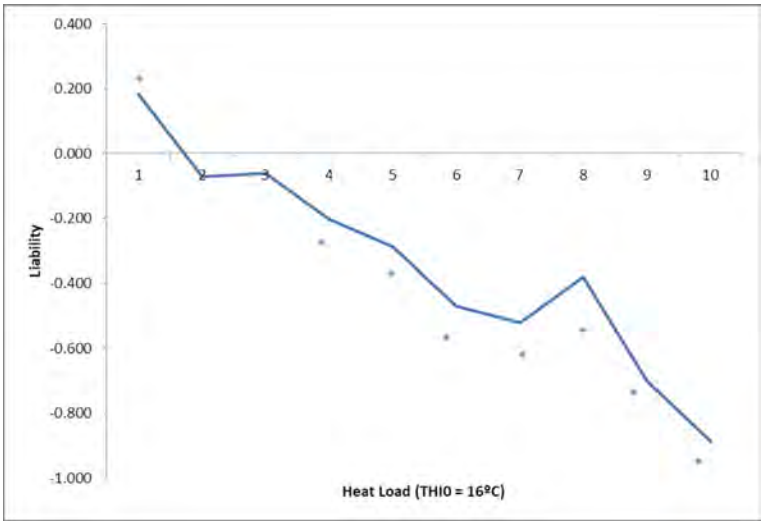


Figure 2. Posterior means of heat load (THI0 = 16 °C) deviated from the lowest category (heat load = 0) on liability for farrowing rate. *Bayesian posterior probability greater or lower than zero (PProb>0).

Discussion

In the population analysed, we observed a 76% of farrowing rate after the first insemination, this is a relevant trait because it is a direct measure of non-productive sow days when a sow is not lactating or gestating (Wilson *et al.*, 1986). Therefore, it is important to improve it. The heritability of liability for FR observed in the present study is low (0.03) and it agrees with the values reported by other authors (Holm *et al.*, 2005 and Bloemhof *et al.*, 2012).

The AIC analyses showed that a model including heat load with a THI0 of 16 °C had a better quality than those with THI0 of 18, 20 or 22 °C. The inclusion of heat load in the Bayesian threshold model revealed that heat load affects the farrowing rate. Although an increase in liability was observed when heat load was equal to 1 °C respect heat load equal to 0 °C. The liability for farrowing rate heavily decreases for heat load values equal or higher than 4 °C which agrees with the results reported by Bloemhof *et al.* (2012) in other pig breeds.

In addition to this, a lower farrowing rate was associated to Duroc semen. Duroc semen unlike the Iberian one, was refrigerated and then shipped to the farm, therefore, it cannot be discriminated whether the effects observed on the farrowing rate are due to Duroc breed itself or the semen handling. The results showed here point out that THI has to be taken into account for improving farrowing rate and other reproductive traits in Iberian pigs.

Acknowledgements

This work was financially supported by RTI2018-096189-J-I00. We want to thank to Sánchez Romero Carvajal (SRC) enterprise and to the staff of the farm for all the technical support.

References

Bloemhof, S., Kause, A., Knol, E. F., Van Arendonk, J. A. M., & Misztal, I. (2012) J. Anim. Sci., 90(7):2109-2119. https://doi.org/10.2527/jas.2011-4650.

Hansen, P. J., Drost, M., Rivera, R. M., Paula-Lopes, F. F., Al-Katanani, Y. M., *et al.* (2001) Theriogenology, 55(1), 91-103. https://doi.org/10.1016/S0093-691X(00)00448-9.

Hoffmann, I. (2010) Anim. Genet. 41(Suppl. 1):32–46. https://doi.org/10.1111/j.1365-2052.2010.02043.x.

Holm, B., M. Bakken, O. Vangen, and R. Rekaya. (2005) J. Anim. Sci. 83:41–48. https://doi.org/10.2527/2005.83141x.

Legarra, A., L. Varona, and E. López De Maturana. (2011). TM Threshold Model. Available at: http://snp.toulouse.inra.fr/~alegarra/manualtm.pdf.

Piñán, J., Alegre, B., Kirkwood, R.N., Soriano-Úbeda, C., Maj, M., *et al.* (2021) Effect of Season and Parity on Reproduction Performance of Iberian Sows Bred with Duroc Semen. Animals 11: 3275. https://doi.org/10.3390/ani11113275.

Ravagnolo, O., and Misztal, I. (2000). J. Dairy Sci. 83(9), 2126-2130. https://doi.org/10.3168/jds.S0022-0302(00)75095-8.

RStudio Team (2020). RStudio: Integrated Development for R. RStudio, PBC, Boston, MA URL http://www.rstudio.com/.

Wilson, M. R., R. M. Friendship, I. McMillan, R. R. Hacker, R. Piper *et al.* (1986) J. Anim. Sci 62:576–582. https://doi.org/10.2527/jas1986.623576x.

796. Whole-genome screening for resilience against PRRSV outbreaks in breeding sows

M. Laplana*, R. Ros-Freixedes, J. Estany, L.J. Fraile and R.N. Pena

Department of Animal Science, University of Lleida-Agrotecnio-CERCA Center, 25198 Lleida, Catalonia, Spain; marina.laplana@udl.cat

Abstract

The porcine reproductive and respiratory syndrome virus (PRRSV) has a world-wide impact in health and productivity of growing pigs and sows. A genetic component in the resilience of sows facing a PRRSV outbreak has been reported. In this study, we aimed to identify genetic variants associated with resilience in sows using data on the stability of reproductive performance (SRP) during a PRRSV outbreak as a proxy for resilience. Forty-eight Landrace × Large White sows with extreme SRP phenotypes underwent whole-genome sequencing at a coverge of 7X. A genome-wide association study (GWAS) for the SRP phenotype identified 44 variants associated with the 'stable' phenotype. Some candidate genes are related to disease, cell structure, proliferation and migration and immune response. The identification of genetic markers can be useful to identify resilient sows that can successfully cope with a PRRSV outbreak without compromising their health status or productivity.

Introduction

Pork is one of the most common sources of animal protein for human consumption, thus, making necessary a sustained increased in pig production to provide the growing global food demand. One strategy to achieve this is improving the efficiency of the pork production sector with animals that can face stressors and recover in a short period of time without altering their normal physiology and their productivity, also known as resilient animals.

Among the potential stressors that can challenge a farm, infectious agents can have dramatic consequences leading to large losses and compromising pig welfare. One of the common viral pathogens circulating in the swine production systems is the Porcine Reproductive and Respiratory Syndrome Virus (PRRSV), which can cause endemic and/or epidemic infections. In sows, PRRSV infection can lead to respiratory problems as well as reproductive failure, including embryonic death in early gestation (Lunney *et al.* 2010; Chand *et al.* 2012), foetal death and abortions in late gestation, early farrowings, and elevated preweaning mortality (Karniychuk *et al.* 2011; Karniychuk and Nauwynck 2013). Although a PRRSV vaccination scheme is a common control measure, vaccines have a limited power to control this disease (Nan *et al.* 2017) due, in part, to the high mutation rate of the virus and the kinetic of the immune responses.

The role of genetic variability of the host in the outcome of PRRSV infection in pregnant sows has been demonstrated (Serão *et al.* 2014; Abella *et al.* 2019; Pena *et al.* 2019). In addition, several DNA markers associated with resilience to PRRSV infection in reproductive sows have also been reported (Laplana *et al.* 2020). In this line, we aimed to provide a global characterization of the genetic variability of pig resilience to PRRSV infection and their combined effect in reproductive traits. For this, we used a low-coverage (7X) whole-genome sequencing approach to unravel to most promising genetic DNA markers that could be implemented in the future in selection or crossbreeding schemes to improve resilience under the challenge of a PRRSV outbreak.

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_796, © M. Laplana *et al.* 2022

Materials & methods

Animals. A group of 305 Landrace × Large White sows from a PRRSV-positive stable production farm were included in the study. Main characteristics of the animals and farm have been previously described (Abella *et al.* 2019). Reproductive data was collected from 2016 to 2018, with no major pig diseases with the exception of a PRRSV outbreak at the end of the screened period, which lasted 15 weeks. PRRSV exposure during the outbreak was confirmed by antibody detection in all sows as previously described (Abella *et al.* 2019). Data on 1,464 parities were recorded including the number of stillborn (NSB) and mummified (NMU) piglets per litter. The total number of lost piglets per litter (NLP) was calculated as the sum of NSB and NMU.

In order to discover genetic markers related to resilience, the stability of reproductive performance (SRP) was used as a proxy of PRRSV resilience. SRP was assessed by the ability of maintaining stable NLP during the PRRSV outbreak. SRP score was calculated for each sow as follows: NLP in outbreak – mean NLP in non-outbreak farrowings. Animals were ranked according to their SRP score and were classified as 'stable' sows when SRP score <1 and 'non-stable' sows when SRP score >1. Finally, a total of 48 sows with extreme reproductive phenotypes, including 22 stable and 26 non-stable sows, were selected from the top and bottom of the ranking for whole-genome sequencing (WGS).

Sequencing. Low-coverage (7X) WGS was performed in NovaSeq 6000 sequencing system (Illumina) with 150 bp paired-end reads using standard Illumina protocols for library preparation, sequencing and initial quality check in the Centro Nacional de Análisis Genómico (CNAG-CRG, Barcelona, Spain). Sequence reads were processed with Trimmomatic (Bolger *et al.* 2014) to remove adapter sequences and mapped to the reference genome *Sscrofa11.1* using the BWA-MEM algorithm (Li 2013). Duplicates were marked with Picard. Single nucleotide polymorphisms (SNPs) and short insertions and deletions were identified with the variant caller GATK HaplotypeCaller (GATK 3.8.0) (DePristo *et al.* 2011; Poplin *et al.* 2017) using default settings separately for each individual. After filtering out variants with a mean coverage 3 times greater than the average global coverage with VCFtools (Danecek *et al.* 2011), we retained biallelic variant positions for further analyses.

Genome-wide association analysis and annotation. Plink software version 1.9 was used to filter out variants with minor allele frequency below 0.1 and call rate below 0.8. Next, a relatedness matrix was constructed in GEMMA v0.98.1 software from a reduced Plink dataset pruned based on linkage disequilibrium (windows of 200 variants, sliding by 25, linkage disequilibrium r^2 of 0.1 and genotyping rate ≥99.99%). Two genome-wide association analyses (GWAS) were performed in GEMMA including 14,689,130 variants. The first GWAS (GWAS-1) was performed considering SRP as a categorical score (stable and non-stable). The second GWAS (GWAS-2) was performed using the quantitative value of SRP. Variants were considered associated with the phenotype when P-value$<10^{-6}$. Genomic regions were defined by the overlap of associated variants expanding the variant coordinates ±1 Mb. Genes were retrieved using Biomart tool from Ensembl (EMBL-EBI), release 104, and Pig Sus scrofa 11.1 genes and annotated with Ensembl Variant Effect Predictor and the Enrichr tool.

Results

We identified a total of 25,568,592 biallelic variants, adding up to a total of 21,469,571 SNPs and 4,099,571 indels in the sequenced animals. With the GWAS-1 we identified 44 variants associated with the classification by the stability of reproductive performance (SRP) distributed in 13 chromosomal regions (Figure 1). To note that 27.3% of the associated variants have not been previously annotated in the Ensemble database.

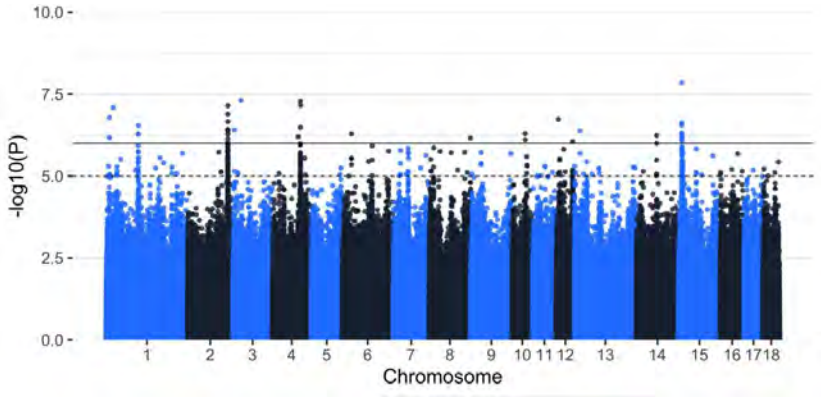


Figure 1. Manhattan plot of genome-wide -log10(*P*-values) for the stability of reproductive performance (SRP) in GWAS-1 (stable vs non-stable sows). The horizontal solid line represents the significance threshold (*P*-value=10^{-6}), the dashed line represents the suggestive threshold *P*-value (10^{-5}).

Next, associated variants were summarized in 16 genomic regions. Four regions had the strongest evidence of association. Briefly, in region 3 (1:9887943-11899689) the lead SNP, 1:107892069, also known as rs329563388, is located in the third intron of the *peptidylprolyl isomerase B* gene (*PPIB*) and is also a 3'UTR variant of *sorting nexin 22* gene (*SNX22*), which is related to intracellular trafficking, degradation of the extracellular matrix and collagen chain trimerization pathways as well as to osteogenesis imperfecta disease among others. Region 4 (2:133860837-136261458) contains twelve variants associated with SRP. The lead SNP in this region is 2:135261458, known as rs336589103, and located in the second intron of *AF4/FMR2 family member 4* gene (*AFF4*). Region 8 (4:93237379-96309123) contains 3 associated variants and expands 3.07 Mb. The lead SNP is 4:94237379 or rs322037463, which is located in intron 23 of *Gon 4 like* gene (*GON4L)* and relates to transcriptional repressor activity, which is required for B cell lymphopoiesis. Finally, region 16 (15:10187282-12903678) contains twelve associated variants. The lead SNP, 15:11187282, also known as rs692795800, is located in intron 42 of the *LDL receptor related protein 1B* gene (*LRP1B*). This gene has been reported as a putative tumor suppressor gene in humans as it is involved in cell proliferation and migration. All these genes are positional and functional candidates for SRP.

In order to identify more candidate genes for reproductive performance stability, we studied the functional annotation of genes positioned in the 16 genomic regions. GO analysis reported enrichment of genes that participated in immunoglobulin and other immune and growth factor receptor activity, catalytic activity acting on tRNA and ephrin receptor binding. Furthermore, we found enrichment in KEGG pathways involved in the regulation of the immune response and cell proliferation, division and survival, growth, angiogenesis and immunity.

Moreover, in GWAS-2 using SRP phenotype values we identified 4 variants associated with the reproductive outcome of the sow during the PRRSV outbreak, which overlapped with the region in chromosome 2 (2:133598288-135771079) associated with the SRP categorical classification.

Conclusions

Our study provides evidence of the genetic background on reproductive stability during a PRRSV outbreak. Identified associated genomic regions contain genes with functions related to immune response regulation, cell proliferation, division, survival and growth, all of them with putative impact on reproductive outcome.

Further studies are required to refine and validate these findings in larger populations, as well as their applicability to identify the sows that can be more resilient against PRRSV infection.

References

Abella G, Novell E, Tarancon V, Varona L, Pena RN, *et al.* 2019. J Anim Sci; 97(8): 3228–3236. https://doi.org/10.1093/jas/skz192

Bolger AM, Lohse M, Usadel B. 2014. Bioinformatics; 30(15): 2114–2120. https://doi.org/10.1093/bioinformatics/btu170

Chand RJ, Trible BR, Rowland RR. 2012. Curr Opin Virol; 2(3): 256–263. https://doi.org/10.1016/j.coviro.2012.02.002

Danecek P, Auton A, Abecasis G, Albers CA, Banks E, *et al.* 2011. Bioinformatics; 27(15): 2156–2158. https://doi.org/10.1093/bioinformatics/btr330

DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, *et al.* 2011. Nat Gen; 43(5): 491–498. https://doi.org/10.1038/ng.806

Karniychuk UU, Saha D, Geldhof M, Vanhee M, Cornillie P, *et al.* 2011. Microb Pathogenesis; 51(3): 194–202. https://doi.org/10.1016/j.micpath.2011.04.001

Karniychuk UU, Nauwynck HJ. 2013. Vet Res; 44(1): 95. https://doi.org/10.1186/1297-9716-44-95

Laplana M, Estany J, Fraile LJ, Pena RN.2020. Animals; 2020;10(5): 902. https://doi.org/10.3390/ani10050902

Li H. 2013. arXiv:1303.3997

Lunney JK, Benfield DA, Rowland RRR. 2010. Virus Res; 154(1-2): 1–6. https://doi.org/10.1016/j.virusres.2010.10.009

Nan Y, Wu C, Gu G, Sun W, Zhang Y-J, *et al.* 2017. Fron Microbiol; 8. https://doi.org/10.3389/fmicb.2017.01635

Pena RN, Fernández C, Blasco-Felip M, Fraile LJ, Estany J. 2019. Viruses; 11(8): 706. https://doi.org/10.3390/v11080706

Poplin R, Ruano-Rubio V, DePristo MA, Fennell TJ, Carneiro MO, *et al.* 2017. bioRxiv. https://doi.org/10.1101/201178

Serão NVL, Matika O, Kemp RA, Harding JCS, Bishop SC, *et al.* 2014. J Anim Sci; 2905–2921. https://doi.org/10.2527/jas.2014-7821

797. Genetic relationships among immune response traits of young healthy pigs evaluated by immunoassays

V. Bhatia[1*], J. Schmied[2], J. Cheng[1], X. Bai[3], B. Mallard[2], F. Fortin[4], J.C.S. Harding[5], M.K. Dyck[3], G.S. Plastow[3], C.J. Field[3], C. Rogel-Gaillard[6], F. Blanc[6], PigGen Canada and J.C.M. Dekkers[1]

[1]Iowa State University, Ames, IA, USA; [2]University of Guelph, ON, Canada; [3]University of Alberta, Edmonton, AB, Canada; [4]Centre de Développement du Porc du Québec, Inc., Quebec City, QC, Canada; [5]University of Saskatchewan, Saskatoon, SK, Canada; [6]Université Paris-Saclay, INRAE, AgroParisTech, GABI, France; vishesh@iastate.edu

Abstract

This study estimates phenotypic and genetic correlations of different components of the immune system of young healthy pigs, evaluated by various immunoassays, with an aim towards revealing a deeper understanding of the immune system of pigs and the role of immunoassays in predicting disease resilience. The composition of circulating immune cells, their phagocytic activity, and immune response measured by the High Immune Response (HIR™) technology were found to be heritable. The percentages of granulocytes and neutrophils in blood had high positive phenotypic and genetic correlations with each other and with the percentage of cells that showed phagocytosis activity. The HIR tests showed antibody-mediated immune response to be positively genetically correlated with cell composition traits, while cell-mediated immune response had a positive genetic correlation with only % eosinophils. These results suggest that immunoassays conducted on young healthy pigs provide important genetic information on their immune response profiles.

Introduction

Infectious diseases reduce animal welfare and disrupt production goals in pork production, thereby imposing a significant economic challenge to the industry. Thus, selecting pigs that are more resilient to disease i.e. with the ability to maintain performance while diseased (Albers *et al.*, 1987, Mulder and Rashidi, 2017), has gained traction. Over the years, many immunoassays have been developed that evaluate different aspects of the immune system, either *in vitro*, such as phagocytosis tests, or *in vivo*, such as the High Immune Response (HIR™) technology, which measures both antibody-mediated immune response (AMIR) and cell-mediated immune response (CMIR) (Schmied *et al.*, 2018), capturing critical aspects of the immune system to improve resilience (Wilkie and Mallard, 1999). Schmied *et al.* (2018) showed that HIR tests can identify pigs with superior capacity for immune response, and decreased mortality when disease challenged. Gorlov *et al.* (2021) showed that high phagocytic activity of cells is associated with a high innate immune status. Against this background, the objective of this study was to estimate phenotypic and genetic correlations among immune response traits in young healthy pigs based on various immune assays, to enhance our knowledge of the different aspects of the immune system regarding their potential as genetic indicator traits for disease resilience.

Materials & methods

The data used were obtained in a large research project that aims to understand the genetic basis of, and identify indicators of, disease resilience in a wean-to-finish polymicrobial natural disease challenge model (Putz *et al.*, 2019, Cheng *et al.*, 2020). The immune assays investigated here were conducted on young healthy Landrace × Yorkshire barrows from seven breeding companies, comprising PigGen Canada, while housed in a quarantine nursery before disease challenge. All pigs were genotyped using a 650K Affymetrix Axiom Porcine Genotyping Array. Details on genotype quality control are in Putz *et al.* (2019).

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_797, © V. Bhatia *et al.* 2022

Immunoassays. The immunoassays investigated in the quarantine nursery were: the proportion of immune cells in blood, the phagocytotic activity of those cells, and the HIR test. Blood samples for cell composition and phagocytosis tests were collected on the 994 pigs described by Lim *et al.* (2021) at ~26 days of age. Using flow cytometry, the phagocytosis test quantified the % of cells able to capture FITC-labelled opsonized *Escherichia coli* bacteria (Phag%) and phagocytic activity of those cells based on mean fluorescence intensity (PhagMFI). Based on morphology, leukocyte subpopulations of monocytes, lymphocytes, neutrophils, granulocytes, and eosinophils were also quantified and analyzed for phagocytic activity. For the HIR test, AMIR was evaluated by an enzyme linked immunosorbent assay and CMIR by measuring the change in skin-fold thickness due to cutaneous delayed-type hypersensitivity (Schmied *et al.*, 2012,2018). HIR phenotypes of 1397 pigs (cycles 5-7 of Cheng *et al.*, 2020) were used. In total, 708 pigs had both phagocytosis and HIR data.

Statistical analyses. Variance components and correlations were estimated by GBLUP using ASReml 4.1, with the fixed effect of batch, the covariate of entry age, and random effects of pen by batch, additive genetics, and litter, with the genomic relationship matrix of Cheng *et al.* (2020). For phagocytosis test traits, the fixed effect of assay date was also included. Technician was included as a fixed effect for AMIR and CMIR, along with serum antibody activity before antigen injection as a covariate for AMIR and a negative control for CMIR.

Results & discussion

Table 1 shows estimates of heritability and of phenotypic and genetic correlations among the cell composition and phagocytosis traits. Heritability estimates were similar to those reported on the same data by Bai *et al.* (2022). The phenotypic correlation of % lymphocytes was highly positive with % granulocytes (the genetic correlation estimate did not converge) and highly negative with % neutrophils, as was the genetic correlation. These findings agree with Koomkrong *et al.* (2017), who found a highly negative relationship between % neutrophils and % lymphocytes. Within cell type, %Phag and PhagMFI showed moderate positive phenotypic and genetic correlations (green cells in Table 1). The only strong positive genetic correlation of cell composition with phagocytic activity was observed for monocytes (0.39 for % and 0.49 for MFI). Phenotypic correlations were low between most phagocytosis traits, but genetic correlations ranged from low to high, but with generally high SE. Phenotypic and genetic correlations of the % phagocytizing cells were highly positive with % granulocytes and % neutrophils, and highly negative with % lymphocytes. Thus, it is consistent with Uribe-Querol and Rosales (2020) that show the essential role of granulocytes and neutrophils is that of phagocytosis.

Both AMIR and CMIR had weak phenotypic correlations with cell composition and phagocytosis traits (Table 2). However, AMIR showed moderate positive genetic correlations with most cell composition traits, while CMIR was moderately genetically correlated with only % eosinophils (0.24±0.30). Both AMIR and CMIR had moderate genetic correlations with phagocytizing activity of eosinophils (% and MFI) but low genetic correlations with phagocytizing activity of other cells. Note that, although AMIR is based on circulating antibody levels, AMIR itself occurs in secondary lymphoid organs and, therefore, may not relate directly with cell composition and phagocytosis activity in blood.

Significant estimates of heritability and of genetic correlations of immunoassay traits collected on young healthy pigs clearly indicate the substantial genetic variation that exists for immune response. Uncovering immune traits measured on young healthy pigs that are indicators of resilience will allow for the selection of pigs capable of maintaining performance traits during disease. Ongoing analyses investigate genetic relations with other immunoassay traits that were measured on young healthy pigs in this project, such as proliferation of peripheral blood mononuclear cells after mitogen stimulation (Jeon *et al.*, 2021) and natural antibody levels (Chen *et al.*, 2020) to provide a full understanding of all aspects of a young healthy pig's immune system.

Table 1. Estimates (SE) of heritabilities (on diagonal) and of genetic (above diagonal) and phenotypic (below diagonal) correlations among phagocytosis assay.

	Gran %	Eos %	Neu %	Mon %	Lym %	Phag %	Phag MFI	GPhag %	GPhag MFI	EPhag %	EPhag MFI	NPhag %	NPha gMFI	MPhag %	MPhag MFI	LPhag %	LPhag MFI
Gran %	***0.42*** ***(0.08)***	-0.03 (0.21)	DNC	-0.36 (0.16)	DNC	0.96 (0.02)	0.59 (0.14)	0.33 (0.15)	-0.34 (0.25)	0.58 (0.20)	0.52 (0.47)	0.00	-0.33 (0.24)	0.36 (0.17)	0.30 (0.19)	0.20 (0.16)	-0.02 (0.2)
Eos[2] %	0.04 (0.04)	***0.22*** ***(0.08)***	-0.10 (0.21)	0.25 (0.24)	-0.05 (0.21)	DNC	-0.10 (0.24)	0.00	0.02 (0.33)	DNC	0.00 (0.53)	0.00	0.03 (0.32)	-0.03 (0.25)	0.09 (0.26)	0.41 (0.19)	0.55 (0.26)
Neu %	1.00 (0.00)	-0.05 (0.04)	***0.42*** ***(0.08)***	-0.38 (0.16)	-0.97 (0.01)	0.95 (0.02)	0.60 (0.14)	0.32 (0.15)	-0.35 (0.24)	DNC	0.45 (0.46)	0.00	-0.35 (0.24)	0.33 (0.17)	0.29 (0.19)	0.11 (0.16)	-0.11 (0.19)
Mon %	-0.37 (0.03)	0.07 (0.04)	-0.37 (0.04)	***0.27*** ***(0.08)***	0.12 (0.19)	-0.13 (0.17)	-0.24 (0.21)	0.05 (0.18)	0.36 (0.26)	-0.54 (0.22)	-0.35 (0.60)	0.10 (0.17)	0.35 (0.26)	0.39 (0.21)	0.49 (0.19)	0.10 (0.18)	0.07 (0.23)
Lym %	0.97 (0.00)	-0.04 (0.04)	-0.96 (0.00)	0.13 (0.04)	***0.42*** ***(0.09)***	-0.76 (0.11)	-0.59 (0.14)	0.00	0.25 (0.26)	-0.75 (0.24)	-0.40 (0.41)	-0.39 (0.12)	0.24 (0.25)	-0.47 (0.16)	-0.39 (0.17)	-0.17 (0.16)	0.03 (0.2)
Phag %	0.94 (0.00)	-0.02 (0.03)	0.94 (0.00)	-0.18 (0.04)	-0.94 (0.00)	***0.41*** ***(0.08)***	0.50 (0.15)	0.47 (0.13)	-0.14 (0.27)	0.89 (0.19)	0.40 (0.43)	0.00	-0.12 (0.26)	0.53 (0.15)	0.39 (0.17)	0.33 (0.15)	0.06 (0.2)
Phag MFI	0.42 (0.03)	-0.06 (0.04)	0.43 (0.03)	-0.26 (0.04)	-0.39 (0.03)	0.42 (0.03)	***0.30*** ***(0.09)***	0.40 (0.18)	DNC	0.67 (0.21)	DNC	0.21 (0.24)	DNC	-0.10 (0.24)	0.08 (0.26)	-0.33 (0.18)	-0.21 (0.23)
GPhag %[2]	0.30 (0.04)	0.00	0.35 (0.04)	-0.10 (0.04)	0.00	0.46 (0.03)	0.47 (0.03)	***0.36*** ***(0.09)***	0.22 (0.29)	DNC	0.87 (3.27)	DNC	0.21 (0.29)	0.46 (0.18)	0.34 (0.2)	0.13 (0.17)	DNC
GPha gMFI	-0.19 (0.38)	0.03 (0.04)	-0.19 (0.04)	0.09 (0.04)	0.16 (0.04)	-0.08 (0.04)	0.74 (0.02)	0.31 (0.03)	***0.14*** ***(0.08)***	0.28 (0.35)	DNC	0.48 (0.26)	1.00 (0.00)	-0.21 (0.34)	-0.29 (0.35)	0.04 (0.26)	0.39 (0.3)
EPhag %[2]	0.25 (0.04)	-0.20 (0.03)	0.24 (0.03)	-0.10 (0.04)	-0.23 (0.04)	0.29 (0.03)	0.20 (0.04)	0.41 (0.03)	0.01 (0.04)	***0.19*** ***(0.08)***	0.64 (0.37)	0.35 (0.26)	0.27 (0.34)	DNC	0.19 (0.27)	0.46 (0.24)	0.51 (0.29)
EPhag MFI	-0.01 (0.04)	-0.16 (0.04)	0.00 (0.04)	-0.01 (0.04)	0.02 (0.04)	0.04 (0.04)	0.27 (0.03)	0.29 (0.03)	0.27 (0.03)	0.30 (0.03)	***0.05*** ***(0.06)***	DNC	0.75 (1.0)	-0.08 (0.52)	0.47 (0.64)	-0.52 (0.60)	0.34 (0.69)
NPhag %[2]	0.12 (0.03)	-0.01 (0.03)	0.13 (0.03)	0.00	-0.14 (0.04)	0.26 (0.03)	0.46 (0.03)	0.84 (0.01)	0.48 (0.03)	0.19 (0.04)	0.26 (0.03)	***0.28*** ***(0.09)***	0.47 (0.26)	0.67 (0.17)	0.52 (0.22)	0.30 (0.19)	0.37 (0.24)
NPhag MFI	-0.18 (0.04)	0.04 (0.04)	-0.19 (0.04)	0.09 (0.04)	0.16 (0.04)	-0.08 (0.04)	0.74 (0.02)	0.30 (0.03)	1.00 (0.00)	0.02 (0.04)	0.28 (0.03)	0.48 (0.03)	***0.15*** ***(0.08)***	-0.21 (0.33)	-0.30 (0.35)	0.06 (0.25)	0.42 (0.29)
MPhag %	0.30 (0.04)	-0.02 (0.04)	0.29 (0.04)	0.16 (0.04)	-0.35 (0.04)	0.49 (0.03)	0.10 (0.04)	0.34 (0.03)	0.02 (0.04)	0.11 (0.03)	0.09 (0.04)	0.37 (0.03)	0.03 (0.04)	***0.31*** ***(0.09)***	0.58 (0.18)	0.48 (0.16)	0.44 (0.22)
MPhag MFI	0.26 (0.04)	-0.01 (0.04)	0.26 (0.04)	0.16 (0.04)	-0.32 (0.04)	0.37 (0.03)	0.47 (0.03)	0.25 (0.04)	0.25 (0.04)	0.15 (0.04)	0.13 (0.04)	0.24 (0.04)	0.25 (0.04)	0.50 (0.03)	***0.26*** ***(0.09)***	0.07 (0.24)	0.34 (0.2)
LPhag %[2]	0.13 (0.04)	0.11 (0.04)	0.11 (0.04)	0.15 (0.04)	-0.15 (0.04)	0.29 (0.04)	-0.22 (0.04)	0.08 (0.04)	-0.01 (0.04)	0.02 (0.04)	-0.05 (0.04)	0.11 (0.04)	-0.01 (0.04)	0.41 (0.03)	0.03 (0.04)	***0.48*** ***(0.09)***	0.68 (0.14)
LPhag MFI	0.10 (0.04)	0.05 (0.04)	0.08 (0.04)	0.12 (0.04)	-0.13 (0.04)	0.20 (0.04)	0.02 (0.04)	0.12 (0.03)	0.09 (0.04)	0.04 (0.04)	0.02 (0.04)	0.14 (0.04)	0.09 (0.04)	0.24 (0.04)	0.28 (0.04)	0.48 (0.03)	***0.28*** ***(0.09)***

[1] Gran%: % granulocytes; Eos%: % eosinophils; Neu%: % neutrophils; Mon%: % monocytes; Lym%: % lymphocytes; Phag% and PhagMFI: % phagocytizing cells and their activity; GPhag%, EPhag%, NPhag%, MPhag%, and LPhag% (MFI): %Phag and PhagMFI of granulocytes, eosinophils, neutrophils, monocytes, and lymphocytes.

[2] Log Transformed due to non-normality of residuals.

[3] Did not converge.

Table 2. Estimates (SE) of phenotypic and genetic correlations of phagocytosis assays with AMIR and CMIR. Heritability was 0.25±0.07 for AMIR and 0.20±0.06 for CMIR.[1]

Traits[1]	AMIR		CMIR	
	Genetic	Phenotypic	Genetic	Phenotypic
Gran%	0.29 (0.19)	0.00 (0.04)	-0.09 (0.21)	0.01 (0.04)
Eos%[2]	0.25 (0.27)	-0.01 (0.04)	0.24 (0.30)	0.01 (0.04)
Neu%	0.31 (0.19)	0.00 (0.04)	-0.10 (0.21)	0.00 (0.04)
Mon%	0.10 (0.23)	-0.01 (0.04)	0.19 (0.25)	0.01 (0.04)
Lym%	-0.27 (0.19)	0.01 (0.04)	0.08 (0.21)	-0.02 (0.04)
Phag%	0.18 (0.18)	0.00 (0.04)	-0.19 (0.21)	0.02 (0.04)
PhagMFI	-0.45 (0.24)	-0.08 (0.04)	-0.10 (0.25)	0.01 (0.04)
GPhag%[2]	0.00 (0.00)	0.00 (0.03)	0.00 (0.00)	0.03 (0.03)
GPhagMFI	DNC[3]	-0.03 (0.04)	-0.10 (0.33)	-0.01 (0.04)
EPhag%[2]	-0.20 (0.30)	-0.03 (0.04)	-0.25 (0.32)	-0.05 (0.04)
EPhagMFI	-0.38 (0.52)	-0.02 (0.04)	0.23 (0.44)	0.05 (0.04)
NPhag%[2]	0.00 (0.00)	0.01 (0.01)	0.00 (0.00)	0.04 (0.04)
NPhagMFI	DNC	-0.03 (0.04)	-0.01 (0.33)	-0.01 (0.04)
MPhag%	-0.24 (0.26)	0.00 (0.04)	-0.04 (0.30)	0.02 (0.04)
MPhagMFI	-0.04 (0.25)	-0.02 (0.04)	-0.04 (0.27)	0.01 (0.04)
LPhag%[2]	0.03 (0.19)	0.06 (0.04)	-0.13 (0.21)	0.00 (0.04)
LPhagMFI	0.03 (0.24)	0.08 (0.04)	0.49 (0.24)	0.07 (0.04)

[1] For Footnotes, see Table 1.

Acknowledgements

Funding from Genome Canada, Genome Alberta, PigGen Canada, and USDA-NIFA grant # 2017-67007-26144 is acknowledged.

References

Bai X., Putz A.M., Wang Z., Fortin F., Harding J.C.S., *et al.* (2020) Front Genet 11:216. https://doi.org/10.3389/fgene.2020.00216.

Chen, Y., Cortes, L.E.T., Ashley, C., Putz, A.M., Lim, K.S., *et al.* (2020) BMC genomics 21(1):1-17. https://doi.org/10.1186/s12864-020-06994-0

Cheng J., Putz A.M., Harding J.C.S., Dyck M.K., Fortin F., *et al.* (2020) J Anim Sci 98(8). https://doi.org/10.1093/jas/skaa244.

Gorlov F., *et al.* (2021) IOP Conf. Ser.: Earth Environ. Sci. 677 032061

Jeon, R. L., Gilbert, C., Cheng, J., Putz, A. M., Dyck, M. K., *et al.* (2021) J Anim Sci, 99(8). https://doi.org/10.1093/jas/skab084

Koomkrong, N., Boonkaewwan, C., Laenoi, W., & Kayan, A. (2017) Asian-Australas J Anim Sci 30(12):1751–1755. https://doi.org/10.5713/ajas.17.0037

Lim, KS., Cheng, J., Putz, A. *et al.* (2021) BMC Genomics 22:614 https://doi.org/10.1186/s12864-021-07912-8

Putz A.M., Harding J.C.S., Dyck M.K., Fortin F., Plastow G.S., *et al.* (2019) Front Genet 9:660. https://doi.org/10.3389/fgene.2018.00660

Schmied, J., Hamilton, K., Rupa, P., Oh, S. Y., & Wilkie, B. (2012). Vet. Immunol. Immunopathol, 149(1-2):11–19. https://doi.org/10.1016/j.vetimm.2012.05.020

Schmied, J.D., Putz, A., Dekkers, J., Dyck, M., Fortin, F., *et al.* (2018). Proc. of the 11th WCGALP, Auckland, New Zealand.

Uribe-Querol and E., Rosales, C. (2020). Front Immunol 11:1066 https://doi.org/10.3389/fimmu.2020.01066

Wilkie, B., Mallard, B., 1999. Vet. Immunol. Immunopathol. 72: 231–235.

798. Sequencing-based genome-wide association study for resilience indicators in growing pigs

H. Laghouaouta*, R.N. Pena, R. Ros-Freixedes, L. Fraile and M. Laplana

Department of Animal Science, University of Lleida-Agrotecnio-CERCA Center, 25198 Lleida, Catalonia, Spain; houda.laghouaouta@udl.cat

Abstract

Resilience refers to the ability of animals to withstand stressors and rapidly recover their production. There is a growing interest in the improvement of resilience through selective breeding. However, there is no straightforward quantifying method. The body weight deviation and the increment of the acute-phase protein haptoglobin after applying a common vaccine challenge were used as resilience indicators in growing pigs. We assumed that resilient pigs would maintain their productivity and show a low activation of haptoglobin. Whole-genome sequencing-based association analyses were performed for the suggested resilience indicators of the most resilient (n=40) and susceptible (n=40) pigs. We identified a total of 28 associated genomic regions harbouring relevant candidate genes involved in the immune response (*CCL20*, *CD82*, *INAVA*…) and growth (*IGFBPL1*) pathways. Our study advances our understanding of the genetic background of resilience and corroborates the strong relationship between resilience, immune response and growth.

Introduction

Selective breeding for improved resilience would provide resistant animals with robust phenotypes (Mulder and Rashidi, 2017), reducing the economic losses and increasing the profitability of the production systems. However, there is no direct measure of resilience. Productivity-related traits and immunophenotypes reflecting perturbations caused by stressors have been proposed as resilience indicators in several livestock species (Berghof *et al.*, 2019; Chen *et al.*, 2020; Poppe *et al.*, 2021; Putz *et al.*, 2019). We suggested the body weight deviation (ΔBW) and the increment of the acute-phase protein haptoglobin (ΔHP) after applying a common vaccine challenge as resilience indicators in growing pigs. Pigs that maintained a steady production (high ΔBW) and had a minor activation of haptoglobin (low ΔHP) were deemed resilient. In contrast, pigs with a depressed production (low ΔBW) and a high activation of haptoglobin (high ΔHP) were deemed susceptible (Laghouaouta *et al.*, 2021). Heritability estimates were moderate for ΔBW (0.33) and low-to-moderate for ΔHP (0.16), indicating that they are genetically controlled and may be improved through selective breeding.

There is scarce information about the genetic background of resilience. A few association analyses were performed using resilience indicators in pigs such as antibody levels (Chen *et al.*, 2020; Sanglard *et al.*, 2020) and health-related traits (Ballester *et al.*, 2020). These studies detected promising candidate genes related to the immune response pathway. In this study, we used whole-genome sequencing data of pigs with extreme resilience indicators (ΔBW and ΔHP) to unravel the genetic background of resilience by identifying associated genomic regions and candidate genes.

Materials & methods

Animals and phenotypes. A total of 540 Duroc pigs at 10 weeks of age (71.4±2.4 days) were randomly placed into five batches of 104 to 111 pigs each. Pigs were reared under the same conditions and had *ad libitum* access to commercial diets. After two weeks (85.6±2.4 days of age), 445 pigs were challenged with an attenuated Aujeszky vaccine (Auskipra, Laboratorios Hipra, Amer, Girona) and 95 control pigs were inoculated with phosphate-buffered saline. The resilience indicators ΔBW and ΔHP were calculated for each challenged pig as described in Laghouaouta *et al.* (2021).

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_798, © H. Laghouaouta *et al.* 2022

Both ΔBW and ΔHP were simultaneously considered to select 80 pigs with extreme response towards the challenge (40 resilient and 40 susceptible). For each batch, the eight most resilient pigs with the highest ΔBW and the lowest ΔHP values and the eight most susceptible pigs with the lowest ΔBW and the highest ΔHP values were selected.

Genotypes and quality control. Pigs with extreme response to the challenge were sequenced with paired-end reads using the NovaSeq 6000 platform (Illumina, San Diego, CA, USA) at an average coverage of 8.5X. Adapter sequences were removed using Trimmomatic (Bolger *et al.*, 2014). Reads were aligned to the *Sscrofa11.1* reference genome using the BWA-MEM algorithm (Li, 2013) and duplicates were tagged using Picard Tools. Variant calling was performed with the GATK Haplotype Caller to identify single nucleotide polymorphisms, insertions and deletions (Depristo *et al.*, 2011). Further, variants with a minor allele frequency below 0.2 and a missing genotyping frequency above 0.1 were filtered out using PLINK (Chang *et al.*, 2015). After quality control, a total of 80 individuals and 7,986,234 variants remained in the dataset.

Whole-genome association study. Association analyses were performed for the resilience indicators by fitting a univariate linear mixed model using GEMMA (Zhou and Stephens, 2012). Data from the extreme pigs were analysed to investigate the genetic background of resilience using the continuous resilience indicators (ΔBW and ΔHP) and the categorical pigs' response to the challenge (resilient or susceptible). Phenotypes were adjusted for the batch fixed effect (5 levels). Associated variants were identified at a *P*-value threshold of 10^{-5}. Associated genomic regions were determined by grouping significant variants within 500 kb and extending each genomic region by 200 kb on each side to consider nearby genes. Candidate genes within the associated genomic regions were retrieved from the Ensembl database. Gene ontologies were investigated using DAVID database (Jiao *et al.*, 2012) and a literature search.

Results

Whole-genome association analyses identified 61, 32 and 69 associated variants with ΔBW, ΔHP and the classification of each pig by their response to the challenge (resilient/susceptible groups), respectively (Figure 1). Six genomic regions at pig chromosomes (SSC) 2, SSC3, SSC9, SSC14, SSC15 and SSC18 were associated with ΔBW and 13 genomic regions at SSC1, SSC2, SSC4, SSC8, SSC9, SSC10, SSC11 and SSC15 were deemed associated with ΔHP. Among them, two genomic regions at SSC9 (27.4-27.8 Mb) and SSC15 (128.8-129.2 Mb) were associated with both ΔBW and ΔHP. Besides, sixteen genomic regions were found to be associated with the pigs' response to the challenge, some of which overlapped with the associated genomic regions obtained for ΔBW (27.4-27.8 Mb at SSC9 and 128.6-129.2 Mb at SSC15) and ΔHP (238.9-239.3 Mb at SSC1, 18.2-18.9 Mb, 58.7-59.3 Mb and 129.0-129.7 Mb at SSC8, 27.4-27.8 Mb at SSC9, 22.6-23.4 Mb at SSC10, 64.8-65.2 Mb at SSC11 and 128.6-129.1 at SSC15). The associated genomic regions harbour several candidate genes involved in immune response and growth pathways such as C-C motif chemokine ligand 20 (*CCL20*), CD82 antigen (*CD82*), innate immunity activator (*INAVA*) and insulin-like growth factor binding protein-like 1 (*IGFBPL1*).

Discussion

The aim of this study was to unravel the genetic background of resilience in pigs by identifying genomic regions and candidate genes associated with resilience indicators. Association analyses were conducted using whole-genome sequencing data of the most resilient and susceptible pigs to increase the detection power of associations between variants and phenotypes.

A total of 28 associated genomic regions were detected across the pig genome, indicating the polygenic nature of resilience. That was expected since resilience is a complex trait regulated by various biological processes (Colditz and Hine, 2016). Since pigs were selected using simultaneously ΔBW and ΔHP, there

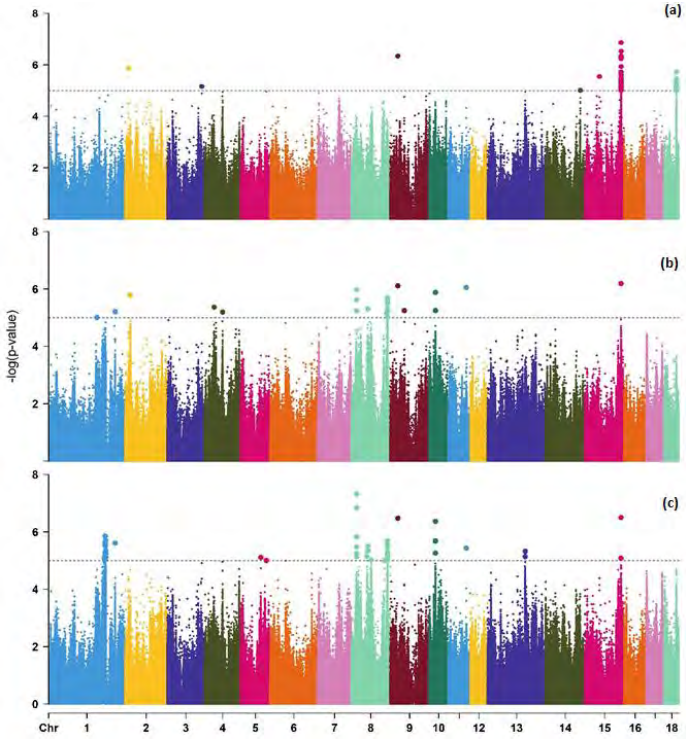


Figure 1. Manhattan plot for the genome-wide association study for (a) the body weight deviation (ΔBW), (b) the haptoglobin increment (ΔHP) and (c) the pigs' response after the vaccine challenge.

was a partial overlap between the associated genomic regions. The associated genomic regions harboured relevant candidate genes related to the immune response and growth pathways, which corroborates the relationship between resilience, the immune system and growth (Dantzer *et al.*, 2018; Zheng *et al.*, 2021). In addition, the associated genomic regions at SSC5 (71.8-72.2 Mb) and SSC8 (18.2-18.9 Mb, 55.4-55.8 Mb and 58.7-59.3 Mb) overlap with haptoglobin concentration QTL in pigs (Wimmers *et al.*, 2009). Further validation of the identified associations is needed in order to apply these markers in selection breeding programs under field conditions.

References

Berghof T.V.L., Bovenhuis H., and Mulder H.A. (2019) Front. Genet. 10:1216. https://doi.org/10.3389/fgene.2019.01216

Ballester M., Ramayo-Caldas Y., González-Rodríguez O., Pascual M., Reixach J., *et al.* (2020) Sci. Rep. 10:1–15. https://doi.org/10.1038/s41598-020-75417-7

Bolger A.M., Lohse M., and Usadel B. (2014) Bioinformatics 30:2114–2120. https://doi.org/10.1093/BIOINFORMATICS/BTU170

Chang C.C., Chow C.C., Tellier L.C., Vattikuti S. Purcell S.M., *et al.* (2015) Gigascience 4:7. https://doi.org/10.1186/s13742-015-0047-8

Chen Y., Tibbs Cortes L.E., Ashley C., Putz A.M., Lim K.S., *et al.* (2020) BMC Genomics 21:1–17. https://doi.org/10.1186/s12864-020-06994-0

Colditz I.G. and Hine B.C. (2016) Anim. Prod. Sci. 56:1961–1983. https://doi.org/10.1071/AN15297

Dantzer R., Cohen S., Russo S.J., and Dinan, T.G. (2018) Brain. Behav. Immun. 74:28–42. https://doi.org/10.1016/j.bbi.2018.08.010

Depristo M.A., Banks E., Poplin R., Garimella K. V., Maguire J.R., *et al.* (2011) Nat. Genet. 43: 491–501. https://doi.org/10.1038/ng.806

Jiao X., Sherman B.T., Huang D.W., Stephens R., Baseler M.W., *et al.* (2012) Bioinformatics 28:1805–1806. https://doi.org/10.1093/bioinformatics/bts251

Laghouaouta H., Pena R.N., Ros-Freixedes R., Reixach J., Díaz M., *et al.* (2021) Animals 11: 2970. https://doi.org/10.3390/ANI11102970

Li H., (2013) arXiv:1303.3997.

Mulder H.A., and Rashidi, H., (2017) J. Anim. Sci. 95: 3346–3358. https://doi.org/10.2527/JAS.2017.1479

Poppe M., Bonekamp G., van Pelt M.L., and Mulder H.A. (2021) J. Dairy Sci. 104:1967–1981 (2021). https://doi.org/10.3168/jds.2020-19245

Putz A.M., Harding J.C.S., Dyck M.K., Fortin F., Plastow G.S., *et al.* (2019) Front. Genet. 9: 660. https://doi.org/10.3389/fgene.2018.00660

Sanglard L.P., Fernando R.L., Gray K.A., Linhares D.C.L., Dekkers J.C.M., *et al.* (2020) Front. Genet. 11:1276. https://doi.org/10.3389/FGENE.2020.01011

Wimmers K., Murani E., Schellander K. and Ponsuksili S. (2009) Int. J. Immunogenet. 36: 141–151.

Zheng A., Zhang A., Chen Z., Pirzado S.A., Chang W., *et al.* (2021) J. Anim. Sci. Biotechnol.12:1–19. https://doi.org/10.1186/S40104-021-00591-1

Zhou, X., and Stephens, M., (2012) Nat. Genet. 44:821–824. https://doi.org/10.1038/ng.2310

799. First parity indicator traits for longevity in crossbred sows

M.W. Iversen* and E.G. Enger

Norsvin, Storhamargata 44, 2317 Hamar, Norway; maja.iversen@norsvin.no

Abstract

Longevity is an important trait in sows, for economical as well as animal welfare reasons. Due to early replacement based on breeding values, longevity data collected in purebred nucleus populations is often useless. Collecting data in crossbreds is an alternative. This study looked at alternative traits measured in first parity crossbred sows and their genetic correlations to longevity traits. Traits investigated were litter size, weight and fat depth at insemination and farrowing. Results show positive genetic correlations (range 0.38-0.76) between longevity and first parity observations of litter size, backfat and weight at insemination, backfat at farrowing, and gain in backfat during pregnancy. A reduction in backfat during first lactation had a negative correlation with longevity. These results indicate that backfat and weight at first insemination, or backfat at first farrowing can be used as indicator traits for longevity in sows, in addition to performance traits such as litter size.

Introduction

Longevity in sows is a very important trait because keeping the sow in the herd for an extra parity is cheaper than replacing it (Hoge and Bates, 2011), provided it still performs adequately. It is challenging to measure in purebred nucleus herds however, because sows are often replaced based on breeding values early in life, before they have a chance to show their potential for longevity. Therefore, collecting longevity information in crossbred sows may be a good alternative. Indicator traits measured early in life that are correlated to longevity may also be a good alternative and will improve the accuracy of the breeding values of related purebreds for longevity. Another challenge with longevity or culling is that it is affected by many factors other than genetics such as health, behaviour and reproduction, as well as production traits having an influence on culling decisions (Yazdi *et al.*, 2000). The aim of this study was to investigate new potential indicator traits for longevity in crossbred sows.

Materials & methods

Animals and data. Data was from one purebred dam line (A) and their F1 crossbred (X) with purebred dam line B. For purebred animals (A), only longevity data was used (53,186 animals). Longevity was defined as; LGY12 – whether (0 or 1) the sow was inseminated for a second litter within 85 days after first farrowing, LGY13 – how many litters the sow had within 270 days after first farrowing (1-3), and LGY14 – how many litters the sow had within 420 days after first farrowing (1-4). Crossbred sows had data from five farms on total number born (1,805 animals), total number born alive (1,805 animals), weight (in kg, 874 animals) and backfat (in mm, 858 animals) at insemination, backfat at farrowing (in mm, 1,016 animals), and longevity (69,595 animals) as described above.

Statistical analysis. All models were analysed in ASReml (Gilmour *et al.*, 2009). Pedigree-based relationships were used for variance component estimation. Traits were analysed by single-trait models to determine initial variance components, and then each of the first-parity crossbred traits were analysed bivariate with each of the longevity traits. The following model was applied:

$$y = Xb + Zu + Vv + e,$$

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_799, © M.W. Iversen and E.G. Enger 2022

where y was a vector of observations, X, Z, and V, known incidence matrices; b, a vector of fixed effects (Table 1); u, a vector of random additive genetic effects, with $u \sim N(0, A\sigma_u^2)$, where σ_u^2 was the additive genetic variance and A was the additive relationship matrix (pedigree); v, was a vector of common litter effects, with $v \sim N(0, I_v\sigma_v^2)$, where σ_v^2 was the variance of common litter effects; and e, a vector of residuals, with $e \sim N(0, I_e\sigma_e^2)$, where σ_e^2 was the residual variance. Matrices I_v and I_e were identity matrices of the appropriate dimensions.

Results

Weight of the sows increased over time, with drops in weight after each farrowing due to piglets being born and loss of weight during lactation (Figure 1). All farms had similar trends, but one farm had lower weights on average compared to other farms. Backfat decreased with time (within pregnancy-phase) and was always highest at first insemination compared to later inseminations (Figure 2). Farm 2 had higher variations in changes in backfat and higher levels overall than other farms, and Farm 1 had lower variation and levels.

Using only crossbred data, estimates of heritability for the longevity traits were between 0.01 and 0.025. When purebred data was included, heritabilities for longevity were in the range 0.06-0.08. Genetic correlations between the longevity traits were 0.87 (0.06), 0.85 (0.07) and 0.99 (0.01) for LGY12 with LGY13, LGY12 with LGY14 and LGY13 with LGY14, respectively. Litter size, backfat depth at farrowing and insemination, weight at insemination and fat gain in pregnancy were positively correlated to longevity (Table 2). Fat reduction during first lactation was negatively correlated to LGY13 and LGY14 but did not have a significant correlation to LGY12.

Discussion

This study investigated the use of data from first parity crossbred sows as indicators for longevity. Backfat depth at first insemination and farrowing had a positive genetic correlation with longevity traits. This agrees with findings by others (Knauer *et al.*, 2010; Stalder *et al.*, 2005), who found that sows with higher

Table 1. Fixed and random effects[1] for all traits.

Trait[2]	Litter nr dam	Farm	Year first insemination	Age (days)	Days weighing to insemination	Animal	Year-season farrowing	Litter id
LGY12	F	F	F			G		R
LGY13	F	F	F			G		R
LGY14	F	F	F			G		R
TNB	F	F		FR		G	R	
NBA	F	F		FR		G	R	
BFF		F		FR		G	R	
BFI	F	F		FR		G		R
WI	F	F		FR	FR	G		
FGP		F		FR		G	R	
FLL		F		FR		G	R	

[1] F = fixed effect; FR = fixed regression; R = random effect; G = genetic effect.

[2] Trait: LGY12, whether (0 or 1) the sow was inseminated for a second litter within 85 days after first farrowing; LGY13, how many litters the sow had within 270 days after first farrowing (1-3); and LGY14, how many litters the sow had within 420 days after first farrowing (1-4). TNB = total number born, NBA = number born alive, BFF = backfat depth at farrowing (mm), BFI = backfat depth at insemination (mm), WI = weight at insemination, FGP = backfat gain in pregnancy (mm), FLL = fat loss in lactation.

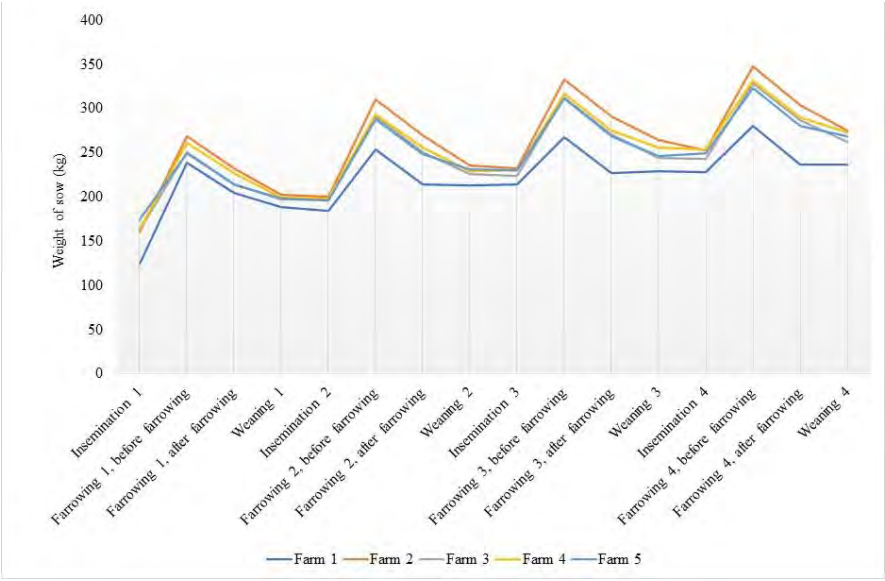


Figure 1. Average weight (kg) of sows measured for 4 litters across 5 farms.

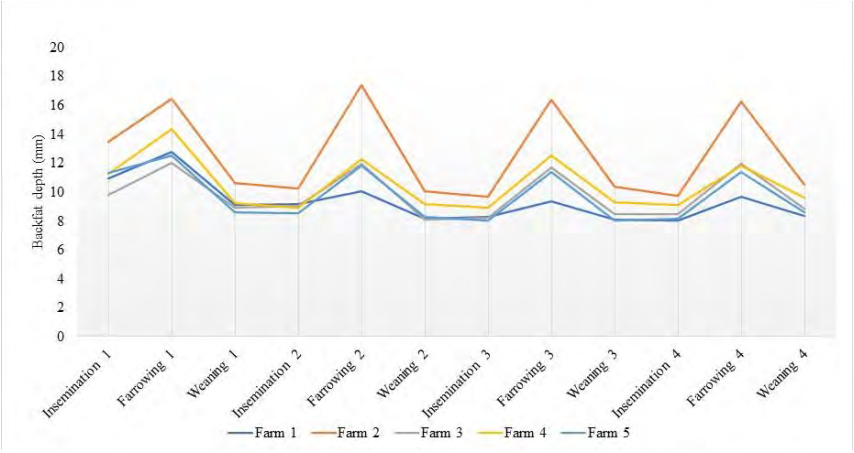


Figure 2. Average backfat depth (mm) of sows measured at different times for 4 litters across 5 farms.

backfat scored higher for number of lifetime litters and stayability. Backfat of sows decreased with age, and therefore, a focus on getting sows sufficiently fat before first insemination seems important to support longevity. Weight at first insemination was also positively correlated with longevity, thus larger sows (within age) may be favoured to improve longevity. However, selecting for these traits would result in larger adult size of sows, which is not necessarily positive. Fat reduction in lactation did not have a significant genetic correlation to LGY12 but was significant (and negative) for later parities. This may suggest that this is not an important factor for whether the sow gets inseminated for a second litter, but because backfat of sows decrease with time, it may have a bigger impact later in life. High weight loss in first lactation may lead to a smaller second litter (Schenkel *et al.*, 2010), which may influence selection decisions between second and third parity. Litter size at first parity was positively correlated with longevity and agrees with results found

Table 2. Heritabilities and genetic correlations (s.e.) between first parity traits and longevity.

Trait[1]	LGY12	LGY13	LGY14	h^2
Total number born	0.469 (0.170)	0.385 (0.174)	NS[2]	0.11 (0.04)
Number born alive	0.504 (0.177)	0.480 (0.178)	0.457 (0.184)	0.10 (0.04)
Backfat farrowing (mm)	0.487 (0.154)	0.426 (0.171)	0.469 (0.169)	0.38 (0.11)
Backfat insemination (mm)	0.569 (0.287)	NS	NS	0.16 (0.12)
Weight insemination (kg)	0.594 (0.116)	0.461 (0.140)	0.498 (0.136)	0.41 (0.11)
Fat gain pregnancy (mm)	NS	0.677 (0.230)	0.762 (0.229)	0.20 (0.10)
Fat loss lactation (mm)	NS	-0.5305 (0.186)	-0.588 (0.179)	0.32 (0.12)

[1] Traits: LGY12, whether (0 or 1) the sow was inseminated for a second litter within 85 days after first farrowing; LGY13, how many litters the sow had within 270 days after first farrowing (1-3); and LGY14, how many litters the sow had within 420 days after first farrowing (1-4).
[2] NS = not significant (estimates not significantly different from zero).

by others (Guo *et al.*, 2001). Not surprising, as sows with small litters at first parity are less likely to be inseminated for a second parity. In conclusion, backfat depth and weight at first insemination, and backfat depth at first farrowing are moderately to highly genetically correlated to longevity in sows and could be used as indicator traits to select for longevity.

References

Gilmour, A.R., Gogel, B., Cullis, B., Thompson, R. (2009) ASReml User Guide 3.0.

Guo, S.-F., Gianola, D., Rekaya, R., and Short, T. (2001) Livest. Prod. Sci. 72:243–252. https://doi.org/10.1016/S0301-6226(01)00219-6

Hoge, M., and Bates, R. (2011). J. Anim. Sci. 89:1238–1245. https://doi.org/10.2527/jas.2010-3175

Knauer, M., Stalder, K., Serenius, T., Baas, T., Berger *et al.* (2010) J. Anim. Sci. 88:3486–3492. https://doi.org/10.2527/jas.2009-2319

Schenkel, A.C., Bernardi, M.L., Bortolozzo, F.P., and Wentz, I. (2010) Livest. Sci. 132:165–172. https://doi.org/10.1016/j.livsci.2010.06.002

Stalder, K.J., Saxton, A.M., Conatser, G.E., and Serenius, T.V. (2005) Livest. Prod. Sci. 97:151–159. https://doi.org/10.1016/j.livprodsci.2005.03.009

Yazdi, M., Rydhmer, L., Ringmar-Cederberg, E., Lundeheim, N., and Johansson, K. (2000) Livest. Prod. Sci. 63:255–264. https://doi.org/10.1016/S0301-6226(99)00133-5

800. protiPig: Genomic analyses of traits related to nitrogen utilisation efficiency in pigs

M. Schmid*, R. Weishaar, D. Berghaus, E. Haese, N. Sarpong, A. Kurz, J. Seifert, A. Camarinha-Silva, M. Rodehutscord and J. Bennewitz

Institute of Animal Science, University of Hohenheim, Garbenstrasse 17, 70599 Stuttgart, Germany; markus_schmid@uni-hohenheim.de

Abstract

One important goal of sustainable pig production is to reduce nitrogen emissions by improving the protein utilisation efficiency of the pigs. The presented study investigated the suitability of traits related to protein efficiency for pig breeding as well as their genetic architecture. A total of 508 F1-crosses descending from 20 Piétrain sires and Landrace dams were generated and phenotyped for nitrogen utilisation efficiency, blood urea nitrogen, and production traits including feed intake and conversion and daily gain. All individuals were genotyped for 60k SNPs. Variance components were estimated by applying mixed linear models and revealed moderate heritabilities for all traits. Significant genomic regions were mapped for blood urea nitrogen traits in genome-wide association studies. It appeared that measuring the blood urea nitrogen might be a suitable indicator trait to breed for an improved nitrogen utilisation efficiency.

Introduction

Nitrogen (N) emissions have negative impact on the environment and the health of individuals as they lead to air pollution, water eutrophication as well as many more consequences on ecosystems which compromise e.g. biodiversity. In agriculture, a great proportion of N emissions result from N excretion of farmed animals such as pigs. Ammonia is the major product of metabolic amino acid degradation and it is transformed into urea in the liver and excreted via the urine of pigs. Measuring the N intake and excretion enables to calculate the N retention and the N utilisation efficiency (NUE). Thereby, the blood urea nitrogen (BUN) concentration provides a relatively easy-to-measure indicator trait of NUE (Kohn *et al.* 2005). Besides effects of the diet, age and sex, NUE is moderately impacted by the genetics of the individuals (Ruiz-Ascacibar *et al.* 2017, Kasper *et al.* 2020).

protiPig is a multi-disciplinary project at the University of Hohenheim with the overall aim to improve the protein efficiency in order to reduce the negative impacts of N emissions and simultaneously reducing the required amount of feed protein that is edible for humans. As breeding for an improved protein efficiency of the animals is expected to be one potential way of achieving such goals (e.g. Sorg *et al.* 2021), first results regarding the genetic architecture and genetic parameters of traits related to this subject obtained in a large experimental study are presented.

Materials & methods

Experimental design and data set. A F1 cross population consisting of 508 individuals (females:barrows = 1:1) that descended from 20 sires of the German Piétrain population and German landrace sows was established within a period of 2.5 years and using 21 cohorts of pigs. The animals were born and housed at the Hohenheim University's experimental station Unterer Lindenhof under standardised conditions. Between 76 days of age and the day of slaughtering (on average 149 days), they were kept in single cages and fed *ad libitum* with barley-wheat-soybean meal-based diets with marginal lysine concentration in a 2-phase fattening period. The feed intake (FI) was recorded during the whole experimental phase. Body weights were determined was well as blood samples at three consecutive days were taken for BUN measuring in week 3 and 6 of the experiment. Thus, the sampling periods were placed in feeding-phase 1 and 2, respectively. The

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_800, © M. Schmid *et al.* 2022

average BUN value across the three samples in sampling period 1 (2) were considered as phenotypes and are denoted as BUN1 (BUN2) in the following. Further, the feed conversion rate (FCR) and daily gain (DG) were calculated. The NUE of all individuals was approximated by using formulas derived from N-balance data and blood metabolites of a subset of 56 individuals for which N input and output was quantified. All individuals were genotyped using the Illumina PorcineSNP60 BeadChip (Illumina, San Diego, CA). After data filtering, the phenotypes for the traits FI, FCR, DG, BUN and NUE of 475 individuals and their genotype information of 48,748 SNPs remained as the data basis for further analyses.

Statistical analyses. The variance components of the traits were estimated using ASReml (Butler *et al.* 2009) and the following mixed model:

$$y = X_b b + Z_{co} co + a + e \quad (1)$$

Thereby, y denotes the vector of phenotypic records, vector b contains the fixed effects (sex and age at slaughtering) and co the random cohort effects. The design matrices X_b and Z_{co} relate the effects to the observation and vector e contains the residuals. The random animal effect a was modelled as $a \sim N(0, G \times \sigma_a^2)$, with σ_a^2 being the additive-genetic variance and G the genomic relationship matrix according to VanRaden (2008). In order to estimate genetic correlations between the traits, the model was extended towards a bivariate model. The likelihood-ratio test was applied to test the significance of the heritabilities. Single marker genome-wide association studies (GWAS) were conducted for the N related traits by means of GCTA software (Yang *et al.* 2011). The model was

$$y = X_b b + g + e, \quad (2)$$

where b contains the fixed effect of the SNP to be tested, the effects of the sex, cohort and age at slaughtering, $g \sim N(0, G \times \sigma_g^2)$ denotes the random polygenetic effect and e is the residual vector. A leave-one-chromosome-out approach was applied to avoid a double-fitting of the tested SNP. Because a genome-wide significance level using Bonferroni methods is very stringent, additionally a nominal significance threshold was defined. SNPs were considered to be significantly trait associated when their *P*-values were $P<5\times10^{-5}$ or they exceeded the Bonferroni genome-wide threshold.

Results

All investigated traits were significantly heritable ($P<0.001$), with heritabilities ranging from $h^2=0.25$ for NUE to $h^2=0.47$ for DG. Because the heritability of BUN1 and BUN2 did not differ ($h^2=0.41$) and their genetic correlation was 0.99, the average values of all 6 BUN measures were also treated as one single trait (meanBUN). The heritability of meanBUN and its genetic and phenotypic correlations to the other traits as well as the standard errors are presented in Table 1. A significant negative genetic correlation was observed between meanBUN and NUE.

Genome-wide association studies revealed typical quantitative traits. Three genome-wide significant SNPs on chromosome 7 were detected for BUN2 and meanBUN, respectively. When the nominal significance level was applied, 26 SNPs were significantly associated with BUN2, of which 24 were located on chromosome 7. The other were on chromosome 2 and 5. All 16 nominally significant SNPs for meanBUN were detected on chromosome 7. Figure 1 shows the GWAS results for the BUN traits. No SNP was significantly associated with the trait NUE.

Table 1. Heritabilities[1] (diagonal), genetic[2] (below diagonal) and phenotypic (above diagonal) correlations as well as their standard errors (in parentheses) of the investigated traits average blood urea nitrogen (meanBUN), efficiency of nitrogen utilisation (NUE), feed intake (FI), feed conversion rate (FCR) and daily gain (DG).

	meanBUN	NUE	FI	FCR	DG
meanBUN	0.49 (0.09)	-0.39 (0.07)	0.07 (0.08)	0.15 (0.08)	0.18 (0.08)
NUE	-0.60* (0.13)	0.25 (0.07)	-0.13 (0.07)	<0.01 (0.08)	-0.11 (0.19)
FI	0.44* (0.14)	-0.29* (0.20)	0.40 (0.09)	-0.25 (0.07)	0.75 (0.03)
FCR	-0.06 (0.17)	0.01 (0.21)	<0.01 (0.19)	0.27 (0.07)	0.44 (0.06)
DG	0.33* (0.14)	-0.24 (0.19)	0.86* (0.05)	0.52* (0.14)	0.47 (0.09)

[1] Obtained from univariate analyses.

[2] Obtained from bivariate analyses.

* Significant genetic correlation ($P<0.05$).

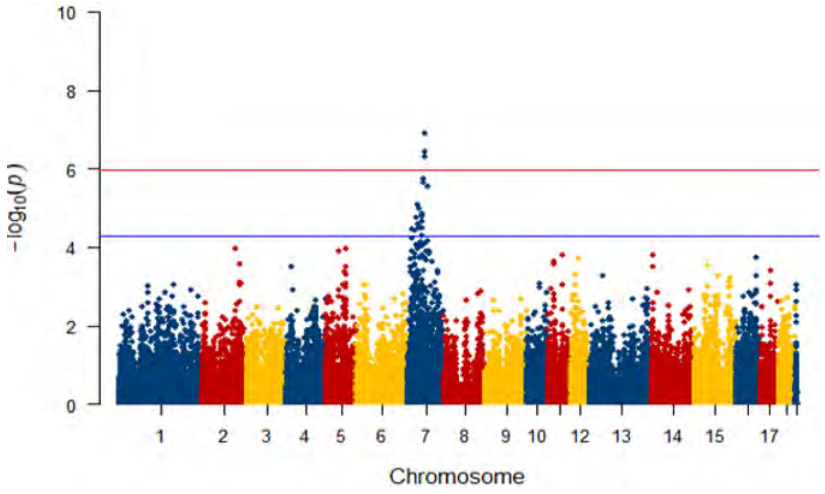


Figure 1. GWAS results of the blood urea nitrogen (averaged values across the two sampling periods, meanBUN).

Discussion

Moderate heritabilities with low standard errors were estimated for the N related traits BUN1, BUN2 and NUE, indicating that these traits have the potential as selection traits to breed for an improved utilisation of N and hence for protein efficiency in pigs. This agrees with Klindt *et al.* (2006) who estimated heritabilities for BUN at different timepoints and Kasper *et al.* (2020) who estimated population genetic parameters for NUE. However, due to the high degree of standardisation in the present study it should be further evaluated whether the findings are also valid in field data or in purebred populations. Of note, the present study implied a marginal lysine supply in order to allow the animals for expressing their potential of N utilisation. The genetic correlations between BUN1 and BUN2 were close to unity implying that the timepoint of phenotyping animals for BUN is less relevant. The results showed that low BUN levels come along with an increased NUE and are also correlated on a genetic scale which underlines the suitability of BUN as a heritable indicator trait for NUE. Genome-wide association studies revealed a polygenic architecture of the traits with many genomic regions, each having a small effect on the phenotype expression. Although strong

association signals were detected for BUN2 and meanBUN, suggesting candidate genes is not possible at this stage as the mapping precision was very low.

Adding gut microbiota data to further zoom into the mechanisms of protein efficiency in pigs is in progress within the *protiPig* project.

Acknowledgements

Funding by the H. Wilhelm Schaumann Stiftung, Hamburg, Germany, is gratefully acknowledged.

References

Butler D.G., Cullis B.R., Gilmour A.R., and Gogel B.J. (2009) ASReml-R reference manual. The state of Queensland, Department of Primary Industries and Fisheries, Brisbane, Australia.

Kasper C., Ruiz I., Stoll P., and Bee G. (2020) J Anim Breed Genet 137:545-558. https://doi.org/10.1111/jbg.12472

Klindt J., Thallman R.M., and Wise T. (2006) JAS 84(6):1323-1330. https://doi.org/10.2527/2006.8461323x

Kohn R.A., Dinneen M.M., and Russek-Cohen E. (2005) JAS 83(4):879-889. https://doi.org/10.2527/2005.834879x.

Ruiz-Ascacibar I., Stoll P., Kreuzer M., Boillat V., Spring P., and Bee G. (2017) Animal 11(3): 394-404. https://doi.org/10.1017/S1751731116001634

Sorg D., Klatt A., Plambeck N.O., and Koeder L. (2021), Perspektiven fuer eine umweltvertraegliche Nutztierhaltung in Deutschland. Available at: https://www.umweltbundesamt.de/sites/default/files/medien/5750/publikationen/2021-04-14_texte_33-2021_tierhaltung_bf_0.pdf.

vanRaden P.M. (2008) J Dairy Sci 91(11):4414-4423. https://doi.org/10.3168/jds.2007-0980

Yang J., Lee S.H., Goddard M.E., and Visscher P.M. (2011) Am J Hum Genet 88(1): 76-82. https://doi.org/10.1016/j.ajhg.2010.11.011

801. Genomic prediction of ham weight loss during dry-curing in heavy pigs

V. Bonfatti*, S. Faggion and P. Carnier

Department of Comparative Biomedicine and Food Science, University of Padova, viale dell'Università 16, 35020 Legnaro, Italy; valentina.bonfatti@unipd.it

Abstract

Excessive ham weight losses during dry-curing (WL) represent a loss of marketable product, hindering the quality of dry-cured hams. Bayesian regression models were developed for observed (OB) and infrared-predicted WL (IR), from the EBV of 640 and 956 crossbred pigs, respectively. The accuracies of direct genomic values of the purebred sires (DGV-PB) of crossbred pigs were estimated in a leave-one-family-out validation scheme. Accuracies were satisfactory and very similar for both traits. The correlation between the DGV-PB for IR and sire EBV for OB was slightly lower (0.318) than the correlation between the DGV-PB for OB and the sire EBV for OB (0.383). While genomic prediction of OB and IR can be equally suggested to be incorporated in future selection programs aiming at reducing WL, the use of IR allows a reduction in phenotyping costs, favouring the establishment of larger reference populations, with accuracies comparable to those achievable using OB phenotype.

Introduction

Genetic selection of pigs aiming at reducing excessive ham weight losses (WL) during dry-curing is challenging because measures of WL: (1) require the establishment of ham individual traceability systems throughout the dry-curing process, largely limiting the availability of phenotypes; and (2) can be obtained only after the completion of the entire dry-curing process, resulting in long generation intervals, with consequent penalizing effects on the response to selection. Recently, on-site infrared spectroscopy has demonstrated its effectiveness as a large-scale phenotyping tool for WL, by enabling cost-effective, high-throughput, early predictions of WL, using spectral data acquired at the slaughterhouse on green hams, combined with carcass and green ham quality traits (Bonfatti and Carnier, 2020). In addition, it provides infrared predictions of WL (IR) that are highly genetically correlated with the measured trait (OB) and characterized by a higher heritability than the corresponding measured trait (Bonfatti and Carnier, 2020). The incorporation of IR into pig genomic selection programs aiming at reducing WL might be a promising alternative to the genomic prediction of OB, as it would allow periodical phenotype collection at virtually no additional costs and the creation of larger reference populations. The objective of this study was to compare the accuracy of genomic prediction models for OB and IR and evaluate the ability of genomic predictions for IR to predict the genetic merit for OB.

Materials & methods

Phenotypes. A measure of WL was obtained on the left ham at the end of dry-curing (12 mo) from 1,888 crossbred (CB) pigs produced in the sib-testing program of the C21 Goland sire line (Gorzagri, Fonzaso, Italy). Age at slaughter and body weight were those typical of heavy pigs (270 days, and 167 kg, respectively). Of these pigs, 1,624 had also spectra records and were used to develop infrared prediction models for WL, as detailed by Bonfatti and Carnier (2020). Briefly, the prediction model included the following explanatory variables: (1) visible-infrared spectra, collected at the slaughterhouse on the transversal section of the subcutaneous fat of the raw ham using a LabSpec®5000 (ASD Inc., Boulder, Colorado, USA) equipped with a fibre optic contact probe; (2) sex; (3) carcass traits (carcass weight, backfat depth, lean meat content, weight of the hams); (4) green ham quality traits (ham subcutaneous fat depth, and linear scores for ham round shape, subcutaneous fat thickness, and ham marbling).

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_801, © V. Bonfatti *et al.* 2022

To obtain the IR values for the 1,624 pigs with available records of OB, a set of prediction models was developed excluding, at each round, all the samples from a given slaughter batch from the training set and applying the obtained model to the samples of the batch that was left out. The process was repeated until all the slaughter batches (n=25) were used in prediction. Finally, a model trained on the entire dataset (n=1,624) was applied to obtain records of IR in a larger dataset, extracted from the historical database of the genetic evaluation program of the Goland C21 sire line. The larger dataset included records of carcass traits, green ham quality traits and spectral variables from 8,048 CB pigs. Hence, a total of 9,672 records of IR were available for this study.

Pseudo-phenotypes. Genomic prediction models were developed following a multiple-step genomic evaluation approach, where the estimated breeding values (EBV) of the CB pigs were used as pseudo-phenotypes. The EBV for OB and IR were obtained using BLUPF90 (Misztal *et al.*, 2002). The complete pedigree information was available for all CB pigs with phenotypic records and all their purebred (PB) C21 Goland sires. The model included the fixed effect of sex (female and castrated male) and slaughter batch and the random additive genetic effect of the animal.

Genotypes. Genotypes were available for 1,029 CB pigs, spanning four generations. Of these, 640 had a record of OB and were offspring of 57 sires, and 956 had a record of IR and were offspring of 104 sires. Sire families with records of both OB and IR were 53. Genotypes of the sires were also available. Genotypes at 8,826 SNP were obtained using the GGP Porcine LD chip (Genomic Profiler for Porcine LD, GeneSeek Inc., a Neogen Co., Lincoln, NE) according to the manufacturer's protocol, and subsequently imputed to the GeneSeek® Genomic Profiler™ Porcine 50K (Neogen, Lansing, MI, USA) using Fimpute v. 2.2 (Sargolzaei *et al.*, 2014). All genotypes were then subjected to quality control using PLINK v. 1.9 software (http://www.cog-genomics.org/plink2). After editing, a total of 29,559 SNPs were retained for the genomic prediction models.

Genomic prediction model. Bayesian Ridge Regression was fitted to the CB data using the BGLR package (De los Campos and Rodriguez, 2021) in the R software (R Core Team, 2021). A total of 500,000 iterations were generated to estimate the parameters of the models, with a burn-in period of 100,000 iterations, and a thinning interval of 100 samples. To simulate a scenario under which a reference population including only CB pigs with phenotypic information is used to provide genomic predictions of the genetic merit of PB candidates, the estimated SNP substitution effects were used to calculate the direct genomic values of the PB sires (DGV-PB) of the CB pigs.

In particular, models were trained on the EBV of CB pigs and tested on the EBV of PB sires (EBV-PB) according to the following scenarios:

1. The model was trained on the EBV of CB pigs for OB and tested on the EBV-PB for OB;
2. The model was trained on the EBV of CB pigs for IR and tested on the EBV-PB for IR;
3. The model was trained on the EBV of CB pigs for IR and tested on the EBV-PB for OB.

For each scenario, a leave-one-family-out training-validation scheme was performed: models were trained excluding, at each round, the members of a sire family, a Bayesian Ridge Regression model was fitted to the training set, and the estimated SNP substitution effects obtained at each round were used to calculate the DGV-PB of the sire that was excluded from training. The correlation between EBV-PB and DGV-PB was used as a measure of model accuracy. The bias was calculated as the regression coefficient of the EBV-PB on DGV-PB.

Results

On average, OB was 27.8±2.4%, whereas IR averaged 26.5%, and exhibited a lower variability (SD=1.8%) compared to the corresponding measured trait. The accuracies of the genomic predictions obtained for the investigated traits are summarized in Table 1. Values of accuracies were very similar for OB and IR (0.383 and 0.396, in scenarios 1 and 2, respectively), whereas a lower accuracy (0.318) was obtained in scenario 3. All the correlation coefficients were significantly different from 0. Regression coefficients were highly variable across cross-validation folds and not significantly different from 1.

Discussion

The leave-one-family-out training-validation scheme allowed to mimic as far as possible the circumstances under which models are going to be applied in practice, where selection is based on the DGV estimated for PB breeding candidates and estimates are based on a reference population of CB finishing pigs belonging to families related to, but different from, those of the PB candidates.

Scenario 1 had the aim to evaluate the accuracy attainable by models exploiting OB phenotyping, whereas scenario 2 allowed testing the ability of models trained on IR pseudo-phenotype to predict the EBV-PB for the same trait. Results obtained in the two scenarios were comparable, probably due to the larger training set and higher heritability of IR compared to OB (0.39 vs 0.31, data not reported in tables).

The correlation between DGV-PB and EBV-PB in scenario 3 corresponds to the accuracy of models in predicting the genetic merit of PB sire from a reference population based on CB pigs with no close relationships with the sires, thus representing the most realistic scenario. The accuracy under this scenario was lower than the accuracy achievable if the model was trained on OB (scenario 1), but comparable to that value for practical applications. In pigs the accuracy of the DGV tends to decline rapidly over generations, and periodical phenotype collections, especially in animals related to selection candidates, are likely to be needed to update genomic prediction models (van Grevenhof and van der Werf, 2015). Routine phenotyping for OB is currently unfeasible, and the periodical phenotype collection is expensive. In addition, when the target is the prediction of PB genetic merit for CB animals, an increased number of training animals with phenotypes is needed. Under these circumstances, the use of IR in place of OB can allow a reduction of the phenotyping costs, favouring the construction of larger reference populations, a greater flexibility in the collection of phenotypes, with only a minimal impact on model accuracy compared to the use of OB phenotypes.

Considering the limited phenotyping costs for IR and the relatively short generation length achievable using IR phenotypes, the genetic gain attainable by a genomic selection program based on IR phenotyping

Table 1. Accuracy of direct genomic values of sires (DGV-PB) for observed ham weight loss (OB) or its infrared prediction (IR) in predicting the EBV of sires (EBV-PB) for OB or IR.[1]

Scenario	DGV-PB	EBV-PB		Accuracy		Bias	
			n	*r*	*P*-value	B	*P*-value[2]
1	OB	OB	57	0.383±0.125	<0.01	1.190±0.387	0.626
2	IR	IR	104	0.396±0.091	<0.001	0.838±0.192	0.400
3	IR	OB	104	0.318±0.094	<0.001	0.739±0.218	0.234

[1] n: number of animals used to calculate the accuracy; *r*: correlation coefficient between the DGV-PB and the EBV-PB; β: regression coefficients of EBV-PB on DGV-PB.
[2] H_0: β = 1.ß

of the reference population should be compared to that achievable by a traditional selective program based on routine IR phenotyping.

References

Bonfatti V., and Carnier P. (2020) Prediction of dry-cured ham weight loss and prospects of use in a pig breeding program. Animal 14, 1128-1138, https://doi.org/10.1017/s1751731120000026.

De los Campos G., and Rodriguez P.P. (2021) BGLR: Bayesian generalized linear regression. Available at: https://cran.r-project.org/web/packages/BGLR/index.html.

Misztal I., Tsuruta, S., Strabel, T., Auvray, B., Druet, T., *et al.* BLUPF90 and related programs (BGF90) (2002) Proc. of the 7th World Congress on Genetics Applied to Livestock Production, Montpellier, France.

R Core Team 2021. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Available at: https://www.R-project.org/.

Sargolzaei M., Chesnais J.P., Schenkel F.S. (2014) A new approach for efficient genotype imputation using information from relatives. BMC Genomics, 15, 478, https://doi.org/10.1186/1471-2164-15-478.

Van Grevenhof I., Van der Werf J. (2015) Design of reference populations for genomic selection in crossbreeding programs. Gen. Sel. Evol. 47, 14, https://doi.org/10.1186/s12711-015-0104-x.

802. Meta-analyses for boar taint compounds in two purebred maternal lines and Piétrain-sired crosses

I. Brinke[1*], C. Große-Brinkhaus[1,2], K. Roth[1], M.J. Pröll-Cornelissen[1], I. Schiefler[2] and E. Tholen[1]

[1]Institute of Animal Science, University of Bonn, Endenicher Allee 15, Bonn, Germany; [2]Association for Bioeconomy Research (FBF e.V.), Adenauerallee 17, 53115 Bonn, Germany; ibri@itw.uni-bonn.de

Abstract

Breeding against boar taint becomes more important since surgical castration of young piglets was banned by law in Germany from 2021. Main compounds responsible for boar taint are androstenone and skatole. Previous studies underline the polygenetic background of at least androstenone in the genome and pointed out breed-specific differences. Genotype data from previous genome wide association studies were used to perform meta-analyses in order to identify overlapping gene regions between lines, associated with androstenone and/or skatole. Data from three different origins was used: two maternal purebred breeds (Landrace and Large White) and a Piétrain-sired commercial endproduct population (Piétrain×F1/F2 sow). Results of the meta-analyses showed eleven overlapping markers for skatole in all three populations in a previously identified region on *Sus scrofa* chromosome (SSC) 14 around 141 Mega base pairs (Mb) which underlines the importance of this region for skatole for further analyses.

Introduction

Castration of young male piglets in their first week of life without anaesthesia to avoid boar taint is banned by law in Germany from 2021 (Deutscher Bundestag 06.11.2018). Boar taint is mainly caused by two compounds: androstenone (AND) (Patterson 1968), a sex steroid hormone which is produced in the Leydig cells of a boars' testis, and skatole (SKA), which is synthesized as a metabolite of the amino acids tryptophan in the colon (Yokoyama und Carlson 1979). Accumulation of both compounds in the backfat starts with the puberty as the sexual maturation begins (Bonneau 1982). Breeding against boar taint is a sustainable and beneficial long-term alternative to surgical castration regarding animal welfare, resource efficiency (Lundström *et al.* 2009) and carbon footprint (Stefanski *et al.* 2018).

Both compounds were identified to be moderate to high heritable with a range of 0.25 to 0.88 for AND and 0.19 to 0.54 for SKA as reviewed by Robic *et al.* (2008) To establish breeding against boar taint in modern breeding programmes, it is mandatory to identify causal gene regions or quantitative trait loci (QTL) that are responsible for the formation, metabolism or degradation of AND and SKA. For AND, previous studies showed significant regions across the whole autosome within and above different breeds (Große-Brinkhaus *et al.* 2015), which underlines the polygenetic genetic architecture of this trait. In contrast to that, SKA is assumed to have a more monogenetic background with focus on SSC 6, 7, 10 and 14.

The aim of this study was to conduct meta-analyses from data and results of earlier genome-wide association studies (GWAS) in different pig populations to identify overlapping regions for AND and SKA.

Materials & methods

The datasets for this study were based on 3,211 boar taint phenotypes of the purebred nucleus populations Landrace (LR) and Large White (LW) and a Pi-sired crossbred (Pi×$F_{1/2}$) population. The crossbred sows had different compositions: a 2-way cross of type LW×LR and a 3-way cross of type LW×(Leicoma×LR) (Große-Brinkhaus *et al.* 2015). All AND and SKA concentrations of the nucleus populations were analysed from backfat tissue samples by using a standardized stable isotope dilution analysis-headspace solid-phase microextraction-gas chromatography / mass spectrometry (Fischer *et al.* 2011). For the Pi×$F_{1/2}$ trials, AND and SKA were

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_802, © I. Brinke *et al.* 2022

evaluated from backfat tissue samples using gas chromatography/mass spectrometry (GC/MS) (García-Regueiro und Diaz 1989), as this data was analysed before the method of Fischer *et al.* (2011) was established. All animals were genotyped with the Illumina PorcineSNP60 Bead Chip (Illumina, San Diego, CA, USA).

Genome wide association study. GWAS was performed univariate per trait and population by using the R-package GenABEL (Aulchenko *et al.* 2007b). Concentrations of AND and SKA were used as log-transformed values. SNPs and individuals with a call-rate of less than 0.95 and SNPs with a minor allele frequency less than 0.05 were excluded from further analyses. In the nucleus populations, population stratification was corrected by application of the GRAMMAR approach (Aulchenko *et al.* 2007a) whereas the stratification in the crossbred population was corrected by the genomic control approach by Devlin und Roeder (1999). Association analyses were performed under the following model for the nucleus populations:

$$y = X\beta + Zu + e \quad (1)$$

where y contains the phenotypes AND and SKA. The Equation 1 consisted of year-season of slaughter as fixed environmental effect (β) and animal (u) and error (e) as random effects. Age and weight at slaughter were used as fixed covariates. The model for the Pi×$F_{1/2}$ data differs only slightly with a combined factor including the effects testing station-year-season and housing system (group- or single-penned).

Meta-analysis. For the meta-analysis, the results of the performed GWAS described above, were compared trait specific among the populations. Therefore, a weighted inverse variance method was used, as implemented in the R-package MetABEL (Aulchenko *et al.* 2007b). In brief, the weight (w_i) for each GWAS study (i) was equal to one over the square of the standard error (s_i^2) of the estimate for that GWAS study:

$$w_i = \frac{1}{s_i^2} \quad (2)$$

The pooled effect estimate was then calculated by multiplying the GWAS study effect estimate (β_i) by the GWAS study weight (w_i, Equation 2), summing over all GWAS studies (Equation 3).

$$\beta_{pool} = \sum_{i=1}^{N} w_i \beta_i \quad (3)$$

In the last step, the pooled effect estimate (β_{pool}, Equation 3) was divided by the sum of the GWAS study weights in order to calculate the χ^2 statistics of the meta-analysis (Equation 4) (Hartung *et al.* 2008).

$$\chi^2 = \frac{\beta_{pool}}{\sum_{i=1}^{N} w_i} \quad (4)$$

As the datasets were already corrected for population stratification for the GWAS, further correction for stratification was not necessary. Results from the meta-analyses were corrected after Bonferroni.

Only the results of two GWAS studies were investigated at once within the meta-analysis, leading to following comparisons: {LR:LW}, {LR:Pi×$F_{1/2}$} and {LW:Pi×$F_{1/2}$}.

Results

The number of animals, overall means and standard deviations of raw phenotypes are shown in Table 1, separated by population.

Population specific GWAS for AND revealed nine significant associations on SSC 5 in the LR nucleus population whereas no significant associations were found for the other populations. All markers were

Table 1. Descriptive statistics of the analysed traits.[1]

	LR			LW			Pi×$F_{1/2}$		
Trait	n	x̄	SD	n	x̄	SD	n	x̄	SD
AND	1,293	1,923	1,433.29	1,317	1,305	1,107.46	601	569	664.05
SKA	1,293	203.8	221.67	1,317	84.79	122.58	601	169.4	182.84

[1] AND = androstenone in ng/g fat; SKA = skatole in ng/g fat; x̄ = mean; SD = standard deviation.

located in a region from 21.5 to 22.3 Mb. For SKA, five significant associations were found on SSC 6 and 14 in LR, 15 in LW on SSC 14, and none in the Pi×$F_{1/2}$ population. The most prominent region was at SSC 14 around 141 Mb where 15 of the 19 significant markers of both breeds were located.

A considerable larger number of QTLs were found by applying the meta-analyses using the combined results of two breeds. In total meta-analyses for SKA revealed 27 markers for {LR:LW} (24 located on SSC 14, three on SSC 6), 5 or 10 markers for {LW:Pi×$F_{1/2}$} and {LR:Pi×$F_{1/2}$} respectively (Figure 1b). Of particular importance are three markers which were found in all combined breed scenarios. These QTL were located on SSC 14 around 141 Mb which covers the well-known CYP2E1 gene. Moreover, six identical QTL were found in the meta-analysis of the {LR:LW} and {LW:Pi×$F_{1/2}$} combinations. These novel SKA influencing QTLs were located on SSC 6 and 14. For SKA, nine markers were found in meta-analyses but not in single analyses (0 vice versa).For AND one significant marker was found in the combined analysis {LR:LW}, seven for {LR: Pi×$F_{1/2}$} (all located on SSC 5 in a 21.5 to 22.3 Mb interval) and 35 for {LW:Pi×$F_{1/2}$}(Figure 1a). Only one QTL was found in the meta-analyses of {LR:LW} as well as in {LR:Pi×$F_{1/2}$}. This novel QTL is located on SSC 5 at 21.5 Mb. Meta-analyses for AND revealed two markers, that were only found in the single analyses but not in the meta-analyses (0 *vice versa*).

Discussion

Breeding against boar taint is an appropriate method to replace surgical castration without anaesthesia if the amount of odorous boars at slaughterhouse can be reduced. The descriptive summary (Table 1) showed that concentrations of AND and SKA in fat were on average greater in LR than in LW. The Pi×$F_{1/2}$ population revealed the lowest AND concentration, but was situated between LR and LW regarding the mean SKA concentration.

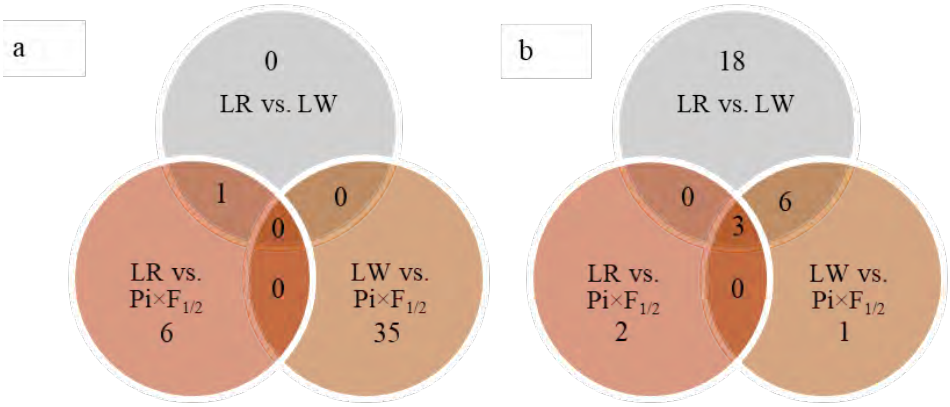


Figure 1. Number of shared significant markers in meta-analyses between the three datasets for androstenone (a) and skatole (b).

Different studies identified an extensive range of QTL associated with AND across the whole genome which differs among breeds as summarised by Robic *et al.* (2008), not including the region at 21.5 to 22.3 Mb that was identified in this study. In contrast to that, a candidate gene for SKA metabolism (CYP2E1) in the liver is already identified on SSC 14 in several purebred and crossbred lines (Squires und Lundström 1997; Skinner *et al.* 2005; Moe *et al.* 2009; Mörlein *et al.* 2012; Wiercinska *et al.* 2012). Meta-analyses in this study confirmed this region to be involved with SKA as 15 of a total of 19 significant associated markers were located in the promoter region of CYP2E1 on SSC 14 at ~141 Mb. This shows the potential of this QTL for breeding against boar taint as this region seems to be highly conserved among breeds. Further analyses, especially multivariate GWAS is necessary to further evaluate this regions potential on SSC 14 for further (genomic) selection against boar taint. In a next step, a pooled meta-analysis of all three populations {LR:LW:- $Pi \times F_{1/2}$} will be proven.

Acknowledgements

Project were supported by funds of the Federal Ministry of Food and Agriculture (BMEL) based on a decision of the Parliament of the Federal Republic of Germany via the Federal Office for Agriculture and Food (BLE) under the innovation support programme (Grant numbers: 2817904115 and 313-06.01-28-1-68.024-11).

References

Aulchenko, Y. S.; Koning, D.-J. de; Haley, C. (2007a). In: *Genetics* 177 (1), S. 577–585. https://doi.org/10.1534/genetics.107.075614.

Aulchenko, Y. S.; Ripke, S.; Isaacs, A.; van Duijn, C. M. (2007b). In: *Bioinformatics* 23 (10), S. 1294–1296. https://doi.org/10.1093/bioinformatics/btm108.

Bonneau, M. (1982). In: *Livest. Prod. Sci.* 9 (6), S. 687–705. https://doi.org/10.1016/0301-6226(82)90017-3.

Deutscher Bundestag (06.11.2018): Entwurf eines Vierten Gesetzes zur Änderung des Tierschutzgesetzes. Drucksache 19/5522.

Devlin, B.; Roeder, K. (1999). In: *Biometrics* (55), S. 997–1004.

Fischer, J.; Elsinghorst, P. W.; Bücking, M.; Tholen, E.; Petersen, B. *et al.* (2011). In: *Anal. Chem.* 83 (17), S. 6785–6791. https://doi.org/10.1021/ac201465q.

García-Regueiro, J. A.; Diaz, I. (1989). In: *Meat Science* 25 (4), S. 307–316. https://doi.org/10.1016/0309-1740(89)90048-x.

Groβe-Brinkhaus, C.; Storck, L. C.; Frieden, L.; Neuhoff, C.; Schellander, K. *et al.* (2015). In: *BMC Genet.* 16, S. 36. https://doi.org/10.1186/s12863-015-0194-z.

Hartung, J.; Knapp, G.; Sinha, B. K. (2008): Statistical Meta-Analysis with Applications: Wiley.

Lundström, K.; Matthews, K. R.; Haugen, J.-E. (2009). In: *animal* 3 (11), S. 1497–1507. https://doi.org/10.1017/S1751731109990693.

Moe, M.; Lien, S.; Aasmundstad, T.; Meuwissen, T. H. E.; Hansen, M. H. S. *et al.* (2009). In: *BMC Genet.* 10, S. 32. https://doi.org/10.1186/1471-2156-10-32.

Mörlein, D.; Lungershausen, M.; Steinke, K.; Sharifi, A. R.; Knorr, C. (2012). In: *Meat Sci.* 92 (4), S. 739–744. https://doi.org/10.1016/j.meatsci.2012.06.031.

Patterson, R. L. S. (1968). In: *J. Sci. Food Agric.* (19).

Robic, A.; Larzul, C.; Bonneau, M. (2008). In: *Genet. Sel. Evol.* 40 (1), S. 129–143. https://doi.org/10.1051/gse:2007040.

Skinner, T. M.; Doran, E.; McGivan, J. D.; Haley, C. S.; Archibald, A. L. (2005). In: *Anim. Genet.* 36 (5), S. 417–422. https://doi.org/10.1111/j.1365-2052.2005.01342.x.

Squires, E. J.; Lundström, K. (1997). In: *J. Anim. Sci.* (75), S. 2506–2511.

Stefanski, V.; Labussière, E.; Millet, S.; Weiler, U. (2018). In: Wageningen Academic Publishers (Hg.). 69th Annual Meeting of the European Federation of Animal Science. Dubrovnik, 27.08.2018-31.08.2018. European Federation of Animal Science. Wageningen, S. 229.

Wiercinska, P.; Lou, Y.; Squires, E. J. (2012). In: *animal* 6 (5), S. 834–845. https://doi.org/10.1017/S1751731111002175.

Yokoyama, M. T.; Carlson, J. R. (1979). In: *Am. J. Clin. Nutr.* (32), S. 173–178.

Part 44

Free communications

803. A data-architecture to monitor and collect cow-individual methane emissions real-time from commercial dairy farms

I. Hulsegge*, D. Schokker, G.G.J. Klandermans and C. Kamphuis

Wageningen University & Research, Animal Breeding and Genomics, P.O. Box 338, 6700 AH Wageningen, the Netherlands; ina.hulsegge@wur.nl

Abstract

Real-time monitoring and collecting cow-individual methane emissions will help individual farmers in identifying which management interventions affect methane emissions, and the dairy sector in developing strategies to reduce methane emissions. To achieve this real-time monitoring and collecting of cow-individual methane data, an architecture was developed that automates both aspects in a commercial setting. The first step in the data architecture was to collect the analogue signal of sniffers real-time, and transform this signal into methane concentrations. Because these sniffers do not record cow identification, the second step was to combine the methane concentration data with real-time cow identification. The cow identification is retrieved by automatically detecting the ear tag in a live-stream video footage, to subsequently 'read' the 4-digit number of that ear tag. With the deployed architecture, we developed a universal, low-tech, and scalable method to collect cow-individual sensor data real-time for technologies that do not record cow identification by default.

Background

Precision Livestock Farming (PLF) includes the integration and interpretation of relevant sensor information leading to new opportunities in the management and breeding of livestock through continuous real-time monitoring of health, behaviour, production, reproduction, and environmental impact (Berckmans 2017; Herlin *et al.* 2021). PLF technologies are entering the livestock domain in a fast pace, offering the potential to collect information that was not possible a decade ago (Buller *et al.* 2020). Currently, PLF technologies are available to monitor greenhouse gasses (GHG) emissions at cow-individual level (de Haas *et al.* 2021). One of these technologies is a non-dispersive infrared unit or 'sniffer' to measure GHG emission (de Haas *et al.* 2021). These relatively cheap sniffers offer opportunities to meet the Dutch government's objective to reduce GHG emissions by 49% by 2030, compared to 1990 levels (Goverment of the Netherlands 2021). Real-time monitoring of cow-individual methane emissions will help in identifying which management interventions affect methane emissions (e.g. new feeding ration), and in developing strategies to reduce methane emissions in the dairy sector (e.g. through breeding). Currently, the methane emission measurements with sniffers take place in the milking robot as described by Huhtanen *et al.* (2015), which enables information to be collected on each individual cow. The sniffer monitors a constant analogue signal that is linearly related to gas concentrations, which can be transformed into a methane and carbon dioxide concentration. The sniffer measures continuously, and the assumption is that when a cow is in the milking robot and eating her concentrates while being milked, that the sniffer is measuring that cow's methane concentration. At a later stage, sniffer data and robot data are combined to add cow identification to sniffer data. To collect sniffer information real-time, the first step of our data-architecture involves the connection of the sniffer to an Arduino, a small and inexpensive computer and accompanying software to program this computer (https://www.arduino.cc). The Arduino is programmed to push the sniffer data to the cloud service platform Azure (https://azure.microsoft.com) via the Internet of Things (IoT) network. This IoT system was selected to avoid potential problems with low coverage of e.g. a WIFI signal (Schokker *et al.* 2020), allowing future upscaling independent of WIFI coverage. In Azure, an IoT hub was built which was subscribed to the data stream. This means that this IoT hub was used to acquire and ingest (load), and transform the raw sniffer analogue signal into a methane concentration data point. At this stage, we

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_803, © I. Hulsegge *et al.* 2022

thus have cow-individual methane concentrations measured and collected in real-time. However, sniffers do not record cow identification, and thus, the next step of our data-architecture involved the real-time information which cow the sniffer is actually monitoring.

Automatic and real-time animal identification

Animal identification, which refers to the process of accurately recognizing individual animals, plays an important role in automatic and real-time analysis of activities and productivity in PLF. All cows in the Netherlands are uniquely identified with a lifetime registration number at birth. This number is printed on the ear tags that cows are obligatory to wear. Four digits of this ear tag are highlighted using a larger font size. These four digits are typically used by farmers as the cow's work number. In addition to these mandatory ear tags, Radio Frequency Identification Device (RFID) transponders that are connected to the neck-collar of the cow are also a method readily used to identify cows (e.g. RFID is used by milking robots for cow identification). However, not all farmers have milking robots. Also, milking robots may store the transponder identification together with a datetime stamp, but accessing and matching this identification to other data is not always straightforward, and therefore error-prone. Identification of animals by automatically reading the ear tags may be a more universal solution in precision dairy farming. This automation can be done using video footage or camera images. Methods for individual cow identification system using ear tags were described by Ilestrand (2017) and Zin *et al.* (2020). These methods have been applied on a static dataset of previously collected images. But for PLF, a system is required that continuously reads the ear tag number and makes this cow identification available real-time. The use of video is an attractive approach to achieve this because it offers 24/7 monitoring, and is a relatively cheap investment. We developed a system that enables real-time cow identification through the use of vision technology and artificial intelligence (AI). The system was developed and implemented in Python 3.8 using the Open Computer Vision Library (OpenCV 4.1.2), an existing library for the analysis of image or video datasets in AI and deep learning (Culjak *et al.* 2012). The system 'detects' the ear tag area from a live-stream, 'segments' the ear tag area from the live-stream, 'improves' the image quality using image processing techniques, 'rotates' the ear-tag such that the baseline of the tag is horizontal in the image, and subsequently 'reads' the 4-digit number of the ear-tag by applying a number recognition model. This whole process is depicted in Figure 1. The number recognition model was developed by transfer learning of a pre-trained household number model (Goodfellow *et al.* 2014), a Convolutional Neural Network (CNN) classifier. Recognizing house numbers is a quite similar problem to recognizing ear tag numbers. The house number model is retrained with ear tag numbers, which increased the accuracy of a test set from 80% using the pre-trained model to 87% using the retrained model. The developed data-infrastructure consists of a video station involving a Raspberry Pi 4 Model B (a small, inexpensive computer developed by the Raspberry Pi Foundation in the United Kingdom), connected to a security camera (DS-2CD2T43G0) via ethernet (LAN) connection. The video station is installed in the milking robot as a place where also the sniffer sensor is continuously recording the analogue signal. The Raspberry Pi is programmed to collect, and process the live-stream of the camera, to push just the ear tag number together with a timestamp to the Arduino where the methane data is also collected. From the Arduino the methane data, ear tag number and timestamp are streamed onto the Azure cloud platform using Azure Streaming Analytics (Figure 1). Streaming Analytics is a querying, alerting and monitoring tool that monitors data streams.

Challenges

At this stage, sniffer and ear tag data streams come-in simultaneously and (near) real-time in the cloud platform. However there are still challenges to overcome. The next step will be to integrate and synchronize the sniffer- and ear tag data streams real-time. Furthermore, during a cow's visit to the milking robot the ear tag is not always (clearly) visible, especially if the cow is moving her head. As a result, ear tag recognition is not robust enough to deduct the starting- and/or end-time of the robot visit. To overcome this problem

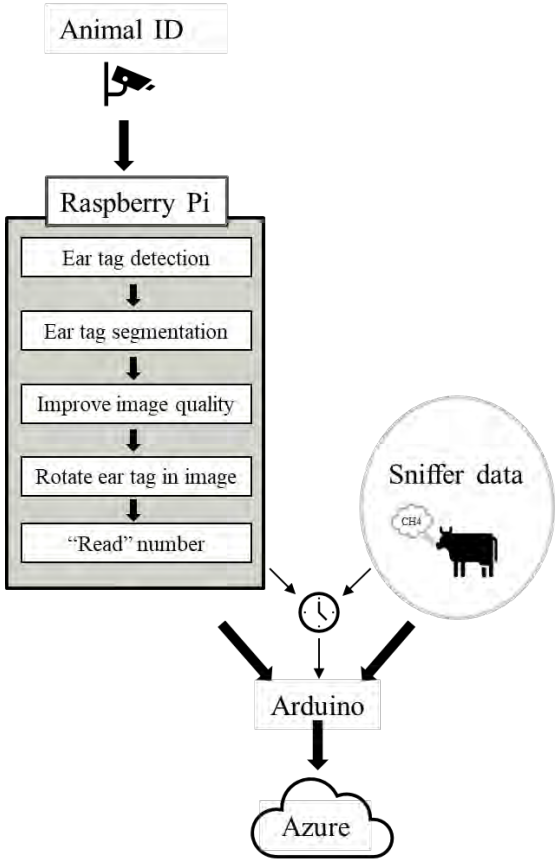


Figure 1. Schematic overview infrastructure.

template matching will be used to check if there is a cow in the milking robot, and to signal the starting and endpoint of a cow visit. Improvement in the downstream process, after 'reading' the 4 digits need also to be investigated. One of current investigation is to improve the number recognition by comparing consecutive predicted numbers in time. Such a system could be developed by using a majority vote of the prediction for in a sliding window of consecutive frames. Other improvements could be made by comparing the predicted number with a list of all cow ear tag numbers present in the herd.

Conclusions

A universal, low-tech and scalable method to collect cow-individual sensor data real-time for technologies that do not record cow identification by default is developed and deployed in a commercial setting. The system can be applied anywhere on a farm, not just in the milking robot as described in this paper. The method offers perspective to be used by other PLF techniques that do not register animal IDs by default. The data architecture is flexible in accepting additional data streams, e.g. other sensor data, but also animal and farm data that are often stored in separate farm management systems. Another part of the flexibility of the infrastructure links to the scalability, i.e. handling data from new farms in case the number of farms where methane is monitored, increases (Schokker *et al.* 2020).

Acknowledgements

This research was conducted by Wageningen Livestock Research, commissioned and funded by the Dutch Ministry of Agriculture, Nature and Food Quality, within the framework of Policy Support Research theme 'Data Driven & High Tech' (project number KB-38-001-004).

References

Berckmans D. (2017) Animal Front. 7(1): 6-11. https://doi.org/10.2527/af.2017.0102

Buller H., Blokhuis H., Lokhorst K., Silberberg M. and Veissier I. (2020) Animals 10(10): 1-12. https://doi.org/10.3390/ani10101779

Culjak I., Abram D., Pribanic T., Dzapo H. and Cifrek M. (2012) A brief introduction to OpenCV, pp. 1725-1730 in *35th MIPRO*, Opatija, Croatia.

De Haas Y., Veerkamp R.F., De Jong G. and Aldridge M.N. (2021) Animal. https://doi.org/10.1016/j.animal.2021.100294

Goodfellow I.J., Bulatov Y., Ibarz J., Arnoud S. and Shet V. (2014) Multi-digit number recognition from street view imagery using deep convolutional neural networks, pp. in *2nd ICLR*. International Conference on Learning Representations, ICLR, Banff, Canada.

Goverment of the Netherlands (2021) Climate policy, pp. Available at: https://www.government.nl/topics/climate-change/climate-policy.

Herlin A., Brunberg E., Hultgren J., Högberg N., Rydberg A. *et al.* (2021) Animals 11(3): 1-20. https://doi.org/10.3390/ani11030829

Huhtanen P., Cabezas-Garcia E.H., Utsumi S. and Zimmerman S. (2015) Comparison of methods to determine methane emissions from dairy cows in farm conditions, pp. 3394-3409 in *J. Dairy Sci.* Elsevier Inc.

Ilestrand M. (2017) Automatic Eartag Recognition on Dairy Cows in Real Barn Environment, pp. 55.

Schokker D., Mollenhorst H., Seigers G., De Haas Y., Veerkamp R.F. *et al.* (2020) Real-Time Visualization of Methane Emission at Commercial Dairy Farms, pp. 194-200 in *13th IFIP WG 5.11 International Symposium on Environmental Software Systems, ISESS 2020*, edited by Athanasiadis I.N., Knibbe W.J., Frysinger S.P. and Schimak G. Springer.

Zin T.T., Pwint M.Z., Seint P.T., Thant S., Misawa S. *et al.* (2020) Sensors 20(12): 1-18. https://doi.org/10.3390/s20123564

804. An ontology approach for the collection and management of genetic traits and milk protein data

C. Franzoni Migliorati[1*], J.B.C.H.M. van Kaam[1], R. Finocchiaro[1], C.A. Lombardi[1], M. Marusi[1] and M. Cassandro[1,2]

[1]Associazione Nazionale Allevatori della Razza Frisona, Bruna e Jersey Italiana (ANAFIBJ), Via Bergamo 192, 26100 Cremona (CR), Italy; [2]Department of Agronomy, Food, Natural resources, Animals and Environment (DAFNAE), University of Padova, Viale dell'Università 16, 35020 Legnaro (PD), Italy; chiarafranzonimigliorati@anafi.it

Abstract

The aim of this study is to propose a new approach for the collection of data from genetic traits and milk proteins and to manage this information in a unique virtual environment. Information about genetic traits and milk proteins can come from different sources and can be determined using different types of genetic tests. All input data are recorded in the database with a priority that is defined by the database update algorithm, taking into account the source of the data and the type of genetic test. In this way, the most reliable information is considered official herd book data, according to the established priorities. In conclusion, this new management of genetic traits and milk proteins aims to lay the foundations for a process of harmonization that supports international exchanges of data and that makes the interpretation of genetic traits clearer, in particular for breeders and for the livestock industry.

Introduction

In recent years, the advent of genomic analysis, new technologies and advanced automated systems have completely changed the way we approach livestock data, allowing an increasing exchange of genetic trait information on a global scale (Gutierrez-Reinoso *et al.*, 2021). Today millions of data on genetic traits (monogenetic traits) are available but only through an efficient management and sharing system can this information be a real resource for breeders, for the livestock industry and for the international scientific community (Niklitschek *et al.*, 2016). The creation of this new database for the collection and management of data on genetic traits and milk proteins genetic variants, arises from the need to make this information always available and clear, in order to promote the exchange of data and harmonization with international standards. The management of genetic traits implemented until now allowed to record only the diagnosis of the genetic trait (e.g. tested carrier or free), neglecting other information, such as the source and the historicization of the type of genetic tests, which can have an important influence on the reliability of the data. Aim of this project is to design, with an ontology approach, a new database for management and tracking of the source of the data and the type of analyses.

Materials & methods

The new database (DB) is designed to manage data on genetic traits and milk proteins in a unified virtual environment, using an ontology approach. An ontological approach looks at the things (entities) the data is about and uses them as the basis for the structure of the data. If the things that are important to the activities and the relationships between them are correctly identified, then the data model is universally useful (Cui and O'Brien, 2000). Being a relational database the key element for its operation are the entities. The DB entities have specific properties some of which relate entities to each other.

The entities. The ontology DB has five entities: Gene entity, Diagnosis entity, Source entity, Analysis entity and Exposition sequence entity. The entity Gene contains the list of genetic traits and milk proteins and identifies them with a unique progressive code. The genetic traits are divided into two categories according

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_804, © C. Franzoni Migliorati *et al.* 2022

to the type of genetic test: the Gene category is assigned to diagnoses resulting from direct tests while the Haplotype category is assigned to diagnoses resulting from indirect tests. At the moment 30 genes and 21 haplotypes are included. The diagnosis entity lists all possible diagnoses for each genetic trait and milk protein and establishes the nomenclature with which the result of the analysis is identified. Both the international nomenclature of WHFF (World Holstein Friesian Federation) and the nomenclature used today in Italy are provided. A total of 164 diagnoses are expected, including those identified by imputation. The Source provides a list of sources from which data can be received and each of them is assigned a unique code. Each source is assigned a priority level; a higher priority is given to more reliable data sources (Table 1).

In the entity Analysis all incoming analyses are recorded and each is assigned a unique code. For each incoming analysis, the serial number of the animal, the unique code related to the genetic trait or milk protein, the unique code of the diagnosis, the unique code of the source are recorded and it is specified when the analysis comes from imputation. The entity Exposition sequence establishes the sequence in which diagnoses are to be displayed on official documents. The results of genetic traits will be presented first, then coat colour genes followed by milk proteins.

The database update algorithm. Each animal can have one or more tests for the same genetic character. All of them must be entered in the DB with the priority determined at the time of entry. For each test to be entered, the algorithm considers 3 different priority levels:

1. Direct/indirect test: for tests involving the same genetic traits the gene category always has priority over the haplotype category whatever the source.
2. Sources: analysis of the same category (gene or haplotype) the algorithm establishes priority according to the data source.
3. Imputation: when the analysis is determined without having the animals' genome but based on ancestry, it is considered 'from imputation'. These analyses are identified in the Analysis entity and the algorithm considers them with the lowest priority level so that they are not exposed on official documents.

Results

The new ontology DB is structured to handle millions of records and tests to verify its performance were carried out by loading batches of data. As shown in Table 2, the data loading tests included 497,097 records from genomic files and 110,591 records from the old genetic traits management system, for a total of 607,688 incoming data. 598,815 analyses were correctly registered in the database (488,979 from genomics and 109,836 from the old management system). There were 8,873 records that the update algorithm correctly did not include in the database because they were already present with the same source and the same diagnosis. 2,743 records were indicated with a different result from the one already in the database. In this case, the algorithm inserts the data but indicates as official analysis the one with the most reliable source, according to the established priorities. Correctly, for the 2,737 records from genomic files the DB reports

Table 1. Source and priority level.

Source	Priority level
Specific test by accredited laboratories	100
Genomic test files	90
Official documents by foreign herdbooks	70
Old genetic traits management system	50
Unofficial documents or self-declarations	30

Table 2. Results of data uploading tests.

Source	Genomic files			Old genetic traits management system	Total
	Marker-based tests	Haplotypes	Both		
No. of input data	188,627	308,470	497,097	110,591	607,688
No. of data entered	180,509	308,470	488,979	109,836	598,815
No. of data not entered because already present	8,118	0	8,118	755	8,873
No. of data notified for different diagnosis in DB	2,737	0	2,737	6	2,743

that the official diagnosis has been changed because a higher priority is assigned to that source. While for the 6 records derived from the old system was notified that there is a different diagnosis but the official diagnosis is still the one in the DB. The haplotypes were all correctly entered into the database (308,470 records) without any analyses already in the database or analyses notified due to different results.

Discussion

The new ontology database makes it possible to keep track of all the tests registered for the same animal and for each type of test it is possible to consult, through a web application: the genetic trait, the result, the source, the type of test, the priority and the date of inclusion in the database. Priority levels have been designed to represent the reliability of the data based on the type of genetic test and the source of the data. Taking into account the scientific evidence, types of direct analysis (marker-based tests) have been considered more reliable (and therefore with higher priority) than indirect tests that do not detect the causal allele, such as haplotypes (Sonstegard *et al.*, 2013). Even lower priority has been assigned to imputation analyses that are determined without having the animals' genome but based on ancestry (Nicolazzi *et al.*, 2013). Concerning the source of the incoming data, it was decided to give the highest priority to genomic test files (Italian or from data exchanges with other countries) and to the results that come from official documents issued by internationally accredited laboratories because these sources certify the scientific evidence of the execution of the test. It should be emphasized that the diagnoses derived from genomic files are considered more reliable also because our procedure, before uploading the data in DB, provides Mendelian controls that compare, for biallelic genes, the genotype of the animal with that of the verified parents (if available). Slightly lower priority has been assigned to the information reported on official documents by foreign herdbooks because in this case we do not have an analysis or a genome of the animal but only an information and human error cannot be excluded. Data from the old genetic traits management system were given lower priority because the source cannot be traced. Tracking all analyses for the same genetic character allows the analysis with the highest priority, and therefore the most reliable, to be automatically considered official. In addition, the recording of all incoming data allows immediate and automatic identification of any conflicting results for the same animal. The nomenclature used to identify genetic traits was a central element of the project; the new management system includes both the nomenclature used today in Italy and the international WHFF standard with the aim of encouraging future use of the international nomenclature in Italy. Furthermore, the algorithm of the database makes it possible to determine the genetic traits that must be included in official documents and the expression codes with which they must be displayed. Analyses that are not to be displayed in official documents (e.g. non-royalty free genetic traits, haplotypes or imputation analyses) are determined, but it is a flexible system that allows them to be exposed in the future if necessary. The flexibility of the system is decisive because it is designed to allow technical staff to choose and change, autonomously and without acting on the software, but only a change into the reference DB table, which genetic traits are to be displayed on official documents and

on our online consultation applications, ensuring that the order in which genetic traits are displayed can be chosen and changed to highlight, for example, the diagnoses of genetic traits deemed most interesting.

Conclusions

In conclusion, the new ontology data management system enables automatic acquisition of information on genetic traits and reduces human errors as much as possible. The reliability of the data is the key element to ensure a process of international harmonisation and data exchange for the benefit of the breeders, the livestock industry and the scientific community.

References

Cui, Z. and O'Brien, P. (2000). Proc. of the 33rd Hawaii International Conference on System Sciences, Maui, Hawaii.

Gutierrez-Reinoso, M.A., Aponte, P.M., Garcia-Herreros, M. (2021). Animals 11:599. https://doi.org/10.3390/ani11030599

Nicolazzi, E.L., Biffani, S., Jansen, G. (2013). J. Dairy Sci. 96:2649-2653. https://doi.org/10.3168/jds.2012-6062

Niklitschek, S., Lama, J., Trejo, C., Wickham, B., Burke, M., *et al.* (2016). Proc. of 40th ICAR Biennial Session, Puerto Varas, Chile.

Sonstegard, T.S., Cole, J.B., VanRaden, P.M., *et al.* (2013). Plos ONE 8:e54872 https://doi.org/10.1371/journal.pone.0054872

World Holstein Friesian Federation (WHFF), Standardized Labelling for Monogenetic Trait Coding (2019), Available at: http://www.whff.info/documentation/documents/StandardizeLabellingofGeneticTraitCoding2019.pdf

805. Genotyping errors in SNP markers: types, consequences, and a novel detection method

W.M. Rauw*, M.G. Gil, O. Torres and L. Gomez-Raya

Departamento de Mejóra Genética Animal, Instituto Nacional de Investigación y Tecnología Agraria y Alimentaria (INIA-CSIC), Ctra de la Coruña km 7.5, 28040 Madrid, Spain; wendy.rauw@csic.es

Abstract

Genotyping errors consist in misclassification of genotypes at a genetic marker. Most research using genotyping arrays only report generic screening of SNPs due to low calling rate, departures of Hardy-Weinberg equilibrium or because of reduced allele frequency (Minimum Allele Frequency, MAF). Detection of genotyping errors for a particular experimental dataset remains generally unreported. We propose to relate the probability of genotyping errors to the probability of the rate of rejection of true parent-offspring. Only single nucleotide polymorphisms (SNPs) at intermediate frequencies are useful to determine genotyping errors. Genotyping errors are detected mostly for SNPs at allele frequencies between 0.18 and 0.50. Our method can estimate the probability of generic (across all SNPs) and specific (affecting some SNPs) genotyping errors. For the rate of rejection of 0.000735 in a porcine HD array we estimated a genotyping error of 6 out of 1000 SNPs.

Introduction

SNPs are distributed throughout the genome at a rate of approximately 1 in 1000 base pairs, and may be located within a gene locus (thus altering amino acids), but are mostly located in the non-coding DNA between genes (Shastry, 2002). Because the dense coverage of the genome by SNPs, they can be used for linkage disequilibrium mapping of SNP markers that are in close proximity to (or are themselves) the causal mutation. To accomplish this, the vast amount of SNPs detected by recent cost-effective advances in next-generation sequencing (NGS) technologies can be used in the development of SNP arrays that are used to interrogate the entire genome for associations between the phenotype and around 50.000 to one million SNPs at a time in Genome Wide Association Studies (GWAS). In genomic selection, estimates of the effects of all genotyped markers across the genome are used to estimate an individual's breeding value, rather than estimates only of the effects of significant markers. In addition, pedigree-based relationships among animals can be replaced by genomic relationships that are computed based on SNP genotypes in genomic best linear unbiased prediction (GBLUP). Virtually every genetic study includes genotyping errors, quantified by the Quality scores probability of error provided by genomics companies, resulting from calling algorithms that misidentify and misclassify the individual's genotype. It is likely that genotyping errors vary between experiments due to variation in DNA quality or problems associated to lab sample handling; genotyping errors are usually not reported in scientific studies. We define genotyping errors as the misclassification of genotypes. Our goal is to determine the probability of genotyping errors in large scale genotyping experiments.

Calculation of the genotyping errors in genetic research

In the last decades, the number of animals being genotyped as well as the number of SNPs typed per animal has increased dramatically. Genotyping and/or DNA sequencing is now the gold standard for genetic research. For example, a porcine High Density SNP array provided by ThermoFisher has more than 600K SNPs, however, genotyping errors reduce the true total number of available SNPs for data analysis. After cleaning data based on calling rate, genotyping errors are typically approached by excluding SNPs that deviate from Hardy-Weinberg Equilibrium and those with minor allele frequency (MAF) (Fialkowski *et al.*, 2019).

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_805, © W.M. Rauw *et al.* 2022

Two types of genotyping errors can be distinguished: generic errors affect all SNPs randomly and equally, while specific errors are inherent to the marker itself, i.e. they do not affect all SNPs equally. The latter may influence GWAS when a particular erroneously typed SNP rejects a truly existing association to a causal mutation, while generic errors may particularly influence genomic selection accuracy because it results in erroneous estimation of the effects associated with an individual's genotype at the affected SNP markers (the impact on genomic evaluations should be small if a genotyping error probability is not large and a large number of SNPs is used). It is possible to test for genotyping errors through repeated testing with the same or different molecular genotyping methods. However, this is unpractical for most routine genotyping operations. Alternatively, genotyping errors can be detected using inconsistencies in Mendelian inheritance between parents and offspring (not common for most GWAS studies). There are two possibilities to detect erroneously genotyped parent-offspring: a) detection based on rejection of a parent-offspring relationship if one parent and its offspring are homozygous for alternative alleles (Gomez-Raya *et al.*, 2022), and 2) detection based on rejection of a parent-offspring relationship if both parents are homozygous for the same allele while progeny is heterozygous. We will focus on the first, because it is the easiest and more common situation.

Inconsistent Mendelian inheritance between a parent and its offspring

In livestock production, genotyped individuals in a population are often related. A first-degree relative is an individual that shares ~50% of their alleles with a particular other individual: parent-offspring and full-sibs. In most cases, individuals with first-degree parent-offspring relationships are part of the same selection candidates in genomic selection or crossbreeding programs. Under strict Mendelian inheritance, individuals with a first-degree parent-offspring relationship cannot be homozygous for alternative alleles at a given locus. For example, an individual offspring cannot be homozygous CC at a locus, when a parent is GG at the same locus. Following Gomez-Raya *et al.* (2022), inconsistent Mendelian inheritance between parents and offspring can be used to assign true maternities and paternities based on the rate of rejection of parent-offspring relationships. Furthermore, this method can be used to detect genotyping errors after true parent-offspring relationships are resolved, i.e. two individuals appear to be homozygous for alternative alleles at a SNP marker when in reality they are not.

Rate of rejection of first-degree relationships and probability of genotyping errors

The probability of rejecting a true first-degree relationship at a locus is related to genotyping errors and can easily be measured when parents and offspring co-exist in a genotyped data set (Gomez-Raya *et al.*, 2022). A model of genotyping errors for parents with genotype *AA* is given in Table 1; models for parents with genotypes *Aa* or *aa* are constructed in the same fashion. Parents with genotype *AA* will be at a population frequency of f_{AA}. The probability of a genotyping error, *ge*, results in erroneously observed genotypes *AA*, *Aa,* and *aa*, at frequencies f_{AA}, f_{Aa}, and f_{aa}, respectively. Note that this model assumes that erroneous genotyping occurs given the probability of the genotype frequencies in the population; that is, misclassification of genotypes depends on the population frequency. True *AA* parents produce two types of offspring: *AA* at frequency f_A, and *Aa* at frequency f_a. True *AA* offspring with a probability of genotyping error *ge* will appear in the population at genotype frequencies $ge\,f_{AA}$, $ge\,f_{Aa}$, and $ge\,f_{aa}$. Using probabilities in Table 1, the probability of rejection for parents with true genotype *AA* is:

$$r_{AA} = gef_{aa}[(1-ge) + 2ge\ f_{AA} + (1-ge)f_A\ + gef_Af_{AA}]$$

The same arguments can be used to relate rate of rejection of parents with true genotypes *Aa* and *aa*:

$$r_{Aa} = gef_{AA}[ge\ f_{aa} +\ 0.5f_a(1-ge)] + gef_{aa}[ge\ f_{AA} +\ 0.5f_A(1-ge)]$$

Table 1. Modeling genotyping errors for parents with genotype *AA*. Frequencies of genotypes are between brackets; *ge*: genotyping error or probability of incorrect genotype.

True genotype parent (freq)	Prob. of error	Observed genotype parent (freq)	True genotype offspring (freq)	Prob. of error	Observed genotype offspring (freq)
AA (f_{AA})	*1-ge*	*AA* (1)	*AA* (f_A)	*1-ge*	*AA* (1)
	ge	*AA* (f_{AA})		*ge*	*AA* (f_{AA})
	ge	*Aa* (f_{Aa})		*ge*	*Aa* (f_{Aa})
	ge	*aa* (f_{aa})		*ge*	*aa* (f_{aa})
			Aa (f_a)	*1-ge*	*AA*
				ge	*AA* (f_{AA})
				ge	*Aa* (f_{Aa})
				ge	*aa* (f_{aa})

$$r_{aa} = ge f_{AA}[(ge\ f_{aa}) + (1 - ge) f_a\ + (1 - ge) + ge\ f_{aa}]$$

The joint rate of rejection for the three parental genotypes is $r_r = r_{AA} f_{AA} + r_{Aa} f_{Aa} + r_{aa} f_{aa}$. Figure 1 shows that the relationship between the rate of rejection of first-degree relationships in Equation 1 depends strongly on the allele frequency. Only SNPs at intermediate allele frequencies are likely to reject true parent-offspring relationships. The rate of rejection is a lower bound of the true genotyping error.

In an example of nine Duroc sires, 55 Iberian dams, and 214 Duroc × Iberian offspring genotyped with the HD porcine Affymetrix array (658,692 SNPs), true paternity of 75 and maternity of 85 piglets were established with a rate of rejection of 0.000735 (Gomez-Raya *et al.*, 2022). Figure 2B shows that those SNPs do not correspond to the genotyping errors detected by the Inconsistent Mendelian Inheritance method, instead, genotyping errors are detected mostly for SNPs at allele frequencies between 0.18 and 0.50. On the contrary, the distribution of the allele frequencies of 592,053 autosomal SNPs in the array shows a typical U distribution, with a majority of SNPs close to fixation (Figure 2A).

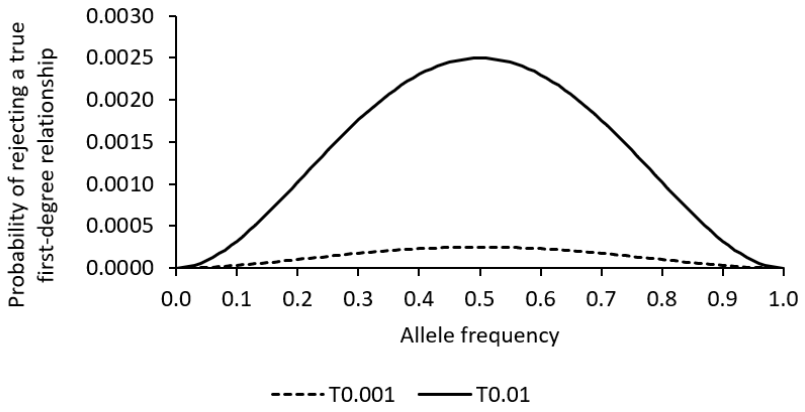


Figure 1. The probability of rejecting a true first-degree relationship as a function of allele frequency when the true probability of genotyping error is 0.001 or 0.01%.

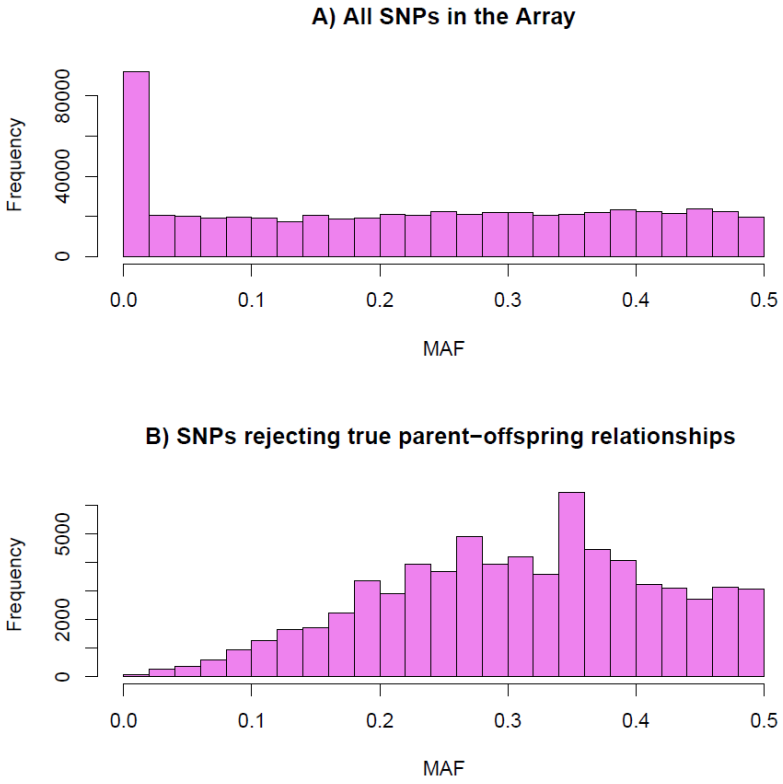


Figure 2. Distribution of SNPs as a function of minimum allele frequency (MAF). (A) Allele frequency distribution of 592,053 autosomal SNPs in the array; (B) Allele frequency distribution of 69,882 SNPs rejecting true parent-offspring relationships.

The generic genotyping error of an array can be computed by setting *ge* at an initial value and by computing the average r_r using all SNPs and their frequencies in the array. If the average r_r differs from that observed in our dataset, then *ge* is increased or reduced until the average r_r approaches the approximate same value as the actual estimated value in data r_r. For the rate of rejection of 0.000735 we estimated a *ge* = 0.006. That is, six out 1000 genotypes were genotyping errors. In addition to the generic genotyping errors we can identify specific SNPs that repeatedly reject true parent-offspring relationships. In our analysis we identified 7,744 SNPs rejecting ≥5 true parent-offspring relationships (Gomez-Raya *et al.*, 2022). Discarding SNPs based merely on MAFs implies profiling based on the fact that clustering algorithms are more prone to error with SNPs at low MAF. Instead, SNPs detected by the Inconsistent Mendelian Inheritance method should be the result of true genotyping errors.

References

Fialkowski A., Beasley T.M., Tiwari H.K. (2019). Multifactorial Inheritance and Complex Diseases. In: Pyeritz R.E., Korf B.R., Grody W.W. (eds) Emery and Rimoin's Principles and Practice of Medical Genetics and Genomics: Foundations. Academic Press, London, UK, pp. 323-358.

Gomez-Raya L., Gómez Izquierdo E., de Mercado de la Peña E., García Ruiz F., Rauw W.M. (2022). BMC Genomic Data: accepted. https://doi.org/10.21203/rs.3.rs-861738/v1

Shastry B.S. (2002). J. Hum. Genet. 47:561-566. https://doi.org/10.1007/s100380200086

806. LncRNAs variability in skeletal muscle of Polish Landrace boars

B. Hofman[1], J. Szyda[1,2], M. Frąszczak[1*] and M. Mielczarek[1,2]

[1]Biostatistics group, Department of Genetics, Wroclaw University of Environmental and Life Sciences, Kożuchowska 7, 51-631 Wroclaw, Poland; [2]National Research Institute of Animal Production, Krakowska 1, 32-083 Balice, Poland; magdalena.fraszczak@upwr.edu.pl

Abstract

Recently, numerous studies including various tissues have been carried out on long non coding RNAs (lncRNAs) but still, lncRNAs variability has not yet been fully understood. In this study, we characterized the interindividual variability of lncRNAs in pigs, in the context of number, length, and expression. Six transcriptomes of Polish Landrace boars (S1-S6), including half-brothers (S1-S3), were investigated using bioinformatics (lncRNAs identification) and statistical methods (lncRNAs variability). The number of lncRNA ranged from 1,289 to 3,500 per animal. The number, length, and expression of lncRNAs significantly varied between individuals. S1 exhibited significantly higher, while S3 and S6 exhibited lower expression than other animals. Based on the pairwise analysis two groups with no difference in expression level were distinguished: S2, S4, and S5 as well as S3 and S6. No consistent pattern has been found between pairs of half-brothers. This work might contribute to further research leading to understanding the role of lncRNA variability.

Introduction

Long non coding RNAs (lncRNAs) are RNA molecules longer than 200 nucleotides that are not translated into functional proteins. According to Statello *et al.* (2020) lncRNAs play a role in gene expression regulation by affecting transcription, DNA replication, and the response to DNA damage and repair. Other functions involve structural and regulatory roles in mRNA metabolism, including splicing, turnover and translation, as well as signaling pathways. Since alterations in lncRNAs expression affect important cellular functions they have the potential to contribute to a phenotypic variation including numerous diseases (Statello *et al.* 2020). Domestic animals are an essential resource in studies involving RNA, including lncRNAs (Lagarrigue *et al.* 2021). Still, lncRNAs have not yet been fully characterised in pig genomes. In our study, we investigated interindividual variability of lncRNAs in pigs, in the context of number, length, and expression.

Materials & methods

RNA-seq data. Transcriptomes were collected from muscle tissue (*longissimus dorsi*) samples, which were taken from six Polish Landrace boars (S1-S6), where three boars were half-brothers (S1-S3). DNA isolation and reverse transcription were performed to obtain DNA and cDNA templates required for library preparation. Libraries were generated from 50 ng of RNA fragments using the KAPA RNA HyperPrep Library Preparation Kit (Roche) and then sequenced using the Illumina NovaSeq 6000 in a paired-end mode with the designated single read length of 101 bp. The number of raw reads per individual ranged from 120,802,823 to 168,508,615.

Bioinformatic pipeline. Quality control of raw reads was carried out for each sample using the FASTQC software (Andrews, 2020). Data cleaning was done by Trimmomatic (Bolger *et al.* 2014). Edited reads were then mapped to the Sscrofa11.1 reference genome (Warr *et al.* 2020) using HISAT2 (Kim *et al.* 2019). The Samtools package (Li *et al.* 2009) was used for post-alignment processing. The next step comprised the quantification of gene expression with the StringTie software (Kovaka *et al.* 2019) based on information from aligned reads, reference genome, and reference gene annotation (Warr *et al.*, 2020), which resulted in

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_806, © B. Hofman *et al.* 2022

transcripts abundances expressed in Fragments per Kilobase Million (FPKM). Transcripts were then filtered out to retain only putative lncRNA by applying the following sequence of removal criteria: (1) transcripts containing a single exon using the gffread tool from GFF Utilities (Pertea and Pertea, 2020); (2) transcripts shorter than 200 bp or longer than 10,000 bp; (3) transcripts consisting of at least 50% of low complexity regions identified using the sDUST software (Morgulis *et al.* 2006); (4) transcripts with an expression lower than 0.3 FPKM (Cai *et al.* 2018); (5) transcripts containing known protein-coding domains; and (6) transcripts with high coding potential using the Coding Potential Assessment Tool (CPAT) software (Wang *et al.* 2013). Among the remaining transcripts that represent lncRNA, the novel lncRNAs were identified by comparing sequences with the non-coding RNA sequence database RNA-Central (Petrov *et al.* 2015).

Interindividual variability of number, length, and expression level of lncRNAs. The χ^2 of the goodness of fit test was used to test whether the *number* of lncRNA is uniformly distributed between individuals. The normality of the distribution of both the length of lncRNA and the expression level measured by FPKM was assessed with the Kolmogorov-Lilliefors test. The underlying null hypothesis was defined as: lncRNA length/expression follows a normal distribution. Due to the non-normality of considered data, non-parametric methods were applied. The underlying null hypothesis assuming that the distributions of lncRNA *lengths* were the same for all animals was verified using the Kruskal-Wallis test, which is a rank test dedicated to independent samples. Then the Dunn test was used to compare the distribution of lncRNAs in pairs of animals. The nonparametric Quade test for dependent data was used to verify if there is interindividual variability in the *expression* level between all individuals. To check if any boar has a higher/lower expression of lncRNA than other individuals the pairwise comparison using Wilcoxon signed-rank test was performed. In both cases the length and expression analysis the Bonferroni procedure was applied for multiple testing correction across all pairs of individuals.

Results

The number of lncRNAs significantly varied between individuals ($P<2.2\times10^{-16}$) and ranged from 1,289 to 3,500 (Figure 1). Moreover, only 232 lncRNAs were common among all boars (Figure 2). Not only the counts but also lncRNA lengths significantly varied between individuals ($P<2.2\times10^{-16}$) (Figure 3).

In particular, individual S1 differed from the other boars. Moreover, there was significant interindividual variability in the expression level ($P=8.43\times10^{-266}$ in the Quade test). Pairwise Wilcoxon test with Bonferroni correction showed that lncRNAs of S1 exhibit higher expression than those belonging to the other boars (P varied from 5.48×10^{-89} to 2.22×10^{-139}), while S3 and S6 exhibit much lower expression than 4 other animals

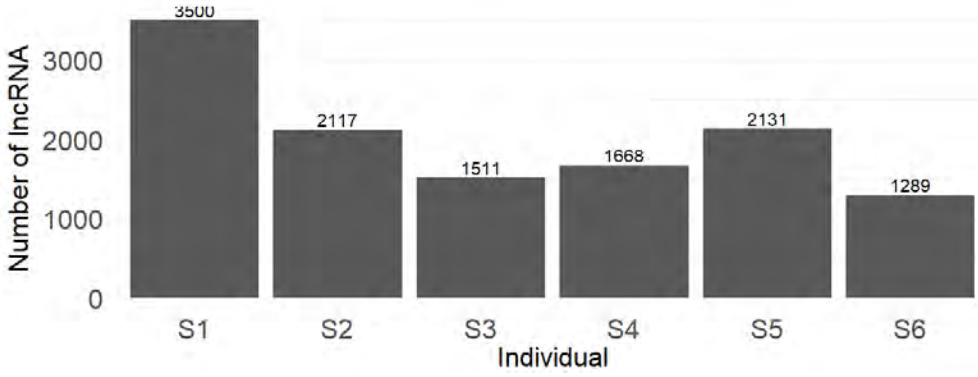


Figure 1. The total number of lncRNA per individual.

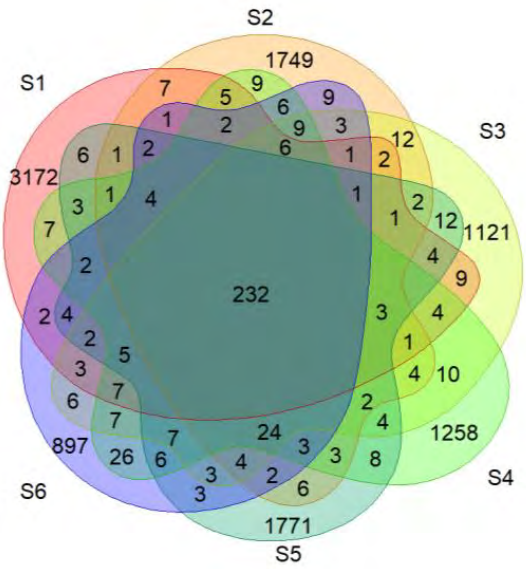


Figure 2. The number of common LncRNAs between boars.

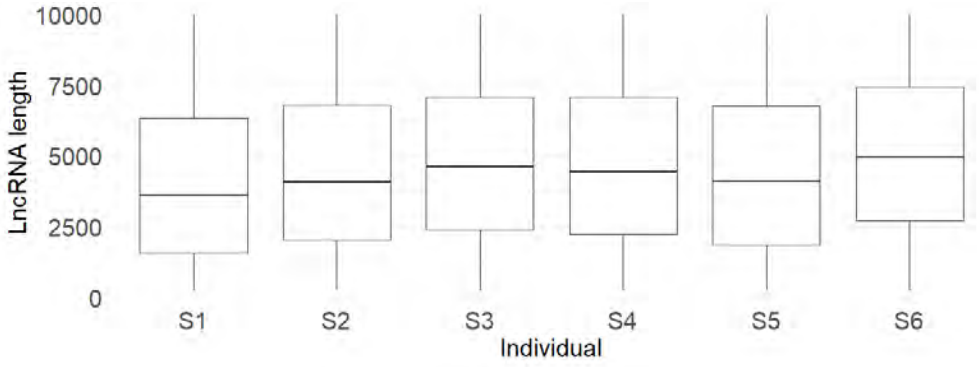


Figure 3. The lncRNA length per individual.

(*P*-values varied from 0.001 to 9.4×10^{-135} for pairs including S3 animal and from 1.11×10^{-139} to 1.21×10^{-7} for pairs including S6). Based on the pairwise analysis two groups with no difference in expression level were distinguished: S2, S4, and S5 as well as S3 and S6.

Discussion

It is already known that lncRNAs are abundant in multiple organisms, e.g. human, other vertebrates, plants and simple model species such as *C. elegans* and yeast (Kornienko *et al.* 2016). Multiple studies including numerous tissues have been carried out on lncRNAs also in pigs (Kosinska-Selbi *et al.* 2020, Yang *et al.* 2021) and the improvements in the understanding of non-coding RNAs variation has been observed by e.g. recognizing that lncRNAs display higher expression variation than protein-coding genes (Hu *et al.* 2020, Kornienko *et al.* 2016). Nevertheless, not much is known about the interindividual variability of lncRNA number, length and expression. In this work, we found that individuals significantly varied in all three

terms. Moreover, no consistent pattern has been found between pairs of half-brothers. This work might be a good start to understanding lncRNA variability between individuals and contribute to further research pointing to the role of this variability.

Acknowledgements

This work was supported by the Wroclaw University of Environmental and Life Sciences (Poland) as the Ph.D. research program 'Innowacyjny Naukowiec, no. N060/0035/20'. The computational power was provided by Poznan Supercomputing and Networking Centre. Muscle samples of Polish Landrace boars were kindly provided by slaughterhouse RTA SZWAGRZAK in Zloczew, Poland.

References

Andrews S. (2020) Available at: Babraham Bioinformatics - FastQC A Quality Control tool for High Throughput Sequence Data

Bolger, A.M., Lohse, M. and Usadel, B. (2014) Bioinformatics 30(15). https://doi.org/10.1093/bioinformatics/btu170

Cai, W., Li, C., Liu, S., Zhou Ch., Yin H. *et al.* (2018) Front. genet 9:281. https://doi.org/10.3389/fgene.2018.00281

Hu, H., Jia, Q., Xi, J., Zhou, Bo., Li, Z. (2020) BMC Genet. 21(636). https://doi.org/10.1186/s12864-020-07055-2

Kim, D., Paggi, J.M., Park, Ch. Benett, Ch., Salzberg, S.L. (2019) Nat. Biotechnol 37(8). https://doi.org/10.1038/s41587-019-0201-4

Kornienko, A.E., Dotter, Ch.P., Guenzl, P.M., Gisslinger, H., Gisslinger B., *et al.* (2016) Genome Biol. 17(14). https://doi.org/10.1186/s13059-016-0873-8

Kosinska-Selbi, B., Mielczarek, M., Szyda, J. (2020) Animal 14(10). https://doi.org/10.1017/S1751731120000841

Kovaka, S., Zimin, A.V., Pertea, G.M., Razaghi R., Salzberg S.L. *et al.* (2019) Genome Biol.(1). https://doi.org/10.1186/s13059-019-1910-1

Lehmann, E.L, Romano J.P. (2005) Testing Statistical Hypotheses, Springer, New York, NY

Li, H., Handsaker, B., Wysoker, A., Fennel, T., Ruan, J. *et al.* (2009) Bioinformatics 25(16). https://doi.org/10.1093/bioinformatics/btp352

Lagarrigue, S., Lorthiois, M., Degalez, F., Gilot, D., Derrien, T. (2021) Mamm Genome. https://doi.org/10.1007/s00335-021-09928-7

Morgulis, A., Gertz, E.M., Schaffer, A.A., Agarwala, R. (2006) J. Comput. Biol. 13(5). https://doi.org/10.1089/cmb.2006.13.1028

Pertea, G. and Pertea, M. (2020) F1000Res 9. https://doi.org/10.12688/f1000research.23297.2.

Petrov, A.I., Kay, S.J.E., Kalvari, I., Howe, K.L., Gray, K.A. *et al.* (2015) Nucleic Acids Res. 43(D1). https://doi.org/10.1093/nar/gkw1008

Quade, D. (1979) J Am Stat Assoc, 74(367). https://doi.org/10.1080/01621459.1979.10481670

Statello, L., Guo, Ch.J., Chen, L.L., Huarte, M. (2021) Nat. Rev. Mol. Cell Biol. 22(96-118). https://doi.org/10.1038/s41580-020-00315-9

Wang, L., Park, H.J., Dasari, S., Wang, S., Kocher, J.P. *et al.* (2013) Nucleic Acids Res 41(6). https://doi.org/10.1093/nar/gkt006

Warr A., Affara N., Aken B., Beiki H., Bickhart D.M., *et al.* (2020) Gigascience 1:9(6). https://doi.org/10.1093/gigascience/giaa051

Yang, Y., Yan, J., Fan, X., Chen, J., Wang, Z. et al (2021) PLoS Genet 17(11). https://doi.org/10.1371/journal.pgen.1009910.

807. Inclusion of average rainfall in genetic evaluation of SA Holsteins to mitigate genotype by environment interaction

M. van Niekerk[1,2*], F.W.C. Neser[1], J.B. van Wyk[1] and V. Ducrocq[1,2]

[1]Department of Animal, Wildlife and Grassland Sciences, University of the Free State, 205 Nelson Mandela Drive, Park West, Bloemfontein 9301, South Africa; [2]Université Paris-Saclay, INRAE, AgroParisTech, UMR GABI, 78350 Jouy-en-Josas, France; michiel@ganna.co.za

Abstract

Genetic improvement for persistency of milk production in South African Holstein cattle in a total mixed ration or pasture production system will be impeded because of the existence of genotype by environment interaction between the production systems. Previous studies where herds' production system were known (unlike reality) showed that rainfall level indicate the production system preferred by producers. The study aimed to determine the effect of including average rainfall in a random regression model as a possible proxy for production system used. The animal-additive-genetic effect due to annual rainfall (class) was added to the existing additive-genetic-average-production and -persistency effect. Heritability estimates of rainfall class 1 ('low' rainfall) coincided with total mixed ration estimates while rainfall class 5 ('high' rainfall) coincided with pasture estimates in later lactations. Results suggest that genetic background responsible for average production level and persistency over the lactations are not consistent over different rainfall areas.

Introduction

Genotype by environment interaction has been shown to exist in South African (SA) Holstein dairy cattle in previous studies by Ducrocq *et al.* (unpublished data) and Van Niekerk *et al.* (2019, 2021). In the latter two studies, clear differences in genetic parameters between herds using a total mixed ration (TMR) or pasture (PSTR) production system were shown. Specifically, genotype by environment interaction for persistency of production between the two production systems were indicated with genetic correlations for the first three lactations being below 0.64 throughout (Van Niekerk *et al.*, 2021). This mean that animals whose progeny perform well in TMR will not necessary perform well under PSTR conditions with regards to milk production (MilkPr) traits. Selecting sires whose progeny will perform satisfactory or superior in both or one of the production systems will thus not be accurate, as the production system used by herds are not known and thus ignored in genetic evaluations of SA Holsteins. Ducrocq *et al.* (2015) showed that the production system used by herds are dependent on the rainfall (long term annual average) of the herd's location (nearest town) in SA. They also showed that rainfall had the largest effect on the genetic variability of MilkPr using a reaction norm model. The aim of this study was to investigate the effects of including average rainfall of the towns nearest to the applicable herds as a proxy for the production system used, utilising a previously developed alternative random regression model (aRRM; Van Niekerk *et al.*, 2019).

Materials & methods

Full access to the complete pedigree and milk recording data that were used in this study were given by the SA Holstein Cattle Breeders' Society.

Data. Test-day (TD) records of MilkPr (kg) from 1988 to 2016 were used for the analysis. The dataset consisted of the first three lactations of cows from 37 herds, of which 26 utilize a TMR and 11 a PSTR production system. Cows with seven or more TD records per lactation and who completed their first lactation were used in the study. Sires with less than 10 daughters and contemporary groups with less than five cows were excluded from the dataset.

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_807, © M. van Niekerk *et al.* 2022

Statistical analyses. Restricted maximum likelihood was used to analyse the data, making use of the Wombat software (Meyer, 2007). MilkPr was analysed using multi-lactation records and the aRRM (Van Niekerk *et al.* (2019). This model was created as a better substitute to the fixed regression model used for genetic evaluations of SA Holsteins (Interbull, 2022). In the aRRM the animal-additive-genetic random effects were defined as separate for each lactation and expressed as a function of days in milk, containing a constant plus a regression slope multiplied by current days in milk. Consequently, each lactation is a combination of an (additive) genetic-average-production (GPr) and -persistency (GPe) effects (first and second eigenvector per lactation, respectively). In the analysis for this study, a third term (regression) over each day in milk was added for the animal-additive-genetic effect (third eigenvector). This term represents the genetic effect due to rainfall (GR), as a proxy to dominant production system. Rainfall level (long term annual average) was categorised in five classes as follow: Class 1: 413 to 492 mm; Class 2: 493 to 597 mm; Class 3: 598 to 672 mm; Class 4: 673 to 787 mm; Class 5: 788 to 1,008 mm. Classes were defined according to a relative balanced distribution of herds over rainfall level. The permanent environmental effect was included in the same way as the animal-additive-genetic effect. Residual variance was considered heterogeneous over and between lactations and partitioned in 10 classes per lactation. This aRRM, TD and multi-lactation model that included the GR effect (GR model) for MilkPr can be described as follows:

$$y = Tt + Ss + Hh + Ll + Mm + Aa + Cc + Qu + Zp + e \quad (1)$$

where y is a vector of 24-h test-day MilkPr observations for the lactations; t is a vector of fixed Herd × Test-day × Number of milkings interaction (contemporary group) effects; s is a vector of fixed Herd × Calving year × Calving season × Lactation number interaction effects. The following vectors of fixed regression coefficients are included in Equation 1: h for Herd effects; l for Calving year × Lactation number interaction effects; m for Calving month effects; a for Calving age class effects (8 classes); c for previous calving interval class effects (8 classes). The random regression coefficients for the animal additive genetic and permanent environmental effects were u and p, respectively and e is a vector of residual effects. The matrices T, S, H, L, M, A, C, Q and Z are incidence matrices that relate observations to their respective effects.

Results & discussion

The model that included the GR effect in this study did not have a better Bayesian Information criterion (increase of 27,316.38) compared to results from Van Niekerk *et al.* (2021), where the TD records from TMR and PSTR were treated as different traits (i.e. multi-trait analyses) and without the GR effect (TMR-PSTR model). However, there was an improvement of the Bayesian information criterion (decrease of 2,879.55) compared to results from Van Niekerk *et al.* (2021), where TD records from TMR and PSTR were treated as the same trait, also without the GR effect. This model can be seen as a 'national SA Holstein genetic evaluation' (nGE model). The addition of the GR effect in the current model explained 7.8% of the animal-additive-genetic effect for MilkPr in lactation 1 and decreased to 0.9 and 0.0% for lactation 2 and 3, respectively.

In lactation 1 (Figure 1), heritability estimates from the GR model were similar between ('low') rainfall classes 1, ('medium) 3 and ('high') 5, and in close agreement to TMR and nGE estimates. In lactation 2 the rainfall class 1 and 3 estimates were still close to each other and to TMR and nGE estimates. However, rainfall class 5 estimates were higher during early- to mid-lactation and closer to heritability estimates from the PSTR production system. In lactation 3 the differences between the heritability estimates of rainfall class 1, 3 and 5 are more pronounced during mid- to late-lactation with estimates from rainfall class 5 being closest to PSTR estimates, and rainfall class 1 estimates still being closest to TMR estimates (as in lactation 1 and 2). (Van Niekerk *et al.*, 2021).

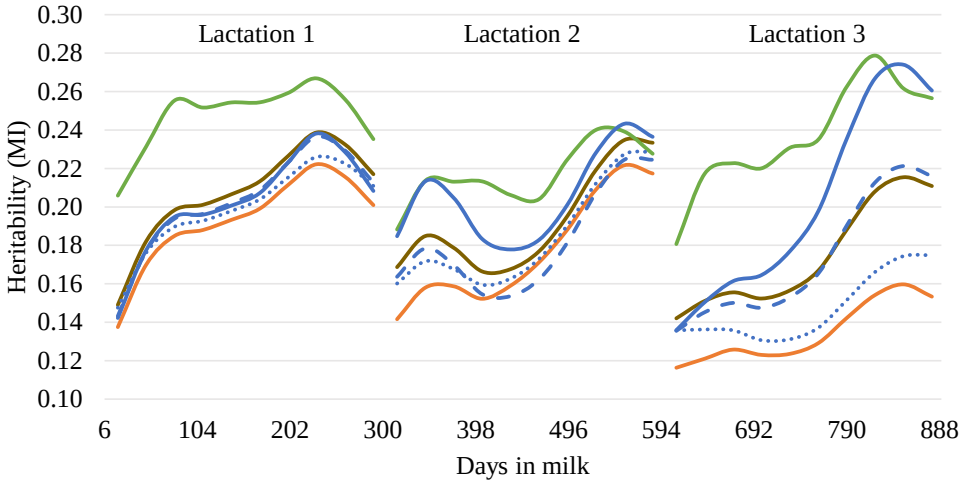


Figure 1. Heritability estimates for milk production over the first three lactations using the model including the genetic effect due to rainfall (class 1: blue dotted; class 3: blue dashed; class 5: blue solid; GR model), compared to a previous model where the two production systems were defined as different traits (total mixed ration: orange; pasture: green; TMR-PSTR model), as well as a 'national SA Holstein genetic evaluation' model excluding the genetic effect due to rainfall and TD records undefined for production system (brown; nGE model).

The nGE heritability estimates are similar to the TMR estimates as most records analysed with the nGE model came from herds using a TMR production system (Van Niekerk *et al.*, 2021). Rainfall class 1 heritability estimates being close to previous TMR heritability estimates make sense as 100% of herds in this class use a TMR production system. This also make sense to a lesser degree, for rainfall class 3 estimates as 78% of herds utilise a TMR production system. 75% of herds employing a PSTR production system in rainfall class 5 explains the similarity of these estimates with previous PSTR estimates. Rainfall classes 3 and 5 heritability estimates seem to follow nGE and PSTR estimates more precisely in later lactations, specifically lactation 3. (Van Niekerk *et al.*, 2021).

Genetic correlations (Table 1) between the GR and GPr effects were positive but weak, ranging from 0.43 to 0.32 for lactation 1 and 2, and 0.19 for lactation 3. The genetic correlation between the GR and GPe effects were negative and also weak for lactation 1 and 2 (-0.28 and -0.30, respectively), but positive and close to zero for lactation 3 (0.10).

The GPr and GPe components are made up of the average-additive-genetic level and persistency respectively, of performance of animals and their progeny, irrespective of the rainfall area(s) in which they produce and are measured in, over each of the lactations. When the additive-genetic performance of animals and their progeny are further regressed for average rainfall, the genetic correlations mentioned suggest that the genes involved in MilkPr are not consistent and reliable over different rainfall areas.

Conclusions

Considering the heritability estimates in this study and supported by previous research, including rainfall level as an important climatic variable in the proposed aRRM for MilkPr seem to be a good proxy for production system used by Holstein herds in SA, especially for later lactations. The way in which the rainfall classes are constructed may have an important effect on the heritability estimates. More herds in

Table 1. Genetic correlations between the different animal-additive-genetic effects for milk production (MilkPr) over the first three lactations.[1]

MilkPr		Lac1			Lac2			Lac3		
		GPr	GPe	GR	GPr	GPe	GR	GPr	GPe	GR
Lac1	GPr	1.00								
	GPe	-0.29	1.00							
	GR	0.43	-0.28	1.00						
Lac2	GPr	0.90	-0.41	0.43	1.00					
	GPe	-0.07	0.12	-0.03	-0.15	1.00				
	GR	0.20	-0.41	0.88	0.32	-0.30	1.00			
Lac3	GPr	0.88	-0.35	0.32	0.98	-0.24	0.25	1.00		
	GPe	-0.08	0.03	-0.07	-0.14	0.95	-0.26	-0.22	1.00	
	GR	0.26	-0.15	0.52	0.27	0.03	0.38	0.19	0.10	1.00

[1] Lac_i = lactation 1, 2 or 3; GPr = (additive) genetic-average-production effect; GPe = (additive) genetic-average-persistency effect; GR = (additive) genetic effect due to long-term-annual, average rainfall.

high rainfall areas that utilise a PSTR production system should be identified and incorporated in the study to increase accuracy of genetic parameters.

References

Ducrocq, V., Cadet, A., Patry, C., Pienaar, L., Van Wyk, J.B. *et al.* (2015) Interbull Meeting, Orlando, Florida, United States of America. https://interbull.org/static/web/Ducrocq_GxE2015.pdf

Interbull. (2022) https://interbull.org/ib/geforms

Meyer, K. (2007) WOMBAT – A tool for mixed model analyses in quantitative genetics by REML. J. Zhejiang Uni. SCIENCE B 8:815-821. https://doi.org/10.1631/jzus.2007.B0815

Van Niekerk, M., Neser, F.W.C., and Ducrocq, V. (2019) 51st South African Society for Animal Science (SASAS) congress, University of the Free State, Bloemfontein, South Africa. http://doi.org/10.13140/RG.2.2.17663.15520

Van Niekerk, M., Neser, F.W.C., Van Wyk, J. B., and Ducrocq, V. (2021) Proc. of the 72nd Annual Meeting of the European Federation of Animal Science, Davos, Switzerland. http://doi.org/10.13140/RG.2.2.17981.38883

808. Development of an automatic method to estimate IMF for fattening pigs in breeding program

L. Flatrès-Grall[1,2,4*], C. Ribas[4], K. Kashefifard[1,4], C. Herbert[3,4] and G. Lenoir[1,4]

[1]AXIOM, 37310 Azay-sur-Indre, France; [2]Chrysalide Loïc Flatrès-Grall, 29150 Dinéault, France; [3]HYCOLE, 59159 Marcoing, France; [4]GIS Nord Sélection, 37310 Azay-sur-Indre, France; loic.fg@mailo.fr

Abstract

Intramuscular fat content (IMF) is an important trait to estimate in order to select meat quality in pigs. The IMF estimation, using ultrasound measurement on live pigs (IMFref), is a suitable method but partly manual and so, too time consuming for field application. From a first dataset with 2,311 pigs, we developed an automatic method based on several steps to estimate IMF value from raw ultrasound pictures (IMFauto). This method permitted to obtain high corelations between values obtained by IMFref and IMFauto methods, 0.65 at picture level and 0.83 at animals' ranking within batch level. A genetic parameters estimation was done to validate the two traits (IMFref and IMFauto) using 1,544 animals evaluated from 2020 and 2021. The two traits were genetically identical with genetic corelation equal to 1. The automated estimation gave a more accurate estimation with higher heritability and less rejected pictures. The trait IMFauto is validated to be used routinely in selection but correlation between this trait and IMF measured on meat have to be validated.

Introduction

The intramuscular fat level (IMF) is related to the meat quality in pork due to association with tenderness and juiciness of the meat (Ragland *et al.*, 1998). AXIOM, a pig breeding company, decided to integrate IMF trait in breeding goal of its Duroc line in 2015. The IMF level is estimated on the *longissimus dorsi* muscle by using ultrasound measurements on live animals around 100 kg liveweight. Routinely, the IMF estimation (IMFref) is done using BioSoft Toolbox® II for Swine 2.5 developed by Biotronics company (Newcom *et al.*, 2002). This method is partly manual. Indeed, technician has to done several actions and choices to estimate IMF: image quality control and selection of the area to analyse on ultrasound picture. This process requires 1 week between animal ultrasound measurement and integration of IMF estimation in the BLUP genetic evaluation. This process is not efficient due to this delay. In order to improve efficiency and save, AXIOM decided to develop an automatised method for IMF calculation. To estimate IMF value with ultrasound pictures, several approaches have been studied for the last years. Ragland *et al.* (1998) used multiple regression to predict IMF using linear model. Aass *et al.* (2006) used stepwise regression procedures to develop prediction models. In 2020, AXIOM had estimated more than 2,500 raw ultrasound pictures and values of IMF calculated from linear regression using BioSoft Tollbox ® II for Swine 2.5 (IMFref). At this stage, no IMF measurements on meat were available. The aim of this study was to develop an automatic method to estimate IMF content in a short time and to validate that the new trait estimated was genetically similar to IMFref.

Materials & methods

Animals. Animals used were males and females from AXIOM Duroc line, born in one nucleus farm and raised at the original farm or at the boar testing station of AXIOM. Two datasets were created to develop this method. The first one used to develop the method of the automatic IMF (IMFauto) estimation was composed with 2,311 animals tested from 2016 to 2019. The second dataset, used for genetic parameters estimation on of IMFauto was composed of 1,544 pigs recorded from 2020 to 2021. The pedigree file was traced back 24 generations corresponding to 88,974 animals born between 1994 and 2020. All the known ancestors were integrated in the pedigree file.

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_808, © L. Flatrès-Grall *et al.* 2022

Ultrasound measurements. All pigs, around 100 kg liveweight, were weighted and 6-10 ultrasound images were taken using EXAGO ultrasound system (IMV Imaging; Angoulême, France) equipped with a linear transducer. Measurement site was on the *longissimus dorsi* muscle between the 10th and 13th vertebrae on the left side at almost 7 cm from the dorsal line. Operators who recorded measurements on pigs were AXIOM technicians, trained together at least one time per year to record IMF pictures. All pigs from one batch were tested the same day by the same operator in one group of control (batch).

IMF reference estimation. The reference value of IMF (IMFref) was calculated using a linear regression with BioSoft Toolbox® II for Swine 2.5. In average 5.2 (±1.5) pictures / animal were considered as acceptable after quality control done by the technician. Then operator positioned a 1 square centimeter area on each picture to estimate IMFref (Newcom *et al.*, 2002). The IMFref value considered for each animal was the average of values estimated for all pictures measured from this animal, from 1 to 11 pictures for this study.

IMFauto development. Automatic process of IMF estimation was developed (IMFauto). Each raw ultrasound picture was segmented using scikit-image library on Python software (van der Walt *et al.*, 2014). Firstly, a Canny filter function (Canny,1986 cited by van der Walt *et al.*, 2014) was used to highlight in a binary picture the muscle edge. Secondly, the transform hough line function (Van der Walt *et al.*, 2014) used the result of the Canny filter function to create automatically the limit of the muscle area. From this segmentation, muscle area was extract from the original picture. From muscle area, the parameters described by Newcom *et al.* (2002) were extracted from each picture, representing 50 variables for each picture. From those variables, a random forest analyze was proceeded using the randomForest package (Liaw and Wiener, 2002) on R (R Core Team, 2018) to estimate the best coefficients of the linear equation to predict IMFref from automatic picture treatment. Best model was validated using Spearman corelation using package stats on R (R Core Team, 2018). Several levels of validation have been done: picture level, animal level and ranking within batch. Picture quality control was also automatically done on picture size after segmentation.

IMFauto calculation. IMFauto calculation was done on all the historical data of ultrasound pictures from AXIOM (first dataset). The IMFauto value considered for each animal was the average of values estimated for all pictures validated by the method. IMFauto calculation take less than 1 second per picture and around 15 minutes per batches (30 animals with 6 pictures/animals) for all the process.

Genetic parameters estimation. Genetic parameters were estimated using ASReml 3.0 software (Gilmour *et al.*, 2009) with the restricted maximum likelihood methodology applied to a multitrait animal model with two traits, IMFref and IMFauto. Fixed effects were tested at an α-risk of 5% using the Wald test on ASReml (Gilmour *et al.*, 2009). Testing farm with 2 levels, sex with 2 levels (male and female) and batch of control with 37 levels were significant and included as fixed effect in models. The number of pictures taking in account to calculate IMFauto or IMFref and the weight at testing were significant and added as covariate. Litter effect used as contemporary group effect, was added as random effect. Dataset for genetic validation was composed with 1,544 Duroc animals

Results

Linear equation. The best linear model obtained by random forest to predict IMFref from automatic treatment showed a Spearman correlation of 0.65 at picture level and 0.74 at animal level (Table 1). Correlation of animal's ranking from highest to lowest estimation of IMF value within batch between IMFref and IMFauto was 0.83.

Table 1. Descriptive statistics for both IMF.

	Number of animals	Average weight at testing (kg)	Average number of pictures / animal	Average IMF (%)
IMFref	1,082	105.1	5.2	1.99
IMFauto	1,474	105.6	5.7	1.96

Phenotypic means and distributions. On the dataset used for genetic parameters estimation 70% of animals have value for IMFref and 95% for IMFauto. The difference was due to quality control process with 30% of animals rejected for IMFref and only 4.5% for IMFauto method. The IMF trait was tested at 102 kg for IMFref and 106 kg liveweight for IMFauto with close average estimates (1.99 vs 1.96%).

Genetics parameters. Heritability of IMF calculated was quite high with 0.52 for IMref and 0.66 for IMFauto but could not considered as different due to standard error (Table 2). Genetically, the two traits could be considered as identical, genetic correlation was 0.99, not significatively different from 1. Whereas, absolute value of the phenotype could differ between the two methods with the phenotypic correlation significatively different from 1 (0.84).

Litter effect was similar for both IMF, with 1 and 2% of the total variance explained for IMFref and IMFauto. They were not difference from 0 with a standard error of 0.02.

Discussion

Linear model for prediction seems to be the best approach in the AXIOM's context, it was previously used by Ragland *et al.* (1998) and Aass *et al.* (2005). Kvam *et al.* (2017) showed that deep learning applied to ultrasound images can predict accurately IMF as long as sufficient image quality and large number of images were available. To develop this kind of method AXIOM should invest in a large database with IMF from meat whereas, for a linear approach, row pictures and IMF calculation were already available in AXIOM's databases. Therefore, predicting IMFauto from IMFref appears as the best option. Nevertheless, IMFauto was developed to predict IMFref but not to estimate the real value of intramuscular fat in pork meat. It will be necessary to validate IMFauto with chemical value of IMF or at least visual ranking of IMF realised directly on meat.

Genetic validation of IMFauto. IMFauto and IMFref had very similar values (1.99 vs 1.96% of IMF). Maignel *et al.* (2009) showed that correlation between IMF obtained *in vivo* by ultrasound measurement and IMF on pork meat using chemical analysis was around 0.62, value higher than the study done by Newcom in 2002. In fact, absolute value obtain by ultrasound technology is lower than value in the meat but the rank between animals is correct and permitted to select animals on IMF (Maignel *et al.*, 2009). IMFauto method has to be validated with samples of IMF from chemical analysis of the meat.

Table 2. Genetic parameters of IMFref and IMFauto with heritability on diagonal, phenotypic corelation down to the diagonal and genetic corelation up to the diagonal.

	IMFref	IMFauto
IMFref	0.52 (±0.08)	0.99 (±0.02)
IMFauto	0.84 (±0.02)	0.66 (±0.08)

Heritability estimates of IMFauto and IMFref were close to values found by Maignel *et al.* in 2009 (0.52 and 0.66 vs 0.66), with Duroc and dataset size very similar. Won *et al.* in 2017 found IMF heritability at 0.52 but it was estimated on Yorkshire animals.

Corelation between IMFauto and IMFref was close to 1, meaning that the both approaches evaluated the same trait. Heritability was higher for IMFauto and phenotypic variance were smallers than IMFref. Probably, the automatized process produced less bias than operator evaluation. In the same time, less pictures were rejected during quality control. Indeed, automatic quality control is more objective than human evaluation.

AXIOM developed an automatic process to estimate IMF using ultrasound measurement. Linear equation was based on IMFref which was not an estimation of intramuscular content of the meat but estimation of it using ultrasound pictures. Genetic parameters have shown that IMFauto and IMFref were genetically the same trait. The automatization reduced variability of the estimation and bias due to operators' decisions. IMFauto was a good solution to select animal instead IMFref with less time between measurement and more accuracy. However, no validation has been done on real IMF value in pork meat. The next step will be to evaluate IMF on meat using chemical estimation on animals previously evaluated alive.

References

Aass L., Gresham J.D., and Klemetsdal G. (2006) Livest. Sci. 101:228-241. https://doi.org/10.1016/j.livprodsci.2005.11.014

Canny J. (1986) IEEE Trans. Pattern Anal. Mach. Intell. 8(6):679-698. https://doi.org/10.1109/TPAMI.1986.4767851

Gilmour A.R., Gogel B.J., Cullis B.R. and Thompson R. (2009) ASREML user guide release 3.0. Available at: https://www.vsni.co.uk/downloads/asreml/release3/UserGuide.pdf

Kvam J., and Kongsro J. (2017) Computers and Electronics in Agriculture, 142(B):521-523. https://doi.org/10.1016/j.compag.2017.11.020.

Liaw A., and Wiener M. (2002) R News 2(3):18-22.

Maignel L., Daigle JP., and Sullivan B. (2009) Proc of the 41st Journ. Rech. Porcine, Paris, France.

Newcom D.W., Baas T.J., and Lampe J.F. (2002) J. Anim. Sci., 80:3046-3052. https://doi.org/10.2527/2002.80123046x

Ragland K.D. (1998). https://doi.org/10.31274/rtd-180813-10805

R Core Team (2018) R: A Language and Environment for Statistical Computing. Available from: https://www.R-project.org

Van der Walt S., Schönberger J.L., Nunez-Iglesias J., Boulogne F., Warner J.D. *et al.* (2014) PeerJ 2:e453. https://doi.org/10.7717/peerj.453

Won S, Jung J, Park E, Kim H., (2018) Asian-Australas J Anim Sci. 31(2):157-162. https://doi.org/10.5713/ajas.17.0218. Epub 2017 Jun 27. PMID: 28728355; PMCID: PMC5767496.

809. Assessing genomic prediction of economic trait in alpacas: a simulation study

G.C. Mamani[1*], B.F. Santana[2] and D. Jarquin[3]

[1]Instituto Nacional de Innovación Agraria, Av. La Molina N° 1981, 2791, Lima, Peru; [2]University of Connecticut, Storrs, CT 06269, USA; [3]University of Florida, Gainesville, FL 32611, USA; gerardocmamani@gmail.com

Abstract

The objective of this study was to assess the prediction accuracy of genomic breeding values using different methods. The data were simulated containing phenotypic, pedigree and genomic information mimicking an alpaca population. A total of 70,000 SNP markers were simulated considering the 36 alpaca chromosomes. The heritability of the simulated phenotypic trait (fiber diameter trait) was set to 0.41. Three prediction methods were considered: BLUP, GBLUP and ssGBLUP. The genomic prediction accuracy was computed as the correlation between the true genetic values and the predicted values, also the expected accuracy was estimated via the prediction error variance. GBLUP provided the highest prediction accuracy for non-phenotyped young animals followed by ssGBLUP and BLUP. The obtained results indicate that in the absence of animal performance data, the genomic information allows us to obtain a higher accuracy of breeding values which can be leveraged to obtain greater genetic gains.

Introduction

In animal breeding, the selection of genetically superior individuals is crucial to improve trait performance of populations. The estimation of breeding values through a predigree-based best linear unbiased prediction (BLUP) which method is widely applied, both in academia and for commercial purposes. Recently, the availability of dense molecular markers (genomic data) has revolutionized conventional breeding strategies. However, the high initial cost for implementing this methodology could discourage its use. The use of simulation studies helps to assess the potential of this methodology to take more informed decisions before making a large investment.

Genomic selection (GS) proposed by Meuwissen *et al.* (2001) aims to estimate the genomic breeding values (GEBV) of non-phenotyped individuals and to carry out the selection of the animals using GEBV. This method initially requires the availability of phenotypic and genomic information for a set of already observed individuals to be used from model calibration then the genomic breeding values of non-phenotyped individuals are estimated using exclusively their maker profiles.

The alpaca is an important specie for many farmers in Peru due cultural, economic, and social aspects. To our knowledge, as today there are no alpaca breeding programs using genomic information in the selection process. The objective of this study is to compare the accuracy of estimated breeding values obtained with three prediction methods using simulated alpaca genomic data.

Materials & methods

Simulation. The phenotypes, pedigree and genomic data were generated using QMSim software (Sargolzaei, 2014). The historical population consisted of 100 generations with a gradual decrease in size from 3,000 to 640 individuals. A total of 600 females and 40 males of the last generation of the historical population were the founders of the recent population. The population was expanded by 10 generations based in phenotypic selection. The simulated genome considers the 36 alpaca autosomes. Genotypes were simulated for 70,000

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_809, © G.C. Mamani *et al.* 2022

bi-allelic markers for the five last generations (3,000 individuals). Phenotypic measures were simulated based on real heritability estimates for fiber diameter (0.41, Gutiérrez *et al.*, 2009).

Statistical analysis. *BLUP (Best linear unbiased prediction).* The following model was used to predict the breeding values when using pedigree data solely:

$$y = 1u + Za + e \quad (1)$$

where, y represents the response of the trait, u is the general mean and it is common to all genotypes, Z is the incidence matrix that connects phenotype with additive genetic effects, a is the vector of additive genetic effects with $a \sim N(0, A\sigma^2_a)$ and e is the vector of residuals such that $e \sim N(0, I\sigma^2_e)$ where, A is the numerator relationships matrix, I is the identity matrix, σ^2_a is the additive genetic variance, and σ^2_e is the residual variance.

GBLUP (genomic best linear unbiased prediction). The direct genomic value (DGV) is estimated for all genotyped individuals using the following equation:

$$y = 1u + Zg + e \quad (2)$$

where, DGV (*g*), $g \sim N(0, G\sigma^2_g)$, G is the genomic relationship matrix whose entries contains the genomic similarities between pairs of individuals, σ^2_g is the additive genetic variance explained by markers, and the remaining terms as previously defined. The G matrix was calculated according to VanRaden (2008). The GBLUP analysis was implemented using the Blupf90 program (Misztal, 2014).

ssGBLUP (Single-Step genomic best linear unbiased prediction). The ssGBLUP was implemented to combine the information of genotyped and non-genotyped individuals using the inverse of the hybrid matrix H $(H)^{-1}$ (Aguilar *et al.*, 2010) which joins the matrix of the numerator relationship matrix (A) with the genomic relationship matrix obtained with maker data (G). The ssGBLUP analysis was performed using the same parameterization as the BLUP, only replacing the matrix A^{-1} by H^{-1}. This method was also implemented in the Blupf90 program (Misztal, 2014).

Accuracy. The model's predictive ability was measured by computing the Pearson's correlation between the predicted and the true breeding values for the different prediction methods. In addition, we also measured the expected accuracy by computing the Predicted Error Variance (PEV).

Results

The highest accuracy was obtained with GBLUP method. The density plot (Figure 1) shows the empirical distribution of the estimated breeding values obtained with the three different methods implemented. These empirical distributions are mound shaped and close to a normal distribution. The BLUP method provided EBV with less variability, also the empirical distributions derived from the GBLUP and ssGBLUP methods are overlapped also showing a slightly smaller variance.

Table 1 shows the average, minimum and maximum prediction accuracy for the last two simulated generations (9 and 10). Predicting the breeding values for young genotyped animals without phenotypic records, the GBLUP method returned the highest average correlation (0.82 and 0.81) for these generations compared with the other two alternatives. The BLUP returned the smallest accuracy (0.56 and 0.50) while the ssGBLUP returned intermediate results (0.67 and 0.63). Thus, the relative improvements in predictive ability using GBLUP compared with ssGBLUP were 22 and 29% for generations 9 and 10, respectively.

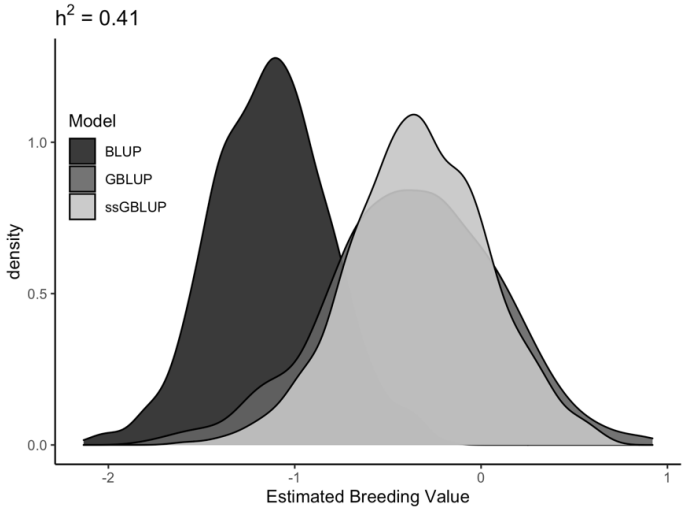


Figure 1. Density of estimated breeding values obtained using BLUP (best linear unbiased prediction), GBLUP (genomic beast linear unbiased prediction), and ssGBLUP (single-step genomic best linear unbiased prediction) methods.

Table 1. Average accuracy values obtained with different methods using A matrix (BLUP), G matrix (GBLUP) and H matrix (ssGBLUP) of the last two generations (9 and 10).

Generation	n	BLUP	GBLUP	ssGBLUP
9	600	0.56	0.82	0.67
10	600	0.5	0.81	0.63

However, the animals without genomic information but with pedigree data are excluded of analysis, except when using the ssGBLUP.

Discussion

The GBLUP method returned the highest mean accuracy, but it only can consider individuals with genomic data while the method that only uses pedigree data (BLUP) produced the lower results. On the other hand, the ssGBLUP had higher accuracies than BLUP, confirming that the integration of maker and pedigree data can be beneficial.

This comparison is between the last two generations of non-phenotyped but genotyped individuals. Garrick (2011), in beef cattle, showed that genomic prediction offers accuracies that exceed those of pedigree-based parent average of young selection candidates, and it can be equivalent to progeny tests based on up to 10 offspring. In addition, the ssGBLUP permits the estimation of genetic values for non-genotyped animals, which is not possible in GBLUP analysis (Tsuruta, Misztal, Aguilar, Lawlor, 2011) and also allows increasing training sets sizes by including genotypes with pedigree and phenotypic data only. The obtained results using BLUP, GBLUP and ssGBLUP of this study are in agreement with those reported by Daetwyler, Villanueva and Woolliams (2007). These authors found that higher accuracies when the heritability of the traits was high due to a high contribution of gene effects for phenotypic variation.

Conclusions

The prediction of breeding values using genomic information showed accuracies higher than pedigree-based estimations in animals without phenotypic information. Between the two methods that use genomic information, the ssGBLUP allows obtaining breeding values for un-genotyped individuals as well.

References

Aguilar I., Misztal I., Johnson D.L., Legarra A., Tsuruta S., *et al.* (2010) J. Dairy Sci. 93(2):743–752. https://doi.org/10.3168/jds.2009–2730.

Calus M.P.L., Veerkamp R.F. (2007) J. Anim. Breed. Genet. 124(6):362-368. https://doi.org/10.1111/j.1439-0388.2007.00691.x

Daetwyler H.D., Villanueva B., Bijma P., Woolliams J.A. (2007) J. Anim. Breed. Genet. 124(6): 369–376. https://doi.org/10.1111/j.1439-0388.2007.00693.x

Garrick, J.D. (2011). Genet. Sel. Evol, 43(17): 1-11. https://doi.org/10.1186/1297-9686-43-17

Gutiérrez J.P., Goyache F., Burgos A., Cervantes I. (2009) Livest. Sci. 123(2-3): 193–197. https://doi.org/10.1016/j.livsci.2008.11.006

Meuwissen T.H.E., Hayes B.J., Goddard M.E. (2001) Genetics. 157(4):1819-1829. https://doi.org/10.1093/genetics/157.4.1819

Misztal, I., S. Tsuruta, D.A.L. Lourenco, I. Aguilar, A. Legarra, and Z. Vitezica. (2014) Manual for BLUPF90 family of programs. (Accessed Date, 2021)

Muir W.M. (2007). J. Anim. Breed. Genet. 124(6):342-355. https://doi.org/10.1111/j.1439-0388.2007.00700.x.

Sargolzaei M., Schenkel F.S. (2009) Bioinformatics 25(5):680–681. https://doi.org/10.1093/bioinformatics/btp045

Tsuruta S., Misztal I., Aguilar I., Lawlor T.J. (2011) J. Dairy Sci. 94(8):4198–4204. https://doi.org/10.3168/jds.2011-4256

VanRaden P.M. (2008) J. Dairy Sci. 91(11):4414–4423. http://doi.org/10.3168/jds.2007-0980

810. Somatic cell count and its relationship with udder type traits in Italian Mediterranean Buffaloes (*Bubalus bubalis*)

M.M. Gómez[1], R. Cimmino[1], Y. Gombia[1], G. Zullo[1] and S. Biffani[2*]

[1]Italian National Association of Buffalo Breeders (ANASB), Via Petrarca 42-44, 81100 Caserta, Italy; [2]National Research Council (CNR), Institute of Agricultural Biology and Biotechnology (IBBA), Via Edoardo Bassini 15, 20133 Milano, Italy; stefano.biffani@ibba.cnr.it

Abstract

Mammary infections in livestock are a major problem which affects animal health, resulting in significant economic losses in the dairy industry. Among the factors which can mitigate or increase the rate of intramammary infections, udder morphology plays a substantial role. The somatic cell score (SCS) is commonly used as predictor in selection for mastitis resistance. The aim of this study was to investigate the relationship between SCS and udder traits in the Italian Mediterranean Buffalo (IMB). A total of 122,717 lactations, SCS and type traits (one composite trait and six linear traits) records from 16,627 primiparous buffalo cows were used. The relationship was investigated fitting two mixed models. Analysis of variance suggested that SCS were significantly (P<0.001) affected by both udder morphology and milk yield. Least square means clearly indicated that having a morphologically correct udder contributes significantly to reducing SCS level in milk.

Introduction

Mastitis is an infection of the mammary gland, a disease of great economic importance resulting in significant economic losses in the dairy industry. Indeed, the economic costs caused by mastitis are of different nature, for example the reduction in milk production, the discarded milk, the increased culling, the increased costs of health care and labour, and the reduced milk quality. Therefore, strategies to control mastitis must include preventive health care, hygiene, veterinary remedies and genetic selection (Juozaitiene *et al.*, 2006). A commonly used phenotypic indicator of udder health is somatic cell count (SCC) (Tiezzi *et al.*, 2020). After being log-transformed to SCS, it is used as a predictor in the selection for mastitis resistance (Martin *et al.*, 2018, Parker Gaddis *et al.*, 2020). Some studies in the buffalo species (Dhakal, 2006, Piccinini *et al.*, 2006) confirmed that the threshold value of 200×103 cells/ml in buffalo milk was indicative of the presence of inflammation. Moreover parity, calving age, stage of lactation, season, stress, milking interval are additional features which can affect SCC level in buffalo milk (Singh and Ludri, 2001). Among the factors which can mitigate or increase the incidence of intramammary infections, udder morphology plays a substantial role (Pantoja *et al.*, 2020). However, few studies have been performed to investigate the relationship between SCS and udder morphology in the buffalo species. Hence, the purpose of this study was to evaluate the relationship among traits related to the mammary system, milk yield and SCS in the Italian Mediterranean Buffalo.

Materials & methods

Data. Three sources of data were available, namely milk yield data, SCC data and type trait data. The production data set included a total of 6,576,041 milk test day samples collected between 1982 and 2019 from 262,254 buffaloes, the SCC file included 2,348,632 records, corresponding to 175,362 buffaloes and the type trait dataset was composed of 81,537 records collected between 2004 and 2020. After joining and editing these three sources of information, only primiparous buffalo cows were retained and the final dataset was composed of 122,717 records from 16,627 primiparous buffalos. The variable SCC was log-transformed into SCS by using the formula SCS = log2(SCC/100,000) + 3 (Ali and Shook, 1980). The udder traits, which included a udder score (US) as composite trait and six linear traits were: fore udder attachment

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_810, © M.M. Gómez *et al.* 2022

(FUA), rear udder width (RUW), udder depth (UD), teat placement (TP), teat length (TL) and teat position (TPOS). US was scored using a scale from 65 to 99, while all linear traits were scored on a scale with 1 and 50. Finally, all traits were transformed into classes, with an interval of 5.

Model. The relationship between SCS and udder traits was investigated fitting the following mixed model:

$$y_{ijklmn} = \mu + DIM30_i + YM_j + MT_k + b_1MP_l + HYS_m + A_n + e_{ijklmn} \quad (1)$$

where y is the dependent variable (SCS test-day); μ is the general mean; DIM30 is the fixed effect of the i^{th} stage of lactation; YM is the fixed effect of the j^{th} year-month combination of milk recording; MT is the fixed effect of the k^{th} score of the morphological trait; b1 is the linear regression coefficient of daily milk production (MP). Animal (A) and Herd-year-season (HYS) were random effects; and e was the random residual term. Furthermore, backward elimination was used to eliminate non-significant effects from the model. Analysis of variance and least square means (LSM) were calculated to identify possible relationships between morphological traits and SCS.

Results

The results of the analysis of variance for SCS based on a model with the composite trait as explanatory variable (M1) and a model with all linear traits as explanatory variables (M2) are in Table 1. All effects included in the model were significant ($P<0.001$). These findings confirm the hypothesis that SCS are significantly affected by both the morphology of the mammary system and the production level.

LSM for both composite and linear traits can be observed in Figure 1a and 1b, respectively. An increase in US score has a favourable effect on SCS, which decrease by approximately 1 unit. Indeed, an udder score higher than 80 reflects a functional udder with strong ligaments, appropriate teat size and position, and proper volume. These findings are also confirmed when the relationship between SCS and morphology is analysed using three udder-related linear traits, namely TP, UD and TL. As expected, an increase in both TP and UD score has a favourable effect on SCS (decrease) while an intermediate score for TL (higher scores = longer teats) is more effective.

Table 1. Analysis of variance for SCS based on a model with the composite trait (M1) and linear traits (M2).

Factor[2]	DF	M1[1]		M2[1]	
		Mean	F value	Mean	F value
DIM30	10	123.87	56.61***	121.37	55.47***
YM	98	126.15	57.66***	126.39	57.77***
MP (kg)	1	2,503.02	1,144.05***	2,627.09	1,200.81***
US[3]	4	11.24	5.13***	-	-
UD[3]	10	-	-	13.93	6.36***
TP[3]	10	-	-	8.60	3.93***
TL[3]	10	-	-	19.19	8.77***

[1] M1, model 1; M2, model 2.

[2] See model (1) for acronym.

[3] US = Udder score; UD = udder depth; TP = teat placement; TL = teat length.

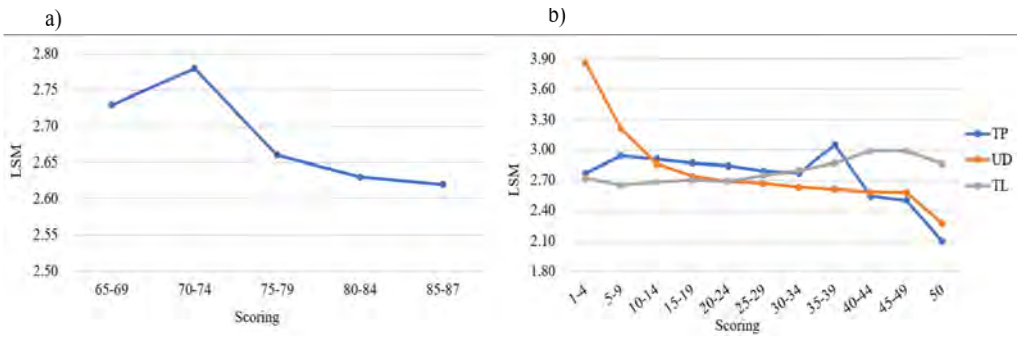


Figure 1. Least square means for (a) composite trait (US) and (b) linear traits (TP, UD, TL).

Discussion

Milk production is a complex process which can possibly exacerbate stress responses. This is particularly true in specialized dairy breeds where udder health problems and the detriment of its conformation have been observed (Sartori *et al.*, 2018). This is the case of the IMB, where an impressive increase in milk production over the last 15 years has been observed (ANASB, 2021). Alongside milk production, an increase in SCS has also been observed, being steeper in buffalo cows born before 2011 than in recent years (Costa *et al.*, 2021). Type traits appraisal in the IMB started at the early 2000s and in 2018 udder and feet score have been included in the IMB breeding objective (Biffani *et al.*, 2021). From the results observed in this study, an udder with the correct length, direction and position of the teats favors low SCC values during lactation. Present findings are also supported by earlier studies, where deeper udders or udders presenting an incorrect teat position were found to be at higher risk of developing mastitis (Juozaitiene *et al.*, 2006, Singh *et al.*, 2014). Moreover, we identified longer teats as potential risk factor for mastitis. This is in agreement with the findings of Berry *et al.* (2004) who also showed that dairy cows with longer teats were genetically predisposed to a higher incidence of mastitis. Rogers and Hargrove (1993) reported that Holsteins with shorter teats were associated with lower SCC showing a lower incidence of mastitis. Accordingly, the overall udder morphology and some additional udder traits such as US, TP, UD or TL could have some degree of association with the udder health in the buffalo cows. Hence, their inclusion in breeding programs as indicator traits may help to control and possibly reduce the incidence of mastitis in the IMB.

Acknowledgements

This research was funded by Italian Ministry of Agriculture (MIPAAF – DISR 07). Project: 'BIG' Prot. N. 0215513 11/05/2021.

References

Ali, A. K. A. and G. E. Shook. 1980. An Optimum Transformation for Somatic Cell Concentration in Milk1. J. Dairy Sci. 63(3):487-490.

ANASB. 2021. Associazione Nazionale Allevatori Specie Bufalina. http://www.anasb.it/statistiche/ [Accessed Dec 11, 2021].

Berry, D. P., F. Buckley, P. Dillon, R. D. Evans, and R. F. Veerkamp. 2004. Genetic relationships among linear type traits, milk yield, body weight, fertility and somatic cell count in primiparous dairy cows. Ir. J. Agric. Food Res.:161-176.

Biffani, S., M. Gómez, and A. Cesarani. 2021. Pitfalls and opportunities of genetic and genomic evaluation in the Buffalo species: experiences from Italy. Interbull Bulletin (56):94-101.

Costa, A., M. De Marchi, G. Neglia, G. Campanile, and M. Penasa. 2021. Milk somatic cell count-derived traits as new indicators to monitor udder health in dairy buffaloes. Ital. J. Anim. Sci. 20(1):548-558.

Dhakal, I. P. 2006. Normal somatic cell count and subclinical mastitis in Murrah buffaloes. J Vet Med B Infect Dis Vet Public Health 53(2):81-86.

Juozaitiene, V., A. Juozaitis, and R. Micikeviciene. 2006. Relationship between somatic cell count and milk production or morphological traits of udder in Black-and-White cows. Turk J V Anim Sci 30(1):47-51.

Martin, P., H. W. Barkema, L. F. Brito, S. G. Narayana, and F. Miglior. 2018. Symposium review: novel strategies to genetically improve mastitis resistance in dairy cattle. J. Dairy Sci. 101(3):2724-2736.

Pantoja, J. C. F., L. B. N. Correia, R. S. Rossi, and G. S. Latosinski. 2020. Association between teat-end hyperkeratosis and mastitis in dairy cows: A systematic review. J. Dairy Sci. 103(2):1843-1855.

Parker Gaddis, K. L., P. M. VanRaden, J. B. Cole, H. D. Norman, E. Nicolazzi, *et al.* 2020. Symposium review: Development, implementation, and perspectives of health evaluations in the United States. J. Dairy Sci. 103(6):5354-5365.

Piccinini, R., M. Miarelli, B. Ferri, C. Tripaldi, M. Belotti, *et al.* 2006. Relationship between cellular and whey components in buffalo milk. J. Dairy Res. 73(2):129-133.

Rogers, G. W. and G. L. Hargrove. 1993. Absence of quadratic relationships between genetic evaluations for somatic cell scores and udder linear traits. J. Dairy Sci. 76(11):3601-3606.

Sartori, C., N. Guzzo, S. Mazza, and R. Mantovani. 2018. Genetic correlations among milk yield, morphology, performance test traits and somatic cells in dual-purpose Rendena breed. Animal 12(5):906-914.

Singh, M. and R. S. Ludri. 2001. Somatic Cell Counts in Marrah Buffaloes (Bubalus bubalis) During Different Stages of Lactation, Parity and Season. Asian-Australas. J. Anim. Sci. 14(2):189-192.

Singh, R. S., B. K. Bansal, and D. K. Gupta. 2014. Udder health in relation to udder and teat morphometry in Holstein Friesian×Sahiwal crossbred dairy cows. Trop. Anim. Health Prod. 46(1):93-98.

Tiezzi, F., A. M. Maisano, S. Chessa, M. Luini, and S. Biffani. 2020. Heritability of Teat Condition in Italian Holstein Friesian and Its Relationship with Milk Production and Somatic Cell Score. Animals 10(12):2271.

811. Estimating vertical path coefficients of transmission of non-genetic inherited effects from a similarity matrix

I. David[1*], C. Larzul[1] and A. Ricard[2,3]

[1]GenPhySE, INRAE, Université de Toulouse, INPT, 31326 Castanet Tolosan, France; [2]GABI, INRAE, AgroParisTech, Université Paris Saclay, 78352 Jouy-en-Josas, France; [3]Institut Français du Cheval et de l'Equitation, Département Recherche et Innovation, 61310 Exmes, France; ingrid.david@inrae.fr

Abstract

In addition to DNA, non-genetic information, that can cause phenotypic differences between animals (microbiota, epigenetic), can be transmitted physically from one generation to the next, thus contributing to the inheritance of a trait. The path coefficients of transmission of these non-genetic inherited factors are unknown. We propose a method to estimate them from the similarity matrix between individuals. This matrix can be derived from the direct measurements of the transmitted factors. The method provided consistent estimates on simulated and real data.

Introduction

Animal selection in livestock species is based on genetic inheritance. However, non-genetic information, that can cause phenotypic differences between animals, can also be transmitted from one generation to the next, thus contributing to the inheritance of a trait. As the genetic information, the non-genetic inherited information can be transmitted physically across generations. This is the case for epigenetic marks (epigenetic inheritance) and other media such as metabolites and symbionts (microbiota inheritance). Because the laws of transmission of the non-genetic inherited factors are unknown, disentangling the different sources of inheritance involved in the phenotypic resemblance between relatives is challenging. To overcome this difficulty, David and Ricard (2019) proposed to combine the different inherited values into a single one in the transmissibility model in the case when phenotypic and pedigree information are available. With the development of new molecular measurement methods, additional low-cost information on inherited factors becomes available, offering new possibilities to study the non-genetic inheritances. The objective of the study is to propose a method to extract, from this additional information, knowledge on the vertical transmission of non-genetic factors useful for selection. The method is evaluated on simulated and real data.

Materials & methods

Methods. Let's consider a similarity matrix B. This similarity matrix is a matrix of resemblance between individuals for a given phenomenon. It can be obtained from knowledge about the law of transmission of a specific phenomenon and population structure or from direct measurements of the transmitted information (measurements of genomic, microbiota or epigenetic information on several individuals). The similarity matrix can be decomposed into 3 parts:

$$B = B_t + B_e + B_E \quad (1)$$

where B_t is the expected vertical transmission relationship matrix, B_e a matrix that accounts for the variability around the expected transmission, and B_E a covariance matrix accounting for the external factors shared by individuals influencing the transmitted inherited factor. The vertical transmission relationship matrix B_t is a symmetric matrix with 1's on the diagonal (without inbreeding) and r_{ij} as off-diagonal entries. For the case of 2 animals i and j with n common ancestors ($l = 1,n$) $r_{ij} = \sum_{l=1}^{n} r_{ij,l}$ with $r_{ij,l} = \lambda_s^{k_{ij,l}} \lambda_d^{q_{ij,l}}$, $k_{ij,l} = k_{il} + k_{jl}$, $q_{ij,l} = q_{il} + q_{jl}$, where k_{il} and q_{il} are the number of sire and dam

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_811, © I. David *et al.* 2022

transmissions between ancestor *l* and animal *I*, respectively; λ_s and λ_d are the sire and dam path coefficients of transmission, respectively. Their values are useful to evaluate if selection is possible (i.e. is the factor transmitted to the next generation? by which way (sire, dam or both)? and with which intensity (value of the path coefficients)?). Depending on the information used to build the similarity matrix and the inherited factor under consideration, all or only part of the decomposition in Equation 1 is to be taken into account. Indeed, to study genetic inheritance, the genetic relationship matrix A is a similarity matrix derived from the pedigree knowing that an animal received half of its DNA from its father and half from its mother (both path coefficients equal to 0.5). It corresponds to the expected Identical-By-Descent relationships (IBD), i.e. the average relationships assuming infinite number of loci: $A = B_t$. Using SNP markers, the 'real' IBD relationships matrix (genomic relationship matrix G) is computed, resulting in a similarity matrix slightly different from A, given the finite genome size (Hill and Weir, 2011), $G = B_t + B_e = A + B_e$. Conversely to genetic, non-genetic inherited factors may be modified by the environment and/or are not stable over time. Thus, microbiota (M) or epigenetic (Γ) similarity matrices, obtained from dedicated measurements (OTUs or methylation) (Camarinha-Silva *et al.*, 2017), include the 3 parts in their decomposition: $M \text{ or } \Gamma = B_t + B_e + B_E$.

The objective of the study is to estimate B_t and its components (the sire and dam path coefficients of transmission) from the similarity matrix B. To estimate components of B_t, the chosen method was to apply the transmissibility model to a realization of pseudo phenotypes y generated by $y \sim N(0,B)$. The transmissibility model is: $y_i = y_i\beta + t_i + e_i$, with $\mathbf{t} \sim N(0, \mathbf{B_t}\sigma_t^2)$ and $\mathbf{e} \sim N(0, \mathbf{I}\sigma_e^2)$, where B_t is the previously defined transmission relationship matrix whose unknown components (path coefficients of transmission) are to be estimated and I the identity matrix. Variance components and path coefficients of the transmissibility model are estimated with ASReml and the freely available OWN Fortran program developed by David (2018). Pseudo phenotypes simulation and estimation of the path coefficients from the transmissibility model are repeated in order to obtain the distribution of the path coefficient estimates.

Application to simulated similarity matrix B. A population consisting of four non-overlapping generations without selection was simulated. The founder generation comprised 6 sires and 50 dams that were randomly mated to give birth to 400 offspring (sex-ratio 0.5). Among the progeny, 6 males and 50 females were sampled randomly to be the parents of the next generation. The same process was repeated until the last generation. The final population comprised 1,256 individuals. The transmission relationship matrix B_t was built based on the corresponding pedigree and considering three different sets of sire and dam path coefficients of transmission ({sire,dam}={0.56,0.30},{0.40,0.40} or {0.25,0.50}). The matrix B_e was set to 0.04I. To structure the B_E matrix, the 1,256 animals were randomly grouped into 6 different groups mimicking different environmental effects influencing the inherited factors. The covariances between animals in the same environment were sampled in a uniform distribution]0,1[. The B_E matrix then consisted in 1's on the diagonal and the value of the sampled covariance as off diagonal entries between animals sharing the same environment. We then considered same weight of the environment and transmission on the inherited factors to build the similarity matrix, i.e.: $B = 0.5B_t + B_e + 0.5B_E$. Pseudo phenotypes were simulated using this similarity matrix.

Application to real similarity matrix B. A first population of 660 genotyped animals was used. They were sires and dams of pigs divergently selected for residual feed intake (Gilbert *et al.*, 2017). After quality control, genotypes for 64,487 SNPs were available to compute Ω as $\frac{\mathbf{ZZ^T}}{k}$ where k is twice the sum of heterozygosities of the markers and Z the matrix of genotypes centered by allele frequencies (VanRaden, 2008). Then the genomic relationship matrix G was obtained by scaling Ω, to make the means of the diagonal and off-diagonal elements of G and A matrices equal to each other. We simulated pseudo phenotypes using the similarity matrix A or G (2 different datasets) and, to prevent convergence issues of the transmissibility

model due to no residual variation, added noise (e ~ $N(0, 0.04I)$) to obtain noised pseudo-phenotypes.

A second population consisted in 801 Large White pigs from a divergent selection on microbiota composition (Larzul *et al.*, 2021). The gut microbiota was characterized by sequencing the V3-V4 variable region of the 16S rRNA gene from fecal samples collected on 60-day-old piglets. From the initial 147 genera, unknown classified genera and genera with more than 50% of missing value were removed. The relative abundances of the 65 remaining genera were used to build the similarity matrix B. Missing values (i.e. zero) were imputed using non-parametric multiplicative simple imputation (Martín-Fernández *et al.*, 2003). Then isometric log-ratio (ilr,(Egozcue *et al.*, 2003)) transformation was applied and B was computed as $\frac{\mathbf{XX^T}}{64}$ where X is the 801×64 matrix of by column scaled ilr coordinates (i.e. $X_{ij} = \frac{ilr_{ij} - \overline{ilr_{.j}}}{sd(ilr_{.j})}$). Pseudo phenotypes were simulated using this similarity matrix.

In each case, path coefficients of B_t were estimated by applying the transmissibility model to the simulated (noised) pseudo phenotypes. Fixed effects included in the model were the environment for *simulated* B, a general mean only for *real* B with the A or G matrix, the batch effect for *real* B with the microbiota similarity matrix. Simulation of (noised) pseudo phenotypes and estimation with the transmissibility model were repeated 100 times.

Results

Variance component and path coefficient estimates obtained for pseudo phenotypes generated using simulated and real similarity matrices are presented in Table 1. For simulated B, real A and G matrices, the true values of the path coefficients to estimate are known (equal to the simulated value or to 0.5 for genetic inheritance). In such cases, path coefficient estimates were close to their true value and had low standard errors (0.03 to 0.08). For A and G matrices, the estimated residual variance was equal to 0.04, which corresponds to the value of the variance of the added noise. Transmissibility variance estimates were in line with simulated values (0.50 for simulated B, 1.00 for A and G). Applied to the similarity matrix based on microbiota data, the method provided estimates of path coefficient equal to 0.34 and 0.48 for the sire and dam, respectively. The residual variance estimate was large (0.72) and consequently the transmissibility variance reduced to 0.21.

Table 1. Path coefficient and variance estimates (s.e. in bracket).

	Expected transmission coefficients		Estimated transmission coefficients		Estimated residual variance	Estimated transmissibility variance
	Sire	Dam	Sire	Dam		
Simulated B	0.56	0.30	0.56 (0.08)	0.30 (0.05)	0.36 (0.09)	0.55 (0.12)
	0.40	0.40	0.40 (0.07)	0.41 (0.06)	0.31 (0.07)	0.53 (0.08)
	0.25	0.50	0.25 (0.06)	0.49 (0.05)	0.30 (0.06)	0.50 (0.07)
Pedigree	0.50	0.50	0.49 (0.04)	0.49 (0.03)	0.04 (0.03)	1.02 (0.34)
Genomic	0.50	0.50	0.49 (0.03)	0.49 (0.03)	0.04 (0.03)	0.99 (0.42)
Microbiota	Sire < Dam		0.34 (0.21)	0.48 (0.19)	0.72 (0.22)	0.21 (0.13)

Discussion

From the results, we show that applying the transmissibility model to pseudo phenotypes is efficient for estimating the path coefficients of transmission from a similarity matrix. Other methods could have been used, such as gradient-based descent minimization applied to the inverse of the similarity matrix, but results (not shown) were not satisfactory. Applied to a real microbiota similarity matrix, results were consistent with expectations: a sire path coefficient of transmission lower than that of the dams. Indeed, in livestock, the transmission from one generation to the next of all or part of the intestinal microbiota is most likely the result of physical contact between newborns and the dam (David *et al.*, 2019). But there is also a resemblance between the microbiota of the sire and the offspring because the genetics of the host may also have an influence on the microbiota (Goodrich *et al.*, 2016). It should be noted that results obtained should be considered with caution since microbiota information at the genus level was used to build the similarity matrix while information at the OTUs level is usually preferred to capture the similarity between individuals (Camarinha-Silva *et al.*, 2017). The estimate of the residual variance of the transmissibility model applied to the microbiota similarity matrix was large, suggesting that important related environmental effects were unaccounted for in the model. In the proposed method, the environmental effects influencing the non-genetic inherited factors are not taken into account in the estimated transmission relationship matrix. It is thus hypothesized that modifications of the transmissible potential due to environmental effects occur after the transmission of the non-genetic factors. Studies are underway to include transmitted environmental effects in the estimated transmission relationship matrix.

Acknowledgements

This research was supported by the Geronimo Project funded by the European Union H2020 Programme under grant agreement EU 101000236.

References

Camarinha-Silva, A., Maushammer, M., Wellmann, R., Vital, M., Preuss, S. *et al.* (2017) Genetics 117.200782. https://doi.org/10.1534/genetics.117.200782

David, I. (2018) zenodo. https://doi.org/10.5281/zenodo.1487869

David, I., Canario, L., Combes, S., Demars, J. (2019) Front. Genet. https://doi.org/10.3389/fgene.2019.01058

David, I., Ricard, A. (2019) Genetics 212, 1075-1099. https://doi.org/10.1534/genetics.119.302375

Egozcue, J.J., Pawlowsky-Glahn, V., Mateus-Figueras, G. (2003) Math. Geol. 35, 279–300. https://doi.org/10.1023/A:1023818214614

Gilbert, H., Billon, Y., Brossard, L., Faure, J., Gatellier, P. *et al.* (2017) Animal 11, 1427–1439. https://doi.org/10.1017/S175173111600286X

Goodrich, J.K., Davenport, E.R., Beaumont, M., Jackson, M.A., Knight, R. *et al.* (2016) Cell Host Microbe 19, 731–743. https://doi.org/10.1016/j.chom.2016.04.017

Hill, W., Weir, B. (2011) Genet. Res. 93, 47–64. https://doi.org/10.1017/S0016672310000480

Larzul, C., Borey, M., Billon, Y., Rossignol, M.N., Lemonnier, G., *et al.* (2021). Proc of ISAG 2021, Virtual.

Martín-Fernández, J.A., Barcelol Visal, C., Pawlowsky-Glahn, V. (2003) Math. Geol. 35, 253–278. https://doi.org/10.1023/A:1023866030544

VanRaden, P.M. (2008) J. Dairy Sci. 91, 4414–4423. https://doi.org/10.3168/jds.2007-098

812. Building on Falconer's work on environmental sensitivity and scale

D.L. Waters*, J.H.J. van der Werf and S.A. Clark

School of Environmental & Rural Science, University of New England, Armidale, NSW, 2351, Australia; dwaters9@myune.edu.au

Abstract

Reaction norms are popular for analysing the genetics of environmental sensitivity in animals. The slope of the reaction norm is directly used as a measure of sensitivity. However, some of the variation in slope might be due to differences in the scale of the genetic variance. This problem was considered in 1990 by Falconer in the context of phenotypic reaction norms. This paper extends his work into the current use of genetic random regression reaction norms. A simple post-hoc method to estimate breeding values for environmental sensitivity independent of scale effects while accommodating for higher-order polynomials is proposed. A small example is used for illustration.

Introduction

There are two classes of genotype by environment interaction (G×E). Rank-type G×E occurs when the ranking of genotypes change in different environments, while scale-type G×E occurs when the genetic variance changes in different environments. In the context of breeding programs, rank-type G×E is more important as it means selection candidates change depending on the environment. In contrast, the effect of scale-type G×E can be completely removed by using an appropriate transformation (Falconer and Mackay, 1996). Both types of G×E often act simultaneously in a population (Falconer, 1990).

Reaction norm models (RNMs) estimated via random regression capture both rank and scale-type G×E. This approach uses a covariate, such as the mean performance of a contemporary group, to measure the environmental quality for each animal. The breeding value of each animal is then calculated as a function of the covariate, using the performance of relatives in different environments to infer how it changes. A linear function is often used, resulting in estimated breeding values (EBVs) for the intercept and slope for each animal. It is popular to treat the slope as an EBV for how sensitive a genotype is to different environments i.e. it's environmental sensitivity (ES) (Knap, 2005). However, it could be important to consider that the variation in slope arises from both rank and scale-type G×E (Falconer, 1990).

Falconer (1990) noted that many traits have higher variances when the means are higher. The consequence of this in a reaction norm model is illustrated in Figure 1. As the environmental mean increases, the genotypes spread out. While the rankings of genotypes do not change, there is still variation in the slope of genotypes. In this case, does the slope represent the ES of a genotype? If we assume it does, we would note that animals with a larger intercept have a larger slope. This frequently results in the conclusion that higher performing animals are more sensitive (Rauw and Gomez-Raya, 2015). However, animals with negative intercepts of the same size have equivalently large slopes (i.e. equally as sensitive), just in the opposite direction. Also, a scale transformation that equalises the variance across the environments would completely remove the variation in slope. If we ascribe the correlation between intercept and slope (r_{IS} = ~0.99) to a scale effect, the variation in slope due to re-ranking would be ($1 - r_{IS}^2$) (Falconer, 1990), or effectively to zero.

Falconer did not offer a strong opinion on the treatment of scale in reaction norms, other than to say scale effects should be considered in the analysis of environmental sensitivity (Falconer, 1990). However in Falconer and Mackay (1996), it is recommended that by recognising a true scale effect, such as larger variances

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_812, © D.L. Waters *et al.* 2022

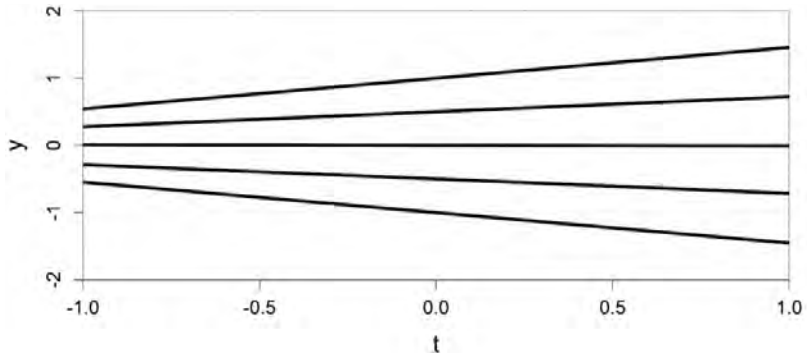


Figure 1. Reaction norm where the variance of y increases with t.

associated with higher body weights in mice, '… there is no need to look deeper into the genetic properties of the character for explanation.' This logic could be applied to the treatment of slope variance in Figure 1.

Hayes *et al.* (2003) accounted for scale in RNMs using a log transformation on phenotypic data. However, if the ratio of genetic and phenotypic variance changes across environments, the genetic variance might not be properly scaled this way (Falconer, 1990). Since random regression RNMs separate genetic and residual variance, a post-analysis focusing only on the genetic variance might be a better course of action. The following method corrects for heterogenous genetic variance on estimates of ES. Ordinary polynomials are used, although it can be easily modified for other polynomials such as Legendre.

Scale-correction method

Consider a random regression reaction norm model that estimates K, a $p \times p$ matrix containing the genetic variances and covariances between the intercept, slope and higher order regression coefficients up to p-1. This also model also estimates breeding values for the intercept (a_0), slope (a_1) and higher order coefficients for each animal. The EBV for animal i in environment j with score t is given by: $\boldsymbol{EBV_{i,j} = a_{0_i} + a_{1_i} \times t_j + \cdots + a_{p-1_i} \times t^{p-1}}$

The genetic variance and covariance matrix (G) between j environments across the trajectory can be obtained by: G = ΦKΦ', where Φ is an $j \times p$ matrix containing a vector of 1's in the first column and a vector of the j environmental scores in the second. The remaining p-2 columns of Φ are given by: $\Phi_{i,q-2} = (\Phi_{i,2})^q$, where q is the order of the polynomial. The diagonal of G gives the genetic variance for each environment. The EBVs for all animals can be standardised (SEBV) within environments based on the diagonal of G:

$$SEBV_{i,j} = \frac{EBV_{i,j}}{\sqrt{G_{j,j}}} \quad (1)$$

When plotted, the SEBVs preserve the points of reranking, but are not affected by scale effects. The SEBVs will not be linear with respect to $\boldsymbol{t}$ (even when using a linear reaction norm), so the average of the first derivative of the curve can be used to approximate the absolute change in SEBV:

$$ES_i = \frac{SEBV_{i,\max(t)} - SEBV_{i,\min(t)}}{\max(t) - \min(t)} \quad (2)$$

This can be interpreted as the 'real' environmental sensitivity (ES_i) of genotype i, representing the change in performance relative to its contemporaries in the absence of heterogenous genetic variance. Another advantage is that it permits higher-order reaction norms to be modelled while still capturing the ES into a single value. The average first derivative could also be explored over different segments of the trajectory.

Applied to a small example

Genetic reaction norms were simulated under zero, medium and high scale effects across environments (t) -1 to 1 units. The variance in intercept and slope was fixed to 1 and 0.4, while the correlation between them was 0, 0.5 and 0.95 for low, medium and high scenarios respectively. For each scenario, 1000 reaction norms were simulated and adjusted for scale following the scale-correction method. A subset of 5 genotypes were drawn from the population in the medium scenario and plotted in Figure 2a. This scenario represented a modest G×E interaction; the genetic correlation between environments -0.75 and 0.75 units was 0.62, while the genetic variance increased from 0.79 to 1.37. For a study with similar G×E levels, see (Madsen *et al.*, 2018). Equation 1 was applied to the population, which created Figure 2b. In this plot, the points of re-ranking were the same as in Figure 2a, however the variance of the SEBV was constant over t and the reaction norms were non-linear. The absolute change in SEBV was then calculated using Equation 2 and is illustrated in a sensitivity plot (Figure 2c). The sensitivity plot should not be interpreted directly as a reaction norm, as it is a linear approximation of the true reaction norm in Figure 2b. Rather, it is only used to visualise the sensitivity of each genotype.

Based on the original reaction norm, genotype 1 would be considered more robust than genotype 2, as it has a flatter slope. However, this is interpretation reverses after correcting for scale effects. This is best explained by looking at the standard deviation (SD) bands in Figure 2a. Genotype 1 is approximately 2.73 SDs above average when t=-1, but this reduces to 1.68 SDs when t=1. Therefore, its performance is relatively more sensitive compared to genotype 2, which is consistently 2.1 SDs above average across the environments. Despite a flat slope in the original reaction norm, genotype 5 improves in rank as t increases and therefore has a positive slope after correction. Meanwhile, genotypes 3 and 4 are not greatly affected by the scale correction, as the effect of scale is not as large closer to the mean.

The relationships between slope before and after correction under the three scenarios are plotted in Figure 3. The scale-correction did not affect the slope when there were no scale effects. As scale effects increased, the correlation between slope before and after correction decreased.

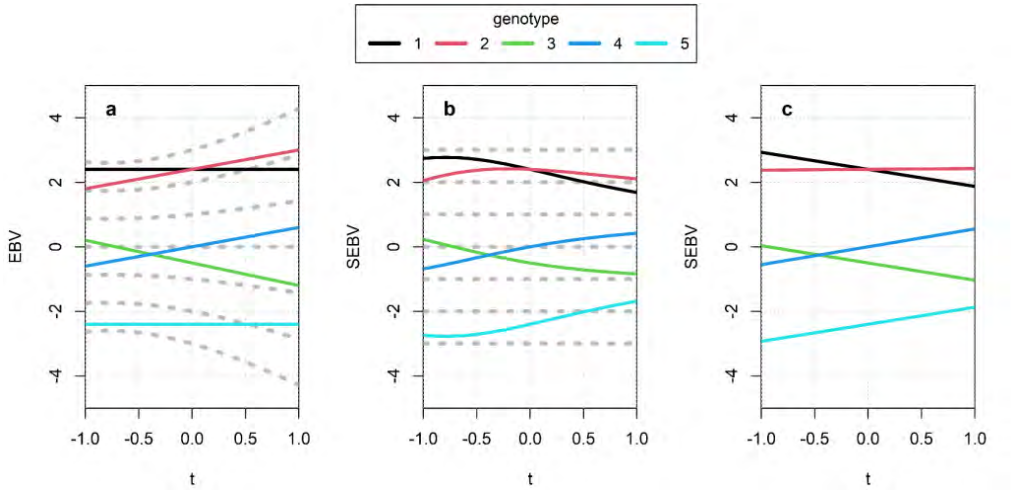


Figure 2. Reaction norms of 5 genotypes coloured individually, with ± standard deviation bands of EBVs given by the broken grey lines a: reaction norms before correction, b: reaction norms with variance standardised, c: sensitivity plot demonstrating change in slope.

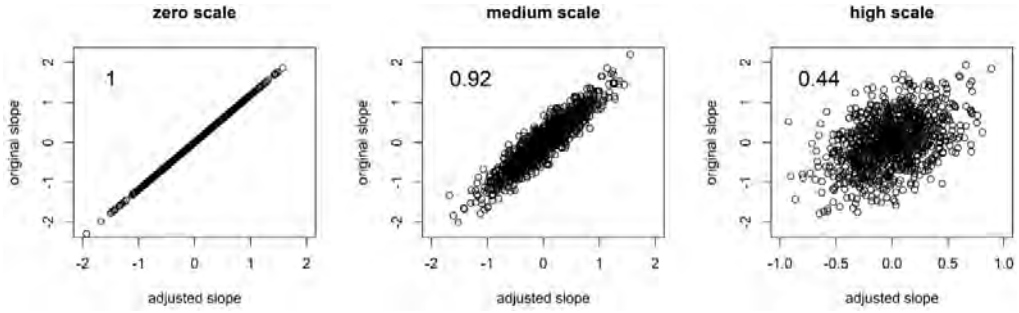


Figure 3. Scatterplot between slope EBVs before (original slope) and after (adjusted slope) being corrected for scale. The correlation is given in the top left.

The method could also be used in the analysis of longitudinal traits, where scale effects are very important. For instance, there is far less variation in weaning weights than in adult weights in livestock. In a linear random regression over time, this would bias the slope of high-weaning weight animals upward and low-weaning weight animals downward. Applying the scale correction would remove this bias and help identify animals that increase in ranking over time (i.e. faster growing animals).

Conclusions

To make progress in genetic selection, it is important to have a clear understanding of the trait being selected for. Environmental sensitivity derived from the slope of a reaction norm was previously not clearly defined, due to the impact of scale effects. A simple method for adjusting for scale effects is given. The resulting estimate of ES captured variation between genotypes in the degree of re-ranking, which could be used directly in breeding programs. The scale-correction could also be useful in longitudinal traits.

References

Falconer D.S. (1990) Genetics Research 56(1):57-70. https://doi.org/10.1017/S0016672300028883

Falconer D.S., and Mackay T.F.C. (1996) Introduction to Quantitative Genetics. Pearson Education Limited, Harlow, England.

Hayes B., Carrick M., Bowman P., and Goddard M.E. (2003) J. Dairy Sci. 86(11):3736-3744. https://doi.org/10.3168/jds.S0022-0302(03)73980-0

Knap P.W. (2005) Aust. J. Exp. Agric. 45(8):763-773. https://doi.org/10.1071/EA05041

Madsen M.D., Madsen P., Nielsen B., Kristensen T.N., Jensen J., and Shirali M. (2018) Sci. J. Anim. Sci. 96(12):4967-4977. https://doi.org/10.1093/jas/sky376

Rauw W.M., Gomez-Raya L. (2015) Front. genet 20(6):310. https://doi.org/10.3389/fgene.2015.00310

813. Impact of a non-Gaussian dependence structure on REML estimation of the bivariate genetic animal model

T. Rohmer[1*], A. Ricard[2,3] and I. David[1]

[1]GenPhySE, Université de Toulouse, INRAE, ENVT, 31326 Castanet Tolosane, France; [2]Université Paris-Saclay, INRAE, AgroParisTech, GABI, Jouy-en-Josas, France; [3]Institut Français du Cheval et de l'Equitation, Pôle développement, Innovation et Recherche, Exmes, France; tom.rohmer@inrare.fr

Abstract

In multiple-trait animal models, variance components are frequently estimated using Restricted Maximum Likelihood method (REML). Such an approach assumes the multivariate normality for the phenotype even if, in practice, this hypothesis is not always realistic. We assessed, using simulation, the impact of a non-Gaussian distribution for the residual term of the mixed model, on the REML estimations. The non-Gaussian distributions were simulated using a copula-based approach. Large populations over 8 generations were simulated using random selection for the 3 first generations and using a truncation selection for the following. Results obtained highlighted the robustness of REML when random selection is performed. On the contrary with a truncation selection process, we observed significant differences with the true parameters, particularly with asymmetric bivariate distributions on the residual part.

Introduction

In a genetic animal context, mixed models are widely used to dissociate the genetic and environmental variance part on the studied phenotypes and perform selection. Extension to the multiple trait model permits to consider the correlations between traits (Meyer (1991), Mrode (2014)).

The principal assumption of this model is the multi normality for the observed phenotype vectors. Since Henderson (1975), the variance-covariance parameters of the model are frequently estimated by REstricted Maximum Likelihood (REML) method, that, under the Gaussian assumption, provides unbiased estimators. However, even if the marginal distributions are Gaussian, the distribution of the multivariate phenotype can be non-Gaussian due to a non-Gaussian dependence structure between the components of the residual vectors of the mixed model.

More precisely let $X = (X_1, X_2)$ be a random vector with cumulative distribution function (c.d.f.) F and marginal c.d.f.s F_1 and F_2 assumed to be continuous. According to Sklar's theorem (Sklar, 1959), it exists a unique function C: $[0,1]^d \rightarrow [0,1]$ such that:

$$F(\boldsymbol{x}) = C\{F_1(x_1), F_2(x_2)\}, \qquad \boldsymbol{x} = (x_1, x_2) \in \mathbb{R}^2$$

The function C is called the copula of the random vector X and characterizes the dependence structure of the vector X. Typical examples of copulas are normal copula (C^N), Clayton copula (C^C), Frank copula (C^F) or Joe copula (C^J), see Nelson (2007). The contour plot of the bivariate distributions with C^C, C^F and C^J copula and Gaussian margins are given in Figure 1. Clayton and Joe distribution are asymmetric (radial sense), and have high dependence in the tail (resp. lower and upper) of the distribution. Frank distribution (with Gaussian margins) is symmetric with slightly more variation in the tails than the Gaussian distribution.

The aim of this paper is to assess the effect of a Gaussian misspecification for the residual part on the REML estimations of the heritabilities and correlations (genetic and residual), when the copula is C^C, C^F or C^J but the margins remain Gaussian.

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_813, © T. Rohmer *et al.* 2022

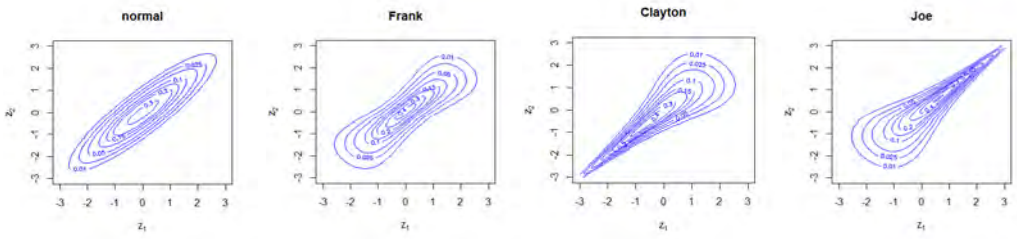


Figure 1. Contour plot of bivariate distributions with Gaussian margins and for copula, the normal copula, Frank's copula, Clayton's copula and Joe's Copula with Kendall's correlation $\tau = 0.7$.

Materials & methods

We simulated simplified pig breeding schemes, following Gonzales *et al.* (2020). The total number of animal animals over 8 generations, including founders was 19,800. In the first 3 generations, choice of reproducers was carried out at random. In the following 5, reproducers were chosen by truncation selection (intra-family selection) from a combination of their 2 breeding values (BVs) estimated by BLUP, with equal weight for each trait, that is to say, from the 3rd generation to the 8th, REML estimation of the variance components and BLUP of the BVs were performed using ASReml software (Gilmour *et al.*, 2015). All the progeny had 2 observed phenotypes and were candidates for selection. Selection rate of 2.9% for the males and 10% for the females were considered (204 breeder females and 10 breeder males by generation including founders). Full-sibs and half-sibs were never mated to limit the inbreeding.

For the founders, the BVs $a_{i,1}$, $a_{i,2}$ for the i^{th} animal were simulated according to a bivariate Gaussian distribution with fixed covariance matrix G. For the other generations, the BVs were simulated through a Mendelian sampling term: $a_{i,j} = 0.5\left(a_{i_f,j} + a_{i_m,j}\right) + M_{ij}$, j = 1,2, where i_f and i_m were respectively the indices of the sire and dam of the i^{th} animal, and $(M_{i,1}, M_{i,2})$ followed a bivariate Gaussian distribution with covariance matrix $G/2$. The residual vectors $(\varepsilon_{i,1}, \varepsilon_{i,2})$ were sampled from a bivariate distribution with standard Gaussian margins and C^N, C^C, C^F or C^J copula whose Kendall's correlation was 0.7 (corresponding Pearson's correlations respectively were 0.891, 0.846, 0.852 and 0.850).

We carried out 1000 Monte Carlo simulations and estimated the heritabilities for the two traits and the genetic and residual correlations. The true genetic variances were 0.18 or 0.67 leading to low and medium heritabilities of 0.153 or 0.401. The true genetic correlations ρ_a were 0.309 or 0.588.

Results

At the end of the 3rd generation (where reproducers were carried out at random), the absolute bias for the heritabilities ranged between 6.77×10^{-6} and 1.43×10^{-3} with SEs between 0.020 and 0.030. The absolute bias for the correlations ranged between 2.4×10^{-7} and 0.010 with SEs between 0.004 and 0.093. No differences were significant (for the t-test at level 5%). The average estimation biases and SEs for the heritabilities evaluated at the end of the 8th generation are shown in Table 1.

When the residual terms followed a bivariate distribution with C^C or C^J copula, we observed significant difference with the true heritability on the trait with medium heritability as soon as the heritability for the two traits are different. In this case, REML over-estimated the heritability for C^J copula and under-estimated the heritability for C^C copula. For C^F copula, we did not observe any significant difference with the theoretical values. The average estimation biases and SEs for the correlations evaluated at the end of the 8th generation are shown in Table 2.

True parameters				Estimated heritabilities							
				Trait 1				Trait 2			
h_1^2	h_2^2	ρ_a		C^N	C^F	C^C	C^J	C^N	C^F	C^C	C^J
0.153	0.153	0.309	bias	0.002	-0.001	-0.004	0.018	0.002	0.000	-0.004	0.019
			SE	0.012	0.012	0.011	0.017	0.012	0.012	0.011	0.016
0.153	0.401	0.309	bias	0.001	0.005	0.015	-0.015	0.004	-0.023	-0.052*	0.057*
			SE	0.014	0.015	0.015	0.013	0.013	0.015	0.014	0.015
0.401	0.401	0.309	bias	0.004	-0.000	-0.004	0.027	0.004	0.000	-0.004	0.027
			SE	0.016	0.016	0.016	0.017	0.015	0.016	0.016	0.018
0.153	0.153	0.588	bias	0.002	0.002	0.002	0.010	0.002	0.003	0.003	0.009
			SE	0.012	0.012	0.011	0.015	0.012	0.012	0.012	0.014
0.153	0.401	0.588	bias	0.001	0.004	0.012	-0.006	0.004	-0.024	-0.051*	0.059*
			SE	0.014	0.015	0.015	0.012	0.014	0.015	0.014	0.015
0.401	0.401	0.588	bias	0.004	0.006	0.008	0.020	0.003	0.005	0.009	0.020
			SE	0.016	0.016	0.016	0.017	0.016	0.016	0.016	0.018

Biases and SEs were obtained from 1 000 simulations. Residual copulas were normal(C^N), Frank(C^F), Clayton(C^C) and Joe(C^J). True genetic correlations were $\rho_a \in \{0.309, 0.588\}$.
'*': significant difference between estimated and true heritability for the t-test at level $\alpha = 0.05$

Table 1. Table of bias and SE of estimated heritabilities.

True parameters				Estimated parameters							
				genetic correlations				residual correlations			
h_1^2	h_2^2	ρ_a		C^N	C^F	C^C	C^J	C^N	C^F	C^C	C^J
0.153	0.153	0.309	bias	-0.005	0.048	0.130*	-0.161*	-0.001	-0.006	-0.016*	0.023*
			SE	0.051	0.051	0.046	0.058	0.003	0.004	0.004	0.007
0.153	0.401	0.309	bias	-0.003	0.009	0.059	0.031	-0.001	-0.012*	-0.031*	0.013*
			SE	0.038	0.038	0.038	0.036	0.004	0.005	0.006	0.006
0.401	0.401	0.309	bias	-0.003	0.060*	0.140*	-0.128*	-0.001	-0.025*	-0.056*	0.060*
			SE	0.029	0.030	0.027	0.032	0.006	0.008	0.007	0.009
0.153	0.153	0.588	bias	-0.003	0.034	0.089*	-0.108*	-0.000	-0.004	-0.010*	0.013*
			SE	0.035	0.034	0.031	0.044	0.003	0.003	0.003	0.005
0.153	0.401	0.588	bias	-0.001	0.003	0.030	0.039	-0.000	-0.009*	-0.023*	0.011*
			SE	0.027	0.028	0.026	0.026	0.004	0.005	0.005	0.005
0.401	0.401	0.588	bias	-0.001	0.049*	0.106*	-0.128*	-0.001	-0.018*	-0.040*	0.052*
			SE	0.021	0.020	0.018	0.027	0.005	0.006	0.006	0.008

Biases and SEs were obtained from 1 000 simulations. Residual copulas were Gaussian(C^N), Frank(C^F), Clayton(C^C) and Joe(C^J). True genetic correlations were $\rho_a \in \{0.309, 0.588\}$. True residual correlations were 0.891, 0.846, 0.852 and 0.850 respectively for C^N, C^F, C^C and C^J.
'*': significant difference between estimated and true correlations for the t-test at level $\alpha = 0.05$

Table 2. Table of bias and SE of estimated correlations.

Concerning the residual correlations, as soon as at least one of the heritability was medium (0.401), all the non-normal copulas led to significant differences with the true residual correlation. In theses case, C^F and C^C copula led to an under-estimation of the residual correlations and C^J led to an over-estimation. Concerning the genetic correlation, as soon as the theoretical heritabilities were equal, we observed significant under-estimation with C^J copula and over-estimation with C^C copula. Using C^F copula, the difference with the theoretical values appeared significant (over-estimation) for medium heritability only.

Discussion

To qualify the robustness of the REML facing deviation from normality in the bivariate case, we look for symmetric (C^F), and asymmetric bivariate distributions (C^C and C^J) for the residual term.

With random selection, the robustness of the REML was remarkable despite the non-Gaussian distribution on the residuals; neither the asymmetry nor the heaviness of the distributions affected the REML estimations.

At the end of the 8th generation (truncation selection from the 3rd generation), we observed systematic bias for the variance-covariance parameters using asymmetric distribution: for the estimated heritability when the true heritabilities for the 2 traits were different and for the estimated correlation, when the true heritabilities were the same and medium. Even for the symmetric non-Gaussian distribution (C^F), we observed significant differences with the true parameter for the estimated correlations. Hence, with truncation selection based on EBVs of 2 observed phenotypes, the asymmetry of the residual part but also the heaviness of the upper tail affects the REML estimations. Clearly, the estimation biases are caused in particular, by the selection process carried out in the upper right tail of the distribution. In the case of same heritability for the 2 traits, high (resp. low) dependences in the upper right tail of the residual distribution (using C^I resp. C^C) will lead to an overestimation (resp. underestimation) of the residual correlations and consequently an underestimation (resp. overestimation) of the genetic correlations. In the case of different heritability for the 2 traits, presumably the selection alters the variance of the BV for the trait with medium heritability which causes under or over estimation of the heritability using C^I and C^C.

References

Gilmour, A. R., Gogel, B. J., Cullis, B. R., Welham, S., & Thompson, R. (2015). ASReml user guide release 4.1 structural specification. *Hemel hempstead: VSN international ltd*. Available at: https://www.hpc.iastate.edu/sites/default/files/uploads/ASREML/UserGuideStructural.pdf

González-Diéguez, D., Tusell, L., Bouquet, A., Legarra, A., & Vitezica, Z. G. (2020). *G3*, *10*(8) 2829-2841. https://doi.org/10.1534/g3.120.401376

Henderson, C. R. (1975). *Biometrics*, 423-447. https://doi.org/10.2307/2529430

Meyer, K. (1991. *GSE*, *23*(1), 67-83. https://doi.org/10.1186/1297-9686-23-1-67

Mrode, R. A. (2014). Linear models for the prediction of animal breeding values. *Cabi*

Nelsen RB. (2007). An introduction to copulas. *Springer Science & Business Media*.

Sklar, M. (1959). Fonctions de répartition à n-dimensions et leurs marges. *Publ. inst. statist. univ. Paris*, *8*, 229-231

814. Impact of genotyping strategy on the accuracy of genomic prediction for survival in pigs

T. Liu[1,2*], B. Nielsen[2,3], O.F. Christensen[2], M.S. Lund[2] and G. Su[2]

[1]Institute of Animal Science, Guangdong Academy of Agricultural Sciences, Dafeng 1st Street 1, Wushan, Tianhe District, 510640 Guangzhou, China; [2]Center for Quantitative Genetics and Genomics, Aarhus University, Blichers Alle 20, 8830 Tjele, Denmark; [3]SEGES Pig Research Centre, Axelborg, Axeltorv 3, 1609 Copenhagen, Denmark; liutfei@gmail.com

Abstract

This study compared four genotyping scenarios of genomic prediction for survival in a simulated pig population with survival rate of 80%. Breeding values were predicted using a linear mixed model and a logit model, based on individual records. The scenarios in order of prediction accuracy were: genotyping all individuals, randomly genotyping 80% individuals, genotyping alive individuals, and no genotyping. Using variance components estimated from pedigree-based model, genomic predictions with genotypes only from alive animals were unbiased. Given the same number of genotyped individuals, random genotyping 80% individuals led to higher prediction accuracy than only genotyping alive individuals, indicating the importance of genotype information of dead pigs. Moreover, the linear and the logit models achieved similar accuracy. We conclude that genotyping dead individuals should be considered, and a linear model can be used for predicting breeding values of survival in pigs.

Introduction

Survival rate can affect not only economic efficiency, but also welfare in pigs. Genetic improvement will be a good solution to improve pig survival. Compared with traditional pedigree-based selection, genomic selection can increase of accuracy of estimated breeding value (EBV) of mortality rate up to 50% (Guo *et al.*, 2015;Knol *et al.*, 2016). Usually, survival rate is recorded as phenotypes of sow. However, survival rate is a complex trait that is also affected by the pig's own genotype. It may therefore be more appropriate to perform genetic evaluation of survival at individual level. Genomic prediction for survival at individual level may encounter problems in the case where dead individuals do not have genotypes. Using only the genotyping data of living individuals may cause bias of genomic prediction. Furthermore, at individual level, survival is a binary trait, and therefore a logit model could be more appropriate, and thus may lead to higher accuracy than a linear mixed model. The objective of this study was to: (1) investigate genetic parameters estimated using a linear model and a logit model; (2) explore the impact of genomic information of dead individuals on genomic prediction; (3) compare the accuracy of genomic prediction of a linear model and the logit model for survival in pigs.

Materials & methods

Data simulation. The data was simulated using QMSim software (Sargolzaei and Schenkel, 2009) mimicking a pig population. We simulated 18 chromosomes, each 100cM with 3,100 markers and 50 QTLs. The simulation started with a founder population of 200 males and 200 females, and went through 300 historical generations to generate linkage disequilibrium between markers and QTLs. After historical generations, a base population was created and went through eight non-overlapping generations. In each generation, 30 sire and 300 dams were randomly selected from alive animals, a sire mated 10 dams randomly, and each dam produced one litter. The litter size was not generated as a genetic trait but randomly assigned to 10, 12, 14, 16, or 18 with the probabilities 0.02, 0.14, 0.68, 0.14, 0.02, and sex ratio was 1:1.

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_814, © T. Liu *et al.* 2022

The liability of an individual to be alive was generated as the sum of direct additive genetic effect (a) of the individual, maternal additive genetic effect (m) of the dam, litter effect and random residual, and independent of litter size. The heritabilities in terms of *a* and *m* as well as proportion of litter variance were all set as 0.04, and genetic correlation between *a* and *m* was 0.3. The phenotype on observed scale was scored as 1 if the liability was the top 80%, and otherwise 0. Four genotyping scenarios were studied: (1) all pigs were genotyped (G_all); (2) only alive pigs (80%) (G_alive); (3) 80% of pigs randomly selected from the whole population (G80_ran); (4) no pig was genotyped (G_none). In total, 30 replicates were simulated for the four scenarios. The reported results were the mean and standard deviation of the 30 replicates.

Statistical analysis. The data were analyzed using a linear mixed model and a logit model.

The linear mixed model is:

$$y = 1\mu + W_l l + Z_a a + Z_m m + e \quad (1)$$

where y is the vector of the observed phenotypes (0 or 1), 1 is a vector of 1s, μ is overall mean, l is the vector of litter effects, a is the vector of direct additive genetic effects, m is the vector of maternal additive genetic effects, and e is the vector of random residuals. W_l, Z_a, Z_m are incidence matrixes associating l, a, m with *y*.

The logit model is:

$$\eta = 1\mu + W_l l + Z_a a + Z_m m \quad (2)$$

where η is the vector of logit function of probabilities of surviving, $\eta_i = \log_e \frac{\mu_i}{1-\mu_i}$ and μ_i is the expected value of y_i. The other notations are the same as in the linear model, but in logit scale.

For both modes, it was assumed that l, a and m have the following distributions: $\boldsymbol{l} \sim N(\mathbf{0}, \boldsymbol{I}\sigma_l^2)$ and $\begin{bmatrix} \boldsymbol{a} \\ \boldsymbol{m} \end{bmatrix} \sim N\left(\mathbf{0}, \boldsymbol{G_0} \begin{bmatrix} \sigma_a^2 & \sigma_{am} \\ \sigma_{am} & \sigma_m^2 \end{bmatrix} \otimes \mathbf{K}\right)$ where σ_l^2, σ_a^2, σ_m^2 and σ_{am} are the variance of *l, a, m* and covariance between *a* and *m*, respectively, and K is an additive genetic relationship matrix. When using traditional BLUP method, K was constructed from pedigree information. When using single-step GBLUP model (ssGBLUP), K was constructed from pedigree and genotypes (Christensen and Lund, 2010). It was assumed that in the linear model, $\boldsymbol{e} \sim N(\mathbf{0}, \boldsymbol{I}\sigma_e^2)$ and in the logit model residuals followed a standard logistic distribution with variance of $\pi^2/3$. In this study, total phenotypic variance was defined as $\sigma_p^2 = \sigma_l^2 + \sigma_a^2 + \sigma_m^2 + \sigma_{am} + \sigma_e^2$.

The variance components estimated from the model with pedigree-based relationship matrix based on data of generations 5-8 were used for prediction of breeding values in all models. The estimation of variances and breeding values were performed using the DMU software (Madsen *et al.*, 2010).

Evaluation of genomic prediction. A validation procedure was performed to assess genomic prediction using different genotyping strategies and statistical models, in which the 5~7th generations were used as reference population, and the 8th generation as validation population. Genomic predictions were evaluated using the following criteria: (1) The correlation between EBV and the true breeding value to assess prediction accuracy; (2) Average true breeding value of the top 5% of individuals in EBVs to assess the realized selection differential; (3) Regression of EBV from whole data on the EBV from reference data (Legarra and Reverter, 2018) to evaluate dispersion bias. Here, EBV from whole data was defined as the EBV obtained from the whole data with genotypes of all individuals.

Results

As shown in Table 1, estimates of parameters were different between the two models due to different scales. By a transformation, the estimates in observed scale from the linear model were consistent with those used in the simulation. For the logit model, the estimated h_a^2 was much lower than the estimated h_m^2, which was unexpected since direct and maternal heritabilities were the same in the simulation. For both models, the estimates of genetic correlations between *a* and *m* (around 0.4) were higher than correlation used in simulation (0.3), and also had large standard error.

Accuracies measured as correlation between total EBV (EBV_a+EBV_m) and total true BV are shown in Table 2. The scenario of G_all had the highest accuracy, and the scenario of no genotyping had the lowest. With the same proportion of genotyped pigs (80%), the accuracy of G80_ran was 4.73 to 11.2 percentage points higher than G_alive, indicating that the genotypes of dead pigs greatly increased the accuracy of EBV. The regression coefficients were around 1 for all genotyping scenarios including G_alive, indicating that genotyping only for alive animal did not lead to biased prediction when using variances estimated from the model with pedigree-based relationship matrix. The results from the logit model were almost the same as those from the linear model.

The 5% top animals selected from the scenario of G_all has the highest true breeding value, followed by G80_ran, then G_alive, and then G_none (Table 3). The results of the linear model and the logit model were very similar, and the difference in true total breeding value of top 5% EBV individuals between the two models was less than 1%.

Table 1. Estimates of variance components in proportion to phenotypic variance using models with pedigree-based relationship matrix.

Parameters	True parameters of liability	Linear model	Logit model
lit^2	0.04	0.019(0.008)	0.026(0.012)
h_a^2	0.04	0.019(0.007)	0.023(0.007)
h_m^2	0.04	0.020(0.010)	0.037(0.016)
r_{am}	0.30	0.401(0.465)	0.400(0.407)

Table 2. Accuracy and bias of EBV.

Model	Genotyping scenario	Accuracy			Unbiasedness		
		All	Genotyped	Nongenotyped	All	Genotyped	Nongenotyped
Linear	G_all	0.511	0.511	/	0.986	0.986	/
	G_alive	0.464	0.477	0.374	1.029	1.006	0.991
	G80_ran	0.487	0.504	0.415	0.983	0.983	0.986
	G_none	0.334	/	0.334	0.985	/	0.985
Logit	G_all	0.511	0.511	/	0.996	0.996	/
	G_alive	0.465	0.479	0.375	1.016	1.015	1.018
	G80_ran	0.487	0.504	0.417	0.997	0.996	0.998
	G_none	0.334	/	0.334	0.993	/	0.993

Table 3. The true breeding value of the top 5% individuals in EBV of survival.

Genotyping scenario	Linear model			Logit model		
	a	*m*	*a+m*	*a*	*m*	*a+m*
G_all	1.847	1.596	3.443	1.781	1.654	3.435
G_alive	1.602	1.604	3.206	1.555	1.658	3.213
G80_ran	1.745	1.529	3.274	1.699	1.609	3.308
G_none	1.104	1.022	2.126	1.122	1.047	2.169

Discussion

In this study, we compared four genotyping scenarios for the accuracy of predictions of survival in pigs. In a pig breeding program, dead animal is usually not genotyped, which may lead to biased estimation of variance components and breeding values when using a genomic model. We carried out a pilot study, and found that a ssGBLUP model with genotypes only from alive animals severely overestimated additive genetic variance and led to a residual variance near to zero (results not shown). Similarly, Wang *et al.* (2020) reported that selective genotyping severely overestimated additive genetic variance. Therefore, in the case of no dead animals being genotyped, genomic prediction should use the variances estimated from the model with pedigree-based relationship matrix to avoid or reduce bias of genomic prediction due to selective genotyping.

With the same size of genotyped individuals, genotyping both alive and dead pigs (G80_ran) has a higher accuracy than genotyping only for alive pigs (G_alive), indicating that the genotypes of dead pigs has an important influence on prediction accuracy. Therefore, it could be a good strategy to genotype dead animals. The predictive powers of the linear model and the logit model are similar, suggesting that linear mixed model is robust to predict breeding value of survival traits. We conclude that the genomic information of dead individuals is very useful, and linear model is a good choice for genomic prediction of survival in pigs.

References

Christensen O.F., and Lund M.S. (2010). Genetics Selection Evolution 42(1): 2. https://doi.org/10.1186/1297-9686-42-2

Guo X., Christensen O.F., Ostersen T., Wang Y., Lund M.S., *et al.* (2015). Journal of Animal Science 93(2): 503-512. https://doi.org/10.2527/jas.2014-8331

Knol E.F., Nielsen B., and Knap P.W. (2016). Animal Frontiers 6(1): 15-22. https://doi.org/10.2527/af.2016-0003

Legarra A., and Reverter A. (2018). Genetics Selection Evolution 50(1): 53. https://doi.org/10.1186/s12711-018-0426-6

Madsen P., Su G., Labouriau R., and Christensen O.F. (Year). Proc. of 9th World Congr. Genet. Appl. Livest. Prod, Leipzig, Germany.

Sargolzaei M., and Schenkel F.S. (2009). Bioinformatics 25(5): 680-681. https://doi.org/10.1093/bioinformatics/btp045

Wang L., Janss L.L., Madsen P., Henshall J., Huang C.-H., *et al.* (2020). Genetics Selection Evolution 52(1): 31. https://doi.org/10.1186/s12711-020-00550-w

815. Effect of a *TLR1A* polymorphism on natural antibodies in poultry

H. Bovenhuis[1*], T.V.L. Berghof[1,2,3] M.H.P.W. Visker[1], J.A.J. Arts[2], J. Visscher[4], J.J. van der Poel[1] and H.K. Parmentier[2]

[1]Animal Breeding and Genomics, Wageningen University & Research, P.O. Box 338, 6700 AH Wageningen, the Netherlands; [2]Adaptation Physiology Group, Wageningen University & Research, P.O. Box 338, 6700 AH Wageningen, the Netherlands; [3]Reproductive Biotechnology, TUM School of Life Sciences, Technical University of Munich, Liesel-Beckmann-Strasse 1, 85354 Freising, Germany; [4]Hendrix Genetics, P.O. Box 114, 5830 AC Boxmeer, the Netherlands; henk.bovenhuis@wur.nl

Abstract

The aim of the current study was to confirm and characterise the effect of a polymorphism in the Toll-like receptor 1 family member A (*TLR1A*) gene on levels of natural antibodies (NAb) in poultry. The study population consisted of a base population and 7 generations of birds from a High and Low selection line. Genotypes for a polymorphism in *TLR1A* were available for chickens in generations 5, 6 and 7. The TLR1A polymorphism had large and significant effects on total NAb levels (IgTotal) and IgM NAb, with full dominance of the *TLR1A* C allele. Significant (*TLR1A*×Generation) interactions were detected for IgM and IgTotal. The frequency of the *TLR1A* C allele was 0.45 in the base population (based on imputed genotypes), 0.66 and 0.04 in generation 7 of the High and Low line, respectively. Results confirm the major effect of this polymorphism in *TLR1A* on NAb levels suggest high sensitivity of the effect to environmental factors.

Introduction

Diseases pose a serious and increasing challenge to livestock production. In recent years, poultry layer production in several countries shifted from battery cages to free roaming housing systems. This increased the number of contacts between chickens, resulting in a higher risk of pathogen spreading. Another recent trend is the restricted use of antibiotics in order to reduce the risk of antibiotic resistance. These developments require other ways to limit the impact of diseases on poultry production and selective breeding might be part of an integrated animal health management approach.

Selective breeding for disease resistance is challenging. Preferably, selection should not be for resistance to one or a few diseases, but for enhanced immune competence resulting in a broad (general) disease resistance. Furthermore, traits that reflect general disease resistance should be easy and cheap to measure and ideally early in the animal's life. NAb could be of value in selection for general disease resistance. NAb show a broad specificity repertoire and act as a first line of defence against infections (Palma *et al.,* 2018). NAb can be measured relatively easily and at a young age in blood samples. We showed in previous studies that selection for high NAb against keyhole limpet hemocyanin (KLH) leads to a higher survival in chickens which are challenged with avian pathogenic *Escherichia coli* (Berghof *et al.,* 2019). Therefore, selection for high NAb levels might contribute to improved general disease resistance in poultry.

In a Genome-Wide Association Study (GWAS) we identified a genomic region on chromosome 4 with a major effect on IgM NAb. This chromosomal region contains the Toll-like receptor 1 family member A (*TLR1A*) gene and a polymorphism in this gene was identified as being the most likely causal variant affecting NAb levels (Berghof *et al.*, 2018). The aim of the current study was to confirm and characterise the effect of this polymorphism in the *TLR1A* gene.

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_815, © H. Bovenhuis *et al.* 2022

Materials & methods

Birds used in this study were from a commercial pure-bred elite white leghorn layer line provided by Hendrix Genetics (Boxmeer, the Netherlands). The study population consisted of a base population, which was unselected for NAb, and 7 subsequent generations of birds from a High and Low selection line. Birds were divergently selected for total KLH-binding NAb titers (IgTotal). Blood plasma samples were taken at 16 weeks of age. An ELISA was performed to determine IgTotal, and KLH-binding IgM and IgG isotypes. A detailed description of the procedure is given by Berghof *et al.* (2018).

The single nucleotide polymorphism (SNP) in *TLR1A* consists of a cytosine (C)/guanine (G) polymorphism, and results in a phenylalanine (F)/leucine (L) amino acid substitution in *TLR1A* at protein position 126. Imputed genotypes for this *TLR1A* SNP were available for the base population (Berghof *et al.,* 2018). These imputed genotypes were used to derive the allele frequency in the base population but were not used in the current study to estimate the effect of the *TLR1A* polymorphism. Genotypes for the *TLR1A* SNP of chickens in generation 5, 6 and 7 of the selection experiment were determined based on a TaqMan assay.

To estimate the effect of the *TLR1A* polymorphism, genotypic and phenotypic information from chickens in generation 5 (n=1105), 6 (n=681) and 7 (n=978) of the selection experiment were available. Genetic differences between selection lines were accounted for in the model by including a line effect. Genotypic effects were estimated using the following model:

$$y_{ijklmn} = \mu + Plate_i + Sex_j + Line_k + TLR1A_l + a_m + d_n + e_{ijklmn} \quad (1)$$

where y_{ijklmn} is IgTotal, IgM, or IgG titer, μ is the mean, $Plate_i$ is the fixed effect of ELISA plate effect (with i = 1 to 166), Sex_j is the fixed effect of sex (with j = 1,2; male or female), $line_k$ is the fixed effect of selection line (with k=1,2; High or Low), $TLR1A_l$ is the effect of the polymorphism in the *TLR1A* gene (with l=1,3; genotypes CC, CG, or GG), a_m is the random additive genetic effect of the m^{th} animal, which is assumed to be distributed as $N(0,A\sigma_A^2)$, d_n is the random effect of the n^{th} dam, which is assumed to be distributed as $N(0,I\sigma_m^2)$, and e_{ijklmn} is the random residual term, which is assumed to be distributed as $N(0,I\sigma_e^2)$. To build the A matrix, pedigree information of 12,443 individuals from 14 generations were used. Generation effects are confounded with ELISA plate effects and therefore 'Generation' was not included as a separate effect in the model.

Additive genetic and maternal environmental variances in these analyses were fixed at the values obtained using all data from the selection experiment consisting of 10,878 birds; heritabilities were fixed at 0.30 for IgM, 0.12 for IgG and 0.12 for IgTotal, maternal environmental effects were fixed at 0.03 for IgM, 0.02 for IgG and 0.03 for IgTotal. Additional analyses were performed to test for possible interactions between *TLR1A* and Sex, TLR1A and Line and TLR1A and Generation. All statistical analyses were performed using ASREML 4 (Gilmour *et al.* 2015).

Theoretically expected changes in *TLR1A* frequencies due to mass selection for IgTotal were approximated using Falconer and Mackay (1996). The coefficient of selection was approximated as:

$$s = i_r \left(\frac{2a}{\sigma_p}\right) \quad (2)$$

where i_r is the standardized realized selection differential, a is the estimated additive genetic effect of the *TLR1A* polymorphism on IgTotal and σ_p is the phenotypic standard deviation of IgTotal. The estimated additive genetic effect of the *TLR1A* polymorphism on IgTotal and the phenotypic standard deviation were

obtained from analyses using model (1). The theoretically expected change in *TLR1A* G frequency was calculated as (Falconer and Mackay, 1996):

$$\Delta q = -sq^2(1 - q) \quad (3)$$

Changes in *TLR1A* G frequency were calculated separately for males and females and subsequently averaged. Approximate changes in *TLR1A* frequencies assume that effects are constant over generations and show complete dominance i.e. assuming that effects of CC and GC on IgTotal are identical.

Results

Table 1 shows that the *TLR1A* polymorphism had a highly significant effect on IgM and IgTotal NAb titers. The CC and CG genotypes had more than 1 titer point higher IgM NAb levels than the GG genotype. The difference between the CC and CG genotypes for NAb IgM titers was small and not significant, suggesting full dominance of the C allele. Similar as for IgM, CC and CG genotypes had higher IgTotal NAb titers than the GG genotypes and the difference between CG and GG genotypes was not significant.

For IgG no significant interactions between the *TLR1A* polymorphism and Sex, Line or Generation were detected.The *TLR1A* effect on IgM and IgTotal suggested full dominance of the C allele and therefore for the interaction analyses we only distinguished two genotype classes: CC/CG and GG. For IgTotal no significant (TLR1A×Sex) interaction was detected. For IgM a significant (TLR1A×Sex) interaction was detected (P=0.004); the effect of the *TLR1A* polymorphism (CC/CG versus GG) in males (1.15) was larger than in females (0.92). No significant (TLR1A×Line) interactions were detected indicating that effects do not differ significantly between the High and Low selection line. Highly significant (TLR1A×Generation) interactions were detected, both for IgM (P<0.001) and IgTotal (P=0.008). For IgTotal the difference between CC/CG and GG was 0.43 titer points in generation 5, 0.23 titer points in generation 6 and 0.71 titer points in generation 7. For IgM the difference between CC/CG and GG was 0.70 titer points in generation 5, 1.22 titer points in generation 6 and 1.28 titer points in generation 7. The frequency of the *TLR1A* C allele was 0.45 in the base population (based on imputed genotypes). The frequency of the C allele in the High line was 0.71 in generation 5, 0.64 in generation 6 and 0.66 in generation 7 (Figure 1). In the Low line the frequency of the C allele was 0.09 in generation 6, 0.06 in generation 6 and 0.04 in generation 7. In generation 5 of the Low line there were 5 chickens with genotype CC and in generation 6 and 7 of the Low line there were no chickens left with the CC genotype. In generation 7 the selection response for IgM due to *TLR1A* allele frequency changes was -0.65 titer points in the Low line and +0.21 titer points in the High line. The segregation of the TLR1A polymorphism therefore resulted in an asymmetrical selection response for IgM.

Table 1. Estimated effects of the *TLR1A* polymorphism on NAb titers based on chickens in generation 5 (n=1,105), 6 (n=681) and 7 (n=978) of a selection experiment.

	IgTotal	IgM	IgG
GG	0	0	0
GC	$0.49_{(0.09)}$	$1.01_{(0.07)}$	$0.22_{(0.10)}$
CC	$0.44_{(0.11)}$	$1.12_{(0.09)}$	$0.07_{(0.13)}$
P	<0.001	<0.001	0.048

Discussion

Highly significant effects of *TLR1A* on IgTotal and IgM reported by Berghof *et al.* (2018) were confirmed in the current study. Estimated genotypic effects of *TLR1A* on IgM and IgTotal were larger than those reported by Berghof *et al.* (2018). The current data allowed testing for the interaction between *TLR1A* and sex, line or generation. Surprisingly, we detected highly significant *TLR1A* by generation interactions for IgM and IgTotal. In the selection experiment, conditions were kept as stable as possible. However, despite this standardization generation-specific variation cannot be avoided. Each generation was raised as a single batch and therefore several factors might be responsible for the observed *TLR1A* by generation interactions. Despite high sanitary status of the research facility, immune status and presence of pathogens can be other major effectors of TLR expression. As far as we know such strong genotype by environment interaction effects have not been reported before for disease resistance-related traits. Possibly because such effects were either not present or not investigated, but it does highlight the importance of the test environment when selecting for KLH-binding NAb, and possibly for disease resistance-related traits in general.

Highly significant effects of *TLR1A* on IgTotal were also confirmed based on the observed allele frequency changes in the High and Low selection line. Results provided unique experimental verification of changes in allele frequencies at a major gene with dominant gene action affecting a quantitative trait that is subjected to mass selection. Theoretically, allele frequencies might not only change due to selection but also due to drift. Increase of inbreeding per generation is low; on average 0.26% in the Low line and 0.31% in the High line.

Therefore, allele frequency changes due to drift are expected to be small. If NAb play an important role in the immune defence of poultry one might expect that purifying (natural) selection would remove alleles with deleterious effects. However, if pure line individuals are kept under strict hygienic conditions in breeding nuclei, their immune response is not seriously challenged, and purifying (natural) selection will not be effective. Reciprocal recurrent selection also might not be effective in removing deleterious alleles

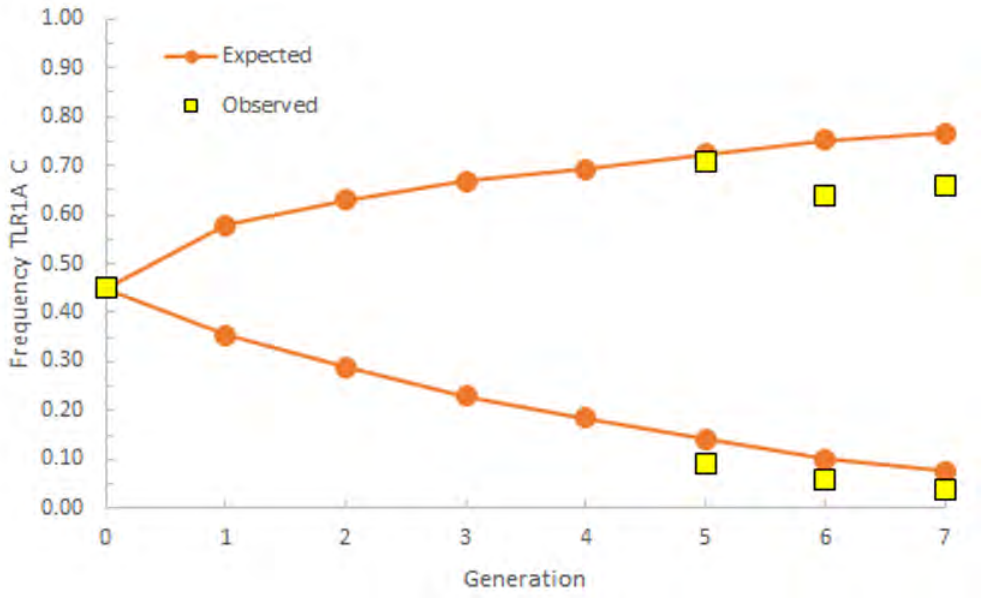


Figure 1. Realized and theoretically expected changes of *TLR1A* C allele frequencies in a High and Low selection line for NAb.

when the other parental line is homozygous for the favourable *TLR1A* allele (CC). Therefore, for situations with genotype by environment interaction and complete dominant mode of gene action, the crossbreeding system might be 'hacked' by deleterious mutations.

References

Berghof T.V.L., Visker M.H.P.W., Arts J.A.J., Parmentier H.K., van der Poel J.J. *et al.* (2018) Front. Immunol. https://doi.org/10.3389/fimmu.2017.01879

Berghof, T.V.L., Matthijs M.G.R., Arts J.A.J, Bovenhuis H., Dwars R.M. *et al.* (2019) Dev. Comp. Immunol. 93:45-57. https://doi.org/10.1016/j.dci.2018.12.007

Falconer, D.S. and T.F.C. Mackay. 1996. Introduction to quantitative genetics. Pearson Education Limited, Essex, England

Gilmour, AR, Gogel BJ, Cullis BR, Welham SJ, Thompson, R. (2015) ASReml User Guide Release 4.1. VSN International, Hemel Hempstead, United Kingdom. 2015

Palma J, Tokarz-Deptuła B, Deptuła J, Deptuła W. (2018) Cent Eur J Immunol. 43:466-475. https://doi.org/10.5114/ceji.2018.81354

816. Geometry meets genomics

L. Gomez-Raya[1], W.M. Rauw[1] and A. Hess[2]

[1]Departamento de Mejora Genética Animal, Instituto Nacional de Investigación y Tecnología Agraria y Alimentaria (INIA-CSIC), Ctra de La Coruña km 7.5, 28040 Madrid, Spain; [2]Department of Agriculture, Veterinary & Rangeland Sciences, University of Nevada, Reno, Mail Stop 202, Reno, NV 89557, USA; luis.gomez.raya@csic.es

Abstract

Geometry of vector spaces is modeled to deal with genomic information. A vector of additive values deviating from the mean is constructed for each SNP (Single Nucleotide Polymorphism). The norm of this vector (length) is the additive standard deviation. We propose the term 'angular linkage disequilibrium (LD)' as the angle, φ_{AB}, between two vectors of additive values corresponding to loci *A* and *B*. In this model, $\cos\varphi_{AB} = r_{AB}$, where r_{AB} is the square root of the standard r^2 measure of LD. Applications of geometry of vector spaces include: (1) the Gram-Schmidt process to eliminate redundant LD information in genomic selection, and (2) estimation of angular and traditional (r^2) LD with a hand calculator. Simulation results showed a full agreement between simulated and estimated LD parameters using either the Expectation-Maximization algorithm or the method using genotype counts suitable for computation with a hand calculator.

Introduction

Geometry is one of the oldest branches of mathematics. Geometry has evolved to include applications in almost all sciences, including art and architecture. One of its most important applications is the geometry of vector spaces. Genomics is the study of genomes. In livestock improvement, this includes genome-wide identification of allelic variants associated with production or resistance to disease (genome-wide association studies, GWAS) and breeding value estimation based on genomic information (Genomic Selection). The objective of this study is to show applications of the geometry of vector spaces to genomics. Linkage disequilibrium (LD) is defined as the angle between two vectors of additive value deviations.

Geometry of abstract vector spaces to define vectors of additive values

A vector is an object that has both a magnitude and a direction. A vector in an *n*-dimensional Euclidean vector space (or R^n) is defined as $\vec{a} = (a_1, a_2, a_3, \ldots, a_n)$, where $a_1, a_2, a_3, \ldots, a_n$ are the components of the vector. We propose to create vectors of additive values $\vec{u}_A$ and $\vec{u}_B$ for two loci *A/a* and *B/b*. For locus *A/a*, genotypes *AA*, *Aa* and *aa* have additive genotypic values α_A, 0, $-\alpha_A$ with frequencies f_{AA}, f_{Aa} and f_{aa}, respectively. The genotypic mean of locus A/a is $\mu_A = \alpha_A f_{AA} - \alpha_A f_{aa} = \alpha_A (f_{AA} - f_{aa})$. The same approach can be used to model locus B/b with effects α_B and with genotypic mean $\mu_B = \alpha_B (f_{BB} - f_{bb})$, where f_{BB} and f_{bb} are the frequencies of genotypes *BB* and *bb*, respectively. When implementing genotypic information in the language of vector spaces in R^n, the components of $\vec{u}_A$ and $\vec{u}_B$ are centered additive genotype values (additive genotype value for the corresponding genotype minus the genotypic mean). The components of vectors $\vec{u}_A$ and $\vec{u}_B$ for individuals with genotypes *AA*, *Aa*, and *aa* or *BB*, *Bb*, *bb* at two loci *A/a* and *B/b* are:

$$u_{AA} = \big(1 - (f_{AA} - f_{aa})\big)\alpha_A;\ u_{Aa} = -(f_{AA} - f_{aa})\alpha_A;\ \ u_{aa} = (-1 - (f_{AA} - f_{aa}))\alpha_A$$

$$u_{BB} = (1 - (f_{BB} - f_{bb}))\alpha_B;\ u_{Bb} = -(f_{BB} - f_{bb})\alpha_B;\ u_{bb} = (-1 - (f_{BB} - f_{bb}))\alpha_B \quad (1)$$

R.F. Veerkamp and Y. de Haas (eds)
CC BY

DOI: 10.3920/978-90-8686-940-4_816, © L. Gomez-Raya *et al.* 2022

The norms or length of $\vec{u}_A$ and $\vec{u}_B$ are the square root of the additive variance of markers *A/a* and *B/b*, respectively. The direction of a vector is the measure of the angle with a horizontal line. The cosine of the angle, φ_{AB}, between two vectors of additive values, $\vec{u}_A$ and $\vec{u}_B$, is:

$$\cos \varphi_{AB} = \frac{\vec{u}_A \cdot \vec{u}_B}{\|\vec{u}_A\| \|\vec{u}_B\|} \quad (2)$$

where $\vec{u}_A \cdot \vec{u}_B$ is the dot product of vectors $\vec{u}_A$ and $\vec{u}_B$, and $\|\vec{u}_A\|$ and $\|\vec{u}_B\|$ are the norms of vectors $\vec{u}_A$ and $\vec{u}_B$, respectively. Equation 2 is the correlation coefficient between a pair of loci A and B, and is usually expressed as its square, r^2_{AB} (Hill and Robertson, 1968).

We define, angular LD, as the angle, φ_{AB}, between vectors $\vec{u}_A$ and $\vec{u}_B$ with value:

$$\varphi_{AB} = arc\ cos\left(\frac{\vec{u}_A \cdot \vec{u}_B}{\|\vec{u}_A\| \|\vec{u}_B\|}\right) \quad (3)$$

Angle φ_{AB} in Equation 3 is in radians but it can be converted to degrees by multiplying φ_{AB} by 180/π. For $\varphi_{AB} = 90^0$, the two vectors are orthogonal and in linkage equilibrium. For $\varphi_{AB} \neq 90^0$, the two vectors are non-orthogonal and the level of dependency is larger for values of φ_{AB} near 0^0 or 180^0; we say that the two loci are in linkage disequilibrium.

To illustrate estimation of angular LD, let us use an example of 10 Iberian sows as shown in Table 1. Columns 2 and 3 list genotypes for two SNPs in positions 17702032 and 18169229 located on SSC1. The vector components for each individual are in columns 4 and 5. The other columns provide dot products and the square of each vector components for the two SNPs necessary to compute angular LD, respectively. From Table 1, we can compute genotype frequencies for the two positions: $f_{AA}^{17702032} = 0.4$; $f_{AG}^{17702032} = 0.5$; $f_{GG}^{17702032} = 0.1$ and $f_{AA}^{18169229} = 0.5$; $f_{AG}^{18169229} = 0.2$; $f_{GG}^{18169229} = 0.3$. The genotypic means in the two positions are $\mu_{17702032} = 0.3\alpha_{17702032}$, and $\mu_{18169229} = 0.2\alpha_{18169229}$, where α_i is the effect of the SNP in position *i*. For this example, the cosine of the angular LD is:

$$\cos \varphi_{AB} = \frac{\vec{u}_A \cdot \vec{u}_B}{\|\vec{u}_A\| \|\vec{u}_B\|} = \frac{2.76}{\sqrt{(4.1)(7.6)}} = 0.494$$

Table 1. Genotypes for 10 individuals at two loci in SSC1 in positions 17702032 and 18169229 of reference map SScroffa v11.1.

Individual	Genotype position		Vector components		Dot product	Square of each vector component in position	
	17702032	18169229	17702032	18169229		17702032	18169229
1	AG	GG	-0.3	-1.2	0.36	0.09	1.44
2	AG	GG	-0.3	-1.2	0.36	0.09	1.44
3	AA	AA	0.7	0.8	0.56	0.49	0.64
4	AG	AA	-0.3	0.8	-0.24	0.09	0.64
5	AG	GG	-0.3	-1.2	0.36	0.09	1.44
6	GG	AG	-1.3	0.2	-0.26	1.69	0.04
7	AA	AA	0.7	0.8	0.56	0.49	0.64
8	AA	AA	0.7	0.8	0.56	0.49	0.64
9	AG	AG	-0.3	0.2	-0.06	0.09	0.04
10	AA	AA	0.7	0.8	0.56	0.49	0.64
Sum					2.76	4.1	7.6

The angular LD is $\varphi_{AB} = arc\ cos(0.494) = 1.054$. In degrees, $\varphi_{AB} = 60.39^0$. It can be shown that under Hardy-Weinberg conditions the equivalence between angular and traditional LD (r^2) is:
$\varphi_{AB} = arc\ cos\left(\frac{\vec{u}_A \cdot \vec{u}_B}{\|\vec{u}_A\|\|\vec{u}_B\|}\right) = arc\cos(r_{AB})$, where r_{AB} is the correlation between additive values for loci A and B (Hill and Robertson, 1968). Some properties of the geometry of vector spaces can be useful when analyzing genomic information. We explore: (1) The Gram-Schmidt process to eliminate redundant LD information in genomic selection; and (2) Estimation of angular and traditional (r^2) LD with a hand calculator.

Gram-Schmidt process to eliminate redundant LD information in genomic selection

Current GBLUP incorporates SNP information independently of other SNPs, ignoring linkage disequilibrium between them. Geometry of vector spaces can mitigate redundant LD information by the so-called Gram-Schmidt process. For example, we can extract information associated to one SNP that is in LD with another SNP. The first step in the Gram-Schmidt process is to create orthogonal vectors $\vec{v}_A$ and $\vec{v}_B$ as follows:

$$\vec{v}_A = \vec{u}_A;\ \vec{v}_B = \vec{u}_B - \left(\frac{\vec{v}_A \cdot \vec{u}_B}{\|\vec{v}_A\|}\right)\vec{v}_A$$

For the example with 10 Iberian genotypes (Table 1), $\vec{v}_B = \vec{u}_B - 0.67317\,\vec{v}_A$ and $\vec{v}_B = (-0.998, -0.998,\ 0.328,\ 1.002, -0.998, 1.077,\ 0.328, 0.328, 0.402, 0.328)$. The Gram-Schmidt process can be applied to multiple SNPs but it is problematic with rounding errors, even with the modified Gram-Schmidt process. We are currently investigating the application of the Gram–Schmidt process removing dependences due to LD locally (affecting a few SNPs in the neighborhood).

Estimation of angular and traditional (r^2) LD with a hand calculator

LD is usually estimated using the Expectation-Maximization (EM) method proposed by Excoffier and Slatkin (1995), which is an iterative method that may be time consuming when estimating LD for hundreds of thousands or millions of SNP pairs. Geometry of vector spaces can facilitate calculations requiring just genotype counts and Equation 1. The dot product $\vec{u}_A \cdot \vec{u}_B$, and the norms of $\|\vec{u}_A\|$ and $\|\vec{u}_B\|$ can be obtained easily after realizing that only three types of elements, corresponding to each of the three genotypes in vectors $\vec{u}_A$ and $\vec{u}_B$, are required. Thus, the dot product in Equation 1 is:

$$\begin{aligned}\vec{u}_A \cdot \vec{u}_B = {} & n_{AABB}(1-\alpha_A)(1-\alpha_B) + n_{AABb}(1-\alpha_A)(-\alpha_B) + n_{AAbb}(1-\alpha_A)(-1-\alpha_B) \\ & + n_{AaBB}(-\alpha_A)(1-\alpha_B) + n_{AaBb}(-\alpha_A)(-\alpha_B) + n_{Aabb}(-\alpha_A)(-1-\alpha_B) \\ & + n_{aaBB}(-1-\alpha_A)(1-\alpha_B) + n_{aaBb}(-1-\alpha_A)(-\alpha_B) + n_{aabb}(-1-\alpha_A)(-1-\alpha_B)\end{aligned} \quad (4)$$

Where n_{ij} is the count of genotypes *ij* (*AABB*, *AABb*, *AAbb*, *AaBB*, *AaBb*, *Aabb, aaBB*, *aaBb*, *aabb*), $\alpha_A = f_{AA} - f_{aa} = \alpha_B = f_{BB} - f_{bb}$. Also, for Equation 1, norms of $\|\vec{u}_A\|$ and $\|\vec{u}_B\|$ can be obtained as:

$$\|\vec{u}_A\| = \sqrt{N[f_{AA}[1-\alpha_A]^2 + [-f_{Aa}(\alpha_A)^2] + f_{aa}[-1-(\alpha_A)]^2]}$$
$$= \sqrt{N[f_{AA} + f_{aa} - [(\alpha_A)]^2]}, \text{ and}$$

$$\|\vec{u}_B\| = \sqrt{N[f_{BB} + f_{bb} - [(\alpha_B)]^2]} \quad (5)$$

where $N = n_{AABB} + n_{AABb} + n_{AAbb} + n_{AaBB} + n_{AaBb} + n_{Aabb} + n_{aaBB} + n_{aaBb} + n_{aabb}$. Therefore, the nine genotypic counts can be used to directly estimate angular and traditional (r^2) LD requiring just a hand calculator when estimating LD for a pair of SNPs. For a large number of SNPs, computation time is much reduced. An example with genotype counts for 435 Iberian sows for genotypes at SNP positions 17702032 and 18169229 on SSC1 (Table 2) yielded an estimate of cos $\varphi_{AB} = 0.3213$ using the method proposed in this study, and

r^2 = 0.1046022 when using the EM algorithm proposed by Excoffier and Slatkin (1995). The angular LD between these two SNPs is 71.26^0.

Results & discussion

Results of a computer simulation with 10,000 replicates for two scenarios regarding allele frequencies and r^2 are given in Table 3. Sample sizes of 100 and 1000 are used. The method using Equations 4 and 5 into 2 and termed *HC-r²* yields the same results as the EM-algorithm (*EM-r²*), and as the simulated parameters. The standard deviation between replicates is slightly higher in method *HC-r²*. Angular LD provides LD information as an angle of the departure of orthogonality, which is 90^0. There is a potential range of applications of geometry of vector spaces to genomics that we are just beginning to delve into, from removing LD in nearby SNPs from Genomic Selection using the Gram-Schmidt process, to construction of blocks of haplotypes using properties of the vector spaces. For example, vector spaces in R^n have the additional property allowing to sum groups of markers that can have an angular LD with the sum of other groups of markers. In a recent paper, we have applied geometry of vectors spaces to construct additive and dominance genomic relationship matrices including orthogonalization by the Gram-Schmidt process (Gomez-Raya *et al.*, 2021).

Table 2. Genotype counts of 435 sows at two loci in in SSC1 in positions 17702032 and 18169229 of reference map SScroffa v11.1.

	BB	*Bb*	*bb*
AA	25	61	26
Aa	77	114	25
aa	68	35	4

Table 3. Average estimates (± s.e.) of angular LD (φ_{AB}), and traditional r^2 estimated by the EM algorithm (*EM-r²*) and by the method in equations (4) and (5) suitable for analysis with a hand calculator (*HC-r²*) in the simulation of 10,000 replicates. The simulated conditions for two loci A and B are: r^2_{AB} = 0.9, f_A = 0.5, f_B = 0.5 and r^2_{AB} = 0.5, f_A = 0.5, f_B = 0.4. n = the total number of individuals.[1]

	n=100			n=1000		
Simulation	φ_{AB}	*EM-r²*	*HC-r²*	φ_{AB}	*EM-r²*	*HC-r²*
S1: r^2_{AB}0.9	17.94^0	0.900	0.900	18.40^0	0.900	0.900
	(±18.04)	(±0.043)	(±0.045)	(±17.95)	(±0.014)	(±0.014)
S2: r^2_{AB}0.5	44.86^0	0.503	0.502	45.00^0	0.500	0.500
	(±44.62)	(±0.071)	(±0.083)	(±44.62)	(±0.022)	(±0.026)

[1]S1: $f_A = 0.5, f_B = 0.5$; S2: $f_A = 0.5, f_B = 0.4$.

Appendix: R-code for the example of 10 Iberian sows

(Remove semicolons in code)

ua <- c(-0.3,-0.3,0.7,-0.3,-0.3,-1.3,0.7,0.7,-0.3,0.7)

ub <- c(-1.2,-1.2,0.8,0.8,-1.2,0.2,0.8,0.8,0.2,0.8)

dotAB=ua%*%ub; dotAA=ua%*%ua; dotBB=ub%*%ub

cos=dotAB/sqrt(dotAA*dotBB); angle= acos(cos)*180/3.1416

b=(dotAB/(dotAA)); vb=ub-as.vector(b)*va

References

Excoffier L., and Slatkin, M. (1995) Molecular biology and evolution 12(5):7. https://doi.org/10.1093/oxfordjournals.molbev.a040269.

Gomez-Raya L., Rauw, W.M., and Dekkers, J.C.M. (2021) Genetics Selection Evolution 53: Article 7:1-18. https://doi.org/10.1186/s12711-020-00589-9.

Hill, W.G., and Robertson, A. (1968). Theoretical and Applied Genetics 38(6):226–231. https://doi.org/10.1007/BF01245622.

Author index

A

Aamand, G.P. 304, 976, 1298, 1314, 1515
Abbink, W. 2395
Abdelsayed, M. 1762
Abdollahi-Arpanahi, R. 1388
Ablondi, M. 3159
Acin-Perez, A. 2399
Acloque, H. 2174, 2281
Addes, M. 3151
Adeleke, M.A. 2464
Adeola, A.C. 2471
Adomako, K. 2441
Adriaens, I. 402, 581, 606, 659
Afonso, J. 2061, 2233, 2293, 2612
Afseth, N.K. 356, 569, 573
Agbeti, W. 2395
Agez, Y. 2568
Aggrey, S.E. 952, 1126, 2441
Agha, S. 524
Aguerre, S. 406, 651
Aguilar, I. 164, 240, 1217, 1335, 1530, 2676, 2741, 2757
Ahariz, N. 2620
Ahmed, R.O. 2471
Aihara, M. 2859
Ajmone-Marsan, P. 741, 886, 1017, 2826
Akanno, E.C. 1262
Alabart, J.L. 3050
Alam, M. 3172
Albuquerque, L.G. 1118, 1254, 1598, 2692
Alcantara, L. 227
Aldridge, M.N. 148, 179, 781
Aleksovski, G. 2508
Alemu, S.W. 1753
Alexandre, P. 1326, 1396
Alexandre, P.A. 1930, 2640
Alexandri, P. 3008
Alex, R. 2729
Aliloo, H. 2628, 2652
Alkhoder, H. 586, 1164, 3155, 3163
Al-Khudhair, A.S. 1463
Allain, C. 288, 520, 1551, 2093
Allain, D. 2924
Allal, F. 2351
Allen, J.M. 2684
Almasi, F. 643
Alom, J. 2069
Alonso-García, M. 3054
Aloryi, K.D. 2471
Alsina, P. 610
Altarriba, J. 786, 1610
Alvarenga, A.B. 1571, 2708
Alves, A.A.C. 1684
Alvestad, R. 569
Amaral, A.J. 1025
Ambike, V.B. 2031
Amer, P.R. 70, 111, 1782, 1897, 1946, 1950, 1958, 2891
Ampudia, F. Tejerina 1021, 1057
Amuzu-Aweh, E.N. 1733, 1917, 2433
Amyes, N. 168
Anacleto, O. 728
Anagnostopoulos, A. 2301, 2782
Andersen-Ranberg, I. 115
Andersen, T. 2875
Andersson, B. 2475, 2479
Andonov, S. 2508, 2839
Andrade, A.N. Nunez 2636
Andrade, B.G.N. 2612
Andrianova, E.L. 3113
Angarita Barajas, B.K. 1205
Angarita, B.K. 1331
An, L. 2456
Antonakos, G. 2920
Antonini, A. 2020
Antonios, S. 1845
Appel, A.K. 1090
Arakawa, A. 3176
Arango, J. 1143
Arari, E. 1848
Araujo, A.C. 1571
Arbeau, T. 1791
Arbizu, C.I. 1114
Arcangioli, M.-A. 2158
Archer, J.A. 2716
Archibald, A.L. 1737, 2281, 3097
Arellano-Vera, W. 2305
Arends, D. 1008, 1567, 3012
Arias, K.D. 1979
Arisman, B.C. 676
Ariyarathne, H.B.P.C. 2822
Arnal, M. 3042
Arranz, J.J. 753, 2909, 2956, 3020, 3054
Arts, J.A.J. 3355
Asadollahpour, H. 886
Asher, L. 508
Aslam, M.L. 2220
Aspeholen Aby, B. 107
Asselstine, V. 2237
Astruc, J-M. 288, 1845, 2960, 3037, 3070
Atashi, H. 156, 814
Athanasiadis, I.N. 1502
Athanassiou, C.G. 2549
Aththar, M.H.F. 2379
Atikpakpe, R. 1791
Atxaerandio, R. 135
Auer, F.J. 466
Auffret, M.D. 2616
Auvray, G. 655
Axford, M.M. 1762, 1770, 2843
Ayin, A. 886
Aymard, P. 708

B

Baazaoui, I. 794, 1848
Baba, T. 1306, 2867
Babigumira, B. 1745
Bach, A. 2257
Baes, C.F. 227, 340, 458, 845, 1181, 2237, 2452, 2798, 2814, 2992
Bahareldin, A.A. 2471
Bahrndorff, S. 2541
Bai, L. 2321
Bailey, R. 728, 810
Bailly, J. 614
Bailly-Salins, A. 296
Bain, W. 168, 211
Baird, H. 168, 886
Bai, X. 696, 3282
Bajek, A. 688
Bakhtin, M. 886
Baldi, F. 2676
Baldi, F.S. 2668, 2672
Balieiro, J.C.C. 2199
Ballan, M. 996, 3249
Baller, J.L. 1856
Ballester, M. 1173, 2174
Bâlteanu, V.A. 886
Bamidele, O. 1757
Bandaranayake, P.C.G. 1881
Banerjee, P. 2612
Banga, C.B. 300, 497, 1725
Banks, R.G. 1804, 2504, 2512, 3062
Banos, G. 1721, 2301, 2782, 2887, 3066
Bapst, B. 647, 2932, 3004
Baranski, M. 2220, 2347, 2355
Barasc, H. 2178, 2333
Barbat, A. 462, 2158, 2162, 2333
Barbey, S. 260, 426
Barden, M. 2301, 2782
Bardou, P. 886
Barfield, D. 886
Bargelloni, L. 2367, 2375, 2383
Barkas, D. 688
Barnard, J. 2069
Barros, C. Pita 2329
Bartolomé, E. 3139
Basarab, J. 1656
Basso, B. 2560, 2568, 2584
Bastiaansen, J.W.M. 590, 1753, 2379, 2407, 2412, 2533
Bastos, A.P. 664
Battagin, M. 1185
Baudement, M.-O. 2265
Baudouin, G. 2528
Baumung, R. 1017, 1057
Baur, A. 1148, 1614, 1837
Becker, D. 2261, 2297, 2786
Becker, G.M. 2913
Bécot, L. 512, 2483
Beder, C. 565
Bédère, N. 512, 2483
Bedford, C. 2782
Bedhiaf-Romdhani, S. 794, 1848
Bed'hom, B. 672
Bee, G. 352
Behrendt, R. 127
Beinat, M. 2281
Bekele Mengistu, S. 2403
Bello, N.M. 1701
Bello, S.F. 2471
Benavides, C. 2475
Ben Braiek, M. 3070
Benfica, L.F. 890
Bengtsson, C. 2802
Bennewitz, J. 336, 1197, 1543, 1626, 1872, 2425, 3294
Ben Zaabza, H. 1515
Benzie, J.A.H. 2403
Benzoni, L. 191
Berger, B. 822, 886
Berghaus, D. 3294
Berghof, T.V.L. 3355
Berglund, B. 2839
Berg, P. 102, 830, 976, 1852, 2769, 3109
Bermann, M. 1217, 1225, 1380, 1523, 1530, 2757
Bernal-Barragán, H. 2605
Bernstein, R. 2556, 2572
Berry, D.P. 53, 131, 143, 316, 745, 1004, 1310, 1547, 1586, 2644, 2720, 2847, 2863, 2948
Bertolini, F. 3217
Bertotto, D. 2367, 2375
Bertrand, C. 231
Bertuzzi, P. 231
Besnard, F. 2158, 2178
Besson, M. 688
Bester-van der Merwe, A. 438
Bestin, A. 688
Bevilacqua, C. 2097
Bhatia, V. 3282
Bhave, K.G. 1893, 2031
Biada, I. 2077
Bidanel, J.P. 2052
Bienefeld, K. 2556, 2572
Bierman, C. 1262
Bierman, C.D. 398
Biffani, S. 373, 553, 598, 906, 2704, 3090, 3335
Bijma, P. 393, 516, 692, 704, 761, 790, 849, 2584, 2588
Billon, Y. 532, 614
Bilton, T.P. 168
Bink, M.C.A.M. 1102, 1233, 1278, 2174, 2329, 3262
Bionda, A. 373
Birkett, M. 410
Biscarini, F. 798, 906, 3090
Bissonnette, N. 2269
Bi, Y. 2456
Bizelis, I. 2224
Bjerring, M. 152

Blackburn, H. 1029, 1033, 1057
Blackburn, H.D. 826
Blair, H.T. 2725
Blaj, I. 1626
Blake, D.P. 2421, 2464
Blanc, F. 442, 696, 2097, 3282
Blanchet, B. 2101, 3245
Blasco, A. 2073, 2077
Blichfeldt, T. 183, 886
Bliznyuk, N. 2879
Bluet, B. 296
Blum, H. 2117
Blunk, I. 2733
Bobbo, T. 553, 598, 906
Boden, M. 2293
Boerman, J.P. 422, 627
Boerner, V. 1428, 1526, 1901, 2765
Boes, J. 1017
Boettcher, P. 1017, 1057
Bogerd, J. 2036
Boggio, G. Martinez 2093
Bohan, A. 2952, 3016
Bohlouli, M. 454
Boichard, D. 231, 256, 260, 426, 712, 1148, 1213, 1420, 1614, 1837, 2158, 2162, 2170, 2178, 2333
Boink, G. 886
Boison, S. 569
Boison, S.A. 2220, 2347, 2355
Boissy, A. 520
Bojkovski, D. 822, 1021
Boland, T.M. 2905
Bolner, M. 3249
Bolormaa, S. 127, 1181
Bonekamp, G. 402, 446, 2040
Bonesmo, H. 115
Bonfatti, V. 2367, 3298
Bonifazi, R. 2704
Bonilha, S.F.M. 474
Bonnafe, G. 520
Bonneau, M. 614
Borchersen, S. 577
Borges, C.A.V. 1697
Bormann, J. 1701
Boshove, A. 2337
Bosse, M. 2249
Bouchez, O. 2568
Boudry, P. 2351
Boulanger, F.X. 2524
Boulot, S. 3245
Bouquet, A. 268, 1680, 2101
Bourguignon, L. 1885
Boussaha, M. 426, 2158, 2162, 2170, 2178
Bouwman, A.C. 537, 561, 1181, 2249, 2329, 2533, 2537
Bovenhuis, H. 418, 1745, 2491, 3355
Bovo, S. 972, 996, 3217, 3249
BovReg consortium 2261, 2297
Bozzi, R. 191
Bradic, L. 2826
Bradley, D.G. 886
Bradley, P.E. 3062
Brajkovic, V. 2826
Brand, Z. 2499
Brard-Fudulea, S. 2524
Brascamp, E.W. 2584, 2588
Brauning, R. 1070
Breen, V. 2445
Breen, V.P. 1684, 1812
Brekke, C. 830
Brenig, B. 2253
Brenot, S. 2924
Breuer, S.A. 3241
Brinke, I. 3302
Brito, L.F. 422, 528, 627, 890, 1571, 1575, 2708, 2932, 2992, 3004
Brochard, M. 1995
Brockmann, G.A. 1008, 1131, 1567, 2495, 3012
Brodsky, L. 3113
Brown, D.J. 2964, 2984, 2996, 3000, 3008, 3062
Brown, H. 410
Brown, W.E. 348
Bruce, A. 2048
Bruce, D.M. 2048
Brügemann, K. 818
Brugman, S. 2036
Bruneau, N. 442
Brunier, J. 688
Bruni, G. 2932, 3004
Brunsvik, P. 284
Bruscadin, J.J. 2061, 2293, 2612
Bryson, B. 168, 211
Bubnič, J. 2580
Büchler, R. 2508
Bugeon, J. 2351
Bugiwati, S.R.A. 886
Buisson, D. 2960, 3037
Buitenhuis, A.J. 152, 493, 1551
Buitkamp, J. 3024
Bulach, T. 2568
Bullen, S.L. 1770
Bullock, K.D. 82
Bunning, H. 223
Bunter, K.L. 3221
Burborough, K.A. 2216
Burgeon, C. 3237
Burke, J.M. 676, 2913, 2964
Burlot, T. 442, 512, 2483
Burns, B.M. 2684
Burns, J.G. 2777
Burton, K.L. 659
Buys, N. 478, 3167
Byrd, M.K. 528, 1575
Byrne, A. 2708
Byrne, T.J. 70, 1958

C

Cachelou, F. 2399
Cai, Y. 802, 886
Cai, Z. 886, 1181, 1286, 1471, 1559, 2126, 2245
Calboli, F.C.F. 2343
Calenge, F. 442
Calle-García, J. 1173
Calus, M. 1278, 1372, 1498
Calus, M.P.L. 53, 761, 849, 1233, 1812, 2491, 2704
Calvo, J.H. 1094, 2012, 3050
Camara, M. 2379, 2407
Camarinha-Silva, A. 2425, 3294
Campbell, D.L.M. 482
Campos, G.S. 1254, 1494
Canario, L. 532, 614
Cancedda, M.G. 716
Canesin, R.C. 474
Cano, M. 2158
Cánovas, A. 2109, 2237, 2798, 2855
Cantão, M.E. 664, 2448
Cantet, R.J.C. 1205, 1331
Cao, J. 276
Capitan, A. 1614, 2158, 2162, 2178, 2333
Cappelloni, M. 972
Capra, E. 741
Carabaño, M.J. 988, 2794
Cardoso, F.F. 1494
Cardoso, T.F. 2061, 2293, 2612
Carillier-Jacquin, C. 1680, 2101, 3245
Carneiro, P.L.S. 1571
Carnier, P. 2367, 3298
Carolan, S. 886
Carolino, I. 960
Carolino, N. 960
Carrara, E.R 890
Carr, C. 1618
Carta, A. 373, 716, 1082, 2972, 3028
Carthy, T.R. 2810
Carvalheiro, R. 890, 910, 1118, 1254, 1598, 2692
Casa, A. 549
Casadio, R. 3217
Casellas, J. 786, 3180
Cason, E.D. 932, 2069, 2749
Cassandro, M. 191, 450, 500, 2704, 2818, 2895, 3311
Cassani, V. 3090
Castilhos, A.M. 308
Castillo-Lopez, E. 430
Casto-Rebollo, C. 2073, 2089
Casu, S. 716, 1082, 2972, 3028
Cate, H.S. 2416
Caulton, A. 1070, 2285
Cauquil, L. 2101
Cavani, L. 348
Cecchinato, A. 1185
Cervantes, I. 1979
Cesarani, A. 373, 968, 1356, 1392, 1523, 2757, 2761, 2928
Cesar, A.S.M. 2199, 2233
Cesbron, N. 2178
Chadyiwa, M.C. 1713
Chagunda, M.G.G. 336, 1729, 1741, 1889
Cha, J. 594
Chakraborty, R. 1897
Chamberlain, A.J. 773, 2216, 2228, 2289, 2597
Chang, H.K. 3172
Chanteloup, L. 2648
Chapard, L. 3167
Chapman, N.C. 2504, 2512
Chapuis, H. 344
Charagu, P. 1102
Charfeddine, N. 135, 1181
Charlier, C. 857, 1622, 2065, 2249, 2261, 2297, 2620
Chase-Topping, M.E. 3212
Chassier, M. 296
Chávez-Galarza, J.C. 1114
Chen, B. 802
Chen, C. 557
Chen, C.J. 545
Chen, C.-Y. 1270, 1490, 1579, 1634
Chen, C.Y. 1135, 1139
Cheng, H. 545, 1143, 1169, 1294, 1482, 1519, 1668
Cheng, H.H. 2456
Cheng, J. 696, 1482, 3282
Cheng, Y. 2130
Cheng, Z. 2065
Chenier, F. 2351
Chen, J.L. 2491
Chen, L.H. 1437
Chen, S.Y. 422, 627
Chen, Y. 312, 802
Chepkwony, M. 1737
Chepsiror, C.K. 215
Chichero, D. 2359
Chinchilla-Vargas, J. 2806, 3241
Chinyere, E. 1909
Chitneedi, P.K. 1181, 2170, 2277
Chiwanga, G.H. 1733
Cho, E. 594, 1660, 2429, 2460
Choi, I. 2121
Choi, T.J. 3172
Cho, S. 594, 1660, 2429, 2460
Chottin, C. 426
Chouicha, N. 1733, 1917, 2433
Christensen, O.F. 1152, 1376, 1384, 1459, 2093, 3204, 3229, 3351
Chud, T.S. 1181
Chuma-Alvarez, J.L. 1983
Chung, K. 2121
Chung, Y. 2121, 2460
Chung, Y.J. 1594
Chu, T.T. 361, 388, 2437, 3208
Ciani, E. 794
Ciappesoni, G. 107, 160, 164, 195
Cichero, D. 2371
Ciconello, F.N. 2199
Cieleń, G. 3253
Čílová, D. 3093

Cimmino, R. 3335
Ciobanu, D.C. 2142, 2166
Civita, M. Di 2818
Çınar Kul, B. 886
Clark, E. 886, 3097
Clark, E.L. 1717, 2261, 2297, 2317
Clarke, R. 1070
Clarke, S. 1070
Clarke, S.M. 2285, 2387
Clark, R. 2261, 2297
Clark, S. 684, 1201, 2628, 2632, 2652, 2700
Clark, S.A. 1975, 3343
Clasen, J.B. 1432, 1828
Clement, V. 3042
Cleveland, M.A. 324, 2616
Cloete, S.W.P. 2499, 3078, 3081
Cocks, B.G. 127
Coffey, M. 541, 1229, 2301, 2782, 2980
Colbry, D. 557
Coldebella, A. 2448
Cole, J.B. 80, 806, 1774
Colleau, J.J. 1841
Colli, L. 886, 1017, 2826
Comte, R. 3245
Congiu, M. 2928
Conington, J. 107, 3066
Connelley, T. 1737
Conte, Y. Le 2568
Cook, J. 1820
Copley, J. 765
Copley, J.P. 2684
Coppieters, W. 857, 2065, 2249, 2620
Corbeau, J. 2162
Corbet, N.J. 2684
Cordeiro, J.M. 1025
Correddu, F. 2928
Corredor, F-A. 1114
Cortellari, M. 373
Cortés-Hernández, J.G. 2753
Costa, A. 418
Costa Monteiro Moreira, G. 1622, 2065
Costilla, R.C. 769
Cotticelli, A. 598
Couldrey, C. 2285, 2790
Couteller, L. 3270
Coutinho, L.L. 2061, 2199, 2233, 2293, 2612
Cozzi, G. 659
Cozzi, P. 553, 798, 906, 3090
Crepaldi, P. 373
Crespo-Piazuelo, D. 2174
Crichan, H. 3151
Croiseau, P. 1148, 1213, 1274, 2170
Cromie, A. 2720
Cromie, A.R. 131, 143
Crooijmans, R. 1037, 1753
Crooijmans, R.P.M.A. 2241, 2329
Crossa, J. 1676
Crosson, P. 143
Crotta, M. 2421
Croue, I. 1148
Crowe, M.A. 2065
Crowley, J. 131, 143, 207
Crysnanto, D. 1066
Cubric-Curik, V. 886, 894, 984, 2826, 3101
Cumer, T. 886
Curik, I. 874, 894, 910, 984, 2826, 3101
Cuyabano, B. 1148, 1193, 1213
Cuyabano, B.C.D. 219, 231, 1420
Cyrillo, J.N.S.G. 890

D

Dachs, N. 2117
Dadousis, C. 2421
Daetwyler, H.D. 127, 737
Dagnachew, B. 2220
Dagnachew, B.S. 1442
Dagong, M.I.A. 886
Dahle, B. 2508
Dahl, M. 1459
Dainat, B. 2568
Dai, X. 2301, 2421, 2782
Dale, L.M. 466
Dall'Olio, S. 3217
Daly, A. 565
Daly, K.G. 886
D'Ambrosio, J. 2399
Danchin-Burge, C. 882, 1045, 1057, 2158
Danchin, C. 1021, 2178
Da Silva, A. 886
da Silva, D.A. 1598
da Silva, V.H. 2061
Dauben, C.M. 1090
Davey, M. 2281
David, I. 272, 614, 1970, 3339, 3347
Davière, J.-B. 462
Davila, K.M. Sarlo 2636, 3192
Davoodi, P. 2467
Davoudi, P. 1652
D.Duclos 1021
de Almeida, V.V. 2199
Dearden, P.K. 2564, 2576
De Barbieri, I. 107, 160, 164, 195
Dechow, C. 777
Dechow, C.D. 1033
Decker, J.E. 82, 1555
De Clercq, K. 668
Decourtye, A. 2568
Deerenberg, R. 2359, 2371
Dee, S. 700
Degano, L. 1392
Degremont, L. 2351
de Haas, Y. 148, 179, 393, 402, 1181
Dehnavi, E. 1962
Deighton, M.H. 127
de Jong, G. 280, 737
De Jong, M.C.M. 692, 704

Dekkers, J. 1290, 1482, 2433
Dekkers, J.C.M. 248, 398, 696, 1688, 1733, 1824, 1917, 1999, 2321, 3212, 3282
de Klerk, B. 508, 537, 561, 781
de Koning, D.J 668
Delaby, L. 316
Delafosse, A. 712
Delanoue, E. 2052
Delany, M. 2456
Delgado, J.V. 1037
de las Heras-Saldana, S. 2632, 3046
de los Campos, G. 252, 1270, 1486, 1490, 1634
Delval, E. 520
Demars, J. 532
Dematawewa, C.M.B. 1745, 1921
Demyda-Peyrás, S. 2020
Denholm, S.J. 541
Deniz, M. 474
Denizot, V. 426
de Oliveira Junior, G.A. 845
De Oliveira, T.A. 553
Deretz, S. 3143
Derežanin, L. 964
Derks, M. 1189, 1753, 1876
Derks, M.F.L. 1278, 2134, 2241, 2313, 2329, 2337, 3225
Deru, V. 1680
Déru, V. 2101
Desai, A. 643
Descamps, D. 426
Desire, S. 1579, 1583
de Sousa, K.T. 474
Dessie, T. 1721, 1753, 1757
De Vries, A. 2879
Dewhurst, R.J. 324, 2616
De Wit, A.A.C. 2016, 2036
Dhanikachalam, V. 1602
Dhorne-Pollet, S. 3151
Dias, L.T. 474
Diaz, C. 988, 2794
Diaz, J. Ramirez 906
Difford, G.F. 176, 284, 569
Dimauro, C. 1392, 2928
Ding, J.Q. 1160
Dinh, P. 2121
Diniz, W.J.S. 2061, 2293, 2612
Djebali, S. 2281
Djikeng, A. 410, 1717, 1741
do Bem, R.D. 890
Dobrzański, J. 2134
Dodd, E. 765
Dodds, K.G. 168, 2285, 2387
Do, D.N. 837, 1652
Doekes, H.P. 516, 790, 3125
Doeschl-Wilson, A. 524, 724, 728, 732, 1237, 1475, 2887, 3212
Dominguez-Gasca, N. 2475
Donkpegan, A.S.L. 2524
Dønnem, I. 107
Doornweerd, J.E. 537, 561
Dorea, J.R.R. 602
Dorji, T. 886
Dorso, L. 2178
Dossa, L.H. 818
Doucette, J.S. 422, 627
Douhard, F. 264, 1987
Douls, S. 520
Doutre, L. 3037
Dovč, P. 822, 3101
Dozias, D. 2648
Drew, L. 886
Drögemüller, C. 886, 1074, 1664
Drouilhet, L. 2924
Druet, T. 857, 1012, 1622, 2065, 2249, 2620
Drzaic, I. 894, 984
Duarte, J.L. Gualdrón 2620
Dube, B. 300, 438, 497
Duchêne, J.-C. 2924
Duclos, D. 1037, 1045
Duclos, R. 2052
Ducos, A. 1987
Ducro, B. 203, 631
Ducrocq, V. 256, 340, 1148, 1213, 1246, 1258, 1885, 1905, 1995, 2749, 3323
Dudemaine, P.L. 2269
Duenk, P. 761, 1233
Duff, C. 1930
Duff, C.J. 1396, 2700
Dufrasne, M. 1318
Dugas, E. 2924
Duijvesteijn, N. 1791
Du, M. 2556, 2572
Dunkelberger, J.R. 700
Dunne, E. 2905, 2944, 2948
Dunne, F.L. 2863
Dunn, I.C. 2475, 2479
Dupont-Nivet, M. 688, 2391, 2399
Dupont, S. 2261
Durán-Alvarez, C. 2871
Durand, C. 520
Durland, E. 569
Dürr, J.W. 1774
Duthie, C.A. 2616
Dyck, M.K. 696, 3282
Dzama, K. 438, 3081

E

Eck, K. 2154
Econogene Consortium 886
Edel, C. 1302, 1416
Eding, H. 1368
Ediriweera, T.K. 1660, 2429, 2460
Eenennaam, A.L. Van 2044, 2056
E, G. 886
Egger-Danner, C. 418, 466
Egidy, G. 2097
Egito, A.A. 2601

Eguiarte, L.E. 2871
Ehsani, A. 1934, 2467
Eiríksson, J.H. 1384
Ekine-Dzivenu, C.C. 1709, 1729
Eleftheriou, E. 2549
Eler, J.P. 2668, 2672
Ellen, E. 2016
Ellen, E.D. 393, 508, 537, 2085, 2533, 2537
Elmes, S.N. 211
El-Ouazizi El-Kahia, L. 1979
Elsen, J.M. 1833
Elzo, M.A. 2636
Elzo, M.E. 1618
Emmerling, R. 1302, 1416
Enez, F. 2351
Engelbrecht, A. 2499
Engelmann, S. 2786
Enger, E.G. 356, 365, 3290
Enger, E. Gjerlaug 115
Engle, B. 1106
Engle, B.N. 769, 2597
Engler, C. 3151
Eory, L. 3097
Eory, V. 2777
Eriksson, S. 944, 2712, 2802, 3159
Esatu, W. 1721
Escouflaire, C. 1614, 2158
Esfandyari, H. 1872
Espigolan, R. 2668, 2672, 2676
Estany, J. 1098, 3278
Esteban Blanco, C. 3054
Esteban-Blanco, C. 2909, 2956, 3020
Evans, K.M. 1643
Evans, R. 1310, 1428, 1506, 1547
Evans, R.D. 143, 207, 565, 1787, 2644, 2680, 2720, 2863, 2891
Ewaoluwagbemiga, E.O. 352
Eynard, S.E. 2568

F

Fabre, S. 3070
Facchini, E. 2545
Facy, M.L. 2660
Faggion, S. 2367, 3298
Fahey, A.G. 1160
Fair, M. 2069
Falchi, L. 968
Falker-Gieske, C. 1086, 1626
Fanalli, S.L. 2199
Fanelli, F. 3217
Fang, L. 1668, 2321
Fang, Y. 3212
Faraut, T. 861, 886
Farrell, L. 107, 2944, 2948, 2952
Farrell, L.J. 3016
Fassier, T. 276, 3074
Fatoba, A.J. 2464
Faverdin, P. 256, 260
Faza, D.R.L.R. 1697
Femenia, M. 2178
Fennessy, P.F. 70, 2564, 2576
Ferdosi, M.H. 737, 1221, 1450, 1820
Ferenčaković, M. 910
Fernandes, A.F.A. 1684
Fernandes, A.R. 1697
Fernandes Júnior, G.A. 1118, 1254
Fernandes Junior, J.A. 2601
Fernandes, L.T. 664, 2448
Fernández, A. 1094, 1181, 2012, 2383
Fernández, J. 732, 841, 882, 2383, 3139
Fernandez-Santos, M.R. 988
Fernando, P.R.M.K. 1921
Fernando, R. 1290, 1482, 1519
Fernando, R.L 1688
Fernando, S.C. 434
Ferraresso, S. 2375
Ferrari, V. 500, 2895
Ferraz, J.B.S. 2668, 2672
Ferreira, B.R. 1737
Ferret, C. 426
Ferro-Mauricio, R.D. 1114
Fickel, J. 964
Field, C.J. 696, 3282
Fikse, W.F. 304, 486, 490, 1432, 2712, 2802
Filée, J. 2528
Filho, A. Luchiari 2199
Filho, I. Carvalho 1598, 2692
Finocchiaro, R. 191, 450, 500, 2818, 2895, 3311
Fisch, A. 1737
Fischer, D. 1181, 2830
Fitzsimmons, C. 1656
Fix, J. 1692
Fix, J.S. 3266
Flatrès-Grall, L. 272, 1672, 3327
Fleming, R.H. 2475
Fleyshman, D.I. 3113
Flori, L. 1012
Floyd, T. 2178
Fodor, I. 537, 561
Foissac, S. 2281
Foister, S. 524
Folch, J. 3050
Fonseca, F.N. 664
Fonseca, L.F.S. 1118, 1254
Fonseca, P.A.S. 1598, 2109, 2798
Fontanesi, L. 972, 996, 2516, 3217, 3249
Fontanillas, E. 655
Fordyce, G. 765, 2684
Formoso-Rafferty, N. 1979
Forneris, N.S. 1205, 1331
Fortes, M.R.S. 1209, 1282, 1563, 2293, 2640
Fortin, F. 696, 3282
Fortuna, G.M. 1266
Fortun-Lamothe, L. 1987
Forutan, M. 1106
Fosso, B. 2421

Foster, S. 410
Fotinga-Meiring, F. 2016
Foucras, G. 2158, 3020
Foulquié, D. 520
Fourichon, C. 712
Foutz, J.C. 2441
Fragomeni, B. 1250, 3105
Fraile, L. 3286
Fraile, L.J. 3278
Franch, R. 2375
François, Y. 688, 2399
Fransen, T. 2790
Fraslin, C. 2343, 2363
Frąszczak, M. 3319
Frayssignes, J. 2924
Freguglia, M. 2367, 2375
Freitas, P.H.F. 528, 1575
Fresco, S. 426
Fresi, P. 2932, 3004
Friedrich, J. 810
Friedrichs, M. 3188
Friggens, N.C. 268, 272, 276, 328, 659
Fritz, S. 712, 1213, 1614, 1837, 2158
Frizzarin, M. 549
Frost, E.A. 2504, 2512
Frutos, P. 2909, 2956
Fuerst-Waltl, B. 418
Fukuzawa, Y. 3176
Fuller, A.L. 2441
Fulton, J. 1143

G

Gaddis, K.L. Parker 348
Gagné, D. 2269
Galimberti, G. 3217
Gallardo, R. 2433
Gallardo, R.A. 1733, 1917
Gallo, M. 972, 3217
Galluzzo, F. 450, 500, 2895
Gama, L.T. 1025, 2668, 2672
Gamsiz, K. 2383
Gao, H. 248, 1534
Gao, Y. 2321
Garcia, A. 1225, 1412, 1539, 1795, 2708
Garcia, A.O. 1697
García-Ballesteros, S. 732
García-Casco, J.M. 3274
Garcia, J.F. 741, 910
García-Rodríguez, A. 119, 135
García-Ruiz, A. 2737, 2753, 2871
Garner, J.B. 127, 2228
Garreau, H. 672, 708
Garrick, D. 1519, 2716, 2790
Garza-Brenner, E. 2605
Garzón, N.A.M. 1118
Gaspa, G. 2928
Gaundare, Y. 1893
Gautason, E. 2769
Gautier, M. 1012
Gava, D. 664
Gayraud, J. 2924
Gebhardt-Henrich, S.G. 635
Gebregiwergis, G.T. 573
Gebreyesus, G. 2024
Gebreyohanes, G. 1400, 1909
Gehrke, L.J. 2117
Geibel, J. 1053, 1467, 2325
Gemert, C.J.A. van 3109
Genestout, L. 2568
Gengler, N. 156, 312, 814, 865, 1318, 2608, 3237
Georges, M. 857, 1622, 2065, 2249, 2620
Gerdts, V. 3212
Geretschläger, A. 3117
Gerez, T. 2560
Ghaderi-Zefreh, M. 1237
Ghareeb, A.F.A. 2441
Gheyas, A.A. 1757
Giambra, I.J. 818
Gianola, D. 1676
GIANT Consortium 61
Gibson, J.P. 1602, 1905, 1909
Gilbert, H. 187, 264, 344, 708, 1274, 1680, 1970, 1987, 2101
Gillard, G.B. 2265
Gille, A. 2924
Gillet, E. 708
Gil, M.G. 3315
Gilthorpe, J. D. 1086
Gimeno, D. 195
Ginja, C. 1017
Girard, C.G. 1221
Girardie, O. 614
Girma, M. 1721
Githaka, N. 410
Giuffra, E. 2097, 2281
Gjerde, B. 284, 569
Gjuvsland, A.B. 830, 1189, 1647, 2337
Gkelia, D. 2920
Glantz, M. 2191
Glavaš, V. 3101
Gløersen, M.O. 183
Gloria, L.S. 422, 627
Glorieux, G. 865
Goddard, M. 2216, 2228, 2289
Goddard, M.E. 737, 773, 992, 1106, 1122, 2834
Gòdia, M. 72, 2313
Goiri, I. 119, 135
Gökçek, E.Ö. 2383
Golden, B.L. 82, 1966, 2007
Gol, S. 3258
Gombia, Y. 3335
Gomès, M. Bérodier 1995
Gómez, G. 3274
Gómez-Guzmán, J.A. 2305
Gómez, M.M. 3335
Gomez Proto, G. 369
Gomez-Raya, L. 3315, 3360

Gondo, A. 2601
Gondro, C. 219, 1156, 1193, 1420, 1598, 2105
Gonen, S. 2220
Gonzales, J. 1114
González, C. 1094, 2012
Gonzalez, M.E.M. 627
Gonzalez, M.E. Montes 422
González, O. 610
Gonzalez-Recio, O. 1676
González-Recio, O. 135, 1181, 2012, 2901
González-Recio, Ó. 2257
González-Rodríguez, O. 320, 2174
Gonzalez-Verdejo, C. 2794
Gori, A.S. 1622
Gorjanc, G. 845, 1135, 1266, 1408, 1579, 1725, 1934, 1942, 2089, 2552, 2580, 3121
Gormley, I.C. 316, 549
Gorssen, W. 478
Gotoh, Y. 2867
Goto, Y. 1306
Götz, K.-U. 1302, 1416
Goulas, E. 3151
Gourichon, D. 442
Gousseau, V. 1841
Gowane, G.R. 2729
GplusE consortium 2065
Grammes, F. 2359
Granado-Tajada, I. 119, 753
Grande, S. 373
Granleese, T. 90
Graverand, Q. Le 107
Gray, C. 2790
Gray, K. 2081
Gredler-Grandl, B. 280, 340, 1181
Greef, K.H. de 2040
Greer, G.J. 211
Grelet, C. 312
Griffiths, B.E. 2301, 2782
Grindflek, E. 115, 356, 365, 1189, 2337, 3225
Groenen, M.A.M. 2241, 2249, 2313, 2329, 2337, 2403
Grohs, C. 2158, 2162, 2178
Gromboni, C.F. 2293
Grønvold, L. 2265
Grose, J. 2656
Grøva, L. 906
Groβe-Brinkhaus, C. 1090, 3188, 3302
Grueneberg, A. 1270, 1490, 1634
Gualdron Duarte, J.L. 1622
Gualdron, J.L. 2065
Guan, D. 2456
Guatteo, R. 712, 2162, 2178
Gudkov, A.V. 3113
Guéméné, D. 2052
Guichard, M. 2568
Guigue, A. 2524
Guillaume, F. 462, 2568
Guillenea, A. 749, 1428
Guilliet, J. 2528
Guinan, F.L. 1774
Guintard, A. 1885, 2158
Guitian, J. 2421
Guitton, E. 672
Guizard, S. 2281
Guldbrandtsen, B. 1868, 2769
Gulzari, B. 2407, 2412
Gunadi, B. 2379
Gunia, M. 672, 708
Guo, J. 802, 886
Guo, X. 3200, 3229, 3233
Gupta, I.D. 2729
Gurman, P.M. 1110, 1348
Gustmann, K. 1086
Gutiérrez, G. 1913
Gutiérrez-Gil, B. 2909, 2956, 3020, 3054
Gutiérrez, J.P. 1979
Gutiérrez-Rivas, M. 135, 2257
Guyomar, C. 2281
Guy, S.Z.Y. 3000, 3062

H

Haach, V. 664
Haas, V. 2425
Haas, Y. de 508, 659, 781
Habashy, W.S. 2441
Hadjigeorgiou, I. 2920
Hadlich, F. 2114, 2277
Haegeman, A. 668
Haese, E. 3294
Haffray, P. 688, 2351, 2399, 2524
Häfliger, I.M. 1074
Hager, A. 2920
Hager-Theodorides, A.L. 3130
Haile, A. 1848
Hailemariam, D. 1181
Haile-Mariam, M 2843
Haile-Mariam, M. 737, 1770
Halle, I. 2487
Halli, K. 454
Hallsson, J. 886
Hambrook, D. 1741
Hamelin, C. 2162
Hammond, J.A. 1737
Han, J. 557, 802
Hannemann, E. 2117
Hanotte, O. 1721, 1753, 1757
Hanrahan, J.P. 2810
Hansen, L.B. 332
Hansen, L.S. 2541
Hanson, D. 700
Hansson, I. 486, 490
Han, Y. 2879
Harding, J. 3212, 3282
Harding, J.C.S. 696
Harland, C. 2790
Harlizius, B. 1876, 2313, 2329, 2337, 3225
Harpur, B.A. 2520

Harris, B.L. 1360
Harrison, P.J. 2416
Harrison, P.W. 2281
Harrison, S.J. 923
Hartono, E. 952, 1126
Hassanpour, A. 1860
Hassen, A. 1400
Hassenfratz, C. 3245
Hassoun, P. 288
Hatlen, B. 284, 569
Hauser, A. 2117
Hawken, R. 1684, 2445
Hay, E. 826
Hay, E.H. 952, 1126
Hayes, B. 1479, 2309
Hayes, B.J. 765, 769, 1106, 1177, 1282, 2130, 2597, 2684
Hazard, D. 520, 861
Heams, T. 102
Hebart, M.L. 2656, 2660
Hébrard, W. 532
Hegarty, R. 123
Heia, K. 569
Heidaritabar, M. 1102
Heimes, A. 2786
Heins, B.J. 332, 777
Heinsbroek, L.T.N. 2028
Helies, V. 672, 708
Helloin, E. 672
Hely, F. 111
Hely, F.S. 2564
Hemdène, M. 1848
Hemmink, J.D. 1737
Henne, H. 964, 1090
Hennig, S.L. 2056
Henry, H. 168
Henryon, M. 3200, 3208, 3229, 3233
Henshall, J. 393, 1812
Henshall, J.M. 781
Heo, J.M. 2460
Heo, S. 594
Herbert, C. 672, 708, 1672, 3327
Heringstad, B. 176
Hermesch, S. 66, 187, 1749, 1901, 1938, 2504, 3208, 3221
Hermida, M. 786
Hernandez, A. 2636
Hernández, P. 2073, 3180
Herring, W. 1135, 1139, 1490, 1579, 1634
Herring, W.O. 1270, 1408
Herron, J. 143, 2952
Hervás, G. 2909, 2956
Hess, A. 1070, 3360
Hess, A.S. 168
Hesse, D. 2495
Hess, M.K. 168
He, Y. 2081
Heywood, A. 2576
He, Z.X. 1160
Hickey, J. 1579
Hickey, J.M. 1135, 1139, 2208
Hickey, S.M. 168, 211
Hickson, R.E. 2716, 2822, 2851
Hidalgo, J. 1225, 2445
Hiemstra, S.J. 882, 1021, 1057
Hijas, M. Martín de 3180
Hille, M.M. 434
Hiltpold, M. 2273
Hindle, M.M. 1454
Hine, B. 1930
Hine, B.C. 1396
Hinrichs, D. 1872
Hinsu, A.T. 2421
Hirooka, H. 2688, 2696
Hirschfeld, M. 2253
Hodge, M.J. 3046
Hoedemaker, M. 2786
Hofer, A. 1630
Hofman, B. 3319
Hofmanova, B. 894
Hogeveen, H. 446
Hollander, C.A. de 1812
Holle, D.M. 1643
Hollifield, M.K. 1380
Holroyd, S.E. 2725
Holtenius, K. 2839
Honaker, C.F. 923
Honerlagen, H. 199
Honkatukia, M. 1017
Hoppe, A. 2556, 2572
Horvath, S. 2285
Horvat, S. 822, 886
Hosseini, S. 2253
Houaga, I. 1725
Houlahan, K. 227, 458, 2814
Houston, R. 2375
Houston, R.D. 2343, 2363, 2383
Howard, J. 2081
Hozé, C. 1614, 1837, 2158, 2170, 2333, 3070
Huang, C.H. 2437
Huang, W. 2105
Huang, Y. 528, 1575, 2081
Huau, C. 276, 3074
Hubin, X. 156, 865, 2178
Hu, G. 837, 1652
Hu, H. 312
Huisman, A. 1102
Huisman, A.E. 2174, 2329, 3262
Hulsegge, I. 581, 606, 3307
Hulst, A.D. 692, 704
Huson, H.J. 777, 1029, 1643, 3113
Hutchison, J.L. 2773
Hu, Z-L. 1446

I

Ibañez-Escriche, N. 2073

Ibáñez-Escriche, N. 2077, 2089, 3180, 3258
Ibeagha-Awemu, E.M. 2269
Ibelli, A.M.G. 664, 2448
Id Bella, M. 861
Id-Lahoucine, S. 2798
Iheshiulor, O.O.M. 3262
Ilatsia, E.D. 215, 336, 2138, 2917
Iliopoulou, E. 3130
Illa, S.K. 902
Imron, I. 2379
Ingham, A. 1930
Ingham, A.B. 1396
Inoue, K. 2688, 2696
Inoue, Y. 2688, 2696
Isberg, S.R. 1749
Ishii, K. 3176
Iso-Touru, T. 1181, 2830
Ithurbide, M. 276
Ivanković, A. 919
Ivask, M. 2024
Iversen, M.W. 3290
Ivkić, Z. 919
Izquierdo, M. 688

J

Jaafar, M.A. 777, 1029
Jacques, A. 1045
Jacquin, L. 1680
Jafarikia, M. 2992
Jagusiak, W. 1590
Jahnel, R.E. 2733
Jakimowicz, M. 1590
Jakobsen, J. 107, 183
Jaman, F. 2508
James, P. 2309
Janes, M. 3121
Jang, S. 1139
Jansen, S. 2487
Janssen, P.W.M. 2725
Janssens, S. 478, 3167
Janss, L. 1286, 1471
Januarie, D.A. 932
Jarquin, D. 3331
Jayakumar, S. 2147
Jayawardana, J.M.D.R. 2851
Jéhannet, P. 2028, 2036
Jenkins, G.M. 1782
Jenko, J. 1424, 1934
Jensen, J. 361, 388, 2437
Jeon, B. 1479
Jerry, D.R. 2416
Jewell, N. 2178
Jhadav, S. 1602
Jiang, J. 1482, 1692
Jiang, Y. 802, 886
Ji, C. 2471
Jiménez, M.A. 753, 2901
Jiménez-Montero, J.A. 135
Jin, D. 594
Jivanji, S. 2790
Johanns, S. 430
Johansen, K. 152
Johansen, L. Buhelt 2191
Johansson, A.M. 853, 906, 944, 2936
Johnson, D.D. 1618
Johnson, J.S. 528, 1575
Johnson, P. 107, 168
Johnsson, M. 853, 1135, 1266, 1579, 2208, 2475
Johnston, D.J. 1221, 1820
Johnston, S.E. 720, 830, 861, 940
Jones, D.B. 2416
Jonsson, N. 2309
Jordan, R. 168
Joshi, A. 1602, 1893
Joshi, C.G. 2421
Joshi, P. 2729
Joshi, S. 1893
Jouneau, L. 426
Jourdain, J. 2162, 2333
Jourdan, A. 2351
Journaux, L. 2052
Jovanovska, M. 2508
Jufar, B. 1400
Junes, P. 2545
Jung, H. 1479
Jungnickel, M. 1579
Jungnickel, M.K. 2208
Junior, L.A. Glatzl 1697
Junqueira, V.S. 1494
Jurado, N. Vargas 2996
Jurquet, J. 260
Jyotsana, B. 2194

K

Kaam, J.T. van 2818
Kachman, S.D. 2142
Kadri, N. 2170, 2212
Kadri, N.K. 1630
Kahumbu, S. 1909
Kalaldeh, M. Al 1602
Kalds, P. 802
Kale, D.S. 2203
Kalm, E. 3155, 3163
Kamidi, C.M. 2917
Kamphuis, C. 402, 446, 581, 606, 659, 1502, 3307
Kang, J.M. 1594
Kanlisi, R.A. 2433
Kantanen, J. 886, 1017
Kapsona, V. 2782
Karagianni, V. 3130
Karahan, B. 2383
Karaman, E. 749, 1148, 1286, 1428, 1432, 1471
Kargo, M. 152, 268, 976, 1041, 1049, 1828, 1852, 2541
Karimi, K. 837
Karim, K. 2620
Karim, L. 857, 2249

Karlau, A. 2020
Karlsson, J. 2839
Karrow, N.A. 2237, 2855
Kar, S.K. 148, 2085
Karthickeyan, S.M.K. 2031
Kasap, A. 2968, 2988
Kasarda, R. 894, 3058, 3086, 3093
Kaseja, K. 3066
Kashefifard, K. 1672, 3327
Kašná, E. 414, 504
Kasper, C. 352, 3270
Kate, R.R. 2203
Kause, A. 2343, 2363
Kawaguchi, F. 886
Kawahara, T. 1306, 2867
Kawakami, J. 2867
Kayang, B.B. 1917, 2433
Kazymbet, P. 886
Keane, O.M. 2810
Kebede, F.G. 1753
Keele, J.W. 1404
Kelleher, M.M. 2891
Kelly, D.N. 2863
Kelly, M.J. 470
Kelly, S. 219
Kelly, T. 1733, 1917
Kelly, T.R. 2433
Kempe, R.M. 304
Kemp, R.A 1999
Kemp, R.A. 248, 398, 1262
Kemp, S. 1753, 1909
Kenny, D. 1547
Kenny, D.A. 292
Kent, M.P. 2265
Keogh, K. 292
Kernéis, T. 2391
Kerr, B.J. 3241
Keshavarzi, H. 482
Kettunen, A. 1991
Khanal, P. 252
Khansefid, M. 643, 737, 773, 1450, 2843
Khayatzadeh, N. 822, 886
Kho, E. 2597
Kidane, A. 573
Kimata, M. 3184
Kim, H.J. 2632
Kim, M. 594, 1660, 2429, 2460
Kim, N. 886
Kim, Y.K. 1594
Kinghorn, A.J. 1856
Kinghorn, B.P. 1856
Kinghorn, M.G. 2749
Kirk, T. 131
Kirwan, S.F. 207
Kistler, T. 2560, 2584
Kızılkaya, K. 1639
Kjøglum, S. 2371
Klandermans, G.G.J. 3307
Klawatsch, J. 2224
Klerk, B. de 393
Klingström, T. 98, 668, 853, 1808
Kloet, S.L. 2241
Klosa, J. 1606
Knap, P. 80
Knight, M.I. 127
Knol, E. 1290, 2016, 2337
Knol, E.F. 380, 700, 1189, 1688, 1876, 3253
Knowler, K.J. 211
Knutsen, T. 2359
Köck, A. 466
Koga, S. 2220
Koh, Y.J. 594
Koivula, M. 304, 1298, 1314
Kok, J. 1782
Kokkonen, T. 236
Koltes, D. 2199
Koltes, J. 2199
Koltes, J.E. 2806
Komen, H. 102, 590, 1753, 2379, 2403, 2407, 2412
Komen, J. 86, 2395
Kominakis, A. 2920, 3130
Komwihangilo, D. 1709, 1909
König, S. 454, 818, 1008, 1131, 1567
Koning, D.-J. de 2191, 2325, 2475
Koning, D.J. de 1808
Koning, L. 179
Konjačić, M. 919
Kootstra, G. 537
Kooverjee, B.B. 948
Kooverjee, B. Bhika 2187
Koringa, P.G. 2421
Korkuć, P. 1008, 1131, 1567
Korsvoll, S.A. 2371
Kosinska-Selbi, B. 2126
Kosińska-Selbi, B. 1590
Koskinen, H. 2343, 2363
Kotlarz, K. 2126
Kour, A. 2147
Koutsouli, P. 2224
Kovačić, M. 2508
Kramer, K. 2048
Kramer, L. 1824
Kramer, L.M. 2321, 2806
Kranis, A. 1454, 1583
Krattenmacher, N. 3155, 3163
Krebs, S. 2117
Kreuzer-Redmer, S. 430, 2495
Kring, R.D. 1828
Kristensen, T.N. 2541
Kritikos, N. 3130
Kröhl, Z.M. 3188
Kronqvist, C. 2839
Kruijt, L. 2085
Krupová, Z. 414, 504
Kudinov, A. 1534
Kudinov, A.A. 1314

Kuehn, C. 1181
Kuehn, L.A. 826, 1404, 1966, 2593, 2624
Kuhla, B. 199
Kühn, C. 2170, 2261, 2277, 2297, 2786, 3155, 3163
Kumar, S. 1897
Kunz, E. 2154
Kuo, R. 2130, 2281
Kupisiewicz, K. 2024
Kurukulasuriya, M.S. 1881, 1921
Kurylo, C. 2281
Kurz, A. 3294
Kyläniemi, M. 2830

L

Labadie, K. 2549
Labarthe, E. 2568
Labbé, L. 2391
Lacroix-Lamandé, S. 426
Lagarde, H. 2399
Laghouaouta, H. 3286
Lagler, D.K. 2154
Lagriffoul, G. 288, 2960
Lague, M. 344
Lahmar, M. 1848
Lahoz, B. 3050
Laing, A.R. 2684
Lakamp, A.D. 434
Lakhal, W. 2281
Lakhssassi, K. 3050
Lallias, D. 2391, 2399
Lambe, N. 107
Lambe, N.R. 2940
Lamb, H.J. 1177
Lamers, K. 2855
Lamont, S. 2433
Lamont, S.J. 1733, 1917
Lamy, J. 2016
Lantier, F. 672
Lan, Y.H. 2745
Laplana, M. 3278, 3286
Lara-Rivera, A.L. 2305
Larkin, D. 1078
Larroque, H. 288, 2924, 2932, 3004
Larsberg, F. 2495
Larsen, T. 276
Larzul, C. 3245, 3339
Lashmar, S.F. 870, 878
Lassen, J. 340, 577, 2875
Laterrière, M. 2269
Laureau, S. 2375
Laurenco, D. Lino 1392
Laurent, F. 426
Laursen, S.F. 2541
Lavrenčič, E. 1725
Lawlor, T.J. 898, 1000
Lazzari, B. 741, 906
Leal-Gutiérrez, J.D. 1618
Leandro, F.D. 1697
Lebepe, J. 2464
Leclerc, H. 462, 1148, 2158, 2178
Lecoeur, A. 442
Leder, E. 2220
Ledur, M.C. 664, 2448
Lee, C. 482
Lee, D.H. 1594
Lee, D.J. 1594
Lee, J.H. 594, 1660, 2429, 2460
Lee, S. 1352, 1594, 2121
Lee, S.H. 1193
Lee, S.S. 3172
Lee, Y.L. 2249, 2329
Lefebvre, R. 260, 426
Lefebvre, T. 2549
Legarra, A. 1000, 1217, 1335, 1376, 1530, 1845, 2093, 2757, 2761, 2960
Le Graverand, Q. 618
Lehnert, S. 1326
Lei, C. 886
Leighton, E. 3105
Leino, A.-M. 304, 1339
Leitão, A. 1025
Leite, N.G. 3196
Leite, N.L. 380
Lemal, P. 2608
Lema, M.O. 164
Lembo, G. 2395
Lenoir, G. 272, 1672, 3327
Lenstra, J.A. 886, 1017
Leonard, A. 1062, 1066
Leroy, G. 1017, 1057
Leroy, P. 2351
Le Roy, P. 512
Letayef, N. 1848
Letuka, M. 438
Lewerentz, F. 2191
Lewis, R.M. 676, 923, 1404, 2913, 2964, 2996
Lewis, T.W. 3121
Liao, Y. 886
Li, B. 2301, 2782
Li, C. 802, 1181, 1656
Lidauer, M. 1372, 1400, 1498, 1534
Lidauer, M.H. 236, 304, 1181
Lien, S. 2265
Ligda, C. 1057
Ligios, C. 1082
Li, J. 1143, 1668
Li, L. 1110, 1348
Lillehammer, M. 2416
Lima, A.L.R. 2601
Lima, A.O. 2293
Lima, J. 2616
Lim, K.-S. 1917
Lim, K.S. 2321, 2806
Limon, G. 2421
Lindberg, S.K. 569
Lindburg, J.J. 77

Lindgren, F. 1408
Lin, E.C. 2745
Ling, A. 952, 1126
Lin, K.H. 2745
Link, W. 2487
Linneflaatten, L.-B. 183
Li, Q.H. 1160
Li, S. 203, 1160
Little, E. 700
Littlejohn, M. 2790
Littlejohn, M.D. 2216, 2289
Liu, G.E. 2321
Liu, H. 1049, 1152, 1852
Liu, L. 631
Liu, R.R. 1160
Liu, T. 3204, 3351
Liu, Z. 1164, 1343, 1364, 3163
Li, W. 1160
Li, Y. 1326, 1396
Li, Z. 757
Lizana, C. 1983
Lloret-Villas, A. 2273
Lôbo, R.B. 2676
Lobo, R.N.B. 1926
Loftus, J.P. 3113
Loh, G. 2495
Loh, Z. 1201
Loke, B. 659
Lombardi, C.A. 3311
Lont, D. 86, 102
Lopdell, T. 2065
Lopdell, T.J. 2216, 2289
Lopes, L.S.F. 227
Lopes, M.S. 761, 1189, 1876, 2134, 2337, 3225, 3253
López-Carbonell, D. 786, 1610
López-Catalina, A. 2257
López-Correa, R. 2676
López-Correa, R.D. 1335
Lopez-Cruz, M. 252, 1490
López-García, A. 135, 2012
Lopez, P. 2371
López-Paredes, J. 135
López-Villalobos, N. 2725, 2822, 2851
Lora, I. 659
Loughnan, S. 2416
Lourenco, D. 380, 898, 1000, 1139, 1217, 1225, 1254, 1380, 1388, 1412, 1523, 1530, 1774, 2445, 2761, 2883, 2886, 3196
Lourenco, D.A.L. 1799, 1983
Louveau, I. 3245
Lou, W. 203, 631
Løvendahl, P. 152, 493
Loy, D.D. 82
Loy, J.D. 434
Lozada-Soto, E.A. 806
Lubritz, D. 1143
Lu, D. 1412, 1539, 1795
Lühken, G. 2154
Lund, M.S. 749, 1286, 1428, 1471, 2024, 3204, 3351
Lyatuu, E. 1709, 1909
Lynch, C. 458
Lynn, A.M. 2628
Lyons, R.E. 2684

M

Maak, S. 2170
Macciotta, N. 373
Macciotta, N.P.P. 968, 1392, 2928
MacGibbon, A.K.H. 2725
Machado, C.H.C. 1697
Machado, M.A. 960, 1697, 2182
Machefert, C. 288
MacLeod, I.M. 127, 737, 773, 2228
MacNeil, M.D. 1705, 1713, 1816
Madenci, D. 2887
Madilindi, M.A. 300, 497
Madoui, MA. 2549
Madsen, M.D. 684
Madsen, O. 2241, 2281, 2313
Madsen, P. 2437
Maestrale, C. 1082
Magwaba, T. 1733
Mahla, R. 2568
Maicelo, J.L. 1114
Maignel, L. 2992
Maiwashe, A. 1322
Maiwashe, N. 2151
Maizon, D.O. 1778
Majolino, G. 3090
Makanjuola, B. 219
Makanjuola, B.O. 1193
Makgahlela, M.L. 1705, 2151
Mäki-Tanila, A.V. 1864
Ma, L. 1482
Malan, I.N. 3078
Malheiros, J.M. 2612
Mallard, B. 3282
Mallard, B.A. 2855
Maltecca, C. 806, 1482, 1692, 2081, 2101
Mamani, G.C. 3331
Manafiazar, G. 1652, 1656
Manciaux, L. 462
Mancin, E. 369, 1242, 3135
Manjula, P. 2429, 2460
Mannen, H. 886
Manosalva, C. 2355
Mante, J. 655
Mantovani, R. 369, 1242, 3135
Mäntysaari, E.A. 236, 304, 1298, 1314, 1339, 1506, 1515, 1534
Mäntysaari, P. 236
Manunza, A. 553, 598, 798, 906
Manzanilla-Pech, C.I.V. 340, 1181, 2875
Mapel, X.M. 2273
Mapholi, N.O. 438, 497, 878
Marcon, D. 244, 520, 618
Marett, L.C. 127, 2228

Margarita, M.L. 3125
Margetin, M. 894
Margetín, M. 3058
Marie-Etancelin, C. 107, 244, 618, 1551, 2093
Marina, H. 753, 2909, 2956, 3020, 3054
Marín, B. 3050
Marinč, A. 2552
Marion, G. 724, 728, 1475
Marjanovic, J. 781
Markey, A. 3237
Marle-Koster, E. Van 2151
Marle-Köster, E. van 878, 1586
Marois, D. 2437
Marques, C.B. 160, 164
Marshall, K. 810
Martelli, P.L. 3217
Martikainen, K. 980
Martin, A.A.A. 2798
Martínez, A. 886
Martínez-Álvaro, M. 324, 2616
Martinez, A. Martinez 1037
Martínez-Blanch, J. 1094
Martinez Boggio, G. 1551
Martínez-González, J.C. 2305
Martinez, N. Garcia 1791
Martínez, P. 786
Martin, F. 2281
Martin, P. 260, 340, 426, 2648
Martinsen, K. 115, 356, 365
Martin, S.J. 90
Martins, M.F. 1697
Marusi, M. 191, 450, 500, 2818, 2895, 3311
Masangkay, J.S. 886
Masaoka, M. 886
Maskal, J.M. 528, 1575
Mason, B.A. 2228
Masoodi, S. 1450
Masoudi, A.A 2467
Massault, C. 2416
Massender, E. 2992
Mastranestasis, I. 2920
Mastrangelo, S. 794, 968
Masuda, Y. 898, 1000, 1306, 1343, 1364, 1530, 2757
Mateescu, R.G. 1618, 2636
Matera, R. 598
Mather, M. 757
Mathur, P.K. 700
Matika, O. 410, 438, 497, 1237, 1583
Matos, G. 3274
Mattalia, S. 231, 406, 651, 1258, 2158, 2178
Matthews, D. 565, 2810
Mattock, J. 324, 2616
Maugan, L.-H. 1246, 1258
Maupetit, D. 2648
Maupin, M. 672
Ma, W. 3192
Mayerhofer, M. 466
Mayes, M.S. 2806
May, K. 1008, 1131, 1567
Mazza, R. 886
McCarthy, J. 2720, 2847
McCormack, H.A. 2475, 2479
McCrabb, E.J. 3062
McCulloch, R. 1326
McDermott, K. 107, 1004, 2948, 2952, 3016
McDonald, J. 1326
McEwan, J. 886, 2810
McEwan, J.C. 123, 168, 183, 211
McEwin, R.A. 2656, 2660
McGee, M. 292
McGilchrist, P. 2628, 2700, 3000
McGovern, F. 107
McGovern, F.M. 2905, 2944, 2948
McGowan, M.R. 2684
McHugh, N. 107, 1004, 2644, 2905, 2944, 2948, 2952, 3016
McIntosh, K. 1583
McKay, S. 2109
McMillan, A.J. 2964
McNabb, B.R. 2056
McNaughton, L. 2790
McNaughton, L.R. 2851
McNeilly, T.N. 720
McParland, S. 316, 549
McRae, K. 1070
McRae, K.M. 211, 2285
McWhorter, T.M. 2883
McWilliam, S. 1326, 2640
Mdyogolo, S. 1705
Medina, M. 2371
Medrano, J.F. 2109, 2237
Medugorac, I. 822
Međugorac, I. 2117, 2154, 2224
Mee, J. 2810
Megens, H.J. 2403
Mehrotra, A. 1630
Mehtiö, T. 236, 304
Meijboom, F.L.B. 2048
Mei, Q. 1152
Mejía-Melchor, E.I. 2737
Melbo Tajet, H. 1934
Mellanby, R.J. 1579
Mencarelli, A. 2412
Mendel, C. 2154
Menezes, G.R.O. 2601
Menghak, P. 1889
Mercadante, M.E.Z. 474, 890
Mercat, M.-J. 2174, 3245
Mercat, M.J. 1045
Meseret, S. 1400, 1909
Mészáros, G. 94, 102, 3058, 3117
Mészárosová, M. 3058, 3117
Meutchieye, F. 1717
Meuwissen, T.H.E 573
Meuwissen, T.H.E. 284, 356, 928, 1122, 1442
Meyerholz, M.M. 2786
Meyer, K. 1510

Meyermans, R. 478
Meynadier, A. 618, 1551, 2093
Miari, S. 1082
Miar, Y. 837, 1652, 1656
Michenet, A. 655
Miedzinska, K. 2281
Mielczarek, M. 3319
Migliorati, C. Franzoni 3311
Miglior, F. 227, 458, 2237, 2798, 2814
Mikko, S. 3159
Mikko, S.C.M. 3147
Mikulec, N. 984
Milanesi, M. 741, 886
Miles, A.M. 2773
Milfort, M.C. 2441
Milgen, J. van 1290
Milkeviych, V. 1286, 1471
Milkevych, V. 1049, 1852
Miller, J.E. 2913
Miller, L.C. 3192
Miller, R. 2387
Miller, S. 1412, 1539, 1571
Miller, S.P. 384, 1795, 2708
Minéry, S. 256
Minozzi, G. 553, 3090
Misztal, I. 898, 1000, 1139, 1217, 1225, 1380, 1388, 1412, 1523, 1530, 1774, 1799, 2445, 2757, 2761, 2883
Mitchell, A.P. 2887
Miyumo, S.A. 336
Miyunga, A. 1737
Mkhize, N.E. 497
Moate, P.J. 127
Moen, T. 2359, 2371
Mogaka, D. 1909
Moghaddar, N. 1110, 1348, 2984
Molano, E.S. 2782
Molina, A. 2020
Molinero, E. 1098
Mollandin, F. 1274
Mollel, E. 1733
Molotsi, A.H. 3078
Momin, Md.M. 1352
Monau, P.I. 956
Mondet, F. 2568
Monehan, T. 2790
Mongellaz, M. 2097, 2174, 2281
Montalcini, C.M. 635
Montaldo, H.H. 1983
Monteagudo, L. 1037
Moody, J. 2216
Moore, K.L. 1221
Morales-González, E. 841
Morales, R.A. Alonso 2871
Mora, M. 320, 1970, 2073
Moran, D. 1741
Moravcikova, N. 894
Moravčíková, N. 3058, 3086, 3093
Moreira, G.C.M. 857, 2199, 2261, 2297
Morek-Kopeć, M. 1590
Moreno-García, M. 1094
Moreno, J.A.H. 1701
Moreno-Romieux, C. 3070
Morés, M A.Z. 664
Morés, M.A.Z. 2448
Moret, Y. 2549
Morgan, J.L. 2964
Morgan, J.L.M. 676, 2913
Morin, T. 688
Mörlein, D. 2487
Morosini, N.S. 2233
Morota, G. 545
Morris, K.M. 1721
Morrison, L. 1737
Morrison, S.J. 2905
Morris, S.T. 2716
Mortier, J. 2162
Mortimer, S.I. 3000
Morvezen, R. 688, 2351
Moser, C. 3105
Moser, G. 470
Mosnier, F. 296
Mostert, B.E. 3032
Mostin, L. 668
Mota, R.R. 1318, 1356
Mota, V. 569
Mourão, G.B. 2199, 2233, 2293, 2612
Mrode, R. 1229, 1400, 1709, 1729, 1741, 1909, 2777
Mrode, R.A. 80
Msoffe, P.L. 1733, 1917
Msuta, G. 1709, 1909
Muchadeyi, F.C. 878
Mucha, S. 3066
Mueller, M.L. 2044, 2056
Muhairwa, A.P. 1733, 1917
Muir, S.K. 127
Mukaratirwa, S. 2464
Mukherjee, A. 902
Mukherjee, S. 902
Mukiibi, R. 2375
Mulas, G. 716
Mulder, H.A. 203, 446, 631, 680, 2329, 2403
Mulhall, S.A. 1787
Mullaart, E. 446, 2016
Mumtaz, S. 902
Munilla, S. 1205, 1778
Muñoz, I. 610
Muñoz, M. 3274
Muñoz-Tamayo, R. 272
Murani, E. 2114
Murdoch, B.M. 2913
Murphy, C. 2644, 2847
Murphy, C.P. 1547
Musa, A.A. 936
Mushi, J.R 1733
Mushi, J.S. 1917
Mutch, E. 2980

Mwai, A.O. 1729
Mwai, O.A. 1400
Mwendia, C. 410
Mydland, L.T. 356

N

Naazie, A. 1917, 2433
Naderi, S. 1310, 1506, 2680
Nakagawa, S. 2867
Nam, S. 2460
Napel, J. ten 1372, 1498, 1502, 1506
Nath, S. 902
Navajas, E.A. 107, 160, 164, 195, 240
Nawaz, M.Y. 1156
Nazari-Ghadikolaei, A. 2712
Negera, T. 1400
Neglia, G. 598
Negrini, R. 996
Negro, A. 373
Negro, S. 3180
Negussie, E. 236, 304, 1400
Nel, C.L. 3081
Nesengani, L.T. 438
Neser, F.W.C 2187
Neser, F.W.C. 932, 1705, 1713, 1816, 2069, 2749, 3323
Neshat, M. 1352
Neto, A.Z. 2233
Neudischko, M. 2568
Neuenschwander, S. 1630
Neujahr, A.C. 434
Neumann, G.B. 1008, 1131, 1567
Neupane, M. 1029
Neuteboom, M. 882, 1021
Neves, H.H.R. 1118, 1254, 1598
Newman, S. 1966
Newton, J.E. 72, 1770
Ngeno, D.K. 2917
Ngeno, K. 2138
Nguyen, H. 643
Nguyen, L. 2309
Nguyen, L.T. 1177, 2130, 2597
Nguyen, T. 2765
Nguyen, T.T.T. 111
Ni, A.X. 2491
Niazi, A. 668
Nicholas, F.W. 757
Nicolazzi, E.L. 2761
Nielsen, B. 3204, 3351
Nielsen, H.M. 1041, 1828, 1852, 1991, 2541
Nielsen, U.S. 304
Nie, Q. 2471
Nieuwhof, G. 2765
Nieuwhof, G.J. 111, 1762, 1766
Nijman, I.J. 886
Nilforooshan, M.A. 1360
Nimbkar, C. 1897
Niranjan, S.K. 2147
Nishio, M. 2688, 2696, 3176
Nishiura, A. 2859
Ni, Z. 1688
Njeru, R. 1737
Noble, C. 569
Nogueira, A.R.A. 2293
Nogueira, G.dP. 741
Noguera, J.L. 3180
Nolte, T. 2487
Nolte, W. 3155, 3163
Nomura, Y. 886
Norberg, E. 176
Nordbø, Ø. 365, 1424
Norman, H.D. 1774
Norris, A. 284, 569, 2220, 2355
Norton, T. 557
Nosková, A. 1630
Nothman, J. 757
Notter, D.R. 2913, 2996
Nousianen, A. 2343, 2363
Novosel, D. 2826
Nsoso, S.J. 956
Nugroho, J.E. 2537
Null, D.J. 1463
Nuñez-Dominguez, R. 2668, 2672
Nuñez-Domínguez, R. 2305
Núñez, P. 3258
Nuphaus, S. 380
Nussey, D.H. 720
Nxumalo, K.S. 2151

O

Oakey, H. 2660
Obsteter, J. 1934
Obšteter, J. 1942, 2552, 2580
O'Connor, E. 737, 2905, 2944, 2948
Ødegård, J. 2359, 2371
O'Donovan, M. 316
Oelschlägel, D. 639
Ogawa, S. 3176, 3184
Oget-Ebrad, C. 857, 2065, 2249
Ohlsson, I. 1808
Ohlsson, J.I. 98
Oikonomou, G. 2301, 2782
Ojango, J.M. 1709, 1729, 1909
Ojeda-Marín, C. 1979
Ojo, O.E. 430
Okamura, T. 3176
Oka, T. 2867
Okeno, T.O. 215, 2138
Okeyo, A.M. 1709, 1909
Okpeku, M. 878
Olasege, B.S. 1209, 1282
Olesen, I. 1991
Olivares, M.E. 1983
Oliveira, E.C.M. 2668, 2672
Oliveira, H.R. 1571, 2708, 2798, 2814, 2992
Oliveira, J.C. 1697
Oliveira Jr., G.A. 227, 2814

Oliveira, M.H.V. 308
Oliveira, T.P. 1725, 1942
Olivier, W.J. 2976, 3081
Oloo, R.D. 1709, 1729
Olsen, D. 115, 365
Olsen, H.F. 139
Olson, K.M. 80
Oltenacu, P.A. 2636
Omar, F. Md. 886
Ong, L.T. 1766
Onzima, R. 3012
Ooi, E. 2834
Oosting, S.J. 203, 631
Opoola, O. 1741
Opriessnig, T. 3212
Orehovački, V. 984
Ortiz, E.G. 834
Osawa, T. 1306
Østbye, T.K.K. 2220
Østergård, S. 1828
Oster, M. 199
Ostersen, T. 3200, 3229, 3233
Ostrovski, H. 2105
Otsyina, H.R. 1917, 2433
Ouchene-Khelifi, N.-A. 886
Ouweltjes, W. 581, 606
Owen, J.R. 2056
Owens, P. 143

P

Pabiou, T. 207, 565, 745, 2664, 2680, 2952, 3016
Pacheco, H.A. 1185
Pacífico, C. 430
Pagotto, U. 3217
Paim, T. 826
Palhière, I. 276, 296, 1833, 1841, 3037, 3042
Palma-Vera, S.E. 964
Palpant, N. 2293
Palstra, A. 2395
Palstra, A.P. 2028, 2036, 2403
Pandit, R.J. 2421
Paneru, U. 2984
Panetto, J.C.C. 960, 1697, 2182
Pannebakker, B.A. 2537
Pannier, L. 3000
Pant, S.D. 3046
Pan, Z. 1668, 2456
Papachristou, D. 2224
Pappas, D. 3130
Paredes-Sánchez, F.A. 2605
Parisot, S. 288, 520
Park, C.A. 1446
Park, J. 594
Parmentier, H.K. 3355
Parra-Bracamonte, G.M. 2305, 2605
Pascoli, F. 2375
Pascual, M. 610
Patil, D.V. 2203
Patrice, P. 2399
Pauciullo, A. 2928
Paul, N.-F. 1086
Paulsson, M. 2191
Pausch, H. 802, 1062, 1066, 1630, 2170, 2212, 2273
Pavez, M.L. 2355
Pavlík, I. 3058
Pavlov, B. 2508
Paxton, E. 1737
Péchoux, C. 2158
Pećina, M. 919
Pedota, G. 598
Pedrosa, V.B. 422, 627
Peers-Adams, J. 168
Peeters, K. 781, 2545
Peiró-Pastor, R. 2257, 2383
Peixoto, J.O. 664, 2448
Pelayo, R. 2909, 2956, 3020, 3054
Pemberton, J.M. 57, 720, 940
Peñagaricano, F. 252, 348, 1185
Peñaloza, C. 2375, 2383
Pena, R.N. 1098, 3278, 3286
Penasa, M. 418, 2895
Pepke, F. 2097
Peraza, P. 164
Pereira, F. 886
Pereira, G.M. 332
Pereira, M.A. 1697
Perez, B.C. 1233, 1278
Pérez-Enciso, M. 622, 1173
Pérez-Rodríguez, P. 914, 1486
Periasamy, K. 1745
Perrault, L. 1791
Persichilli, C. 2818
Perucho, O. 610
Petelle, M.B. 635
Petersen, G.E.L. 2564, 2576
Peters, L. 940
Peters, S.O. 1639, 2464
Pethick, D. 3000
Petrini, J. 2061, 2612
Petzl, W. 2786
Peynot, N. 1037
Pezier, T. 426
Philippe, M. 462, 2333
Philippe, R. 2170
Phocas, F. 2351, 2391, 2399, 2560, 2568, 2584
Pietarinen, J. 1864
Pietralunga, V. 426
PigGen Canada 3282
Pigozzi, G. 3135
Pile, G. 168
Piles, M. 320, 610, 1970
Pilkington, J.G. 720
Pilla, F. 968, 2818
Pimentel, E.C.G. 1302, 1416
Pinard-van der Laan, M.H. 442
Pineda-Quiroga, C. 119, 753

Pinedo, P. 2879
Piñero, D. 2871
Pinton, A. 2178, 2333
Pintus, D. 716
Pires, J. 276
Pita Barros, C. 1753
Pitchford, W.S. 2656, 2660
Pitkänen, T. 304, 1339, 1515, 1534
Pizzol, M.S. Dal 2448
Plassard, V. 2158, 2162
Plastow, G. 1102, 1181, 1262, 1437, 1656, 2261, 2297, 3212
Plastow, G.S. 248, 696, 2321, 3282
Plisson-Petit, F. 2924
Pocrnic, I. 1266, 1408, 1725, 1934, 1942, 2089
Podgorniak, T. 2265
Podtar, V. 1602
Poemape, C. 1114
Pogorevc, N. 822
Polkinghorne, R. 2628
Pollet, N. 2528
Pompanon, F. 886
Pong Wong, R. 728
Pong-Wong, R. 1237
Ponsuksili, S. 199, 2114
Pook, T. 1053, 1467, 1860
Pooley, C.M. 724, 728, 1475
Popoola, M.A. 2471
Poppe, M. 402, 446, 651, 1502
Popplewell, G.I. 2660
Poquet, Y. 2568
Portes, D. 288
Portolano, B. 968
Porto-Neto, L.R. 1209, 1282, 1326, 1396, 1563, 1930, 2640
Porto, T. 2612
Poslon, V. 2968
Pösö, J. 304
Post, M. 111
Potdar, V. 1893
Pouil, S. 2391
Poulsen, B.G. 3196
Powell, J. 1737
Prakash, V. 2194
Pravia, M.I. 164, 240
Prchal, M. 2399
Prendergast, J. 810, 1737
Prendes, R. Gonzalez 1037
Prentice, J.C. 728
Prešern, J. 2508, 2552
Prestløkken, E. 573
Priest, B. Dumont Saint 3143
Prieur, V. 655
Pröll-Cornelissen, M.J. 3302
Proto, G.G. 1242
Proto, G. Gomez 3135
Prowse-Wilkins, C.P. 773, 2216, 2228, 2289
Prunier, A. 3245
Pryce, J.E. 53, 127, 643, 737, 773, 1181, 1770, 2834, 2843
Psifidi, A. 1721, 2301, 2421, 2782
Pszczola, M. 172
Puillet, L. 268, 328
Pukhrambam, M. 2147
Punde, N. 1893
Purfield, D.C. 207, 745
Puškadija, Z. 2508
Pyoos-Daniels, G.M. 1816

Q

Quilcate, C. 1114
Quillet, E. 2391
Quintanilla, R. 1173
Quinton, C. 1946
Quinton, C.D. 1958
Qu, J. 1294, 1519

R

Rachman, M. 1757
Rae, A. 1583
Rafter, P. 1004, 2948
Rahmatalla, S.A. 3012
Ramayo-Caldas, Y. 320, 1173, 2174
Ramirez-Diaz, J. 553, 598
Ramirez, J.F. 328
Ramírez, M. 786, 1610
Ramírez-Valverde, R. 2305
Ramljak, J. 919, 2968, 2988
Ramón, M. 753, 988, 2012, 2794, 2901
Randhawa, I.A.S. 1177
Raoul, J. 1833
Raphaka, K. 956
Ratnakar, V. 1893
Rau, A. 1274
Rauw, W.M. 3315, 3360
Ravagnolo, O. 164, 240, 2741
Raymond, B. 72, 1181
Raza, A. 2309
Real-García, V.M. Del 2605
Reding, J.J. 1586, 3032
Reecy, J. 2199
Reecy, J.M. 1446
Reents, R. 1164, 3163
Regitano, L.C.A. 2061, 2199, 2233, 2293, 2612
Reich, C.M. 2228
Reimer, C. 1053, 2325
Reinoso-Peláez, E.L. 2012
Reinsch, N. 936, 964, 2733
Reisinger, N. 430
Reixach, J. 3258
Rekaya, R. 952, 1126, 2441
Rekik, M. 1848
Relun, A. 2158
Renand, G. 2648
Renckly, J. 1791
Restoux, G. 1037, 1045, 1885
Retallick, K. 1225, 1412, 1539, 1571, 2708
Retallick, K.J. 384, 1795
Reverter, A. 1326, 1396, 1563, 1930, 2640

Reyer, H. 199, 964
Rezende, F.M. 1618, 2636
Riazi, M. 1656
Ribani, A. 972, 2516, 3217
Ribas, C. 1672, 3327
Ricard, A. 3143, 3151, 3339, 3347
Ricci, S. 430
Richardson, C.M. 111
Richter, J. 2445
Riffault, S. 426
Rigaud, T. 2549
Riggio, V. 1237
Rikkers, R.S.C. 2533
Rindfleish, S.J. 3046
Ring, S.C. 2720, 2891
Riser, M. 1250
Riser, M.M. 3105
Ristov, S. 2826
Rius-Vilarrasa, E. 853
Rivera-Chacon, R. 430
Riviere, J. 2097
Rivière, J. 2158
Robert-Granié, C. 288, 2932, 3004
Robert, R. 672, 708
Roberts, A.J. 826
Robinson, N. 2220
Robinson, N.A. 2416
Robledo, D. 2375
Robson, J.F. 541
Rocha, M.I.P. 2293
Roch, L. 3270
Rochus, C.M. 2452, 2798
Rocland, M. 651
Rodehutscord, M. 3294
Rodenburg, T.B. 393, 508
Rodríguez-Almeida, F.A. 2605
Rodriguez, E. 1618
Rodríguez, J.D. 2676
Rodriguez-Navarro, A.B. 2475
Rodríguez-Ramilo, S.T. 1833, 1845, 3139
Roehe, R. 324, 524, 2616
Rogel-Gaillard, C. 696, 3282
Rohmer, T. 3347
Rohr, P. von 647
Rojo-Guerra, M. 1094
Rolf, M. 1701
Rolf, M.M. 82
Romashov, G. 1078
Rombouts, T. 2620
Romé, H. 2437
Rönnegård, L. 486, 490
Roozeboom, C. 2407, 2412
Roques, S. 148
Rosa, G.J.M. 602, 1676, 1684, 1970
Rosas, J.P. 3180
Rose, I.J. 3062
Rosen, B.D. 886, 1717, 2913
Rosendal, G.K. 1991
Ros-Freixedes, R. 1098, 1135, 1139, 1270, 1579, 3278, 3286
Rosner, F. 639
Ross, E. 1479, 2309
Ross, E.M. 765, 1177, 2130, 2597
Rossoni, A. 1185
Ross, P.J. 2056
Rostellato, R. 1246, 1258
Roth, K. 3302
Rouger, R. 2524
Rovere, G. 219, 1193
Rowan, T.N. 82, 1555
Rowe, S. 107, 123, 168, 211, 2810
Rowland, K. 1143
Roy, J.P. 2269
Roy, P. Le 2483
Rubinigg, M. 2588
Rubio-de Juan, A. 988
Ruesche, J. 708
Ruíz-de la Cruz, G. 2605
Ruíz-Flores, A. 914
Ruiz-López, F.J. 1983, 2737, 2753, 2871
Runcie, D.E. 1294
Rupp, R. 264, 276, 296, 1551, 1987, 3074
Russell, C.A. 2593
Russenberger, J. 3105
Russ, I. 2117, 2154, 2224
Ryan, C.A. 745
Ryan, C.V. 207
Rychtarova, J. 894
Rydhmer, L. 139

S

Saborío-Montero, A. 135, 2012
Sader, G. 2545
Sae-Lim, P. 2347, 2355
Sahana, G. 340, 886, 1181, 1286, 1471, 1559, 2126, 2245, 2541, 2769
Saille, S. 256
Saintilan, R. 1148
Saito, Y. 2859
Salaris, S. 716, 1082, 2972, 3028
Salavati, M. 1717, 2065, 2261, 2297, 2317, 3097
Sallam, M. 2475
Samaddar, A. 1634
Samaraweera, A.M. 1901, 2700
Samaraweera, M. 1930
Sánchez, J.P. 320, 610, 1970, 2174
Sánchez-Molano, E. 2887
Sanchez, M.P. 712, 1614, 2170
Sánchez-Ramos, R. 914
Sanchez-Rodriguez, E. 2475
Sandilands, V. 2479
Sandoval, E. 168
Sanglard, L.P. 2624
Sang, Y. 3192
Sanjayaranj, I. 2725
Sann, C. 2568
Santacreu, M.A. 2073, 2077

Santana, B.F. 1250, 3331
Santos, B.F.S. 1782, 1958
Santos, D. 1025
Santos, J.E.P. 252
Sá, P. 1025
Sargolzaei, M. 2883
Sarkar, M. 2147
Sarker, N.R. 1938
Sarpong, N. 3294
Sarto, M.P. 3050
Sartori, C. 369, 1242, 3135
Sarviaho, K. 980
Sasaki, O. 2859
Sasazaki, S. 886
Satoh, M. 3176, 3184
Sattler, C.G. 2883
Saunier, D. 256
Saura, M. 728, 2012, 2383
Savage, D.B. 1820
Savchuk, A. 1233
Savegnago, R.P. 2105
Savietto, D. 708, 1987
Savoia, S. 2704
Sawhney, T. 1897
Scala, A. 3028
Schaefer, R. 664
Schafberg, R. 639
Schauer, J. 2325
Schelt, J. van 2537
Schenkel, F. 340, 2932, 3004
Schenkel, F.S. 227, 458, 845, 1181, 2237, 2798, 2814, 2992
Schettini, G.P. 834
Schiavitto, M. 996
Schiavo, G. 972, 996, 3217, 3249
Schibler, L. 712
Schiefler, I. 3302
Schinckel, A.P. 528, 1575
Schmicke, M. 2786
Schmid, M. 3294
Schmidtmann, C. 1049, 1197
Schmied, J. 3282
Schmitt, E. 2545
Schmutz, M. 2475, 2479
Schnabel, R.D. 1555
Schneider, H. 1543
Schneiders, G.H. 2441
Schoenebeck, J.J. 2479, 3097
Schokker, D. 148, 1502, 2085, 3307
Scholtens, M.R. 2387
Scholtz, M.M 948, 2187
Scholtz, M.M. 1705, 1713, 1816
Schön, J. 964
Schoon, M. 882, 1021
Schopen, G.C.B. 280, 1181
Schrauf, M. 1498
Schrooten, C. 737
Schroyen, M. 2608
Schuberth, H.J. 2786
Schurink, A. 1897
Schwab, C. 1692
Schwab, C.R. 3266
Sdourlias, T. 3130
Sebek, L.B. 179
Secher, J.B. 2024
Sechi, T. 716, 1082
Seefried, F.R. 647, 1074, 1664
See, G.M. 3266
Segelke, D. 199, 586, 1164, 1197, 1543
Seichter, D. 2117, 2154, 2224
Seifert, J. 3294
Sell-Kubiak, E. 2134, 2337, 3253
Semmer, J. 3024
Senczuk, G. 968, 2818
Sener-Aydemir, A. 430
Seo, D. 1660, 2460
Sepulveda, B.J. 127
Serrano, M. 753, 1094, 2012, 2901, 3050
Serre, R.F. 2568
Servin, B. 861, 2568
Seshoka, M.M. 1816
Settar, P. 1143
Setyawan, P. 2379
Sevillano, C.A. 1189, 1876, 3225
Shaffer, W. 1701
Shah, M. Kumar 886
Shalloo, L. 131, 143, 2952
Sharifi, A.R. 2253, 2487
Sharif-Islam, M. 3208
Sharma, A. 2855
Sharma, P. 2879
Sherlock, R.G. 1360
Shihabi, M. 874, 3101
Shim, W.J. 2293
Shi, R. 203, 631
Shirsath, T. 1893
Shirzadifar, A. 1652, 1656
Shrestha, K. 2545
Shull, C. 1692
Shumbusho, F. 1741
Sibersen, C. 2191
Siberski-Cooper, C.J. 2806
Siegel, P.B. 923
Siegford, J. 557
Sifuentes-Rincón, A.M. 2305, 2605
Silva, B.P.M. da 2199
Silva, D.A. 890, 1254
Silva, D.A. da 2692
Silva, G.L.L.P. 1881, 1921
Silva, J.A.II.V. 308, 834
Silva, L.N. 2601
Silva, M.V.G.B. 960, 1697, 2182
Silva, T.L. 1598, 2692
Silva, V. 553
Silva, V.H. da 2233
Silva-Vignato, B. 2233
Silveira, N.F. 2741

Simčič, M. 822, 886, 2224
Simianer, H. 102, 1860, 2325, 2487
Simm, G. 2777, 2980
Simo, G. 1717
Simpson, A. 3130
Sinclair, R. 720
Sinecen, M. 1639
Singh, J. 2203
Siqueira, F. 2601
Sirard, M.A. 2269
Slagboom, M. 268, 1049, 1828, 1852
Slater, G.P. 2520
Slattery, T. 2387
Sleator, R.D. 1547, 1787, 2847
Smet, S. De 2620
Smith, C. 1326
Smith, E. 107, 3066
Smith, J. 1757, 2281, 3097
Snelling, W.M. 1966, 2593, 2624
Snell, R.G. 2216
Snyman, M.A. 3081
Sokolov, A. 2281
Soleimani, T. 187
Sölkner, J. 102, 822, 874, 886, 894, 910, 1745, 2826
Sol, M. 280
Solomon, B. 1721
Soma, P. 948, 2187
Sonesson, A.K. 284, 569, 2383
Son, M. van 3225
Sørensen, A.C. 1852, 3200, 3229, 3233
Sørensen, J.G. 2541
Sørensen, L.P. 976
Sourdioux, M. 2052, 2524
Souza, C.B. 2601
Souza, M.M. de 2061
Soyeurt, H. 156
Spangler, G.L. 1717
Spangler, M.L. 82, 434, 1966, 2007, 2593, 2624, 3266
Sparks, A.M. 720
Spehar, M. 845, 2968
Špehar, M. 2988
Spelman, R. 2790
Spengeler, M. 2170
Spoelstra, M. 1745
Spourita, M. 1086
Sprechert, M. 2495
Srihi, H. 308, 786, 1610, 3180
Srikanth, K. 1029
Stachowicz, K. 1782, 2891
Stalder, K.J. 3241
Stålhammar, H. 2802
Statham, J.M.E. 659
Stear, M. 643
Steensma, M. 2329
Steibel, J.P. 557, 1331
Steinheim, G. 107
Stella, A. 553, 598, 741, 798, 906, 3090
Stephansen, R.B. 304, 340, 2875
Stephen, C.P. 3046
Steyn, Y. 898, 1000
Stijn, T. van 168
Stock, J. 1872
Stock, K.F. 3163
Stock, K.F.S. 3155
Storlien, H. 1934
Stothard, P. 2237
Strabel, T. 172
Strachan, L. 2580
Strandberg, E. 102, 944, 2802, 2839
Strandén, I. 1298, 1314, 1372, 1498, 1506, 1515, 1534
Strugnell, B. 2178
Strugnell, J.M. 2416
Struthers, S. 2479
Suárez-Vega, A. 2909, 2956, 3020, 3054
Šubara, G. 919
Suchocki, T. 1590
Su, C.L. 2745
Su, G. 749, 1384, 1432, 2024, 3204, 3229, 3233, 3351
Su, H. 1824
Sullivan, B. 2992
Sullivan, P. 1364
Sullivan, T.M. 2855
Sumreddee, P. 952
Sun, Y.Y. 2491
Šuran, E. 919
Suthar, S. 2194
Sutherland, A. 886
Sutton, K. 1721
Sutton, K.M. 2142
Swalve, H.H. 639
Swaminathan, M. 1602, 1893, 1905, 2031
Swan, A.A. 1348, 1962, 3000, 3008, 3081
Swinkels, W. 2028
Symonds, J.E. 2387
Szwaczkowski, T. 2134
Szyda, J. 1590, 2126, 3319

T

Tabell, J. 2830
Tabet, K. 2568
Tabor, A. 2309
Tahir, M.S. 1209, 1282
Takahashi, H. 3176
Takeda, H. 2065, 2859
Takeda, M. 2688, 2696
Talenti, A. 810, 1737
Tammen, I. 757
Tang, L. 2065
Tan, W.L.A 1563
Tan, W.L.A. 1209
Tanya, V. 1717
Tarekegn, G.M. 2839, 3012
Tarsani, E. 1583, 2920
Taskinen, M. 1339, 1498, 1515, 1534
Tassell, C.P. Van 1029, 2913
Tatebayashi, R. 2859

Tate, M.L. 2387
Taurisano, V. 2516
Taussat, S. 426, 2648
Tavernari, F.C. 2448
Tavernier, E. 316
Taylor, J.B. 2996
Tearle, R. 2656
Teissier, M. 2932, 3004
Teixeira, C.S. 1598, 2692
Tempelman, R.J. 252, 348, 2814
Teng, J. 2321
Tenhunen, S. 976
ten Napel, J. 402, 2704
Tera, A. 1400
Tessema, E. 1400
Tessi, J.M. 1778
Tetens, J. 102, 1086, 1197, 1543, 1626, 2253, 3155, 3163
Thaller, G. 1197, 1543, 1626, 2117, 3155, 3163
Thallman, R.M. 1966
Thekkoot, D.M. 1262, 1999
Tholen, E. 1090, 3188, 3302
Thomasen, J. 268
Thomasen, J.R. 577, 976, 1049, 1148, 1828, 1852, 2802
Thomas, M.G. 1639
Thomsen, B. 1559
Thorsrud, J.A. 1643
Tiezzi, F. 191, 528, 806, 1575, 1692, 2081, 2101
Tigchelaar, J. 886
Tinarelli, S. 972
Ting, S.H. 2745
Tiplady, K.M. 2216
Tirumurugaan, K.G. 2031
Tixier-Boichard, M. 1021, 1037, 1045
Toffan, A. 2375
Toghiani, S. 826, 952
Toledo-Alvarado, H.O. 2737
Toledo-Alvarado, M.H.O. 72, 252
Tolhurst, D. 1408
Tomiyama, M. 3184
Tomley, F.M. 2421
Toral, P.G. 2909, 2956
Torgersen, J.S. 2371
Toro, M.Á. 841
Toro-Ospina, A.M. 308, 834
Torres Junior, R.A.A. 2601
Torres, O. 3315
Torshizi, R. Vaez 2467
Tortereau, F. 107, 244, 618, 861
Toscano, M.J. 635
Tosser-Klopp, G. 886
Toye, P. 1737
Trakooljul, N. 199, 2253
Trevarton, A. 2216
Tribout, T. 256, 260, 712, 1213, 1258, 2170
Trinh, T. Quoc 2403
Trujano-Chavez, M.Z. 914
Truong, B. 1352
Tsairidou, S. 728, 2887
Tsigenopoulos, C. 2375, 2383
Tsuruta, S. 380, 898, 1000, 1225, 1412, 1530, 1799, 2757, 2761, 2883
Tuggle, C.K. 2321
Tuliozi, B. 369, 1242, 3135
Tulpan, D. 227
Tu, P.A. 2745
Turkalj, K. 2826
Turner, S.P. 524
Tusell, L. 1680
Twomey, A.J. 1787, 2664

U

Ugarte, E. 119, 135, 753
Uimari, P. 980
Upadhyay, M. 2117, 2154
Ureña, I. 3050
Usai, M.G. 716, 1082, 2972, 3028
Utt, M.D. 2883
Utzeri, V.J. 2516, 3217
Uzunov, A. 2508

V

Vacherie, B. 2549
Valasek, H.F. 2007
Valdivieso, E. Mantilla 2309
Valente, B. 1490, 1634
Valente, B.D. 1135, 1270
Valente, J.P.S. 474
Valera, M. 3139
Vallée, R. 231, 256, 462
Vallejo, V.H. 834
van Arendonk, J.A.M. 86, 1791
Van Borm, S. 668
van Breukelen, A.E. 179
VandeHaar, M.J. 252, 348, 2814
van den Berg, I. 773
van den Heuvel, J. 849
Vandenplas, J. 1372, 1498, 2704
Vandeputte, M. 688
van der Beek, S. 280
Vanderick, S. 156, 312, 814
Vander Jagt, C.J. 2289
van der Linden, A. 203, 280, 631
van der Nest, M.A. 2187
van der Poel, J.J. 3355
van der Sluis, M. 393, 508, 537, 2040
van der Valk, E.G.T. 2085
van der Werf, J.H.J. 90, 684, 1110 , 1201, 1348, 1352, 1901, 1975, 2628, 2632, 2984, 3008, 3208, 3343
Van der Westhuizen, R.R. 1586
van den Boer, E. 2545
Van Eenennaam, A.L. 66, 82
Vanhala, T.K. 2191
Vanichkina, D.P. 757
van Kaam, J.B.C.H.M. 191, 450, 500, 2895, 3311
van Kessel, J. 3212
van Niekerk, M. 3323

van Pelt, M.L. 446
VanRaden, P.M. 806, 1356, 1463, 2761, 2773
van Son, M. 1189, 1876
van Staaveren, N. 458
Van Tassell, C.P. 777, 1717, 1774
Vanvanhossou, S.F.U. 454, 818
Vargas, G. 1118, 1598
Vargas Jurado, N. 1404
VarGoats Consortium 886
Vargovic, L. 3221
Varona, L. 308, 786, 1610, 3180
Vašek, J. 3093
Vásquez, H.V. 1114
Vásquez-Peláez, C. 2737
Vassilief, H. 861
Vázquez-Armijo, J.F. 2305
Vázquez-Gómez, M. 3180
Vedder, L. 659
Veenvliet, B. 168
Veerkamp, R. 1372, 1498, 1502, 1506
Veerkamp, R.F. 179, 280, 340, 446, 537, 1181, 2249, 2588, 2704
Veiga, F.C. da 2233
Velasco-Galilea, M. 320
Velazco, J. 160, 164, 195
Venkataramanan, R. 2031
Ventura, H.T. 1697
Vera, B. 164
Verardo, L.L. 960, 2182
Verma, A. 2729
Vermeulen, P.D. 2069
Verneque, R.S. 1697
Verschuren, L.M.G. 2085
Verstringe, S. 2085
Vervelde, L. 1721
Vetois, E. 2351
Vicario, D. 1392
Vignal, A. 2568
Vila, J. 3262
Vilkki, J. 2830
Villanueva, B. 728, 732, 841, 1181, 2383
Villumsen, T.M. 361, 388
Vilotte, M. 2097
Vinet, A. 231, 406, 426, 2648
Virta, A. 2830
Visscher, J. 1278, 3355
Visser, C. 870, 878
Vitezica, Z.G. 849, 1845
Vogelzang, R. 380
Vollmar, S. 2425
von Depka-Prondzinski, M. 3163
von Rohr, P. 1664
Vosgerau, S. 3155, 3163
Vos, J. de 2241, 2281
Vostra-Vydrova, H. 874, 894
Vostrá-Vydrová, H. 3093
Vostrý, L. 874, 894, 984, 3093
Vu, H. 2166

W

Wabbersen, J. 586
Waineina, R.W. 2138, 2917
Wakshlag, J.J. 3113
Wales, W.J. 127
Walker, L.R. 2166
Walker, S.P. 2387
Walkom, S.F. 1221, 2504, 3008
Wall, E. 223, 2777, 2944
Wallenbeck, A. 376
Wallen, S. 356
Waller, E. 211
Wallin, L.E. 183
Walmsley, B.J. 1938, 2003
Walsh, S. 2644
Walters, A.C.L. 2576
Walugembe, M. 1733, 1917, 2433
Wang, H. 276
Wang, M.Q. 2269
Wang, X. 802
Wang, Y. 203, 631, 1181, 1917, 2170, 2456
Wang, Z. 516
Warburton, C.L. 769
Warin, L. 708
Wasike, C.B. 336, 2917
Waters, D.L. 3343
Wathes, D.C. 2065
Watson, K. 410
Watson, K.A. 1583
Watson, M. 324, 2616
Watson, T. 168
Watt, K. 720
Weaber, R. 1701, 1966
Weaber, R.L. 82
Weepener, H.L. 1816
Wegener, J. 2508
Weigel, K. 2814
Weigel, K.A. 348
Weigend, A. 2325
Weigend, S. 1037, 1053, 2325, 2487
Weikard, R. 2786
Weik, F. 2716
Weisbecker, J.L 244
Weisbecker, J.L. 618
Weishaar, R. 3294
Wellmann, R. 1872, 1954
Weng, Z. 324
Wen, H. 528, 1575
Wen, J. 1160
Werling, D. 2301
Werner, A. 466
Wethal, K.B. 176
Whalen, A. 1135
White, C. 2855
White, H.M. 348
White, J.J. 82
Wicki, M. 3074

Widmer, S. 1664
Wiener, P. 810, 1717
Wientjes, Y.C.J. 761, 849
Wiggans, G.R. 1774
Wijebandara, K.G.C.B. 1881
Wijesena, H.R. 2142
Wijnrocx, K. 1318
Williams, M. 2847
Williams, R.O. 127
Wilmot, H. 156, 814, 865
Wilson, C.S. 1033
Wilson, P.W. 2475, 2479
Wimborne, E.A. 720
Wimmers, K. 199, 2114, 2253
Windig, J.J. 790, 882, 3109, 3125
Wing, J. 211
Winnberg, K. 176
Winters, M. 2887
Winters, M.W. 1229
Wit, A.A.C. de 2085
Wittenburg, D. 1606
Wobbe, M. 3155, 3163
Wockner, K. 1209
Woelders, H. 2016
Wolc, A. 1143, 1733
Wold, J.P. 569
Wolfe, C.W. 1029
Wolf, M.J. 1008, 1131, 1567
Woloszyn, F. 2924
Wongso, D. 2537
Wood, B.J. 2452, 3208
Woolley, S.A. 2317
Woolliams, J.A. 928, 3121
Worku, D. 2729
Worth, G. 2790
Wragg, D. 438, 1737, 2552
Wurzinger, M. 1913
Wu, X. 1559
Wu, Y. 802
Wyk, J.B. van 3323

X

Xia, D. 2301
Xiang, R. 773, 2228, 2289
Xiang, T. 1152
Xue, Y. 590
Xu, H. 2471

Y

Yamaguchi, S. 2867
Yáñez, J.M. 2371
Yates, J. 107, 3066
Yeates, A. 2790
Yengo, L. 61
Yin, H. 2321
Yin, T. 454, 818
Young, E.A. 211
Yuan, C. 2065
Yudin, N. 1078
Yu, H. 1290, 1688, 1824
Yu, M. 2460
Yurchenko, A. 1078
Yu, X. 2403
Yu, X-J. 1442

Z

Zaalberg, R.M. 361, 388
Zak, L. 2016
Żarnecki, A. 1590
Zavadilová, L. 414, 504
Zebeli, Q. 430
Zemb, O. 2101
Zenger, K.R. 2416
Zeng, J. 1169
Zerbe, H. 2786
Zerlotini, A. 2061, 2293, 2612
Zetouni, L. 148
Zhang, C. 248, 1262
Zhang, D. 2471
Zhang, H. 886
Zhang, S. 2471
Zhang, W. 3097
Zhang, W.-Y. 886
Zhang, X. 1946, 2471
Zhang, Y. 2036
Zhang, Z. 2321
Zhao, G.P. 1160
Zhao, T. 1169, 1519, 1668
Zhao, X. 2269
Zheng, M.Q. 1160
Zheng, Z. 886
Zhou, H. 1668, 1733, 1917, 2433, 2456
Zhu, W. 2471
Zingaretti, L.M. 622, 1173
Zira, S. 139
Zishiri, O.T. 300, 2464
Zorc, M. 822, 3101
Zubiri-Gaitán, A. 2073
Zullo, G. 3335
Zumbach, B.J. 1266
Zwaan, B.J. 849, 2537
Zwane, A.A. 2151